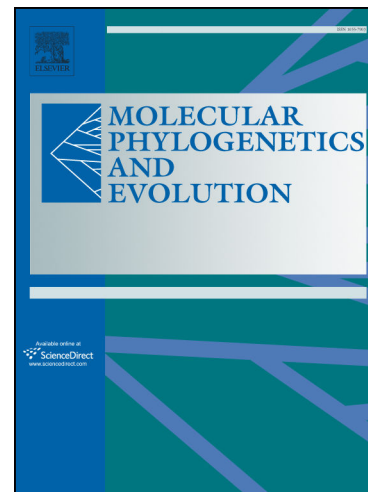


Journal Pre-proofs

Diversification dynamics of freshwater bivalves (Unionidae: Parreysiinae: Coelaturini) indicate historic hydrographic connections throughout the East African Rift System

Claudia M. Ortiz-Sepulveda, Björn Stelbrink, Xavier Vekemans, Christian Albrecht, Frank Riedel, Jonathan A. Todd, Bert Van Bocxlaer

PII: S1055-7903(20)30088-9
DOI: <https://doi.org/10.1016/j.ympev.2020.106816>
Reference: YMPEV 106816



To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 1 October 2019
Revised Date: 24 March 2020
Accepted Date: 1 April 2020

Please cite this article as: Ortiz-Sepulveda, C.M., Stelbrink, B., Vekemans, X., Albrecht, C., Riedel, F., Todd, J.A., Van Bocxlaer, B., Diversification dynamics of freshwater bivalves (Unionidae: Parreysiinae: Coelaturini) indicate historic hydrographic connections throughout the East African Rift System, *Molecular Phylogenetics and Evolution* (2020), doi: <https://doi.org/10.1016/j.ympev.2020.106816>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Diversification dynamics of freshwater bivalves (Unionidae: Parreysiinae: Coelaturini)
indicate historic hydrographic connections throughout the East African Rift System**

Claudia M. Ortiz-Sepulveda^{1,*}, Björn Stelbrink^{2,3}, Xavier Vekemans¹, Christian Albrecht², Frank Riedel⁴, Jonathan A. Todd⁵, Bert Van Bocxlaer¹

¹Univ. Lille, CNRS, UMR 8198 – Evo-Eco-Paleo, F-59000 Lille, France.

²Department of Animal Ecology and Systematics, Justus Liebig University Giessen, Giessen, Germany.

³Zoological Institute, Department of Environmental Sciences, University of Basel, Basel, Switzerland.

⁴Department of Paleontology, Institute of Geological Sciences, Freie Universität Berlin, Berlin, Germany

⁵Department of Earth Sciences, Natural History Museum, London, UK

*Corresponding author; email: claudia.ortiz-sepulveda@univ-lille.fr; tel: +33 3 20 43 40 15; address: Université de Lille, CNRS UMR 8198 Evo-Eco-Paleo; Cité scientifique, Bat. SN2; 59655 Villeneuve d'Ascq cedex, France. ORCID: 0000-0003-0072-719X

Abstract

Invertebrates are exceptionally diverse, but many are in decline because of anthropogenic changes to their habitat. This situation is particularly problematic for taxa that are not well monitored or taxonomically poorly understood, because the lack of knowledge hampers conservation. Despite their important functional role in freshwater ecosystems, African bivalves of the family Unionidae remain poorly studied compared to their highly threatened relatives in Europe, the U. S. A. and Canada. To resolve relationships and to study diversification dynamics in space and time, we performed time-calibrated phylogenetic studies and biogeographical modeling on the unionids from the East African Rift System and surroundings, including representatives of all currently recognized Afrotropical genera except for *Brazzaea* (and *Unio* from southern Africa). Our analyses indicate that all sampled Afrotropical unionids belong to the tribe Coelaturini (subfamily Parreysiinae), as does the genus *Moncetia* from Lake Tanganyika, which is currently attributed to the family Iridinidae. Colonization of Africa from Eurasia by Parreysiinae occurred ~17 Ma ago, and the subsequent diversification of Coelaturini in Africa continued at a steady pace, although net diversification decreased over time as more niches and ecoregions became occupied. Clades in Coelaturini largely reflect drainage basins, with the oldest lineages and highest regional diversity occurring in Lake Tanganyika, followed by the Congo Basin watershed in general. The species assemblage of Lake Tanganyika reflects multiple independent events of colonization and intralacustrine diversification since the Late Miocene or Early Pliocene. The clades of other regions, including that containing the species from Lake Malawi, are comparatively young. Biogeographical analyses indicate that the colonization history was mainly driven by cladogenesis in sympatry, whereas few anagenetic events contributed to the modern distribution of Coelaturini. Ancestral range estimations demonstrate

that Coelaturini originated in the Victoria and/or Tanganyika ecoregions, and that the Congo Basin played an essential role in the colonization of Africa by Coelaturini.

Keywords: Freshwater mussels, *Coelatura*, Africa, systematics, phylogeography, biogeographic modeling

Journal Pre-proofs

1. Introduction

Invertebrates are exceptionally diverse and central to the functioning of many ecosystems (Wilson 1987; Eisenhauer et al. 2019), but substantial parts of their biodiversity remain poorly studied. Many invertebrates across ecosystems are vulnerable to anthropological stress and two thirds of the monitored populations show a mean abundance decline of 45% since the 16th century (Dirzo et al. 2014). This decline is especially problematic for those taxa for which little systematic knowledge exists because lack of data hampers the implementation of conservation policies, which makes poorly known biodiversity vulnerable to irretrievable loss.

Freshwater bivalves of the family Unionidae, which contains at least 620 extant species (Bogan and Roe 2008; Graf and Cummings 2019), are among the most endangered of freshwater biota (Lydeard et al. 2004; Lopes-Lima et al. 2017) as a result of the continued anthropogenic degradation of their ecosystems (Pereira et al. 2014). Unionids have complex life histories that include parental care and larval parasitism (Barnhart et al. 2008) and they perform many valuable ecosystem functions (Strayer et al. 1994; Lopes-Lima et al. 2017; Vaughn 2018).

Important spatial bias occurs, however, in knowledge on unionid biodiversity: A long research tradition exists on the ecology (Ortmann 1920; Dillon 2000) and systematics (Simpson 1900; Ortmann 1920; Haas 1969; Graf and Cummings 2007; Lopes-Lima et al. 2017) of the unionids of the U. S. A. and Europe, whereas knowledge on Asian Unionidae has started to develop only much more recently (Brandt 1974; Bolotov et al. 2017; Bolotov et al. 2018; Pfeiffer et al. 2018; Wu et al. 2018; Zieritz et al. 2018). Knowledge on Unionidae of sub-Saharan Africa remains much more restricted in that populations are not monitored and insights into their species richness and taxonomy are still primarily based on shell morphology and distributional data (Mandahl-Barth 1988; Daget 1998; Scholz and Glaubrecht 2004; Graf and Cummings 2011;

Seddon et al. 2011). Due to a scarcity of phylogenetic studies, the taxonomic validity and phylogenetic position of many Afrotropical genera is uncertain (Graf and Cummings 2007), although hypothetical systematic reclassifications have been proposed (Lopes-Lima et al. 2017; Graf and Cummings 2019).

Nevertheless, the combined efforts of previous studies have helped delineate the deeper phylogenetic relationships among Unionidae and the placement of some African taxa. Whelan et al. (2011) recovered Unionidae as monophyletic and retrieved the African genera *Coelatura*, *Prisodontopsis* and *Nitia* as a well-supported clade, the tribe Coelaturini, which is sister to the Asian Indochinellini, as has later been confirmed by Bolotov et al. (2018). These tribes belong to the subfamily Parreysiinae (Bolotov et al. 2017; Lopes-Lima et al. 2017), which is sister to the Unioninae (Pfeiffer et al. 2019). Coelaturini supposedly contains 38 nominal species that belong to 8 genera (Graf and Cummings 2007, 2019): *Brazzaea* Bourguignat, 1885, *Coelatura* Conrad, 1853, *Grandidieria* Bourguignat, 1885, *Mweruella* Haas, 1936, *Nitia* Pallary, 1924, *Nyassunio* Haas, 1936, *Prisodontopsis* Tomlin, 1928 and *Pseudospatha* Simpson, 1900, of which *Mweruella* and *Prisodontopsis* may have to be synonymized (Graf et al. 2014). However, the potential attribution of *Brazzaea*, *Grandidieria*, *Nyassunio* and *Pseudospatha* to Coelaturini has not been substantiated. Indeed, some authors (Kat 1987) have suggested that, among others, *Brazzaea* and *Pseudospatha* would be more closely related to *Unio* (*Cafferia*) *caffer*, a unionid from southern Africa belonging to the subfamily Unioninae. This hypothesis remains to be tested, however.

The endangered state of Unionidae worldwide combined with the poor knowledge on the biodiversity and systematics of African unionids and the lack of conservation studies on African populations is alarming, especially considering the large-scale ecosystem change observed in

many African freshwater systems (Thieme et al. 2010; Darwall et al. 2011). Here, we aim to fill the existing knowledge gap by testing the monophyly of Coelaturini with extended taxon sampling and by evaluating phylogenetic support for the identification system of Mandahl-Barth (1988) based on the morphological-geographical recognition of species. We reconstruct phylogenetic relationships of Unionidae occurring throughout Central and East Africa, with a focus on the East African Rift System (EARS). The EARS contains ~90 % of the species diversity of Afrotropical unionids and representatives of all currently recognized genera (Mandahl-Barth 1988; Daget 1998; Graf and Cummings 2007). Finally, we examine diversification dynamics in Coelaturini in their spatial-temporal context via fossil-based time-calibration and ancestral range estimations to reconstruct major colonization pathways among African freshwater ecoregions. We also compare biogeographical patterns in Coelaturini to those of their fish hosts.

2. Material and methods

2.1. Material and taxonomy sampling

The material used for this study was obtained from the Systematics and Biodiversity Collection of the Justus Liebig University (UGSB), the Natural History Museum London (NHMUK) and the Museum d'Histoire Naturelle de Genève (MHNG), together with data from previous studies (Whelan et al. 2011; Graf et al. 2014; Pfeiffer and Graf 2015; Bolotov et al. 2017) and almost a decade of fieldwork (between 2006 and 2015). We obtained samples from all 8 genera that supposedly belong to Coelaturini—in this respect *Unio caffer*, which is widely distributed in southern Africa, was not included in our study (see Whelan et al. 2011). For *Brazzaea* only historic material was available, however, and because the extracted DNA was of very poor

quantity and quality, this taxon is not further considered here. Unfortunately, sequencing of specimens from West Africa housed at the MHNG resulted in similar issues. Material yielding good quality DNA and genetic sequences is enlisted in Table S1. It includes specimens from all major lakes of the EARS that are currently inhabited by Afrotropical unionids and from, among others, the Congo River and several of its tributaries, the Nile, including the Victoria Nile, the Zambezi, the Okavango, and the Cunene.

We used an extensive set of Asian Parreysinae from previous studies (Whelan et al. 2011; Graf et al. 2014; Pfeiffer and Graf 2015; Bolotov et al. 2017) as outgroup, including specimens of *Indonaia* Prashad, 1918, *Radiatula* Simpson, 1900, *Indochinella* Bolotov, Pfeiffer, Vikhrev & Konopleva, 2018 and *Parreysia tavoyensis* (Gould, 1843) (see Table S1).

2.2. DNA extraction and sequencing

Genomic DNA was extracted from ~1 mm³ of muscle or mantle tissue using a CTAB protocol (Wilke et al. 2006). Four molecular markers were sequenced, i.e. fragments of the mitochondrial cytochrome c oxidase subunit I (COI) and the large ribosomal subunit (16S) complemented with fragments of the nuclear histone 3 (H3) and the large ribosomal subunit (28S). We used the following primers: LCO1490 and HCO2198 (COI; Folmer et al. 1994); 16sar-L and 16sbr-H (16S; Palumbi et al. 2002); D23F and D4RB (28S; Park and Ó Foighil 2000); H3F and H3R (H3; Colgan et al. 1998). DNA vouchers for newly sequenced material (~ 240 specimens) were deposited at the UGSB; data on other specimens was retrieved from NCBI GenBank (Table S1). PCR cycling condition are specified in Table S2. Bidirectional DNA sequencing was performed on a 16-capillary 3730xl Genetic Analyzer (Applied Biosystems). New sequences were deposited in GenBank (Table S1).

2.3. Phylogenetic inference

All four gene fragments were aligned individually using the online version of MAFFT v. 7 (Kato et al. 2017). Subsequently, we reduced the dataset to unique haplotypes and examined substitutional saturation (Xia et al. 2003; Xia and Lemey 2009) for each gene fragment with DAMBE v. 6.4.101 (Xia 2017). Fragments of the protein-coding genes (COI: 660 bp; H3: 328 bp) were translated to amino acid sequences with MESQUITE v. 3.51 (Maddison and Maddison 2018) to evaluate the presence of stop codons. Fragments of the ribosomal genes (16S, 28S) were analyzed with GBlocks v. 0.91b (Castresana 2000; Talavera and Castresana 2007) after alignment to exclude hypervariable sections from the aligned sequences. We allowed gap positions within the final blocks, but we excluded sections with multiple contiguous non-conserved positions. These criteria reduced the length of the alignment from 541 bp to 453 bp and from 928 bp to 407 bp for 16S and 28S, respectively.

All gene fragments were concatenated using MESQUITE and unique haplotypes were recognized with DAMBE. Subsequently, we used Partition Finder v. 2.1.1 (Lanfear et al. 2016) and a Bayesian information criterion (BIC) to infer the best-fit partitioning scheme and substitution models (Table S3). Phylogenetic analyses on this concatenated dataset were performed using Bayesian inference (BI) and maximum likelihood (ML) on the CIPRES Science Gateway v. 3.3 (Miller et al. 2010). ML was performed with RAxML-HPC BlackBox v. 8.2.12 (Stamatakis 2014) using the GTR + Γ substitution model applied to the partitions indicated in Table S3, simultaneous best-tree search and bootstrapping with recommended stop rule. BI was performed using MrBayes v. 3.2.3 (Ronquist and Huelsenbeck 2003) with the partitions and substitution models indicated in Table S3. Two independent analyses were performed, each by

running four chains simultaneously for 30,000,000 Markov chain Monte Carlo (MCMC) generations, sampling each 1000th tree. We examined whether stationarity of MCMC runs was reached in the Bayesian analyses using Tracer v. 1.7.1 (Rambaut et al. 2018). The resulting trees were visualized with FigTree v. 1.4.3 (Rambaut 2009).

2.4. Estimation of divergence times

We estimated divergence times with BEAST v. 1.10.2 (Suchard et al 2018; Drummond et al 2012) on the CIPRES Science Gateway. We conducted three independent runs, with substitution and clock models unlinked among the gene fragments, an uncorrelated lognormal relaxed-clock model (Drummond et al. 2006), a Yule speciation process for the tree prior (Yule 1925; Gernhard 2008), 80,000,000 MCMC generations, and sampling each 1,000th tree. Initially, best-fit substitution models as identified by PartitionFinder were implemented (see Table S3), however, because of potential overparameterization of complex substitution models a stationary state was not reached in preliminary runs. Therefore, we repeated the analyses with the simpler HKY model, which also performed consistently well in model comparisons. Site heterogeneity and invariant site parameters (+ Γ , + I, + Γ + I) were set to the best performing variant.

Divergence times were estimated with three fossil calibration points, which we established based on the guidelines of Parham et al. (2012; see supplementary text for more details): The oldest fossil of *Nitia* was used to calibrate the MRCA of the *Nitia* clade (lognormal prior with 7.5 ± 1.3 Ma [mean \pm SD; 95% confidence interval (CI): 5.3-10.4 Ma]), the oldest fossil of *Coelatura* cf. *hauttecoeuri* was used to calibrate the MRCA of the *C. hauttecoeuri*/*C. bakeri*/*C. aegyptiaca* clade (lognormal prior with 5.2 ± 0.5 Ma [mean \pm SD; 95% CI: 4.3-6.3 Ma]), whereas the oldest

fossil of *Coelatura* was used to calibrate the MRCA of *Coelatura sensu lato* (lognormal prior with 10.0 ± 1.6 Ma [mean \pm SD; 95% CI: 7.2-13.5 Ma]) (Van Damme and Pickford 2010).

The results of the three independent runs were examined in Tracer after discarding 8,000,000 runs (10%) as burn-in. In individual runs, the ESS values of some parameters indicated a potential lack of convergence, but upon combining the results of all three runs, parameter values consistently showed ESS values >200 indicating proper mixing of the MCMC. From the 80,000 trees obtained per run 10 % was discarded as burn-in after which the remaining trees were subsampled in LogCombiner of the BEAST package to yield 36,000 trees per run. For each of these subsampled tree files we then identified the maximum clade credibility (MCC) tree with TreeAnnotator (BEAST package). These three MCC trees allowed us to evaluate the robustness of the outcome (topology and node ages) of our analyses for independent runs. The tree with the highest log clade credibility was selected as the best supported tree for subsequent analyses.

2.5. Species delimitation

To enable biogeographical analyses, we first pruned the gene tree into a species tree using the `prune_specimens_to_species` function of the package BioGeoBEARS v. 1.1.2. (Matzke 2014, 2018) in R v. 3.4.3 (R Core Team 2018) after removing the outgroup from the phylogeny.

Pruning requires an indication of species entities, which we approximated with operational taxonomic units (OTUs). These OTUs were defined based on the following information: First, we used three automated species recognition methods, i.e. automated barcode gap discovery (ABGD) (Puillandre et al. 2012), the Poisson tree processes (PTP) model (Zhang et al. 2013) and the generalized mixed Yule-coalescent (GMYC) approach (Fujisawa and Barraclough 2013).

ABGD is a fast single-locus method that was performed on the web server

(<http://wwabi.snv.jussieu.fr/public/abgd/abgdweb.html>) using our COI alignment, because of its exceptional suitability for species identification, delimitation and discovery in Metazoa (Hebert et al. 2003). We used default parameters, except for relative gap width, which was set to 1.0. The other two methods are tree-based. We used the BEAST MCC tree for the PTP analysis, which was ran with BI and ML on the bPTP server (<https://species.h-its.org/ptp>) using 500,000 MCMC generations (other parameters as default) after removing the outgroup taxon *P. tavoyensis*. Finally, GMYC was performed on the GMYC server (<https://species.h-its.org/gmyc>) with the BEAST MCC tree and a single threshold. Subsequently, the proposed OTUs were compared between methods, after which we related them to the morphology-based taxonomy using the relevant literature (e.g. Mandahl-Barth 1988) and comparison with museum records and the MUSSELp database (<http://mussel-project.uwsp.edu/fmuotwaolcb/index.html>).

2.6. Diversification dynamics

We used LogCombiner to sample 1,000 trees from the BEAST posterior distribution. For each of these trees we removed the outgroup and reconstructed the corresponding OTU-based tree. The posterior distribution of OTU-based trees was examined with DensiTree v. 2.2.5 (Bouckaert and Heled 2014). Subsequently, we constructed lineage-through-time (LTT) plots and their 95% confidence interval with functions of the R packages ape v. 5.0 (Paradis et al. 2004) and phytools v. 0.6-44 (Revell 2012). We also extracted parameters for birth-death (BD) and pure-birth (PB) models from our dataset accounting for incomplete sampling, which we used to simulate 100 trees under each model. We used these simulations to establish 95% confidence intervals on the LTTs for the BD and PB models to compare with the empirical dataset. Additionally, we tested for rate changes in the diversification dynamics of Coelaturini since their MRCA using functions

from the R package TreePar v. 3.3 (Stadler 2015), and with a Bayesian analysis of macroevolutionary mixtures (BAMM) in BAMM v. 2.5.0 (Rabosky et al. 2013; Rabosky et al. 2017). Priors for speciation and extinction rates (i.e. λ and μ , respectively) in the BAMM analysis were defined using the setBAMMpriors function of the R package BAMMtools v. 2.1.6 (Rabosky et al. 2014) and the expected number of shifts was set to 1.0. The analysis was conducted with 10,000,000 generations, sampled each 10,000 generations, and with a burn-in of 10%. The sampling fraction was defined as the number of OTUs divided by the total number of nominal valid species ($26/38 = 68.4\%$).

2.7. Ancestral range estimation

We examined how Coelaturini were distributed over the African continent with a phylogenetic approach to biogeography based on reticulate models as implemented in BioGeoBEARS. Specifically, we fit six models, i.e. DEC, DEC+J, DIVALIKE, DIVALIKE+J, BayAreaLIKE, BayAreaLIKE+J (Massana et al. 2015; Matzke 2018). DEC is the dispersal-extinction-cladogenesis model of Ree and Smith (2008) as implemented in the software package Lagrange. DIVALIKE is a model that represents a likelihood interpretation of parsimony-based dispersal-vicariance analyses (Ronquist 1997). The BayAreaLIKE model is a simplified likelihood interpretation of Bayesian analyses of biogeography including many areas, as implemented in the program BayArea (Landis et al. 2013). This last model is similar to DEC, but it assumes that no particular events are associated with cladogenesis. Comparison of DEC and BayAreaLIKE therefore indicates the importance of the cladogenesis model on biogeographical estimations for a dataset. Finally, these three models were fit with and without the parameter J, which represents the frequency of founder-event jump dispersal, which is a consequence of founder-event

speciation. Founder-event speciation is a crucial process of dispersal, at least in island systems (Matzke 2014), to which the East African lakes may belong (Salzburger et al. 2014). After analyzing these six models we compared the model fit using a corrected Akaike information criterion (AICc). Biogeographical modeling was performed for OTU-trees based on the BEAST MCC tree and two ultrametricized versions of the MrBayes consensus tree. The latter two trees were: 1) constructed from the MrBayes consensus tree with functions of phytools, and 2) by sampling a tree from the BEAST posterior distribution that has an identical topology to the MrBayes consensus tree. To examine the role of anagenetic and cladogenetic biogeographical events in shaping the diversity and distribution of Coelaturini, we additionally performed biogeographical stochastic mapping (BSM; see Dupin et al. 2017). This analysis was performed in BioGeoBEARS with 1,000 replicates and parameters of the best-fit biogeographical model used as priors.

We used African freshwater ecoregions that were established based on the composition and distribution of the freshwater fish fauna (Abell et al. 2008) as geographic units for our analyses. These ecoregions have been used to describe patterns in unionid diversity before, which is significantly correlated with species richness in fish (Graf and Cummings 2011). For practical purposes, however, some ecoregions were modified for this study, because for some ecoregions no bivalve records were available, but more importantly some bivalve OTUs were widely distributed (covering up to 5 ecoregions), which made biogeographical analyses computationally intensive. In all these cases, it was possible to combine several small fish ecoregions into a broader unionid ecoregion without losing information, e.g. by combining subzones within the Okavango, Congo and Nile watersheds. Finally, we defined a total of 12 geographic areas and

we used the occurrence information of specimens to define the distribution of OTUs throughout these ecoregions.

3. Results

3.1. Phylogenetic inference

Examination of the sequence alignments for each gene fragment revealed no evidence of substitutional saturation. The concatenated dataset contained 1,848 bp and 205 unique haplotypes. Phylogenetic reconstruction with RAxML and MrBayes resulted in highly congruent topologies and an overall strongly supported backbone (Fig. 1). These analyses indicated that the seven included genera belong to Coelaturini, which was recovered as a highly supported, monophyletic taxon (BPP=1.00; BS=100). Additionally, our analyses indicate that *Moncetia*, supposedly an iridiniid (see 4.1. Evolutionary and taxonomic implications), also belongs to Coelaturini, and that it is part of the clade including *Pseudospatha tanganyikense*. The Tanganyikan genera *Grandidieria* (BPP=1.00; BS=100) and *Pseudospatha* + *Moncetia* (BPP=1.0; BS=100) are monophyletic, genetically distinct and have basal positions in Coelaturini. Contrasting topologies, however, were recovered in the MrBayes + RAxML trees versus the BEAST MCC tree: In the MrBayes consensus tree and the RAxML tree, *Grandidieria* is sister to all other Coelaturini, whereas it is sister to *Pseudospatha* in the BEAST MCC tree. In each case a related node is weakly supported (BPP= 0.93; BS = 67 versus BPP = 0.66), however. The genus *Nitia* is also monophyletic and highly supported (BPP=1.00; BS=100) and was recovered as the sister group to *Coelatura sensu lato* in the Bayesian analysis (BPP=0.99), whereas in the RAxML analysis it was recovered (albeit very poorly supported; BS < 50) as the sister group to the highly supported clade (BPP=1.00, BS=97) of *Coelatura* from the Nile

watershed, including among others Lakes Albert, Edward and Victoria. We consider this latter grouping in the RAxML analysis uninformative given the low associated BS values. *Coelatura sensu lato* contains all other Coelaturini, including the genera *Mweruella*, *Prisodontopsis*, and *Nyassunio*, which renders the current definition of the widespread *Coelatura* paraphyletic. Within *Coelatura sensu lato*, as for Coelaturini in general, the spatial distribution of clades coincides well with watershed boundaries, with various deeply divergent clades occurring in the Congo watershed (Fig. 1), which are all highly supported (BPP=1.00; BS=94-100). Three of these clades occur directly in the Congo Basin, whereas the fourth occurs mainly in Lake Mweru, although it has also colonized the Luvua River in the Upper Congo Basin. Finally, the fifth clade occurs in Lake Tanganyika and surroundings, indicating that Lake Tanganyika has been colonized at least twice by Coelaturini. The other two recovered groups are the clade geographically distributed throughout the Okavango and Zambezi (BPP= 1.00; BS=95) and the clade including taxa from the Luapula River and Lake Malawi (BPP= 1.00; BS=82). Analysis of our genetic data did not recover *Nyassunio* from Lake Malawi as monophyletic. Instead these specimens form a polytomy with *Coelatura* from Zambia and the Luapula River. This polytomy is one of three major polytomies in the current phylogeny, the others representing *Coelatura* from the Nile drainage and those from the Upper Zambezi. At least two of these polytomies supposedly contain several species.

3.2. Molecular-clock analyses

The fossil-calibrated BEAST phylogeny (Fig. 2) overall produced the same topology to that of RAxML and MrBayes, apart from the issue of *Grandidieria* and *Pseudospatha* mentioned above (compare Figs. 1 and 2) and the uncertain position of *Nitia* in the RAxML analysis. Based on this

calibration, Coelaturini would have diverged from Indochinellini ~23 Ma (95 % highest probability density (HPD): 15.25-31.95 Ma), with a MRCA of Coelaturini originating ~17 Ma (95% HPD: 11.98-23.28 Ma). The diversification between *Grandidieria* and *Pseudospatha* occurred ~14 Ma (95 % HPD: 8.61-19.57 Ma). Multiple cladogenetic events in the Congo Basin occurred in the Middle to Late Miocene (95% HPD: 7.42-12.56, 6.63-11.72, 5.54-10.46). Diversification within clades from Lake Tanganyika ensued since the Miocene-Pliocene transition (95% HPD: 2.64-9.75, 3.89-11.20, 2.99-8.07 Ma), whereas lineage diversification within the Okavango-Zambezi region, the Lake Mweru clade and the Lake Malawi clade occurred simultaneously at ~3 Ma (95 % HPD: 1.68-4.98, 1.50-4.89, 1.74-4.52 Ma).

3.3. Species delimitation

Although the OTUs proposed by our automated species recognition methods were not identical the results display a high level of congruence (Table S4). ABGD and PTP suggested 11-48 OTUs in the ingroup, but the best-fit ABGD estimation (31 OTUs) was very similar to the ML estimation of the PTP method (29 OTUs) and to the GMYC solution (32 OTUs). This similarity highlights the overall robustness of these methods for our dataset, and it allowed us to synthesize these results in a single hypothesis of OTUs (Table S4). This synthetic hypothesis includes 26 OTUs, because some proposed OTUs divided well-supported clades into hypothetically reciprocally monophyletic subclades that have limited BPP or BS support. Acknowledging that the results of OTU methods are often conflicting, we performed subsequent analyses with 24 to 27 OTUs to investigate the robustness of downstream analyses. These analyses indicated highly similar results under various hypothetical OTU assignments. In what follows OTUs are indicated

with species names, although these names reflect morphological similarity in the absence of molecular data on topotypical specimens for many of the studied species.

3.4. Analyses of diversification dynamics

Examination of 1,000 OTU-based trees (Fig. 3) indicates a limited effect of topological changes among gene trees on the construction of the OTU-based tree, which consolidates the robustness of downstream OTU-based analyses. The LTT plot (Fig. 3) indicates a steady increase in diversity since the MRCA of Coelaturini. This accumulation closely follows the expected diversity under pure-birth and birth-death models, which are almost identical for our dataset. Over the last 5 Ma, lineages accumulated somewhat faster in the empirical dataset compared to the pure-birth and birth-death simulations, although their 95% CIs continue to overlap broadly. We observed evidence for a single shift in diversification rates along the phylogeny of Coelaturini ($p = 0.002$) in TreePar, i.e. a decrease in the number of lineages around 3 Ma ago. The BAMM analyses are generally congruent with this result and indicate that speciation rates gradually decreased over time (Fig. 4B). The extinction rate is very low, and almost constant, resulting in a ~2-fold decrease in net diversification rates over time.

3.5. Biogeographical analyses

Analyzing biogeographical models for our OTU-based trees indicated a substantially better fit for the DEC model than for DIVALIKE and BayAreaLIKE (AICc weight for DEC and DEC+J = 0.86-0.99, Table 1), regardless of the topology or the number of OTUs considered. This result implies that changes upon cladogenesis are important for the estimation of historic biogeography in Coelaturini. Biogeographical analyses using the BEAST topology gave highest support for the

DEC model (AICc weight = 0.57), whereas it was DEC+J for the MrBayes topology (AICc weight = 0.67). In each case, the other variant of the DEC model received considerable support too (AICc weight = 0.29-0.33). For the BEAST topology, the difference between both models is not significant (one-tailed χ^2 : $p = 0.270$), whereas for the MrBayes analysis the support of DEC+J over DEC trends towards significance (one-tailed χ^2 : $p = 0.046$). As such the importance of jump dispersal upon founder-event speciation in Coelaturini remains ambiguous, with a low jump-dispersal parameter for DEC+J models (Table 1). The extinction parameter was also low, which is consistent with low extinction rates in the BAMM analysis (Fig. 4B), and which may explain why LTT plots under pure-birth and birth-dead models are very similar for this empirical dataset (Fig. 3; see 3.4. Analyses of diversification dynamics). Ancestral range estimations under DEC and DEC+J models are highly similar, regardless of the topology; we report DEC here. Stochastic mapping indicated that few anagenetic events have occurred: dispersal was reconstructed for 7 out of 50 branches (Fig. 4C). The majority of cladogenetic events ($n = 14-16$ out of 25) occurred in sympatry, with fewer events relating to subset sympatry ($n = 5-7$) and vicariance ($n = 4$). The MRCA of Coelaturini occurred in the ecoregions of Lakes Tanganyika, Victoria, or less likely, that of the Central Congo River. Subsequently, within *Coelatura sensu lato*, colonization occurred from the northern part of the East African Rift throughout the Nile watershed, and gradually southward towards the ecoregions of the Okavango + Zambezi, Mweru + Upper Lualaba + Lake Malawi.

4. Discussion

4.1. Evolutionary and taxonomic implications

This first rift-wide analysis of Coelaturini has several evolutionary and taxonomic implications. First, it indicates that *Moncetia*, a monotypic genus from Lake Tanganyika currently attributed to the family Iridinidae (Leloup 1950; Mandahl-Barth 1988; Daget 1998; Graf and Cummings 2007), is a unionid. This affiliation was suggested before by Kat (1987), although he did not substantiate it with data. Kondo (1984) attributed *Moncetia* to Mutelidae [= Iridinidae], but reported it to have larval glochidia, which are considered to be an apomorphy of Unionidae + Margaritiferidae + Hyriidae (Parodiz and Bonetto 1963; Graf and Cummings 2006), whereas Iridinidae produce lasidium larvae. It remains unclear, however, how Kondo (1984) attributed glochidia to *Moncetia*, because he reported to have collected them directly from fish hosts and two other glochidium-bearing bivalve genera, i.e. *Pseudospatha* and *Coelatura*, co-occur with *Moncetia* in Lake Tanganyika. Beyond revealing that *Moncetia* belongs to Coelaturini and thus that it should have glochidium larvae indeed, we additionally show that it belongs to the same clade as *Pseudospatha*. Interestingly, Leloup (1950) considered *Moncetia* and *Pseudospatha* to be morphologically continuous, although he placed both in Iridinidae. Studies with faster evolving markers and more comprehensive geographic sampling of Lake Tanganyika are required to examine how *Pseudospatha* and *Moncetia* are related to each other. Our results also call for an in-depth comparative study of the reproductive biology of Coelaturini altogether, especially of its representatives from Lake Tanganyika, given that *Grandidieria* has secondarily lost its parasitic larval stage (Leloup 1950; Kondo 1990), and that both *Grandidieria* and *Moncetia* were reported to brood their eggs within the inner demibranchs of their gills only (Kondo 1984, 1990). Such endobranchy has to our knowledge not been observed in other Unionidae, which are known to brood eggs in all four demibranchs (tetrageny) or the outer pair of demibranchs only (ectobranchy) (Graf and Cummings 2006; Wu et al. 2018). Endobranchy is

the common condition in Hyrioidea (Hyriidae) and Etherioidea (Etheriidae + Mycetopodidae + Iridinidae; Graf and Cummings 2006). In any case, our results and those of Whelan et al. (2011) jointly disprove the previous hypothesis that *Pseudospatha*, *Moncetia* and *Prisodontopsis* are more closely related to the South African *Unio caffer* than to *Coelatura* (Kat 1987).

Interestingly, the widespread genus *Coelatura* belongs to the crown-group of Coelaturini rather than the stem, implying that its morphological traits, such as its general habitus, larval and brooding morphology, the position of the umbo and umbonal sculpture, may be derived rather than ancestral for Coelaturini. This finding increases the likelihood that Late Cenozoic African fossils with seemingly atypical features, such as *Pseudobovaria* with its strongly anteriorly projected umbo (see Van Damme and Pickford 2010), in fact represent Coelaturini.

Taxonomic revision of *Coelatura sensu lato* is thus required, but beyond the scope of the current paper. Conceivable strategies are either to resurrect/erect genus names for individual clades, or alternatively, to lump *Coelatura*, *Mweruella*, *Prisodontopsis*, *Nyassunio* and many deeply divergent lineages into *Coelatura*, formalizing the concept of *Coelatura sensu lato* as used here. Ideally, the taxonomic decision would be informed further with complementary studies on morphology, anatomy and life-history to ensure that taxa are readily diagnosable. Beyond the need to resolve the paraphyly of *Coelatura sensu lato*, a striking result of our analysis is that multiple nominal species that are morphologically distinct and that occupy distinct regions of watersheds are recovered in large polytomies. One example is the clade of the Nile watershed, which includes specimens of *C. aegyptiaca* (Nile), *C. bakeri* (Lake Albert), *C. stuhlmanni* (Lake Edward), *C. hauttecoeuri*, *C. alluaudi* (Lake Victoria), and possibly other species from Lake Victoria that were not sampled (*C. cridlandi*). A second example is the clade including the Coelaturini of Lake Malawi, which contains *C. choziensis* (Lake Bangweulu, Upper Congo), *C.*

luapulaensis (Upper Congo), *C. mossambicensis* (Lake Malawi and Lower Zambezi), *C. hypsiprymna* (Lake Malawi) and *N. nyassaensis* (Lake Malawi). A similar issue was observed in gastropods (Viviparidae) from the Nile drainage: morphologically distinct nominal *Bellamyia* species that occupy separate geographic regions were not recovered as separate entities in phylogenetic analyses with a limited number of gene fragments (Schultheiß et al. 2014). However, subsequent analysis with 15 microsatellites recovered several of these nominal species as distinct molecular groups of which each occupies a geographically separated, though hydrographically interconnected region within the Nile watershed (Van Bocxlaer et al. unpublished data). Whereas our phylogenetic analyses of Coelaturini resolve relationships between taxa from distinct watersheds to large extent, our markers did not fully resolve the relationships of nominal species within drainage basins, except perhaps among multiple deeply separated species in the Congo Basin. Population genomic analyses will be required to examine the reality of previously proposed species entities. This current lack of resolution is one of the main reasons why we currently refrain from revising *Coelatura sensu lato*. It also implies that some of our OTUs may comprise multiple, cryptic species. As a result, we used our robust phylogenetic backbone to reconstruct biogeographical patterns and processes among but not within watersheds.

4.2. Diversification dynamics within Coelaturini

Basal cladogenetic events within Coelaturini arose between taxa that are currently living in Lake Tanganyika. The MRCA of the clade containing *Grandidieria*, *Pseudospatha* and *Moncetia* is ~14 Ma (95% HPD: 8.61-19.57 Ma), which is broadly consistent with age estimates for the formation of Lake Tanganyika (~ 9 - 12 Ma; Cohen et al. 1993). Whether biodiversity within

Lake Tanganyika has resulted from intralacustrine radiation versus independent colonizations has been the subject of considerable debate for freshwater gastropods (West and Michel 2000; Wilson et al. 2004). Our analyses indicate that at least two independent colonization events of the lake occurred by Coelaturini, followed by considerable intralacustrine diversification. Intralacustrine diversification is also observed in other Tanganyikan taxa, such as the spiny eel *Mastacembelus* (Brown et al. 2010) and the cichlid *Tropheus* (Baric et al. 2003). The diversification of Coelaturini throughout the EARS resulted in a steady accumulation of lineages over time, with an apparent, but non-significant increase ~5 Ma. Assuming that it is not an artifact related to the species delimitation methods, this increase may reflect 1) Pliocene-Pleistocene climate change (Cane and Molnar 2001), or 2) the creation of new ecological opportunities after the extinction of several unionid lineages in sub-Saharan Africa, most notably *Pseudobovaria*, *Pseudodiplodon* and *Gautieraia* (Van Damme and Pickford 2010). Overall, however, a decrease in net diversification rates over time is observed in the BAMM analyses, which suggests rapid diversification of Coelaturini upon colonization of Africa, with a subsequent deceleration of diversification as niches and ecoregions become progressively occupied. In-depth examination of molecular diversity in several drainage basins, including the Lake Victoria Basin, the Okavango-Zambezi region and the Lake Malawi Basin, will be required to reconstruct the dynamics of species diversity in Coelaturini in more detail.

Another aspect that needs clarification is that our analyses of diversification dynamics and our best-fit biogeographical models all reconstruct very low levels of extinction. Whereas this finding is perhaps not surprising for recently diversifying clades (see e.g. Condamine et al. 2015), several Afrotropical unionid genera have gone extinct since the Middle Miocene (Van

Damme and Pickford 2010). Further studies are required to examine whether the reconstructed level of extinction is congruent with the paleontological record.

4.3. Historical biogeography

Our phylogenetic framework suggests that patterns of colonization and continent-wide diversification dynamics in Africa are similar for Coelaturini and several fish families (Rüber et al. 2006; Brown et al. 2010). This finding is not surprising given the general correlation between African bivalve and fish diversity (Graf and Cummings 2011), and the fact that most Coelaturini have a fish-parasitizing larval stage (which also represents the phase of greatest mobility in their life cycle). As such, Parreysiinae probably colonized Africa attached to a fish host, like other parasites did (Pariselle 2003), and this colonization by Parreysiinae from Eurasia is thus expected to reflect the patterns of their freshwater fish hosts.

4.3.1. Colonization of Africa

Our fossil-calibrated phylogeny indicates that the MRCA of Coelaturini dates to ~17 Ma (95% HPD: 11.98-23.28 Ma), which suggests that the colonization of Africa from Eurasia by Parreysiinae likely occurred in the Early or Middle Miocene. This estimate is consistent with geological evidence for the closure of the Tethys seaway ~18-20 Ma (Okay et al. 2010), and with the first evidence of mammal exchanges between both continents (Rögl 1999; Harzhauser et al. 2007; Harzhauser and Piller 2007). Additionally, rifting in southwestern Ethiopia seems to have occurred around the same time (~18 Ma; Ebinger et al. 2000; Pik et al. 2008), although much of the eastern branch of the EARS was established ~20 Ma (Pickford 1982; Chorowicz 2005). These events have formed a corridor that has facilitated the colonization of Africa from Eurasia by freshwater biota, as evidenced by the fish families Mastacembelidae (Brown et al. 2010),

Anabantidae (Rüber et al. 2006), Clariidae (Agnese and Teugels 2005), Bagridae and Cyprinidae (Stewart and Murray 2017) and viviparid gastropods (Schultheiß et al. 2014; Sil et al. 2019).

Despite the congruence with geological and other biogeographical data, our estimated timeframe for the colonization of Africa by Parreysiinae is substantially younger than the Paleocene-Eocene age estimates of Bolotov et al. (2017). These latter authors used a calibration strategy based on the assignment of Paleogene Asian fossils to modern genera and fossil constraints on outgroup taxa, which has been criticized before (Pfeiffer et al. 2018). Early Cenozoic estimates render the absence of Coelaturini in the Oligocene and Early to Middle Miocene fossil record of the East African Rift problematic (Kat 1987).

The route by which Parreysiinae colonized Africa from Eurasia remains equivocal. Colonization probably occurred via the Arabian Peninsula and the Bab al-Mandab Strait towards the Awash River in Ethiopia. Miocene deposits along this river contain freshwater fish fossils whose origins were in Asia, such as *Labeo*, *Barbus*, *Labeobarbus*, *Bagrus* and *Clarias*, which corroborates this hypothesis (Stewart and Murray 2017), but the associated freshwater mollusk fauna is currently poorly surveyed. The alternative hypothesis is that the colonization of Africa from Asia took place via a northern route through the Sinai Peninsula and Northern Egypt, as is suggested for some terrestrial vertebrates (e.g. Pook et al. 2009; Gilbert et al. 2014). This colonization route has not been substantiated for freshwater biota because during much of the Middle and Late Miocene the Sinai Peninsula was occupied by marine waters (Westcott et al. 2000; Van Damme and Van Bocxlaer 2009; Tsigenopoulos et al. 2010; Stewart and Murray 2017). The large-scale colonization of Africa from Eurasia in the Miocene, and early dispersal within Africa, seem to have coincided with the Middle Miocene Climatic Optimum (~15-17 Ma). This period was warm

with high precipitation (Zachos et al. 2001) that would have enhanced hydrographic connectivity in Africa, and therefore dispersal and diversification of aquatic biota (Day et al. 2017).

4.3.2. Dispersal throughout the EARS

We outline hydrographic connections in the Late Miocene and Early Pliocene in Fig. 5 together with a synthesis of the implications of our biogeographical analyses. Our biogeographical model suggests that the MRCA of Coelaturini occurred within the ecoregions of Lake Tanganyika, Victoria and (less likely) the Upper Congo. Regardless of the route that Parreysiinae used to colonize Africa from Eurasia, it is unlikely that Lake Tanganyika was reached without passing through areas of the Lake Victoria ecoregion first, implying that the MRCA of Coelaturini probably occurred in central East Africa and then gave rise to separate lineages that colonized the Tanganyika and Victoria regions ~14 Ma (as also supported by subset sympatry in our BSM). Early occupation of the Lake Victoria area is congruent with the fossil record, as the oldest fossils of Coelaturini occur in this ecoregion (i.e. Baringo Basin and Albertine Basin; De Groeve 2005; Van Damme and Pickford 2010), although these fossils are Late Miocene and thus substantially younger than ~14 Ma. The alternative scenario is that Africa was primarily colonized via the Lake Tanganyika ecoregion and subsequently via the Lake Victoria and Congo ecoregions. The fossil record and BSM provide less evidence for this scenario, but rapid colonization of the Lake Tanganyika ecoregion after migration to Africa is also recorded for Viviparidae: The currently monotypic Tanganyikan viviparid genus *Neothauma* diverged rapidly from other, more widespread African viviparid lineages (Schultheiß et al. 2014). However, the fossil viviparid record indicates that several extinct species of *Neothauma* occupied the Lake Victoria ecoregion in the Miocene (Van Damme and Pickford 1999; Salzburger et al. 2014).

Neothauma tanganyicense, and perhaps some of the Tanganyikan Coelaturini, may thus be relicts in Lake Tanganyika from earlier more widespread taxa, suggesting that the lake had a refugial function at a continental scale. Faunal exchange between the Lake Victoria and Lake Tanganyika ecoregions could have occurred via the proto-Rusizi River, a precursor of the current Rusizi, or alternatively via the Malagarasi River (Danley et al. 2012).

Our biogeographical analyses further suggest that Coelaturini colonized the Congo River from the larger Lake Victoria ecoregion, which includes the developing Ethiopian rift. During the Early and Middle Miocene the proto-Aruwimi River was diverted towards the Nile, which could have allowed the colonization of Nilotic ecoregions and the Upper Congo from the larger Lake Victoria area (Flügel et al. 2015). In the Late Miocene, the waters of the Lake Victoria ecoregion drained towards the Albertine Basin, and then via a precursor of the Aruwimi River into the Congo River (Van Damme and Pickford 1999). At that time, rifting resulted in the development of Palaeolake Obweruka (Van Damme and Pickford 1999), which may have provided some habitats in which Coelaturini persisted locally. The Late Miocene-Early Pliocene *Coelatura* cf. *stanleyvillensis* from the Albertine Basin displays morphological features that are similar to those of taxa currently inhabiting the Congo Basin (Van Damme and Pickford 2010). As rifting proceeded, the hydrographic connectivity of the Congo and Victoria ecoregions may have decreased, facilitating allopatric speciation in *Coelatura* ~10 Ma. Around the same time, viviparid gastropods occurring in the Congo and Victoria ecoregions became isolated as well (Schultheiß et al. 2014). Other evidence for aquatic colonization from the Victoria ecoregion towards the Congo is observed in the haplochromine cichlid “Yaekama”, which is distributed in the northeastern part of the Congo River System near Kisangani, but groups with the Lake Victoria superflocks (Schwartz et al. 2012). Furthermore, the distribution of the poeciliid genus

Hypsopanchax also points to the connectivity of the Victoria ecoregion with the Congo Basin drainage via a precursor of the Aruwimi (Poll and Lambert 1965; Roller et al. 2010; Bragança and Costa 2019).

Subsequently, several diversification events occurred in Coelaturini within the Congo Basin ~8-10 Ma, and some of the descendant lineages (*C. cf. leopoldvillensis*1 & 2, *C. cf. gabonensis*1 & 2) are currently sympatric. The potential biogeographical setting for the accumulation of diversity in this hotspot is currently uncertain, but diversification caused by the hydrographic separation of regions within the Congo Basin is conceivable. Alternatively, it is possible that these lineages did not originate in the Congo Basin itself, but that they derive from independent colonization events from separate source regions as in the spiny eel *Mastacembelus* (Day et al. 2017) and the fish fauna of the Upper Luapula River (Van Steenberge et al. 2014).

Coelatura colonized Lake Tanganyika from the Congo Basin relatively late ~5-9 Ma, while *Grandidieria* and *Pseudospatha* + *Moncetia* already inhabited the lake. This pathway implies the colonization of the Upper Congo waters (Lualaba and Lukuga Rivers) by *Coelatura* from the proto-Aruwimi and Lower Congo. Fossil evidence that would substantiate the pathway and time of this colonization of Lake Tanganyika has not been discovered yet. However, various fish taxa of Lake Tanganyika have relatives in the Congo River basin, e.g. Lamprologini and *Lacustricola* (Bragança and Costa 2019), and the Malagarasi River, a major tributary of Lake Tanganyika that probably existed prior to rifting (Cohen et al. 1997), is also inhabited by a fish fauna that is similar to that of the Congo Basin (Goodier et al. 2011; Kullander and Roberts 2011). A similar distribution is also observed for the gastropod *Potadomoides* (West et al. 2003). As our biogeographical reconstruction indicates, *Coelatura* also colonized the Malagarasi River and coastal East Africa from Lake Tanganyika later.

Between ~6-8 Ma, *Coelatura* also dispersed from the Congo catchment to the Okavango region. This colonization could have occurred by a connection from the Lower Congo system to the headwaters of the Upper Zambezi via the Kasai River (Veatch 1935; Stankiewics and de Wit 2006; Moore et al. 2007). Viviparid gastropods have been hypothesized to have used this route to colonize the Okavango region from the Congo River ~ 6.3 Ma (Schultheiß et al. 2014), and the route has been invoked for various fish taxa (Bell-Cross 1966), such as catfish (Day et al. 2009), tigerfish (Goodier et al. 2011), and cichlids (Schwartzter et al. 2012). The BSM analysis suggests that parts of Zambia including the areas of Lake Bangweulu and Lake Mweru were colonized simultaneously to the Okavango region by Coelaturini, followed by vicariance ~4.5-6.0 Ma. The Mweru ecoregion was probably colonized from the Okavango ecoregion via the Chambeshi River, which may have drained eastward into the Rufiji system in the Pliocene (Stankiewics and de Wit 2006). Although the Chambeshi River is currently confluent with the Luapula, it represents the former upper reaches of the Kafue, a major Upper Zambezi affluent (Moore and Larkin 2001; Goudie 2005; Stankiewics and de Wit 2006) and a connection between the Chambeshi and the Kafue probably existed until the early Pleistocene (0.78-2.58 Ma) (Cotterill and de Wit 2011). A biogeographical pattern similar to that of Coelaturini was also recovered by Schultheiß et al. (2014) for viviparids of their 'clade II', which were broadly distributed throughout Zambia, and for the fishes of the Upper Luapula area (Van Steenberge et al. 2014). Colonization of the Malawi Basin by Coelaturini is substantiated by fossils of ~2.5 Ma from Paleolake Chiwondo (Schrenk et al. 1995; Van Damme and Gautier 2013). The colonization of the coastal East African ecoregion may have occurred through the Malawi Basin, ~0.8-1.2 Ma ago via the proto-Ruhuhu River, which would have been the outlet of Lake Malawi at that time (Ivory et al. 2016). Finally, a recolonization of the Congo River occurred from Lake Mweru

~1.85-3.04 Ma by *Coelatura briarti*, which may have taken place via the Luvua River, a tributary of the Lualaba River.

As mentioned before, the proto-Aruwimi River was diverted towards the Nile during the Early and Middle Miocene, which could have allowed an early colonization of Nilotic ecoregions (Flügel et al. 2015). Currently, we have no evidence of such a colonization for Coelaturini and the extant Nilotic fauna results from Early Pliocene dispersal, which is congruent with the fossil record (see Van Damme and Van Bocxlaer 2009). Several fish taxa from Early Pliocene deposits (lower Tinde Member) of Paleolake Manonga in Northern Tanzania have Nilotic affinities, indicating connectivity of the southern Lake Victoria ecoregion to Nilotic ecoregions at that time, either via the Albertine Basin, the Omo-Turkana Basin or both (Stewart 1997). Coelaturini have not been recovered from the deposits of Paleolake Manonga (Van Damme and Gautier 1997), however, the Early Pliocene fossils of *C. aegyptiaca* and *C. bakeri* from the Omo-Turkana Basin substantiate faunal exchange between the Lake Victoria ecoregion and Nilotic ecoregions at that time (Van Bocxlaer 2011; supplementary text on fossil calibrations).

In summary, knowledge on the historic hydrographic connections throughout the EARS and surroundings and our results on distribution patterns in Coelaturini are highly congruent. Many particularities of the biogeography of Coelaturini are furthermore substantiated by similar patterns in other freshwater invertebrates and various fish taxa. A striking result of our study is the ancient colonization of the Lake Tanganyika ecoregion, where various genera of Coelaturini appear to have evolved, persisted and coexisted. Our analyses furthermore suggest that each of the lineages colonizing the Tanganyika ecoregion underwent diversification in the Late Miocene-Early Pliocene. For major clades of Coelaturini elsewhere, population genetic analyses will be

required to study biodiversity patterns within drainage basins, and the evolutionary processes that underpin this diversification.

Acknowledgements

We thank Daniel Engelhard, Thies Geertz, Adrian Indermauer, Alidor Kankonda, Nicolas Lichilin, Koen Martens, Ellinor Michel, Jean-Papy Mongindo Etimosundja, Walter Salzburger, Ernest Tambwe Lukosha, Emmanuel Vreven and Oscar Wembo Ndeo for collecting specimens, and Jon Ablett (NHMUK) and Emmanuel Tardy (MNHG) for access to specimens and the permission to sample tissues. Friedemann Schrenk and Erik Verheyen provided logistical support. Pierre-Alexandre Gagnaire, Torsten Hauffe, Claude Monnet, John Pfeiffer, Camille Roux and an anonymous referee provided constructive advice that has strongly improved the current manuscript.

Funding

CMOS received an internship from the Justus Liebig University in Giessen and a PhD fellowship from Lille University and the project ANR-JCJC-EVOLINK of the French Agence Nationale de la Recherche (ANR, to BVB); BS, CA and FR were supported by grants of the Deutsche Forschungsgemeinschaft (BS: DFG STE 2460/2-1; CA: DFG AL 1076/5-2, DFG AL 1076/6-2); JT was supported by the Nyanza Project (National Science Foundation-REU grant: 0223920). This work is a contribution to the CPER research project CLIMIBIO and CMOS, XV and BVB thank the French Ministère de l'Enseignement Supérieur et de la Recherche, the Hauts de France Region and the European Funds for Regional Economic Development for their financial support.

Competing interests

The authors have no competing interests to declare.

References

- Abell, R., M. L. Thieme, C. Revenga, M. Bryer, M. Kottelat, N. Bogutskaya, B. Coad, N. Mandrak, S. Contreras Balderas, W. Bussing, M. L. J. Stiassny, P. Skelton, G. R. Allen, P. Unmack, A. Naseka, R. Ng, N. Sindorf, J. Robertson, E. Armijo, J. V. Higgins, T. J. Heibel, E. Wikramanayake, D. Olson, H. L. López, R. E. Reis, J. G. Lundberg, M. H. Sabaj Pérez, and P. Petry. 2008. Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience* 58:403-414.
- Agnese, J.-F. and G. G. Teugels. 2005. Insight into the phylogeny of African Clariidae (Teleostei, Siluriformes): implications for their body shape evolution, biogeography, and taxonomy. *Molecular Phylogenetics and Evolution* 36:546-553.
- Baric, S., W. Salzburger, and C. Sturmbauer. 2003. Phylogeography and evolution of the Tanganyikan cichlid genus *Tropheus* based upon mitochondrial DNA sequences. *Journal of Molecular Evolution* 56:54-68.
- Barnhart, M. C., W. R. Haag, and W. R. Roston. 2008. Adaptations to host infection and larval parasitism in Unionoida. *Journal of the North American Benthological Society* 27:370-394.
- Bell-Cross, G. 1966. The distribution of fishes in Central Africa. *Fisheries Research Bulletin Zambia* 4:3-20.

- Bogan, A. E. and K. J. Roe. 2008. Freshwater bivalve (Unioniformes) diversity, systematics, and evolution: status and future directions. *Journal of the North American Benthological Society* 27:349-369.
- Bolotov, I. N., A. Kondakov, I. V. Vikhrev, O. V. Aksenova, Y. V. Bespalaya, M. Y. Gofarov, Y. S. Kolosova, E. S. Konopleva, V. M. Spitsyn, K. Tanmuangpak, and S. Tumpeesuwan. 2017. Ancient river inference explains exceptional oriental freshwater mussel radiations. *Scientific Reports* 7:e2135 (doi: 2110.1038/s41598-41017-02312-z).
- Bolotov, I. N., J. M. Pfeiffer, E. S. Konopleva, I. V. Vikhrev, A. V. Kondakov, O. V. Aksenova, M. Y. Gofarov, S. Tumpeesuwan, and T. Win. 2018. A new genus and tribe of freshwater mussel (Unionidae) from Southeast Asia. *Scientific Reports* 8:e10030 (doi: 10010.11038/s41598-10018-28385-y).
- Bouckaert, R. R. and J. Heled. 2014. DensiTree 2: seeing trees through the forest. *BioRxiv* (<http://dx.doi.org/10.1101/012401>).
- Bragança, P. H. N. and W. J. E. M. Costa. 2019. Multigene fossil-calibrated analysis of the African lampeyes (Cyprinodontoidei: Procatopodidae) reveals an early Oligocene origin and Neogene diversification driven by palaeogeographic and palaeoclimatic events. *Organisms Diversity & Evolution* 19:303-320.
- Brandt, R. A. M. 1974. The non-marine aquatic Mollusca of Thailand. *Archiv für Molluskenkunde* 105:1-423.
- Brown, K. J., L. Rüber, R. Bills, and J. J. Day. 2010. Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. *BMC Evolutionary Biology* 10:e188 (doi: 10.1186/1471-2148-10-188).

- Cane, M. A. and P. Molnar. 2001. Closing of the Indonesian seaway as a precursor to east African aridification around 3-4 million years ago. *Nature* 411:157-162.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17:540-552.
- Chorowicz, J. 2005. The East African rift system. *Journal of African Earth Sciences* 43:379-410.
- Cohen, A. S., K.-E. Lezzar, J.-J. Tiercelin, and M. J. Soreghan. 1997. New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. *Basin Research* 9:107-132.
- Cohen, A. S., M. J. Soreghan, and C. A. Scholz. 1993. Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. *Geology* 21:511-514.
- Colgan, D. J., A. McLauchlan, G. D. F. Wilson, S. P. Livingston, G. D. Edgecombe, J. Macaranas, G. Cassis, and M. R. Gray. 1998. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Australian Journal of Zoology* 46:419-437.
- Condamine, F. L., N. S. Nagalingum, C. R. Marshall, and H. Morlon. 2015. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evolutionary Biology* 15:e65 (doi: 10.1186/s12862-12015-10347-12868).
- Cotterill, F. P. D. and M. J. de Wit. 2011. Geocodynamics and the Kalahari epeirogeny: linking its genomic record, tree of life and palimpsest into a unified narrative of landscape evolution. *South African Journal of Geology* 114:489-514.
- Daget, J. 1998. *Catalogue raisonné des mollusques bivalves d'eau douce africains*. Backhuys Publishers and ORSTOM, Leiden and Paris.

- Danley, P. D., M. Husemann, B. Ding, L. M. DiPietro, E. J. Beverly, and D. J. Peppe. 2012. The impact of the geological history and paleoclimate on the diversification of East African cichlids. *International Journal of Evolutionary Biology* 2012:e574851 (doi: 574810.571155/572012/574851).
- Darwall, W., K. Smith, D. Allen, R. Holland, I. Harrison, and E. Brooks. 2011. The diversity of life in African freshwaters: underwater, under threat. An analysis of the status and distribution of freshwater species throughout mainland Africa. IUCN, Cambridge, UK and Gland, Switzerland.
- Day, J. J., R. Bills, and J. P. Friel. 2009. Lacustrine radiations in African *Synodontis* catfish. *Journal of Evolutionary Biology* 22:805-817.
- Day, J. J., A. Fages, K. J. Brown, E. J. Vreven, M. L. J. Stiassny, R. Bills, J. P. Friel, and L. Rüber. 2017. Multiple independent colonizations into the Congo Basin during the continental radiation of African *Mastacembelus* spiny eels. *Journal of Biogeography* 44:2308-2318.
- De Groeve, E. 2005. De laat-Cenozoïsche zoetwatermollusken van de Tugen Hills (Kenia): taxonomie, paleoecologie en paleozoögeografie. Department Geology and Soil Science. Ghent University, Belgium, unpublished MSc. thesis, 121 pages.
- Dillon, R. T., Jr. 2000. The ecology of freshwater molluscs. Cambridge Univ. Press, Cambridge, UK.
- Dirzo, R., H. S. Young, M. Galletti, G. Ceballos, N. J. B. Isaac, and B. Collen. 2014. Defaunation in the Anthropocene. *Science* 345:401-406.
- Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4:e88: 699-710.

- Dupin, J., N. J. Matzke, T. Särkinen, S. Knapp, R. G. Olmstead, L. Bohs, and S. D. Smith. 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography* 44:887-899.
- Ebinger, C. J., T. Yemane, D. J. Harding, S. Tesfaye, S. Kelley, and D. C. Rex. 2000. Rift deflection, migration, and propagation: Linkage of the Ethiopian and Eastern rifts, Africa. *Geological Society of America Bulletin* 112:163-176.
- Eisenhauer, N., A. Bonn, and C. A. Guerra. 2019. Recognizing the quiet extinction of invertebrates. *Nature Communications* 10:1-3 (doi: 10.1038/s41467-41018-07916-41461).
- Flügel, T. J., F. D. Eckardt, and F. P. D. Cotterill. 2015. The present day drainage patterns of the Congo River system and their Neogene evolution *in* M. J. de Wit, F. Guillocheau, and M. C. J. de Wit, eds. *Geology and Resource potential of the Congo Basin*. Springer, Berlin & Heidelberg.
- Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3:294-299.
- Fujisawa, T. and T. G. Barraclough. 2013. Delimiting species using single-locus data and the generalized mixed Yule coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology* 62:707-724.
- Gernhard, T. 2008. The conditioned reconstructed process. *Journal of Theoretical Biology* 253:769-778.

- Gilbert, C. C., F. Bibi, A. Hill, and M. J. Beech. 2014. Early guenon from the late Baynunah Formation, Abu Dhabi, with implications for cercopithecoid biogeography and evolution. *Proceedings of the Academy of Natural Sciences of the U. S. A.* 111:10119-10124.
- Goodier, S. A. M., F. P. D. Cotterill, C. O'Ryan, P. H. Skelton, and M. J. de Wit. 2011. Cryptic diversity of African tigerfish (genus *Hydrocynus*) reveals palaeogeographic signatures of linked Neogene geotectonic events. *PLoS One* 6:e28775 (doi: 10.1371/journal.pone.0028775).
- Goudie, A. S. 2005. The drainage of Africa since the Cretaceous. *Geomorphology* 67:437-456.
- Graf, D. L. and K. S. Cummings. 2006. Palaeoheterodont diversity (Mollusca: Trigonioidea + Unionoidea): what we know and what we wish we knew about freshwater mussel evolution. *Zoological Journal of the Linnean Society* 148:343-394.
- Graf, D. L. and K. S. Cummings. 2007. Review of the systematics and global diversity of freshwater mussel species (Bivalvia: Unionoidea). *Journal of Molluscan Studies* 73:291-314.
- Graf, D. L. and K. S. Cummings. 2011. Freshwater mussel (Mollusca: Bivalvia: Unionoidea) richness and endemism in the ecoregions of Africa and Madagascar based on comprehensive museum sampling. *Hydrobiologia* 678:17-36.
- Graf, D. L. and K. S. Cummings. 2019. Unionidae Rafinesque, 1820, and the General Unionida. Pp. 202-209 in C. Lydeard, and K. S. Cummings, eds. *Freshwater mollusks of the world: a distribution atlas*. Johns Hopkins Univ. Press, Baltimore.
- Graf, D. L., A. J. Geneva, J. M. I. Pfeiffer, and A. D. Chilala. 2014. Phylogenetic analysis of *Prisodontopsis* Tomlin, 1928 and *Mweruella* Haas, 1936 (Bivalvia: Unionidae) from

- Lake Mweru (Congo basin) supports a Quaternary radiation in the Zambian Congo.
Journal of Molluscan Studies 80:303-314.
- Haas, F. 1969. Superfamily Unionacea. *Das Tierreich*. Walter de Gruyter, Berlin.
- Harzhauser, M., A. Kroh, O. Mandic, W. E. Piller, U. Göhlich, M. Reuter, and B. Berning. 2007.
Biogeographic responses to geodynamics: a key study all around the Oligo-Miocene
Tethyan Seaway. *Zoologischer Anzeiger* 246:241-256.
- Harzhauser, M. and W. E. Piller. 2007. Benchmark data of a changing seapalaeogeography,
palaeobiogeography and events in the Central Paratethys during the Miocene.
Palaeogeography, Palaeoclimatology, Palaeoecology 253:8-31.
- Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications
through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences*
270:313-321.
- Ivory, S. J., M. W. Blome, J. W. King, M. M. McGlue, J. E. Cole, and A. S. Cohen. 2016.
Environmental change explains cichlid adaptive radiation at Lake Malawi over the past
1.2 million years. *Proceedings of the National Academy of Sciences of the U. S. A.*
113:11895-11900.
- Kat, P. W. 1987. Biogeography and evolution of African freshwater molluscs: implications of a
Miocene assemblage from Rusinga Island, Kenya. *Palaeontology* 30:733-742.
- Katoh, K., J. Rozewicki, and K. D. Yamada. 2017. MAFFT online service: multiple sequence
alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*:1-7
(doi: 10.1093/bib/bbx1108).
- Kondo, T. 1984. Hosts of the larvae of *Moncetia lavigeriana* (Bivalvia: Mutelidae) in Lake
Tanganyika. *Venus* 43:347-352.

- Kondo, T. 1990. Reproductive biology of a small bivalve *Grandidieria burtoni* in Lake Tanganyika. *Venus* 49:120-125.
- Kullander, S. O. and T. R. Roberts. 2011. Out of Lake Tanganyika: endemic lake fishes inhabit rapids of the Lukuga River. *Ichthyological Exploration of Freshwaters* 22:355-376.
- Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology* 62:789-804.
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34:772-773.
- Leloup, E. 1950. Lamellibranches. Institut Royal des Sciences Naturelles de Belgique, Bruxelles.
- Lopes-Lima, M., E. Froufe, V. T. Do, M. Ghamizi, K. E. Mock, Ü. Kebapçı, O. Klishko, S. Kovitvadhi, U. Kovitvadhi, O. S. Paulo, J. M. I. Pfeiffer, M. Raley, N. Riccardi, H. Sereflisan, R. Sousa, A. Teixeira, S. Varandas, X. Wu, D. T. Zanatta, A. Zieritz, and A. E. Bogan. 2017. Phylogeny of the most species-rich freshwater bivalve family (Bivalvia: Unionida: Unionidae): defining modern subfamilies and tribes. *Molecular Phylogenetics and Evolution* 106:174-191.
- Lydeard, C., R. H. Cowie, W. F. Ponder, A. E. Bogan, P. Bouchet, S. A. Clark, K. S. Cummings, T. J. Frest, O. Gargominy, D. G. Herbert, R. Hershler, K. E. Perez, B. Roth, M. Seddon, E. E. Strong, and F. G. Thompson. 2004. The global decline of nonmarine mollusks. *Bioscience* 54:321-330.
- Maddison, W. P. and D. R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis. Version 3.51, <http://www.mesquiteproject.org>.

- Mandahl-Barth, G. 1988. Studies on African freshwater bivalves. Danish Bilharziasis Laboratory, Charlottenlund.
- Massana, K. A., J. M. Beaulieu, N. J. Matzke, and B. C. O'Meara. 2015. Non-null effects of the null range in biogeographic models: exploring parameter estimation in the DEC model. bioRxiv (<http://biorxiv.org/content/early/2015/09/16/026914>).
- Matzke, N. J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* 63:951-970.
- Matzke, N. J. 2018. BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R scripts. Version 1.1.1., GitHub, 6 November 2018 (doi: <http://dx.doi.org/10.5281/zenodo.1478250>).
- Miller, M., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 1-8. Gateway Computing Environments Workshop. Institute of Electrical and Electronics Engineers (IEEE), New Orleans.
- Moore, A. E., F. P. D. Cotterill, M. P. L. Main, and H. B. Williams. 2007. The Zambezi River. Pp. 311-332 in A. Gupta, ed. *Large Rivers*. John Wiley & Sons, Chichester, U. K.
- Moore, A. E. and P. A. Larkin. 2001. Drainage evolution in south-central Africa since the breakup of Gondwana. *South African Journal of Geology* 104:47-68.
- Okay, A. I., M. Zattin, and W. Cavazza. 2010. Apatite fission-track data for the Miocene Arabia-Eurasia collision. *Geology* 38:35-38.
- Ortmann, A. E. 1920. Correlation of shape and station in fresh-water mussels (naiades). *Proceedings of the American Philosophical Society* 59:269-312.

- Palumbi, S., A. Martin, S. Romano, W. O. McMillan, L. Stice, and G. Grabowski. 2002. The simple fool's guide to PCR, version 2.0. Department of Zoology and Kewalo Marine Laboratory, University of Hawaii, Honolulu.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289-290.
- Parham, J. F., P. C. J. Donoghue, C. J. Bell, T. D. Calway, J. J. Head, P. A. Holroyd, J. G. Inoue, R. B. Irmis, W. G. Joyce, D. T. Ksepka, J. S. L. Patané, N. D. Smith, J. E. Traver, M. van Tuinen, Z. Yang, K. D. Angielczyk, J. M. Greenwood, C. A. Hipsley, L. Jacobs, P. J. Makovicky, J. Müller, K. T. Smith, J. M. Theodor, R. C. M. Warnock, and M. J. Benton. 2012. Best practices for justifying fossil calibrations. *Systematic Biology* 61:346-359.
- Pariselle, A. 2003. Considérations systématiques et évolutives à propos des Monogènes parasites de poissons des eaux douces tropicales. Habilitation à diriger les recherches. France, Perpignan University.
- Park, J.-K. and D. Ó Foighil. 2000. Sphaeriid and corbiculid clams represent separate heterodont bivalve radiations into freshwater environments. *Molecular Phylogenetics and Evolution* 14:75-88.
- Parodiz, J. J. and A. A. Bonetto. 1963. Taxonomy and zoogeographic relationships of the South American Naiades (Pelecypoda: Unionacea and Mutelacea). *Malacologia* 1:179-214.
- Pereira, D., M. C. D. Mansur, L. D. S. Duarte, A. S. Oliveira, D. M. Pimpão, C. T. Callil, C. Ituarte, E. Parada, S. Peredo, G. Darrigran, F. Scarabino, C. Clavijo, G. Lara, I. C. Miyahira, M. T. R. Rodriguez, and C. Lasso. 2014. Bivalve distribution in hydrographic regions in South America: historical overview and conservation. *Hydrobiologia* 735:15-44.

- Pfeiffer, J. M., D. L. Graf, K. S. Cummings, and L. M. Page. 2018. Molecular phylogeny and taxonomic revision of two enigmatic freshwater mussel genera (Bivalvia: Unionidae *incertae sedis*: *Harmandia* and *Unionetta*) reveals a diverse clade of Southeast Asian Parreysiinae. *Journal of Molluscan Studies* 84:404-416.
- Pfeiffer, J. M. I., J. W. Breinholt, and L. M. Page. 2019. Unioverse: A phylogenomic resource for reconstructing the evolution of freshwater mussels (Bivalvia, Unionoida). *Molecular Phylogenetics and Evolution* 137:114-126.
- Pfeiffer, J. M. I. and D. L. Graf. 2015. Evolution of bilaterally asymmetrical larvae in freshwater mussels (Bivalvia: Unionoida: Unionidae). *Zoological Journal of the Linnean Society* 175:307-318.
- Pickford, M. 1982. The tectonic, volcanics and sediments of the Nyanza rift valley. *Zeitschrift für Geomorphologie N. F.* 42:1-33.
- Pik, R., B. Marty, J. Carignan, G. Yirgu, and T. Ayalew. 2008. Timing of East African Rift development in southern Ethiopia: implication for mantle plume activity and evolution of topography. *Geology* 36:167-170.
- Poll, M. and J. G. Lambert. 1965. Contribution à l'étude systématique et zoogéographique des Procatopodinae de l'Afrique centrale (Pisces, Cyprinodontidae) *Bulletin de l'Académie royale des Sciences d'Outre-Mer* 2:615-631.
- Pook, C. E., U. Joger, N. Stümpel, and W. Wüster. 2009. When continents collide: phylogeny, historical biogeography and systematics of the medically important viper *Echis* (Squamata: Serpentes: Viperidae). *Molecular Phylogenetics and Evolution* 53:792-807.
- Puillandre, N., A. Lambert, S. Brouillet, and G. Achaz. 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* 21:1864-1877.

- R Core Team. 2018. R: A language and environment for statistical computing. <https://www.R-project.org/>. R Foundation for Statistical Computing, Vienna, Austria.
- Rabosky, D. L., M. C. Grundler, C. J. Anderson, P. O. Title, J. J. Shi, J. W. Brown, H. Huang, and J. G. Larson. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5:701-707.
- Rabosky, D. L., J. S. Mitchell, and J. Chang. 2017. Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Systematic Biology* 66:477-498.
- Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* 4:e1958 (doi: 10.1038/ncomms2958).
- Rambaut, A. 2009. FigTree, version 1.4.3, <http://tree.bio.ed.ac.uk/software/figtree>.
- Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian Phylogenetics using Tracer 1.7. *Systematic Biology* 67:901-904.
- Ree, R. H. and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57:4-14.
- Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217-223.
- Rögl, F. 1999. Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography: short overview. *Geologica Carpathica* 50:339-349.

- Roller, S., J. Hornung, M. Hinderer, and I. Ssemmanda. 2010. Middle Miocene to Pleistocene sedimentary record of rift evolution in the southern Albert Rift (Uganda). *International Journal of Earth Sciences* 99:1643-1661.
- Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* 46:195-203.
- Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Rüber, L., R. Britz, and R. Zardoya. 2006. Molecular phylogenetics and evolutionary diversification of labyrinth fishes (Perciformes: Anabantoidei). *Systematic Biology* 55:374-397.
- Salzburger, W., B. Van Bocxlaer, and A. S. Cohen. 2014. Ecology and evolution of the African Great Lakes and their faunas. *Annual Review of Ecology, Evolution and Systematics* 45:519-545.
- Scholz, H. and M. Glaubrecht. 2004. Evaluating limnic diversity: toward a revision of the unionid bivalve *Coelatura* Conrad, 1853 in the Great Lakes of East African and adjacent drainage systems (Mollusca, Bivalvia, Unionidae). *Mitteilungen aus dem Museum für Naturkunde in Berlin. Zoologische Reihe* 80:89-121.
- Schrenk, F., T. G. Bromage, A. Gorthner, and O. Sandrock. 1995. Paleoecology of the Malawi Rift: vertebrate and invertebrate faunal contexts of the Chiwondo Beds, northern Malawi. *Journal of Human Evolution* 28:59-70.
- Schultheiß, R., B. Van Bocxlaer, F. Riedel, T. von Rintelen, and C. Albrecht. 2014. Disjunct distributions of freshwater snails testify to a central role of the Congo system in shaping

- biogeographical patterns in Africa. *BMC Evolutionary Biology* 14:e42 (doi: 10.1186/1471-2148-14-42).
- Schwartz, J., E. R. Schwartz, E. Vreven, J. Snoeks, F. P. D. Cotterill, B. Misof, and U. K. Schliewen. 2012. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) triggered by Neogene landscape evolution. *Proceedings of the Royal Society B-Biological Sciences* 279:4389-4398.
- Seddon, M., C. Appleton, D. Van Damme, and D. L. Graf. 2011. Freshwater molluscs of Africa: diversity, distribution and conservation. Pp. 92-125 in W. Darwall, K. Smith, D. Allen, R. Holland, I. Harrison, and E. Brooks, eds. *The diversity of life in African freshwaters: underwater, under threat*. IUCN, Cambridge, UK, and Gland, Switzerland.
- Sil, M., N. A. Aravind, and K. P. Karanth. 2019. Role of geography and climatic oscillations in governing into-India dispersal of freshwater snails of the family: Viviparidae. *Molecular Phylogenetics and Evolution* 138:174-181.
- Simpson, C. T. 1900. Synopsis of the Naiades: or pearly fresh-water mussels. *Proceedings of the United States National Museum* 22:501-1044.
- Stadler, T. 2015. TreePar: Estimating birth and death rates based on phylogenies. R package version 3.3. <https://CRAN.R-project.org/package=TreePar>.
- Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
- Stankiewicz, J. and M. J. de Wit. 2006. A proposed drainage evolution model for Central Africa-Did the Congo flow east? *Journal of African Earth Sciences* 44:75-84.
- Stewart, K. M. 1997. Fossil fish from the Manonga Valley, Tanzania: description, paleoecology, and biogeographic relationships. Pp. 334-349 in T. Harrison, ed. *Neogene paleontology*

- of the Manonga valley, Tanzania. A window into evolutionary history of East Africa. *Topics in Geobiology* 14:333-349.
- Stewart, K. M. and A. M. Murray. 2017. Biogeographic implications of fossil fishes from the Awash River, Ethiopia. *Journal of Vertebrate Paleontology* 37:e1269115 (doi: 1269110.1261080/02724634.02722017.01269115).
- Strayer, D. L., D. C. Hunter, L. C. Smith, and C. K. Borg. 1994. Distribution, abundance, and roles of freshwater clams (*Bivalvia*, *Unionidae*) in the freshwater tidal Hudson River. *Freshwater Biology* 31:239-248.
- Talavera, G. and J. Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56:564-577.
- Thieme, M. L., B. Lehner, R. Abell, and J. Matthews. 2010. Exposure of Africa's freshwater biodiversity to a changing climate. *Conservation Letters* 3:324-331.
- Tsigenopoulos, C. S., P. Kasapidis, and P. Berrebi. 2010. Phylogenetic relationships of hexaploid large-sized barbs (genus *Labeobarbus*, *Cyprinidae*) based on mtDNA data. *Molecular Phylogenetics and Evolution* 56:851-856.
- Van Bocxlaer, B. 2011. Palaeobiology and evolution of the late Cenozoic freshwater molluscs of the Turkana Basin: *Unionidae* Rafinesque, 1820, partim *Coelatura* (*Bivalvia*: *Unionoidea*). *Journal of Systematic Palaeontology* 9:523-550.
- Van Damme, D. and A. Gautier. 1997. Late Cenozoic freshwater mollusks of the Wembere-Manonga Formation, Manonga valley, Tanzania. Pp. 351-360 in T. Harrison, ed. *Neogene paleontology of the Manonga valley, Tanzania. A window into evolutionary history of East Africa.* *Topics in Geobiology* 14:351-360.

- Van Damme, D. and A. Gautier. 2013. Lacustrine mollusc radiations in the Lake Malawi Basin: experiments in a natural laboratory for evolution. *Biogeosciences* 10:5767-5778.
- Van Damme, D. and M. Pickford. 1999. The late Cenozoic Viviparidae (Mollusca, Gastropoda) of the Albertine Rift Valley (Uganda-Congo). *Hydrobiologia* 390:171-217.
- Van Damme, D. and M. Pickford. 2010. The Late Cenozoic bivalves of the Albertine Basin (Uganda-Congo). *Geo-Pal Uganda* 2:1-121.
- Van Damme, D. and B. Van Bocxlaer. 2009. Freshwater molluscs of the Nile Basin, past and present. Pp. 585-629 *in* H. J. Dumont, ed. *The Nile: Origin, environments, limnology and human use*. Springer, New York.
- Van Steenberge, M., E. Vreven, and J. Snoeks. 2014. The fishes of the Upper Luapula area (Congo basin): a fauna of mixed origin. *Ichthyological Exploration of Freshwaters* 24:329-345.
- Vaughn, C. C. 2018. Ecosystem services provided by freshwater mussels. *Hydrobiologia* 810:15-27.
- Veatch, A. C. 1935. *Evolution of the Congo Basin*. The Geological Society of America, New York.
- West, K. and E. Michel. 2000. The dynamics of endemic diversification: Molecular phylogeny suggests explosive origin of the thiarid gastropods of Lake Tanganyika. Pp. 331-354 *in* A. Rossiter, and H. Kawanabe, eds. *Biology of ancient lakes: biodiversity, ecology and evolution*. Academic, London/San Diego.
- West, K., E. Michel, J. A. Todd, D. S. Brown, and J. Clabaugh. 2003. The gastropods of Lake Tanganyika: Diagnostic key, classification and notes on the fauna. Centre for African

- Wetlands, Ghana on behalf of International Association of Theoretical and Applied Limnology, Chapel Hill, NC.
- Westcott, W. A., W. N. Krebs, P. A. Bentham, and D. T. Pocknall. 2000. Miocene brackish water and lacustrine deposition in the Suez Rift, Sinai, Egypt. *Palaios* 15:65-72.
- Whelan, N. V., A. J. Geneva, and D. L. Graf. 2011. Molecular phylogenetic analysis of tropical freshwater mussels (Mollusca: Bivalvia: Unionoida) resolves the position of *Coelatura* and supports a monophyletic Unionidae. *Molecular Phylogenetics and Evolution* 61:504-514.
- Wilke, T., G. M. Davis, D. C. Qiu, and R. C. Spear. 2006. Extreme mitochondrial sequence diversity in the intermediate schistosomiasis host *Oncomelania hupensis robertsoni*: Another case of ancestral polymorphism? *Malacologia* 48:143-157.
- Wilson, A. B., M. Glaubrecht, and A. Meyer. 2004. Ancient lakes as evolutionary reservoirs: evidence from the thalassoid gastropods of Lake Tanganyika. *Proceedings of the Royal Society B-Biological Sciences* 271:529-536.
- Wilson, E. O. 1987. The little things that run the world (the importance and conservation of invertebrates). *Conservation Biology* 1:344-346.
- Wu, R.-W., T.-H. Chen, D. T. Zanatta, S. Ouyang, and X.-P. Wu. 2018. Reproductive traits of nine freshwater mussel species (Mollusca: Unionidae) from Poyang Lake, China. *Journal of Molluscan Studies* 84:324-332.
- Xia, X. 2017. Dambe6: New tools for microbial genomics, phylogenetics and molecular evolution. *Journal of Heredity* 108:431-437.
- Xia, X. and P. Lemey. 2009. Assessing substitution saturation with DAMBE. Pp. 615-630 *in* P. Lemey, M. Salemi, and A. M. Vandamme, eds. *The Phylogenetic Handbook: A Practical*

- Approach to DNA and Protein Phylogeny. 2nd edition. Cambridge University Press, Cambridge, UK.
- Xia, X., Z. Xie, M. Salemi, L. Chen, and Y. Wang. 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* 26:1-7.
- Yule, G. U. 1925. A mathematical theory of evolution: based on the conclusions of Dr. J.C. Willis. *Philosophical Transactions of the Royal Society of London, series B* 213:21-87.
- Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:686-693.
- Zhang, J., P. Kapli, P. Pavlidis, and A. Stamatakis. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29:2869-2876.
- Zieritz, A., A. E. Bogan, E. Froufe, O. Klishko, T. Kondo, U. Kovitvadhi, S. Kovitvadhi, J. H. Lee, M. Lopes-Lima, J. M. Pfeiffer, R. Sousa, V. Tu Do, I. Vikhrev, and D. T. Zanatta. 2018. Diversity, biogeography and conservation of freshwater mussels (Bivalvia: Unionida) in East and Southeast Asia. *Hydrobiologia* 810:29-44.

Tables

Table 1. Comparison of the six biogeographical models that were fit to our OTU phylogeny with their log-likelihood (LnL), number of parameters (NumPar), parameter values (d = diversification, e = extinction and j = jump-dispersal) and relative AICc support with weights (AICc_wt). The best-fit model is highlighted in bold.

Model	LnL	NumPar	d	e	j	AICc	AICc_wt
BEAST MCC tree							
DEC	-67.51	2	0.0058	1.00E-12	-	139.5	0.5700
DEC+J	-66.9	3	0.0050	1.00E-12	0.0077	140.9	0.2900
DIVALIKE	-69.69	2	0.0070	1.00E-12	-	143.9	0.0640
DIVALIKE+J	-68.28	3	0.0055	1.00E-12	0.0076	143.6	0.0730
BAYAREALIKE	-82.14	2	0.0073	8.60E-02	-	168.8	0.0000
BAYAREALIKE+J	-73.4	3	0.0045	1.00E-07	0.0150	153.9	0.0004
MrBayes consensus tree							
DEC	-72.95	2	0.1400	2.00E-01	-	150.4	0.3300
DEC+J	-70.97	3	0.1200	1.00E-12	0.0094	149.0	0.6700
DIVALIKE	-81.46	2	0.2000	8.10E-01	-	167.4	0.0001
DIVALIKE+J	-77.31	3	0.1300	1.00E-12	0.0090	161.7	0.0012
BAYAREALIKE	-86.86	2	0.2100	2.02E+00	-	178.2	0.0000
BAYAREALIKE+J	-87.07	3	0.3100	3.24E+00	0.0040	181.2	0.0000

Figures

Fig. 1. Phylogenetic relationships in Coelaturini reconstructed with Bayesian inference (MrBayes) using data from four gene fragments (COI, 16S rRNA, H3 and 28S rRNA). As the maximum likelihood reconstruction (RAxML) is very similar we indicate Bayesian posterior probabilities and bootstrap support values for relevant branches. Colored symbols for the 11 ingroup clades are used in fig. 4A. Representative taxa are displayed in the inset: a) *Pseudospatha tanganyicensis*, b) *Grandidieria burtoni*, c) *Prisodontopsis aviculaeformis*, d) *Mweruella mweruensis*, e) *Coelatura aegyptiaca*, f) *Nitia acuminata*, g) *Nyassunio nyassaensis*, h) *Moncetia anceyi*.

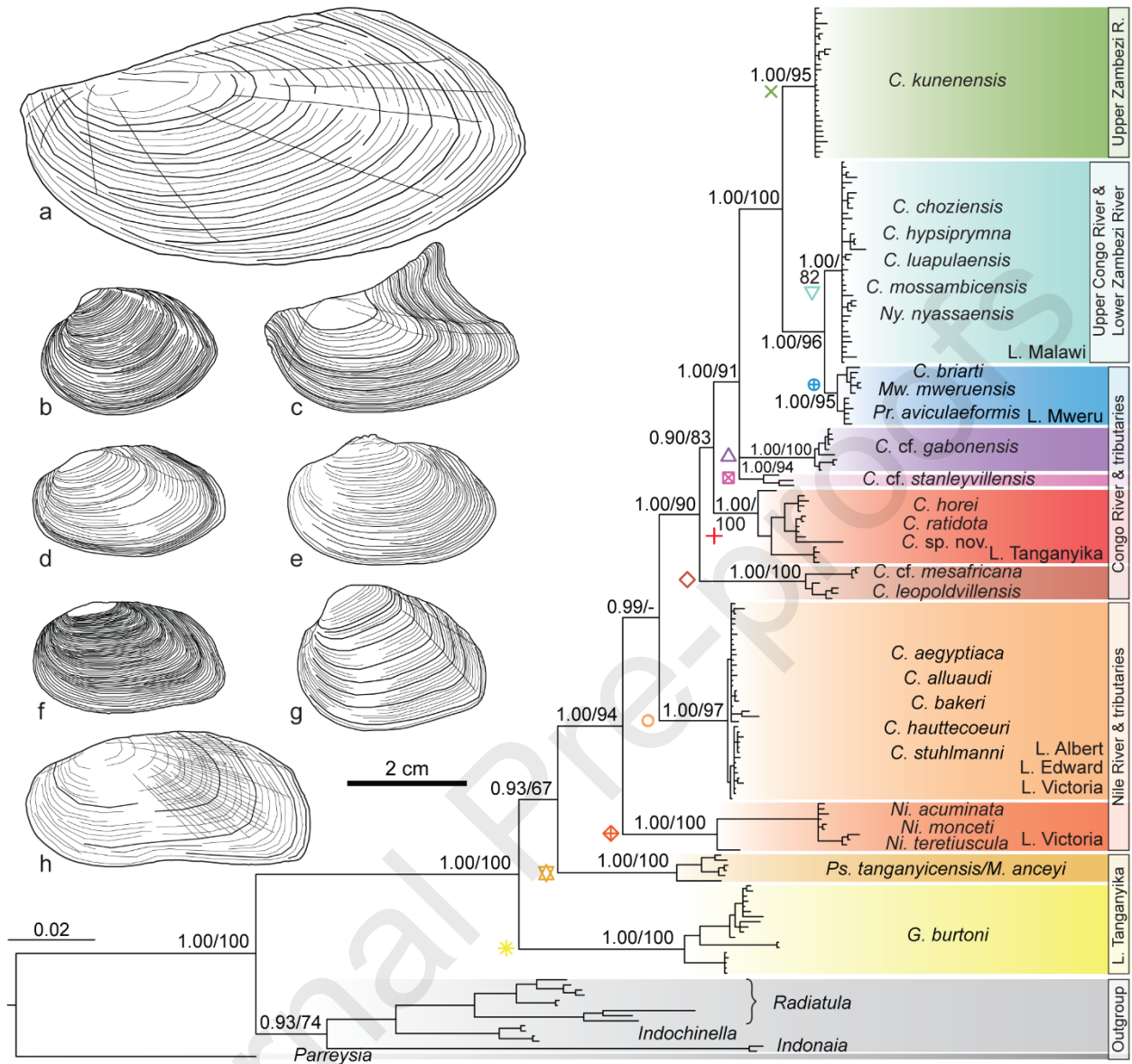
Fig. 2. Fossil-calibrated MCC tree reconstructed with BEAST. Posterior probabilities (BPP) are indicated with colored circles, node bars denote 95% HPD intervals of node ages and asterisks indicate the nodes that were used for fossil calibration. *Parreysia* was included in the analysis but removed from the tree to increase readability for the ingroup.

Fig. 3. OTU-based trees reconstructed from 1,000 trees that were randomly sampled from the posterior distribution of BEAST. The inset indicates how lineages accumulated through time (in blue), which compares well to pure-birth and birth-death models for this dataset (in red). A decrease in diversification rates was observed ~ 3 Ma ago in TreePar.

Fig. 4. Diversification and biogeographical analyses. A) The distribution of Coelaturini across Africa is highlighted in transparent purple, together with the sampled localities; the boundaries of freshwater ecoregions are marked in grey; sampling localities are indicated with the clade-

specific color-coded symbols of Fig. 1 to reconstruct the geographic distribution of the 11 clades of Coelaturini from Figs 1 & 2. B) Speciation and extinction rates as inferred from the BAMM analysis. C) OTU-based phylogeny with ancestral range estimations derived from the DEC model and the results of biogeographical stochastic mapping. Dashed branches indicate anagenetic dispersal events, whereas letters at nodes indicate cladogenetic events of s = subset sympatry, y = sympatry, and v = vicariance.

Fig. 5. Reconstruction of hydrographic connections throughout the East African Rift System (EARS) during A) the Late Miocene (~10 Ma) and B) Early Pliocene (~4 Ma), with arrows representing colonization patterns of Coelaturini during these (sub-)epochs. In the Early to Middle Miocene Parreysiinae colonized the EARS from Eurasia. The MRCA of Coelaturini subsequently colonized the Victoria and Tanganyika ecoregions (black arrows). A secondary colonization of the Tanganyika ecoregion occurred via the Congo, from where the Okavango was also colonized (grey arrows). During the Pliocene, Coelaturini were established in three faunal provinces directly within the East African Rift (bioregions **S** = Upper Congo, **T** = Tanganyika, and **V** = Victoria; indicated in bold). The Nile was colonized from the larger Victoria ecoregion, whereas the Mweru and Malawi regions from the Okavango via the Upper Zambezi (black arrows). Finally, in the Pleistocene, a recolonization of the Congo headwaters from Lake Mweru and dispersal to the Malagarasi and Eastern coastal Africa occurred from Lakes Tanganyika and Malawi, respectively (grey arrows). Paleolakes are indicated in dark grey whereas wetlands in light grey.



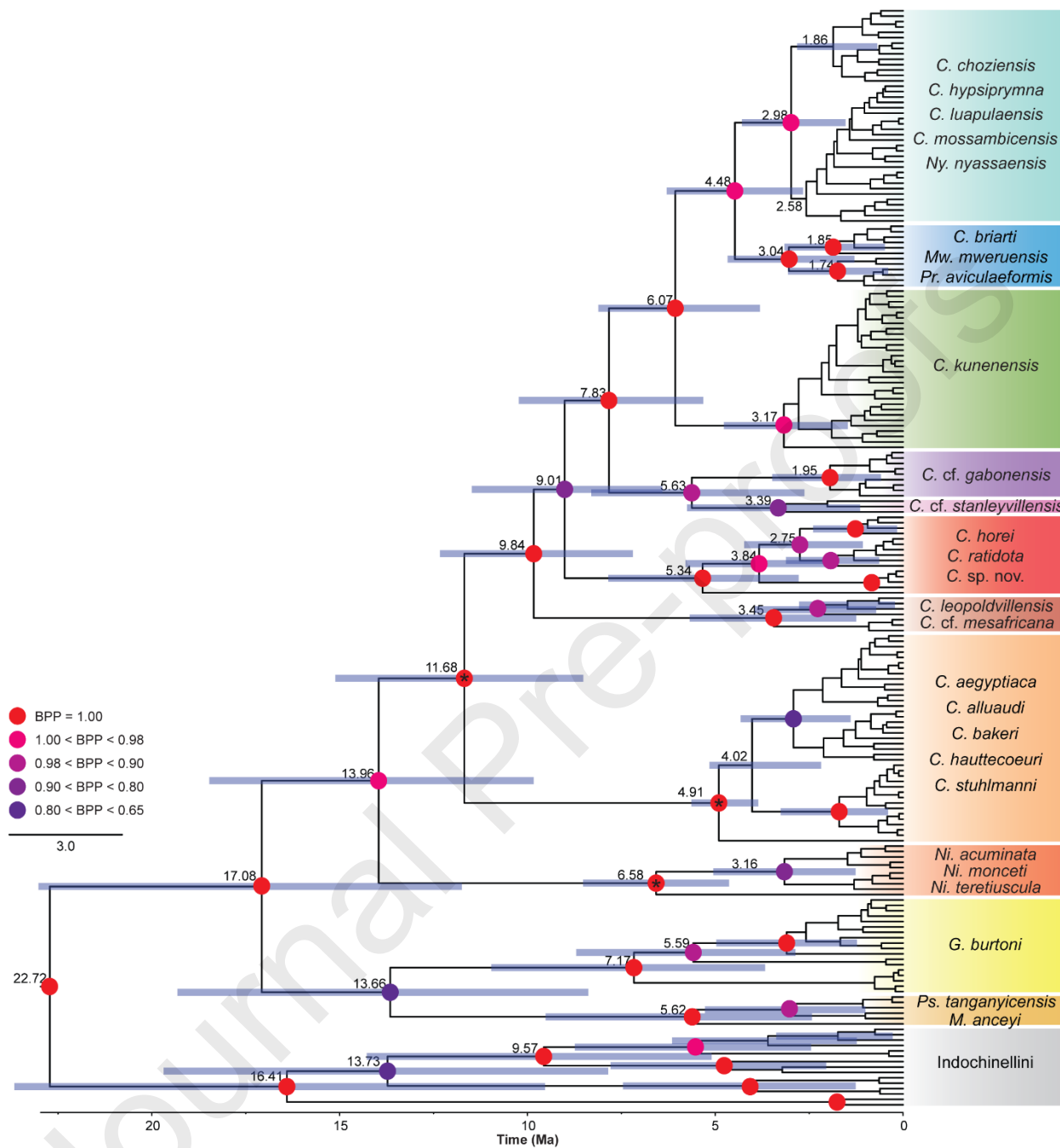


Figure 3

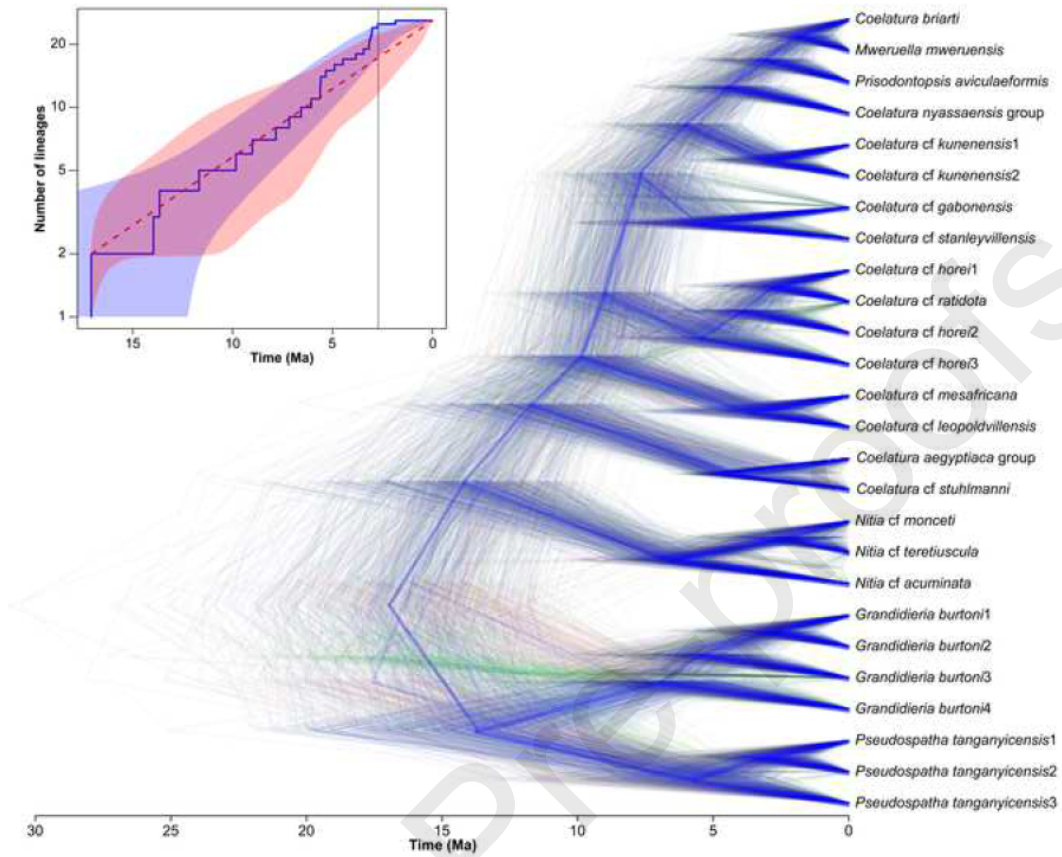
[Click here to download Figure Ortiz_et_al_Figure3.tif](#)


Figure 4

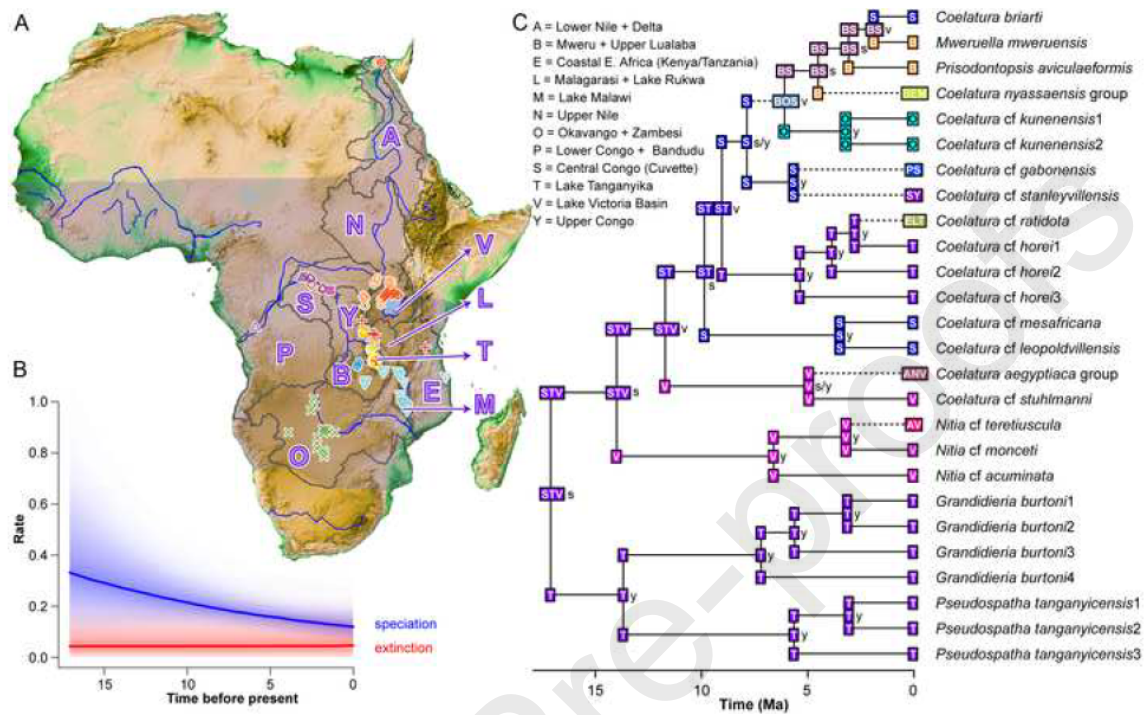
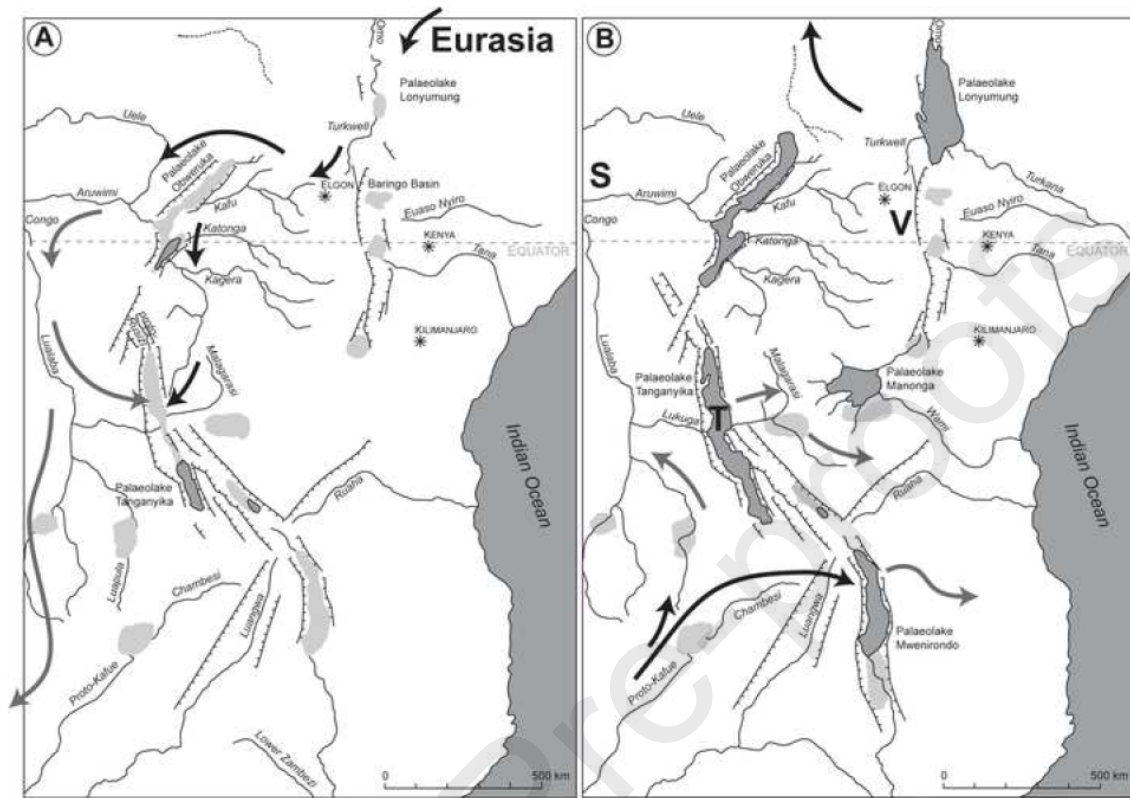
[Click here to download Figure Ortiz_et_al_Figure4_revised.tif](#)


Figure 5

[Click here to download Figure Ortiz_et_al_Figure5.tif](#)


**Diversification dynamics of freshwater bivalves (Unionidae: Parreysiinae: Coelaturini)
indicate historic hydrographic connections throughout the East African Rift System**

Claudia M. Ortiz-Sepulveda^{1,*}, Björn Stelbrink^{2,3}, Xavier Vekemans¹, Christian Albrecht², Frank Riedel⁴, Jonathan A. Todd⁵, Bert Van Bocxlaer¹

Highlights

Aftropical Unionidae and *Moncetia* belong to Coelaturini

Coelaturini diversified at a steady pace after colonization of Africa ~17 Ma

Most diversification events occurred in sympatry and clades coincide with drainage basins

Lake Tanganyika was colonized at least twice followed by intralacustrine diversification

Biogeographic patterns in Coelaturini resemble those in other taxa, including fishes

Author Statement

BVB and CA conceived and designed the study; BVB, XV and CA supervised the project; CMOS, CA, FR, JAT and BVB collected material; CMOS performed molecular biology; CMOS performed identifications, which were verified by BVB; CMOS led data curation; CMOS, BS and BVB performed analyses; CMOS produced the figures, with help of BVB; all authors contributed to the interpretation of data; CMOS and BVB drafted the initial manuscript and all authors contributed to later versions. All authors read and approved the final manuscript.

Graphical abstract

