Phylogenomics of the adaptive radiation of *Triturus* newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle

B. Wielstra[a,b,c,d,1,*], E. McCartney-Melstad[a,e,f,1], J.W. Arntzen[c], R.K. Butlin[b,g], H.B. Shaffer[a,e,f]

[a] *Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA.* [b] *Department of Animal and Plant Sciences, University of Sheffield, S10 2TN Sheffield, UK.* [c] *Naturalis Biodiversity Center, 2300 RA Leiden, The Netherlands.* [d] *Institute of Biology, Leiden University, 2300 RA, Leiden, The Netherlands.* [e] *La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA.* [f] *Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA.* [g] *Department of Marine Sciences, University of Gothenburg, Gothenburg 405 30, Sweden.* [1] *These authors contributed equally to this work.* * *Corresponding author; e-mail: ben.wielstra@naturalis.nl*

Abstract

Understanding the course of eco-morphological evolution in adaptive radiations is challenging as the phylogenetic relationships among the species involved are typically difficult to resolve. Newts of the genus *Triturus* (marbled and crested newts) are a well-studied case: they exhibit substantial variation in the number of trunk vertebrae (NTV) and a higher NTV corresponds to a longer annual aquatic period. Because the *Triturus* phylogeny is still unresolved, the evolutionary pathway for NTV and annual aquatic period is unclear. To resolve the phylogeny of *Triturus*, we generate a c. 6,000 transcriptome-derived marker data set using a custom target enrichment probe set, and conduct phylogenetic analyses including: 1) data concatenation with RAxML, 2) gene tree summary with ASTRAL, and 3) species tree estimation with SNAPP. All analyses consistently result in the same, highly supported topology. Our new phylogenetic hypothesis only requires the minimal number of inferred changes in NTV count to explain the NTV radiation observed today. This suggests that, while diversification in body shape allowed ecological expansion in *Triturus* to encompass an increasingly aquatic life style, body shape evolution was phylogenetically constrained.

Keywords: morphology; phylogeny; sequence capture; systematics; target enrichment; transcriptome

1. Background

In adaptive radiations, reproductive isolating barriers between nascent species evolve in response to rapid ecological specialization [1] and that ecological speciation typically correlates with pronounced morphological differentiation [2, 3]. Adaptive radiations are known throughout the Tree of Life and illustrate the power of natural selection to drive speciation [4]. While adaptive radiations represent some of the best-known examples of evolution in action – most famously Darwin's finches [5] and Lake Victoria cichlid fishes [6] – the phylogenetic relationships between the species involved are notoriously difficult to decipher [7-9]. Yet, to accurately retrace the evolution of phenotypic diversity in adaptive radiations, requires well-established phylogenies.

Inferring the true branching order in adaptive radiations is hampered by the short time frame over which they unfold, which provides little opportunity between splitting events for phylogenetically informative substitutions to become established (resulting in low phylogenetic resolution [10, 11]) and fixed (resulting in incomplete lineage sorting and discordance among gene trees [12-14]). Resolving the phylogeny of rapidly multiplying lineages becomes more complicated the further back in time the radiation occurred, as the accumulation of uninformative substitutions along terminal branches leads to long-branch attraction [15, 16]. A final impediment is reticulation between closely related (not necessarily sister-) species through past or ongoing hybridization, resulting in additional gene-tree/species-tree discordance [17-19].

Phylogenomics can help. Consulting a large number of markers spread throughout the genome has proven successful in resolving both recent (e.g. [20-26]) and ancient (e.g. [27-31]) evolutionary radiations. Advances in laboratory and sequencing techniques, bioinformatics and tree-building methods, facilitate phylogenetic reconstruction based on thousands of homologous loci for a large number of individuals, and promise to help reveal the evolution of eco-morphological characters involved in adaptive radiations [32, 33]. Here we conduct a phylogenomic analysis for an adaptive radiation that moderately-sized multilocus nuclear DNA datasets [34-36] have consistently failed to resolve: the Eurasian newt genus *Triturus* (Amphibia: Urodela: Salamandridae; vernacularly known as the marbled and crested newts).

One of the most intriguing features of *Triturus* evolution is the correlation between ecology and the number of trunk vertebrae (NTV). Species characterized by a higher modal NTV (which translates into a more elongate body build with shorter limbs) are associated with a more aquatic lifestyle [37-44]; the number of months a *Triturus* species spends in the water (defined at the population level as the peak in emigration minus the peak in immigration) roughly equals NTV minus 10 (Fig. 1). The intrageneric variation in NTV shown by *Triturus*, ranging from 12 to 17, is unparalleled in the family Salamandridae [44, 45]. Although a causal relationship between NTV expansion and an increasingly aquatic lifestyle has been presumed [37-44], the evolutionary pathway of NTV and aquaticness in the adaptive radiation of *Triturus* is unclear. A resolved species tree is required to address this issue.

Our goal is to obtain a genome-enabled phylogeny for *Triturus* use it to reconstruct the eco-morphological evolution of NTV and aquatic/terrestrial ecology across the genus. As the large size of salamander genomes hampers whole genome sequencing (but see [46-48]), we employ a genome reduction approach in which we capture and sequence a set of transcriptome-derived markers using target enrichment, a technique which affords extremely high resolution at multiple taxonomic levels [49-54]. Using data concatenation (with RAxML), gene tree summarization (with ASTRAL) and species tree estimation (with SNAPP), we fully resolve the *Triturus* phylogeny and place the extreme body shape and ecological variation observed in this adaptive radiation into an evolutionary context.

2. Methods

Target capture array design

Nine *Triturus* newts (seven crested and two marbled newt species) and one banded newt (*Ommatotriton*) were subjected to transcriptome sequencing. Transcriptome assemblies for each species were generated using Trinity v2.2.0 [55], clustered at 90% using usearch v9.1.13 [56], and subjected to reciprocal best blast hit analysis [57-59] to produce a set of *T. dobrogicus* transcripts (the species with the highest quality transcriptome assembly) that had putative orthologues present in the nine other transcriptome assemblies. These transcripts were then annotated using blastx to *Xenopus tropicalis* proteins, retaining one annotated transcript per protein. We attempted to discern splice sites in the transcripts, as probes

spanning splice boundaries may perform poorly [60], by mapping transcripts iteratively to the genomes of *Chrysemys picta* [61], *X. tropicalis* [62], *Nanorana parkerii* [63] and *Rana catesbeiana* [64]. A single exon ≥ 200bp and ≤ 450bp was retained for each transcript target. To increase the utility of the target set to all *Triturus* species, orthologous sequences from multiple species were included for targets with > 5% sequence divergence from *T. dobrogicus* [49]. We generated a target set of 7,102 genomic regions for a total target length of approximately 2.3 million bp. A total of 39,143 unique RNA probes were synthesized as a MyBaits-II kit for this target set at approximately 2.6X tiling density by Arbor Biosciences (Ann Arbor, MI, Ref# 170210-32). A detailed outline of the target capture array design process is presented in Supplementary Text S1.

Sampling scheme

We sampled 23 individual *Triturus* newts (Fig. 2; Supplementary Table S1) for which tissues were available from previous studies [65-67]. Because the sister relationship between the two marbled and seven crested newts is well established, while the relationships among the crested newt species are unclear, we sampled the crested newt species more densely, including three individuals per species to capture intraspecific differentiation and avoid misleading phylogenies resulting from single exemplar sampling [68] (Fig. 1). As *Triturus* species show introgressive hybridization at contact zones [69], we aimed to reduce the impact of interspecific gene flow by only including individuals that originate away from hybrid zones and have previously been interpreted as unaffected by interspecific genetic admixture [65, 66]. A test for the phylogenetic utility of the transcripts used for marker design underscores the reality of phylogenetic distortion by interspecific gene flow (details in Supplementary Text S1).

Laboratory methods

DNA was extracted from samples using a salt extraction protocol [70], and 10,000ng per sample was sheared to approximately 200bp-500bp on a BioRuptor NGS (Diagenode) and dual-end size selected (0.8X-1.0X) with SPRI beads. Dual-indexed libraries were prepared from 375-2000ng of size selected DNA using KAPA LTP library prep kits [71]. These libraries were pooled (with samples from other projects) into batches of 16 samples at 250ng per sample (4,000ng total) and enriched in the presence

of 30,000ng of c0t-1 repetitive sequence blocker [50] derived from *T. carnifex* (casualties from a removal action of an invasive population [72]) by hybridizing blockers with libraries for 30 minutes and probes with libraries/blockers for 30 hours. Enriched libraries were subjected to 14 cycles of PCR with KAPA HiFi HotStart ReadyMix and pooled at an equimolar ratio for 150bp paired-end sequencing across multiple Illumina HiSeq 4000 lanes (receiving an aggregate of 18% of one lane, for a multiplexing equivalent of 128 samples per lane).

Processing of target capture data

Sequences from the sample receiving the greatest number of reads were used to *de novo* assemble target sequences for each target region using the assembly by reduced complexity (ARC) pipeline [73]. A single assembled contig was selected for each original target region by means of reciprocal best blast hit (RBBH) [74] and these were used as a reference assembly for all downstream analyses. Adapter contamination was removed from sample reads using skewer v0.2.2 [75] and reads were then mapped to the reference assembly using BWA-MEM v0.7.15-r1140 [76]. Picard tools v2.9.2 (https://broadinstitute.github.io/picard/) was used to add read group information and mark PCR duplicates, and HaplotypeCaller and GenotypeGVCFs from GATK v3.8 [77] were used to jointly genotype the relevant groups of samples (either crested newts or crested newts + marbled newts depending on the analysis; see below). SNPs that failed any of the following hard filters were removed: QD < 2, MQ < 40, FS > 60, MQRankSum < -12.5, ReadPosRankSum < -8, and QUAL < 30 [78]. We next attempted to remove paralogous targets from our dataset with a Hardy Weinberg Equilibrium (HWE) filter for heterozygote excess. Heterozygote excess p-values were calculated for every SNP using vcftools 0.1.15 [79], and any target containing at least one SNP with a heterozygote excess p-value < 0.05 was removed from downstream analysis. More detail on the processing of the target capture data can be found in Supplementary Text S2.

Phylogenetic analyses

For data concatenation, a maximum likelihood phylogeny was inferred with RAxML version 8.2.11 [80] based on an alignment of 133,601 SNPs across 5,866 different targets. We included all 23 *Triturus*

individuals in this analysis. For gene tree summary, ASTRAL v5.6.1 [81] was used to estimate the crested newt species tree from 5,610 gene trees generated in RAxML. The 21 crested newt samples were assigned species membership and no marbled newts were included because estimating terminal branch lengths is not possible for species with a single representative. For species tree estimation, SNAPP v1.3.0 [82] within the BEAST v2.4.8 [83] environment was used to infer the crested newt species tree from biallelic SNPs randomly selected from each of 5,581 post-filtering targets. All three individuals per crested newt species were treated as a single terminal and marbled newts were again excluded because sampling one individual per species violates the Yule speciation prior assumption. We also estimated divergence times in SNAPP. A detailed description of our strategy for phylogenetic analyses is available in Supplementary Text S3.

3. Results

The concatenated analysis with RAxML supports a basal bifurcation in *Triturus* between the marbled and crested newts (Fig. 3), consistent with the prevailing view that they are reciprocally monophyletic [34-36]. RAxML recovers each of the crested newt species as monophyletic, validating our decision to collapse the three individuals sampled per species in a single terminal in ASTRAL and SNAPP. Furthermore, all five *Triturus* body builds are recovered as monophyletic (cf. [34-36]). The greatest intraspecific divergence is observed in *T. carnifex* (Supplementary Text S1; Supplementary Fig. S1; Supplementary Table S2).

Phylogenetic inference based on data concatenation with RAxML (Fig. 3), gene tree summary with ASTRAL (Fig. 4a) and species tree estimation with SNAPP (Fig. 4b) all recover the same crested newt topology, with a basal bifurcation between the *T. karelinii*-group (NTV = 13; *T. ivanbureschi* sister to *T. anatolicus* + *T. karelinii)* and the remaining taxa, which themselves are resolved into the species pairs *T. carnifex* + *T. macedonicus* (NTV=14; the *T. carnifex*-group), and *T. cristatus* (NTV=15) + *T. dobrogicus* (NTV=16/17). In addition, the bifurcation giving rise to the four crested newt species groups (cf. Fig. 1) occurred in a relatively short time frame (Supplementary Fig. S2), reflected by two particularly short, but resolvable internal branches (Fig. 3; Fig. 4).

The phylogenomic analyses suggest considerable gene tree/species tree discordance in *Triturus*. The normalized quartet score of the ASTRAL tree (Fig. 4a), which reflects the proportion of input gene tree quartets satisfied by the species tree, is 0.63, indicating a high degree of incomplete lineage sorting. Furthermore, the only node in the SNAPP tree with a posterior probability below 1.0 (i.e. 0.99) is subtended by a very short branch (Fig. 4b). We also observe highly supported topological incongruence with the full mtDNA-based phylogeny of *Triturus* (Supplementary Text S4; Supplementary Fig. S3) [38].

Considering an NTV count of 12, as observed in the marbled newts as well as the most closely related newt genera, as the ancestral state for *Triturus* [44, 84], three sequential single-vertebral additions to NTV along internal branches, and one or two additions along the terminal branch leading to *T. dobrogicus* (in which NTV = 16 and NTV = 17 occur at approximately equal frequency [44, 85]), are required to explain the present-day variation in NTV observed in *Triturus* (Fig. 3). This is the minimum possible number of inferred changes in NTV count required to explain the NTV radiation observed today (Supplementary Fig. S4). No NTV deletions or reversals have to be inferred, implying a linear, single-addition progression rule for vertebral addition in *Triturus.*

4. Discussion

We use phylogenomic data to study the evolution of ecological and phenotypic diversity within the adaptive radiation of *Triturus* newts. In contrast to previous attempts to recover a multilocus species tree [34-36], we recover full phylogenetic resolution with strong support. Despite a high degree of gene tree/species tree discordance, independent phylogenetic approaches based on data concatenation (RAxML), gene tree summarization (ASTRAL) and species tree estimation (SNAPP), all recover the same, highly supported topology for *Triturus* (Fig. 3; Fig. 4). The *Triturus* case study underscores that sequence capture by target enrichment is a promising approach to resolve the phylogenetic challenges associated with adaptive radiations, particularly for taxa with large and complicated genomes where other genomic approaches are impractical, including salamanders [50].

Our new phylogenetic hypothesis allows us to place the eco-morphological differentiation shown by *Triturus* into a coherent evolutionary context. Over time, *Triturus* expanded its range of NTV

to encompass higher counts (Fig. 3). The *Triturus* tree is consistent with a maximally parsimonious scenario, under which four to five character state changes are required to explain the radiation in NTV observed today. Any other possible topology would necessitate a higher number of NTV changes to be inferred (Supplementary Fig. S4). Three of these inferred changes are positioned on internal branches, of which two are particularly short, suggesting that changes in NTV count can evolve in a relatively short time. The fourth and fifth inferred change are situated on the external branch leading to *T. dobrogicus*, the only *Triturus* species with substantial intraspecific variation in NTV count [44, 85].

Newts annually alternate between an aquatic and a terrestrial habitat and the functional trade-off between adaptation to life in water or on land likely poses contrasting demands on body build [86-89]. Assuming that the observed relationship between one additional trunk vertebra and an extra month annually spent in the water (Fig. 1) is causal, then the NTV flexibility expressed by *Triturus* suggests enhanced ecological opportunities by an ability to exploit a wider range in hydroperiod (i.e. the annual availability of standing water) more efficiently. Despite the evolvability of NTV count [44], NTV evolution has been phylogenetically constrained in *Triturus*; apparently the change in NTV count was directional and involved the addition of one trunk vertebra at a time (Fig. 3; Supplementary Fig. S4). Species with a more derived body build have a relatively prolonged aquatic period and, because species with transitional NTV counts remain extant, the end result is an eco-morphological radiation.

Triturus newts show a certain degree of intraspecific variation in NTV today. Such variation is partially explained by interspecific hybridization (emphasizing the genetic basis of NTV count) [69], but there is standing variation in NTV count within all *Triturus* species [42]. This suggests that, during *Triturus* evolution, there has always been intraspecific NTV count polymorphism for natural selection to work with. Whether the directional, parsimonious evolution of higher NTV and the equally parsimonious evolutionary increase in aquatic lifestyle is causal, and which of these two may be the actual target of selection, remain important open questions. A proper understanding of the functional relationship between body build and aquaticness in *Triturus* is still lacking [86]. The recent availability of the first salamander genomes [46-48] offers the prospect of sequencing the genome of each *Triturus* species and exploring the developmental basis for NTV and its functional consequences in the diversification of the genus.

Ethics. For sampling for transcriptome sequencing permits were provided by the Italian Ministry of the Environment (DPN-2009-0026530), the Environment Protection Agency of Montenegro (no. UPI-328/4), the Ministry of Energy, Development and Environmental Protection of Republic of Serbia (no. 353-01-75/2014-08), and TÜBİTAK, Turkey (no. 113Z752). RAVON & Natuurbalans-Limes Divergens provided the *T. carnifex* used to create c0t-1.

Data availability. Raw sequence read data for the sequence capture libraries of the 23 *Triturus* samples and the transcriptome libraries are available at SRA (PRJNA498336). Transcriptome assemblies and genotype calls (VCF) for the 21- and 23-sample datasets are available at Zenodo (https://doi.org/10.5281/zenodo.1470914).

Author contributions. BW, EMM, JWA, RKB, HBS designed the research; BW and EMM performed the research; BW and EMM wrote the paper with input from JWA, RKB and HBS. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 655487.

Acknowledgements. Andrea Chiocchio, Daniele Canestrelli, Michael Fahrbach, Ana Ivanović, Raymond van der Lans, and Kurtuluş Olgun helped obtain samples for transcriptome sequencing. Tara Luckau helped in the lab. Peter Scott provided valuable suggestions on methodology. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. Computing resources were provided by XSEDE [90] and the Texas Advanced Computing Center (TACC) Stampede2 cluster at The University of Texas at Austin.

References

[1] Schluter, D. 2000 *The ecology of adaptive radiation*, OUP Oxford.

[2] Gavrilets, S. & Losos, J.B. 2009 Adaptive radiation: contrasting theory with data. *Science* **323**, 732. (doi:10.1126/science.1157966).

[3] Glor, R.E. 2010 Phylogenetic insights on adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* **41**, 251-270. (doi:10.1146/annurev.ecolsys.39.110707.173447).

[4] Losos, Jonathan B. 2010 Adaptive radiation, ecological opportunity, and evolutionary determinism. *Am. Nat.* **175**, 623-639. (doi:10.1086/652433).

[5] Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., et al. 2015 Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371. (doi:10.1038/nature14181).

[6] Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., et al. 2014 The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375. (doi:10.1038/nature13726).

[7] Rokas, A. & Carroll, S.B. 2006 Bushes in the Tree of Life. *PLoS Biol.* **4**, e352. (doi:10.1371/journal.pbio.0040352).

[8] Patel, S., Kimball, R.T. & Braun, E.L. 2013 Error in phylogenetic estimation for bushes in the tree of life. *Journal of Phylogenetics & Evolutionary Biology* **1**, 110. (doi:10.4172/2329-9002.1000110).

[9] Shaffer, H.B. & Thomson, R.C. 2007 Delimiting species in recent radiations. *Syst. Biol.* **56**, 896-906. (doi:10.1080/10635150701772563).

[10] Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G. & Baurain, D. 2011 Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* **9**, e1000602. (doi:10.1371/journal.pbio.1000602).

[11] Whitfield, J.B. & Lockhart, P.J. 2007 Deciphering ancient rapid radiations. *Trends Ecol. Evol.* **22**, 258-265. (doi:10.1016/j.tree.2007.01.012).

[12] Pamilo, P. & Nei, M. 1988 Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568-583. (doi:10.1093/oxfordjournals.molbev.a040517).

[13] 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. (doi:10.1371/journal.pgen.0020173).

[14] Degnan, J.H. & Rosenberg, N.A. 2006 Discordance of species trees with their most likely gene trees. *PLoS Genet.* **2**, e68. (doi:10.1371/journal.pgen.0020068).

[15] Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O. & Rogers, J.S. 2001 Bias in phylogenetic estimation and Its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* **50**, 525-539. (doi:10.1080/10635150117959).

[16] Felsenstein, J. 1978 Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Biol.* **27**, 401-410. (doi:10.1093/sysbio/27.4.401).

[17] Mallet, J., Besansky, N. & Hahn, M.W. 2016 How reticulated are species? *Bioessays* **38**, 140-149. (doi:10.1002/bies.201500149).
[18] Leaché, A.D., Harris, R.B., Rannala, B. & Yang, Z. 2014 The influence of gene flow on species tree estimation: a simulation study. *Syst. Biol.* **63**, 17-30. (doi:10.1093/sysbio/syt049).
[19] Kutschera, V.E., Bidon, T., Hailer, F., Rodi, J.L., Fain, S.R. & Janke, A. 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. (doi:10.1093/molbev/msu186).
[20] Scott, P.A., Glenn, T.C. & Rissler, L.J. 2018 Resolving taxonomic turbulence and uncovering cryptic diversity in the musk turtles (*Sternotherus*) using robust demographic modeling. *Mol. Phylogenet. Evol.* **120**, 1-15. (doi:10.1016/j.ympev.2017.11.008).
[21] Giarla, T.C. & Esselstyn, J.A. 2015 The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Syst. Biol.* **64**, 727-740. (doi:10.1093/sysbio/syv029).
[22] Shi, C.-M. & Yang, Z. 2018 Coalescent-based analyses of genomic sequence data provide a robust resolution of phylogenetic relationships among major groups of gibbons. *Mol. Biol. Evol.* **35**, 159-179. (doi:10.1093/molbev/msx277).
[23] Nater, A., Burri, R., Kawakami, T., Smeds, L. & Ellegren, H. 2015 Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Syst. Biol.* **64**, 1000-1017. (doi:10.1093/sysbio/syv045).
[24] Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T. & Braun, E.L. 2016 Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. *Syst. Biol.* **65**, 612-627. (doi:10.1093/sysbio/syw014).
[25] Léveillé-Bourret, É., Starr, J.R., Ford, B.A., Moriarty Lemmon, E. & Lemmon, A.R. 2018 Resolving rapid radiations within angiosperm families using anchored phylogenomics. *Syst. Biol.* **67**, 94-112. (doi:10.1093/sysbio/syx050).
[26] Leaché, A.D., Banbury, B.L., Linkem, C.W. & de Oca, A.N.-M. 2016 Phylogenomics of a rapid radiation: is chromosomal evolution linked to increased diversification in north american spiny lizards (Genus Sceloporus)? *BMC Evol. Biol.* **16**, 63. (doi:10.1186/s12862-016-0628-x).
[27] Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K. & Glenn, T.C. 2012 More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol. Lett.* **8**, 783-786. (doi:10.1098/rsbl.2012.0331).
[28] Jarvis, E.D. & Mirarab, S. & Aberer, A.J. & Li, B. & Houde, P. & Li, C. & Ho, S.Y.W. & Faircloth, B.C. & Nabholz, B. & Howard, J.T., et al. 2014 Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320. (doi:10.1126/science.1253451).
[29] McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T. & Glenn, T.C. 2012 Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* **22**, 746-754. (doi:10.1101/gr.125864.111).
[30] Irisarri, I. & Meyer, A. 2016 The identification of the closest living relative(s) of tetrapods: phylogenomic lessons for resolving short ancient internodes. *Syst. Biol.* **65**, 1057-1075. (doi:10.1093/sysbio/syw057).
[31] Song, S., Liu, L., Edwards, S.V. & Wu, S. 2012 Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14942-14947. (doi:10.1073/pnas.1211733109).
[32] Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C.B., Glor, R.E., Jaffe, J.D., et al. 2011 The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587. (doi:10.1038/nature10390).
[33] Stroud, J.T. & Losos, J.B. 2016 Ecological opportunity and adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* **47**, 507-532. (doi:10.1146/annurev-ecolsys-121415-032254).
[34] Arntzen, J.W., 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.
[35] Espregueira Themudo, G., Wielstra, B. & Arntzen, J.W. 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. (doi:10.1016/j.ympev.2009.03.024).
[36] Wielstra, B., Arntzen, J.W., van der Gaag, K., 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**, e111011. (doi:10.1371/journal.pone.0111011).
[37] Arntzen, J.W. 2003 *Triturus cristatus* Superspecies - Kammolch-Artenkreis (*Triturus cristatus* (Laurenti, 1768) - Nördlicher Kammolch, *Triturus carnifex* (Laurenti, 1768) - Italienischer Kammolch, *Triturus dobrogicus* (Kiritzescu, 1903) - Donau-Kammolch, *Triturus karelinii* (Strauch, 1870) - Südlicher Kammolch). In *Handbuch der Reptilien und Amphibien Europas. Schwanzlurche IIA* (eds. K. Grossenbacher & B. Thiesmeier), pp. 421-514. Wiebelsheim, Aula-Verlag.
[38] Wielstra, B. & Arntzen, J.W. 2011 Unraveling the rapid radiation of crested newts (*Triturus cristatus* superspecies) using complete mitogenomic sequences. *BMC Evol. Biol.* **11**, 162. (doi:10.1186/1471-2148-11-162).
[39] Arntzen, J.W. & Wallis, G.P. 1999 Geographic variation and taxonomy of crested newts (*Triturus cristatus* superspecies): morphological and mitochondrial data. *Contrib. Zool.* **68**, 181-203.
[40] Urošević, A., Slijepčević, M.D., Arntzen, J.W. & Ivanović, A. 2016 Vertebral shape and body elongation in *Triturus* newts. *Zoology* **119**, 439-446. (doi:10.1016/j.zool.2016.05.003).
[41] Govedarica, P., Cvijanović, M., Slijepčević, M. & Ivanović, A. 2017 Trunk elongation and ontogenetic changes in the axial skeleton of *Triturus* newts. *J. Morphol.* **278**, 1577-1585. (doi:10.1002/jmor.20733).
[42] Slijepčević, M., Galis, F., Arntzen, J.W. & Ivanović, A. 2015 Homeotic transformations and number changes in the vertebral column of *Triturus* newts. *PeerJ* **3**, e1397. (doi:10.7717/peerj.1397).

[43] Vukov, T.D., Sotiropoulos, K., Wielstra, B., Džukić, G. & Kalezić, M.L. 2011 The evolution of the adult body form of the crested newt (*Triturus cristatus* superspecies, Caudata, Salamandridae). *Journal of Zoological Systematics and Evolutionary Research* **49**, 324-334. (doi:10.1111/j.1439-0469.2011.00633.x).
[44] Arntzen, J.W., Beukema, W., Galis, F. & Ivanović, A. 2015 Vertebral number is highly evolvable in salamanders and newts (family Salamandridae) and variably associated with climatic parameters. *Contrib. Zool.* **84**, 85-113.
[45] Lanza, B., Arntzen, J.W. & Gentile, E. 2010 Vertebral numbers in the Caudata of the Western Palearctic (Amphibia). *Atti Mus. Civ. Stor. Nat. Trieste* **54**, 3-114.
[46] Elewa, A., Wang, H., Talavera-López, C., Joven, A., Brito, G., Kumar, A., Hameed, L.S., Penrad-Mobayed, M., Yao, Z., Zamani, N., et al. 2017 Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. *Nat. Commun.* **8**, 2286. (doi:10.1038/s41467-017-01964-9).
[47] Nowoshilow, S., Schloissnig, S., Fei, J.-F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. 2018 The axolotl genome and the evolution of key tissue formation regulators. *Nature* **554**, 50-55. (doi:10.1038/nature25458).
[48] Smith, J.J., Timoshevskaya, N., Timoshevskiy, V.A., Keinath, M.C., Hardy, D. & Voss, S.R. 2018 A chromosome-scale assembly of the enormous (32 Gb) *Axolotl* genome. *bioRxiv*. (doi:10.1101/373548).
[49] Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C. & Good, J.M. 2012 Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics* **13**, 403. (doi:10.1186/1471-2164-13-403).
[50] McCartney-Melstad, E., Mount, G.G. & Bradley Shaffer, H. 2016 Exon capture optimization in amphibians with large genomes. *Mol. Ecol. Resour.* **16**, 1084-1094. (doi:10.1111/1755-0998.12538).
[51] Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., et al. 2009 Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182. (doi:10.1038/nbt.1523).
[52] McCartney-Melstad, E., Vu, J.K. & Shaffer, H.B. 2018 Genomic data recover previously undetectable fragmentation effects in an endangered amphibian. *Mol. Ecol.*, TBA. (doi:10.1101/306340).
[53] Abdelkrim, J., Aznar-Cormano, L., Fedosov, A., Kantor, Y., Lozouet, P., Phuong, M., Zaharias, P. & Puillandre, N. 2018 Exon-capture based phylogeny and diversification of the venomous gastropods (Neogastropoda, Conoidea). *Mol. Biol. Evol.* **35**, 2355–2374. (doi:10.1093/molbev/msy144).
[54] Bragg, J.G., Potter, S., Bi, K. & Moritz, C. 2016 Exon capture phylogenomics: efficacy across scales of divergence. *Mol. Ecol. Resour.* **16**, 1059-1068. (doi:10.1111/1755-0998.12449).
[55] Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652. (doi:10.1038/nbt.1883).
[56] Edgar, R.C. 2010 Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461. (doi:10.1093/bioinformatics/btq461).
[57] Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. & Yuan, Y. 1998 Predicting function: from genes to genomes and back. *J. Mol. Biol.* **283**, 707-725. (doi:10.1006/jmbi.1998.2144).
[58] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. 2009 BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421-421. (doi:10.1186/1471-2105-10-421).
[59] Tatusov, R.L., Koonin, E.V. & Lipman, D.J. 1997 A genomic perspective on protein families. *Science* **278**, 631. (doi:10.1126/science.278.5338.631).
[60] Neves, L.G., Davis, J.M., Barbazuk, W.B. & Kirst, M. 2013 Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* **75**, 146-156. (doi:10.1111/tpj.12193).
[61] Shaffer, H.B., Minx, P., Warren, D.E., Shedlock, A.M., Thomson, R.C., Valenzuela, N., Abramyan, J., Amemiya, C.T., Badenhorst, D., Biggar, K.K., et al. 2013 The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28. (doi:10.1186/gb-2013-14-3-r28).
[62] Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., et al. 2010 The genome of the western clawed frog *Xenopus tropicalis*. *Science* **328**, 633. (doi:10.1126/science.1183670).
[63] Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., Zhong, L., Wang, L., Wu, D.-D., Zhang, B.-L., et al. 2015 Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1257. (doi:10.1073/pnas.1501764112).
[64] Hammond, S.A., Warren, R.L., Vandervalk, B.P., Kucuk, E., Khan, H., Gibb, E.A., Pandoh, P., Kirk, H., Zhao, Y., Jones, M., et al. 2017 The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. *Nat. Commun.* **8**, 1433. (doi:10.1038/s41467-017-01316-7).
[65] Wielstra, B., Burke, T., Butlin, R.K., Avcı, A., Üzüm, N., Bozkurt, E., Olgun, K. & Arntzen, J.W. 2017 A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters* **1**, 93-101. (doi:10.1002/evl3.9).
[66] Wielstra, B., Burke, T., Butlin, R.K. & Arntzen, J.W. 2017 A signature of dynamic biogeography: enclaves indicate past species replacement. *Proc. Royal Soc. B* **284**, 20172014. (doi:10.1098/rspb.2017.2014).
[67] Wielstra, B., Crnobrnja-Isailović, J., Litvinchuk, S.N., Reijnen, B.T., Skidmore, A.K., Sotiropoulis, K., Toxopeus, A.G., Tzankov, N., Vukov, T. & Arntzen, J.W. 2013 Tracing glacial refugia of *Triturus* newts based on mitochondrial DNA phylogeography and species distribution modeling. *Front. Zool.* **10**, 13. (doi:10.1186/1742-9994-10-13).
[68] Spinks, P.Q., Thomson, R.C., Pauly, G.B., Newman, C.E., Mount, G. & Shaffer, H.B. 2013 Misleading phylogenetic inferences based on single-exemplar sampling in the turtle genus *Pseudemys*. *Mol. Phylogenet. Evol.* **68**, 269-281. (doi:10.1016/j.ympev.2013.03.031).
[69] Arntzen, J.W., Wielstra, B. & Wallis, G.P. 2014 The modality of nine *Triturus* newt hybrid zones, assessed with nuclear, mitochondrial and morphological data. *Biol. J. Linn. Soc.* **113**, 604–622. (doi:10.1111/bij.12358).

[70] Sambrook, J. & Russell, D.W. 2001 Molecular cloning: a laboratory manual 3rd edition. *Coldspring-Harbour Laboratory Press, UK*.

[71] Glenn, T.C., Bayona-Vásquez, N.J., Kieran, T.J., Pierson, T.W., Hoffberg, S.L., Scott, P.A., Louha, S., Bentley, K.E., Finger Jr., J.W., Troendle, N., et al. 2017 Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about $1USD per Sample (3RAD). *BioRxiv*. (doi:10.1101/205799).

[72] Meilink, W.R.M., Arntzen, J.W., van Delft, J.J.C.W. & Wielstra, B. 2015 Genetic pollution of a threatened native crested newt species through hybridization with an invasive congener in the Netherlands. *Biol. Conserv.* **184**, 145-153. (doi:10.1016/j.biocon.2015.01.022).

[73] Hunter, S.S., Lyon, R.T., Sarver, B.A.J., Hardwick, K., Forney, L.J. & Settles, M.L. 2015 Assembly by Reduced Complexity (ARC): a hybrid approach for targeted assembly of homologous sequences. *bioRxiv*. (doi:10.1101/014662).

[74] Rivera, M.C., Jain, R., Moore, J.E. & Lake, J.A. 1998 Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6239-6244. (doi:10.1073/pnas.95.11.6239).

[75] Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. 2014 Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182. (doi:10.1186/1471-2105-15-182).

[76] Li, H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

[77] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303. (doi:10.1101/gr.107524.110).

[78] Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. 2017 Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. (doi:10.1101/201178).

[79] Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158. (doi:10.1093/bioinformatics/btr330).

[80] Stamatakis, A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313. (doi:10.1093/bioinformatics/btu033).

[81] Zhang, C., Sayyari, E. & Mirarab, S. 2017 ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches. In *Comparative Genomics* (eds. J. Meidanis & L. Nakhleh), pp. 53-75. Cham, Springer International Publishing.

[82] Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. & RoyChoudhury, A. 2012 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29**, 1917-1932. (doi:10.1093/molbev/mss086).

[83] 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. (doi:10.1371/journal.pcbi.1003537).

[84] Veith, M., Bogaerts, S., Pasmans, F. & Kieren, S. 2018 The changing views on the evolutionary relationships of extant Salamandridae (Amphibia: Urodela). *PLoS ONE* **13**, e0198237. (doi:10.1371/journal.pone.0198237).

[85] Wielstra, B., Vörös, J. & Arntzen, J.W. 2016 Is the Danube crested newt *Triturus dobrogicus* polytypic? A review and new nuclear DNA data. *Amphib.-Reptil.* **37**, 167-177. (doi:10.1163/15685381-00003041).

[86] Gvoždík, L. & van Damme, R. 2006 *Triturus* newts defy the running-swimming dilemma. *Evolution* **60**, 2110-2121. (doi:10.1111/j.0014-3820.2006.tb01848.x).

[87] Fish, F.E. & Baudinette, R.V. 1999 Energetics of locomotion by the Australian water rat (*Hydromys chrysogaster*): a comparison of swimming and running in a semi-aquatic mammal. *The Journal of Experimental Biology* **202**, 353.

[88] Shine, R. & Shetty, S. 2001 Moving in two worlds: aquatic and terrestrial locomotion in sea snakes (*Laticauda colubrina*, Laticaudidae). *J. Evol. Biol.* **14**, 338-346. (doi:10.1046/j.1420-9101.2001.00265.x).

[89] Gillis, G.B. & Blob, R.W. 2001 How muscles accommodate movement in different physical environments: aquatic vs. terrestrial locomotion in vertebrates. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **131**, 61-75. (doi:10.1016/S1095-6433(01)00466-4).

[90] Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., et al. 2014 XSEDE: accelerating acientific discovery. *Computing in Science & Engineering* **16**, 62-74. (doi:10.1109/MCSE.2014.80).

Figures

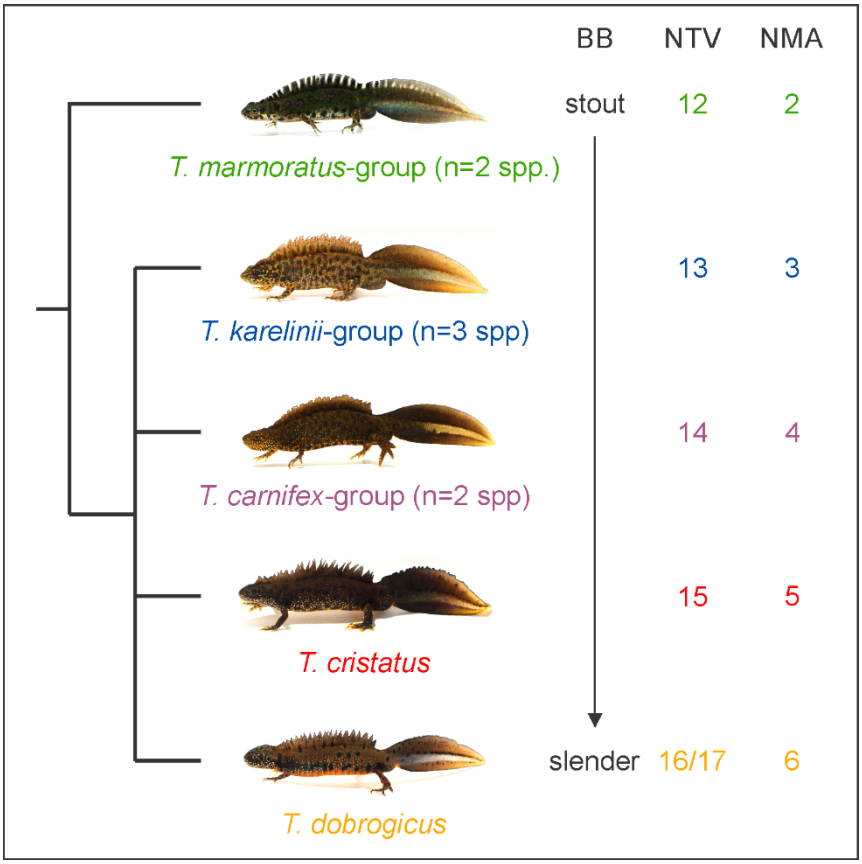

Fig. 1. The adaptive radiation of *Triturus* newts. Five body builds (BB) from stout to slender are observed in *Triturus* that are also characterized by an increasing number of trunk vertebrae (NTV) and number of annual aquatic months (NMA). The marbled newts (*T. marmoratus*-group) and crested newts (remaining four BBs) are sister clades. Relationships among the crested newts are not yet resolved and are the main focus of the present study.

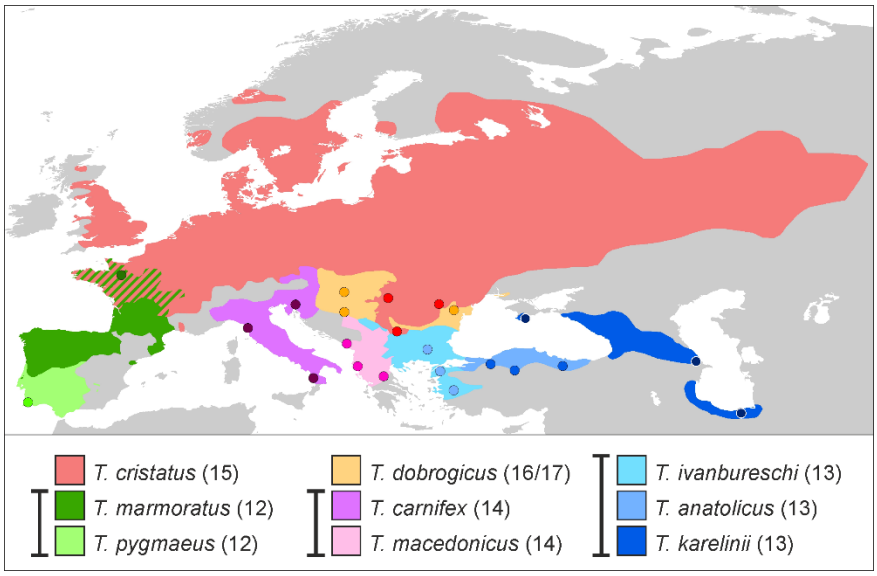

Fig. 2. Distribution and sampling scheme for *Triturus*. Dots represent sample localities (details in Supplementary Table S1). For the marbled newts (in green) a single individual is sampled for each of the two species and for the crested newts (other colours) three individuals are sampled for all seven species. The number in parentheses reflects each species' characteristic number of trunk vertebrae and whiskers link species that possess the same body build (see Fig. 1).

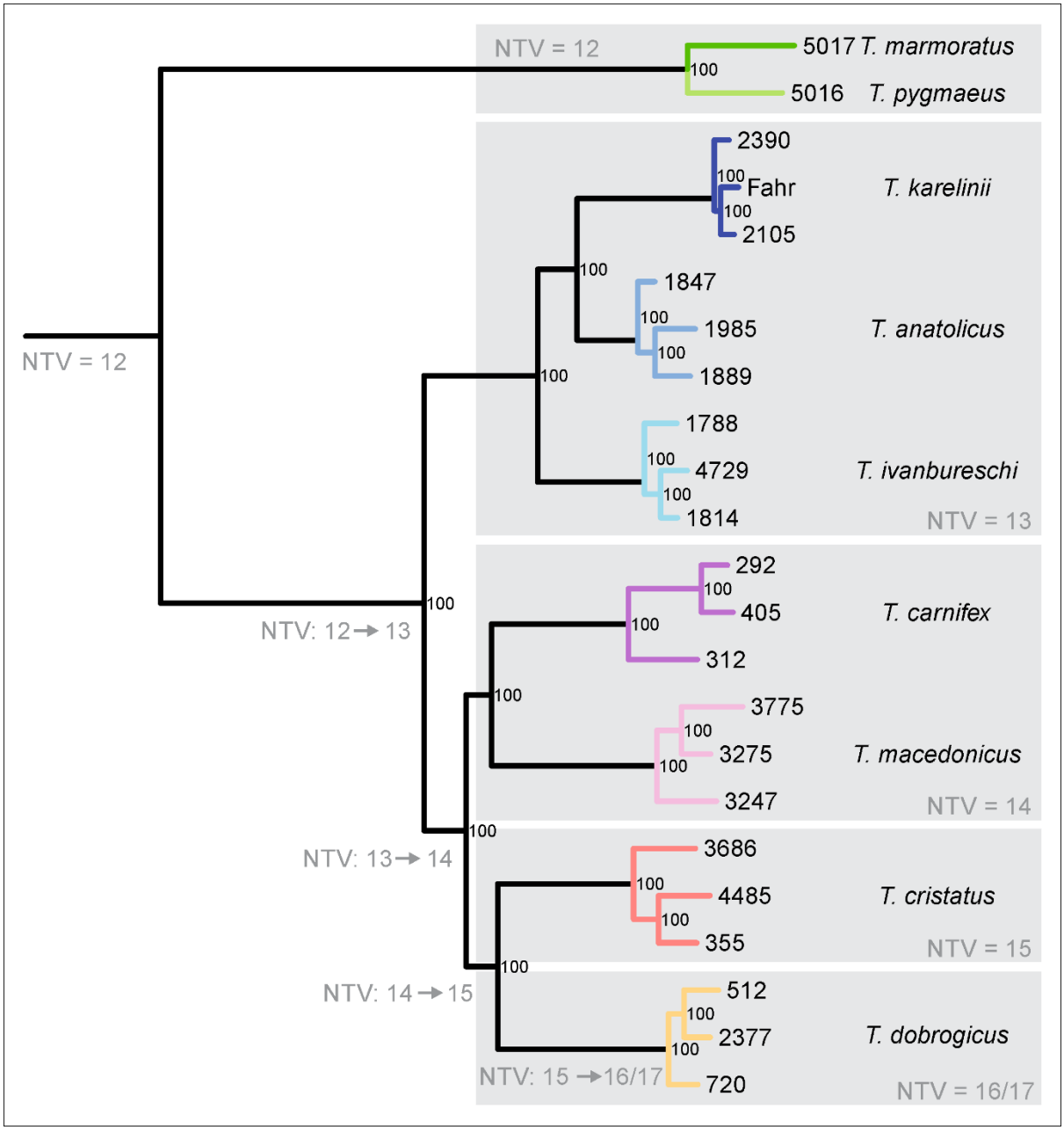

Fig. 3. *Triturus* newt phylogeny based on data concatenation with RAxML. This maximum likelihood phylogeny is based on 133,601 SNPs derived from 5,866 nuclear markers. Numbers at nodes indicate bootstrap support from 100 rapid bootstrap replicates. The five *Triturus* body builds (see Fig. 1) are delineated by grey boxes, with their characteristic number of trunk vertebrae (NTV) noted. Inferred changes in NTV are noted along internal branches. Colours reflect species and correspond to Fig. 2. Tip labels correspond to Supplementary Table S1.

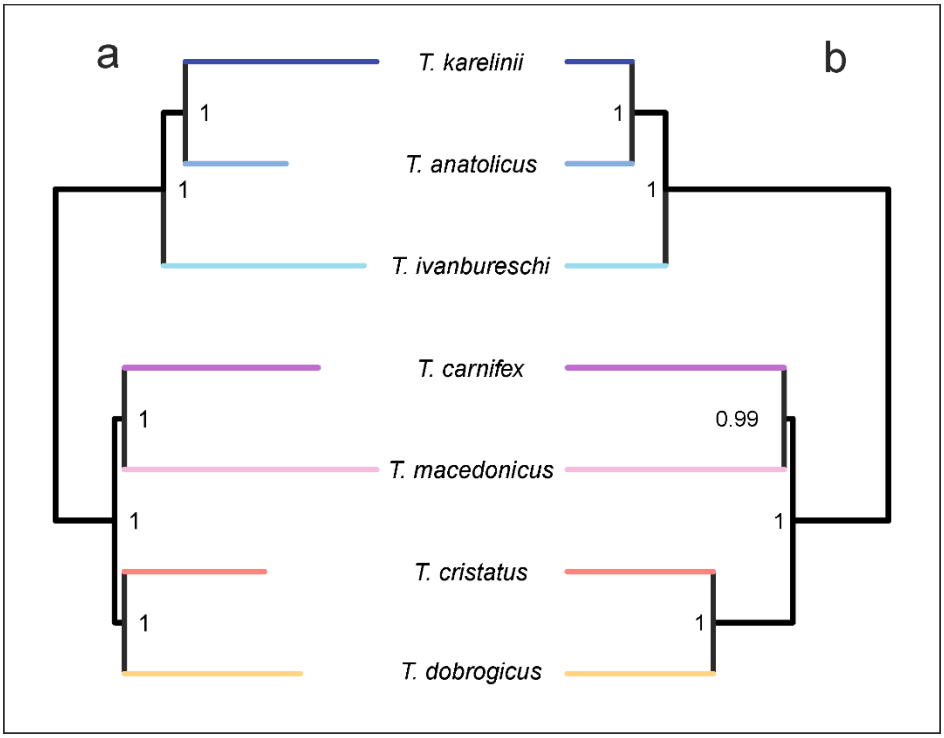

Fig. 4. Crested newt phylogeny based on gene tree summary with ASTRAL and species tree estimation with SNAPP. The ASTRAL tree (a) is based on 5,610 gene trees. Numbers at nodes indicate local quartet support posterior probabilities. The SNAPP tree (b) is based on single biallelic SNPs taken from 5,581 nuclear markers. Numbers at nodes indicate posterior probabilities. Colours reflect species and correspond to Fig. 2. Note that both topologies are identical to the phylogeny based on data concatenation (Fig. 3).

Cover image

Male Balkan crested newt (*Triturus ivanbureschi*) in breeding condition. A relatively stocky *Triturus* newt with a relatively short annual aquatic period. Photo credit: Michael Fahrbach.

Phylogenomics of the adaptive radiation of *Triturus* newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle

B. Wielstra, E. McCartney-Melstad, J.W. Arntzen, R.K. Butlin, H.B. Shaffer

SUPPLEMENTARY TEXT, FIGURES AND TABLES

SUPPLEMENTAL TEXT S1-S3

S1: Array Design

Transcriptome sequencing – Liver tissue samples in RNAlater from ten newts (one each of *Triturus anatolicus*, *T. carnifex*, *T. cristatus*, *T. dobrogicus*, *T. ivanbureschi*, *T. karelinii*, *T. macedonicus*, *T. marmoratus*, *T. pygmaeus*, and *Ommatotriton nesterovi*; Supplementary Table S1) were sent to ZF-Genomics (Leiden, The Netherlands) for RNA extraction and sequencing on a HiSeq 2500. Samples received an average of 43,810,415 clusters (SD=9,744,176) in 150bp paired-end configuration.

QC and Assembly – Paired-end sequencing reads were trimmed for adapter contamination and sequence quality using a 4-bp sliding window in Trimmomatic 0.32 [1], clipping the 3' ends of reads when the average sequence quality within the window dropped below 20. Leading bases with a quality score less than 5 and trailing bases with a quality score less than 15 were also removed, and reads shorter than 40bp were discarded after trimming.

A median of 38,575,204 read pairs were input into the Trinity assembler for each of the ten species (min=27,572,854, max=54,993,188, sd=8,916,227), and a median of 18.6% of these were retained after *in silico* normalization (min=15.8%, max=22.8%, sd=2.3%). Each transcriptome was individually assembled using Trinity 2.2.0 with read coverage normalized to a maximum of 50 [2]. Individual trinity assemblies were clustered at 90% identity using usearch v9.1.13 to reduce redundancy [3]. Assemblies contained a median of 157,608 contigs after clustering at 90% similarity (min=80,803 for *T. karelinii*, and max=182,488 for *T. carnifex*).

These clustered assemblies were then used for pairwise comparison between *T. dobrogicus* and the other nine species using *blastn* v2.2.30 [4]. The reciprocal best blast hits (RBBH) method was used to determine orthology between the assembled transcripts for each pairwise species comparison [5, 6]. *T. dobrogicus* transcripts that possessed reciprocal best blast hits to all of the nine other species were retained and all other transcripts were discarded.

Transcriptome comparison – The remaining set of 10,333 *T. dobrogicus* transcripts was self-blasted to attempt to further reduce redundancy, which may help reduce the inclusion of multiple isoforms of the same gene, chimeric transcripts assembled by trinity, and transcripts with truly similar regions that may complicate downstream bioinformatics. As a conservative measure, both the subject and query transcript were discarded if any transcript showed significant similarity (blast e-value < 0.001) to a different transcript or to different regions of itself.

Annotation – The remaining set of 9,214 *T. dobrogicus* transcripts described above were annotated using a translated blastx search to known *X. tropicalis* proteins with an e-value cutoff of 0.1 [7]. Transcripts that did not have a positive blastx hit to the *Xenopus* protein database were discarded, and only a single transcript annotating to a particular *Xenopus* protein was retained.

Splice site prediction – For the remaining set of 7,228 *T. dobrogicus* transcripts we attempted to infer splice sites in the candidate targets to avoid designing baits that span such boundaries, as these baits may perform poorly [8] and because targeting a single exon for each transcript simplifies downstream analyses. Splice sites were predicted by attempting to map each transcript to the *Chrysemys picta* genome [9] using exonerate's est2genome model [10] with a DNA word length of 10. Approximately 93% of all transcripts (n=6,758) successfully mapped to the *C. picta* genome, and for regions that mapped, the longest contiguous section of the mapped transcript was harvested. If the longest contiguous segment was less than 200bp, the first high-scoring segment pair (HSP) was extended towards the 5' end until reaching 200bp, followed by extending the final HSP towards the 3' end until reaching 200bp if necessary. Of the 6,758 transcripts that mapped to *C. picta*, 69 transcripts did not have an HSP longer than 200bp and could not be extended to 200bp in the 5' or 3' direction and were dropped as prospective targets.

The 470 transcripts that did not align to the *C. picta* genome were sequentially aligned to the genomes of *X. tropicalis* [7], *Nanorana parkerii* [11], and *Rana catesbeiana* [12] to attempt to find splice sites, taking the first successful species alignment from the list. Of these 470 transcripts, 125 mapped to *X. tropicalis*, 39 mapped to *N. parkerii*, and 36 mapped to *R. catesbeiana*. Again the longest contiguous aligned segment of each transcript was retained as a possible target, and transcripts with no aligned HSP of at least 200bp had their alignments extended in the 5' then 3' directions to attain targets of at least 200bp. For the 270 transcripts that did not map to any genome, the first (leftmost) 300bp of the assembled transcript was selected for a target region (except for the one transcript that was only 231bp long—for this target the entire 231bp transcript was used). It is possible that this leftmost orientation may enrich these targets for UTR sequence, assuming that the transcript was fully assembled by Trinity.

All exon targets were trimmed to a maximum of 450bp (from the 3' edge) and checked again for complementarity using a self BLAST in blastn. The first qualifying target from each unique Trinity cluster-gene identifier was retained, and any other targets that arose from the same Trinity gene identifier were discarded (n=19). This target set contains subsequences from 7,139 different transcripts for a total length of 2,272,851bp (mean=318bp, min=200bp, max=400bp, median=300bp).

As we are interested in capturing these loci from all *Triturus* taxa, including both crested and marbled newts, we decided to include probes designed from multiple species for the same target if divergence between representative species in the two main clades was greater than 5% [13]. Since the bulk of the target sequences were designed from *T. dobrogicus*, which together with *T. carnifex*, *T. cristatus* and *T. macedonicus* encompasses one of two main clades in the crested newts [14, 15], the three remaining species of crested newts encompassing the other clade (*T. karelinii*, *T. anatolicus, and T. ivanbureschi*) were used to determine if greater than 5% divergence existed between the two major clades for that target. First, the *T. dobrogicus* targets were blasted against *T. karelinii*, enforcing a full-length HSP with respect to the query sequence, yielding 2,850 hits; 30 of these were found to have a divergence greater than 5% and were added to the 7,139 *T. dobrogicus* targets. Then the remaining 4,289 *T. dobrogicus* targets were blasted to *T. anatolicus*, yielding 2,883 hits and an additional 35 targets. Finally the remaining 1,406 *T. dobrogicus* targets were blasted to *T. ivanbureschi*, yielding 631 hits and 10 more targets. Subsequently the process was repeated for the marbled newts *T. pygmaeus* and *T. marmoratus*, which constitute the sister lineage of the two crested newt clades, yielding an additional 222 and 27 targets after positive hits for 5,544 of 7,139 targets and 440 of 1,595 residual targets, respectively. Overall, an additional 324 orthologous targets that were more than 5% divergent between *T. dobrogicus* and other *Triturus* species were added to attempt to generate a set of probes that would perform well across the genus.

A set of 7,463 target sequences (average length=317bp, min=175bp, max=474bp) was sent to Arbor Biosciences for probe tiling and synthesis. After removing any probes softmasked by RepeatMasker and the Amphibia database, 39,143 unique RNA probes were synthesized at approximately 2.6X tiling density across 7,418 target sequences by Arbor Biosciences (Ann Arbor, MI) as a MyBaits-II kit.

98 *Test for phylogenetic utility* – The phylogenetic utility of the genomic transcript markers was validated by
99 building a phylogeny from the transcript sequences with RAxML. Trinity-assembled transcriptomes were
100 clustered at 90% identity using usearch v9.1.13 [3], and the sequence capture targets were aligned to these
101 clusters using blastn v2.2.31 [4]. The sequences corresponding to each target were extracted for each sample
102 and aligned using mafft v7.313 [16] and all 7,139 sequence alignments (1 per target) were concatenated.
103 RAxML v8.2.11[17] was used to generate a maximum likelihood phylogeny using 100 rapid bootstrap
104 replicates and the GTRCAT model of sequence evolution. Results suggested sufficient phylogenetic
105 resolution, but one unexpected finding was the placement of *T. carnifex* as sister to *T. dobrogicus*
106 (Supplementary Fig. S1a). Yet, in our main experiment, *T. carnifex* was more closely related to *T.*
107 *macedonicus* (see Results). The fact that the *T. carnifex* sample used for transcriptome sequencing
108 originated from close to the documented hybrid zone with *T. dobrogicus* [18, 19] suggests that substantial
109 interspecific gene flow might underlie this relationship. To test this scenario we obtained transcriptomes
110 from two additional *T. carnifex* individuals, sampled away from the hybrid zone with *T. dobrogicus*,
111 representing the distinct Balkan and Italian mtDNA clades [20, 21]. We processed these two individuals as
112 above and reran RAxML, replacing the *T. carnifex* sample from the hybrid zone, and now *T. carnifex*
113 individuals was recovered as sister to *T. macedonicus* (Supplementary Fig. S1b). This reflects the risk of
114 single-exemplar sampling [22] and the distorting influence interspecific gene flow has on phylogenetic
115 inference [23]. These findings support our decision to include multiple samples per species and exclude
116 samples from near known hybrid zones in our main experiment.

S2: Processing of Sequence Capture Data

Reference assembly – Sequence reads from the sample with the most reads (*T. carnifex* 292 with 3,937,346 read pairs) were used to *de novo* assemble target sequences for each target region. Trimmomatic v0.36 [1] was first used to remove adapter contamination and trim leading bases with scores < 5, trailing bases with scores < 15, also employing a 4bp sliding window from 5' to 3', trimming the window and downstream sequence when the average quality of the window dropped < 20. Reads < 40bp were discarded. Trimmed reads were input into PEAR v0.9.10 [24] with default settings to merge overlapping paired end reads into longer single-end fragments.

Unmerged reads and merged read pairs were input into the assembly by reduced complexity (ARC) pipeline [25], which performs alternating tasks of mapping reads to target sequences, followed by per-target *de novo* assembly of mapped reads, replacing the original target sequences with the target assembly at each iteration. Six iterations were performed to generate a set of reference contigs assembled from reads relevant to each target region. A single assembled contig was then selected for each original target region by means of reciprocal best blast hit (RBBH) [26]. These RBBHs were then blasted against one another to determine self-complementary regions, which can indicate chimeric assembly regions, and regions found to be similar to other target regions were trimmed to the nearest terminus of the contig [27]. This set of chimera-trimmed RBBHs was used as a target reference assembly for all downstream analyses.

QC, SNP calling and genotyping – Adapter contamination from library DNA inserts < 150bp was removed from reads using skewer v0.2.2 [28]. Reads were mapped to the reference assembly using BWA-MEM v0.7.15-r1140 [29]. Picard tools v2.9.2 (https://broadinstitute.github.io/picard/) was used to add read group information and mark PCR duplicates, and GATK v3.8 was used to generate gVCFs for each sample using HaplotypeCaller. GenotypeGVCFs was used for groups of samples (crested newts or crested + marbled newts, depending on the analysis) to call SNPs/genotypes, removing SNPs flagged by the following hard filters: QD < 2, MQ < 40, FS > 60, MQRankSum < -12.5, ReadPosRankSum < -8, QUAL < 30 [30, 31].

The *de novo* assembly followed by RBBH approach is susceptible to the inclusion of paralogous loci as putatively single copy targets. Because fixed differences between paralogues will appear as consistently heterozygous SNPs, we next attempted to remove paralogous targets from our dataset through the use of a Hardy Weinberg Equilibrium (HWE) filter for heterozygote excess. Heterozygote excess p-values were calculated for every SNP using vcftools 0.1.15 [32], and any target containing at least one SNP with a heterozygote excess p-value less than 0.05 was removed from downstream analysis.

Reference assembly and genotyping – A total of 4,932,636 reads (including 2,579,319 merged read pairs with an average length of 196bp) were used as input in the ARC assembly pipeline. After six iterations of mapping and assembly, 6,970 targets finished with an average of 295 reads apiece (median=167, sd=1,152), and 6,686 of the original targets had RBBHs to the assembly. After self-blast and trimming to remove potentially chimeric assemblies, a reference assembly of 5,593,497bp was generated for subsequent read mapping and SNP calling.

A median of 44.1% of trimmed reads aligned to the reference assembly (min=41.0%, max=50.5%), and an average of 22.6% of mapped reads were flagged as PCR duplicates, yielding a median unique reads on target of 34.2% (min=31.3%, max=39.4%). For the 23-sample dataset including the two marbled newt species, a total of 370,007 SNPs were recovered that passed hard filters. Of the 6,686 starting targets, 798 were found to contain at least 1 SNP with a HWE heterozygote excess p-value less than 0.05 and were removed. For the 21-sample dataset that did not contain the marbled newts, a total of 286,691 SNPs passed the hard filters and 814 targets were removed because they failed the HWE filter. Pairwise F84 divergences calculated with Phylip 3.697 [33] and based on the 23-sample dataset (including all *Triturus* species) are provided in Supplementary Table S2. The highest intraspecific divergence was observed between the Italian and Balkan clades comprising *T. carnifex*.

S3: Phylogenetic Analyses

Data concatenation with RAxML – RAxML version 8.2.11 [17] was used to infer phylogenies from concatenated alignments of SNPs. All biallelic SNPs in the 23-sample dataset that had genotype qualities of at least 20 in at least 50% of the samples and that fix RAxML's definition of variable (133,601 SNPs total across 5,866 different targets) were used for maximum likelihood phylogenetic analysis. 100 rapid bootstrap replicates and 20 maximum likelihood searches were conducted with the ASC_GTRGAMMA model with Lewis ascertainment correction for SNP analysis [34]. The resulting phylogeny with bootstrap support values was plotted in R using phytools [35].

The mean depth of passing genotype calls across all samples was 42.4X, and median per-site missingness was 4.3%, which corresponds to a single missing genotype (mean=10.1%, sd=14.0%). All crested newt species (for which three individuals were included) were recovered as monophyletic, and all bootstrap values on the tree were 100 (Fig. 3). The longest branch was between the marbled and crested newts and was used to root the tree. Within the crested newts, *T. ivanbureschi* was sister to a clade consisting of *T. anatolicus* and *T. karelinii*. The remaining four species were sister to this assemblage, with *T. carnifex* most closely related to *T. macedonicus* and *T. cristatus* most closely related to *T. dobrogicus*. Since the monophyly of all species was strongly supported, species designations were fixed for subsequent species tree inference.

Gene tree summary with ASTRAL – The species tree inference method ASTRAL v5.6.1 was used to estimate the crested newt phylogeny and explore incomplete lineage sorting from a collection of gene trees [36-38]. No marbled newts were included because estimating terminal branch lengths is not possible for species with a single representative (note that the reciprocal monophyly of crested and marbled newts is well established [15, 39, 40] and also strongly supported by our concatenated RAxML analysis). Separate polymorphic SNP alignments were first generated for each target using SnpSift 4.3 [41] and PGDSpider 2.1.1.2 [42], omitting SNPs with > 50% missing data across the 21 crested newt samples and removing targets that contained one or more samples with 100% missing data across the target using trimal v1.4.1 [43]. RAxML v8.2.11 [17] was used to infer a maximum likelihood gene tree for each target with the ASC_GTRGAMMA model and Lewis ascertainment bias correction [34].

After setting genotypes with quality scores less than 20 to missing data and filtering out sites with > 50% missing data, a total of 143,571 SNPs remained across 5,861 targets to build gene trees. After removing targets that contained samples with 100% missing data and removing sites that RAxML determined to be monomorphic, maximum likelihood gene trees were built for 5,610 targets. These gene trees were used as input into ASTRAL, constraining the seven crested newt species to be monophyletic (as supported by our concatenated RAxML analysis) and outputting local posterior probabilities and inferring terminal branch lengths. Midpoint rooting was used to determine the root. ASTRAL yielded a final normalized quartet score of 0.63. The same topology as in the concatenated RAxML analysis was recovered, with local posterior probabilities of 1 for all nodes (Fig. 4a). Branch lengths in ASTRAL are measured in coalescent units and indicate the degree of discordance among gene trees (within taxa for terminal branches and among taxa for internal branches). The longest terminal branch was recovered for *T. macedonicus*, and the shortest belonged to *T. anatolicus*. The shortest internal branches were those separating the sister lineages *T. carnifex* + *T. macedonicus* from *T. cristatus* + *T. dobrogicus*.

Species tree estimation with SNAPP – The coalescent species tree inference method SNAPP v1.3.0 was used to infer the crested newt species tree from biallelic SNPs [44]. Again, marbled newts were not included because sampling one individual per species violates the Yule speciation prior assumption. Polymorphic biallelic SNPs with genotype phred scores ≥ 20 across all 21 crested newts were first collected. Then, a single SNP from each of the 5,581 remaining loci was randomly selected to reduce the impacts of physical genetic linkage. These SNPs were used as input into SNAPP within the BEAST v2.4.8 environment [45] with the following parameters: species assignment=7 respective species, mutation rate U=1.0, mutation rate V=1.0, coalescence rate=10.0 (and sampled), use log likelihood correction=True, lambda prior=Gamma (initial=10[0.0,inf]) with alpha=2.0 and beta=200.0, snapprior.alpha=1.0, snapprior.beta=250.0, snapprior.kappa=1.0, snapprior.lambda=10.0 (and sampled), snapprior.rateprior=gamma, chain length=10,000,000, store every=1000 (and logging every 1000), and pre burnin=0. A 10% burnin was used and convergence and mixing were assessed with Tracer v1.7.1 [46]. ESS values for all parameters were > 400. A maximum clade credibility tree was constructed with common ancestor heights using TreeAnnotator v2.4.8 [45]. Note that BEAST infers the root as part of the analysis. The same topology as in the RAxML and ASTRAL analyses was recovered (Fig. 4b). All posterior probabilities were 1, except for the node subtending *T. carnifex* and *T. macedonicus*, which was 0.99.

A time-calibrated phylogeny was estimated with SNAPP using the same input SNP file as above. For calibration we interpreted the origin of the Adriatic Sea at the end of the Messinian Salinity Crisis at 5.33 million years ago [47] as the vicariance event causing the *T. carnifex* versus *T. macedonicus* split [14, 39] and set the age of their most recent common ancestor to a uniform distribution between 5.32 and 5.34 million years ago [48]. The output tree from the original, undated SNAPP analysis was used as a starting tree, scaling the entire tree so that the starting age of the calibration node was 5.33 million years ago. The topology was fixed to that recovered by the original SNAPP analysis and dates of remaining nodes were estimated using 1,000,000 MCMC steps, sampling every 500 steps and removing a 10% burn-in. ESS values for parameters were confirmed > 400 with Tracer. A maximum clade credibility tree with median node heights was generated with TreeAnnotator (Supplementary Fig. S2).

S4: Comparison with full mtDNA-based phylogeny

MtDNA has proven misleading at both recent [49] and deeper [50] nodes in the Salamandridae phylogeny and our genome-enabled phylogeny shows a highly supported deviation with a previous full mtDNA (i.e. single marker) phylogeny as well [14]. The deviation concerns the relationship among the three species constituting the '*T. karelinii*-group'; we here recover *T. anatolicus* as sister to *T. karelinii*, rather than to *T. ivanbureschi* as suggested by mtDNA (Supplementary Fig. S3). While such gene tree discordance could reflect incomplete lineage sorting of mtDNA [51], we consider ancient mtDNA introgression more likely, as *T. ivanbureschi* and *T. anatolicus* show geographically extensive introgressive hybridization today [52]. A scenario of ancient introgression is in line with the high degree of gene tree/species tree discordance in the nuclear genome in *T. anatolicus*, as suggested by the short branch in the ASTRAL tree (Fig. 4a). However, as all members of the '*T. karelinii*-group' possess an identical NTV count, the mtDNA-nuDNA mismatch does not influence our interpretation of character evolution (Supplementary Fig. S4).

References

[1] Bolger, A.M., Lohse, M. & Usadel, B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120. (doi:10.1093/bioinformatics/btu170).

[2] Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652. (doi:10.1038/nbt.1883).

[3] Edgar, R.C. 2010 Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461. (doi:10.1093/bioinformatics/btq461).

[4] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. 2009 BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421-421. (doi:10.1186/1471-2105-10-421).

[5] Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. & Yuan, Y. 1998 Predicting function: from genes to genomes and back. *J. Mol. Biol.* **283**, 707-725. (doi:10.1006/jmbi.1998.2144).

[6] Tatusov, R.L., Koonin, E.V. & Lipman, D.J. 1997 A genomic perspective on protein families. *Science* **278**, 631. (doi:10.1126/science.278.5338.631).

[7] Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., et al. 2010 The genome of the western clawed frog *Xenopus tropicalis*. *Science* **328**, 633. (doi:10.1126/science.1183670).

[8] Neves, L.G., Davis, J.M., Barbazuk, W.B. & Kirst, M. 2013 Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* **75**, 146-156. (doi:10.1111/tpj.12193).

[9] Shaffer, H.B., Minx, P., Warren, D.E., Shedlock, A.M., Thomson, R.C., Valenzuela, N., Abramyan, J., Amemiya, C.T., Badenhorst, D., Biggar, K.K., et al. 2013 The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28. (doi:10.1186/gb-2013-14-3-r28).

[10] Slater, G.S.C. & Birney, E. 2005 Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31. (doi:10.1186/1471-2105-6-31).

[11] Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., Zhong, L., Wang, L., Wu, D.-D., Zhang, B.-L., et al. 2015 Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1257. (doi:10.1073/pnas.1501764112).
[12] Hammond, S.A., Warren, R.L., Vandervalk, B.P., Kucuk, E., Khan, H., Gibb, E.A., Pandoh, P., Kirk, H., Zhao, Y., Jones, M., et al. 2017 The North American bullfrog draft genome provides insight into hormonal regulation of long noncoding RNA. *Nat. Commun.* **8**, 1433. (doi:10.1038/s41467-017-01316-7).
[13] Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C. & Good, J.M. 2012 Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics* **13**, 403. (doi:10.1186/1471-2164-13-403).
[14] Wielstra, B. & Arntzen, J.W. 2011 Unraveling the rapid radiation of crested newts (*Triturus cristatus* superspecies) using complete mitogenomic sequences. *BMC Evol. Biol.* **11**, 162. (doi:10.1186/1471-2148-11-162).
[15] Wielstra, B., Arntzen, J.W., van der Gaag, K., 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**, e111011. (doi:10.1371/journal.pone.0111011).
[16] Katoh, K. & Standley, D.M. 2013 MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772-780. (doi:10.1093/molbev/mst010).
[17] Stamatakis, A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313. (doi:10.1093/bioinformatics/btu033).
[18] Arntzen, J.W., Wielstra, B. & Wallis, G.P. 2014 The modality of nine *Triturus* newt hybrid zones, assessed with nuclear, mitochondrial and morphological data. *Biol. J. Linn. Soc.* **113**, 604–622. (doi:10.1111/bij.12358).
[19] Wielstra, B., Sillero, N., Vörös, J. & Arntzen, J.W. 2014 The distribution of the crested and marbled newt species (Amphibia: Salamandridae: *Triturus*) – an addition to the New Atlas of Amphibians and Reptiles of Europe. *Amphib.-Reptil.* **35**, 376-381. (doi:10.1163/15685381-00002960).
[20] Canestrelli, D., Salvi, D., Maura, M., Bologna, M.A. & Nascetti, G. 2012 One Species, three Pleistocene evolutionary histories: phylogeography of the Italian crested newt, *Triturus carnifex*. *PLoS ONE* **7**, e41754. (doi:10.1371/journal.pone.0041754).
[21] Wielstra, B., Crnobrnja-Isailović, J., Litvinchuk, S.N., Reijnen, B.T., Skidmore, A.K., Sotiropoulis, K., Toxopeus, A.G., Tzankov, N., Vukov, T. & Arntzen, J.W. 2013 Tracing glacial refugia of *Triturus* newts based on mitochondrial DNA phylogeography and species distribution modeling. *Front. Zool.* **10**, 13. (doi:10.1186/1742-9994-10-13).
[22] Spinks, P.Q., Thomson, R.C., Pauly, G.B., Newman, C.E., Mount, G. & Shaffer, H.B. 2013 Misleading phylogenetic inferences based on single-exemplar sampling in the turtle genus *Pseudemys*. *Mol. Phylogenet. Evol.* **68**, 269-281. (doi:10.1016/j.ympev.2013.03.031).
[23] Leaché, A.D., Harris, R.B., Rannala, B. & Yang, Z. 2014 The influence of gene flow on species tree estimation: a simulation study. *Syst. Biol.* **63**, 17-30. (doi:10.1093/sysbio/syt049).
[24] Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. 2014 PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620. (doi:10.1093/bioinformatics/btt593).
[25] Hunter, S.S., Lyon, R.T., Sarver, B.A.J., Hardwick, K., Forney, L.J. & Settles, M.L. 2015 Assembly by Reduced Complexity (ARC): a hybrid approach for targeted assembly of homologous sequences. *bioRxiv*. (doi:10.1101/014662).
[26] Rivera, M.C., Jain, R., Moore, J.E. & Lake, J.A. 1998 Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6239-6244. (doi:10.1073/pnas.95.11.6239).
[27] McCartney-Melstad, E., Mount, G.G. & Bradley Shaffer, H. 2016 Exon capture optimization in amphibians with large genomes. *Mol. Ecol. Resour.* **16**, 1084-1094. (doi:10.1111/1755-0998.12538).
[28] Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. 2014 Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182. (doi:10.1186/1471-2105-15-182).
[29] Li, H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
[30] DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498. (doi:10.1038/ng.806).
[31] Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. 2017 Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. (doi:10.1101/201178).

[32] Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158. (doi:10.1093/bioinformatics/btr330).
[33] Felsenstein, J. 1989 PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166. (doi:10.1111/j.1096-0031.1989.tb00562.x).
[34] Lewis, P.O. 2001 A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913-925. (doi:10.1080/106351501753462876).
[35] Revell, L.J. 2011 phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217-223. (doi:10.1111/j.2041-210X.2011.00169.x).
[36] Zhang, C., Sayyari, E. & Mirarab, S. 2017 ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches. In *Comparative Genomics* (eds. J. Meidanis & L. Nakhleh), pp. 53-75. Cham, Springer International Publishing.
[37] 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. (doi:10.1093/bioinformatics/btu462).
[38] Sayyari, E. & Mirarab, S. 2016 Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **33**, 1654-1668. (doi:10.1093/molbev/msw079).
[39] Arntzen, J.W., 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.
[40] Espregueira Themudo, G., Wielstra, B. & Arntzen, J.W. 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. (doi:10.1016/j.ympev.2009.03.024).
[41] Ruden, D., Cingolani, P., Patel, V., Coon, M., Nguyen, T., Land, S. & Lu, X. 2012 Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* **3**. (doi:10.3389/fgene.2012.00035).
[42] Lischer, H.E.L. & Excoffier, L. 2012 PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298-299. (doi:10.1093/bioinformatics/btr642).
[43] Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973. (doi:10.1093/bioinformatics/btp348).
[44] Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. & RoyChoudhury, A. 2012 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29**, 1917-1932. (doi:10.1093/molbev/mss086).
[45] 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. (doi:10.1371/journal.pcbi.1003537).
[46] Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. 2018 Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.*, syy032-syy032. (doi:10.1093/sysbio/syy032).
[47] Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J. & Wilson, D.S. 1999 Chronology, causes and progression of the Messinian salinity crisis. *Nature* **400**, 652-655.
[48] Stange, M., Sánchez-Villagra, M.R., Salzburger, W. & Matschiner, M. 2018 Bayesian divergence-time estimation with genome-wide Single-Nucleotide Polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian isthmus. *Syst. Biol.* **67**, 681-699. (doi:10.1093/sysbio/syy006).
[49] Rodríguez, A., Burgon, J.D., Lyra, M., Irisarri, I., Baurain, D., Blaustein, L., Göçmen, B., Künzel, S., Mable, B.K., Nolte, A.W., et al. 2017 Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. *Mol. Phylogenet. Evol.* **115**, 16-26. (doi:10.1016/j.ympev.2017.07.009).
[50] Veith, M., Bogaerts, S., Pasmans, F. & Kieren, S. 2018 The changing views on the evolutionary relationships of extant Salamandridae (Amphibia: Urodela). *PLoS ONE* **13**, e0198237. (doi:10.1371/journal.pone.0198237).
[51] Platt, I.I.R.N., Faircloth, B.C., Sullivan, K.A.M., Kieran, T.J., Glenn, T.C., Vandewege, M.W., Lee, J.T.E., Baker, R.J., Stevens, R.D. & Ray, D.A. 2018 Conflicting evolutionary histories of the mitochondrial and nuclear genomes in New World *Myotis* bats. *Syst. Biol.* **67**, 236-249. (doi:10.1093/sysbio/syx070).
[52] Wielstra, B., Burke, T., Butlin, R.K., Avcı, A., Üzüm, N., Bozkurt, E., Olgun, K. & Arntzen, J.W. 2017 A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters* **1**, 93-101. (doi:10.1002/evl3.9).
[53] Arntzen, J.W., Beukema, W., Galis, F. & Ivanović, A. 2015 Vertebral number is highly evolvable in salamanders and newts (family Salamandridae) and variably associated with climatic parameters. *Contrib. Zool.* **84**, 85-113.

SUPPLEMENTARY FIGURES S1-S3

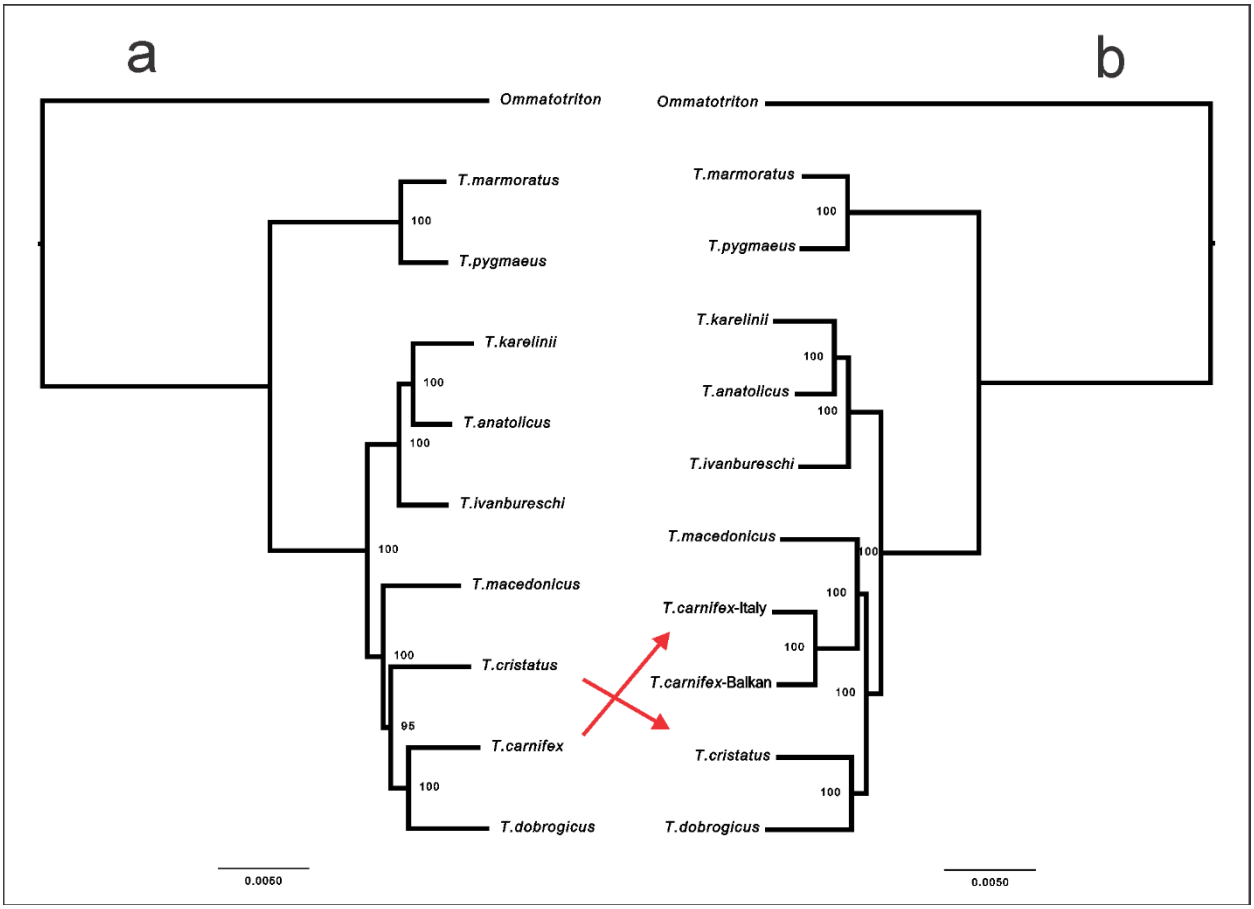

Fig. S1. *Triturus* newt phylogenies based on based on data concatenation of transcriptome data with RAxML. In a) a *T. carnifex* individual is included that is suspected to be admixed with *T. dobrogicus* and in b) this is replaced by two other *T. carnifex* individuals assumed to not be affected by genetic admixture, one from the Balkans and one from Italy, away from the contact zone with *T. dobrogicus*. Note the differences in sister species relationships (reflected by red arrows), with the phylogeny in b) being in full agreement with the one based on target capture data (Fig. 3; Fig. 4).

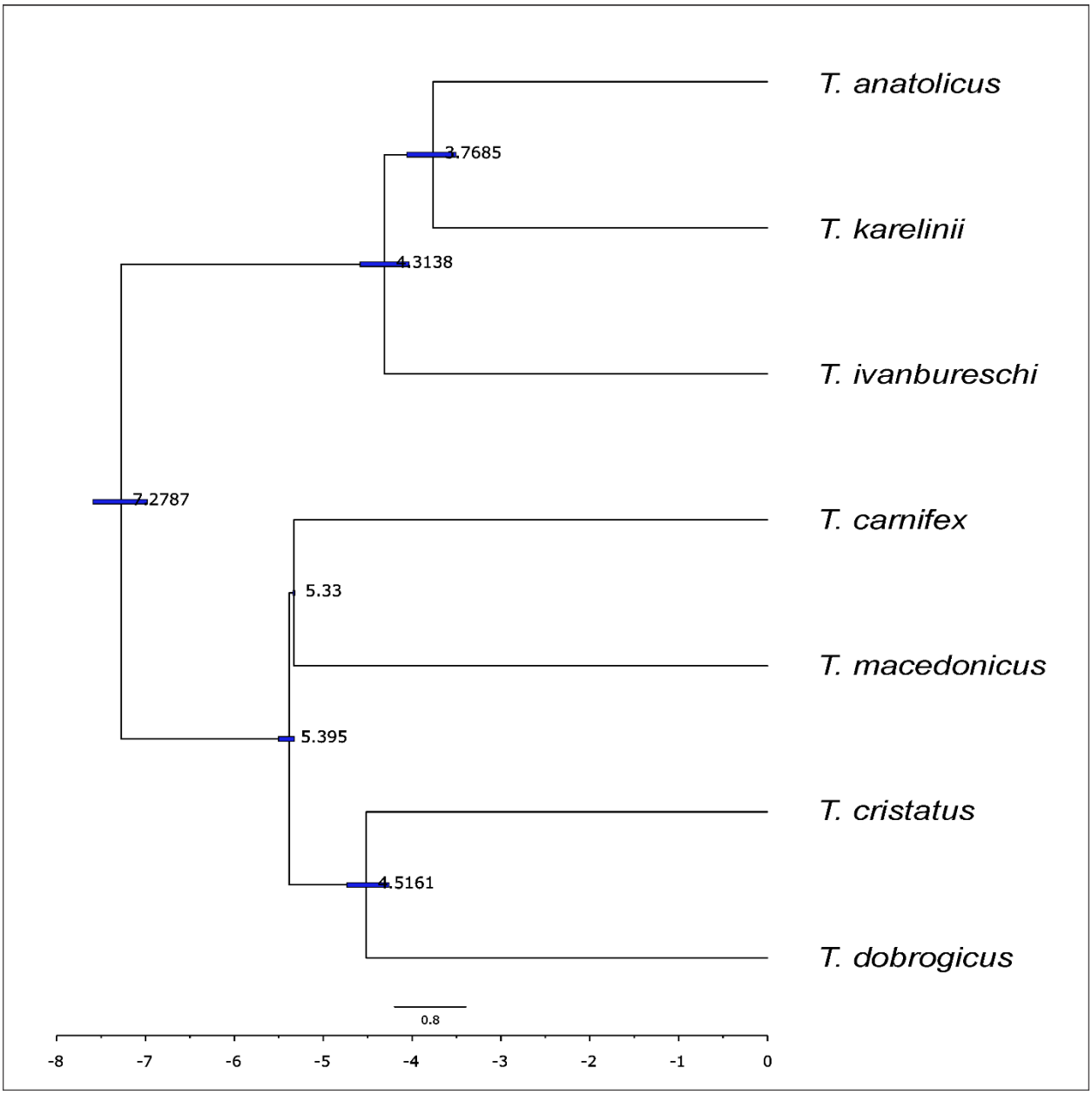

Fig. S2. Dated species tree for the crested newts (*Triturus cristatus* superspecies). SNAPP was used to calibrate divergence times among crested newt species. The split between *T. carnifex* and *T. macedonicus*, assumed to correspond to the origin of the Adriatic Sea at the end of the Messinian Salinity Crisis 5.33 million years ago, was used as a calibration point.

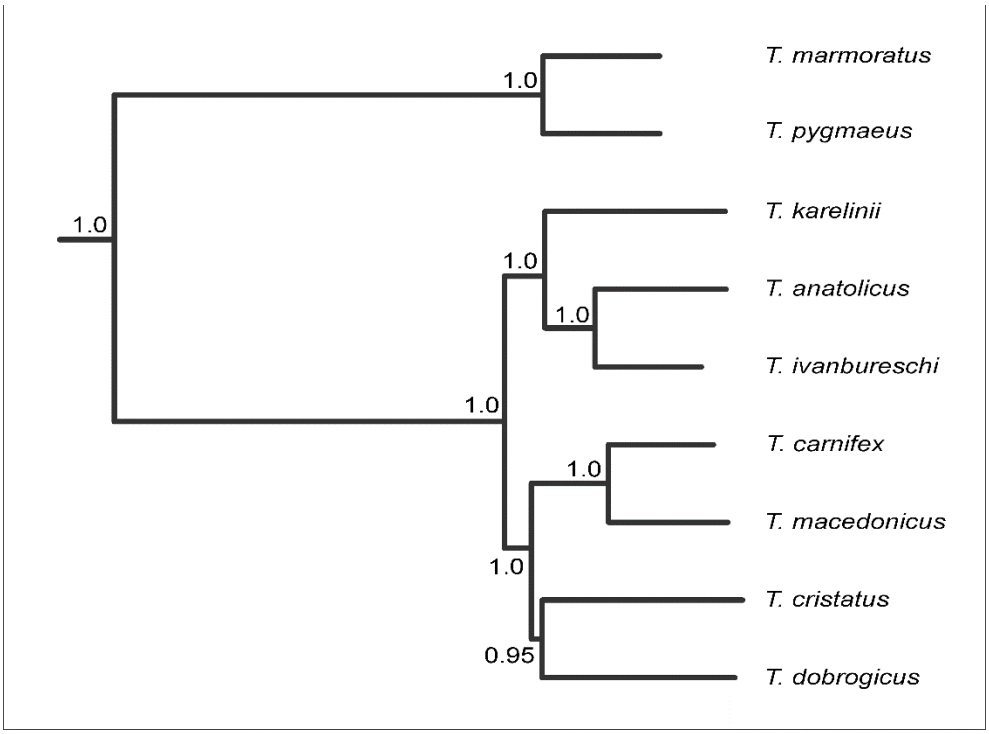

Fig. S3. Full mtDNA phylogeny for *Triturus*. The genome-enabled *Triturus* phylogeny (Fig. 3; Fig. 4) deviates from the phylogeny based on full mtDNA (taken from [14]) for the species relationships in the *T. karelinii*-group of crested newts (with *T. anatolicus* being sister to *T. karelinii* rather than *T. ivanbureschi*). Numbers at nodes indicate posterior probabilities. Note the relatively low support for the sister relationship between *T. cristatus* and *T. dobrogicus*.

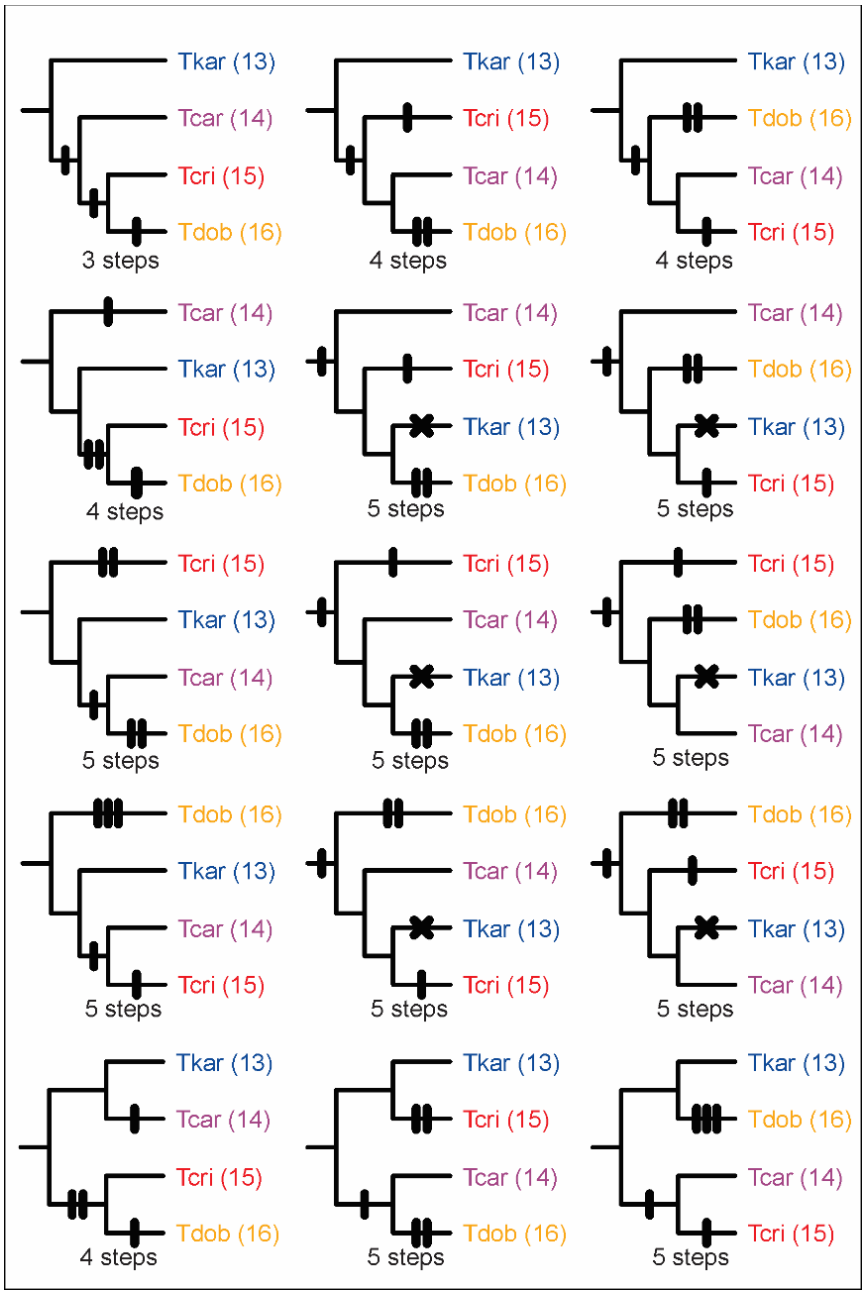

Fig. S4. All 15 topologies possible for a fully bifurcating phylogeny of the four body builds observed in the crested newts (*Triturus cristatus* superspecies). Abbreviations: Tkar = *T. karelinii*-group; Tcar = *T. carnifex*-group; Tcri = *T. cristatus*; Tdob = *T. dobrogicus*. The number of trunk vertebrae (NTV) for each body build is provided in parentheses. A bar is an NTV addition and a cross an NTV deletion. NTV = 12 as observed in the sister lineage the marbled newts (the *T. marmoratus*-group), as well as the most closely related genus *Lissotriton*, is considered to be the ancestral state [50, 53]. For convenience an NTV count of 16 is used for *T. dobrogicus*, but note that NTV = 17 also occurs at roughly equal frequency in this species, which does not influence our interpretation.

SUPPLEMENTARY TABLES S1-S2

Table S1. Sampling details. Individuals are identified with a code that refers to complete specimens (ID starting with ZMA) or tail tips (remaining samples). All material is stored at Naturalis Biodiversity Center, Leiden, The Netherlands.

Target capture

ID	Species	Locality	Latitude	Longitude
5017	*Triturus marmoratus*	France: Jublains	48.252	-0.473
5016	*Triturus pygmaeus*	Portugal: Serra de Monchique	37.335	-8.506
4729	*Triturus ivanbureschi*	Bulgaria: Ostar Kamak	41.878	25.853
1814	*Triturus ivanbureschi*	Turkey: Karakadılar	40.010	26.940
1788	*Triturus ivanbureschi*	Turkey: Bozdağ	38.367	28.103
1847	*Triturus anatolicus*	Turkey: Abanta Gölu	40.612	31.288
1889	*Triturus anatolicus*	Turkey: Gölköy	40.083	33.347
1985	*Triturus anatolicus*	Turkey: Çakırlı	40.446	37.483
2105	*Triturus karelinii*	Ukraine: Nikita	44.538	34.243
6719	*Triturus karelinii*	Azerbaijan: Altiagac	40.854	48.935
RMNH RenA 46931-2390	*Triturus karelinii*	Iran: Qu'Am Shahr	36.436	52.803
ZMA9108-405	*Triturus carnifex*	Italy: Fuscaldo	39.417	16.033
ZMA9145-292	*Triturus carnifex*	Italy: Pisa	43.717	10.400
ZMA9132-312	*Triturus carnifex*	Slovenia: Kramplje	45.733	14.500
3247	*Triturus macedonicus*	Montenegro: Bjeloši	42.374	18.907
3275	*Triturus macedonicus*	Albania: Bejar	40.429	19.850
3775	*Triturus macedonicus*	Greece: Kerameia	39.562	22.081
4485	*Triturus cristatus*	Bulgaria: Montana	43.416	23.222
3686	*Triturus cristatus*	Romania: Budeni	45.768	26.839
ZMA9167-355	*Triturus cristatus*	Romania: Virfuri	46.283	22.467
ZMA9083-512	*Triturus dobrogicus*	Hungary: Alap	46.800	18.683
ZMA9172-720	*Triturus dobrogicus*	Croatia: Zupanja	45.083	18.700
2377	*Triturus dobrogicus*	Romania: Măcin	45.251	28.121

Transcriptomes

ID	Species	Locality	Latitude	Longitude
6720	*Triturus marmoratus*	Portugal: Valongo	41.168	-8.500
6721	*Triturus pygmaeus*	Portugal: Serra de Monchique	37.335	-8.506
6722	*Triturus karelinii*	Azerbaijan: Altiagac	40.854	48.935
6723	*Triturus anatolicus*	Turkey: Hacılar	41.495	32.088
6724	*Triturus ivanbureschi*	Turkey: Keşan	40.924	26.635
6725	*Triturus carnifex*	Croatia: Prkovac	45.569	16.094
6726	*Triturus carnifex*	Croatia: Radetići	45.146	13.842
6727	*Triturus carnifex*	Italy: Viterbo	42.703	13.325
6728	*Triturus macedonicus*	Montenegro: Ceklin	42.367	18.982
6729	*Triturus cristatus*	France: Belgeard	48.259	-0.574
6730	*Triturus dobrogicus*	Serbia: Sremski Karlovski	45.175	19.991
6731	*Ommatotriton nesterovi*	Turkey: Hürriyet	40.276	28.650

Table S2. Inter- and intraspecific divergence in *Triturus* newts. Shown are pairwise F84 divergences calculated with Phylip. Intraspecific distances are in italics. IDs correspond to Supplementary Table S1.

		5017	**5016**	**4729**	**1814**	**1788**	**1847**	**1889**	**1985**	**2105**	**6719**	**2390**	**405**	**292**	**312**	**3247**	**3275**	**3775**	**4485**	**3686**	**355**	**512**	**720**	**2377**
T. marmoratus	**5017**	-																						
T. pygmaeus	**5016**	0.10	-																					
T. ivanbureschi	**4729**	0.72	0.70	-																				
	1814	0.72	0.70	*0.02*	-																			
	1788	0.72	0.71	*0.03*	*0.02*	-																		
T. anatolicus	**1847**	0.68	0.66	0.08	0.08	0.09	-																	
	1889	0.72	0.70	0.11	0.11	0.12	*0.02*	-																
	1985	0.72	0.70	0.11	0.11	0.12	*0.03*	*0.03*	-															
T. karelinii	**2105**	0.71	0.69	0.14	0.14	0.14	0.09	0.11	0.11	-														
	6719	0.74	0.72	0.14	0.14	0.15	0.10	0.12	0.11	*0.01*	-													
	2390	0.73	0.71	0.14	0.14	0.15	0.10	0.12	0.11	*0.01*	*0.02*	-												
T. carnifex	**405**	0.74	0.72	0.25	0.25	0.26	0.23	0.26	0.25	0.27	0.27	0.27	-											
	292	0.75	0.73	0.25	0.25	0.26	0.23	0.26	0.25	0.26	0.27	*0.03*	-											
	312	0.71	0.69	0.23	0.23	0.24	0.21	0.24	0.23	0.25	0.25	0.25	*0.08*	*0.07*	-									
T. macedonicus	**3247**	0.74	0.72	0.24	0.24	0.25	0.22	0.25	0.25	0.25	0.26	0.26	0.22	0.21	0.19	-								
	3275	0.72	0.70	0.23	0.23	0.23	0.20	0.23	0.23	0.24	0.24	0.24	0.20	0.20	0.18	*0.04*	-							
	3775	0.75	0.73	0.24	0.24	0.25	0.22	0.25	0.25	0.26	0.26	0.26	0.22	0.22	0.20	*0.06*	*0.04*	-						
T. cristatus	**4485**	0.71	0.69	0.21	0.21	0.22	0.19	0.22	0.21	0.23	0.23	0.23	0.21	0.21	0.19	0.20	0.18	0.20	-					
	3686	0.72	0.70	0.22	0.22	0.22	0.20	0.22	0.22	0.23	0.24	0.24	0.21	0.21	0.19	0.20	0.18	0.20	*0.05*	-				
	355	0.69	0.67	0.21	0.21	0.21	0.19	0.21	0.21	0.22	0.23	0.23	0.21	0.21	0.19	0.19	0.17	0.19	*0.04*	*0.05*	-			
T. dobrogicus	**512**	0.71	0.69	0.23	0.23	0.24	0.21	0.24	0.23	0.25	0.25	0.25	0.21	0.20	0.18	0.21	0.19	0.21	0.17	0.17	0.17	-		
	720	0.69	0.67	0.21	0.21	0.22	0.20	0.22	0.22	0.23	0.23	0.23	0.19	0.18	0.17	0.19	0.17	0.19	0.16	0.16	0.16	*0.03*	-	
	2377	0.70	0.68	0.22	0.22	0.22	0.20	0.22	0.22	0.23	0.24	0.24	0.19	0.19	0.17	0.19	0.18	0.19	0.15	0.16	0.15	*0.02*	*0.03*	-