How did subterranean amphipods cross the Adriatic Sea? Phylogenetic evidence for the dispersal-vicariance interplay mediated by marine regression-transgression cycles

Running title: Biogeography of transadriatic niphargids

Teo Delić¹*, Fabio Stoch², Špela Borko¹, Jean-François Flot²,³, Cene Fišer¹

¹SubBio Lab, Department of Biology, Biotechnical faculty, University of Ljubljana, Slovenia
²Evolutionary Biology & Ecology, Université libre de Bruxelles (ULB), Belgium
³Interuniversity Institute of Bioinformatics in Brussels – (IB)², Brussels, Belgium

*corresponding author

Acknowledgments: The authors are thankful to Simone Cianfanelli, Mauro Giardini, Salvatore Inguscio, Branko Jalžić, Francesco Lamanna, Tonči Radu, Mauro Rampini, Stefano Taiti, Maja Zagmajster, David Škufca, Gregor Bračko, Valerija Zakšek, Marko Lukić, Jana Bedek, Martina Pavlek, Boris Sket, Andrej Drevenšek and Peter Trontelj who accompanied the authors during the field work or contributed Niphargus samples. We thank Vedran Jalžić for the photography of N. pectecoronata. The study was funded by the Slovenian Research Agency (program P1-0184, project N1-0069, postdoc project Z1-9164 to Teo Delić and a PhD Grant to Špela Borko) and the Belgian Fonds de la Recherche Scientifique (research credit J.0272.17 to Jean-François Flot).

The authors declare they do not have any conflict of interest.
Abstract

**Aim:** We tested the hypothesis that historical marine regression-transgression cycles shaped the distribution patterns of subterranean amphipods through repeated cycles of dispersal and vicariance against the hypothesis that subterranean amphipods colonized both sides of the Adriatic Sea independently.

**Location:** Western Balkan Peninsula, Adriatic Sea Islands and Apennine Peninsula, Europe

**Taxon:** Genus *Niphargus*, a clade of freshwater subterranean amphipods (Crustacea: Amphipoda).

**Methods:** The taxonomic structure of the studied clade was revised using unilocus species delimitation methods. The timeframe of cladogenetic events was inferred using a multilocus, time-calibrated phylogeny and compared to the main regression-transgression events in the Miocene and Pleistocene. The geographic origin of the studied clade, species’ range expansions and contractions, as well as vicariance events were assessed through modelling of historical biogeography.

**Results:** Subterranean amphipods of the genus *Niphargus*, found on both sides of the Adriatic Sea, form a monophylum. The reconstructions of ancestral ranges suggest that the clade emerged in the Balkan Peninsula, dispersed three times independently to the Apennine Peninsula and once back to the Balkans. Adriatic Islands were colonized multiple times, predominantly from the Balkan Peninsula. The dispersal-vicariance events correspond to historical regression-transgression cycles in Miocene and Pleistocene.

**Main conclusions:** Marine regression-transgression cycles apparently shaped the distribution patterns of subterranean amphipods, while the alternative hypothesis received no support. The actual distribution of subterranean faunas apparently reflects old biogeographic events.

**Keywords:** amphipod, Apennine Peninsula, Dinaric Karst, dispersal, land-bridge, *Niphargus*, subterranean, transadriatic, vicariance
Introduction

Marine regressions and transgressions have shaped many biodiversity patterns of terrestrial and freshwater faunas (Woodruff, 2003; Chamberland et al., 2018; Hou & Li, 2018; Desiderato et al., 2019). During marine regressions, when sea levels drop, new dispersal routes emerge connecting islands with each other and the mainland (Woodruff, 2003; Gargani & Rigollet, 2007). By contrast, marine transgressions with deep incursions of salty water into coastal lowlands terminate dispersal routes, reshape island sizes and archipelago structures, modify local climates and alter marine currents (Fernández-Palacios et al., 2016). From a biogeographic perspective, marine regression-transgression cycles correspond to respective periods of dispersal and vicariance, which can plausibly explain the distribution patterns of some clades along the coastal regions (Yang, Hou, & Li, 2013; Cánovas et al., 2016; White, Reimer, & Lorion, 2016; Liu, Li, Ugolini, Momtazi, & Hou, 2018; Desiderato et al., 2019).

However, biologists studying subterranean species have proposed another, different role of marine regression-transgression cycles on faunas. Fluctuations of sea level presumably play an essential role in the colonization of the continental groundwater from ancient shallow seas. Briefly, ancestral marine benthic species colonized the marine interstitial, spread into the continent during transgressions, established contact with, and became stranded in subterranean freshwater during marine regressions (Stock, 1980; Notenboom, 1991). This scenario, as elaborated with several alternatives (Boutin & Coineau, 1990; Holsinger, 1994, 2000), has been used as an explanation for the origin and distribution of many subterranean taxa (Coineau, 1994; Boulanour, Yacoubi, Messouli, & Coineau, 1995; Kupriyanova et al., 2009; Sket & Zakšek, 2009; Baratti, Filippelli, Nardi, & Messana, 2010; Page et al., 2016). By contrast, the more classical view of marine regressions-transgressions as an alternation between processes of dispersal and vicariance has received little attention in subterranean biology, most likely due to the presumed low dispersal capacity of subterranean organisms (Trontelj et al., 2009; Bregović, Fišer, & Zagmajster, 2019). Few studies to date have invoked this scenario in order to explain species distributions (Bauzà-Ribot, Jaume, Fornós, Juan, & Pons, 2011; Cánovas et al., 2016; Guy-Haim, Simon-Blecher, Frumkin, Naaman, & Achituv, 2018; Stokkan et al., 2018), and none of these studies explicitly modelled biogeographic patterns using iterative cycles of regression and transgression.

In this study, we investigated whether regression-transgression cycles might explain the distribution of clades of subterranean amphipods in the Mediterranean region. We studied the amphipod genus *Niphargus*, a species-rich group distributed in the Western Palearctic (Fišer, 2019). As a freshwater taxon, *Niphargus* represents an appropriate model system for studying the role of the marine regression-transgression cycles. In North-Western Europe, *Niphargus* colonized freshwater during the Cretaceous, and the Mediterranean clades presumably derived from freshwater ancestors (McInerney et al., 2014). Today virtually all *Niphargus* species live in subterranean freshwater, with few exceptions found in anchialine caves and rivers (Karaman & Sket, 1989; Sket & Karaman, 1990; Copilaș-Ciocianu, Fišer, Borza, & Petrușek, 2018). The closest relative of *Niphargus* is the freshwater and anchialine genus *Pseudoniphargus* (Jurado-Rivera et al., 2017; Copilaș-Ciocianu, Borko, & Fišer, 2020), whereas their putative marine relatives are not known. Early taxonomic works and phylogenies identified some species groups distributed on both sides of the Adriatic Sea on the Apennine and Balkan Peninsulas as well as on islands in the Adriatic Sea (Ruffo & Vigna Taglianti, 1967; Fišer, Sket, & Trontelj, 2008; Ntakis, Anastasiadou, Zakšek, & Fišer, 2015; Delić, Švara, Coleman, Trontelj, & Fišer, 2017), making it likely that these clades experienced dispersal and/or vicariance due to regression-transgression cycles.

The Balkan Peninsula is the older of the two peninsulas. The western part of the Balkans is a carbonate mountainous region, the Dinaric Karst. The uplift of the Dinaric Karst started approximately 30 million years (My) ago, resulting in a major landmass in South-Eastern Europe (Rögl, 1998; Popov et al., 2004;
Park, 2014). By contrast, the Apennine Peninsula emerged as a group of islands approximately 20 My ago. The shape of the modern Apennine Peninsula was first attained some 10 My ago (Rögl, 1998; Popov et al., 2004). The first land-bridges between the two peninsulas were ephemeral archipelagos that connected the central part of the Dinaric Karst with Apulia about 13-15 My ago (Mazza & Rustioni, 2008). Later on, the two peninsulas were connected with at least two extensive land-bridges. The older land-bridge dates back to the Miocene, when the orogenic uplift of Gibraltar arc seaway triggered the so-called Messinian Salinity Crisis (5.96-5.33 My ago) (García-Castellanos and Villasenor, 2011). As a consequence, large parts of the Mediterranean, including the entire Adriatic basin, completely dried up (Gargani & Rigollet, 2007). The younger and the last existing land-bridges connecting both sides of the Adriatic Sea, date to the Pleistocene, when glacial maxima caused drying of the northern half of the Adriatic Sea (Correggiari, Roveri, & Trincardi, 1996).

The distribution of *Niphargus* species on both sides of the Adriatic Sea can be explained by two hypotheses. The first one is based on the generally low dispersal capacity of subterranean organisms. Species on both sides of the Adriatic Sea originated through single colonization of the ancient landmasses, with no subsequent dispersal across the Adriatic basin. If so, we would expect that these species originated at the contact of both peninsulas, nowadays Northern Italy. Furthermore, lineages found on the Apennine Peninsula cannot be older than 10 My, and – because dispersal is unlikely – present a single, monophyletic lineage, sister to the lineages from the Balkans. The alternative hypothesis states that the distribution of *Niphargus* species on both sides of the Adriatic Sea was shaped by dispersal and vicariance, corresponding to regression-transgression cycles. This hypothesis relaxes the prediction of a maximum clade age of 10 My, and allows these species to emerge in the older Dinaric Karst. In addition, this hypothesis predicts that the ancestors of the modern species crossed the Adriatic basin and colonized the Apennine Peninsula during one or more marine regressions, when the land-bridges connecting the two peninsulas were formed. Hence, it can be expected that species from both sides of the Adriatic Sea belong to several clades. To test these predictions, we performed the present study in three steps. First, we revised the taxonomic structure of the so-called transadriatic clade using molecular species delimitation methods. Second, we assessed the number and the timeframe of cladogenetic events among molecularly defined species by reconstructing a multilocus calibrated phylogeny. Finally, using spatial information and the calibrated phylogeny, we assessed where the studied species originated from and analysed the dispersal-vicariance events. We employed reconstruction methods and modelled the historical biogeography of these species using a series of models, which accounted for anagenetic (dispersal) and cladogenetic (vicariance) processes. We identified four major dispersal events, three from the Dinaric to the Apennine region, and one in the reverse direction.

**Materials and Methods**

**Taxon sampling and DNA isolation**

Taxon sampling aimed to include a wide range of *Niphargus* species, distributed from Ireland to Iran, with comprehensively sampled species distributed on the both sides of the Adriatic Sea. The ingroup sampling included also the genera nested within the genus *Niphargus*, namely *Carinurella* Sket, 1971; *Chaetoniphargus* Karaman & Sket, 2019; *Haploginglymus* Mateus & Mateus, 1958 and *Niphargobates* Sket, 1981. The outgroup comprised five species of *Pseudoniphargus* Chevreux, 1901 (based on Copiaș-Ciocianu, Borko, & Fišer, 2020). We sampled altogether 214 localities and 494 individuals, and stored
them in 96% ethanol. The remaining parts of the samples are deposited in the Zoological collection of the Department of Biology, Biotechnical Faculty, University of Ljubljana, Slovenia, in the collection of the Evolutionary Biology and Ecology Unit of the Université libre de Bruxelles (ULB), Belgium, and in the collection of the Natural History Museum of Verona, Italy. All the information on the specimens used in the analyses, collection sites, DNA vouchers and museum collections are available in Supplementary Information 1.

Each specimen was identified to the closest morphologically resembling morphospecies using original species descriptions. Then, genomic DNA was isolated from one of the pereiopods (thoracic appendages) using the GenElute Mammalian Genomic DNA (Sigma–Aldrich, USA) in Ljubljana and the DNEasy DNA extraction kit (Qiagen, The Netherlands) in Brussels. The remaining part of the specimen was retained for further morphological studies. We amplified nuclear DNA – two parts of the 28S rRNA gene (28S rRNA I and 28S rRNA II), histone H3 (H3), and the mitochondrial cytochrome c oxidase subunit I (COI). A list of primers and PCR amplification protocols used is available in Supplementary Information 2 and 3. In Ljubljana, PCR products were purified using Exonuclease I and FastAP (Thermo Fisher Scientific Inc., USA) according to the manufacturer’s instructions, and sequenced in both directions by Macrogen Europe (Amsterdam, The Netherlands), using the amplification primers; the resulting chromatograms were assembled and edited using Geneious 8.1.9. (Biomatters, New Zealand). In Brussels, PCR products were directly sequenced (without purification) in both directions by Genoscreen (Lille, France), and the resulting chromatograms were assembled and edited using Sequencher (GeneCodes, USA). Ambiguous nucleotide bases were coded using IUPAC nucleotide ambiguity characters.

**Taxonomic structure and molecular species delimitation**

In order to account for only partially resolved taxonomy and the possible presence of morphologically cryptic species, we revised the taxonomic structure of the study dataset prior to the main analyses. We assembled a dataset containing altogether 463 COI sequences of *Niphargus*, the genera nested within it (*Carinurella, Haploginglymus, Niphargobates*) and *Pseudoniphargus* (Supplementary Information 1). The sequences were acquired in this study (120 sequences) or retrieved from GenBank. Putative species structure, represented by the so-called molecular taxonomic operational units (hereafter MOTUs) was inferred using distance and tree-based delimitation methods, Automatic Barcode Gap Discovery (hereafter ABGD) (Puillandre, Lambert, Brouillet, & Achaz, 2012) and Poisson Tree Processes (hereafter PTP) (Zhang, Kapli, Pavlidis, & Stamatakis, 2013), respectively.

ABGD (Puillandre et al., 2012) assigns the sequences to the specific MOTUs based on the assumption that intraspecific genetic distances are smaller than interspecific ones, and without a priori species hypotheses. We ran this analysis on the complete COI dataset, using the ABGD web server [http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html](http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html) with default parameters, namely Kimura two-parameter substitution model, prior for maximum value of intraspecific divergence between 0.001 and 0.1, 10 recursive steps and a gap width of 1.0.

An alternative taxonomic structure was assessed using PTP (Zhang et al., 2013), a phylogeny-based method that delimits species at nodes where presumed intraspecific nucleotide substitution rates switch to interspecific substitution rates. The two nucleotide substitution rates are modelled using two different Poisson processes. For the PTP analyses, we removed duplicate sequences using a custom Perl script (Eme, Malard, Konecny-Dupré, Lefébure, & Douady, 2013), and ran the analysis using a dataset of 300 unique haplotypes. Phylogenetic relationships were estimated in two separate maximum likelihood analyses. The first analysis was performed using PhyML 3.1 (Guindon et al., 2010), relying on the GTR substitution model with estimated gamma shape parameter and proportion of invariant sites. In order
to assess the robustness of maximum likelihood tree, we repeated the analysis in IQ-tree 1.6.7 (Nguyen, Schmidt, von Haeseler, & Minh, 2015), using the invertebrate mitochondrial codon model (GY+F+R7) selected in ModelFinder (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, 2017). The resulting trees were then used to run the PTP analysis on the species delimitation server http://species.h-its.org/ within the Bayesian and maximum likelihood framework. Bayesian posterior probabilities for tentative species were acquired after running 500,000 generations, sampling every 100 generations, and discarding the first 20% of the samples as a burn-in. The putative species were also delimited using multi-rate PTP (mPTP, Kapli et al., 2017), which accounts for multiple evolutionary rates. This analysis lumped geographically and ecologically distant species as well as sympatric species that were separated in the multilocus framework of recent studies (Delić et al., 2017). For these reasons, the results were discarded from further analyses.

**Phylogenetic analyses**

Phylogenetic structure of the studied dataset was inferred using 195 Niphargidae species (Niphargus and the genera nested within), and were rooted with five species of *Pseudoniphargus*, which was shown to be a sister group in two previous studies (Jurado-Rivera et al., 2017; Copilaş-Ciocianu et al., 2020). The dataset included 169 MOTUs delimited in the PTP analysis (see above and Results section) and supplemented with 31 described species (excluded from the above described delimitation analyses because of the missing COI fragment). For the phylogenetic analysis, we assembled a multilocus molecular dataset that included two fragments of 28S rDNA, H3 and COI sequences. All sequences were aligned using the Q-INS-I algorithm implemented in MAFFT 7 (Katoh & Standley, 2013). The best partitioning scheme and the optimal substitution models for the codon positions were obtained using PartitionFinder 2.1.1 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017), and are available in Supplementary Information 2.

Phylogenetic relationships were reconstructed using two alternative approaches: i) maximum likelihood with partition-specific setting and ultrafast bootstrapping in IQ-tree 1.6.7 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015), and ii) Bayesian inference with partition-specific settings in MrBayes 3.2.6 (Ronquist et al., 2012). The node supports for the maximum likelihood analysis were assessed in 1,000 ultrafast bootstrap replicates (Hoang et al., 2018). Alternatively, the Bayesian Markov chain Monte Carlo (MCMC) tree was inferred after running two independent runs with four chains for 30 million generations each, and sampled every 400 generations. After reaching the stationarity phase, the first 25% of trees were discarded as burn-in and the remaining trees were used to calculate a 50% majority rule consensus tree.

**Estimation of divergence times**

In order to define the timeframe of splits between the lineages distributed in the Dinaric Karst and the Apennine Peninsula, we reconstructed a time-calibrated multilocus phylogeny using the package BEAST 2.5.1 (Bouckaert et al. 2018). The dataset comprising 200 individuals (same as above) and four molecular markers were analysed using partitioning schemes with distinct parameters settings, following the best-fit models of evolution proposed by bModelTest 1.2.1 (Bouckaert & Drummond, 2017). Due to the initial issues with the over parametrisation, we reduced the initial number of partitions (seven) to three, which corresponded to the three genes sequenced. The data on the partitioning schemes and the substitution models used in BEAST analyses are available in Supplementary information 2. Substitution and clock models were unlinked for all partitions, while the tree partitions were linked to generate a single dated
phylogeny. The marginal likelihood of Birth death and Yule speciation tree priors were compared in BEAST’s extension Path Sampler. Based on the marginal likelihood, we used a Yule speciation tree prior for the final analyses (results not shown). In order to account for lineage-specific rate heterogeneity, we employed a lognormal relaxed clock (Drummond, Ho, Phillips, & Rambaut, 2006). The tree was calibrated based on two lines of evidence, derived from fossilized specimens and biogeography. Two fossilized Niphargus from Baltic amber are of presumed age 35-50 Ma (Coleman & Myers, 2000; Jazdzewski & Kupryjanowicz, 2010); hence the calibration point (the node where morphologically similar species evolved for the first time) was set to a lognormal distribution with a 95% confidence interval between 34 and 95 Ma, and a mean value of 45 Ma. Furthermore, based on the absence of Niphargus from North America and Greenland, we assumed that Niphargus cannot be older than the initial opening of the Atlantic Ocean, and that this genus most likely evolved after the final submergence of the land-bridges between Eurasia and North America between 57 and 71 Ma (Brikiatis, 2014). Therefore, the second calibration point (root of the family Niphargidae) was set to a lognormal distribution with a 95% confidence interval between 35 and 154 Ma, and a mean value of 64 Ma. Four independent runs of 100 million generations, sampling every 10,000 steps, were performed and combined using LogCombiner 2.6.1. The stationarity of each single run was checked in Tracer 1.7 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) and the first 10% trees were discarded as burn-in. All the remaining samples from the posterior distribution were summarized using TreeAnnotator 1.8 in the so-called maximum clade credibility (MCC) tree, which was visualized in FigTree 1.4. Calibration points were cross-validated in two alternative analyses in which either of the calibration points was excluded.

Modelling past dispersal, vicariance and reconstruction of ancestral ranges

In order to elucidate the historical biogeography of the main transadriatic clade, we modelled dispersal, extinction, vicariance, cladogenesis and ancestral ranges (Fig. 1) using the R 3.5.5 (R Core Team, 2019) package BioGeoBEARS 1.1.2 (Matzke, 2013).

The modelling requires a time-calibrated phylogeny and the distributions of the extant species. In order to avoid the effects of biogeographic events outside the study region on parameter estimation of our models, we pruned the tree and ran the analyses on the focal clade (see Results). To account for phylogenetic uncertainty, we ran all the analysis on a sample of 100 random time-calibrated trees drawn from the stationary phase in the MCMC analysis. Species distributions were extracted from the European Groundwater Crustacean Dataset (Zagmajster et al., 2014), distributional database SubBio (http://subbio.net/db/) and our own, unpublished data. In turn, the distributions were categorized into three regions defined with respect to their paleogeographic origin: The Dinaric Karst, the Apennine Peninsula and the Adriatic Islands (Fig. 2).

In order to explore historical biogeographic events, we ran several biogeographic models. Using a stochastic continuous-time Markov process, these models decompose range evolution at the time of cladogenesis (at the nodes) and anagenesis (along the branches), and test their fit using the Akaike Information Criterion (AIC; REF Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723) or likelihood values. Biogeographical models we used are illustrated in Figure 1. Briefly, all models assume anagenetic range extension and contraction, technically called dispersal (parameter d) and extinction (parameter e), respectively. Range evolution at cladogenetic events is treated differently between the three basic models (DEC, DIVA, BAYAREA, see Fig. 1 for comparison). All the basic models, however, assume that all descending species inherited parts of the ancestral ranges (Fig. 1). An important extension of basic models is the so-called jump-dispersal speciation (labeled as “basic model+J”), modelled with the parameter j that is estimated simultaneously.
with the other parameters (Matzke, 2013). Jump dispersal assumes colonization of a novel biogeographic area at the time of cladogenesis (founder event speciation) and is, arguably, especially important in island systems (Matzke, 2014; see also Ree & Sanmartín, 2018 for criticism of the jump dispersal models). In BioGeoBEARS, all the basic models can be extended with the jump dispersal parameter, and the fit of these models can be assessed using likelihood values or AIC.

**Figure 1.** A summary of biogeographic models as implemented in the R package BioGeoBEARS: hypothetical anagenetic and cladogenetic processes. Redrawn from Matzke (2013). All implemented models allow for founder event (jump dispersal) speciation by addition of parameter “j”.

We ran all possible basic models (DEC, DIVA, BAYAREA), and models with additional jump dispersal speciation (+J). This ended in altogether six models. Biogeographic models were tested using AIC and
Akaike weights to interpret how well different models fit the geographic distributions and the proposed phylogeny. However, given that jump dispersal models (+J) may increase the number of alternative cladogenetic events and disproportionately inflate their explanatory power against anagenetic events, we did not compare the basic and +J models directly (Ree & Sanmartin, 2018). The relative contribution of dispersal and vicariance at the point of cladogenesis in the diversification of the transadriatic *Niphargus* was obtained by running 100 biogeographic stochastic mappings, based on the two models with the best fit (the best basic model and the best +J extended model), the calibrated phylogeny and the geographical distribution.

**Results**

*Species delimitations and phylogenetic analyses*

Both species delimitations yielded similar results. ABGD hypothesized that 463 COI sequences should be classified into 162 MOTUs, while both PTP analyses suggested that the dataset comprised 169 MOTUs (Supplementary Information 4). The difference in MOTU numbers indicated different levels of splitting in three broadly distributed and genetically diverse Dinaric species (*N. subtypicus*, *N. kolombatovici* and *N. hebereri*), where the distinction between population and species level was not clear. However, the alternative structure of MOTUs did not yield conflicting results, i.e. alternative haplotype segregations. Given that biogeographic events (dispersal, extinction, cladogenesis) took place already in the early stages of speciation, sometimes even at a population level, we based all subsequent analyses on MOTUs delimited by the PTP analysis.

Both, maximum likelihood and Bayesian inference resulted in similar trees (Fig. 2, Supplementary Information 5). Species from both sides of the Adriatic Sea were nested within a large clade (ultrafast bootstrap support/posterior probabilities = 97/0.69) distributed predominantly in the Dinaric Karst and altogether including 76 MOTUs. Hereinafter this clade is referred to as the main transadriatic clade (red clade in Fig. 2).
Figure 2. Maximum likelihood (IQtree) molecular phylogeny of the genus *Niphargus*, with the representatives of the so-called main transadriatic clade (ultrafast bootstrap support = 97 %) indicated in red (left). The tree was built using 28S rRNA I and II, H3 and COI gene sequences. The map on the right shows the distribution of the main Transadriatic clade representatives (red circles) and other *Niphargus* species (black circles) used in the species delimitation analyses. Bottom right: *Niphargus pectencoronatae*, an iconic species known from only a few anchialine caves in the Kornati archipelago of the Adriatic Sea (Photo courtesy of Vedran Jalžić).

Within the main transadriatic clade are nested three strongly supported sub-clades (ultrafast bootstrap support/posterior probabilities = 100 %/1), with species distributed on both sides of the Adriatic Sea and on the Adriatic Islands (Fig. 3). For convenience, we named them after the species comprising the clades as follows: The *N. hebereri* – *N. ictus* clade, the *N. steueri* – *N. patrizii* clade and the *N. hvarensis* – *N. salernianus* clade (Fig. 3). In addition, several species pairs (e.g., *N. pachytelson* – *N. pectencoronatae*, *N. hvarensis* – *N. miljeticus*, *N. doli* – *N. arbiter*, (*N. arethusa + N. alpheus*) – *N. salonitanus*, *N. steueri* islands – *N. steueri* mainland, *N. liburnicus* – *N. sp.*) were distributed in the Dinaric Karst and on the Adriatic Islands (Figs. 3 and 5).

The age of splits

The estimated ages of the main nodes between the alternative calibration schemes varied ± 2 Ma, implying concordance between the calibration points used. Here we report the results using both calibration points, whereas detailed results of cross-validation analyses are available in Supplementary Material 6. The age of the common ancestor of the main transadriatic clade was dated to the late Oligocene – early Miocene, with a mean estimate of 23.8 Ma (95 % confidence interval 18.7-29.6 Ma), likely falling into the time window of the Dinaric Karst uplift, which started 30 Ma. The splits between the species distributed in the Dinaric Karst, Adriatic Islands and Apennine Peninsula fell into three time windows (Fig. 3 and 5).
Figure 3. Time-calibrated maximum clade credibility tree of the main transadriatic clade derived from BEAST analysis, using 28S rDNA I and II, H3 and COI gene sequences (full phylogeny is available in Supplement Information 6). The three clades nested within the main transadriatic clade that are distributed on the both sides of the Adriatic Sea are indicated in purple (N. steueri – N. patrizii), orange (N. hebereri – N. ictus) and green (N. hvarensis – N. salernianus). Island populations are indicated as red stars. The posterior probabilities of the node supports are denoted as white > 0.90, grey > 0.95 and black > 0.99 circles and the 95 % confidence intervals are shown as blue node bars (left). Geographical distribution of the main transadriatic clade representatives, with the three clades distributed on both sides of the Adriatic Sea are indicated with purple (N. steueri – N. patrizii), orange (N. hebereri – N. ictus) and green circles (N. hvarensis – N. salernianus). The rest of the main transadriatic clade MOTUs are labelled with black circles (right).

The earliest splits took place between 10 and 20 My ago, hence predating the Messinian Salinity Crisis. The ancestor of the N. hebereri (Dinaric Karst and Adriatic Islands) – N. ictus (Apennine Peninsula) clade split from its Dinaric relatives approximately 14.4 My ago (95% confidence interval 10.4 – 18.5 My ago). Roughly in the same period, the N. salernianus clade (Apennine Peninsula) split from the N. hvarensis clade (Dinaric Karst) (13.3 My ago, 95% confidence interval 9.1 – 15.6 Ma). Somewhat younger is the split between the N. steueri clade (Dinaric Karst) and N. patrizii (Apennine Peninsula), dating to 11 My ago (95% confidence interval 7.2 – 14.2 My ago, Fig. 4).

Later on, and roughly corresponding to the Messinian Salinity Crisis, the N. hebereri clade (Adriatic Islands and Dinaric Karst) split from the N. ictus clade (Apennine Peninsula) (6.9 My ago, 95% confidence interval 4.4 – 8.9 My ago, Fig. 4).

The time of the most recent series of splits coincided with Pleistocene glaciations. These include splits among cryptic species of the N. hebereri complex (Adriatic Islands and Dinaric Karst), and six splits between species distributed on the Adriatic Islands and the Dinaric Karst (Fig. 3 and 4).
Figure 4. Biogeographic reconstruction of the peri-Adriatic region through time; i) Middle Miocene, ii) Messinian Salinity Crisis and iii) Pleistocene (enlarged view on the Dinarides). The main hypothetical dispersal routes of subterranean faunas, as reconstructed using the R package BioGeoBEARS, are indicated as black arrows and the present geography is outlined in background. The reconstructed land-bridges are based on Correggiari et al. (1996), Popov et al. (2004) and Mazza and Rustioni (2008).

Biogeographical models and their spatial implications

Regardless of the model of ancestral range reconstruction used, the main trans Adriatic clade originated in the Dinaric Karst with high probability (Fig. 5). The biogeographic models which included jump-dispersal had higher likelihood than the basic models. Among the basic models, DIVA received the highest support, whereas among the models with jump dispersal (+J), DEC+J received the highest support (Table 1). These two models suggest different biogeographical histories of the main trans Adriatic clade. According to DIVA, the main process explaining species distribution was range extension, modelled by the parameter \( d \) (technically termed as dispersal), which was estimated to be an order of magnitude higher than range contraction (technically termed as extinction, parameter \( e \)) (Table 1). By contrast, the DEC+J model suggested that range evolution was predominantly shaped by jump dispersal, and that founder event speciation was the main driver of the clade’s biogeographical history. These two contrasting models were additionally supported by 100 stochastic models. According to this, the DIVA and DEC+J models estimated that dispersal and jump dispersal had the largest contribution to the existing distribution patterns, respectively. The simulations estimated on average 6.33 (DIVA) / 6.17 (DEC+J) transitions from the Dinaric Karst to the Adriatic Islands, 3.03 (DIVA) / 2.95 (DEC+J) events from the Dinaric Karst to the Apennine Peninsula, 0.51 (DIVA) / 0.61 (DEC+J) events from the Apennine Peninsula to the Dinaric Karst and 0.54 (DIVA) / 0.57 (DEC+J) events from the Apennine Peninsula to the Adriatic Islands.

Table 1. Comparison of the ancestral range estimation models for the main trans Adriatic clade of *Niphargus*, as estimated in BioGeoBEARS (Matzke, 2013).

<table>
<thead>
<tr>
<th>model name</th>
<th>LnL(^1)</th>
<th>P(^2)</th>
<th>dispersion</th>
<th>extinction</th>
<th>jump-dispersal</th>
<th>AIC(^3)</th>
<th>AICc(^4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEC+J</td>
<td>-45.5</td>
<td>3</td>
<td>1.00E-12</td>
<td>1.00E-12</td>
<td>4.47E-02</td>
<td>96.99</td>
<td>97.3</td>
</tr>
<tr>
<td>DIVAlike+J</td>
<td>-45.51</td>
<td>3</td>
<td>1.00E-12</td>
<td>1.00E-12</td>
<td>4.47E-02</td>
<td>97.03</td>
<td>97.35</td>
</tr>
<tr>
<td>BAYAREAlike+J</td>
<td>-45.51</td>
<td>3</td>
<td>1.00E-07</td>
<td>1.00E-07</td>
<td>4.34E-02</td>
<td>97.03</td>
<td>97.35</td>
</tr>
<tr>
<td>DIVAlike</td>
<td>-64.72</td>
<td>2</td>
<td>7.06E-03</td>
<td>2.39E-04</td>
<td>0.00E+00</td>
<td>133.40</td>
<td>133.6</td>
</tr>
<tr>
<td>DEC</td>
<td>-66.00</td>
<td>2</td>
<td>7.06E-03</td>
<td>2.39E-04</td>
<td>0.00E+00</td>
<td>136.0</td>
<td>136.16</td>
</tr>
<tr>
<td>BAYAREAlike</td>
<td>-96.60</td>
<td>2</td>
<td>1.15E-02</td>
<td>2.59E-02</td>
<td>0.00E+00</td>
<td>197.36</td>
<td>197.36</td>
</tr>
</tbody>
</table>

\(^1\)Log-likelihood

\(^2\)Proportion

\(^3\)AIC

\(^4\)AICc
Number of parameters
Akaike information criterion
Akaike information criterion corrected for small sample size

Figure 5. Distribution of the main transadriatic clade categorized into three regions, Dinaric Karst (light green), Adriatic Islands (dark green) and Apennine Peninsula (purple), used in the ancestral range reconstructions (left). Ancestral ranges for the main *Niphargus* transadriatic clade were estimated using the DEC+J and DIVAlike models (right). The DIVAlike model shows only sections of ancestral range reconstructions, differing from the DEC+J model. Pie charts indicate relative probabilities of all possible ancestral ranges. Bars in the middle indicate the present distribution of the main transadriatic clade representatives. The periods when the three historical land-bridges between the Dinaric Karst and the Apennine Peninsula existed are labelled as orange bars on the chronogram. The earliest land-bridge emerged in the Oligocene (Mazza & Rustioni, 2008), followed by the Messinian Salinity Crisis (ca. 5.96-5.33 My ago; Garcia-Castellanos and Villaseñor, 2011; Gargani & Rigollet, 2007) and the Pleistocene glacial periods (Correggiari, Roveri, & Trincardi, 1996).

Discussion

The results unveil a complex historical biogeography of the main transadriatic clade, reflecting the major paleogeographic events in the Adriatic basin (Fig. 4 and 5). The clade apparently originated on the emerging Dinaric Karst, about 19-29 My ago. Following the emergence of land-bridges between the Dinarides and Calabria/Apulia ca. 10-15 My ago (Mazza & Rustioni, 2008), the ancestral representatives
of all three transadriatic sub-clades, namely the *N. hebereri* – *N. ictus* clade, the *N. steueri* – *N. patrizii* clade and the *N. hvarensis* – *N. salernianus* clade, spread across the sea. Subsequent submergence of the land-bridge gave rise to *N. patrizii*, the complex of *N. salernianus* and to the ancestor of a clade comprising *N. stefanellii*, *N. ictus*, *N. cornicanus*, *N. parenzani* and *N. hebereri*. During the Messinian Salinity Crisis (Krijgsman, Hilgen, Raffi, Sierro, & Wilson, 1999), approximately 5.96-5.33 My ago, the ancestor of the *N. hebereri* species clade likely spread from the Apennine Peninsula and colonized the Adriatic Islands as well as the mainland of the Dinaric Karst. Finally, during the Pleistocene (2.5 My -11.7 Ky ago) (Correggiari et al., 1996), at least six lineages colonized the Adriatic Islands from the Dinaric Karst (Figs. 3-5). The timing of the marine regressions corresponded to the periods of dispersal. The details on the dispersal remain unclear, given that models incorporating jump dispersal cannot be statistically compared with the models without this parameter (Ree and Sanmartin, 2018). The results of DIVA imply longer periods of dispersal during marine regressions, followed by transgression-driven vicariance. By contrast, jump dispersal models imply that the dispersal took place in narrow time windows during regressions, simultaneously with the cessation of gene flow between ancestral and descending populations (i.e., founder event speciation) (Matzke, 2013; 2014). In spite of these uncertainties, all the presented results favour the hypothesis that marine regression-transgression cycles shaped the biogeographical history of the main transadriatic *Niphargus* clade. By contrast, the results provide no evidence for the hypothesis that the main transadriatic clade originated through a single colonization of the ancient landmasses with no subsequent dispersal across the Adriatic basin. This is because we estimated the age of this clade as older than 10 My and inferred multiple bidirectional dispersal events across the Adriatic basin.

The biogeographic role of marine regression-transgression cycles proposed herein also makes sense from a paleogeographic perspective. The first colonisations are a logical follow-up of the geological evolution of the wider area in that the origin of the Dinaric Karst preceded the origin of the Apennine Peninsula (Rosenbaum, Lister, & Duboz, 2002; Carminati, Luštrino, & Doglioni, 2012) and that the older landmass was a source of Apennine peninsular faunas.

The robustness of our conclusions depends on the accuracy of the molecular clock. We recognize that the ages of the genus and the focal clade are younger than suggested in previous genus-wide studies (Esmaeili-Rineh, Sari, Delić, Moškrič, & Fišer, 2015; McInerney et al., 2014). However, the calibrations in McInerney et al. (2014) relied on external calibration points that reflected biogeographic splits within the distantly related epigean genus *Gammarus* (Hou, Sket, Fišer, & Li, 2011). Application of a molecular clock calibrated on epigean and distantly related species is potentially problematic given the longer lifespan and slower generation turnover of subterranean species (Saclier et al., 2019). For this reason, we calibrated the phylogeny using exclusively internal calibration points and suggest that our scheme more accurately reflects the timing of cladogenetic events.

The question remains how subterranean species, predominantly massive bodied and cave dwelling, crossed such long distances in relatively short temporal periods. Despite the enhanced connectivity of surface freshwaters, subterranean species hardly disperse across them, presumably due to competitive interactions and predation from surface species (Romero, 1985; Tobler, Schlupp, & Plath, 2007). We can think of two, mutually non-exclusive explanations. First, in early phases of marine regressions, the communities in emerging freshwater habitats underwent dramatic changes (Hou et al. 2011). In the transitional phase, these habitats were similar to salty marshes (Popov et al., 2004), where both freshwater and marine faunas were depauperate and the competitive and predatory pressures were less severe than in fully assembled communities. Some members of the main transadriatic clade tolerate brackish water (Delić et al., 2017; Gottstein et al., 2012; Sket, 1994) or sulphide-rich waters (Fišer, Luštrik, Sarbu, Flot, & Trontelj, 2015; Flot, Wörheide, & Dattagupta, 2010). Hence, within a relatively
narrow time window, the ancestors of the modern clades were possibly competitive enough to spread across species-poor areas, with crustacean communities inhabiting the moderately saline bottom of the Adriatic basin. Second, we cannot rule out subterranean connections across the major parts of the Adriatic basin, either as patches of shallow subterranean habitats (Culver, & Pipan, 2014), or as deep subterranean habitats that developed during a marine regression and filled with sediments during a transgression. Such transitional development of freshwater subterranean connections is supposed to exist due to the paleo-hydrology of Adriatic basin rivers and the evolution of the peri-Mediterranean region during the Neogene (Audra et al., 2007; Mocochain et al., 2009; de Leeuw, Mandic, Krijgsman, Kuiper, & Hrvatović, 2011; Pavelić et al., 2014).

Our study emphasizes the importance of marine regression-transgression cycles in shaping the extant biogeography patterns of subterranean freshwater fauna. In the same area, time-calibrated phylogenetic studies found similar biogeographic patterns among epigean invertebrate and vertebrate taxa (decapods, gastropods, mammals), most of them presumably shaped during the Pleistocene or Messinian Salinity Crisis (Jesse et al., 2009; Korábek, Juričková, & Petrusek, 2014; Jelić et al., 2016; Castiglia, Annesi, Amori, Solano, & Aloise, 2017). By contrast, only few of those studies inferred biogeographic patterns shaped by the presence of the ancient Middle Miocene land-bridge between the central part of the Dinaric Karst and Apulia (Mazza & Rustioni, 2008). This paucity of data for more ancient biogeographic events in the epigean taxa may be due to the region's vivid geological history that wiped out many ancestral species living in the epigean environments.

It has been suggested that subterranean faunas, especially crustaceans, living in stable and physically remote environment, comprise relict and rare species that bear testimony to past faunas (Humphreys 2000). We agree with this view and suggest that the value of subterranean species in historical biogeography is currently underestimated, and we argue that life in remote environments makes subterranean species an interesting model for the study of past biogeographic events.
Data availability statement

All the relevant data has been archived in Dryad Digital Repository and is reachable through supporting Information link https://datadryad.org/stash/share/PgnDvOEr2qr4e5qLcmrp2zG3Pz4ol2MdhdQO-J1a2Jl. Please note that the link is only temporary, and will be replaced with a permanent link after manuscript acceptance.
References


Independent and recent large-scale dispersal into surface waters by two species of the groundwater amphipod genus \textit{Niphargus}. \textit{Molecular Phylogenetics and Evolution}, 119, 37–49.


**Biosketch**

The authors are broadly interested in the biogeography of the Dinaric karst and evolutionary mechanisms underlying phylogeographic and diversification patterns of subterranean animals. They collaborate on questions of subterranean amphipod biodiversity based at SubBio Lab, University of Ljubljana, Slovenia ([http://subbio.net/](http://subbio.net/)) and Evolutionary Biology & Ecology, Université libre de Bruxelles ([http://ebe.ulb.ac.be/](http://ebe.ulb.ac.be/)).