

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/340408164>

Sea water shaping the freshwater biota: Hidden diversity and biogeographic history in the *Paracanthocobitis zonalternans* species complex (Teleostei: Nemacheilidae) in western South...

Article in *Molecular Phylogenetics and Evolution* · April 2020

DOI: 10.1016/j.ympev.2020.106806

CITATIONS

12

READS

548

4 authors, including:



Joerg Bohlen

Institute of Animal Physiology and Genetics CAS, v. v. i.

178 PUBLICATIONS 2,918 CITATIONS

SEE PROFILE



Tomáš Dvořák

Charles University in Prague

13 PUBLICATIONS 49 CITATIONS

SEE PROFILE



Vendula Šlechtová

The Czech Academy of Sciences

97 PUBLICATIONS 1,599 CITATIONS

SEE PROFILE



Sea water shaping the freshwater biota: Hidden diversity and biogeographic history in the *Paracanthocobitis zonalternans* species complex (Teleostei: Nemacheilidae) in western Southeast Asia

Jörg Bohlen^{1,*}, Tomáš Dvořák^{1,2}, Vlastimil Šlechta¹ and Vendula Šlechtová¹

¹ Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Rumburská 89, 277 21 Liběchov, Czech Republic.

² Department of Zoology, Charles University, Viničná 7, 128 00 Prague, Czech Republic.

* Correspondence: joerg_bohlen@yahoo.de

<https://doi.org/10.1016/j.ympev.2020.106806>

Highlights

- A comprehensive phylogeny and biogeographic analysis of a benthic freshwater fish group, the *Paracanthocobitis zonalternans* species complex, from Central Myanmar to northern Malaysia using one mtDNA gene and two nDNA genes.
- Six major evolutionary units supported in all analyses; nearly all with exclusive distribution area.
- Origin of *P. zonalternans* during early Miocene in Tenasserim area and massive dispersal during periods of lowered global sea level (late Miocene).
- Dispersal into three biogeographic units and crossing all known biogeographic barriers in the region.
- Shaping of major lineages by fragmentation of distribution area during periods of higher global sea level (esp. Pliocene and Pleistocene).
- Global sea level was the most important factor in the evolutionary history of these freshwater fishes, although stream capture and ecological speciation also played a role.
- The global sea level should have similar effects in all coastal regions worldwide.

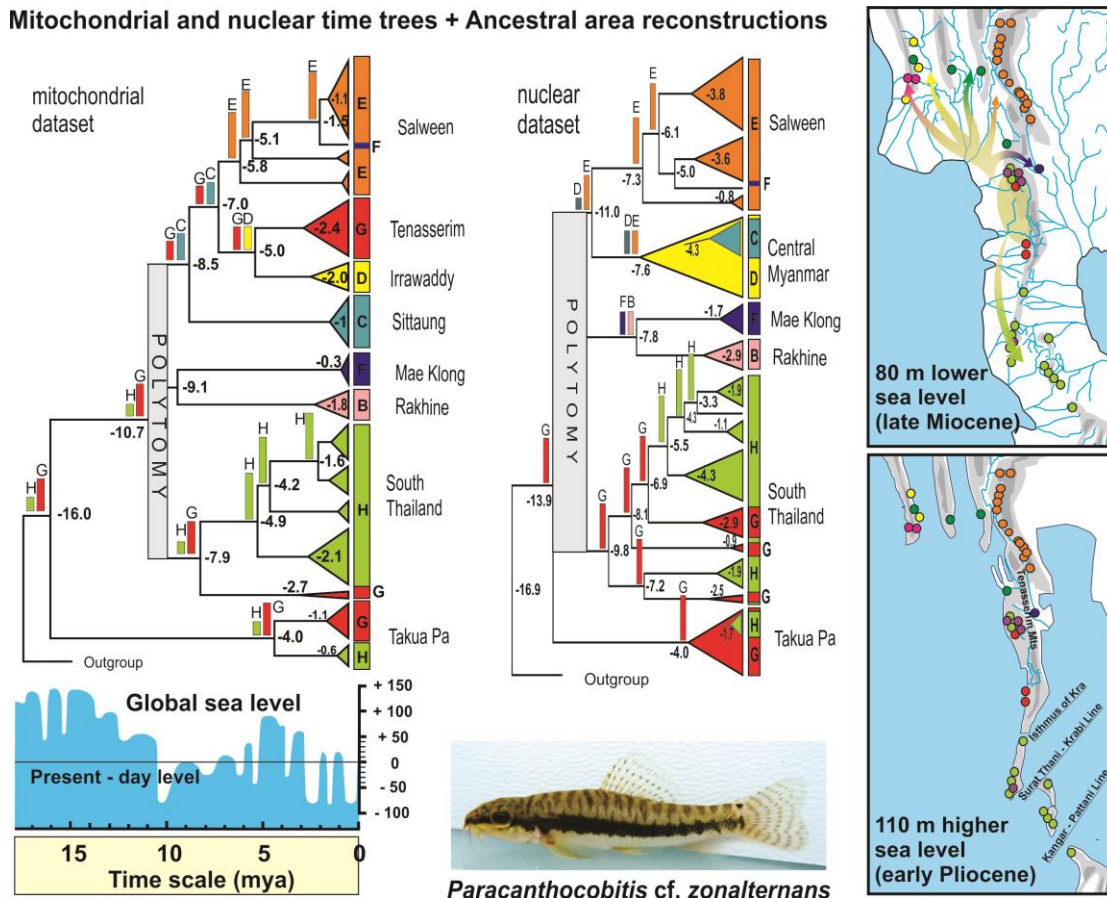
Abstract

Western Southeast Asia is hosting one of the world's most diverse faunas, and one of the reasons for this huge diversity is the complex geologic past of the area, increasing the

frequency of isolation and expansion events over evolutionary time scale. As an example case, the present study reveals the phylogeny and biogeographic history of the *Paracanthocobitis zonalternans* species complex, small benthic freshwater fish (Teleostei: Nemacheilidae) that are commonly occurring across western Southeast Asia (from central Myanmar through western and southern Thailand to northern Malaysia). The group is particularly interesting since it occurs in three biogeographic subdivisions (Burmese, Indochinese, Malay/Sundaic) and across all of the major biogeographic barriers in the region. Basing on mitochondrial and nuclear sequence data of 93 samples from about 50 localities we found six major clades, most with exclusive geographic distribution. Divergence time dated the origin of the *P. zonalternans* species complex to early Miocene (17.8 MYA) and a biogeographic analysis identified the Tenasserim region as the ancestral region. From this region the fish spread during periods of lowered global sea level, particularly during late Miocene (11-8 MYA) northwards into all Burmese river basins and southwards into south Thailand and northern Malaysia. Besides lowered global sea level periods, local stream capture events allowed the complex to expand, e.g. into the Mae Klong basin. Strong fragmentations during periods with elevated sea level during the Pliocene and Pleistocene repeatedly restricted populations to refuges and shaped the observed major lineages. Our results document a higher diversity within the *P. zonalternans* species complex than formerly believed and a strong impact of global sea level on its evolutionary history. Low sea levels promoted dispersal and elevated sea levels fragmentation events. A very similar impact of sea level changes can be expected in all stationary fauna (freshwater and terrestrial) in all non-mountainous coastal regions worldwide.

Graphical abstract

Mitochondrial and nuclear time trees + Ancestral area reconstructions



1. Introduction

Western Southeast Asia contains one of the world's most diverse faunas, both aquatic as well as terrestrial, and it has been demonstrated that in situ diversification as well as migration events have contributed significantly to the origin of such high diversity (de Bruyn et al., 2013). However, little is known about the details of the colonisation history and the biogeographic events that have shaped the diversity in this area (de Bruyn et al., 2014). Moreover, only a part of the extant biodiversity is known (Hughes, 2014; Sodhi et al., 2004). For the identification of diversification mechanisms as well as the investigation of hidden diversity, the most suited model organisms are widespread taxa with limited dispersal ability. In particular strictly freshwater organisms represent suitable models due to their restriction to water systems that reflect evolutionary events on a finer geographic scale than do habitats of terrestrial animals and plants.

Freshwater fishes of the *Paracanthocobitis zonalternans* species complex are very widely distributed from Bangladesh to northern Malaysia and are common in most rivers. For a long time the complex was considered to include a single species, *P. zonalternans* Kottelat, 1990, although Kottelat (1990) pointed on the high morphological variability of these fishes, and suggested later that several species could be hidden under the name *P. zonalternans* (Kottelat, 2012). Recently, Singer et al. (2017) proposed splitting *P. zonalternans* into five species.

Three of these species occur in western Southeast Asia from central Myanmar through western Thailand to northern Malaysia. However, the three species are morphologically similar, making it difficult to identify the evolutionary units within the species complex by morphological characters. We here investigate the possibility of resolving the internal diversity within the *P. zonalternans* species complex by the use of molecular genetic data.

The *P. zonalternans* species complex is moreover interesting biogeographically. Its distribution crosses most of the major biogeographic barriers in western Southeast Asia and reaches into three faunistic regions, the Indian, the Indochinese and the Malay region (Zakaria-Ismail, 1994). These regions generally have different faunas (Lim & Tan, 2002), and their borders are considered to be important biogeographic barriers. Indian and Indochinese faunas are separated mainly by the Tenasserim Mountain Ridge, which stretches in a north-south direction from the Shan Plateau in northeastern Myanmar onto the Malay Peninsula to the Isthmus of Kra, which is the narrowest point of the Malay Peninsula at about 10°30'N (Fig. 1). The Tenasserim Mountain Ridge separates the drainage system of the Andaman Sea from that of the Gulf of Thailand and represents the limit of distribution for several freshwater fish genera (the eastern border for e.g. *Botia*, *Neonoemacheilus*, *Osteobrama* and *Psilorhynchus* from the Indian fauna and the western border for e.g. *Boraras*, *Nemacheilus* and *Yasuhikotakia* from the Indochinese fauna). The fauna of the Malay Peninsula changes from north to south and three lines have been considered to mark the transition zone between the Indochinese fauna in the north and the Malay (Sundaic) fauna in the south (Parnell, 2013). The most northern line marks the Isthmus of Kra, where a transition in birds and terrestrial animals exists together with a shift in vegetation from semi-evergreen to evergreen monsoon rain forest (Li & Li, 2018; Richards, 1996). A pilot study in freshwater prawns identified the Surat Thani – Krabi line as a transition zone where broad marine transgressions flooded the Malay Peninsula in a west-east direction during periods of increased global sea level (de Bruyn et al. 2005; Woodruff, 2003). Botanists identified as a transition zone the Kangar – Pattani line, crossing the Malay Peninsula close to the Thai – Malay border (Corlett, 2014). It marks the transition from a perhumid climate in the south to a seasonally dry climate in the north and is associated with a shift from aseasonal to seasonal evergreen tropical forest (Baltzer et al., 2007). The Tenasserim Mountain ridge and the three transition zones represent biogeographic barriers, but the distribution of the *P. zonalternans* species complex crosses them all, making the group an interesting model in which to study the biogeographic impact of each of these barriers. Up to now, the phylogeny of only one freshwater fish group with a distribution that crosses several of these barriers has been reconstructed and used to address the biogeographic history of the region, the genus *Badis* from the family Badidae (Rueber et al, 2004). However, in contrast to members of the *P. zonalternans* complex, populations of *Badis* are rare and locally restricted, making it difficult to reach the fine geographic coverage needed in a biogeographic study. Due to the strong barrier function of the mentioned biogeographic transition zones and mountain ridges only a few other freshwater fish groups have a distribution from the Irrawaddy to northern Malaysia and to our knowledge no phylogenetic relations or biogeographic scenarios have been proposed. The fact that the *P. zonalternans* complex has such extensive distribution plus is very common and abundant makes it the ideal freshwater fish model in the region.

In the present study we use 93 samples of the *P. zonalternans* species complex from nearly the complete distribution in western Southeast Asia to reconstruct the phylogenetic relationships among populations using mitochondrial and nuclear sequence data to identify evolutionary units. Divergence times and ancestral area reconstructions provide data on the evolutionary history of the complex and allow identification of those factors that have had significant influence on the evolution of these freshwater fishes.

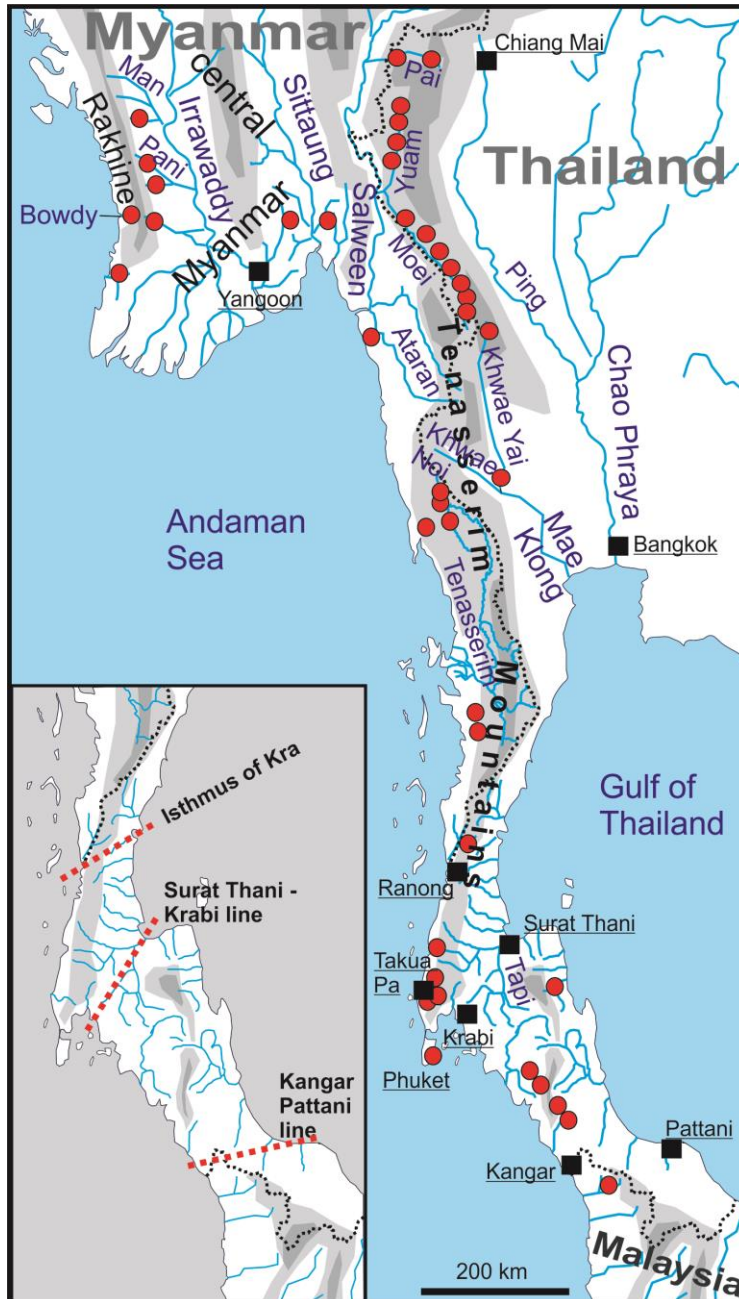


Fig. 1. Geographic overview of the study area and sampling sites. Red dots indicate sampling localities of the *Paracanthocobitis zonalternans* species complex, black squares and underlined names mark cities, names of rivers are given in dark blue and names of mountains in bold. Mountain ranges are shaded in grey (dark grey for the highest parts), and dotted lines represent country borders. The inlet map visualises the geographic position of the Isthmus of Kra, the Surat Thani - Krabi line, and the Kangar - Pattani line.

2. Material and methods

2.1. Sampling and DNA sequencing

Complete specimens or tissue samples fixed and stored in 96% ethanol came from scientific collections (CMK, Collection Maurice Kottelat; NRM, Natural History Museum, Stockholm; ZRC, Zoological Reference Collection, Singapore), from ornamental fish exporters, naturalists and scientists. Our analyses are based on original sequence data of 93 samples of the *P. zonalternans* species complex, 15 samples of other species of *Acanthocobitis* and *Paracanthocobitis* and as outgroups 16 taxa from the family Nemacheilidae and one sequence from the family Cobitidae (Supplementary material 1). Sampling covered nearly the complete distribution area in western Southeast Asia (Fig. 1).

Total genomic DNA was isolated using the DNeasy Blood & Tissue kit (Qiagen) following the manufacturer's instructions. The mitochondrial cytochrome *b* was amplified with primers Glu-L.Ca14337–14359: 5'-GAA GAA CCA CCG TTG TTA TTC AA-3' and Thr-H.Ca15568–15548: 5'-ACC TCC RAT CTY CGG ATT ACA-3' (Šlechtová et al., 2006), the 910 bp region of the nuclear RAG-1 gene was amplified with primers RAG-1F: 5'-AGC TGT AGT CAG TAY CAC AAR ATG-3' (Quenouille et al., 2004) and RAG-RV1: 5'-TCC TGR AAG ATY TTG TAG AA-3' (Šlechtová et al., 2007), and for amplification of nuclear IRBP 2 we used primers IRBP 109F 5'-AAC TAC TGC TCR CCA GAA AAR C-3' and IRBP 1001R 5'-GGA AAT GCA TAG TTG TCT GCA A-3' (Chen et al., 2008). PCRs were performed in a 25 µl reaction volume containing 10 mM Tris-HCl, 50 mM (HN4) 2SO₄, 0.1% Triton X-100, 1.2–1.8 mM MgCl₂ (PCR Blue Buffer), 2 mM TMA oxalate (PCR enhancer), 5 nmol of each nucleotide (PCR dNTP mix), 1.25 U of Taq DNA polymerase Unis (all chemicals by Top-Bio) and 12.5 pmol of each primer. For cytochrome *b* and RAG 1 we used a program with a touch-down profile (temperature is reduced by 1° C per cycle) of 1 min 30 s at 60 – 55° C and 2 min at 72° C followed by 30 cycles with annealing temperature held at 54° C. The amplification of IRBP2 included 35 cycles with annealing temperature at 59° C. For direct Sanger sequencing of PCR products we used BigDye™ Terminator Cycle Sequencing Ready Reaction Kit 1.1 (Applied Biosystems). Sequencing was performed on ABI Prism 3130 GA.

2.2. Phylogenetic analyses

The complementary strands were assembled in SeqMan II module of DNASTAR Lasergene software package, and sequences were aligned using the Clustal W algorithm implemented in BioEdit (Hall, 1999). The alignments were checked by eye for potential mistakes. The sequences were deposited in GenBank under the accession numbers MK608027 – MK608315 (Supplementary material 1). The basic statistics of alignments and among group differences in percentages were estimated in MEGA7 (Kumar et al., 2016). The best-fit nucleotide substitution model and partitioning scheme for every codon position were determined in PartitionFinder 2 (Lanfear et al., 2016) under the Bayesian information criterion (BIC), with the branch length linked and greedy search algorithm. An overview about the used models at different positions is given in Supplementary Table 2.

The phylogenies were inferred using the partitioned Bayesian analyses (BI) in MrBayes 3.2 (Ronquist and Huelsenbeck, 2003). First, each of the alignments was analysed separately to

see the degree of congruency among the dataset. Subsequently, the mitochondrial (mtDNA) and nuclear (nDNA) datasets were analysed separately. For all analyses, the data were partitioned as suggested by PartitionFinder 2. For each dataset the analysis was performed in two independent runs of 10×10^6 generations, with six Metropolis Coupled Markov Chains Monte Carlo (MCMCMC), default heating conditions, random starting trees and a sampling frequency of each 100 generations. The parameter settings corresponded to the best-fit models. The log-likelihood score distribution was examined in order to assess if stationarity was reached. For all the analyses, we applied the relative burnin (relburnin) of 25% and used the remaining 150 002 trees to build a 50 % majority rule consensus tree with posterior probabilities indicating the statistical supports of the clades.

2.3. Divergence time and ancestral range estimations

The age of the clades was estimated for the mtDNA dataset and the nDNA dataset separately using BEAST 2.4.8 (Bouckaert et al., 2014). The partitions were unlinked and assigned to the estimated evolutionary models, clocks and trees linked. We selected the Yule speciation process and relaxed lognormal molecular clock as priors. For the analysis we used three calibration points: 1) maximum age of Nemacheilidae is 60 MYA (set in analysis with uniform distribution 60 – 0), based on the fact that this highly adaptable group is not present in North America, while the related Catostomidae and Cyprinidae managed to colonise North America via the Bering land bridge about 65-60 MYA (Bagley et al., 2018; Cavender, 1991); 2) isolation of the Indian freshwater fauna from the Southeast Asian fauna 24 MYA (set in analysis with uniform distribution 45 - 24) due to the uplift of the Himalayas and the Tibetan Plateau (Ali and Aitchison, 2008; Clark et al., 2004; Klaus et al., 2010; Wang, 2004); 3) minimum age of the only known fossil record of Nemacheilidae, *Triplophysa opinata* from Kyrgyzstan in the middle-upper Miocene (16.0 to 5.3 MYA) (set the minimum age of the genus *Triplophysa* in analysis as 5.3 mya) (Böhme and Ilg, 2003; Prokofiev, 2007).

We conducted MCMC analyses of 50×10^6 generations with sampling of every 1000 generations. For some of the analyses we used the CIPRES on-line portal (www.phylo.org) (Miller et al., 2010). The results were checked in Tracer to check for the convergence, stationarity and effective sampling sizes (ESS) of all parameters. A maximum clade credibility (MCC) tree was built in TreeAnnotator v.1.8.0 (Rambaut and Drummond, 2010) after discarding the first 10% of the trees.

The biogeographical analyses were conducted with use of BioGeoBEARS package (Matzke, 2013) implemented in RASP 4.0 (Reconstruct Ancestral Stage in Phylogenies, Yu et al., 2015). Eight distributional units were set on the basis of known differences in freshwater fish composition (Jayaram, 2010, Kottelat, 1990, 2004, Kullander and Fang, 2004, Ng, 2007): A: India, B: Rakhine, C: Irrawaddy basin, D: Sittaung basin, E: Salween basin, F: Mae Klong basin, G: Tenasserim region, H: south Thailand. Besides the distribution areas definition, the input files included the MCC tree and 200 sample trees from the post burn-in trees generated by BEAST. Three simultaneous unit areas were allowed. Six biogeographic models were compared (DEC, DIVALIKE, BAYAREALIKE, plus each of them with J parameter). Based on the AIC values, the DEC+J model was selected as the best-fit (see Supplementary material 3) and consequently used for the final analysis. However, since DEC as well as DEC+J have

recently been criticised (Ree & Sanmartin, 2018), we compared the ancestral range reconstruction calculated under the DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE and BAYAREALIKE+J.

3. Results

3.1. Phylogenetic reconstructions

The phylogenies of the *P. zonalternans* complex were reconstructed using the 367 sequences of three genes (Table 1). The lengths of the final alignments of cytochrome b, RAG 1 and IRBP 2 were 1120, 950 and 810 bp, respectively. The length of the final alignment of the concatenated dataset of the nuclear genes RAG 1 and IRBP2 was 1760 bp.

The alignment of the complete cytochrome b dataset contained 47% variable (VP) and 42% parsimony informative (PI) positions, 34% VP and 32% PI inside the genus *Paracanthocobitis*, and 31% VP and 29% PI within the ingroup, the *P. zonalternans* complex. Within the complete IRBP 2 dataset 35% VP and 20% of PI were found, inside the genus *Paracanthocobitis* 10% VP and 8% PI, and inside the *P. zonalternans* complex 6% VP and 5% PI. The complete RAG 1 dataset contains 34% VP and 26% PI, the genus *Paracanthocobitis* 14% VP and 11% PI, and the ingroup 11% VP and 8% PI sites.

The genetic differences among the main lineages within the *P. zonalternans* species complex are comparably high, especially in the mitochondrial dataset (Supplementary material 4). Here the sequence divergence between the Takua Pa lineage and the other lineages ranges from 11.3 to 12.2% and from 5.7 to 9.6% among the remaining major lineages. In the nuclear genes the sequence divergences are much smaller (1.0 to 1.8%). The genetic differences between the *P. zonalternans* species complex and other species of *Paracanthocobitis* range from 13.1% to 13.8% in cytochrome b and from 2.9% to 5.4% in the nuclear genes.

The two datasets based on the nuclear genes RAG 1 and IRBP 2 identified the same major lineages with high statistical support and consequently have been combined. On the other hand, the phylogenies based on the analyses of the nuclear and mitochondrial datasets were not congruent, several cases of conflicts were recorded, and it is necessary to present the phylogenetic trees resulting from the nuclear and mitochondrial dataset analyses separately.

3.2. Major lineages

The analysis of the mitochondrial dataset separates the sample set into eight major lineages with high statistical supports, while the analysis of the nuclear dataset recovered six major lineages in the material (Fig. 2). Two major differences are found between the results from the mtDNA and nDNA dataset: First, one of the lineages in the mtDNA dataset was embedded into another lineage in the nDNA dataset (Marked by green line in Fig. 2). Second, two distinct clades in the mtDNA dataset merged to one in the nDNA dataset (yellow line in Fig. 2). The two cases where mtDNA clades were fused with another clade in the nDNA dataset are interpreted here as acquisitions by hybridisation in secondary contact (details in discussion). These two mtDNA clades are therefore eradicated as evolutionary units. Consequently, our data reveal the existence of six major clades.

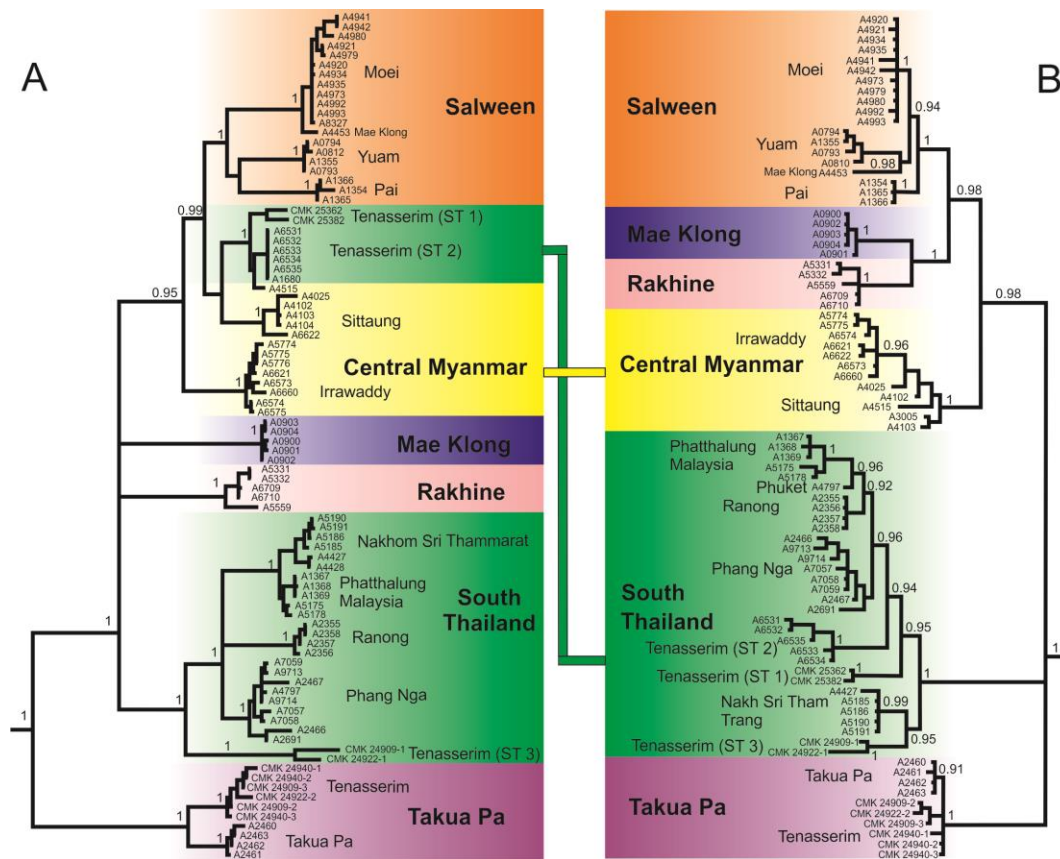


Fig. 2. Phylogenetic reconstructions of the *Paracanthocobitis zonalternans* species complex. Bayesian trees based on A) on the mitochondrial (cytochrome b) dataset, and B) on the nuclear (RAG-1 and IRBP combined) dataset. Nodes with support less than 0.9 are not shown. Major lineages indicated by different colours and bold font, sublineages by simple font. Green and yellow lines point on important differences between the two datasets.

In general the clades have exclusive geographic distribution areas (Fig. 3); therefore we use their distribution areas for naming them. One of the clades (from here onwards named ‘Takua Pa lineage’) is sister to all remaining and collects some of the samples from the Tenasserim region in southernmost Myanmar and some of the samples from the Takua Pa River basin in the Phang Nga province in southwestern Thailand. All other samples from southern Thailand, northern Malaysia and from the central Tenasserim region in southern Myanmar together form the ‘South Thailand lineage’. A third lineage (‘Rakhine lineage’) is composed from all samples from two streams in the Rakhine province of Myanmar. All specimens from the Mae Klong basin in Kanchanaburi (tributary Khwae Noi) together form the ‘Mae Klong lineage’. The fifth clade is the ‘Central Myanmar lineage’ that collects all samples from the Irrawaddy and Sittaung River drainages and the only specimen from northern Tenasserim. The last clade is the ‘Salween lineage’ that is composed from all specimens from the Salween River basin plus one specimen caught in the upper Mae Klong (tributary Khwae Yai) drainage.

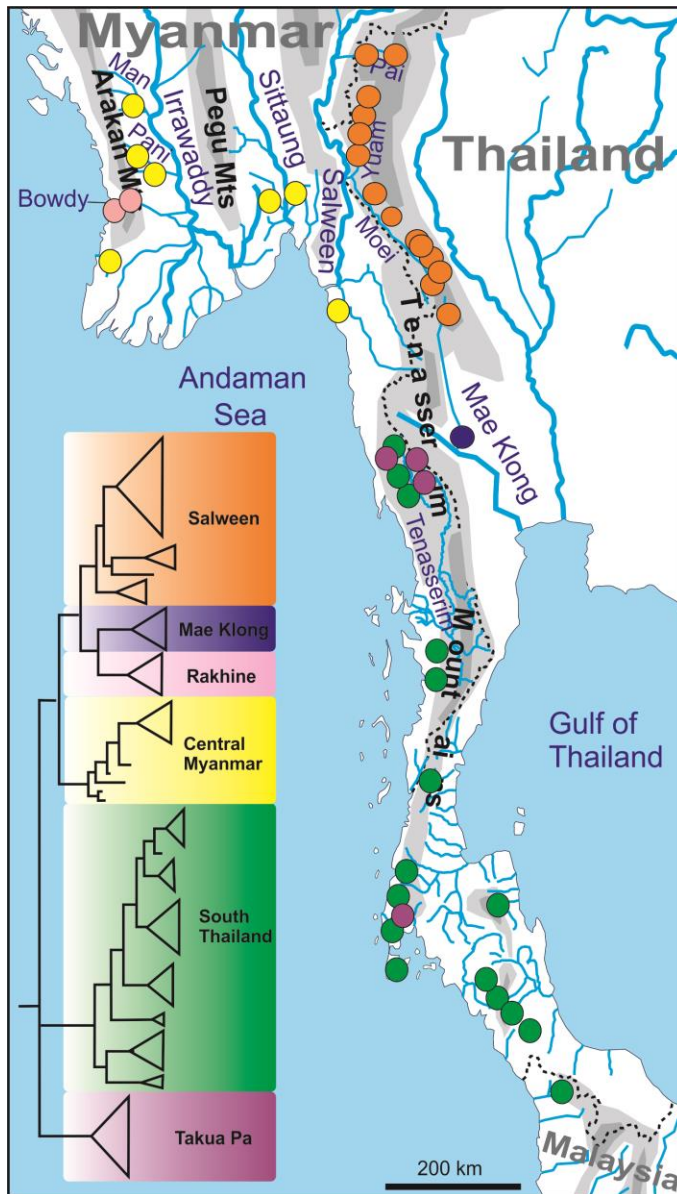


Fig. 3. Bayesian tree based on the nuclear dataset of the *Paracanthocobitis zonalternans* species complex and geographic distribution of the identified lineages. Colours in the tree match colours in the map.

3.3. Sublineages

In the mitochondrial dataset, four of the major clades contain sublineages that have distinct and exclusive geographic distribution areas. The Takua Pa lineage is separated into two sublineages, one collecting all specimens from the Takua Pa River basin, the other all samples from the Tenasserim River basin. The Central Myanmar lineage has two sublineages, one collecting all but one sample from the Irrawaddy basin and the other all samples from the Sittoung basin plus one specimen (A4515) from the northern Tenasserim region. The specimen A4515 was assigned to the Central Myanmar lineage according to the results of the nDNA analysis. The Salween lineage is composed of three subclades, each collecting the samples from one tributary of the Salween (Pai River, Yuam River and Moei River). The

Moei River sublineage additionally contains the single specimen that had been collected in the uppermost part of the Mae Klong basin, just at the watershed with the Moei basin. The South Thailand lineage consists of five subclades, one of them occurring in the central Tenasserim region in Myanmar, this means north of the Isthmus of Kra, and the other four south of it. The Ranong sublineage and the Phang Nga sublineage are distributed along the western (Andaman) coast of southern Thailand south of the Isthmus of Kra, but north of the Surat Thani – Krabi line, while the sister sublineages Nakhon Si Thammarat and Phattalung/Malaysia are located south of the Surat Thani – Krabi line. The Nakhom Sri Thammarat sublineage collects all but one samples from the region between the Surat Thani – Krabi line and the Kangar – Pattani line and the Patthalung/Malaysia sublineage all samples from south of the Kangar – Pattani line.

In the nuclear dataset only the three sublineages within the Salween clade were detected. In the South Thailand clade some groups of samples from certain subregions were visible, but with lower support and different topology than in the mtDNA dataset. Moreover, no geographic pattern in the topology of these groups was recognisable. In the Takua Pa clade and the Central Myanmar lineage one of the subclades (Takua Pa and Irrawaddy, respectively) that were found in the mitochondrial dataset were recovered as monophyletic in the nuclear dataset, but stemmed out of a paraphyletic assemble of the second subclade (Tenasserim and Sittaung, respectively).

3.4. Divergence time estimation

Analysis of the mtDNA dataset: The BEAST analyses based on the mitochondrial dataset dated the origin of the *P. zonalternans* species complex to about 17.9 MYA (Fig. 4, full MCC tree including also outgroup taxa and confidence intervals provided in Supplementary material 5). The Takua Pa lineage separated from the remaining major clades about 16.0 MYA. Since the average cytochrome *b* sequence divergence of the Takua Pa lineage from the remaining lineages is 11.8%, the mutational rate would correspond to 0.37% per lineage per million years (0.74% divergence per pairwise comparison). This mutation rate is very similar to the rate of 0.34% obtained in species of the closely related family Cobitidae by Doadrio and Perdices (2005), which used for molecular clock calibration the opening of the strait of Gibraltar.

Around 10.7 MYA the remaining group split into four clades. Their exact branching order and times varied between analyses and nodes had always very low statistical support, indicating that these splits occurred within a narrow time window. This polytomy appeared in all analyses of both our datasets and is marked in Fig. 4 by grey boxes. One of the resulting clades corresponds to the South Thailand clades, another two contain the clades Mae Klong and Rakhine and the last one contains the clades Salween, Sittaung, Tenasserim and Irrawaddy clades. The sister relation of Mae Klong and Rakhine clades shown in Fig. 4 is an artefact of the biogeographic analysis, which does not allow polytomies. The Irrawaddy lineage separated from Salween, Sittaung and Tenasserim clades 8.5 MYA, the Salween lineage separated 7.0 MYA and the sister lineages Sittaung and Tenasserim shared the most recent common ancestor 5.0 my. The three sublineages within the Salween clade formed 5.8 to 5.1 MYA and the single specimen that had been collected in the upper Mae Klong basin is

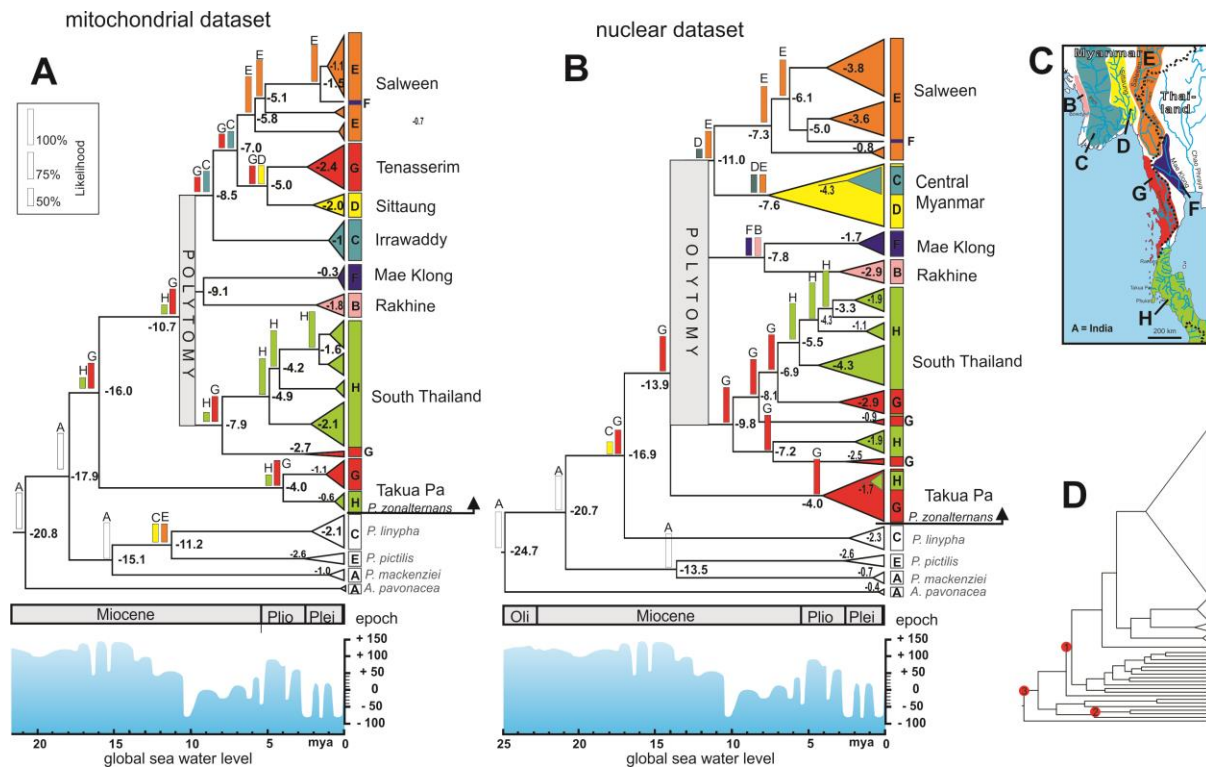


Fig. 4. Divergence time estimations, ancestral area reconstruction and comparison with global sea level in the *Paracanthocobitis zonalternans* species complex. **A)** Ultrametric tree based on the mitochondrial dataset; node labels give the estimated time of divergence; small bars left of nodes give the most likely reconstructed ancestral area(s) for the node, letters and colours of small bars correspond to Fig. 4 C, the height of the bars is proportional to its likelihood, colours of the major lineages indicate their area of distribution and correspond to colours in Fig. 4 C; tree was calibrated using the three calibration points shown in Fig. 4 D, for better visualisation only part of the outgroup is shown (species of *Acanthocobitis* and *Paracanthocobitis*); time scale and global sea level over the last 22 my given below the tree. **B)** Ultrametric tree based on the nuclear dataset; details as in Fig. 4 A. **C)** Map indicating the geographic areas used for the ancestral area reconstruction, colours correspond to colours in Fig. 4 A and B. **D)** Calibration points used to calibrate the time tree: 1 – age of the Indian Nemacheilidae, age 45–24 MYA, 2 – age of the genus *Triplophysa*, min age 5.6 MYA, 3 – minimum age of Nemacheilidae, 60 MYA.

isolated from the Moei sublineage for 1.5 my. Within the South Thailand clade, the separation between the Tenasserim sublineage and the remaining four, more southerly distributed lineages occurred 7.9 MYA, while the southern sublineages have ages of 4.9 to 1.6 my. The two sublineages within the Takua Pa lineage formed 4.0 MYA.

Analysis of the nDNA dataset: The interesting incongruences between mitochondrial and nuclear dataset in the *P. zonalternans* complex demand a comparison of the datasets. The BEAST analysis based on the nuclear dataset dated the origin of the *P. zonalternans* complex to 16.9 mya, the separation of the Takua Pa lineage to 13.9 mya and assigned the polytomy between major clades to the time span 13.9 to 11.1 mya. In the nuclear dataset the Salween and central Myanmar clades separated 11.0 mya. The South Thailand clade experienced seven separation events in the period 9.8 to 3.3 mya.

3.5. Ancestral range estimations

In the comparison of three biogeographic models (Model Test in RASP), each one tested with and without J factor, all three models with the J factor have very similar Lln values. The low p-values of the Likelihood Ratio Test that compares the J and non-J model indicate that the addition of the J factor significantly influences the likelihood of the given model (Supplementary material 3). The three models reconstructed the same ancestral areas at each node with high likelihood. In contrast, the same models without the J factor led to incongruent results and very low likelihood values. The results show on the one hand that the J factor had a significant impact on the ancestral range estimation and on the other hand that the present dataset can be considered robust to the selection of the model.

The biogeographic analyses applying the DEC+J, DIVALIKE+J and BAYAREA+J models on the mtDNA as well as on the nDNA datasets suggested India as the area of origin for the genera *Acanthocobitis* and *Paracanthocobitis* (Fig. 4). The estimated ancestral area of the *P. zonalternans* species complex was the Tenasserim region. The Tenasserim region is still identified as most likely ancestral region for the major clades Takua Pa and South Thailand. According to the mitochondrial dataset, the ancestral area for the common ancestor of the Salween and central Myanmar clades was either in the Irrawaddy basin or the Tenasserim region; according to the nuclear dataset it could have been in the Salween or Irrawaddy basin. We interpret these results in the way that the *P. zonalternans* complex spread from the Tenasserim region with two major clades (Takua Pa, South Thailand) southwards into the South Thailand region and with three branches (two formed by the ancestral forms of the Rakhine and Mae Klong clades and one formed by the common ancestor of the Salween and central Myanmar clades) northwards into the area of Rakhine and the river basins of Irrawaddy, Sittaung, Salween and Mae Klong. The further formation of major clades and subclades took place in the areas in which these groups are nowadays found.

4. Discussion

4.1. Phylogeny of the *P. zonalternans* species complex

Our study represents the first detailed molecular phylogeny of the *P. zonalternans* species complex and demonstrates its monophyly within the genus *Paracanthocobitis*.

The analyses of the mitochondrial as well as the nuclear dataset identify eight and six clades, respectively. These differences represent cases of genetic takeover by another clade and will be discussed below. Orienting primarily on the nuclear dataset we identify six major clades within the group.

4.2. Evolutionary units within the *P. zonalternans* species complex

Our results demonstrate a strong diversity within the *P. zonalternans* species complex. We identified six major lineages, which we consider to represent evolutionary units. In all but one case they have disjunct geographic distribution areas. Only in the case of Takua Pa and South Thailand clades, specimens of two different clades occur in syntopy. Geographic isolation has long been considered one of the most effective factors leading to diversification of animal

lineages (Worsham et al., 2017). Three of the major clades include sublineages in the mitochondrial dataset, and in all three cases the subclades inhabit separate areas, indicating that the same diversification mechanism that has shaped the major clades still acts also on a smaller geographic scale.

The current taxonomy of the *P. zonalternans* species complex, based on morphological data suggested three species to occur in western Southeast Asia, namely *P. zonalternans* in the Salween River basin, *P. nigrolineata* in the basins of Irrawaddy, Sittaung and Mae Klong, and *P. phuketensis* in southern Thailand (Singer et al., 2017). Mapping this concept on our results, our Salween clade corresponds to *P. zonalternans*, and our South Thailand clade to *P. phuketensis*. In our results the central Myanmar rivers and the Mae Klong are inhabited by two different evolutionary units, rejecting the assumption of a single species in these river basins. Since the type locality of *P. nigrolineata* is in the Mae Klong basin, our Mae Klong clade corresponds to *P. nigrolineata*. Our data additionally identify the Rakhine and the Takua Pa clades, indicating that the actual diversity is much higher than presently considered. A focussed taxonomic analysis will be needed to evaluate the species status of the six identified evolutionary units.

4.3. Evolutionary history of the *P. zonalternans* species complex

Due to the better resolution and the linear mode of inheritance without recombination we will base the discussion of the evolutionary history mainly on the results of the mitochondrial dataset. Differences to the nuclear dataset are visible from Fig. 4. The divergence time analysis revealed that *P. zonalternans* separated from the remaining genus *Paracanthocobitis* about 17.9 MYA in the early Miocene and the biogeographic reconstruction identified the Tenasserim region as the area of origin of *P. zonalternans*. The separation of the Takua Pa lineage from the ancestor of the remaining lineages took place in the Tenasserim region. This can best be imagined as a case of geographic isolation, since Tenasserim is bordered in the east by the Tenasserim Mountain ridge and in the west by the Andaman Sea (Fig 5 a). During the early Miocene the global sea level was significantly higher than today, meaning that Andaman Sea and Tenasserim Mountains met at several places and prevented faunistic exchange between the Tenasserim region and neighbouring areas. At about 16 mya the Takua Pa clade appeared, also in the Tenasserim region (Fig. 5 b). Since this clade occurs in syntopy with the South Thailand clade across all of its known distribution area, the speciation mechanism here most likely was sympatric speciation by ecologic factors. Later, in late Miocene about 10.5-9.5 MYA, the global sea level dropped lower than ever before during Cenozoic to about -90 m (Haq et al., 1987; Pearson and Palmer, 2000). This serious drop of sea water led to the retreat of the Andaman Sea coastline for up to 200 km (Fig. 5 c) and offered suitable conditions for a range extension to the north into central Myanmar and in the south to southern Thailand. Therefore the massive lowering of the global sea level in late Miocene was responsible for the colonisation of nearly the entire area that presently is inhabited.

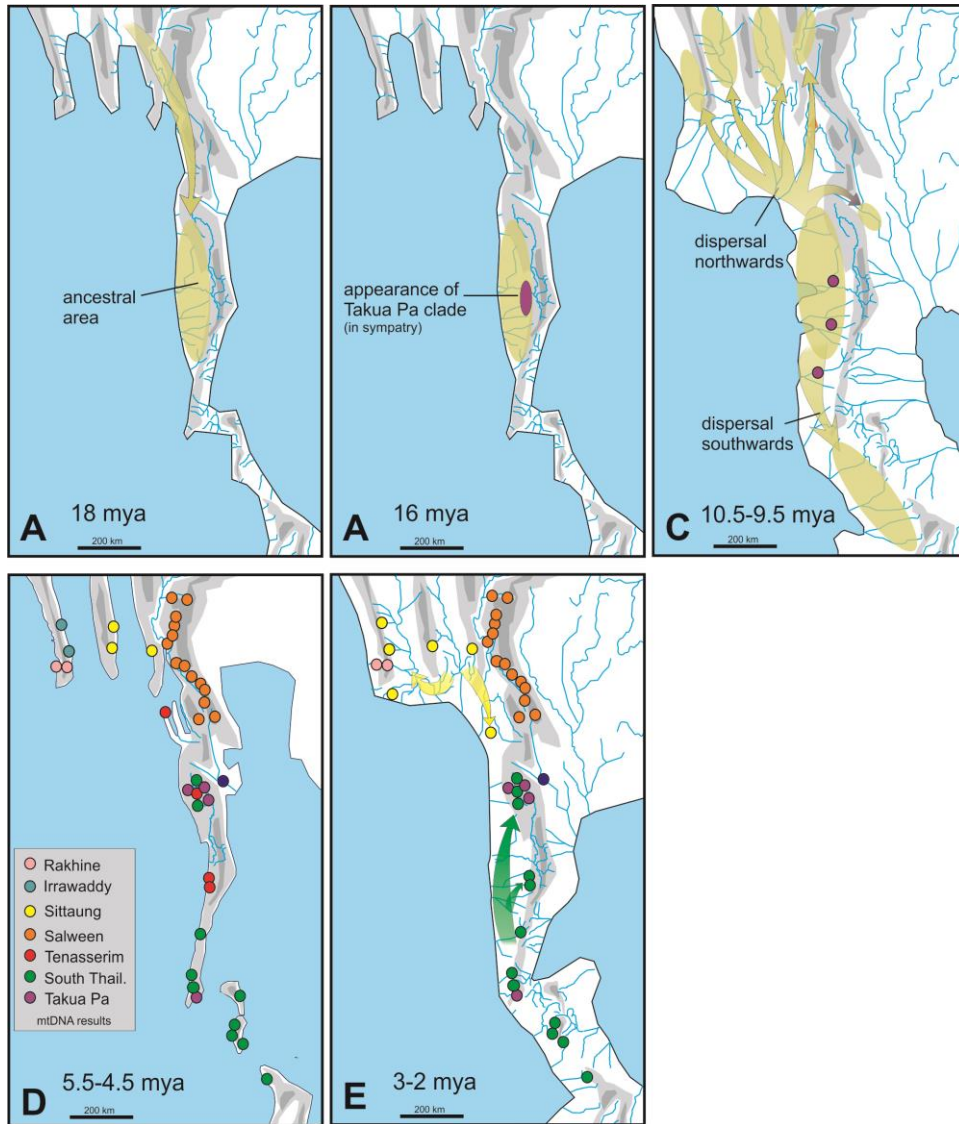


Fig. 5. Visualisation of dispersal during periods of low global sea level and restriction during periods of increased sea level. **A** - 18 mya: colonisation of the Tenasserim area and subsequent isolation of the founder population of the *P. zonalternans* complex; **B** - about 16 mya: separation into ancestral Takua Pa clad and common ancestor of remaining clades (assumed sympatric speciation); **C** - late Miocene (10.5-9.5 mya): strongly lowered global sea level (-90 m below present day level), massive range expansion northwards as well as southwards; **D** - early Pliocene (5.5-4.5 mya): strongly increased global sea level (+100 m higher than present day); serious fragmentation of population leading to the formation of major clades (mtDNA results); **E** - 3-2 mya: slight decrease of global sea level, range expansion of Sittaung lineage and South Thailand clade, leading to secondary contact and nuclear assimilation of Irrawaddy and Tenasserim lineages (nDNA results).

After this initial dispersal period, the *P. zonalternans* species complex underwent a significant diversification. The isolation of the Mae Klong clade from the remaining clades can be explained by the Tenasserim Mountain ridge lying as barrier between the Mae Klong and the remaining clades. The only imaginable pathway for the colonisation of the Mae Klong basin is by local stream capture from the northern Tenasserim region or the Salween

basin. This scenario is supported by the occurrence of other fish species (e.g. *Schistura balteata*) in the upper Mae Klong as well as in the northern Tenasserim region (Kottelat, 1990; Page et al., 2012) and by our specimen A4453 which genetically belongs to the Salween clade, but was collected in the upper Mae Klong basin. In contrast, the separation of the Rakhine, Irrawaddy, Sittaung and Salween clades most likely was driven by a very high global sea level during late Miocene-early Pliocene (5.2-4 MYA), when the oceans rose to about +100 m and the lowland regions and lower river valleys were flooded by marine water (Haq et al. 1987; Hughes et al. 2003). Plotting the localities of the major lineages on a map with an increased sea level during late Miocene-early Pliocene c indicates that the genetic lineages as identified in our analyses must have evolved from populations on upland refuges that were disconnected by sea water during periods with increased ocean level (Fig. 5 d). This fragmentation of the *P. zonalternans* populations in central Myanmar and the Salween together with a subsequent low rate of faunistic exchanges was the foundation of the major lineages in central Myanmar as they were identified in our analyses.

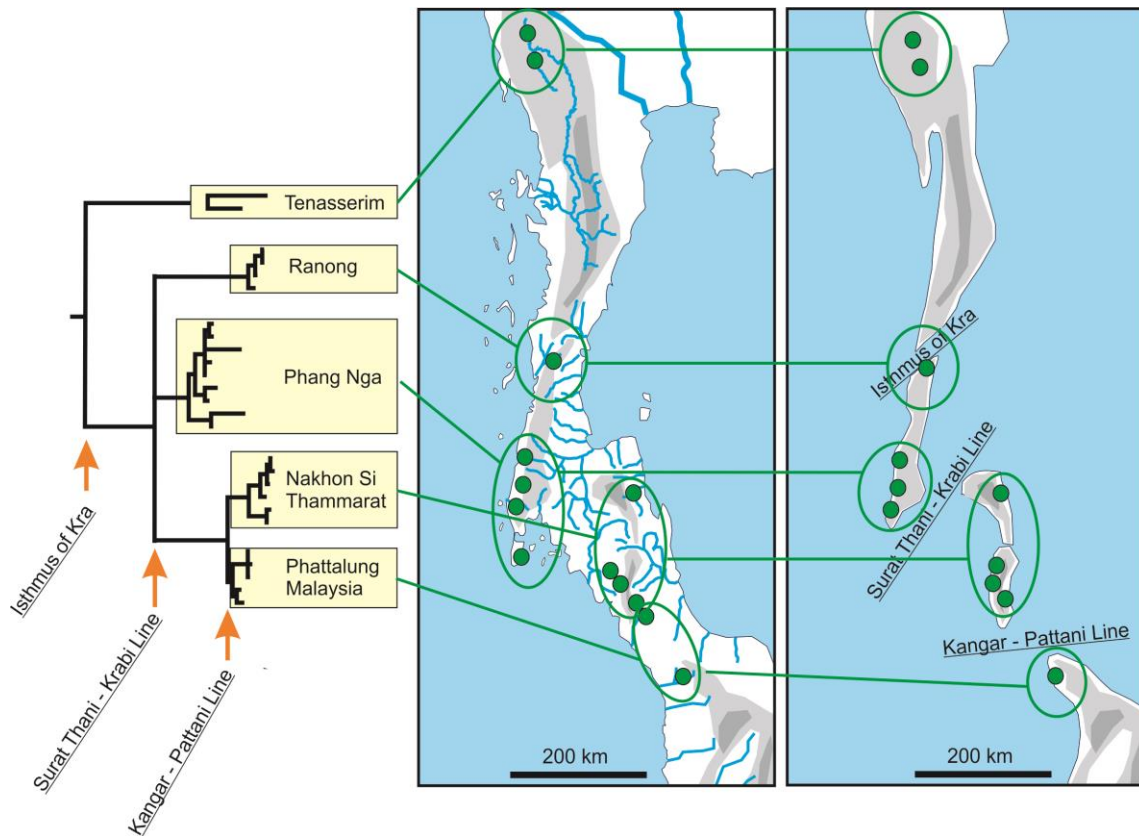


Fig. 6. Sublineages of the South Thailand clade and their distribution area during present-day sea level and during periods of increased sea level. **A)** detail of the Bayesian tree basing on the mitochondrial dataset showing the South Thailand clade; **B)** distribution area of the five subclades on a map showing the present-day sea level; **C)** the same map during a period of increased sea level, e.g. early Pliocene (5.5-4.5 MYA). The separated fragments of the population correspond to the identified subclades in the genetic analysis.

Also in the south of the distribution area a strong fragmentation event during periods of high water level occurred, as depicted in Fig. 6. The most prominent marine transgression was along the three proposed biogeographic transition lines Isthmus of Kra, Surat Thani – Krabi line and Kangar-Pattani line (Li and Li, 2018, Woodruff, 2003). The most important of the observed splits, the separation of the populations north and south of the Surat Thani – Krabi line, took place at – 4.9 my, this means exactly during the period of maximum flooding. The separation of the southern subclades (Nakhom Si Thammarat and Phatthalung/Malaysia) took place – 1.6 mya during early Pleistocene during another high stand of the global sea level (see eustatic curve in Figs. 4). The only proposed biogeographic barrier that cannot functionally be linked with a period of increased sea water in the Isthmus of Kra, but the fact that it anyhow is reflected in the phylogeny of the *P. zonalternans* species complex points on its importance. However, we see the impact of all three proposed zoogeographic barrier zones on the phylogeny and biogeography of our freshwater model and again identify changes in the global sea level as the major factor for diversification.

Although the major clades generally show a pattern of not co-occurring, but on local scale we can detect cases when one single specimens of one clade were found in the distribution area of another clade. The first example is the occurrence of a specimen that in mitochondrial as well as in nuclear dataset was demonstrated to belong to the Salween clade (specimen A4453) in the uppermost Khwae Yai basin, the northern major tributary of the Mae Klong River. The collecting locality of A4453 is located only few km from the watershed between The Khwae Yai (Mae Klong basin) and the Moei (Salween basin). Most likely this occurrence is the result of a local river capture event that, according to our time tree, took place about 1.5 mya. As second example of a local stream capture event we take the occurrence of a specimen that was assigned to the Rakhine clade in both datasets (specimen A5559) in the uppermost stretch of a stream draining into the Irrawaddy basin, just at the watershed to the Bowdy Chaung where only Rakhine specimens were found. Although both cases lead to the situation that two clades occur in the same river basin, these clades are not directly in contact (yet), since both cases were recorded on the very margin of river basins far away from the next known population of the *P. zonalternans* group. However, such local cases might be the beginning of a secondary contact and hybridisation events in the future. Some past hybridisation events as consequence of former cases of secondary contact are discussed in the following paragraph.

4.4. Nuclear versus mitochondrial data

The differences in the topologies between the mtDNA based tree and the nDNA based tree are worth a closer look, since these differences reveal evolutionary events in the past. For better illustration the situation is depicted in Fig. 7. Two differences are most prominent in this respect; the first of them regards the existence of an mtDNA clade that is formed by specimens from the Tenasserim region and which is related to the Central Myanmar and Salween clades, while most of these samples in the nDNA tree are members of the South Thailand clade and one (specimen A4545) is member of the Central Myanmar clade. The results demonstrate that in the Tenasserim region a distinct clade had evolved and was isolated for long enough to develop an own mitochondrial lineage ('Tenasserim clade').

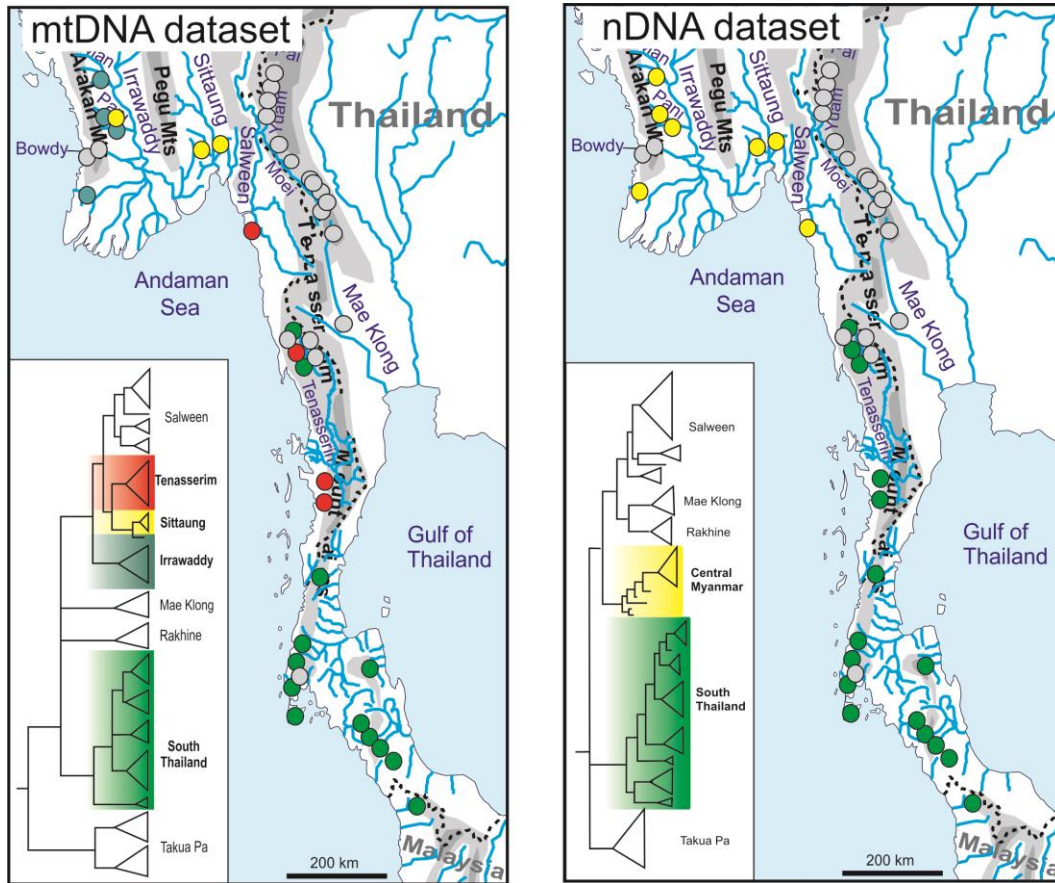


Fig.7. Differences between mtDNA tree and nDNA dataset visualised on the map. In the Tenasserim region and the Irrawaddy basin own clades had evolved according to the mtDNA dataset, but were assimilated by the Sittaung and South Thailand clade during secondary contact.

Later on, the distribution area of the complete clade was most likely colonised by members of the South Thailand clade in the south and of the Central Myanmar clade in the north. During these events of secondary contact, the clades hybridised extensively and the nuclear material of the Tenasserim clade was exchanged for material from the newly arrived clades. By this, the monophyletic mitochondrial lineage is potentially the only ghost that is left from the Tenasserim clade. The clade itself went extinct by massive exchange of nuclear material. The second difference between mtDNA and nDNA datasets we want to point on regard the Central Myanmar samples. In the nuclear dataset the samples of the Central Myanmar clade form a monophyletic lineage, while in the mitochondrial dataset they split into an Irrawaddy clade, a Sittaung clade and the above mentioned specimen from the Tenasserim clade. These mtDNA clades have not even sister-relation in the resulting tree. This case again demonstrates an ancient radiation in the *P. zonalternans* group that was subsequently eliminated by secondary contact. In the case of the Central Myanmar samples, the basins of Irrawaddy and Sittaung must have been isolated from each other long enough to enable two local mtDNA lineages to evolve, most likely during the high sea water stand during early Pliocene. This situation is still present in our mtDNA tree. Later on, the two clades came into secondary contact, hybridised extensively and merged their nuclear material into a single

clade. The fact that in the nuclear dataset the samples from the Irrawaddy basin have a tendency to cluster together into a subclade that stems out of a group formed by the Sittaung samples suggest that the secondary contact came in form of Sittaung specimens colonising the Irrawaddy basin and not reciprocal. This direction of colonisation is further supported by the observation that one specimen from the Irrawaddy basin (specimen A6622) carried an ancestral Sittaung mitochondrion. In both cases the ‘new’ nuclear haplotypes that are found in the area of old mitochondrial lineages started radiating about 2.9 mya, when the sea level sunk for a second time dramatically (see eustatic curve in Fig. 4). We hypothesise that this lowering of sea level opened the colonisation pathways that led to the expansion of the Sittaung and South Thailand clades and subsequently to the assimilation of the Tenasserim clade.

Both cases together demonstrate on the one hand that the clades within the *P. zonalternans* group can hybridise extensively when they come into secondary contact. On the other hand these cases remind us that evolutionary history does not only include radiation events, but also cases of merging in secondary contact and area piracy. Unfortunately such evolutionary events are rarely detectable in datasets that base on present-day samples, and the *P. zonalternans* complex proofed its status as suited biogeographic model by revealing even the ghost of a past clade.

Conclusions

The results of the present study reveal much higher diversity within the geographically widespread *P. zonalternans* species complex than previously believed; opting the presence of six instead of three species and pointing on the need of a taxonomic revision. They further show that the changes in the global sea level have been the most important factor in the evolution of this species group. Low sea levels promoted dispersal and elevated sea levels fragmentation events. Since salt water restricts nearly all freshwater and many terrestrial organisms, we can postulate a similar impact in most other stationary organisms. This would highlight global sea level as one of the most important driving forces for the evolutionary history of not only organisms in the biodiversity hotspot Southeast Asia, but theoretically for all non-mountainous coastal regions worldwide.

Acknowledgements

We would like to thank R. Hoyer, M. Kottelat (CMK), J. Kühne, F. F. Kullander (NRM), K. Lim (ZRC), H. Linke, G. Ott, M. Reichard, I. Seidel, 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. R. Hoyer kindly provided a photo of live *P. zonalternans* ‘Irrawaddy’. We are obliged to two anonymous reviewers for their constructive and helpful advices. The study got financial support from grants 206/08/0637 and 19-18453S of the Grant Agency of the Czech Republic and a support grant from Deutsche Gesellschaft für Ichthyologie.

References

- Ali, J.R., Aitchison, J.C., 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth Sci. Rev.* 88, 145–166. doi:10.1016/j.earscirev.2008.01.007
- Bagley, J.C., Mayden, R.L., Harris, P.M., 2018. Phylogeny and divergence times of suckers (Cypriniformes: Catostomidae) inferred from Bayesian total-evidence analyses of molecules, morphology, and fossils. *PeerJ* 2018, 6: e5168. doi:10.7717/peerj.5168.
- Baltzer, J.L., Davies, S.J., Noor, N.S., Kassim, A.R., LaFrankie, J.V., 2007. Geographical distributions in tropical trees: can geographical range predict performance and habitat association in co-occurring tree species? *J. Biogeogr.* 34, 1916–1926. doi:10.1111/j.1365-2699.2007.01739.x
- Böhme, M., Ilg, A., 2003. fosFARbase, www.wahre-staerke.com/ (accessed 17.Octobre 2018)
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* 10, e1003537. doi: 10.1371/journal.pcbi.1003537
- Cavender, T.M., 1991. The fossil record of the Cyprinidae, in: Winfield, I.J., Nelson, J.S. (Eds.), *Cyprinid fishes: Systematics, biology and exploitation*. Chapman and Hall, London, pp. 34–54.
- Chen, W.-J., Miya, M., Saitoh, K., Mayden, R.L., 2008. 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. doi:10.1016/j.gene.2008.07.016
- Clark, M.K., Schoenbohm, L.M., Royden, L.H., Whipple, K.X., Burchfield, B.C., Zhang, X., Tang, W., Wang, E., Chen, L., 2004. Surface uplift, tectonics, and erosion of eastern Tibet from large-scale draining patterns. *Tectonics* 23, TC1006, doi:10.1029/2002TC001402
- Corlett, R.T., 2014. *The ecology of tropical East Asia*, second ed. Oxford University Press, Oxford.
- de Bruyn, M., Nugroho, E., Hossain, M.M., Wilson, J.C., Mather, P.B., 2005. Phylogeographic evidence for the existence of an ancient biogeographic barrier: the Isthmus of Kra Seaway. *Heredity* 94, 370–378. doi:10.1038/sj.hdy.6800613
- de Bruyn, M., Rüber, L., Nylinder, S., Stelbrink, B., Lovejoy, N.R., Lavoué, S., Tan, H.H., Nugroho, E., Wowor, D., Ng, P.K.L., Siti-Azizah, M.N., von Rintelen, T., Hall, R., Carvalho, G.R., 2013. Paleo-Drainage Basin Connectivity Predicts Evolutionary Relationships across Three Southeast Asian Biodiversity Hotspots. *Syst. Biol.* 62, 398–410. doi: 10.1093/sysbio/syt007
- de Bruyn, M., Stelbrink, B., Morley, R.J., Hall, R., Carvalho, G.R., Cannon, C.H., van den Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L., Maiorano, L., Shoup, R., von Rintelen, T., 2014. Borneo and Indochina are Major Evolutionary Hotspots for Southeast Asian Biodiversity. *Syst Biol* 63, 879–901. doi: 10.1093/sysbio/syu047

- Doadrio, I., Perdices, A., 2005. Phylogenetic relationships among the Ibero-African cobitids (Cobitis, Cobitidae) based on cytochrome b sequence data. *Mol. Phylogen. Evol.* 37, 484-493. doi: 10.1016/j.ympev.2005.07.009
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41, 95-98.
- Haq, B.U., Hardenbol, J., Vail, P.R., 1987. Chronology of fluctuating sea levels since the Triassic. *Science* 235, 1156-1167.
- Hughes, A.C., 2014. Understanding the drivers of Southeast Asian diversity loss. *Ecosphere* 8, 1-33. doi.org/10.1002/ecs2.1624.
- Hughes, J.B., Round, P.D., Woodruff, D.S., 2003. The Indochinese–Sundaic faunal transition at the Isthmus of Kra: an analysis of resident forest bird species distributions. *J. Biogeogr* 30, 569-580. doi:10.1046/j.1365-2699.2003.00847.x
- Jayaram, K.C., 2010. The freshwater fishes of the Indian region. 2nd edition, Narendra, Dehli.
- Klaus, S., Schubart, C.D., Streit, B., Pfenninger, M., 2010. When Indian crabs were not yet Asian - biogeographic evidence for Eocene proximity of India and Southeast Asia. *BMC Evolutionary Biology* 10, 287. doi: 10.1186/1471-2148-10-287
- Kottelat, M., 1990. Indochinese nemacheilines. A revision of nemacheiline loaches (Pisces: Cypriniformes) of Thailand, Burma, Laos, Cambodia and southern Viet Nam. Pfeil Verlag, München.
- Kottelat, M. 2004. *Botia kubotai*, a new species of loach (Teleostei: Cobitidae) from the Ataran River basin (Myanmar), with comments on botiine nomenclature and diagnosis of a new genus. *Zootaxa* 401: 1–18.
- Kottelat, M. 2012. Conspectus cobitidum: an inventory of the loaches of the world (Teleostei: Cypriniformes: Cobitoidei). *The Raffles Bulletin of Zoology, Suppl.* (26):1-199.
- Kullander, S.O., Fang, F., 2004. Seven new species of *Garra* (Cyprinidae: Cyprininae) from the Rakhine Yoma, southern Myanmar. *Ichthyol. Explor. Freshw.* 15, 257-278.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 33, 1870-4. doi: 10.1093/molbev/msw054
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution formolecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* DOI: dx.doi.org/10.1093/molbev/msw260
- Li, F., Li, S., 2018. Palaeocene-Eocene and Plio-Pleistocene sea-level changes as “species pumps” in Southeast Asia: Evidence from *Althepus* spiders. *Mol. Phylogen. Evol.* 127: 545-555. doi: 10.1016/j.ympev.2018.05.014
- Lim, K.K.P., Tan, H.H., 2002. Freshwater fish diversity in Penninsular Malaysia: A review of recent findings, in: Ali, A., Rawi, C.S.M., Mansor, M., Nakamura, R., Ramakrishna, S., Mundkur, T. (Eds.), *The Asian wetlands: bringing partnership into good wetland practices*. Penerbit Universiti Sains Malaysia, Penang, pp. 461-481
- Matzke, N.J., 2013. "Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing." *Front. Biogeogr.* 5, 242-248. doi.org/10.21425/F5FBG19694

- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8.
- Ng, H.H., 2007. *Botia udomritthiruji*, a new species of botiid loach from southern Myanmar (Teleostei: Botiidae) Zootaxa 1608, 41–49. doi: 10.5281/zenodo.178863
- Page, L., Plongsesthee, R., William, F., et al. 2012. *Schistura* (Teleostei: Nemacheilidae) in the Mae Khlong basin in southwestern Thailand with description of a new species. Zootaxa 3586. doi: 10.11646/zootaxa.3586.1.30
- Parnell, J., 2013. The biogeography of the Isthmus of Kra region: a review. Nord. J. Bot. 31, 1-15. doi: 10.1111/j.1756-1051.2012.00121.x
- Pearson, P.N., Palmer, M.R., 2000. Atmospheric carbon dioxide concentrations over the past 60 million years. Nature 406, 695–699. doi: <https://doi.org/10.1038/35021000>
- Prokofiev, A.M., 2007. Redescription of a fossil loach *Triplophysa opinata* (Yakowlew, 1959) from the Miocene of Kirgizia (Balitoridae: Nemacheilinae). J. Ichthyol. 47, 26–31. doi: 10.1134/S0032945207010031
- Quenouille, B., Bermingham, E., Planes, S., 2004. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear sequences. Mol. Phylogen. Evol. 31, 66-88. doi: 10.1016/S1055-7903(03)00278-1
- Rambaut, A., Drummond, A.J., 2010. TreeAnnotator version 1.6.1 [computer program] <http://beast.bio.ed.ac.uk>.
- Ree, R.H., Sanmartin, I., 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. Febr. 2018. doi: 10.1111/jbi.13173
- Richards, P.W., 1996. The tropical rain forest: an ecological study, 2nd edn. Cambridge University Press, Cambridge.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Rueber, L., Britz, R., Kullander, S., Zardoya, R. 2004. Evolutionary and biogeographic patterns of the Badidae (Teleostei: Perciformes) inferred from mitochondrial and nuclear DNA sequence data. Mol. Phylogen. Evol. 32, 1010-22. doi:10.1016/j.ympev.2004.04.020
- Singer, R.A., Pfeifer, J., Page, L.M., 2017. A revision of the *Paracanthocobitis zonalternans* (Cypriniformes: Nemacheilidae) species complex with descriptions of three new species. Zootaxa 4324: 85-107. doi: 10.11646/zootaxa.4324.1.5
- Šlechtová, V., Bohlen, J., Freyhof, J., Ráb, P., 2006. Molecular phylogeny of the Southeast Asian freshwater fish family Botiidae (Teleostei: Cobitoidea) and the origin of polyploidy in their evolution. Mol. Phylogen. Evol. 39, 529-541. doi: 10.1016/j.ympev.2005.09.018
- Šlechtová, V., Bohlen, J., Tan, H.H., 2007. Families of Cobitoidea (Teleostei; Cypriniformes) as revealed from nuclear genetic data and the position of the mysterious genera *Barbucca*, *Psilorhynchus*, *Serpenticobitis* and *Vaillantella*. Mol. Phylogen. Evol. 44, 1358-1365. doi: 10.1016/j.ympev.2007.12.018

- Sodhi, N.S., Koh, L.P., Rahayu, D.L., 2004. Southeast Asian biodiversity: an impending disaster. *Trends Ecol Evol* 19: 654-660. doi:10.1016/j.tree.2004.09.006
- Wang, P., 2004. Cenozoic 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*. Geophysical Monograph 149, 1-22.
- Woodruff, D.S., 2003. Neogene marine transgressions, palaeogeography and biogeographic transitions on the Thai–Malay Peninsula. *J. Biogeogr.* 30, 551-567. doi: 10.1046/j.1365-2699.2003.00846.x
- Worsham, M.L., Julius, E.P., Nice, C.C., Diaz, P.H., Huffman, D.G., 2017. Geographic isolation facilitates the evolution of reproductive isolation and morphological divergence. *Ecol. Evol.* 7, 10278-10288. doi: 10.1002/ece3.3474
- Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historic biogeography. *Mol Phylogenet Evol* 87, 46-49. doi: 10.1016/j.ympev.2015.03.008.
- Zakaria-Ismail, M., 1994. Zoogeography and biodiversity of the freshwater fishes of Southeast Asia. *Hydrobiologia* 285, 41-48. doi: 10.1007/BF00005652

Supplementary material

Supplementary material 1: Table of specimens analysed.

Supplementary material 2: Table of models PartitionFinder.

Supplementary material 3: Table of biogeographic models

Supplementary material 4: Table of genetic distances

Supplementary material 5: Maximum clade credibility tree of all samples resulting from the BEAST analysis. Mean ages and the 95% highest posterior density distributions indicated.