Reconstructing the phylogeny and evolutionary history of freshwater fishes (Nemacheilidae) across Eurasia since early Eocene

Vendula Šlechtová[1], Tomáš Dvořák[1,2], Jörg Freyhof[3], Maurice Kottelat[4], Boris Levin[5,6], Alexander Golubtsov[6], Vlastimil Šlechta[1], Jörg Bohlen[1§]

[1]Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, Liběchov, Czech Republic
[2] Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic
[3] Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, 10115 Berlin, Germany.
[4] Rue des Rauraques 6, 2800 Delémont, Switzerland (permanent address) and Lee Kong Chian Natural History Museum, National University of Singapore, Singapore
[5] Papanin Institute of Biology of Inland Waters, Russian Academy of Sciences, Borok, Yaroslavl Region, Russia
[6] A.N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences, Moscow, Russia

[§] Corresponding author

Author information:
Vendula Šlechtová: 0000-0002-5499-7550; v.slechtova@iapg.cas.cz
Tomáš Dvořák: 0000-0001-8151-2144; dvorak@iapg.cas.cz
Jörg Freyhof: 0000-0002-7042-3127; joerg.freyhof@mfn.berlin
Maurice Kottelat: 0000-0003-3638-2824; mkottelat@dplanet.ch
Boris Levin: 0000-0002-4044-2036; borislyovin@mail.ru
Alexander Golubtsov: 0000-0002-8317-7527; sgolubtsov@gmail.com
Jörg Bohlen: 0000-0003-1259-008x; joerg_bohlen@yahoo.de

Abstract

Eurasia has undergone substantial tectonic, geological, and climatic changes throughout the Cenozoic era, primarily associated with tectonic plate collisions and a global cooling trend. The evolution of present-day biodiversity unfolded in this dynamic environment, characterised by intricate interactions of abiotic factors. However, comprehensive, large-scale reconstructions illustrating the extent of these influences are lacking. We reconstructed the evolutionary history of the freshwater fish family Nemacheilidae across Eurasia and spanning most of the Cenozoic era on the base of 471 specimens representing 279 species and 37 genera. Molecular phylogeny using 6 genes uncovered six major clades within the family, along with numerous unresolved taxonomic issues. Dating of cladogenetic events and ancestral range estimation traced the origin of Nemacheilidae to Indochina around 48 million years ago. Subsequently, one branch of Nemacheilidae colonised eastern, central, and northern Asia, as well as Europe, while another branch expanded into the Burmese region, the Indian subcontinent, the Near East, and northeast Africa. These expansions were facilitated by tectonic connections, favourable climatic conditions, and orogenic processes. Conversely, aridification emerged as the primary cause of extinction events. Our study marks the first comprehensive reconstruction of the evolution of Eurasian freshwater biodiversity on a continental scale and across deep geological time.

Introduction

The present-day species- and genus-level diversity on earth is mainly the result of the evolution during Cenozoic, when biodiversity recovered from the Cretaceous-Tertiary mass extinction event, which eradicated three-quarters of all animal and plant species. During the Cenozoic new lifeforms flourished; often from lineages that had played only a minor role before the mass extinction; like mammals, birds and teleost fishes (Feduccia, 1995; Friedman, 2010).

The Cenozoic is also a period of significant and relatively rapid climatic and geological changes (Summerhayes, 2015). The biggest geological changes occurred in Eurasia, resulting from its collision with the African, Arabian and Indian plates, that caused the closure of the Tethys Sea and the Alpine - Himalayan orogeny (Trifonov et al., 2012). The climate also changed dramatically during Cenozoic; transitioning Earth from a warm and humid hothouse state to a cool and dry icehouse (Westerhald et al., 2020); but with a complex interplay of large-scale changes and smaller-scale fluctuations, creating a diverse range of climatic conditions on the local scale (Mudelsee et al., 2014).

Geological and climatic changes are major drivers of biodiversity evolution (Antonelli et al., 2018; Fritz et al., 2016; Mayhew et al., 2012). Numerous biogeographical studies show distribution patterns and local vicariance and radiation events in the Eurasian biodiversity (e.g. Andreyenkova et al., 2021; Ishii et al., 2023). The majority of these studies focused on terrestrial fauna and flora, but we believe that the best suited model organisms for biogeographical analysis are strict freshwater species. Such hololimnic species are confined to a given water body unless a physical connection with another water body appears. At the same time, the evolution of the earth's hydrographic network

is caused by local geomophological and climatic conditions. The locally ‘locked’ freshwater fauna thus reflects the environmental past of any region on earth in much more detail than most terrestrial animals or plants (e.g. Bănărescu, 1992; Bohlen et al., 2020a).

Despite this potential, there is a paucity of research on the evolutionary history of Eurasian freshwater biodiversity covering the large geographic and temporal scales. While numerous studies have examined certain geological or climatic changes during specific periods of the Cenozoic (e.g. Ballarin and Li, 2018; Gu et al., 2022; Kahlke, 2014) or within limited regions of Eurasia (e.g. Delić et al., 2020; Deng et al., 2020; Saito et al., 2018), a comprehensive analysis encompassing the whole Eurasia and through most of the Cenozoic era and considering climatic as well as geological events as driving forces of biodiversity evolution is lacking.

For the present study we chose the species-rich (presently 838 species) freshwater fish family Nemacheilidae (stone loaches) as model; which is present in almost all water bodies of Europe and Asia (Dvořák et al., 2022; Kottelat, 2012) and stretches from Spain to Japan in west-east direction and from Russia to Indonesia in north-south direction. Most species are small (< 12 cm SL), benthic and inhabit streams and rivers, although a few are found in swamps, caves and lakes. Few even occupy extreme environments: Nemacheilidae include the deepest freshwater cave fish in the world (*Triplophysa gejiuensis*, 400 m below surface) and the fish living at the highest altitude (*Triplophysa stolickai*, 5200 m a.s.l.) (Chu and Chen, 1979; Kottelat, 2012).

The aim of this study is to reconstruct the evolutionary history of the family Nemacheilidae as example for the evolution of freshwater fauna across Eurasia and since early Cenozoic. The research concept includes a phylogenetic reconstruction based on molecular genetic data, followed by the dating of cladogenetic events and a biogeographic

analysis in order to reconstruct ancestral distribution ranges. The results are compared with the known geological and climatic history of Eurasia to identify the causative factors and processes that have influenced the diversification of Eurasian freshwater fishes.

Results

Major clades within Nemacheilidae

To understand the phylogenetic diversity of the family Nemacheilidae and to identify the largest units within, we analysed sequence data for one mitochondrial (cytochrome b) and up to five nuclear (EGR3, IRBP2, Myosin6, RAG1, Rhodopsin1) genes of 471 specimens representing 279 species and 37 genera from across Eurasia (Fig. 1; Supplementary Material Table S1). This is by far the most comprehensive genetic dataset of Nemacheilidae ever published. The outgroup included 21 species from 9 related families (Fig. 2). We applied maximum likelihood analysis (ML) in IQ-TREE and Bayesian inference (BI) in mrBayes on the concatenated dataset. Besides, the individual unrooted ML gene trees were reconstructed in IQ-TREE and subsequently used as input files for ASTRAL III (Zhang et al., 2018) to infer species tree. All analyses revealed Nemacheilidae as monophyletic group and recovered the presence of six monophyletic major clades within the dataset.

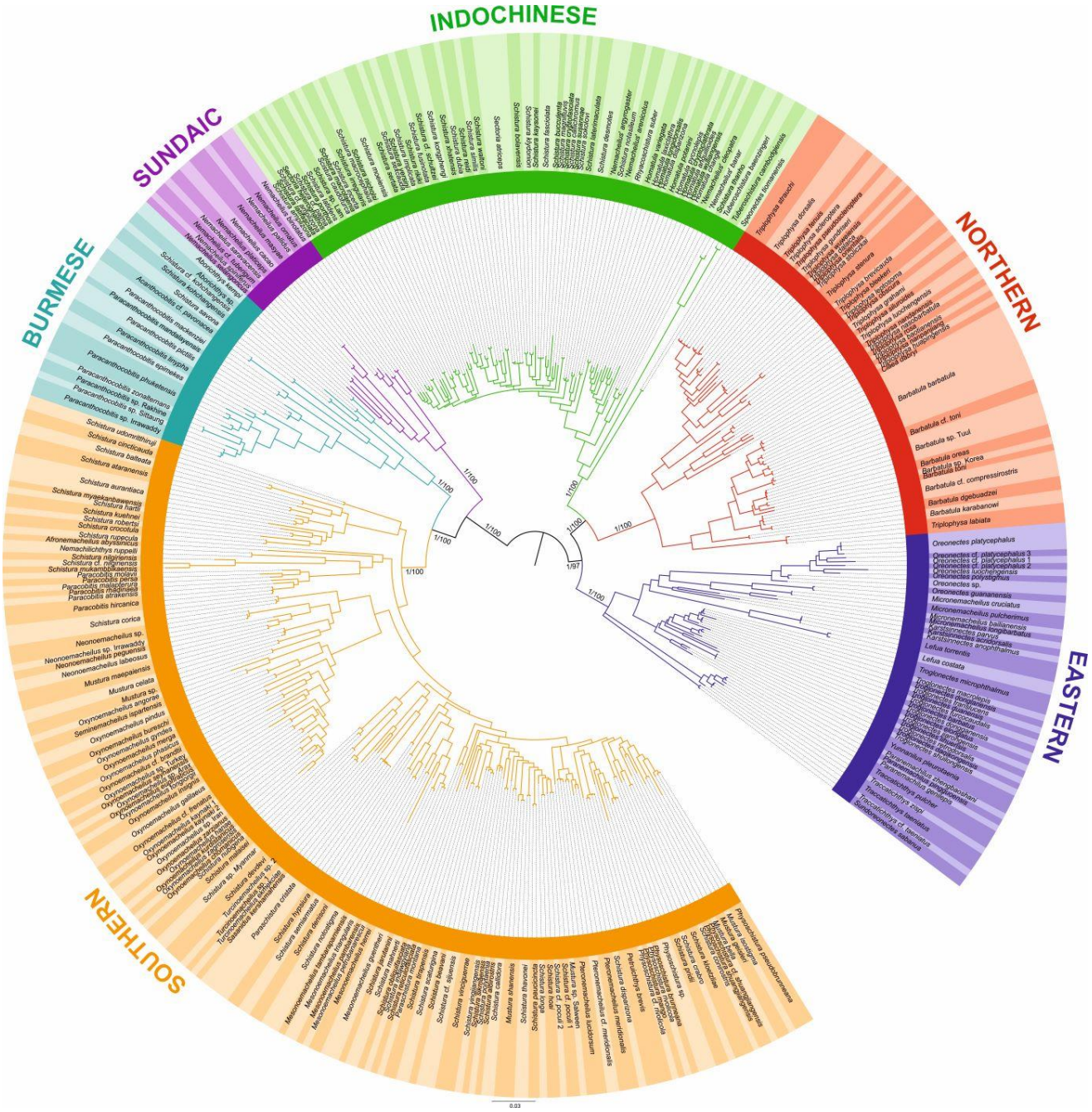

Fig. 1. Phylogenetic tree of Nemacheilidae, including all 471 analysed nemacheilid specimens, reconstructed in MrBayes based on six loci. The topologies of BI and ML trees were congruent. The values at the basal nodes correspond to Bayesian posterior probabilities and ML bootstrap supports, respectively. Only the ingroup is shown.

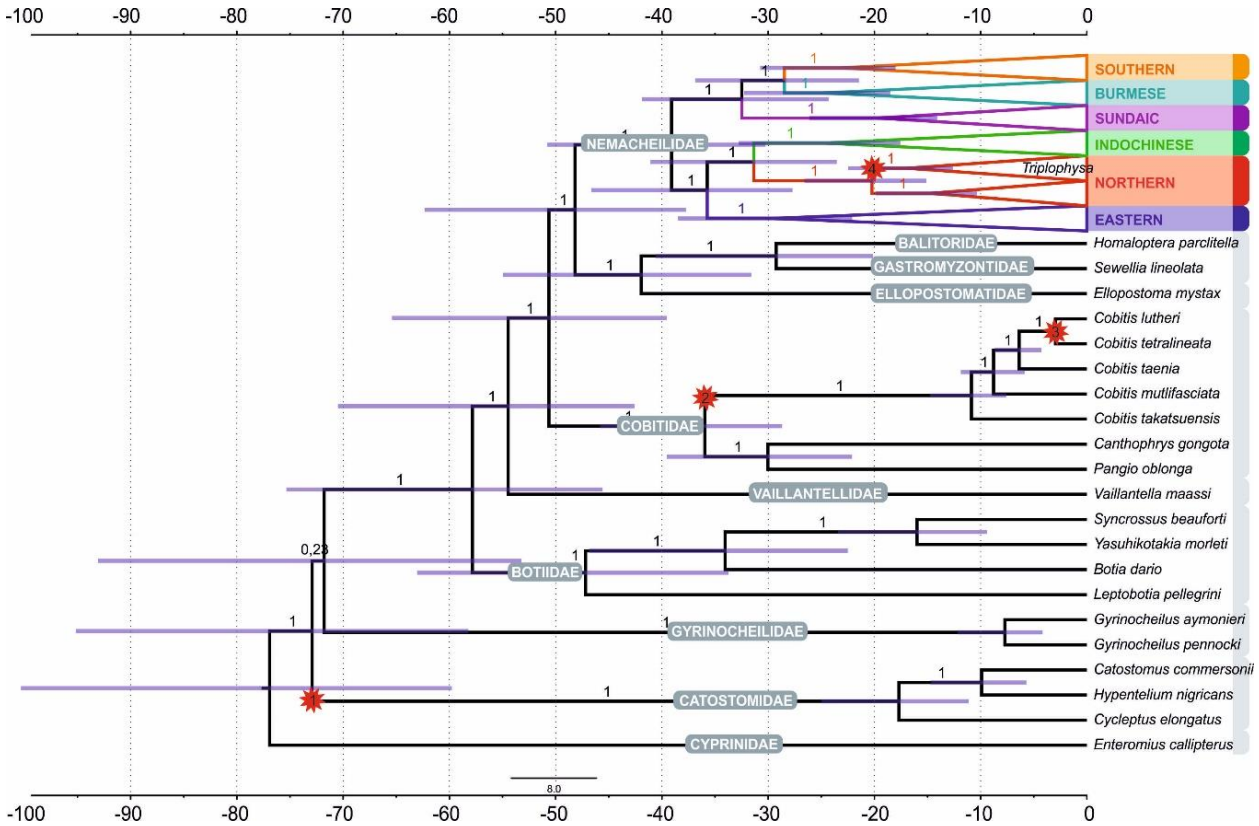

Fig. 2. Divergence time estimation resulting from Bayesian divergence time analysis of concatenated dataset in BEAST 2: maximum clade credibility (MCC) tree of the whole dataset, with the Nemacheilidae ingroup collapsed into main clades. Red stars indicate calibration points, values at branches represent posterior probabilities (PPs lower than 0.90 not shown) and blue horizontal bars denote 95% HPD. The time scale is in millions of years.

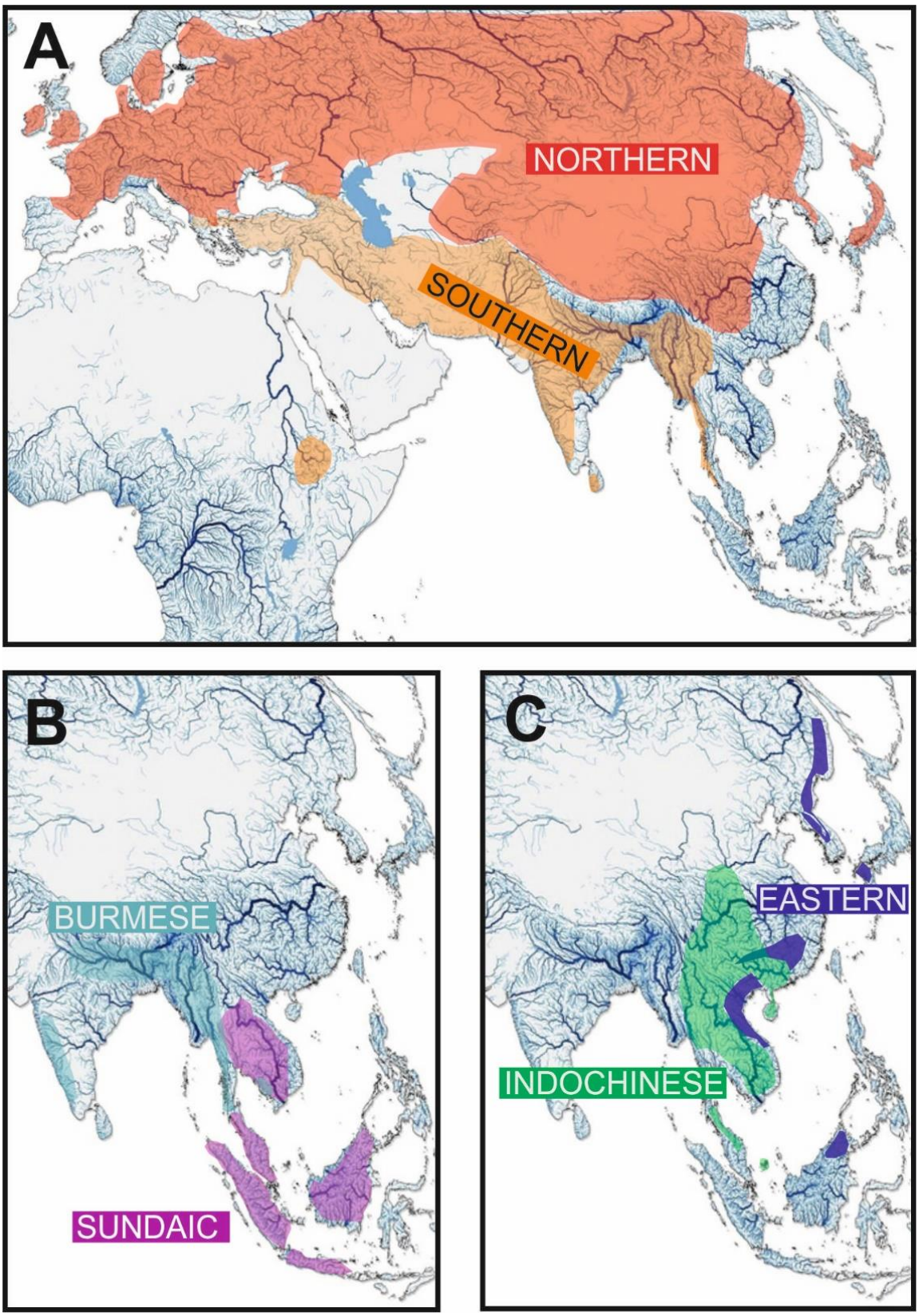

Fig. 3. Geographic distribution of the six major clades of Nemacheilidae.

Each major clade exhibited a distinct distribution (Fig. 3). The most widely distributed is the **'Northern Clade'** (Fig. 3A), which inhabits the northern parts of Asia and most of Europe from Japan to Spain and southwards to Yunnan. In our analyses it includes all species of the genera *Barbatula, Claea* and *Triplophysa*. The **'Southern Clade'** (Fig. 3A) has the second largest distribution and inhabits the Indian subcontinent, the Burmese region and West Asia from the Salween River to southeast Europe; isolated species are found also in Indochina and East Africa (Ethiopia). The Southern Clade is the clade with the highest taxonomic diversity including all analysed species of 14 genera (*Afronemacheilus, Mesonoemacheilus, Mustura, Nemachilichthys, Neonoemacheilus, Oxynoemacheilus, Paracobitis, Paraschistura, Petruichthys, Physoschistura*, *Pteronemacheilus, Sasanidus, Seminoemacheilus* and *Turcinoemacheilus*) plus several species of the genus *Schistura*. Another clade is the **'Burmese Clade'** (Fig. 3B) with a continuous distribution from western Thailand to Northeast India, plus isolated species in the Western Ghats of India and in eastern Thailand and Cambodia. It comprises all analysed species of the genera *Aborichthys, Acanthocobitis* and *Paracanthocobitis,* as well as several species classified as *Schistura.* The **'Sundaic Clade'** (Fig. 3B) includes only the species of the genus *Nemacheilus* and is distributed across Sundaland (the Great Sunda Islands Borneo, Sumatra and Java and the Malay Peninsula) and in the Indochinese rivers draining to the Gulf of Thailand (Mekong, Mae Khlong and Chao Phraya). The **'Eastern Clade'** (Fig. 3C) is found along the eastern margin of Asia from the Russian Far East to eastern Borneo. Its disjunct distribution consists of a single genus (*Sundoreonectes*) in northeast Borneo, the majority of genera and species in northern

Vietnam and southeast China and a single genus (*Lefua*) on the Korean Peninsula, the Japanese Archipelago, northeast China and the Russian Far East. The Eastern Clade comprises all analysed species of the genera *Karstsinnectes, Lefua, Micronemacheilus, Oreonectes, Paranemachilus, Sundoreonectes, Traccatichthys, Troglonectes* and *Yunnanilus.* The sixth clade (**'Indochinese Clade'**, Fig. 3C) has a continuous distribution throughout Thailand, Laos, Cambodia, Vietnam, central Myanmar and southern China, as well as isolated species in the upper/middle Yangtze and on Tioman Island (Malaysia). It is composed of all analysed species belonging to *Homatula, Rhyacoschistura, Sectoria, Spaeonectes* and *Tuberoschistura*, in addition to many species currently assigned to *Schistura* and four species of uncertain generic assignment ('*Nemacheilus*').

The distribution areas of the six major clades are not mutually exclusive. While the Northern, Indochinese, and Southern Clades exhibit limited overlap with each other, the Sundaic Clade shares its northern range with the Indochinese Clade. The Eastern Clade overlaps with the Northern Clade in the north, the Indochinese Clade in the middle, and the Sundaic Clade in the south. The Burmes Clade shares most of its distribution area with the Southern Clade.

Analyses of the concatenated dataset as well as ASTRAL analysis inferring a species tree from single gene trees revealed a basal dichotomy in the family Nemacheilidae (Fig. 1,2,4). The one branch consists of the Eastern Clade as sister to the pair Indochinese plus Northern Clade. The second branch is made of the Sundaic Clade as sister to the pair Burmese and Southern Clade. All basal nodes and most of tip nodes are supported by high statistical supports in all analyses.

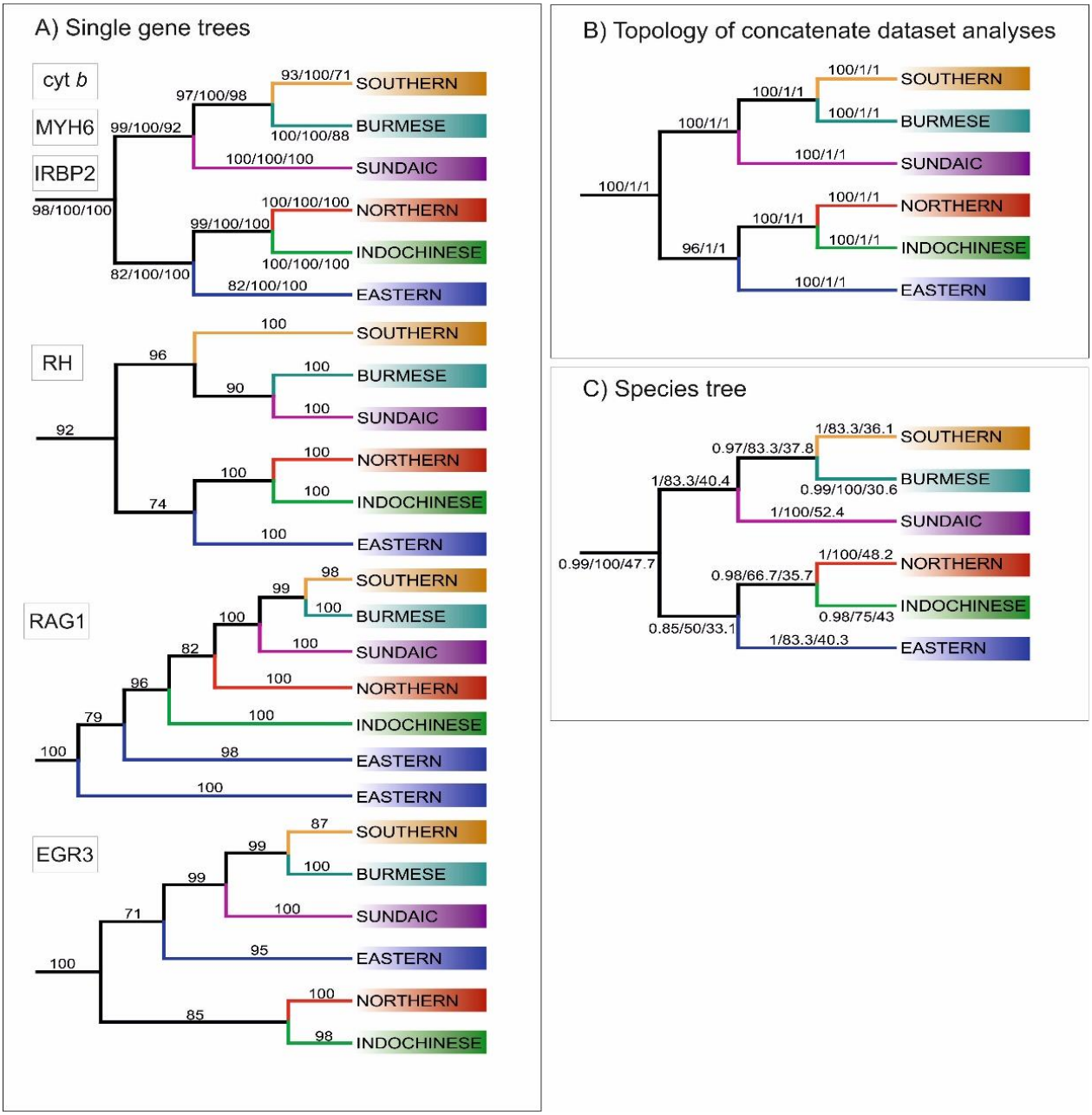

Fig. 4. Topologies of main clades derived from various analyses. A) Topologies recovered by ML analyses of single gene trees, with values at branches indicating bootstrap supports. Topologies congruent to the concatenated data-derived as well as species tree is revealed by Cyt b, MYH6 and IRBP2. Three slightly different topologies were derived from RH, EGR3 and RAG1 B) Topology obtained from analyses of concatenate dataset, the values at the branches correspond to ML/MrBayes PP/BEAST PP supports, respectively. C) Topology inferred by ASTRAL-III using the unrooted ML single gene trees as input. The values at the branches correspond to support values/gCF/sCF.

Divergence times

The ages of the genealogical events were determined using three fossil records and one geological event as calibration points (details of calibration under ‘Methods’). The results indicate that Nemacheilidae shared the last common ancestor with other families of Cobitoidea about 48.5 mya (Fig. 4). The first notable split that separated the family into two big branches took place around 39 mya (for details see Fig. 5).

Biogeographic reconstructions

The biogeographic reconstruction of Nemacheilidae considered 11 geographic regions within Eurasia and northeast Africa. The results show that Nemacheilidae originated in the region of nowadays Indochina and expanded from there in multiple waves. More details are given in Fig. 6.

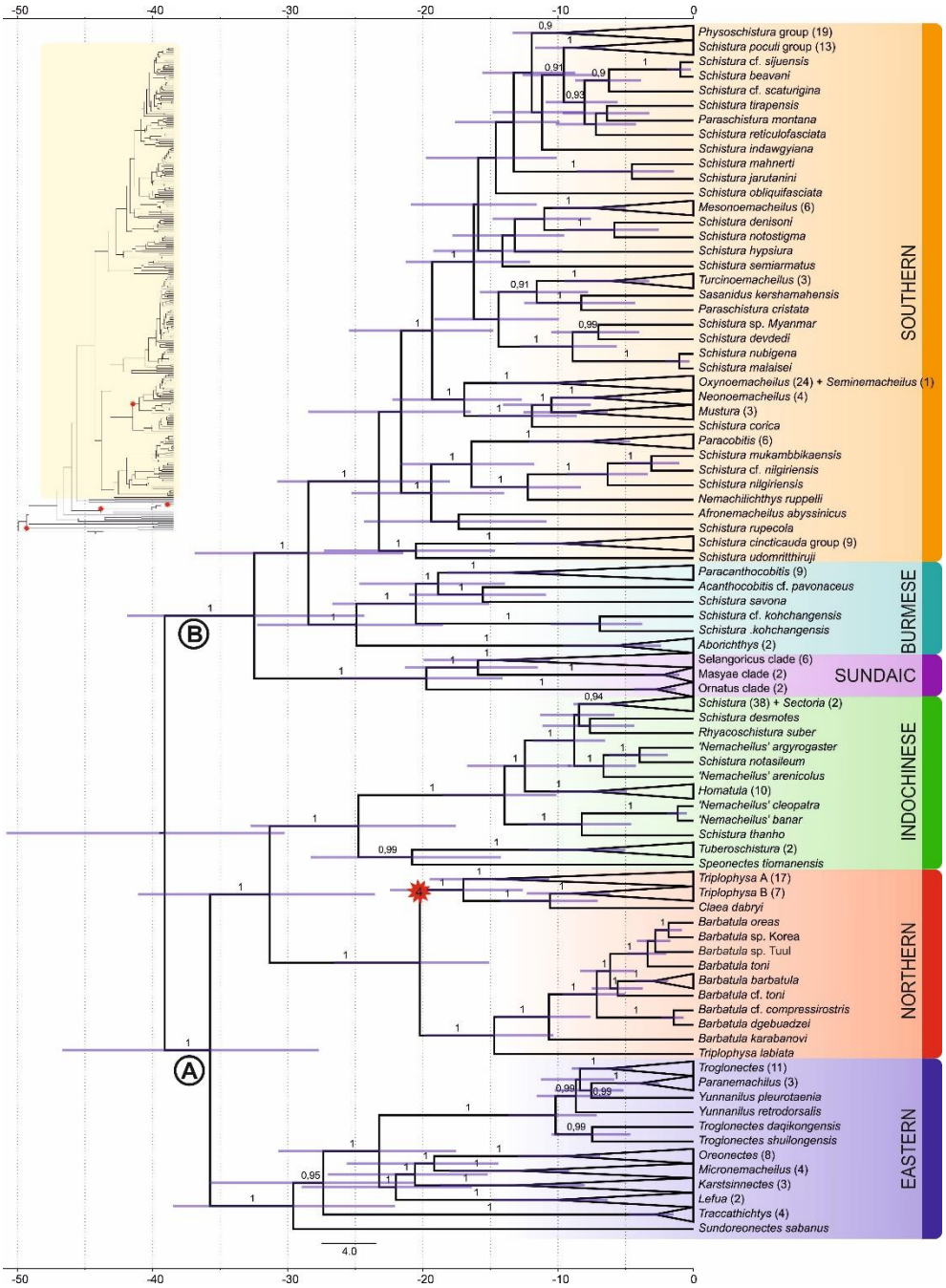

Fig. 5. Divergence time estimation: Ingroup of the maximum clade credibility (MCC) tree resulting from Bayesian divergence time analysis of concatenated dataset in BEAST 2. The red star indicates internal calibration point, the values at the nodes represent posterior probabilities (PPs lower than 0.90 not shown), and the blue bars relevant 95% HPD. The time scale is in millions of years. Inset in top left corner shows complete tree: Nemacheilidae is highlighted in yellow, while the outgroup is not highlighted; all calibration points are indicated by red stars.

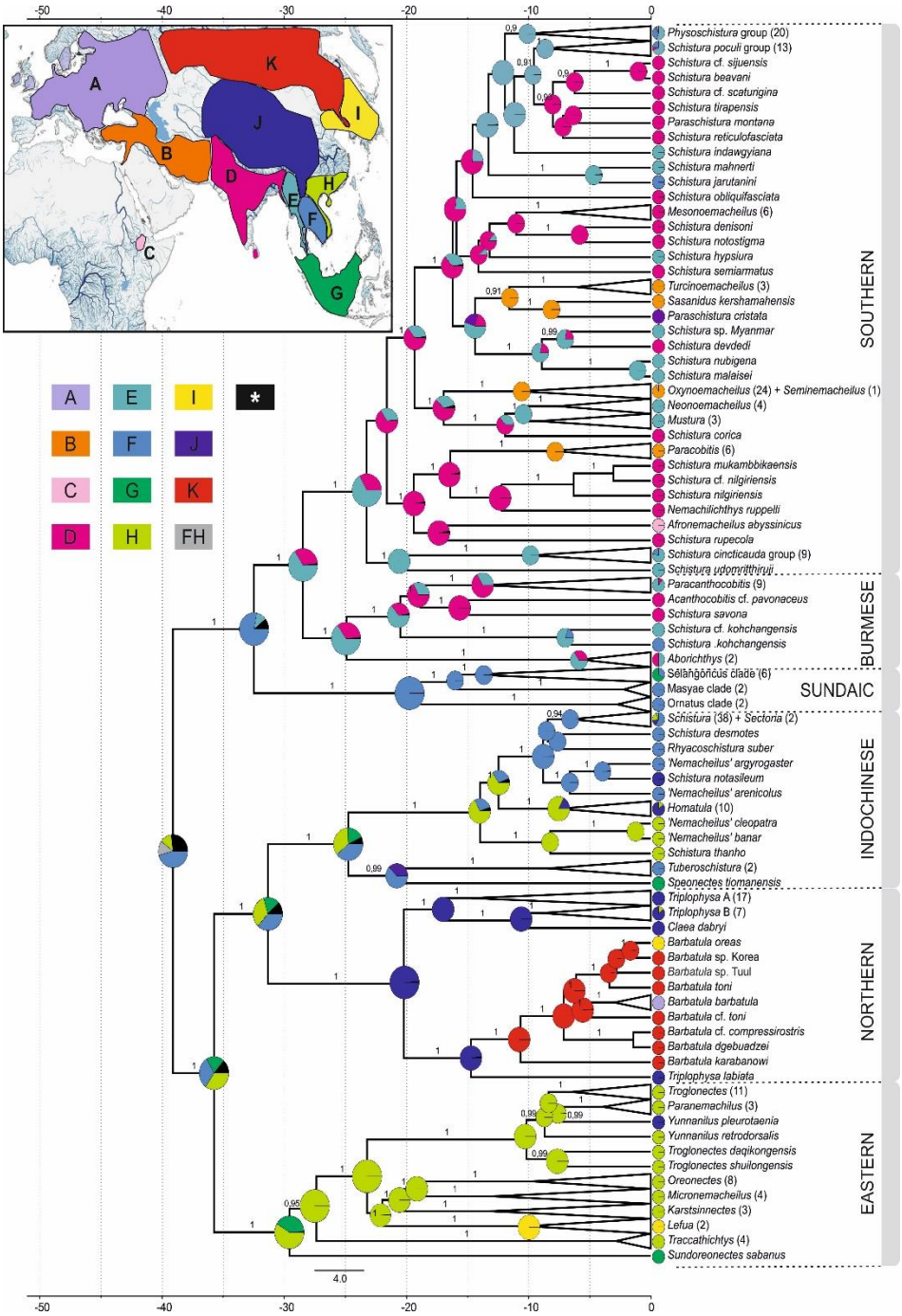

Fig. 6. Reconstruction of the biogeographical history of Nemacheilidae. The inset map in top left corner shows the defined biogeographical regions of Eurasia recognised in the analysis; below the map the colour – region code for the main tree. Grey colour indicates combination of two regions as ancestral (FH); black colour indicates ancestral ranges identified with likelihood < 10%. Main tree: Ancestral range reconstruction with use of DEC+J model based on the phylogeny derived from BEAST2. Pie charts at the nodes indicate relative likelihood of ancestral ranges of most recent common ancestors (MRCA), while the current distribution of taxa is indicated at the tips.

Discussion

The evolutionary history of Nemacheilidae (Fig. 7, 8, Table 1)

Early Eocene (56 – 48 mya) - the origin of Nemacheilidae (Fig. 8A)

The family Nemacheilidae shared the last common ancestor with the other Cypriniformes about 48.5 mya (Fig. 2), and our biogeographic analysis identified the region of nowadays Indochina as most likely place of origin of Nemacheilidae (Fig. 6). During the early Eocene, Southeast Asia experienced a warm and perhumid climate, characterised by both highlands and lowlands with numerous rivers (Morley, 2018); conditions under which stone loaches flourished.

Middle to late Eocene (48 – 34 mya) – first range expansions and formation of the first major clades (Fig. 8B,C)

The first detectable cladogenetic event within the family occurred about 39 mya and separated a lineage formed by the Eastern, Northern and Indochinese Clades (Branch A in Fig. 5) from a lineage including the Sundaic, Burmese and Southern Clades (Branch B). Our biogeographic analysis of the dataset revealed Indochina as most parsimony area of the first cladogenetic event; and consequently as region of origin for branch A as well as branch B.

The Eastern Clade was the first of the six present major clades to form: it separated about 35.8 mya from the common ancestor of the Indochinese and Northern Clades, either in the region of Indochina or Vietnam-south China. Taking into consideration its present-day distribution (Fig. 3) we assume that the young

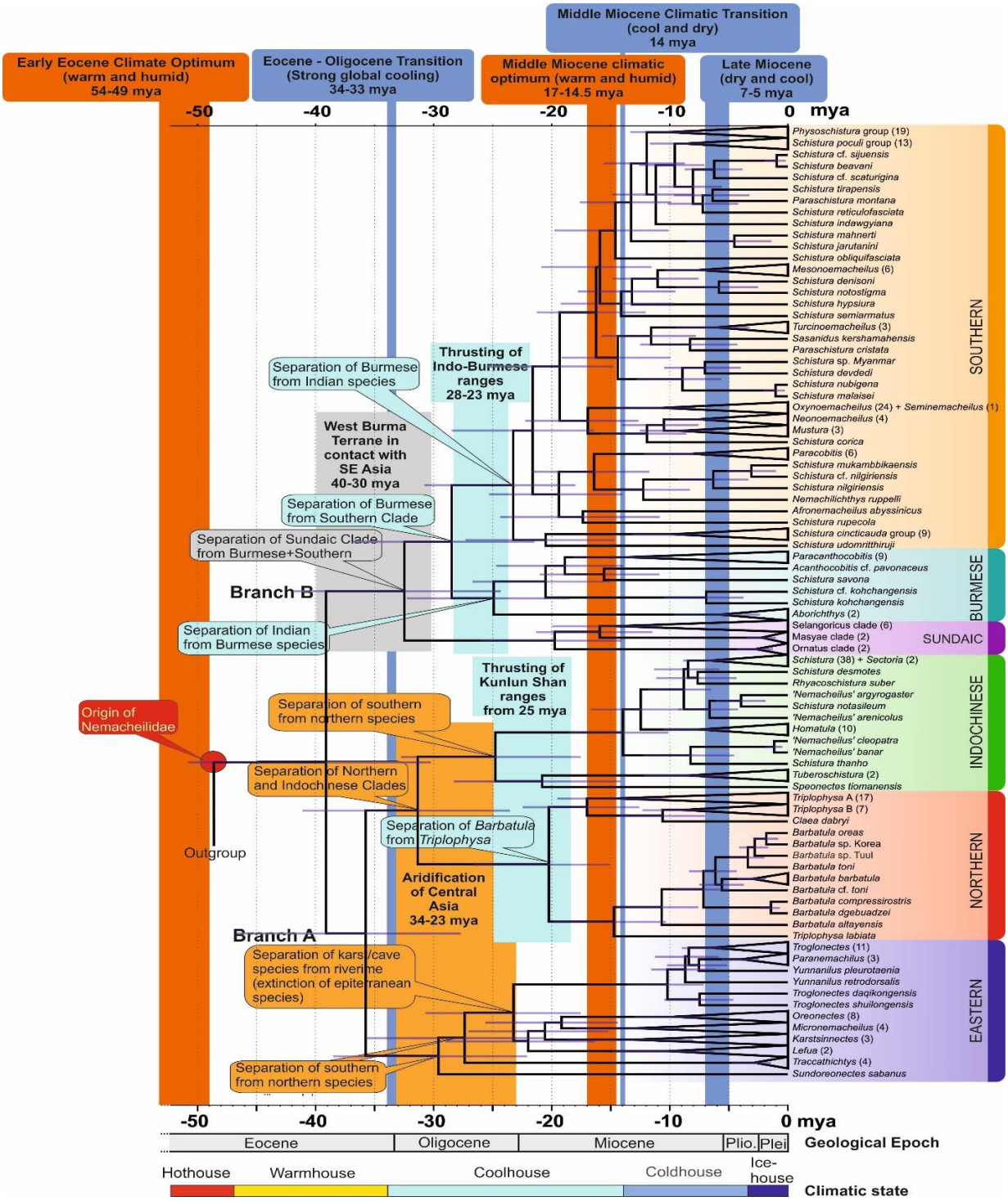

Fig. 7. Time tree of the evolutionary history of Nemacheilidae: The figure presents the evolutionary history of Nemacheildae, highlighting the dating of key geological or climatic key paleo-events that might have significantly impacted their evolution. Orange bars represent exceptionally warm periods, blue bars denote cold periods. Boxes indicate specific geological or climatic event. For detailed explanations of these events and their impact on nemacheilid evolution refer to text. Climatic states from Westerhold et al., 2020.

Eastern Clade spread mainly along the eastern coastline of Asia southwards to Borneo and northwards to northeast China. Geological data support this hypothesis: During the Eocene, Asia was much smaller than today and had a rather flat relief with old mountains mainly along its eastern coast due to the collision with the Pacific Plate since 190 mya (Seton et al., 2012; Wang, 2004), while the upfolding of east-west mountain ridges had just started. The generally warm and humid climate during Eocene (Morley, 2018; Westerhold et al., 2020) additionally promoted the range expansion of the Eastern Clade. It can be assumed that shortly after the range expansion the distribution of the Eastern Clade was more or less continuous and not separated into several disjunct areas as nowadays.

The ancestral Indochinese + Northern Clades also spread northwards, but their present-day distribution (Fig. 3) suggests that they took a route through Central Indochina and China. Most likely, this first geographic expansion led the ancestral Indochinese+Northern Clades northward to at least the Tibetan Protoplateau, which during Eocene epoch had its northern boundary in the region of today's Tanggula mountains (Wang et al., 2008), that means not even half of its present north-south extension. The presence of the two protoclades on the early Tibetan Plateau is indicated by the presence of some of their most basal members in the Yangtze basin and Yunnan: the genus *Homatula* for the Indochinese Clade and several species of *Triplophysa* for the Northern Clade (Figs. 1, 5).

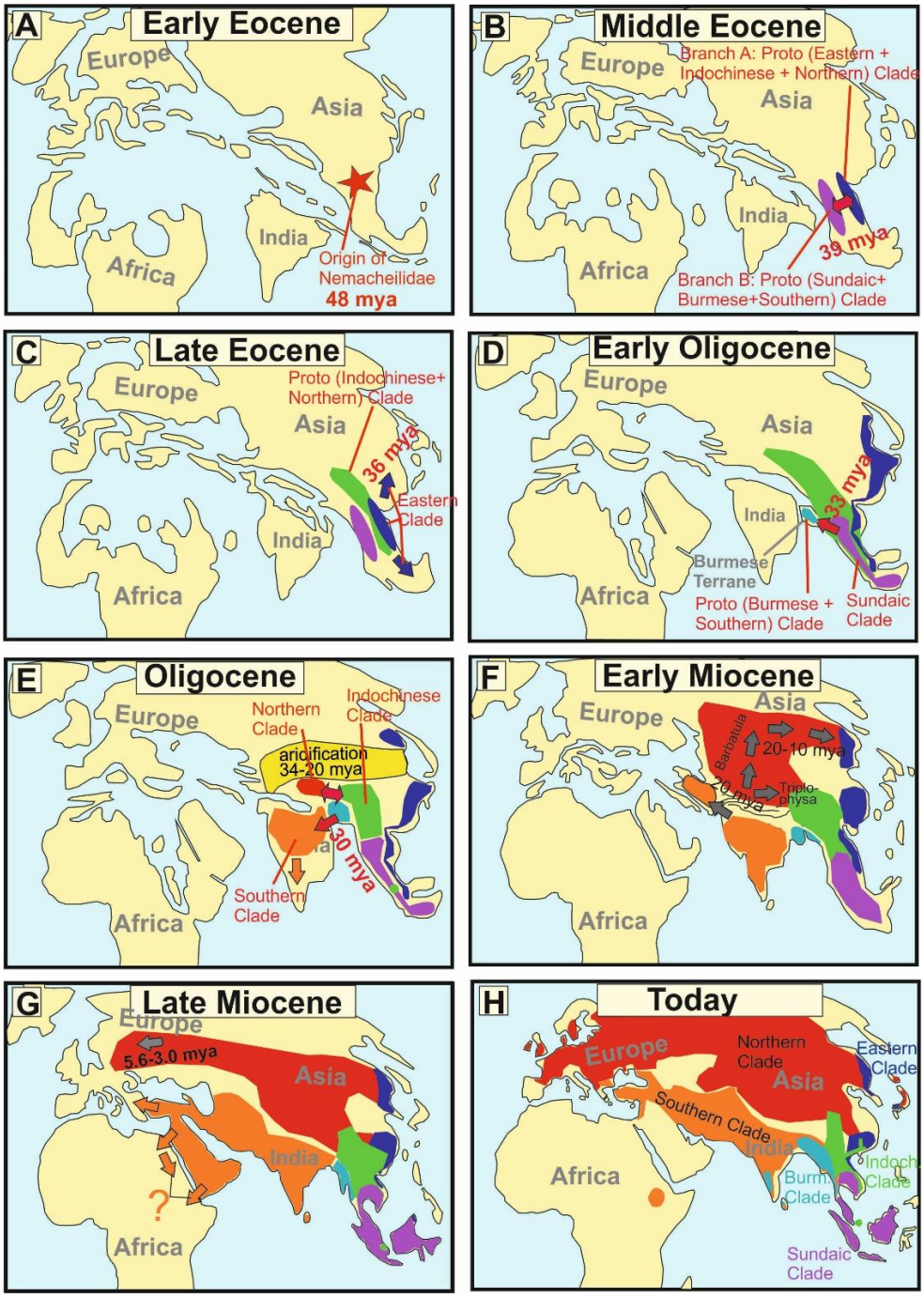

Fig. 8. Paleomaps visualising the geographical and chronological aspects of the main events during the evolutionary history of the freshwater fish family Nemacheilidae across Eurasia. A red star marks the point of origin, arrows indicate dispersal events. Background maps were redrawn according to Blakey (2023), Dierke Atlas (2023), Wang (2004), Westerweel et al. (2020).

Period	Eastern	Indochinese	Northern	Sundaic	Burmese	Southern
Early Eocene 48 mya	Origin of Nemacheilidae in the area of present-day Indochina; probably related to a strong diversification period during the Early Eocene Climatic Optimum (54-49 mya).					
Middle Eocene 39 mya	Split of Nemacheilidae into two branches (Proto Eastern + Indochinese + Northern branch and Proto Sundaic + Burmese + Southern branch). The ultimate cause for this split is unknown.					
Late Eocene 36 mya	Formation of Eastern Clade; spreading along the eastern margin of Asia.	The Proto Indochinese + Northern Clade starts expanding northwards and southwards (at least to present-day southern Malay Peninsula).				
Early Oligocene 33 mya				Formation of the Sundaic Clade by separation of the Proto Burmese + Southern Clade.	The Proto Burmese + Southern Clade forms by colonising the Burmese Terrane, a tectonic microplate that travels northwards with the Indian plate.	
Oligocene 34-23 mya	The distribution of the Eastern Clade becomes fragmented by strong aridification of Central Asia.	Proto Indochinese + Northern Clade in two refuges during strong aridification of Central Asia. The Proto- Northern Clade endures the period on the northern part of the young Tibetan Plateau, the Proto-Indochinese Clade east of the Tibetan Plateau. After this long time of separation they form the Northern and Indochinese Clades.			The Proto Burmese + Southern Clade colonises the Indian Plate (30 mya). Burmese Terrane and India collide with Asian mainland, causing the rise of the Indo-Burmese mountain ridge (from 28 mya), separating the Southern Clade from the Burmese Clade.	
Early and Middle Miocene 23-10 mya			The genus *Barbatula* spreads northwards as the Kunlun, Pamir and Altai mountains form (from 25 mya), then			The Southern Clade colonises the freshly forming Near East in several waves. The Proto-*Afronemacheilus*

			eastwards to Korean Peninsula; the genus *Triplophysa* spreads eastwards into southern China.			separates from Indian species.
Miocene 23-5.3 mya	Period of strong diversification, especially around the time of the Middle Miocene Climatic optimum (17-14.5 mya), Number of lineages in Nemacheilidae rises from 16 to 177					
Pliocene 5-2.5 mya			*Barbatula* spreads westwards due to ice formation in northern Siberia and colonises Europe. In the east it colonises the Japanese Archipelago.			
Pleistocene from 2.6 mya						Extinctions from Arab Peninsula and the lower Nile valley due to strong aridification.

Table 1. Overview about the major evolutionary events in the six main clades within Nemacheilidae from its origin 48 mya until today.

Oligocene (34 – 23 mya) – range expansions, extinctions and the formation of the remaining major clades (Fig. 8D,E)

The Eocene - Oligocene transition brought a period of significant global cooling (Hutchinson et al., 2021) as well as a change in the precipitation regime of Central Asia (Clift and Webb 2018; Ye et al., 2022) due to the ongoing uplift of the Himalayan chain (Ding et al., 2017). A substantial decrease of rainfall in Central Asia caused an aridification from 34 to 23 mya (Bosboom et al., 2014) extending across nearly all of Oligocene Asia from the western to the eastern coast (Wang, 2004). The distribution area of the Eastern Clade became fragmented during this period (Fig. 3).

The Central Asian arid zone also split the Indochinese – Northern protoclade into two parts that evolved into two separate major clades, the Indochinese Clade and the Northern Clade. Moreover, the distribution area of the Indochinese Clade was split and separated the southernmost taxa (*Speonectes* and *Tuberoschistura*) from the taxa presently inhabiting all of Indochina and southern China, including the middle Yangtze basin (Fig. 3). The absence of the Indochinese Clade from any waters north of the Tibetan Plateau suggests that its Oligocene refuge was located southeast of the Tibetan Protoplateau. In contrast, the early Northern Clade found its refuge further north on the Tibetan Protoplateau, the only region that can explain the further history and the present distribution of the clade. By late Oligocene, the climate in Central Asia returned to a warmer and moister state (Zachos et al., 2001) and the Pamir and western Kunlun Mountains (in present day SE Tajikistan, N Pakistan and W China) rapidly thrust upward as an expansion of the northern Tibetan Plateau (Zheng et al., 2015). These favourable conditions coincided with a

major split within the Northern Clade at 20 mya (Fig. 5), when the proto-*Barbatula* clade colonised the Pamir and Kunlun Mountains, while the proto *Triplophysa* branch found its main distribution rather east of the Tibetan Plateau.

Different geological events drove the evolution of branch B: the common ancestor of the Burmese and Southern Clades split at 32.5 mya in Indochina from the Sundaic Clade because the former left Southeast Asia by colonising the West Burma Terrane, a microplate that was pushed northward by the Indian plate (Morley et al., 2021) and made temporary contact with western Sundaland (today's western Malay Peninsula) around Eocene – early Oligocene (Westerweel et al., 2019). Around 30 mya, it made contact with northeast India (Morley et al., 2021), allowing fishes from the West Burma Terrane to colonise the Indian subcontinent. During the Oligocene, the Indian plate and the West Burma Terrane both collided with mainland Asia. The Indian plate presently forms the Indian subcontinent, the West Burma Terrane now forms western Myanmar (Westerweel et al., 2020). Their parallel collision with mainland Asia caused the uplift of the north-south directed Indo-Burmese mountain range (Morley et al., 2020), separating the aquatic fauna of the Brahmaputra basin on the Indian side, from the Chindwin-Irrawaddy basin on the Burmese side, thus separating the Burmese Clade from the Southern Clade.

Miocene (23 – 5.3 mya) – rapid diversification (Fig. 8F,G)

The Miocene is generally an epoch of cooling, with temperatures decreasing from about 5.5º C above present conditions to conditions similar to present (Zachos et al., 2001). But temperatures also underwent strong fluctuations, leading to exceptionally

warm periods like the Middle Miocene Climatic Optimum (17 – 14.5 mya) as well as to drastic cooling events like the Middle Miocene Climatic transition (14 mya) (Morley, 2018). Additionally, the Asian monsoon regime strengthened from the early Miocene (Clift and Webb, 2018), increasing precipitation and altering the distribution of climatic regions across Asia (Sun and Wang, 2005). Moreover, during the Miocene the Himalayan Mountains experienced their main uplift (Ding et al., 2017) and the upfolding area of Central Asia mountain ranges expanded northwards, giving rise to the Tian Shan and Altai ranges (Nissen et al., 2009; Zheng et al., 2015; Liu et al., 2023).

During this complex geological and climatic scenario, the family Nemacheilidae underwent a massive diversification. In our phylogenetic hypothesis, the family comprised 16 lineages at the beginning of Miocene. These lineages diversified during Miocene into 177 lineages. The radiation occurred across all six major clades, albeit to a lesser extent in the Eastern Clade. Interestingly, it manifested similarly in the region of the Indian monsoon and the East Asian monsoon. In both regions, major radiations occurred during the Middle Miocene Climatic Optimum, including the radiation of the *Barbatula* lineage in the Northern Clade, of the Indochinese Clade, the Sundaic Clade and of six branches in the Southern Clade. It appears that under increased temperature and precipitation freshwater fishes flourished, across all of Asia and under different monsoon systems.

One of the radiations that began during the Middle Miocene Climate Optimum involves *Afronemacheilus*, the sole nemacheilid genus found in Africa (Prokofiev and Golubtsov, 2013). Our results confirm that the ancestral *Afronemacheilus* separated

from the Southern Clade about 18 - 17 mya. It could have utilised two potential pathways to Africa: one across the Arabian Peninsula and via the southern Danakil Isthmus, a broad landbridge that stretched from Djibouti to Aden from late Miocene to Pliocene (Bosworth et al., 2005; Steward and Murray, 2017). The second pathway would lead through the Levant and via Sinai into the Nile ('*Gomphotherium* landbridge' Harzhauser et al., 2007), and subsequently upstream into the Ethiopian highlands. Unfortunately, the nearly complete extinction of freshwater fishes from the Arabian Peninsula and from the Nile during Pleistocene aridification (Kotwicki and Salaimani, 2009; Leplongeon, 2021) complicates the reconstruction of its pathway to Africa.

In the Northern Clade, the *Triplophysa* lineage flourished during the Miocene. Its present wide distribution in southern and Central China suggests that it expanded from its Oligocene refuge on the eastern Part of the Tibetan Plateau in an eastern direction (Fig. 3). In contrast, the *Barbatula* branch expanded from its refuge on the western Tibetan Plateau northwards, traversing the uplifting Tian Shan and Altai ranges. This expansion can be traced from the geographic position of the most basal species (*T. labiata* -Tian Shan, *B. altayensis* – Altai), and then eastwards (*B. compressirostris* + *B*. *dgebuadzei* – west Mongolia, *B.* cf. *toni* – east Mongolia; tip species in Korea (*B.* indet Korea) and Japan (*B. oreas*).

During late Miocene, *Barbatula* also expanded westwards through Siberia and colonised Europe between 5.6 mya (separation of the European taxa from the Mongolian sister taxon) and 3.0 mya (oldest split within European *Barbatula*). During this cool and dry period, glaciations occurred on the northern hemisphere (Zachos

et al., 2001), and in Siberia the mean temperature of the coldest month dropped below 0º for the first time during Cenozoic (Popova et al., 2012). Freezing of the northern margin of Eurasia prevented the Siberian rivers (which flow from south to north) from discharging their water into the Northern Ocean and produced extensive wetlands on the shallow relief of Siberia. These wetlands served as pathways for freshwater fishes from East Asia to European river systems during Pleistocene (Kalous et al., 2012), and we assume that the same mechanism brought *Barbatula* to Europe during Pliocene.

Pliocene and Pleistocene (5.3 – 0.01 mya) - Ice ages

During Pliocene and Pleistocene global cooling continued until massive glaciations repeatedly covered the north of Eurasia (de Schepper et al., 2014). However, the Pliocene also brought a much stronger seasonality to Eurasia, especially in the form of cold winters (Shen et al., 2018). Despite these significant changes in climatic conditions, the evolution of Nemacheilidae seems not to have experienced any dramatic negative impact, but reveals mainly local speciation events (e.g. the genus *Barbatula* colonised Eastern Siberia, Japan and the Korean Peninsula and the European *Barbatula* diversified geographically). We know from more detailed studies on Nemacheilidae (Bohlen et al., 2020a,b; Šlechtová et al., 2021) that during Pliocene and Pleistocene also geographic shifts as well as expansions and disappearances occurred, but since these happened on rather local scale they were too insignificant to reflect in the present intercontinental dataset.

Taxonomic implications

Besides its conclusions on biogeography and evolutionary history of the Eurasian freshwater fauna, the present dataset offers insights into the taxonomy of Nemacheilidae. Among the 37 analysed genera, 11 are represented by a single species in the dataset; 18 of the genera with several species in the dataset appear as monophyletic units in our analysis. However, seven genera (*Mustura, Paraschistura, Physoschistura, Oxynoemacheilus, Schistura, Sectoria, Triplophysa*) are identified as para- or polyphyletic based on our results. The most scattered genus is *Schistura*, which is listed in three of the six major clades and in a total of about 20 different genetic lineages.

In fact, this is nothing new. The ‘polyphyly’ of several larger genera, particularly *Schistura*, has been recognised for a long time: already Kottelat (1990) referred to *Schistura* and *Nemacheilus* as ‘wastebasket names’. These are not really cases of polyphyly; it is simply that a large number of new species have been described in the last 40 years and, not knowing in which genus to place them, authors simply 'dumped' them in a few catch-all genera, waiting for more comprehensive studies and knowing well that the used genus names are mere labels. In recent years, more detailed morphological studies have resolved the relations of some groups of species within these catch-all genera (e.g. Conway and Kottelat, 2023; Dvořák et al., 2023; Kottelat, 2018, 2019).Moreover, every genetic study that included more than five species of *Schistura* resulted in polyphyly of the genus (e.g. Chen et al., 2019; Min et al., 2023; Sember et al., 2015; Sgouros et al.,

2019; Tang et al., 2006). A recent detailed phylogenetic investigation of the genus *Nemacheilus* (Šlechtová et al., 2021) found the genus to be polyphyletic, but it became monophyletic after identifying five species that were falsely placed into *Nemacheilus* (these species are labelled '*Nemacheilus*' in the present study). Similarly, revisionary studies of the genera *Paracanthocobitis, Troglonectes, Micronemacheilus, Paranemachilus* and *Oreonectes*, also allowed to resolve their intrarelationships and to redefine them into monophyletic taxa (Singer and Page, 2015; Luo et al., 2023). The main obstacles to a comprehensive taxonomic revision of Nemacheilidae is its vast size, a fast increasing number of new species, and the difficulty to access material from critical areas because of international and local political issues. We believe that results of the present study will facilitate future research on the taxonomic understanding of Nemacheilidae.

Another aspect of Nemacheilidae taxonomy is the existence of numerous undescribed species. Among the 279 species in our dataset, 36 (13%) were either clearly undescribed ('*Genus* sp.') or named with reservations ('*Genus*' cf. *xxx*'). This ratio is an underestimate, since, in general, undescribed species were avoided when composing the dataset. However, the existence of a vast amount of undescribed species within Nemacheilidae is illustrated by the fact that 430 of the 838 valid species of Nemacheilidae known in November 2024 (51%) have been described in the last 25 years (pers. obs.; Kottelat, 2012, 2013, updated).

Conclusions

- We reconstruct the evolutionary history of the Eurasian freshwater fish family Nemacheilidae during most of Cenozoic era. The extensive dataset (471 specimens, 279 species, six genes) from across Eurasia and the low dispersal capacity of the model resulted in high resolution of the phylogenetical and biogeographical analysis.
- Molecular phylogeny uncovered six major clades within the family. Dating of cladogenetic events and ancestral range estimation traced the origin of Nemacheilidae to Indochina around 48 million years ago.
- Major expansions were facilitated by tectonic connections, favourable climatic conditions, and orogenic processes. Conversely, aridification emerged as the primary cause of extinction events.
- Our study represents the first comprehensive and holistic reconstruction of the evolution of Eurasian freshwater biodiversity on a continental scale and through deep geological time.
- Our analyses identified numerous polyphyletic or paraphyletic taxa, highlighting the need for a thorough taxonomic revision within the family Nemacheilidae.

Methods

Sampling and Laboratory procedures

471 samples of more than 250 species from 37 genera of the family Nemacheilidae were examined, including the set of sequences of 364 specimens analysed in our laboratory and sequences of 107 specimens obtained from GenBank. Specimens used for the present study were fixed and stored in 96% ethanol. For more details

about samples, sampling sites and GenBank accession numbers see Supplementary material Table S1.

Total genomic DNA was isolated using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. One mitochondrial gene (cytochrome *b*) and five nuclear genes (RAG1, IRBP2, MYH6, RH1 and EGR3) were selected. For list of primers see Supplementary material Table S2, details about PCRs settings are described in Chen et al., 2003; Chen et al., 2008 and Dvořák et al., 2022.

Sequencing of the PCR products was performed in Laboratory of Fish Genetics on ABI Prism 3130 GA or via sequencing service in Macrogen Europe BV (Amsterdam, Netherlands).

Phylogenetic analyses

Chromatograms were checked and assembled in the SeqMan II module of the DNA Star software package (LASERGENE). Single gene alignments were done in BioEdit (Hall, 1999) with use of ClustalW (Larkin et al., 2007) multiple alignment algorithm. The datasets were concatenated in PhyloSuite v1.2.2 (Zhang et al., 2020).

The best-fit substitution models and partitioning schemes (Supplementary material Table S3) were estimated using Partition Finder 2 (Lanfear et al., 2016) implemented in PhyloSuite 1.2.2 based on the corrected Akaike Information Criterion (AICc).

The 471 specimen phylogenetic trees were inferred from six loci concatenated dataset using the maximum likelihood (ML) and Bayesian inference (BI). The ML analyses were performed using IQ-TREE (Nguyen et al., 2015) implemented in PhyloSuite. The best-fit evolutionary model for each codon partition was automatically determined by ModelFinder (Kalyaanamoorthy, et al., 2017). The node support values were obtained with 5000 ultrafast bootstrap replicates (UFBoot) (Hoang et al., 2018). For the BI analyses we used MrBayes 3.2.7a (Ronquist and Huelsenbeck, 2003) on CIPRES Science Gateway (Miller et al., 2010). The datasets were partitioned into genes and codon positions and analyses were performed in two parallel runs of 10-20 million generations with 8 Metropolis Coupled Markov Chains Monte Carlo (MCMCMC). The relative burnin of 25% was used and from the remaining trees a 50% majority rule consensus trees were constructed.

For reconstructing of the species tree we used the Accurate Species TRee ALgorithm (ASTRAL III) (Zhang et al., 2018). For ASTRAL we used unrooted single gene ML trees reconstructed in IQ-TREE implemented in PhyloSuite. We have used IQ-TREE 2 (Minh at al., 2020) to calculate gene concordance factor (gCF, indicating the percentage of decisive gene trees containing particular branch) and site concordance factor (sCF, indicating the percentage of decisive alignment sites that support a branch in the reference tree) for every branch in the ASTRAL species tree. Distribution maps in Fig. 2 were designed using HYDROSHEDS (Lehner and Grill 2013). All original trees are available from authors upon request.

Divergence time estimations and ancestral range reconstruction

The ages of cladogenetic events were estimated in BEAST 2.6.4 (Bouckaert et al., 2014) via CIPRES Science Gateway with use of four calibration points. The first calibration point is based on the oldest known fossil of the family Catostomidae, *Wilsonium brevipinne*, from early Eocene (Liu, 2021), approximately 56-48 my old. The second calibration point is derived from the only known nemacheilid fossil record, *Triplophysa opinata* from Kyrgyzstan from middle-upper Miocene (16.0 to 5.3 mya) (Böhme and Ilg, 2003; Prokofiev, 2007). For some time, the Miocene fossil species *Nemachilus tener* from Central European was considered a member of Nemacheilidae. However, subsequent research has cast doubt on placement of this species within loaches (Obrhelova, 1967), leaving *Triplophysa opinata* as the only known fossil of Nemacheilidae. Third calibration point is based on the oldest known fossil of the genus *Cobitis*, *C. naningensis*, from early to middle Oligocene in southern China*,* 34-28 my old (Chen et al., 2015). The fourth calibration point, based on the river history of the southern Korean Peninsula, dates the separation between the species *Cobitis tetralineata* and *C. lutheri* to 2.5-3.5 my (Kwan et al., 2014). For the fossil-based calibration points we have used a fossil calibration prior implemented in CladeAge (Matschiner et al., 2017), the fourth calibration point was set to a uniform prior (2.5 – 3.5). The calibration points are depicted in Figs. 2 and 5.

For the final analyses, the partitions were unlinked and assigned the estimated evolutionary models. We used relaxed lognormal molecular clock and birth-death prior. Given the large and complex nature of our dataset, and in light of several preliminary analyses where MCMC chains did not provide satisfactory ESSs even after 1 billion iterations, we made the decision to streamline the dataset by retaining only one specimen per species. To enhance the performance of the analysis, we also implemented parallel tempering using the CoupledMCMC package (Müller and Bouckaert, 2019, 2020). The analysis was configured with four parallel chains of $6x10^8$ generations, with resampling every 1000. Tree and parameter sampling intervals were set to 60.000. The effective sampling sizes (ESS) for all parameters were assessed in Tracer 1.7.1 (Rambaut et al., 2018). A maximum clade credibility (MCC) tree was built in TreeAnnotator 2.6.0 (Rambaut and Drummond, 2010) after discarding the first 25% of trees. The final trees were visualised in FigTree 1.4.4 (Rambaut, 2019). ML and BI were performed with both full as well as the reduced datasets.

The biogeographical analysis was conducted using the BioGeoBEARS package (Matzke, 2013) implemented in RASP 4.0 (Reconstruct Ancestral Stage in Phylogenies, Yu et al., 2015). We set 11 biogeographic regions according to drainage areas separated by major mountain ridges (Ural, Caucasus. Himalayas, Tenasserim Ridge, Indo-Burmese Range, Annamite Cordilliera, Balkan, Hindukush) or saltwater straits (Red Sea) (see Fig. 4A). The selected biogeographic regions are mostly congruent with the classical subregions of the Palaearctic (European, Siberian, East Asia and Mediterranean) and the Oriental (Indian, Malay and Chinese). For the analysis, we utilized the final MCC tree and 100 post burnin trees from the previous BEAST analysis. A maximum of 2-unit areas was allowed to be present in every reconstructed ancestral range. The comparison of six

biogeographic models (DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, BAYAREALIKE+J) suggested that the DEC+J is the best-suited model for the given dataset (as indicated by AICc weight value, Supplementary material Table 4). Likelihood Ratio Test (LRT) p-values, comparing models with and without the J factor, indicated that the addition of the J factor significantly influences the model likelihood. Considering recent criticisms of DEC and DEC+J (Ree and Sanmartin, 2018), we opted to conduct the analysis not only with DEC+J but also with other +J models. Comparison of the results from all applied models revealed only minimal differences.

Data availability statement

Original sequences have been deposited in GenBank under the accession numbers PP259624 – PP259994, PP279842 – PP280528 and PP315675 – 315895. All original trees are available from authors upon request.

References

Andreyenkova, N. G. et al. Phylogeography and demographic history of the black kite *Milvus migrans*, a widespread raptor in Eurasia, Australia and Africa. *J Avian Biol.* **e02822** (2021).

Antonelli, A. et al. Geological and climatic influences on mountain biodiversity. *Nat. Geosci.* **11**, 718-725 (2018).

Ballarin, F. and S. Q. Li. Diversification in tropics and subtropics following the mid-Miocene climate change: a case study of the spider genus *Nesticella*. *Glob. Chang. Biol.* **24** (2), e577–e591 (2018).

Bănărescu, P. *Zoogeography of Fresh Waters, Vol. 2: Distribution and Dispersal of Freshwater Animals in North America and Eurasia.* AULA, Wiesbaden 1992.

Blakey, R. 2023. https://deeptimemaps.com (visited September 2023)

Bohlen, J., Dvořák, T., Šlechta, V. and V. Šlechtová. Resolving an unnoticed diversity within the Schistura robertsi species complex (Teleostei: Nemacheilidae) using molecules and morphology. *Mol. Phylogenet. Evol*. **151**, 106894 (2020b).

Bohlen, J., Dvořák, T., Šlechta, V. and V. Šlechtová. Sea water shaping the freshwater biota: Hidden diversity and biogeographic history in the *Paracanthocobitis zonalternans* species complex (Teleostei: Nemacheilidae) in western Southeast Asia. *Mol. Phylogenet. Evol*. **148**, 106806 (2020a).

Böhme, M. and A. Ilg. 2003. fosFARbase, www.wahre-staerke.com/ (accessed 24. October 2024).

Bosboom, R. et al. Aridification in continental Asia after the Middle Eocene Climatic Optimum (MECO). *Earth and Planet. Sci. Lett.* **389**, 34-42 *(*2014*).*

Bosworth et al. The Red Sea and Gulf of Aden basins. *J. Afr. Earth Sci*. **43,** 334-378 (2005).

Bouckaert, R. et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol*. **10**, e1003537 (2014).

Chen, G.-J. et al. First fossil cobitid (Teleostei: Cypriniformes) from Early-Middle Oligocene deposits of South China. *Vert. PalAs*. **53** (4), 299*–309 (*2015).

Chen, W. T., Yang, J. P., Li, Y. F. and X. Li. Exploring taxonomic diversity and biogeography of the family Nemacheilinae (Cypriniformes). *Ecol. Evol*. **9,** 10343−10353 (2019).

Chen, W.-J., Miya, M., Saitoh, K., and R. L. Mayden. Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: The order Cypriniformes (Ostariophysi) as a case study. *Gene* **423**, 125-134 (2008).

Chen, W.-J., Bonillo, C. and G. Lecointre. Repeatability of clades as criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Mol. Phylogenet. Evol.* **26**, 262-288 (2003).

Chu, S.-L. and Y.-R. Chen. A new blind cobitid fish (Pisces, Cypriniformes) from subterranean waters in Yunnan, China. *Acta Zool. Sin.* **25** (3), 285-287 (1979).

Clift, P. D. and A. A. G. Webb. A history of the Asian monsoon and its interactions with solid Earth tectonics in Cenozoic South Asia. P.J. Treloar, M.P. Searle (Eds.), Himalayan Tectonics: A Modern Synthesis, London, Special Publications, Geological Society (2018).

Conway, K. W. and M. Kottelat. 2023. *Physoschistura mango*, a new miniature species of loach from Myanmar (Teleostei: Nemacheilidae). *Raf. Bull. Zool.*, **71**, 681–701.

De Schepper, S., Gibbard, P. L., Salzmann, U. and J. Ehlers. A global synthesis of the marine and terrestrial evidence for glaciation during the Pliocene Epoch. *Earth Sci. Rev.* **135**, 83-102 (2014).

Delić, T., Stoch, F., Borko, Š., Flot, J. F. and C. Fišer. How did subterranean amphipods cross the Adriatic Sea? Phylogenetic evidence for dispersal–vicariance interplay mediated by marine regression–transgression cycles. *J. Biogeogr.* **47** (9), 1875−1887 (2020).

Deng, T., Wu, F. X., Zhou, Z. K. and T. Su. Tibetan Plateau: an evolutionary junction for the history of modern biodiversity. *Sci. China Earth Sci.* **63** (2), 172−187 (2020).

Dierke Atlas 2023. https://www.westermann.de/reihe/DIWEINT21/Diercke-International-Atlas. (visited September 2023)

Ding, L. et al. Quantifying the rise of the Himalaya orogen and implications for the South Asian monsoon. *Geology* **45** (3), 215–218 (2017).

Dvořák, T., Šlechtová, V. and J. Bohlen. Using species groups to approach the large and taxonomically unresolved freshwater fish family Nemacheilidae (Teleostei: Cypriniformes). *Biology* ***11***, 175 (2022).

Dvořák, T., J. Bohlen, M. Kottelat and V. Šlechtová. Revision of the *Schistura cincticauda* species group (Teleostei, Nemacheilidae) using molecular and morphological markers. *Scient. Rep.*, **13** (16996) (2023).

Feduccia, A. Explosive evolution in Tertiary birds and mammals. *Science.* **267** (5198), 637–638 (1995).

Friedman, M. Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proc. Royal Soc. B.* **277** (1688), 1675–1683 (2010).

Fritz, S. A. et al. Twenty-million-year relationship between mammalian diversity and primary productivity. *Proc. Natl Acad. Sci. USA* **113**, 10908–10913 (2016).

Gu, Q. et al. Phylogeographic relationships and the evolutionary history of the *Carassius auratus* complex with a newly born homodiploid raw fish (2nNCRC). *BMC Genom.* **23**, 242 (2022).

Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser*. **41**, 95-98 (1999).

Harzhauser, M. et al. Biogeographic responses all around the Oligo-Miocene Tethyan seaway. *Zool. Anz*. **246**, 241-256 (2007).

Hoang, D. T. Chernomor, O., von Haeseler, A., Minh B. Q. and L. S. Vinh. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

Hutchinson, D. K. et al. The Eocene–Oligocene transition: a review of marine and terrestrial proxy data, models and model–data comparisons. *Clim. Past* **17**, 269–315 (2021).

Ishii, H. et al. Phylogeography of the stone marten (Martes foina: Mustelidae: Mammalia) in Eurasia, based on a mitochondrial DNA analysis. *Mammal Res.* **68**, 375–381 (2023).

Kahlke, R. D. The origin of Eurasian Mammoth Faunas (*Mammuthus–Coelodonta* Faunal Complex). *Quat. Sci. Rev.* **96**, 32-49 (2014).

Kalous, L., Rylková, K., Petrtýl, M. and J. Bohlen. Hidden diversity within the Prussian carp and designation of a neotype for *Carassius gibelio*. *Ichthyol. Explor. Freshw.* **23**, 11-18 (2012).

Kalyaanamoorthy. S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. and L. S. Jermiin. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* **14** (6), 587–589 (2017).

Kottelat, M. Indochinese nemacheilines. A revision of nemacheiline loaches (Pisces: Cypriniformes) of Thailand, Burma, Laos, Cambodia and southern Viet Nam. Verlag Dr. Friedrich Pfeil, München (1990).

Kottelat, M. Conspectus Cobitidum: An inventory of the loaches of the world (Teleostei: Cypriniformes: Cobitoidei). *Raf. Bull. Zool*. Suppl. **26**, 1– 99 (2012).

Kottelat, M. The fishes of inland waters of Southeast Asia: a catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. *Raf. Bull. Zool.,* Suppl., **27**, 1-663 (2013).

Kottelat, M. *Mustura celata*, a new genus and species of loaches from northern Myanmar, and an overview of *Physoschistura* and related taxa (Teleostei: Nemacheilidae). *Ichthyol. Explor. Freshw.,* **28** (4), 289–314 (2018).

Kottelat, M. *Rhyacoschistura larreci*, a new genus and species of loach from Laos and redescription of *R. suber* (Teleostei: Nemacheilidae). *Zootaxa*, **4612** (2): 151–170 (2019).

Kotwicki, V. and Z. Salaimani. Climates of the Arabian Peninsula – past, present, future. *Int. J. Clim. Chang. Strateg. Manag.* **1**, 297-310 (2009).

Kwan, Y. S., Ko, M. H. and Y. J. Won. Genomic replacement of native *Cobitis lutheri* with introduced *C. tetralineata* through a hybrid swarm following the artificial connection of river systems. *Ecol. Evol.* **4** (8), 1451-65 (2014).

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. and B. Calcott. Partition Finder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol*. **34** (3), 772-773 (2016).

Larkin, M. A. et al. Clustal W and Clustal X. Version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

Lehner, B. and G. Grill. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, **27** (15), 2171–2186 (2013).

Leplongeon, A. The main Nile valley at the end of the Pleistocene (28-15 ka): Dispersal corridor or environmental refugium? *Front. Earth Sci.* 8.607183 (2021).

Liu, J. Redescription of "*Amyzon" brevipinne* and remarks on North American Eocene catostomids (Cypriniformes: Catostomidae)". *J. Syst. Palaeontol.* ***19*** *(9), 1–13 (2021).*

Liu, X. et al. Tectonic uplift of the Tianshan Mountains since Quaternary: evidence from magnetostratigraphy of the Yili Basin, northwestern China. *Int. J. Earth Sci.* **112**, 855–865 (2023).

Luo, T. et al. Phylogenetic relationships of Nemacheilidae cavefish (*Heminoemacheilus*, *Oreonectes*, *Yunnanilus*, *Paranemachilus*, and *Troglonectes*) revealed by analysis of mitochondrial genome and seven nuclear genes. *Zoological Research* **44**, 693–697 (2023).

Matschiner, M. et al. Bayesian node dating based on probabilities of fossil sampling supports trans-Atlantic dispersal of cichlid fishes. *Syst. Biol.* **66**, 3-22 (2017).

Matzke, N. J. Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing. *Front. Biogeogr.* **5**, 242–248 (2013).

Mayhew, P. J., Bell, M. A., Benton, T. G. and McGowan, A. J. Biodiversity tracks temperature over time. *Proc. Natl Acad. Sci. USA* **109**, 15141–15145 (2012).

Miller, M. A., Pfeiffer, W. and T. Schwartz. Creating the CIPRES Science gateway for inference of large phylogenetic trees. Proc. *Gateway Computing Environments Workshop 2010* pages **1–8**, New Orleans, LA, November, 2010.

Min, R., Zhao, Y. H., Kang, B., Chen, X. Y. and J. X. Yang. 2023. Vicariance and monsoon as drivers of diversification of nemacheilid loaches (Teleostei: Cypriniformes) around the Hengduan Mountains of China. *Zool. Res*. **44**, 936-938 (2023).

Minh, B. Q., Hahn, M. and W. R. Lanfear. New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* **37**, 2727–2733 (2020).

Morley, C. K., Naing, T. T., Searle, M. and S. A. Robinson. Structural and tectonic development of the Indo-Burma ranges. *Earth Sci. Rev.* **200**, 102992 (2020).

Morley, R. J. Assembly and division of the South and South-East Asian flora in relation to tectonics and climate change. *J. Trop. Ecol.* **34** (4), 209-234 (2018).

Morley, R. J. et al. Sequence biostratigraphic framework for the Oligocene to Pliocene of Malaysia: High-frequency depositional cycles driven by polar glaciation. *Palaeogreogr. Palaeoclimatol. Palaeoecol.* **561**, 110058 (2021).

Mudelsee, M., Bickert, T., Lear, C. H. and G. Lohmann. Cenozoic climate changes: A review based on time series analysis of marine benthic $\delta^{18}O$ records. *Rev. Geophysics* **52**, 333-374 (2014).

Müller, N. F. and R. Bouckaert. Adaptive Metropolis-coupled MCMC for BEAST 2. *PeerJ* **16**:8:e9473 (2020).

Müller, N. F. and R. Bouckaert. Coupled MCMC in BEAST 2. *bioRxiv* (2019).

Nguyen L.-T., Schmidt, H. A., von Haeseler, A. and B. Q. Minh. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).

Nissen, E., Walker, R., Molor, E., Fattahi, M. and A. Bayasgalan. Late Quaternary rates of uplift and shortening at Baatar Hyarhan (Mongolian Altai) with optically stimulated luminescence. *Geophys. J. Inter.* **177** (1), 259–278 (2009).

Obrhelova, N. Cyprinoidei (Pisces) aus dem hangenden des miozänen Braunkohlenflözes Nordböhmens. *Palaeontographica, Ser. A* **126**, 141–179 (1967).

Popova, S., Utescher, T., Gromyko, D. Bruch, A. and V. Mosbrugger. Palaeoclimate Evolution in Siberia and the Russian Far East from the Oligocene to Pliocene – Evidence from Fruit and Seed Floras. *Turk. J. Earth Sci.* **21** (2), 315–344 (2012).

Prokofiev, A. M. and A. S. Golubtsov. Revision of the loach genus *Afronemacheilus* (Teleostei: Balitoridae: Nemacheilinae) with description of a new species from the Omo-Turkana basin, Ethiopia. *Ichthyol. Explor. Freshw.* **24** (1), 1-14 (2013).

Prokofiev, A. M. Redescription of a fossil loach *Triplophysa opinata* (Yakowlew, 1959) from the Miocene of Kirgizia (Balitoridae: Nemacheilinae). *J. Ichthyol*. **47**, 26–31 (2007).

Rambaut A. *Figtree version 1.4.4* Available from http://tree.bio.ed.ac.uk/software/figtree (2019).

Rambaut, A. and A. J. Drummond. TreeAnnotator version 1.6.1 Available from http://beast.bio.ed.ac.uk/TreeAnnotator (2010).

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. and M. A. Suchard. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol*. **67** (5), 901–904 (2018).

Ree, R.H. and I. Sanmartin. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.* **45** (4), 741–749 (2018).

Ronquist, F. and J. P. Huelsenbeck. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).

Saito, T. et al. Phylogeography of freshwater planorbid snails reveals diversification patterns in Eurasian continental islands. *BMC Evol. Biol.* **18**, 164 (2018).

Sember, A., Bohlen, J., Šlechtová, V., Altmanová, M., Symonová, R. and P. Ráb. Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evol. Biol*. **15,** 251 (2015).

Seton, M. et al. Global continental and ocean basin reconstructions since 200 Ma. *Earth Sci. Rev.* **113** (3-4), 212-270 (2012).

Sgouros, K., Page, L. M., Orlofske, S. A. and R. C. Jadin. 2019. A revised molecular phylogeny reveals polyphyly in *Schistura* (Teleostei: Cypriniformes: Nemacheilidae). *Zootaxa* **4559**, 349-362 (2019).

Shen, X. et al. Increased seasonality and aridity drove the C4 plant expansion in Central Asia since the Miocene–Pliocene boundary. *Earth Planet. Sci. Lett.* **502**, 74-83 (2018).

Singer, R. A. and L. M. Page. Revision of the zipper loaches, *Acanthocobitis* and *Paracanthocobitis* (Teleostei: Nemacheilidae), with descriptions of five new species. *Copeia* **103**, 378-401 (2015).

Šlechtová, V., Musilová, Z., Tan, H. H., Kottelat, M. and J. Bohlen. One northward, one southward: Contrasting biogeographical history in two benthic freshwater fish genera across Southeast Asia (Teleostei: Cobitoidea: *Nemacheilus*, *Pangio*). *Mol. Phylogenet. Evol*. **161**, 107139 (2021).

Steward, K. M. and A. M. Murray. Biogeographic implications of fossil fishes from the Awash River, Ethiopia. *J. Vertebr. Paleont*. **37,** e1269115 (2017).

Summerhayes, C. P. *Earth's climate evolution*. Wiley Blackwell, 2015. ISBN: 978-1-118-89739-3

Sun, X. and P. Wang. How old is the Asian monsoon system? Palaeobotanical records from China. *Palaeogreogr. Palaeoclimatol. Palaeoecol.* **222**, 181-222 (2005).

Tang, Q., Liu, H., Mayden, R. and B. Xiong. Comparison of evolutionary rates in the mitochondrial DNA cytochrome b gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). *Mol. Phylogenet. Evol.* **39,** 347-357 (2006).

Trifonov, V. G., Ivanova, T. P. and D. M. Bachmanov. Recent Mountain Building of the Central Alpine–Himalayan Belt. *Geotectonics* **46** (5), 315–332 (2012).

Wang, C. et al. Constraints on the early uplift history of the Tibetan Plateau. *Proc. Natl Acad. Sci. USA* **105** (13), 4987-4992 (2008).

Wang, P. Cenocoic deformation and the history of sea-land interactions in Asia. In: Clift, P., Wang, P., Kuhnt, W., Hayes, D. (Eds.), Continent-Ocean Interactions in the East Asian Marginal Seas. *Geophys. Monogr. Ser.* **149**, 1-22 (2004).

Westerhold, T. et al. An astronomically dated record of Earth's climate and its predictability over the last 66 million years. *Science* **369** (6509), 1383–1387 (2020).

Westerweel, J. et al. Burma Terrane collision and northward indentation in the Eastern Himalayas recorded in the Eocene-Miocene Chindwin Basin (Myanmar). *Tectonics* **39**, e2020TC006413 (2020).

Westerweel, J. et al. Burma Terrane part of the Trans-Tethyan arc during collision with India according to palaeomagnetic data. *Nat. Geosci.* **12**, 863–868 (2019).

Ye, J. W., Tian, B. and D. Z. Li. Monsoon intensification in East Asia triggered the evolution of its flora. *Front. Plant. Sci.* **13,** 1046538 (2022).

Yu, Y., Harris, A. J., Blair, C. and X. He. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historic biogeography. *Mol. Phylogenet. Evol*. **87**, 46–49 (2015).

Zachos, J. C., Pagani, M., Sloan, L. C., Thomas, E. and K. Billups. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 689–693 (2001).

Zhang, C., Rabiee, M., Sayyari, E. and S. Mirarab. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform*. **19**, 153 (2018).

Zhang, D. et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. **20** (1), 348–355 (2020).

Zheng, H. et al. Late Oligocene–early Miocene birth of the Taklimakan Desert. *Proc. Natl. Acad. Sci.*, **112**, 7662-7667 (2015).

Acknowledgements

VŠ, TD, VŠ and JB would like to thank R. Hoyer, J. Kühne, F.F. Kullander (NRM), J. Kuszniersz, K. Lim (ZRC), H. Linke, A. Nolte, G. Ott, M. Reichard, I. Seidel, S. Somadee, H.H. Tan (ZRC), T.K. Toe, K. Udomritthiruji, T. Win and U. Wolf for their help with obtaining specimens for this study and Š. Pelikanová for help with laboratory work. VŠ and JB were funded by grants 206/08/0637 and 19-18453S of the Grant Agency of the Czech Republic. BL and AG greatly acknowledge S. Cherenkov for his help in sampling. BL was supported by Russian Science Foundation, grant no. 24-44-20019. Material from Laos, Myanmar and Mongolia were obtained by MK under various contracts and over the years he was assisted mainly by Nyein Chan, T. Phommavong and the late M. Erdenebat; the late T. Whitten has been essential in creating these great opportunities.