Newts in time and space: the evolutionary history of Triturus newts at different temporal and spatial scales

Espregueria Themudo, G.

CHAPTER 4

MULTIPLE NUCLEAR AND MITOCHONDRIAL GENES RESOLVE THE BRANCHING ORDER OF A RAPID RADIATION OF CRESTED NEWTS (*TRITURUS*, SALAMANDRIDAE)

Espregueira Themudo, G.[1,2], B. Wielstra[1,3] & J.W. Arntzen[1]

1 - National Museum of Natural History – Naturalis, PO Box 9517, 2300 RA Leiden, the Netherlands.
2 - CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal.
3 - Institute of Biology Leiden - IBL, PO Box 9516, 2300 RA Leiden, the Netherlands.

Contents

Published in *Molecular Phylogenetics and Evolution* **52** (2), 321-328 (2009).

Abstract

Newts of the genus *Triturus* are parapatrically distributed across Europe. Within this group, the crested newts (*Triturus cristatus* superspecies) radiated in a short temporal interval. Given the relatively short timespan in between branching events and to address the gene tree – species tree problem, we sequenced two mitochondrial and five nuclear genes from populations representing the distribution range of all the five crested newt species. We built gene trees using non-hierarchical Bayesian phylogenetics and phylogenetic networks, and a species tree with a recently developed method, which uses a hierarchical Bayesian approach. While the single gene trees did not provide resolution, the hierarchical Bayesian method yielded an almost fully resolved species tree, even though branching events followed one another closely. Results show a previously undetected basal dichotomy between *T. karelinii* and the other four species and a deep differentiation of *T. karelinii* in two lineages, here raised to full species status.

Keywords: Evolutionary radiation, introns, nuclear markers, phylogeny, phylogenetic network, polytomy, species tree, *Triturus cristatus* superspecies, *Triturus karelinii*, *Triturus arntzeni*

Introduction

The task of molecular phylogenetics is to reconstruct the order by which species have originated. However, when the different lineages split over short time spans or when the studied markers contain insufficient information, phylogenies can remain unresolved. One way forward is to collect large amounts of independent data. Independence is essential to resolve the gene tree – species tree problem, as linked *loci* (for example, from mitochondrial DNA) will have the same history. Because of stochasticity in the lineage sorting process, incomplete lineage sorting and horizontal gene transfer, gene trees generally differ from one another and from the actual species tree (which is tried to be recovered) (AVISE and WOLLENBERG, 1997). This also brings up analytical problems. Over the last decades, phylogenetic analyses have mostly followed three strategies (reviewed in HUELSENBECK *et al.*, 1996): the total evidence approach (KLUGE, 1989), the separate analysis for each marker followed by an analysis for congruence (MIYAMOTO and FITCH, 1995), and conditional data combination (DE QUEIROZ, 1993). More recently, phylogenetic methods have been proposed that incorporate biologically more realistic models of evolution, by taking into account processes like incomplete lineage sorting or horizontal gene transfer. Two new methods are phylogenetic networks (see HUSON and BRYANT, 2006) and an approach that calculates species trees explicitly, (hereafter called 'the hierarchical method', EDWARDS *et al.*, 2007).

The genus *Triturus* is composed of two groups, with two and five species, respectively (the marbled newts - *T. marmoratus* and *T. pygmaeus* - from south-western Europe and the crested newts – or *T. cristatus* superspecies - from western, central and eastern Europe). Support for a monophyletic origin is strong for either of the groups, i.e., crested newts, marbled newts and *Triturus* (ARNTZEN *et al.*, 2007; STEINFARTZ *et al.*, 2007). The five crested newt species have mutually exclusive, parapatric ranges across Europe: the Italian crested newt (*T. carnifex*), the northern crested newt (*T. cristatus*), the Dobrogean crested newt (*T. dobrogicus*), the southern crested newt (*T. karelinii* –with two subspecies: *T. k. karelinii* and *T. k. arntzeni*) and the Macedonian crested newt (*T. macedonicus*). All species meet in the Balkans which is considered to be the centre of origin of the group (CRNOBRNJA-ISAILOVIC *et al.*, 1997, see Fig. 1). Crested newts come in four morphologically distinct lineages

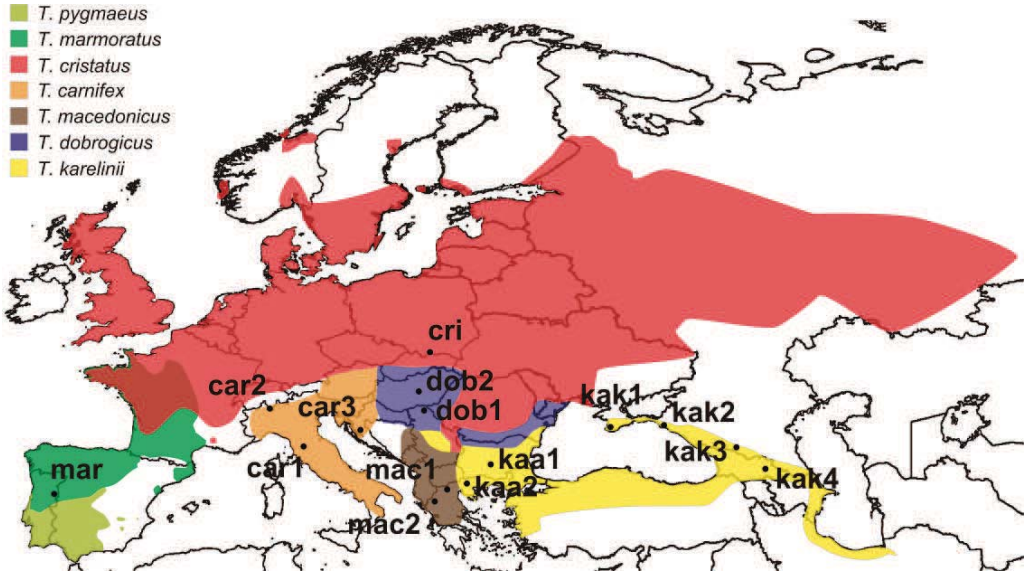

Figure 1- Geographical distribution of *Triturus* species, represented by the following colours: orange - *T. carnifex* (car), red - *T. cristatus* (cri), blue - *T. dobrogicus* (dob), yellow - *T. karelinii* (*T. k. arntzenii* - kaa and *T. k. karelinii* - kak), brown – *T. macedonicus* (mac), dark green - *T. marmoratus* (mar), light green - *T. pygmaeus*. Sample localities are as described in Appendix I.

that differ in the number of presacral vertebrae and relative appendage size (e.g., forelimb length divided by interlimb distance - a ratio known as the 'Wolterstorff-Index' and negatively correlated with the number of presacral vertebrae). Modal presacral vertebrae counts are 14 in *T. karelinii*, 15 in *T. carnifex* and *T. macedonicus*, 16 in *T. cristatus* and 17 or 18 in *T. dobrogicus*, with little intraspecific variation (ARNTZEN, 2003). *Triturus marmoratus* and *T. pygmaeus* have a modal presacral vertebrae count of 13. Note however that the evolutionary tree suggested by this morphological series, as seen from the root (*T. karelinii*, ((*T. carnifex*, *T. macedonicus*), (*T. cristatus*, *T. dobrogicus*))), is not supported by molecular genetic analyses (ARNTZEN *et al.*, 2007).

As a result of that analyses, we forwarded the hypothesis that the speciation of crested newts was near-simultaneous at around 11-10 Ma and occurred in the Balkans (ARNTZEN *et al.*, 2007). The temporal calibration was obtained from a 24 Myr old fossil record ascribed to *T. marmoratus* (ESTES, 1981; STEINFARTZ *et al.*, 2007). We found external support for this hypothesis in the observations that genetic and species diversity is highest in the Balkans and that a polytomous speciation scenario matches with the palaeogeography of the Balkan region (POPOV *et al.*, 2004). It is not expected that more precise paleogeographic reconstructions will

become available soon, so we would suggest, on the contrary, that an established phylogeny - either resolved or polytomous - might actually assist the reconstruction of the paleogeography of south-eastern Europe.

In our previous work on the phylogeny and historical biogeography of the genus *Triturus* (Caudata: Salamandridae) we failed to resolve the basal node of the crested newt radiation (ARNTZEN *et al.*, 2007). Indeed, no firm phylogenetic resolution has been found among five crested newt lineages, other than the sister-species status of *T. carnifex* and *T. macedonicus* (see below). We here apply some of the above methods with a set of seven nuclear and mitochondrial genes, in an attempt to break the crested newt polytomy.

Material and methods

Sampling design

We sampled the five species of crested newts (including both subspecies of *T. karelinii*; *T. k. karelinii* and *T. k. arntzeni*) and one species of marbled newt as outgroup, with altogether 15 individuals from as many populations (Appendix 1). Because each species of crested newt is in spatial contact with two or more other species of the group (Fig. 1) and because hybridization and introgression is known to occur in limited areas (WALLIS and ARNTZEN, 1989), we sampled localities at least 100 km away from documented contact zones (ARNTZEN, 2003). This was done to minimize confounding effects of interspecific gene flow on the phylogenetic reconstruction. Sampling effort reflected the level of intraspecific genetic variation known for the species, which is low in *T. cristatus* (one locality sampled) and high in *T. karelinii* (six localities sampled).

DNA extraction and sequencing

Tail-tip tissue samples were stored in -80 ºC and later transferred to 95% ethanol. Total genomic DNA was extracted with the Qiagen DNeasy tissue kit. PCR conditions consisted of an initial denaturation step at 94 ºC for 4 min, followed by 35 cycles of 30 s denaturation at 94 ºC, 45 s annealing at 57-68 ºC (depending on the fragment) and 90 s extension at 72 ºC and a final extension step of 4 min at 72 ºC. The molecular markers employed are two protein-coding mitochondrial genes (subunits 2 and 4 of the NADH dehydrogenase gene complex - ND2 and ND4) and

five nuclear genes, namely three introns (β-Fibrinogen intron 7 - βfib*int*7, Calreticulin intron C - *Calint*C, Platelet-derived growth factor receptor α - Pdgfrα) and two anonymous markers (Cri1 and Cri4) (ESPREGUEIRA THEMUDO *et al.*, in press). The primers used are described in Table 1. To increase accuracy of the results, both DNA strands were sequenced. Sequences were obtained commercially through Macrogen Inc. BLAST searches at Genbank confirmed the homology of the nuclear introns and the mtDNA genes with published data, while the anonymous nuclear markers remained unidentified.

Data analysis

Chromatograms were checked for basecalling errors with Chromas software (Technelysium Pty Ltd). Sequences were aligned in BioEdit (HALL, 1999). Insertion/deletion polymorphisms (indels) reflecting heterozygosity were reconstructed by comparing chromatograms for the forward and reverse primers (following FLOT *et al.*, 2006). Indels were excluded from the phylogenetic analysis. Because of difficulties in analyzing individuals which are heterozygous for more than one nuclear gene, phased haplotypes were combined into a consensus haplotype by coding polymorphic sites with the IUPAC ambiguity codes (IUPAC, 1974). The partition homogeneity test (or Incongruence length difference test, FARRIS *et al.*, 1994) as implemented in PAUP 4.0b10 (SWOFFORD, 2003) was used to check for congruence in phylogenetics signal among genes. Phylogenetic congruence was rejected for all pairs of markers, except the two mitochondrial genes. Nucleotide composition, transition/transversion ratio and sequence divergence under the Kimura 2-parameter model were calculated with MEGA4 (TAMURA *et al.*, 2007). Translation of coding sequences to amino acids did not reveal any stop codons, validating the authenticity of the mtDNA sequences. To quantify the amount of nucleotide substitution saturation and test if the data was compatible with a molecular clock model, we used DAMBE (XIA and XIE, 2001). We used the Phi test (BRUEN *et al.*, 2006) to check for intragenic recombination. For each of the genes the most appropriate model of nucleotide substitution was selected with MrModelTest (NYLANDER, 2004) as evaluated with the Akaike Information Criterion.

Phylogeny reconstruction

Phylogeny reconstruction was carried out with phylogenetic network, non-

hierarchical and hierarchical Bayesian approaches.

Under the non-hierarchical Bayesian approach, phylogenetic trees were estimated with MrBayes v.3.1. (RONQUIST and HUELSENBECK, 2003). Following the conditional combination method described in the introduction, only the two mitochondrial genes should be concatenated. However, in order to be able to compare the results over the three methods, we still estimated trees from all five nuclear genes concatenated, and from the combined nuclear and mitochondrial genes. For each analysis, the software ran for 25 million generations, with a sample taken every 1000 generation, in two parallel runs of four Metropolis Coupled Monte Carlo Markov Chains (one cold and three incrementally heated). Stabilization of the likelihood occurred well before the runs were terminated. Convergence between runs was good, with an average standard deviation of split frequencies < 0.01 (RONQUIST and HUELSENBECK, 2003). The first quarter of the trees was discarded as burn-in and the remaining trees were summarized under the majority-rule consensus criterion.

Phylogenetic networks were constructed using the Neighbor-Net algorithm as implemented in Splitstree v.4.10 (HUSON and BRYANT, 2006) for 1) each of the five nuclear genes, 2) the five nuclear genes combined, 3) the two mtDNA genes combined, and 4) all seven genes combined. The statistical confidence in inferred nodes was assessed through 1000 bootstrap pseudo-replicates.

A hierarchical Bayesian method was performed with the software BEST 2.1 (LIU and PEARL, 2007). This programme implements the search strategy of Rannala and Yang (2003), which aims to reconcile phylogenetic signal across gene trees into a species tree. Individuals were assigned to a particular species or subspecies, a requirement of the software, based on morphology and geographical origin. Analyses were conducted for only the nuclear genes, and for the complete dataset. BEST was run with topology, branch lengths, gene mutation rates, gamma shape parameter, character state frequencies, transition/transversion ratio, substitution rates under the GTR model, and the proportion of invariable sites set to be unlinked across genes. The chain length used was 300 million generations, with one sample taken every 1000 generations. The last 10,000 trees sampled from the stationary phase were summarized under the 50% majority rule consensus criterion.

In this paper, we will use the term ‘species trees’ to define a phylogenetic

tree calculated from a method that acknowledges heterogeneity in gene trees (EDWARDS, 2009), whereas 'concatenation' is used to define the construction of a supermatrix containing all available data.

Molecular dating

Molecular clock estimates for salamandrids are available for the ND2 gene, but vary considerably. They are contained between the extremes 0.37 (MUELLER, 2006) and 0.64 (WEISROCK *et al.*, 2005) substitutions/site/100 Ma. Several other papers have used fossil calibrations and/or biogeographical data to estimate divergence times in the family Salamandridae. Steinfartz et al (2007) used a fossil calibration of 24 million years before present (Ma) at the root of *Triturus* (their calibration point C4). Zhang et al (2008) estimated the divergence of Salamandridae based on two sets of calibration: one with similar fossil calibrations as Steinfartz and another with an additional biogeographical calibration point for the split between Corsico-Sardinian *Euproctus* and continental *Triturus*. These two sets of calibrations gave an estimate for the root of *Triturus* of 26.6 and 11.3 (ZHANG *et al.*, 2008). We primarily followed the calibration point of 24 Ma of Steinfartz to date divergences on the species tree, and then compared the results to the other possible calibrations. The divergence times of the various lineages were determined from the relative branch lengths in the consensus tree by linear interpolation.

Results

Polymorphism and level of variation

The aligned sequence matrix (with indels removed) has a length of 4336 bp, with 694 variable positions of which 373 are parsimony-informative (Table 2). With 288 (77%) versus 85 (23%) variable positions, phylogenetic information content is higher in the two mitochondrial genes than in the five nuclear genes. The G-content was higher in nuclear genes (17.4 – 26.3 %) than in mitochondrial genes (13.3 – 16.4 %), as expected from the anti-G bias in mtDNA (ZHANG and HEWITT, 1996). Cri4 showed the lowest level of variation and ND2 and ND4 the highest. The level of nucleotide substitution saturation was classified as 'little' for each gene. No

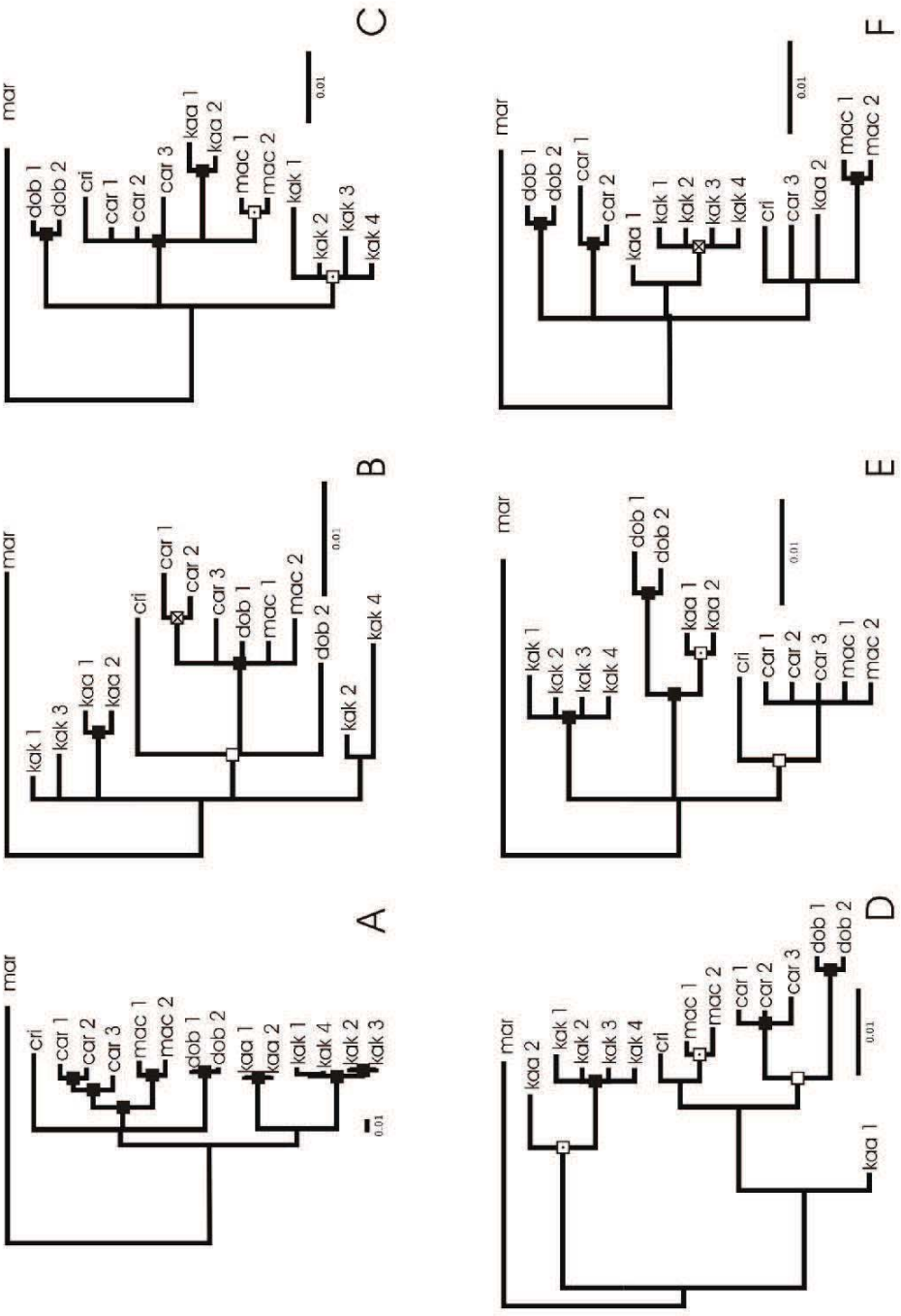

Figure 2 Results of a non-hierarchical Bayesian phylogenetic analysis for the genus *Triturus* from DNA sequence data with *T. marmoratus* as outgroup. The genes studied are of either mitochondrial origin, NADH dehydrogenase subunits 2 and 4 (A) or nuclear origin, β-fibrinogen intron 7 (B), calreticulin intron C (C), platelet-derived growth factor receptor α (D), cristatus anonymous locus 1 (E) and cristatus anonymous locus 4 (F). Posterior probabilities are classified as 1.0 (solid squares), no less than 0.99 (crossed squares), no less then 0.95 (dotted squares) and no less then 0.90 (open squares). Scale bar is in expected changes per site.

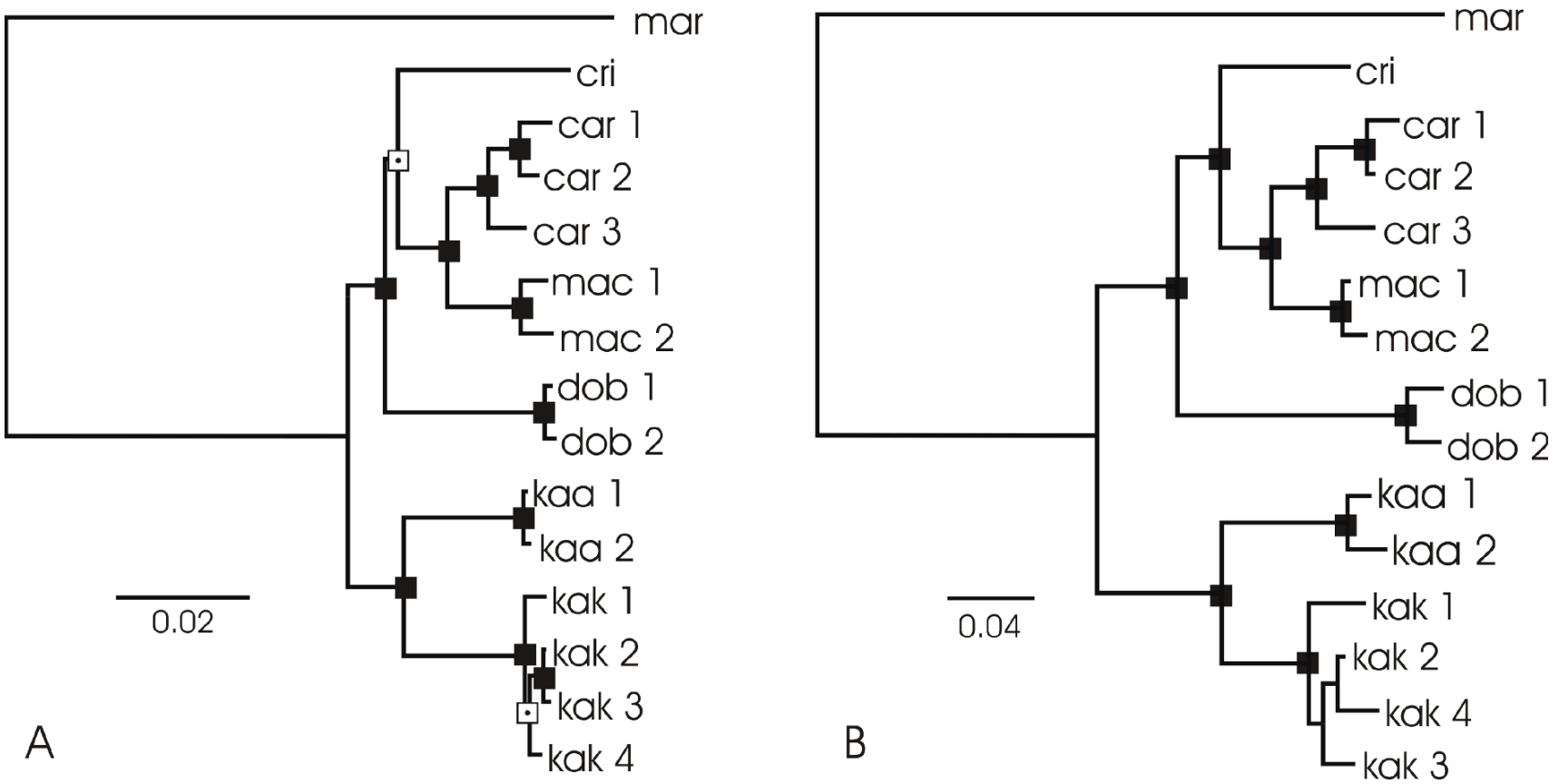

Figure 3 Results of a non-hierarchical Bayesian phylogenetic analysis of *Triturus* with a Total Evidence approach (A) and with a supermatrix of the nuclear genes (B). For details, see text. Scale bar is in expected changes per site.

significant signal for recombination was detected. The selected substitution models were F81+I for βfib*int*7, HKY+G for Cal*int*C, PDGFRα, Cri1 and Cri4, and a mixed model for the combined mtDNA genes: GTR+I+G for ND2 and GTR+I for ND4. Across nuclear genes heterozygosity ranged from 13 to 47%.

Phylogenetic reconstruction

Non-hierarchical Bayesian phylogeny

In the mtDNA gene tree all individuals are grouped into their nominal species and subspecies, with a Bayesian posterior probability (pp) of 1.0 (Fig. 2A). The mtDNA data gives strong support to the sister-group relationship of *T. carnifex* and *T. macedonicus* (pp=1.0) and weak to the two subspecies of *T. karelinii* (pp=0.60). *Triturus karelinii* is the sister species to all other crested newts, but the support for this configuration is weak (pp=0.78).

Across the nuclear genes, the results are not easy to reconcile, neither at the level of recognized species and subspecies, nor in terms of phylogenetic structure (Fig. 2 B-F). Nodes with high posterior probability (pp≥0.95) that also support nominal (sub)species are found for *T. carnifex* (Pdgfrα and Cri4), *T. macedonicus* (*Calint*C, Pdgfrα and Cri4), *T. k. arntzeni* (βfib*int*7, *Calint*C and Cri1) and *T. dobrogicus* and *T. k. karelinii* (all genes except βfib*int*7). Inter(sub)specific phylogenetic relationships supported with pp≥0.95 are *T. dobrogicus* and *T. k. arntzeni* by Cri1 (with exclusion of *T. k. karelinii*) and *T. carnifex*, *T. cristatus*, *T. macedonicus* and *T. k. arntzeni* by Cal*int*C (with exclusion of *T. k. karelinii*).

The combined nuclear tree (Fig. 3B) shows all nodes resolved with a pp of 1.0. *Triturus carnifex* and *T. macedonicus* are sister species. These two species are sister to *T. cristatus*, and *Triturus dobrogicus* is sister to this group. *Triturus. k. karelinii* and *T. k. arntzenii* are sister taxa. The tree resulting from the complete dataset differs from the one in Fig. 3B only by a lower support (pp=0.95) for the ((*T. carnifex* + *T. macedonicus*) + *T. cristatus*) clade (Fig. 3A).

Phylogenetic networks

Networks for the five nuclear genes separately show mixed results, with the occasional high support for the predefined taxa, such as *T. carnifex*, *T. dobrogicus* and *T. macedonicus* by Pdgfrα (results not shown). Conversely, individuals representing a nominal species might also be found scattered over the

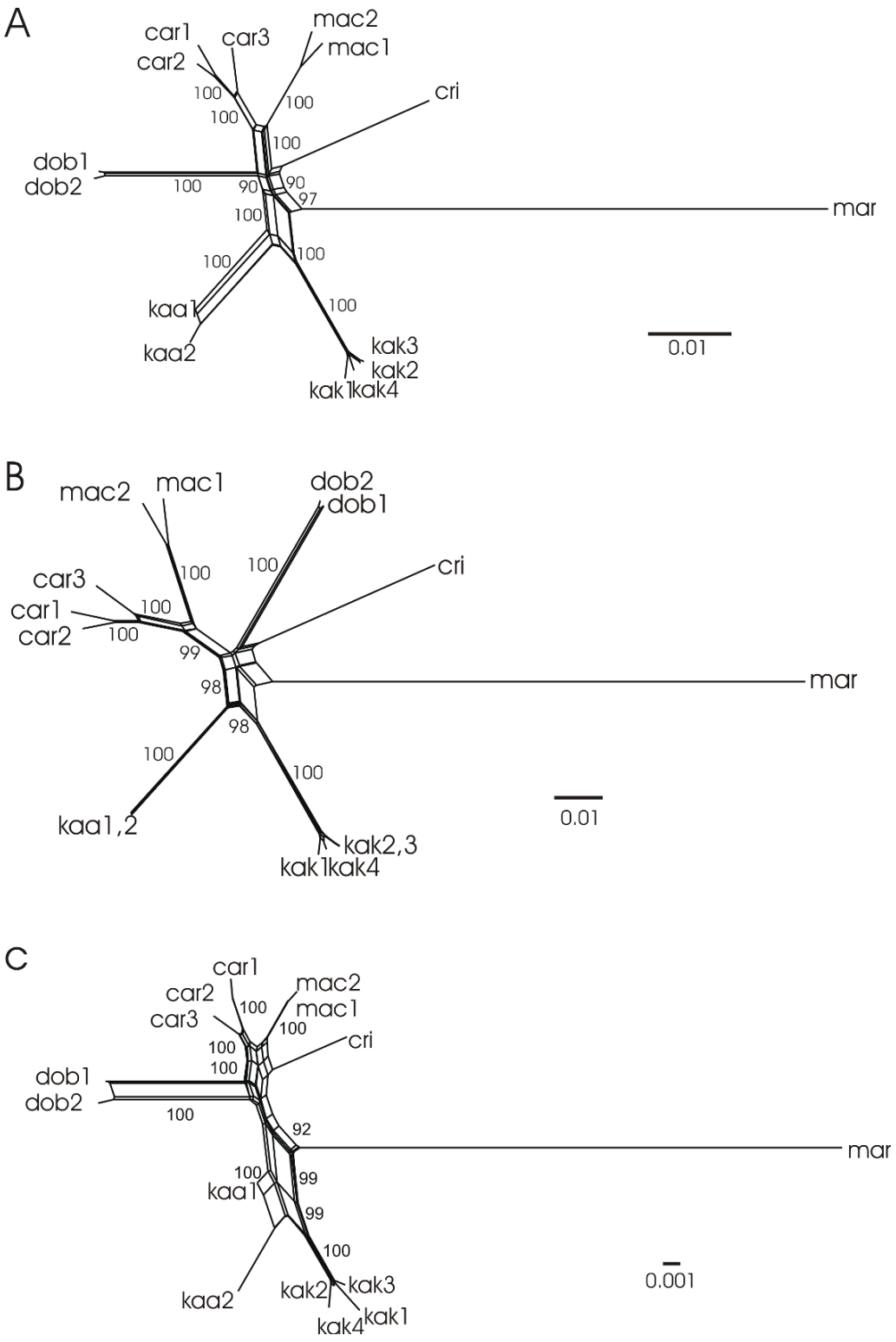

Figure 4 Results of a phylogenetic network analysis for the genus *Triturus* from DNA sequence data with *T. marmoratus* as outgroup. Results are shown for the concatenation of five nuclear genes (A), two mitochondrial genes (B) and all seven genes (c). Values on branches are bootstrap replication scores. Phylogenetic networks do not require congruence between fragments and conflict in the dataset is represented graphically through closed networks (or 'rings'), indicating that interspecific relationships are not fully resolved (Huson and Bryant, 2006).

network, such as *T. carnifex* for Cri4, *T. dobrogicus* for βfib*int*7 and *T. karelinii* for Cal*int*C, Cri1 and Cri4. The networks show generally low phylogenetic structure at the (sub)species level, with the exception of the gene PDGFRα, which supports a clade consisting of *T. carnifex*, *T. cristatus*, *T. dobrogicus* and *T. macedonicus* with a low bootstrap replication scores (brs) = 0.65 (results not shown). The phylogenetic network derived from the combined dataset resolves five species of crested newts in line with current taxonomy (Fig. 4A). Substructuring is found for *T. karelinii* with two groups that concur with current subspecies designations (*T. k. karelinii* and *T. k. arntzeni*). Bootstrap replication scores that would support phylogenetic structure among these taxa vary. *Triturus karelinii* is either monophyletic and differentiated from the other taxa (brs = 1.0), or paraphyletic (brs = 0.97) with just *T. k. karelinii* differentiated from all the other taxa. Moderate support is found for a clade consisting of *T. carnifex*, *T. cristatus*, *T. macedonicus* and *T. dobrogicus* (brs = 0.90). Within this clade, only the sister-group relationship of *T. carnifex* and *T. macedonicus* is strongly supported (brs = 1.0). A network based on the two mtDNA genes has the same topology but differs slightly in bootstrap replication scores (Fig. 4B). Based on the five nuclear genes, the network has the same branching order, with high support for the (sub)species, but low for interspecific relationships (Fig. 4C). The overall distances among taxa are shorter, due to the lower number of informative characters.

Hierarchical Bayesian phylogeny

The species tree is fully resolved (pp > 0.99), except for the position of *T. cristatus* (Fig. 5A). The supported clades are i) *T. carnifex* and *T. macedonicus*, ii) group i and *T. dobrogicus*, and iii) *T. k. karelinii* and *T. k. arntzeni*. When only the nuclear genes are used to build the species tree, the topology is identical, but pp values are lower (Fig. 5B).

Divergence times

The molecular clock test in DAMBE does not reject the null hypothesis of equidistance of all tips to the root of the tree. Linear interpolation on the species tree (Fig. 5A) over the 24 Myr window that covers the radiation of the genus *Triturus*, based on the fossil calibration (Fig. 5A) yields the following estimates of divergence times within the crested newt group: 11.2 Ma for the MRCA of the crested newts

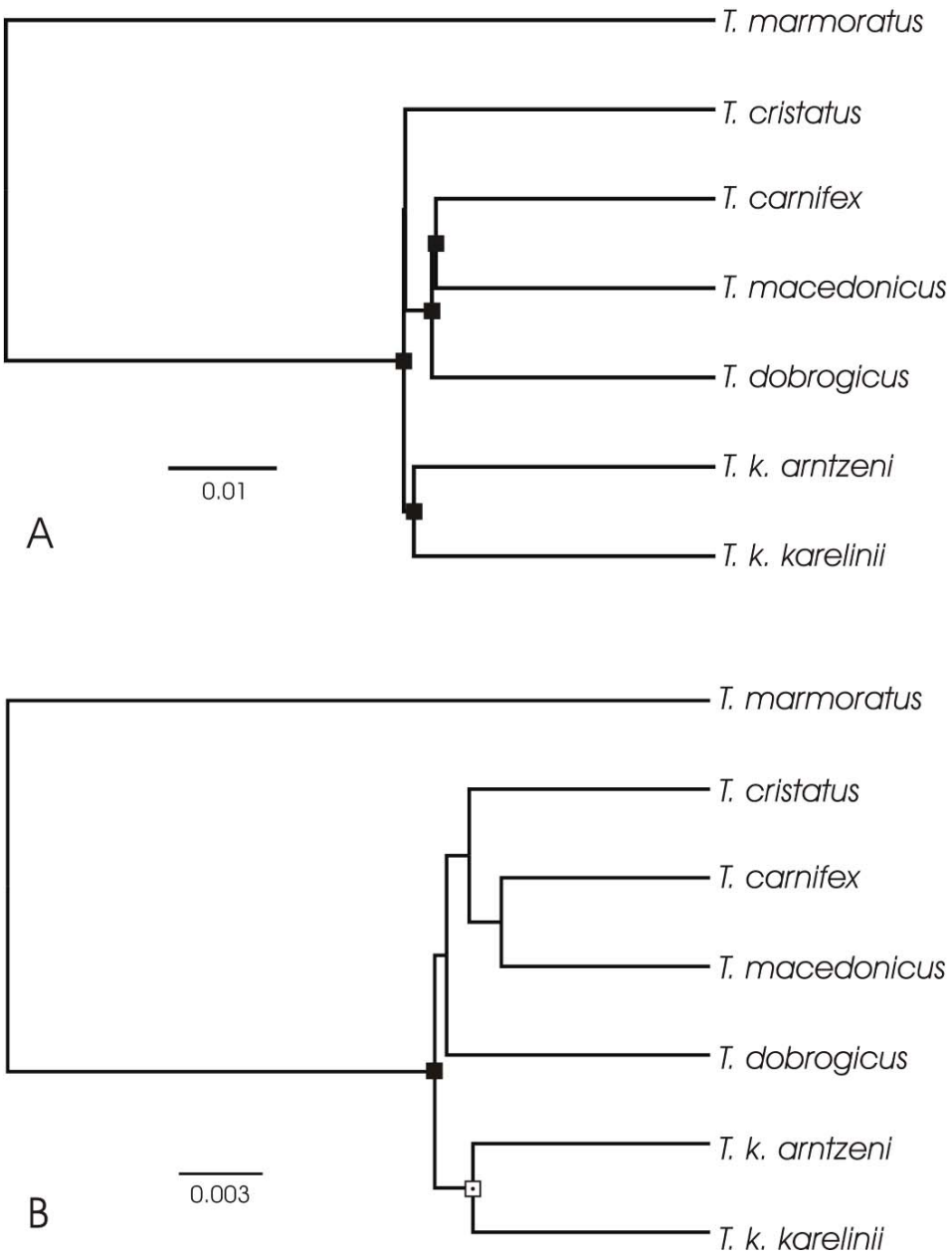

Figure 5 Results of a hierarchical Bayesian phylogenetic analysis for the genus *Triturus*, based upon DNA sequence data from two mitochondrial and five nuclear genes with *T. marmoratus* as outgroup (A); and an analysis with only the five nuclear genes (B). For details, see text. Posterior probabilities are coded as in Fig. 2. Scale bar is in expected changes per site.

(and, by consequence, for the branching off of *T. cristatus*, which stems from the root), 11.1 Ma for the divergence between *T. k. karelinii* and *T. k. arntzeni*, 8.0 Ma for the origin of *T. dobrogicus* and 7.9 Ma for the split between *T. carnifex* and *T. macedonicus*. As molecular clock calibrations for ND2 vary considerably, the time to the MRCA of crested newts is estimated at 14 – 28 Ma under these calibrations, which would multiply each of the above temporal estimates by a factor 1.25-2.5.

Discussion

Rapid radiations, the coalescence, mitochondrial versus nuclear genes, and concatenation

The evolutionary history of rapid radiations is difficult to resolve. During the short timespan between branching events, there is limited opportunity for phylogenetic signal to accumulate. Furthermore, the older the radiation, the more difficult it will be to uncover the order of branching events, as the relative amount of informative phylogenetic change decreases through time (i.e. anagenesis masks cladogenesis). In order to resolve such apparent polytomies, large amounts of data will generally be required.

Coalescence is the merging of genetic lineages backwards in time to a most recent common ancestor. Forward in time, the sorting process taking place during diversification from the ancestral lineages to the pattern observed today is a chance process. Because of this stochasticity, and because lineage sorting may be incomplete, gene trees will generally be different from one another and frequently not represent the species tree (AVISE and WOLLENBERG, 1997). The only way out of this dilemma is to study multiple genes because the more genes involved, the more convincing the convergence towards the species tree will be. By studying many genes together, the 'gene tree versus species tree' problem should ultimately be eliminated.

Animal molecular phylogenetics, in former times, largely relied on just the mitochondrial genome, with two main drawbacks. First, because of its small size (c. 16 kb in vertebrates) it is quickly exhausted in terms of data yield compared to the nuclear genome (for comparison Triturus newts have a nuclear genome size of c. 25 Gb, GREGORY, 2008, even though a large portion is repetitive DNA and phylogenetically not-informative). Second, the mitochondrial genes are linked. The

molecule is inherited as a single unit and mitochondrial gene trees are, hence, not independent. Indeed, recent studies on the phylogeny of the Salamandridae provide similar results whether based upon a small number of representative genes (STEINFARTZ *et al.*, 2007) or on the complete mitochondrial genome (ZHANG *et al.*, 2008). In practice, this means that nuclear genes have to be used to get independent estimations. Unfortunately, primers for nuclear genes will often have to be newly developed (for Triturus see CHAPTER 3).

Multiple independent markers solve the phylogeny of Triturus *newts*

The non-hierarchical Bayesian phylogenetic reconstruction yielded five nuclear gene trees and one mitochondrial tree that were all different from each other, to the extent that and a consensus tree would not be informative. A striking result is that individual sequences do group according to predefined species and subspecies in the mtDNA-based tree, while for nuclear genes this is rarely the case. Possible explanations are i) better phylogenetic resolution due to a larger number of informative characters and ii) lineage sorting has progressed more rapidly, due to a fourfold smaller effective population size (MOORE, 1995). Either way, the observed species level consistency in the individual phylogenetic placement suggests the absence or low frequency of horizontal gene exchange, which we had aimed for through a thoughtful sampling scheme, relying on the well documented species ranges (ARNTZEN, 2003).

The total evidence tree is completely resolved, and is identical to the combined nuclear tree, except for the lower statistical support for the position of *T. cristatus* (Fig. 3B). Compared to the total evidence tree, the mtDNA tree has the basal branches unresolved (Fig. 2A). This means that the nuclear data contributes significantly to the resolution of the species tree, even though this is not obvious from a gene by gene perspective.

The phylogenetic relationships supported by our 1747 bp of mtDNA data are the sister-group status of *T. carnifex* and *T. macedonicus* and *T. k. karelinii* and *T. k. arntzeni* (with pp > 0.99), while moderate support was for the clade including all crested newts that are not *T. karelinii* (pp=0.78). The basal dichotomy between *T. karelinii* versus non-*T. karelinii* was not found by earlier, smaller molecular datasets (e.g. WALLIS and ARNTZEN, 1989; MACGREGOR *et al.*, 1990). Based upon 839 bp of cytochrome-b sequence data, Steinfartz et al. (2007) described the paraphyly of *T.*

karelinii, whereas Arntzen et al. (2007) found no resolution at the basal *Triturus* node, from either 40 allozyme loci or 642 bp of ND4 mtDNA. The two *T. karelinii* subspecies (*arntzeni* and *karelinii*) are almost as old as the *T. karelinii* lineage itself.

In the phylogenetic network analysis, pronounced ambiguities in the total evidence approach are the unresolved branching order (*T. dobrogicus*, *T. cristatus,* (*T. carnifex*, *T. macedonicus*) and the position of the *T. karelinii* subspecies as either monophyletic or paraphyletic (Fig. 4C). The next highest uncertainty refers to the monophyly of the group *T. carnifex*, *T. cristatus*, *T. dobrogicus* and *T. macedonicus* and, following that, the monophyly of *T. carnifex* and *T. macedonicus*.

In contrast with the above results, the phylogenetic tree obtained with the hierarchical Bayesian method is fully resolved, except for the position of *T. cristatus* (Fig. 5A). The newly obtained phylogenetic resolution supports the notion that the radiation of crested newts lineages occurred over a relatively short period. This window of radiation, dated at 11-8 Ma (or 14-10 Ma to 28-20 Ma using the varying ND2 molecular clocks cited above), yielded all six different lineages that we observe today.

This study provides empirical support for the claim (EDWARDS *et al.*, 2007; EDWARDS, 2009) that the method, by accounting for the differential coalescent process across independent markers, estimates the phylogenetic relationships between recently diverged taxa with high precision. Statistical support expressed by posterior probability values is high. This is a remarkable improvement over competing methods (Fig. 2 and 4). We only found a fully bifurcating and supported tree with the non-hierarchical method by concatenating incongruent markers, thus violating the prerequisites of conditional combination. A potential drawback of the hierarchical method is that additional information should be provided in the form of species identification of the individuals included. Crested newt species show distinct morphological characteristics, making species assignment straightforward. Furthermore, our sampling scheme was designed to exclude areas if interspecific gene flow. Indeed, none was detected in the mtDNA marker (Fig. 2A). Despite this, distinguishing between interspecific gene flow and incomplete lineage sorting is a complex issue (AVISE *et al.*, 1983; SANG and ZHONG, 2000; see also HEY and NIELSEN, 2004). We are currently exploring this area of research in *Triturus* newts, by sampling from areas potentially subject to introgression, i.e., within 100 km of

described range borders (G. Espregueira Themudo, A. Bickham Baird & J. W. Arntzen, in prep.).

As predicted by Edwards (2009), statistical support in the species tree from the hierarchical method is less than the with concatenation. The theoretical advantages of the hierarchical method over data concatenation, however, make us believe that the outcome of the hierarchical Bayesian method is more reliable.

Concatenation is also prone to "data swamping", where one or a few of the partitions provides most of the information retained in the phylogenetic tree. This does not seem to be problematic here, as we have shown that the nuclear genes are contributing significantly to the resolution of the total evidence tree (Fig. 3A and B). In the hierarchical method, the effect of the number of informative sites per locus is still not clear (EDWARDS, 2009).

Phylogenetic position and taxonomic status of T. karelinii

Marked levels of genetic differentiation within *T. karelinii* were first apparent from mtDNA, as either monophyletic (WALLIS and ARNTZEN, 1989) or paraphyletic (STEINFARTZ *et al.*, 2007). Interestingly, the phylogenetic network gives high bootstrap support to both the monophyly and the paraphyly of *T. karelinii* (Fig. 4A). This type of uncertainty, expressed in so called phylogenetic 'spider-webs', may be indicative of reticulate evolution, c.q. hybridization (HUSON and BRYANT, 2006).

Based on differences in genome size, protein variation and morphological characteristics, Litvinchuk et al. (1999) described a Balkan population as the representative of a distinct subspecies, *Triturus karelinii arntzeni*. The present study suggest that the two taxa are genetically distinct, at the same level or deeper than other *Triturus* species are to each other, with a date of origin close to the basal radiation of the crested newt group (Fig. 4). Considering their long independent evolutionary history, we suggest that the two taxa should be raised to full species status, i.e., *T. karelinii* (Strauch, 1870) and *T. arntzeni* Litvinchuk, Borkin, Džukić and Kalezić, 1999 (in LITVINCHUK *et al.*, 1999).

Triturus arntzeni and *T. karelinii* may have their range limits at either side of the Bosporus (RAFFAËLLI, 2007), but geological information does not support this view, since the current connection of the Black Sea to the Mediterranean opened a mere 5000 years ago (RYAN *et al.*, 2003). Other data suggest that the species range border could be either in Europe or in Asia. A population from Istanbul (at the

European side of the Bosporus) is classified as *T. karelinii* on the basis of genome size, that is larger in *T. karelinii* than in *T. arntzeni* (LITVINCHUK *et al.*, 1999), whereas mtDNA data suggest that the ranges are in contact around Adapazari in northwestern Anatolia (WALLIS and ARNTZEN, 1989). This region is a highly active tectonic area, with major fault lines going through the Sakarya river valley, called the Adapazari – Karasu Fault, that developed since the Miocene (ELMAS, 2003).A wider analysis on the spatial distribution of morphological and genetic variation in the *T. karelinii* group is underway (K. Olgun et al., in prep.; B. Wielstra et al., in prep.)

Acknowledgments

This work was funded by a PhD grant (Ref. SFRH/BD/16894/2004) to GET from FCT (Fundação para a Ciência e a Tecnologia, Portugal). We thank W. Babik, Ö. Güçlü, K. Olgun, N. Poyarkov and G. Wallis for help in collecting material, and A. Bickham Baird, A. Larson, G. Wallis and two anonymous reviewers for useful comments to earlier versions of this paper.

References

ARNTZEN, J.W. (2003) *Triturus cristatus* Superspecies–Kammolch-Artenkreis. In: *Handbuch der Reptilien und Amphibien Europas. Schwanzlurche IIA* (eds. GROSSENBACHER, K., THIESMEIER B.), pp. 421-514. Aula-Verlag, Wiebelsheim.

ARNTZEN, J.W., ESPREGUEIRA THEMUDO G. and 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. *Contributions to Zoology* **76**, 261-278.

AVISE, J.C., SHAPIRA J.F., DANIEL S.W., AQUADRO C.F. and LANSMAN R.A. (1983) Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Molecular Biology and Evolution* **1**, 38-56.

AVISE, J.C. and WOLLENBERG K. (1997) Phylogenetics and the origin of species. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7748-7755.

BININDA-EMONDS, O.R.P., CARDILLO M., JONES K.E.*, et al.* (2007) The delayed rise of present-day mammals. *Nature* **446**, 507-512.

BRUEN, T.C., PHILIPPE H. and BRYANT D. (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665-2681.

CRNOBRNJA-ISAILOVIC, J., DZUKIC G., KRSTIC N. and KALEZIC M.L. (1997) Evolutionary and paleogeographical effects on the distribution of the *Triturus cristatus* superspecies in the central Balkans. *Amphibia-Reptilia* **18**, 321.

DE QUEIROZ, A. (1993) For consensus (sometimes). *Systematic Biology* **42**, 368-372.

EDWARDS, S.V., LIU L. and PEARL D.K. (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5936-5941.

EDWARDS, S.V. (2009) Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1-19.

ELMAS, A. (2003) Late Cenozoic tectonics and stratigraphy of northwestern Anatolia: the effects of the North Anatolian Fault to the region. *International Journal of Earth Sciences* **92**, 380.

ESPREGUEIRA THEMUDO, G., BABIK W. and ARNTZEN J.W. (in press) A combination of techniques proves useful in the development of nuclear markers in the newt genus *Triturus*. *Molecular Ecology Notes*.

ESTES, R. (1981) Gymnophiona, Caudata. In: *Handbuch der Palaoherpetologie, Part 2* (ed. WELLNHOFER, P.), pp. 1-15. Gustav Fischer, Stuttgart.

FARRIS, J.S., KALLERSJO M., KLUGE A.G. and BULT C. (1994) Testing significance of incongruence. *Cladistics* **10**, 315-319.

FLOT, J.-F., TILLIER A., SAMADI S. and TILLIER S. (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**, 627-630.

GREGORY, T.R. (2008) Animal Genome Size Database. http://www.genomesize.com.

HALL, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

HEY, J. and NIELSEN R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.

HUELSENBECK, J.P., BULL J.J. and CUNNINGHAM C.W. (1996) Combining data in

phylogenetic analysis. *Trends in Ecology & Evolution* **11**, 152-158.

HUSON, D.H. and BRYANT D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267.

IUPAC (1974) Abbreviations and symbols for nucleic acids, polynucleotides and their constituents (Rules approved 1974). *Pure and Applied Chemistry* **40**, 277-331.

KLUGE, A.G. (1989) A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* **38**, 7-25.

LITVINCHUK, S.N., BORKIN L.J., DŽUKIĆ G.*, et al.* (1999) Taxonomic status of *Triturus karelinii* on the Balkans, with some comments about other crested newt taxa. *Russian Journal of Herpetology* **6**, 153-163.

LIU, L. and PEARL D.K. (2007) Species trees from gene trees: reconstructing bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology* **56**, 504-514.

MACGREGOR, H.C., SESSIONS S.K. and ARNTZEN J.W. (1990) An integrative analysis of phylogenetic relationships among newts of the genus *Triturus* (family Salamandridae), using comparative biochemistry, cytogenetics and reproductive interactions. *Journal of Evolutionary Biology* **3**, 329-373.

MIYAMOTO, M.M. and FITCH W.M. (1995) Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* **44**, 64-76.

MOORE, W.S. (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**, 718-718.

MUELLER, R.L. (2006) Evolutionary rates, divergence dates, and the performance of mitochondrial genes in bayesian phylogenetic analysis. *Systematic Biology* **55**, 289-300.

NYLANDER, J.A. (2004) MrModelTest2. Computer program distributed by the author. Evolutionary Biology Centre, Uppsala University, Sweden.

POPOV, S.V., RÖGL F., ROZANOV A.Y.*, et al.* (2004) Lithological-paleogeographic maps of Paratethys: 10 maps Late Eocene to Pliocene. *Courier Forschunginstitut Senckenberg* **250**, 1-46.

RAFFAËLLI, J. (2007) *Les Urodèles du monde* Penclen édition.

RANNALA, B. and YANG Z. (2003) Bayes estimation of species divergence times and

ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645-1656.

RONQUIST, F. and HUELSENBECK J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

RYAN, W.B.F., MAJOR C.O., LERICOLAIS G. and GOLDSTEIN S.L. (2003) Catastrophic Flooding of the Black Sea. *Annual Review of Earth and Planetary Sciences* **31**, 525-554.

SANG, T. and ZHONG Y. (2000) Testing Hybridization Hypotheses Based on Incongruent Gene Trees. *Systematic Biology* **49**, 422 - 434.

STEINFARTZ, S., VICARIO S., ARNTZEN J.W. and CACCONE A. (2007) A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary patterns of the Salamandridae with emphasis on *Triturus* newts. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **308B**, 139-162.

SWOFFORD, D.L. (2003) PAUP*: Phylogenetic Analysis Using Parsimony (and other methods) 4.0 Beta 10. Sunderland (MA): Sinauer Associates.

TAMURA, K., DUDLEY J., NEI M. and KUMAR S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution* **24**, 1596.

WALLIS, G.P. and ARNTZEN J.W. (1989) Mithocondrial-DNA variation in the crested newt superspecies: limited cytoplasmic gene flow among species. *Evolution* **43**, 88-104.

WEISROCK, D.W., HARMON L.J. and LARSON A. (2005) Resolving deep phylogenetic relationships in salamanders: analyses of mitochondrial and nuclear genomic data. *Systematic Biology* **54**, 758-777.

XIA, X. and XIE Z. (2001) DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. *Journal of Heredity* **92**, 371-373.

ZHANG, D.X. and HEWITT G.M. (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution* **11**, 247-251.

ZHANG, P., PAPENFUSS T.J., WAKE M.H., QU L. and WAKE D.B. (2008) Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution* **49**, 586–597.