Molecular phylogeny and biogeography of air sac catfishes of the *Heteropneustes*

***fossilis* species complex (Siluriformes: Heteropneustidae)**

Sahat Ratmuangkhwang^{1,2*}, Prachya Musikasinthorn³ and Yoshinori Kumazawa^{2**}

¹Andaman Coastal Research Station for Development, Kasetsart University Research and Development Institute, Kasetsart University, Suksamran, Ranong 85120, Thailand

²Department of Information and Biological Sciences and Research Center for Biological Diversity, Graduate School of Natural Sciences, Nagoya City University, Nagoya 467-8501, Japan

³Department of Fishery Biology, Faculty of Fishery, Kasetsart University, Bangkok 10900, Thailand (E-mail: ffispcm@ku.ac.th)

* Corresponding author. Email: rdishr@ku.ac.th Fax and Tel: (+66)-77-893-119

** Co-corresponding author. Email: kuma@nsc.nagoya-cu.ac.jp Fax and Tel: (+81)-52-872-5844

Keywords: Molecular phylogeny; Biogeography; *Heteropneustes fossilis*; air sac catfish; river capture

Abstract

The air sac catfish, *Heteropneustes fossilis* (Siluriformes: Heteropneustidae), is widely distributed in freshwaters of the Indian subcontinent and mainland southeast Asia. No comprehensive molecular studies that cover the broad distributional areas have been carried out to date. Here, we conducted molecular phylogenetic analyses using both mitochondrial and nuclear gene sequences to suggest that the *Heteropneustes fossilis* species complex consists of three clades that may potentially be separate species with distinct geographical distribution (southeast Asia, northeastern India, and southwestern India). The first and second clades are more closely related to each other than they are to the third clade. Within the first clade there is a basal divergence of a subclade consisting of individuals from Upper Irrawaddy River basin of Myanmar, which share some morphological traits with members of the Indian clades. Our molecular and morphological data are congruent with hypotheses that the Early-Middle Miocene disconnection between the paleo-Tsangpo River and the Irrawaddy River caused the vicariant divergence between southeast Asian and northeastern Indian clades, and that the southeast Asian *Heteropneustes* originated from the Upper Irrawaddy.

1. Introduction

The air sac catfish, *Heteropneustes* (the only extant genus member of Heteropneustidae in Siluriformes), is widely distributed in freshwater of mainland southeast Asia and the Indian subcontinent (Talwar and Jhingran, 1991; Rainboth, 1996; Rema Devi and Raghunathan, 1999). *Heteropneustes* is found mainly in ponds, ditches and swamps, but sometimes occurs in muddy rivers. The fish is characterized by an elongated body with very short dorsal and long anal fins, a greatly depressed head, four pairs of barbels, pectoral spines with venom gland, and a long air sac serving as an air breathing organ (Nelson, 2006). This catfish is an important food source as well as an occasional target for ecological and physiological studies due to its characteristic breathing ability. However, very little is known about evolution and biogeography of *Heteropneustes*.

Heteropneustes fossilis was first described by Bloch (1794) based on a specimen collected from Tranquebar (= Tharangambadi), Tamil Nadu of southern India. *H. fossilis* has several junior synonyms including *H. microps* (Günther, 1864) which was shown by Pethiyagoda and Bahir (1998) that the confluence of caudal and anal fins used to distinguish *H. microps* from *H. fossilis* (caudal and anal fins separated) resulted from

anomalous fin regeneration of damaged or abnormal hypurals. *H. kemratensis* (Fowler, 1937) was described based on specimens collected from Kemrat (= Khemarat)(the Mekong basin) of northeastern Thailand. The species was recognized as a synonym of *H. fossilis* by Smith (1945) and more recently followed by Menon (1999). Thus, the single widely distributed species *H. fossilis* has been conventionally used by many authors (e.g., Suvatti, 1950; Thiemmedh, 1968; Rainboth, 1996; Monkolprasit et al., 1997; Menon, 1999) until Kottelat (2001) reused both *H. fossilis* (for species distributed from Myanmar to Pakistan and Sri Lanka) and *H. kemratensis* (for species distributed in Thailand and Laos) without showing clear morphological evidence. However, although there is no study clearly diagnosing *H. kemratensis* from *H. fossilis* has been published so far, recently, *H. kemratensis* was recognized as valid by several workers (e.g., Ferraris, 2007; Kottelat, 2013; Hossain et al., 2013).

Two additional species of *Heteropneustes* were recently described from regions within a distributional range of *H. fossilis*. *H. longipectoralis* was described based on specimens collected from the Anamalai Hills in the western Ghats of India (Rema Devi and Raghunathan, 1999). This species can be clearly distinguished from *H. fossilis* by several characters including long pectoral fin, length 3/4th the distance between pectoral

to pelvic fin origins (vs. short to moderate pectoral fin, length 1/2-1/3th the distance between pectoral to pelvic fin origins), larger eyes (6 or less in head length [HL]) (vs. smaller, more than 6 times in HL), pointed caudal fin (vs. rounded) (Rema Devi and Raghunathan, 1999).

Very recently, *H. nani* was described based on specimens collected from Noakhall of Bangladesh (Hossain et al., 2013). Although taxonomic status of *H. nani* needs confirmation, this species was reported to have distinguishable morphological features from other members of *Heteropneustes* in, e.g., small adult size (ca. 98 mm SL) (vs. larger) and united anal and caudal fin (vs. separated). *H. longipectoralis* and *H. nani* were known only from a confined region of western Ghats of India and the Meghna deltaic region of Bangladesh, respectively. Taxonomic status of more widely distributed populations of *Heteropneustes fossilis* is presently uncertain and should be studied as a possibly species complex.

To the best of our knowledge, no comprehensive molecular phylogenetic studies have been carried out on the *Heteropneustes* fish although some genetic data using microsatellite and RAPD markers have been obtained for local Indian or Bangladeshi populations (Nasren et al., 2008; Garg et al., 2009). Agnese and Teugels (2005) and

Nakatani et al. (2011) provided strong molecular evidence that Heteropneustidae having the only genus *Heteropneustes* in it is more closely related to Clariidae than to any other catfish families. Thus, clariid species can be a reasonable outgroup in exploring phylogenetic relationships of *Heteropneustes* populations collected from diverse localities.

In this study, we used both mitochondrial and nuclear gene markers for phylogenetic inference. The mitochondrial NADH dehydrogenase subunit 2 (ND2) gene evolves with a relatively high rate and has been useful in discerning relationships among populations and closely related species (e.g., Duvernell and Turner, 1998; Kumazawa et al., 1999; Li et al., 2006) whereas the nuclear recombination activating gene 1 (RAG1) evolves at a slower rate and might be useful in detecting potential pitfalls of reticulate mitochondrial genealogy from hybridization or introgression among populations and closely related species (see, e.g., Mukai et al., 1997; Sullivan et al., 2006; Egger et al., 2007).

We use an obtained phylogenetic framework not only for gaining taxonomic implications but also for assessing historical biogeography of this fish. The wide distributional areas of *H. fossilis* species complex include landmasses of the different

geological origin (e.g., Indian subcontinent drifted from Gondwanaland) and major rivers clearly disconnected from each other at present but possibly connected in the past under different geological settings. Our molecular data should serve to evaluate hypotheses on how this fish achieved the current wide distribution.

2. Material and methods

2.1. Samples

Seventy three *Heteropneustes* specimens were collected from twelve localities in south and southeast Asia between 2004 and 2009 (Fig. 1 and table 1). We also collected one clariid catfish species, *Clarias batrachus*, as a sister taxon of Heteropneustidae (Diogo et al., 2003). Some of the heteropneustid specimens were obtained from local fish markets or fishermen who could provide details on the locality information. We used samples only when the locality information is reliable.

We collected *H. fossilis* specimens from most major river basins of its distribution (Fig. 1). Specimens from the Indus and Upper Ganges River basins, as well as from dry zone of Central India and Sri Lanka could not be included in this study due to some difficulties in field collection. Heteropneustids in Iran and Iraq are introduced ones (Coad,

1991) and we did not include them. A portion of the right pectoral fin for each specimen was cut and preserved either in 99% ethanol or TNESU8 buffer (Asahida et al., 1996) for molecular experiments. All specimens were then fixed in 10% formalin and preserved in 70% ethanol for morphological investigation. Specimens were deposited in the Research Laboratory of Ichthyology, Faculty of Fisheries, Kasetsart University (RLIKU) with registered numbers shown in table 1.

2.2. Extraction, amplification and sequencing of DNA

DNA was extracted from fin samples using a method described by Asahida et al. (1996). The complete ND2 gene and a partial region of RAG1 gene were amplified by the polymerase chain reaction (PCR) using primers shown in supplementary table S1 (see supplementary figure 1 for the location and orientation of each primer). PCRs were carried out with SpeedStar HS DNA polymerase (Takara) by 30 cycles of denaturation at 98°C for 5 sec, annealing at 55°C for 15 sec, and extension at 72°C for 20 sec.

The amplified products were treated with an ExoSAP-IT (GE) or purified with a Roche High Pure PCR Cleanup Kit and directly sequenced with the amplification primers and/or appropriate internal primers (supplementary table S1). The sequencing reaction

was done with either of the two kits: a DYEnamic ET Terminator Cycle Sequencing Kit (GE) or a BigDye Terminator v3.1 Cycle Sequencing Kit with manufacturers' instruction.

Products of the dye termination reactions were analyzed on Applied Biosystems 373A or 3500 DNA sequencers. A contiguous sequence was obtained by assembling these fragmentary sequences with Sequencher 4.8 (Gene Codes). Heterozygotic or heteroplasmic sites were not found in the determined sequences. All nucleotide sequences determined in this study were deposited to DDBJ/EMBL/GenBank nucleotide sequence databases with accession numbers shown in table 1. ND2 and RAG1 gene sequences were easily alignable by eye with only a few gaps. MacClade 4.08 (Sinauer Associates) was used to edit the DNA sequences.

The complete ND2 and partial RAG1 gene sequences were obtained for 73 and 47 *Heteropneustes* individuals, respectively (table 1). The corresponding gene sequences from the clariid outgroup taxa, *Clarias batrachus* was also obtained in this study (table 1). As other catfish outgroups, the gene sequences for an ictarulid, *Ictarulus punctatus* (AF482987 for ND2 and DQ492511 for RAG1 genes) and a cranoglanid, *Cranoglanis boudierius* (AY898626 for ND2 and DQ492514 for RAG1 genes) were also obtained from GenBank.

2.3. Phylogenetic analyses

Phylogenetic analyses were conducted using complete mitochondrial ND2 and partial nuclear RAG1 gene sequences by Bayesian method. Various analytical conditions were tested (e.g., use of different analytical methods such as the neighbor-joining method, use of different outgroup taxa, inclusion or exclusion of third codon positions of the ND2 genes, and modification of parameters in the Markov chain Monte Carlo (MCMC) process of the Bayesian analyses) to find that these variations did not affect the resultant tree topology with respect to phylogenetic relationships of major clades and subclades.

Bayesian analyses were conducted with MrBayes v3.12 (Ronquist and Huelsenbeck, 2003) using the best model from modeltest 3.0 (Posada and Crandall, 1998) and separating the data into three partitions (first, second, and third codon positions). The MCMC process was set to run four chains simultaneously for at least 4,000,000 generations and until the ASDSF (average standard deviation of split frequency) index becomes less than 0.01. After the initial one fourth of the generations were discarded based on the stationarity of the MCMC process at this stage, trees were collected from every 100 generations. The 50% majority-rule consensus tree was created from this tree pool and the Bayesian posterior probability (PP) values (the frequency of a specific

phylogenetic relationship in the sampled tree population) were obtained as a confidence measure for resultant phylogenetic relationships. In order to supplement the Bayesian PP values, bootstrap (BP) values by the neighbor-joining (NJ) analyses were also obtained. This was conducted by 1,000 BP replications with PAUP*4.0b10 (Swofford, 2003) using models and parameters selected by the modeltest 3.0 (Posada and Crandall, 1998).

Divergence times were estimated using the relaxed-clock Bayesian method with Beast v1.8.0 (Drummond et al., 2012). The Lognormal relaxed clock (Uncorrelated) model with no *a priori* correlation of evolutionary rates between a lineage's rate and that of its ancestor and the Yule tree prior assuming a constant speciation rate per lineage were used. The GTR substitution model coupled with 4 categories of the gamma distribution of site rates was chosen with base frequencies estimated from the data. Two independent Markov chains starting from a random tree were run for ten million generations and sampled every 1000 generations. Tracer v1.5 (Rambaut et al., 2014) was used to confirm more than 200 effective sample sizes for parameters and the convergence of two independent runs after the first 10% samples are removed as burnin. Sampled trees from two runs were then combined and node heights were shown with their median and 95% highest posterior density (HPD) interval values.

For the divergence time estimation, we added several more catfish taxa (the mochokid *Synodontis batesii*, the claroteid *Chrysichthys* sp., the pimelodid *Pimelodus ornatus*, the chacid *Chaca chaca*, the diplomystid *Diplomystes nahuelbutaensis*, and the astroblepid *Astroblepus* sp. 1) to increase nodes on which reliable time priors can be set based on catfish interfamilial relationships and divergence time estimates (Nakatani et al., 2011). The normal prior distribution of node ages was assumed at the representative nodes. The mean and standard deviation of the normal distribution was set as comparable to the mean and the 95% credibility ranges of molecular time estimates from Nakatani et al. (2011). For example, the divergence time between heteropneustids and clariids was estimated to be 63.6 MYA in mean (36.1-95.8 MYA in 95% credible interval) using the independent-rates Bayesian analyses of mitogenomic nucleotide sequences. We thus set an age prior of 63.6 MYA in mean and 15.0 million years in standard deviation at this node.

2.4. Morphological analysis

We selected four characters, two meristics (number of anal fin rays and number of total vertebrae) and two morphometrics (pectoral fin spine length in % of SL and caudal

peduncle depth in % of SL) for morphological investigation. Number of anal fin rays and pectoral fin spine length have been regarded as key taxonomic characters distinguishing species of *Heteropneustes* in the previous studies (e.g., Fowler, 1937; Rema Devi and Raghunathan, 1999)

We examined these morphological characters for specimens listed in table 1. Anal fin rays and vertebrae were counted from radiographs. Only anal fin rays bearing a separate pterygiophore for each were counted. Vertebrae counts included five vertebra comprising the Weberian complex and urostyle. Pectoral fin spine length was measured point to point from an anterior origin (base) to a tip. Caudal peduncle depth was measured point to point at the narrowest portion of caudal peduncle.

Statistical analyses for comparison of morphological data were performed with SPSS 12 for windows (SPSS Inc., Chicago). Student's t-test was conducted for meristic counts. Since there was a possibility that proportional characters increased or decreased with growth, analysis of covariance (ANCOVA), using SL as a covariate, was performed for morphometric measurements. When significance was recognized for slope, ANCOVA was terminated, but was otherwise subsequently performed for Y-intercept (McDonald, 2009).

3. Results

3.1. ND2 and RAG1 gene sequences

All ND2 gene sequences determined in this study (table1) and obtained from GenBank as outgroups were found not to include a full stop codon in the 3' end and addition of poly A to the ND2 mRNA may create a UAA stop codon. Alignment of the complete ND2 genes for 73 *Heteropneustes* individuals produced 48 unique haplotype sequences (table 2). This gene has 1,044 positions excluding the stop codon. Among them 233 (22.3%) were variable and 218 (20.9 %) were parsimony informative (table 2).

Alignment of the 1,466-bp partial RAG1 genes for 47 *Heteropneustes* individuals produced 10 unique haplotypes. Among the 1,466 sites, 56 (3.8%) were variable and 29 (2.0 %) were parsimony informative (table 2). We found some individuals collected from the same localities or closely related drainage systems to show identical haplotypes in both genes. Both variable and parsimony informative characters were the most abundant in the third codon positions for these genes, whereas the two characters were the fewest in the second codon positions (table 2). Reflecting the general trend that mitochondrial genes evolve much faster than nuclear genes, the ND2 genes appear to be much more variable than the RAG1 genes.

3.2. *Phylogenetic analysis*

Figure 2 shows a Bayesian tree constructed using the combined ND2 and RAG1 gene sequences. The heteropneustid individuals were divided into three major clades, A, B, and C with very high confidence values (1.00 for Bayesian PP and 100% for NJ BP). Clades A and B are more closely related to each other than they are to clade C with 1.00 Bayesian PP and 95% NJ BP supports.

Clade A was composed of all individuals from southeast Asia and could be divided into three subclades (A1, A2 and A3) with strong support values (0.97 or higher PP and 92% or higher BP). A1 contains individuals from the Tapi (the locality code Sur; see table 1 and Fig. 1), Chao Phraya (Ayu) and Bang Pakong (Pra) River basins in Thailand. A2 contains individuals from the Sittoung (Bag) and lower Irrawaddy River basins (Rak and Yan) in Myanmar. A3 contains individuals from the Upper Irrawaddy River basins (Put and Myi) in Myanmar. Subclades A1 and A2 are more closely related to each other than they are to subclade A3 (0.99 PP and 77% BP). Clade B was composed of all individuals from northeastern parts of India (Ass and Kol). Clade C was a sister group of Clade A+B composed of all individuals from southwestern parts of India (Kot and Kan).

The Bayesian trees were also constructed using the individual ND2 and RAG1 gene sequences (supplementary figures 2-3). Phylogenetic relationships between major clades and subclades of these trees are in agreement with those in Fig. 2 although minor differences existed mostly in relationships within the major clades and subclades.

3.3. Dating analysis

Figure 3 illustrates a result of the uncorrelated relaxed-clock Bayesian estimation of divergence times using the RAG1 gene sequences. The posterior distribution of ages at 6 nodes with age priors was mostly similar to the prior distribution. Although the posterior mean at a node separating Heteropneustidae and Clariidae (73.9 MYA) was somewhat larger than the prior mean (63.6 MYA), both values are consistent with the fossil-based estimate for this divergence (>50 MYA; Gayet, 1987; Agnese and Teugels, 2005). Divergence times between heteropneustid lineages were estimated to be 39.8 MYA in median (23.1-60.9 MYA in 95% HPD interval) between clades A+B and C, and 26.0 MYA (12.9-43.0 MYA) between clades A and B (Fig. 3). Divergence times between subclades of clade A were estimated to be 4-6 MYA in mean, suggesting that radiation of

southeast Asian heteropneustid lineages occurred relatively recently in Plio-Pleistocene time.

Nakatani et al. (2011) provided somewhat younger time estimates using the autocorrelated-rates Bayesian analyses of mitogenomic nucleotide sequences. We also conducted the dating analysis under this assumption of node age priors. The divergence times between Heteropneustidae and Clariidae, between clades A+B and C, and between clades A and B were, respectively, estimated to be 60.4 MYA (44.4-77.1 MYA), 32.8 MYA (19.5-48.6 MYA), and 21.7 MYA (11.4-34.6 MYA) (data not shown). These values are slightly smaller than those of Fig. 3 but the posterior mean differences between the two prior assumptions are not large as compared to the credibility intervals.

We also conducted the dating analyses using 1st and 2nd codon positions of ND2 gene sequences under similar assumptions. The ND2-based independent-rates Bayesian analyses estimated divergence times between clades A+B and C and between clades A and B to be, respectively, 59.3 MYA (41.9-77.7 MYA) and 49.2 MYA (33.9-66.8 MYA) (data not shown), which were somewhat larger than the corresponding RAG1-based estimates (Fig. 3) in general. Since our age priors are not set at any ingroup (i.e., heteropneustid) nodes, such situation may cause consistently overestimated divergence

times using mitochondrial gene sequences (e.g., Zheng et al., 2011). We therefore place more emphasis on the RAG1-based time estimates (Fig. 3) for later biogeographic discussion.

3.4. Morphological analysis

From four examined characters, number of anal fin rays can separate all examined specimens into two distinct groups without overlapping (supplementary table S2). Individuals of one group, which corresponds to clades A3, B and C (see Fig. 2 and supplementary figures 2-3) had 59-70, 62-74 and 63-74 anal fin rays, respectively, whereas those of another group corresponding to clades A1 and A2 have 75-85 and 75-84 anal fin rays, respectively. Furthermore, most of the rest of examined morphological characters also generally corresponded with the grouping resulted from the number of anal fin rays although partly overlapped (supplementary tables S3 and S4). Statistical tests of all of the four characters clearly showed significant difference ($p < 0.001$) between the two groups (table 3). Figure 4 visually shows the morphological separation of the two groups at the clade/subclade level by two-dimensional plots of the locality-based meristic characters shown in supplementary tables S2 and S3.

We tentatively name these morphological features “*H. fossilis*”-like (green color in Fig. 2 and supplementary figures 2-3) and “*H. kemratensis*”-like (red color) traits, respectively, based on the corresponding morphological trait and geographic distribution of originally described specimens (Bloch, 1794; Fowler, 1937). Note that these morphological traits of specimens used in this study are comparable and morphologically and geographically agree with those from the literature which described *Heteropneustes* from various places of India and mainland southeast Asia including original descriptions of nominal species of *Heteropneustes* (e.g., Bloch, 1794; Hamilton, 1822; Günther, 1864; Day, 1865; Fowler, 1937; Datta Munshi and Srivastava, 1988; Rema Devi and Ragunathan, 1999; Rahman, 2005; Saha, 2007).

4. Discussion

4.1. Taxonomic implications of recognized major clades

In order to gain taxonomic implications, we assessed the depth of genetic divergences between the clades with reference to those between congeneric sister species in other fish. Average pairwise distances of the ND2 and RAG1 gene sequences between clades and subclades are shown in supplementary table 5. Average transversional p-

distances of the ND2 gene sequences range from 0.029 to 0.046 between clades A-C.

These values are larger than or nearly equal to those observed between congeneric sister species of different fish groups (e.g., 0.034 between *Channa gachua* and *Channa orientalis* and 0.029 between *Channa lucius* and *Channa bankanensis* in Channidae; sequence data taken from Li et al., 2006).

p-distances of the RAG1 gene sequences range from 0.018 to 0.028 between clades A-C (supplementary table 5). Again, these values are larger than the RAG1 distances between sister species of other fishes (e.g., 0.003 between *Badis corycaeus* and *Badis pyema* and 0.005 between *Badis ruber* and *Badis khwae* in Badidae; sequence data taken from Rüber et al., 2004). Thus, the clades A-C are separated by large molecular distances almost equivalent to between separate species in other fish groups. Note that our argument for at least three separate species in the *H. fossilis* species complex is not conclusive because there is no direct relationship between genetic divergence and taxonomic status. In future, thorough morphological investigation of additional heteropneustid specimens should be carried out for formal taxonomic revision of Heteropneustidae.

4.2. Morphology vs. molecules

We found two morphological traits that can separate this species complex into two groups (see section 3.4) but the boundary of these traits did not exactly correspond with clades A-C (see green and red colors in Fig. 2 and supplementary figures 2-3). Upon this apparent discrepancy we reinvestigated morphological characters of Indian specimens. We have not so far found any morphological trait that can clearly separate northeast Indian (clade B) and southwest Indian (clade C) populations although clade C individuals may have somewhat distinct characteristics in, e.g., number of gill raker, length of air sac and color pattern on the body (data not shown).

Our molecular phylogeny (Fig. 2 and supplementary figures 2-3) suggests that the “*H. fossilis*”-like (green) traits are ancestral and that the derived “*H. kemratensis*”-like (red) traits emerged more recently in southeast Asian populations. It was unexpected that all specimens from the Upper Irrawaddy River basin in Myanmar (i.e., subclade A3) share the “*H. fossilis*”-like traits but that they are nested in clade A on the basis of strong molecular evidence. Inclusion of the Upper Irrawaddy basin population in the “*H.*

fossilis”-like group is also clearly observed from a graph showing combinations of the two meristic characters (Fig. 4).

We first noticed this based on the ND2 gene tree (supplementary figure 2) and suspected the cross-species hybridization or introgression of mitochondrial DNA into a different species/population. However, the nuclear RAG1 gene tree also supported that the Upper Irrawaddy individuals are nested within clade A (supplementary figure 3). Thus, this phylogenetic relationship was supported by both mitochondrial and nuclear genes although we cannot strictly deny the possibility that other nuclear genes show a different phylogeny.

Then, a question arises regarding why only the Upper Irrawaddy populations retain the “*H. fossilis*”-like traits while they are clearly nested in clade A together with other Southeast Asian populations having the “*H. kemratensis*”-like traits. Our explanation for this is that present Southeast Asian (clade A) populations originated from geographically confined Upper Irrawaddy populations and that morphological transition from the green to red traits took place when ancestors of subclades A1 and A2 diverged from subclade A3 and migrated southward toward their current distribution (see section 4.4 for related discussion). An alternative explanation is that the green traits are under

selection in the Upper Irrawaddy populations. However, four morphological traits examined in this study are among popular ones for fish taxonomy and we do not know any evidence that these traits are under strong selection for particular habitats.

4.3. Origin of Heteropneustidae

Clariidae, which is distributed in both African and Asian freshwaters, was considered as a sister family of Heteropneustidae by several authors (e.g., Diogo et al., 2003; Agnese and Teugels, 2005; Sullivan et al., 2006) (but see Ferraris (2007) who included *Heteropneustes* in the family Clariidae). A recent comprehensive molecular study using complete mitochondrial DNA sequences (Nakatani et al., 2011) confirmed their sister-group relationship very confidently.

Agnese and Teugels (2005) considered that Clariidae diverged from Heteropneustidae in Eurasia around 50 MYA and that some clariid lineages migrated to Africa in Miocene after Africa collided with Eurasia approximately 20 MYA. The divergence time estimation based on complete mitochondrial DNA sequences supported this view by showing that these families diverged about 50-65 MYA (Nakatani et al., 2011). Paleontological data are also congruent with this hypothesis. The oldest known

fossil of Clariidae is from the Lower-Middle Eocene (45–50 MYA) of Pakistan (Gayet, 1987) while the oldest African clariids are from Middle Miocene (10–15 MYA) (Van Couvering, 1977).

To our best, no fossil record from Heteropneustidae has been known and extant heteropneustids are distributed only in Asia (Nelson, 2006). The present study showed that the most basal divergence within the *H. fossilis* species complex occurred between southwest Indian and northeast Indian+southeast Asian taxa in the Eocene-Oligocene (Fig. 3). Taken together with the biogeographic discussion on Clariidae (see above), we assume that Heteropneustidae diverged from Clariidae in drifting Indian subcontinent or in a neighboring Laurasian part before the collision of India to Eurasia, and that the most recent common ancestor of clades A-C occurred in south Asia (Indian subcontinent) in the Eocene-Oligocene after the collision. Geological evidence suggests that the Indian collision to Eurasia had been completed by Early Eocene (e.g., Smith et al., 1994).

We do not think that the most recent common ancestor of clades A-C had such a wide distribution that is equivalent to the current distribution of the *H. fossilis* species complex. Primary freshwater fishes like *H. fossilis* do not usually migrate between major rivers through estuaries and this restricts a continuous gene flow over wide geographical

ranges. *Heteropneustes* fishes could have occurred widely at that time. However, we consider that the direct ancestors of clades A-C probably occurred in a more confined region.

Although our molecular phylogeny does not include *H. longipectoralis* and *H. nani*, they are distributed in a small highland area of southwest India and the Meghna deltaic region of Bangladesh, respectively. Thus, the distribution of *H. longipectoralis* and *H. nani* is consistent with our assumption of the south Asian origin of Heteropneustidae.

4.4. Historical biogeography

We discuss processes of the divergence and migration of heteropneustids based on our molecular phylogenetic evidence and geological history of south and southeast Asia, especially in relation to the history of freshwater river systems. Although some extant catfishes (e.g., Ariidae and Plotosidae) are marine fish and some other families (e.g., Aspredinidae and Bagridae) contain taxa that can tolerate saltwater to some extent (Nelson, 2006), heteropneustid catfishes are highly adapted to the freshwater environment (Hora, 1939). It is thus very unlikely that ancestral heteropneustids migrated

between freshwater river systems via marine or brackish-water dispersal. Heteropneustids and clariids are known to have an accessory air-breathing organ, as well as ability of moving on land, and thus considered to have the capability of traveling short distances on land (Graham, 1997; Nelson, 2006; personal observation [SR and PM]).

The collision of the Indian subcontinent to Eurasia by the Early Eocene (Smith et al., 1994) formed the orogenic belt that created the Tibetan Plateau, the Himalaya Mountains and other highlands, which caused significant changes in climate and drainage systems in south and southeast Asia (e.g., Brookfield, 1998). The northward movement of the Indian plate and the resulting uplift of the Himalayas and the Western Ghats may have contributed to the onset of a drier climate and replacement of the tropical evergreen forests by deciduous ones over the majority of India (Meher-Homji, 1977). The currently disjunct distribution of some animals and plants may represent a relict of a former continuous distribution (Mani, 1974; Karanth, 2003).

In the previous section 4.3, we deduced that the most recent common ancestor of clades A-C occurred in south Asia (Indian subcontinent) in the Eocene-Oligocene. We consider that the divergence of clade C from clades A+B may have been vicariantly associated with the uplift of the Western Ghats and the fragmentation of humid climate.

In the present Indian subcontinent, the Western Ghats separates lowland humid regions of Kerala, Karnataka and Tamil Nadu from relatively arid Deccan Plateau. Some reports (Harris, 2006; Van Bocxlaer et al., 2009) suggest that the climatic changes in these areas started in the Late Oligocene to Early Miocene. This timing seems somewhat more recent than the median divergence time estimates between clades A+B and C (~40 MYA or ~33 MYA depending on the prior assumptions; see section 3.3). However, the vicariant lineage divergence might have preceded the climatic changes to some extent.

The paleo-Tsangpo River of northeastern India ran into the Irrawaddy River of Myanmar by the beginning of Miocene. However, the Early-Middle Miocene uplifts of the eastern Tibet and the Indoburman mountain range caused a river capture which changed the connection of the Tsangpo River to the Brahmaputra River of east India but not to the Irrawaddy River (Fig. 5; Bannert and Helmcke, 1981; Brookfield, 1998; Clark et al., 2004).

We think that the direct ancestor of clades A and B may have been distributed from the paleo-Tsangpo River to the Irrawaddy River. Upon the river capture, populations of these rivers would have been isolated vicariantly (Fig. 5). The estimated divergence times between clades A and B (~26 MYA or ~22 MYA; see section 3.3) are

in agreement with the timing of the Indoburman uplift from the geological evidence (~20 MYA; Bannert and Helmcke, 1981). This view is also consistent with our molecular phylogeny in which the Upper Irrawaddy populations (subclade A3) are genetically close to southeast Asian populations (subclades A1 and A2) but share some morphological features with Indian populations (Fig. 2 and supplementary figures 2-3). The Upper Irrawaddy seems to have been a key source place for the evolution of southeast Asian populations.

The vicariant divergence around the Oligocene-Miocene boundary (19-24 MYA) by the uplift of the Indoburman mountain range was also proposed for baduids (e.g., *Dario dario* of Brahmaputra, Assam, India vs. *Dario hyssginon* of Upper Irrawaddy, Kachin State, Myanmar), anabantoids (e.g., *Colisa* spp. of India vs. *Colisa labiosa* of Myanmar and *Ctenops* of India vs. *Parasphaerichthys* of Myanmar) and cyprinids (e.g., *Danionella priapus* of Brahmaputra, Assam, India vs. its sister group, *D. mirifica* and *D. translucida* of Irrawaddy and Sittang drainages in Myanmar) (Rüber et al., 2004; Britz, 2009). A vicariance between the formerly connected Tsangpo-Brahmaputra and Upper

Irrawaddy was also proposed for glyptosternoid catfishes (Guo et al., 2005) and cyprinids *Schizothorax* (He and Chen, 2006) although the vicariant event for the latter may have occurred somewhat more recently in the Late Miocene (He and Chen, 2006).

Assuming the above-mentioned vicariant divergence between clades A and B, we hypothesize that southeast Asian heteropneustid lineages migrated from the Upper Irrawaddy ancestors relatively recently in Plio-Pleistocene time (Figs. 3 and 5). It seems noteworthy that there appears to be no continuous gene flow between Lower and Upper Irrawaddy populations because they are both morphologically and genetically distinct from each other (see subclades A2 and A3 in Fig. 2 and supplementary figures 2-3). The central lowland of Myanmar was considered to be a distinct climatic zone as the Central Dry Zone (Bender, 1983; Hutchison, 1989). This may set an unknown barrier to hamper natural migrations of heteropneustids between Upper and Lower Irrawaddy River basins.

In conclusion, the present study revealed phylogenetic relationships of the *Heteropneustes fossilis* species complex and suggested their possible relevance to south and southeast Asian paleo-drainages. We found the existence of at least three distinct lineages

in this species complex, which may possibly be recognized as separate species in future. We do not deny the possibility that further sampling of uncollected localities (see section 2.1) may reveal additional *Heteropneustes* lineages. Studies on the historical biogeography of additional freshwater fish that have a similarly wide distribution in south and southeast Asia should also be encouraged to reveal relationships between river captures and freshwater fish evolution more fully.

Acknowledgments

We thank Drs. M. M. Goswami (Gauhati University, Assam, India), G. D. Biji and Messrs. U. Ohn (Forest Resource Environment Development and Conservation Association [FREDA], Myanmar), N. N. Lwin (Green Leaf Travels & Tours, Yangon, Myanmar), M. Bora, R. Pachuau (Assam, India), A. A. Rao (Kolkata, India), S. Thongplu and S. Arbsuwan, Mses. N. Kalsri and S. Laksana for their kind help in collecting fish samples. We also thank Drs. Wirangrong Karinthanyakit, Sébastien Lavoué, and Jun G. Inoue for their kind help in operating softwares for divergence time estimation. Our

gratitude is extended to Drs. M. Nakatani and M. Miya for sending a manuscript and DNA sequences before publication, and to Dr. G. Ortí and two anonymous reviewers for helpful comments. This work was supported by grants from the Ministry of Education, Culture, Sports, and Technology of Japan (grant No. 20405012), partly by grants from the Japan Science and Technology Agency as a part of the CREST Project (Water Use and Management System of the Mekong River) to P.M. and Nagoya City University (the Visiting Scholar from Abroad 2011).

References

- Agnese, J.-F., Teugels, G.G., 2005. Insight into the phylogeny of African Clariidae (Teleostei, Siluriformes): implications for their body shape evolution, biogeography, and taxonomy. *Mol. Phylogenet. Evol.* 36, 546–553.
- Asahida, T., Kobayashi, T., Saitoh, K., and Nakayama, I., 1996. Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffers containing high concentration of urea. *Fish. Sci.* 62, 727-730.
- Bannert, D., Helmcke, D., 1981. The evolution of the Asian plate in Burma. *Geologische Rundschau.* 70, 446–458.
- Bender, F., 1983. *Geology of Burma*. Gebrüder Borntraeger. Berlin, Germany.
- Bloch, M.E., 1794. *Naturgeschichte de ausländischen Fische. Siebenter Theil*. Morino, Berlin, Germany.
- Britz, R., 2009. *Danionella priapus*, a new species of miniature cyprinid fish from West Bengal, India (Teleostei: Cypriniformes: Cyprinidae). *Zootaxa.* 2277, 53-60.
- Brookfield, M.E., 1998. The evolution of the great river systems of southern Asia during the Cenozoic India–Asia collision: rivers draining southwards. *Geomorphology.* 22, 285–312.

Clark, M.K., Schoenbohm, L.M., Royden, L.H., Whipple, K.X., Burchfiel, B.C., Zhang,

X., Tang, W., Wang, E., Chen, L., 2004. Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. *Tectonics*. 23, TC1006.

Coad, B.W., 1991. Fishes of the Tigris-Euphrates basin: A critical check-list (Syllogeus Series No. 68). Canadian Museum of Nature, Ottawa, Canada.

Day, F., 1865. The fishes of Malabar. London. i-xxxii + 1-293, 20 pls.

Datta Munshi, J.S., Srivastava, M.P., 1988. Natural history of fishes and systematics of freshwater fishes of India. Narendra Publ. House, Delhi. i-xviii + 1-403.

Diogo, R., Chardon, M., Vandewalle, P., 2003. On the osteology and myology of the cephalic region and pectoral girdle of *Heteropneustes fossilis* (Siluriformes: Heteropneustidae), with comments on the phylogenetic relationships between *Heteropneustes* and the clariid catfishes. *Animal Biol.* 53, 379-396.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969-1973.

Duvernell, D.D., Turner, B.J., 1998. Evolutionary genetics of Death Valley pupfish populations: mitochondrial DNA sequence variation and population structure. *Mol. Ecol.* 7, 279-288.

Egger, B., Koblmüller, S., Sturmbauer, C., Sefc, K.M., 2007. Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus

Tropheus. BMC Evol. Biol. 7, 137.

Ferraris, C.J.Jr., 2007. Check list of catfishes, recent and fossil

(Osteichthyes:

Siluriformes), and catalogue of siluriform primary types.

Zootaxa. 1418, 1–628.

Fowler, H.W., 1937. Zoological result of the third De Schauensee Siamese Expedition.

Part VIII,--Fishes obtained in 1936. Proc. Acad. Nat. Sci. Phila. 89, 125-264.

Garg, R.K., Sairkar, P., Silawat, N., Vijay, N., Batav, N., Mehrotra, N.N., 2009. Genetic diversity between two populations of *Heteropneustes fossilis* (Bloch) using RAPD profile. Int. J. Zool. Res. 5, 171-177.

Gayet, M., 1987. Lower vertebrates from the Early-Middle Eocene Kuldana Formation of

Kohat (Pakistan): Holostei and Teleostei. Contr. Mus. Paleontol. Univ. Michigan

27, 151–193.

Graham, J.B., 1997. Air breathing fishes: evolution, diversity, and adaptation. Academic

Press, San Diego, USA.

- Günther, A., 1864. Catalogue of the fishes in the British Museum. Catalogue of the Physostomi, containing the families Siluridae, Characinidae, Haplochitonidae, Sternoptychidae, Scopelidae, Stomiatidae in the collection of the British Museum. Cat. Fishes V. 5.
- Guo, X., He, S., Zhang, Y., 2005. Phylogeny and biogeography of Chinese sisorid catfishes re-examined using mitochondrial cytochrome b and 16S rRNA gene sequences. Mol. Phylogenet. Evol. 35, 344–362.
- Hamilton, B.F., 1822. An account of the fishes found in the river Ganges and its branches. Edinburgh and London.
- Harris, N., 2006. The elevation history of the Tibetan Plateau and its implications for the Asian monsoon. Palaeogeogr. Palaeoclimatol. Palaeoecol. 241, 4-15.
- He, D., Chen, Y., 2006. Biogeography and molecular phylogeny of the genus *Schizothorax* (Teleostei: Cyprinidae) in China inferred from cytochrome *b* sequences. J. Biogeogr. 33, 1448-1460.
- Hora, S.L., 1939. Physiology of respiration and evolution of air-breathing fishes. Proc. Nat. Inst. Sci. India. 5, 281-287.

- Hossain, M.S., Sarker, S., Sharifuzzaman, S.M., Chowdhury, S.R., 2013. New species of stinging catfish *Heteropneustes nani* (Siluriformes: Heteropneustidae) from Noakhali, Bangladesh. Vert. Zool. 63, 259-267.
- Hutchison, C.S., 1989. Geological evolution of South-East Asia. Clarendon Press, Oxford, UK.
- Jayaram, K.C., 2006. Catfishes of India. Narendra Publishing House, Delhi, India.
- Karanth, K.P., 2003. Evolution of disjunct distributions among wet-zone species of the Indian subcontinent: Testing various hypotheses using a phylogenetic approach. Curr. Sci. 85, 1276-1283.
- Kottelat, M., 2001. Fishes of Laos. WHT Publications, Colombo, Sri Lanka.
- Kottelat, M., 2013. The fishes of the inland waters of Southeast Asia: a catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. The Raffles Bulletin of Zoology. 27, 1-663.
- Kumazawa, Y., Yamaguchi, M., Nishida, M., 1999. Mitochondrial molecular clocks and the origin of euteleostean biodiversity: familial radiation of perciforms may have

- predated the Cretaceous/Tertiary boundary. In: Kato, M. (Ed.) The biology of biodiversity. Springer-Verlag, Tokyo, pp. 35-52.
- Li, X., Musikasinthorn, P., Kumazawa, Y., 2006. Molecular phylogenetic analyses of snakeheads (Perciformes: Channidae) using mitochondrial DNA sequences. Ichthyol. Res. 53, 148-159.
- Mani, M.S., 1974. Biogeographical evolution in India. In: Mani, M.S. (Ed.), Ecology and biogeography in India. The Hague, Dr. W. Junk B. V. Publishers, Netherlands, pp. 698-724.
- McDonald, J.H., 2009. Handbook of Biological Statistics, 2nd ed. Sparky House Publishing, Baltimore, Maryland.
- Meher-Homji, V.M., 1977. Tropical dry deciduous forests of Peninsular India. Feddes Repert. 88, 113–134.
- Menon, A.G.K., 1999. Check list - Freshwater fishes of India. Records of the Zoological Survey of India. Miscellaneous Publication, Occasional Paper No. 175: i-xxviii + 1-366.
- Monkolprasit, S., Sontirat, S., Vimollohakarn, S., Sonsirikul, T., 1997. Checklist of fishes in Thailand. Office of Environment Policy and Planning, Bangkok, Thailand.

- Mukai, T., Naruse, K., Sato, T., Shima, A., Morisawa, M., 1997. Multiregional introgressions inferred from the mitochondrial DNA phylogeny of a hybridizing species complex of gobiid fishes, genus *Tridentiger*. Mol. Biol. Evol. 14, 1258-1265.
- Nakatani, M., Miya, M., Mabuchi, K., Saitoh, K., Nishida, M., 2011. Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. BMC Evol. Biol. 11, 177.
- Nasren, S., Mohammad, N.I., Mohd, G.O., Khan, Shahidul, I., Samsul, A., 2008. Genetic variation and differentiation of the stinging catfish, *Heteropneustes fossilis* (Bloch), populations assessed by heterologous microsatellite DNA markers. Indian J. Biotechnol. 8, 85-90.
- Nelson, J.S., 2006. Fishes of the World. 4th ed., John Wiley & Sons, New York, USA.
- Pethiyagoda, R., Bahir, M.B., 1998. *Heteropneustes microps*, a junior synonym of *H. fossilis* (Osteichthys: Heteropneustidae). J. South Asian Nat. Hist. 3, 113-114.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14, 817-818.
- Rahman, A.K.A., 2005. Freshwater fishes of Bangladesh. Second Edition. Zoological

Society of Bangladesh, Dhaka. i-xviii + 1-394.

Rainboth, W. J., 1996. Fishes of the Cambodian Mekong. FAO species identification field guide for fishery purposes. United Nations Food and Agriculture Organization, Rome, Italy.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.5, Available from <http://beast.bio.ed.ac.uk/Tracer>. Accessed on March 8, 2014.

Rema Devi, K., Raghunathan, M.B., 1999. *Heteropneustes longiptoralis* (Siluriformes: Heteropneustidae) a new species from the Anamalai Hills, in the Western Ghats. Rec. Zool. Surv. India. 97, 109-115.

Ronquist, F., Huelsenbeck, J.P., 2003. Mrbayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19, 1572-1574.

Rüber, L., Britz, R., Kullander, S.O., Zardoya, R., 2004. Evolutionary and biogeographic patterns of the Badidae (Teleostei: Perciformes) inferred from mitochondrial and nuclear DNA sequence data. Mol. Phylogenet. Evol. 32, 1010-1022.

Saha, B.K., 2007. Family Heteropneustidae. In: Siddiqui, K.U., Islam, M.A., Kabir, S.M.H., Ahmad, M., Ahmad, A.T.A., Rahman, A.K.A., Haque, E.U., Ahmad, Z.U., Begum, Z.N.T., Hassan, M.A., Khondker, M., and Rahman, M.M. (eds.)

- Encyclopedia of Flora and Fauna of Bangladesh. Vol. 23. Freshwater Fishes.
Asiatic Society of Bangladesh, Dhaka. Pp. 159-160.
- Smith, H.M., 1945. The fresh-water fishes of Siam, or Thailand. Bull. U. S. Nat. Mus.
188, 633pp.
- Smith, A.G., Smith, D.G., Funnell, B.M., 1994. Atlas of Mesozoic and Cenozoic
coastlines. Cambridge University Press, New York.
- Sullivan, J., Lundberg, J., Hardman, M., 2006. A phylogenetic analysis of the major
groups of catfishes (Teleostei: Siluriformes) using *rag1* and *rag2* nuclear gene
sequences. Mol. Phylogenet. Evol. 41, 636–662.
- Suvatti, C., 1950. Fauna of Thailand. Department of Fisheries, Kasetsart University,
Bangkok, Thailand.
- Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other
Methods), version 4. Sinauer Associates, Sunderland, MA, USA.
- Talwar, P.K., Jhingran, A.G., 1991. Inland fishes of India and adjacent countries. Oxford
& IBH, New Delhi, India.
- Thiemmedh, J., 1968. Fishes of Thailand: their English, scientific and Thai names.
Kasetsart University Fisheries Research Bull., Bangkok, Thailand.

Van Bocxlaer, I., Biju, S.D., Loader, S.P., Bossuyt, F., 2009. Toad radiation reveals into

India dispersal as a source of endemism in the Western Ghats-Sri Lanka

biodiversity hotspot. *BMC Evol. Biol.* 9, 131.

Van Couvering, J.A.H., 1977. Early records of freshwater fishes in Africa. *Copeia* 1977,

163–166.

Zheng, Y., Peng, R., Kuro-o, M., Zeng, X., 2011. Exploring patterns and extent of bias in

estimating divergence time from mitochondrial DNA sequence data in a particular

lineage: A case study of salamanders (Order Caudata). *Mol. Biol. Evol.* 28, 2521-

2535.

Legends to figures

Fig. 1. Map of India, Bangladesh, Myanmar, and adjacent regions showing the sampling localities, with a sketch of *Heteropneustes fossilis* in the upper right (fin rays omitted).

Abbreviations of the localities are listed in table 1. A thick dotted line represents the Indoburman mountain ranges. Major rivers are highlighted; In, Indus; G, Ganges; T, Tsangpo; B, Brahmaputra; I, Irrawaddy; S, Salween; C, Chaophraya; and M, Mekong. Colors and symbols of localities mean that individuals from these localities have DNA sequences that are nested in the corresponding clades and subclades of molecular phylogeny (refer to Fig. 2 and supplementary figures 2-3).

Fig. 2. A Bayesian tree constructed using all codon positions of combined ND2 and RAG1 gene sequences. Individuals sharing identical sequences are grouped together into a single haplotype for the tree building (see their names at external nodes). The GTR+I+G model selected by the modeltest was used together with its optimized parameters. Numbers on internal nodes indicate NJ BP (shown only when they are more than 50%) and Bayesian PP values at the left and right of a slash, respectively. Colors for branches represent the most parsimonious reconstruction of morphological features that

correspond to “*H. fossilis*”-like (green) and “*H. kemratensis*”-like (red) traits (see text).

Colors and symbols at the right side of the tree correspond to the locality information shown in Fig. 1.

Fig. 3. Divergence times between heteropneustid and other catfish lineages estimated using RAG1 gene sequences. The uncorrelated relaxed-clock Bayesian estimation of divergence times was conducted using Beast v1.8.0 as described in section 2.3. This figure is drawn based on estimated nodal divergence times corresponding to the median value from posterior distributions. Light blue squares spanning nodes show 95% HPD intervals. The mean and standard deviation (in parenthesis) of the normal distribution of several nodal ages as priors are also shown. GenBank accession numbers of RAG1 gene sequences for additional outgroup taxa are: DQ492495 (*Synodontis batesii*), DQ492505 (*Chrysichthys* sp.), DQ492475 (*Pimelodus ornatus*), DQ492469 (*Chaca chaca*), DQ492429 (*Diplomystes nahuelbutaensis*), and DQ492438 (*Astroblepus* sp. 1).

Fig. 4. Two-dimensional plots of numbers of anal fin rays and total vertebrae of *Heteropneustes* individuals examined in this study. Meristic data from supplementary

tables S2 and S3 are plotted to visualize the morphological separation between individuals of subclades A1-A2 and the rest.

Fig. 5. A hypothetical scheme for the relationship between the Tsangpo-Brahmaputra river capture and the heteropneustid divergence. (A) Before the river capture (before the Early-Middle Miocene), the paleo-Tsangpo River flew into the Irrawaddy River. (B) After the river capture (after the Early-Middle Miocene), the Tsangpo River connected to the Brahmaputra River as in the present drainage system. The reconstruction of past river systems is based on Clark et al. (2004). Major rivers are highlighted; T, Tsangpo; B, Brahmaputra; I, Irrawaddy; S, Salween; M, Mekong; R, Red and Y, Yangtze. Arrows indicate the presumed direction of heteropneustid migrations (see text)

Table 1. Specimens of the *Heteropneustes fossilis* species complex and an outgroup used for sequencing and morphological examination in this study

Locality code ^a	Locality details	River or lake basin	No. of specimens ^c	Voucher No. ^d	Accession No. (ND2)	Accession No. (RAG1)
Kot	Kottayam, Kerala, India	Vembanad ^b	11(4)	RLIKU 998.1–6,8–12 (2–5)	AB673255 – 265	AB682946 – 949
Kan	Kanyakumari, Tamil Nadu, India	Tambaraparani	4(4)	RLIKU 1960.1–4 (1–4)	AB673266 – 269	AB682950 – 953
Ass	Assam, Guwahati , India	Brahmaputra ^b	8(6)	RLIKU 966.1–8 (1–4,7–8)	AB673270 – 277	AB682954 – 959
Kol	Kolkata, W. Bengal, India	Ganges-Brahmaputra ^b	5(4)	RLIKU 1897.1–5 (1–4)	AB673278 – 282	AB682960 – 963
Myi	Myitkyina T/S, Kachin, Myanmar	Upper Irrawaddy	14(6)	RLIKU 1898.1, RLIKU 1899.2–6,8–10 (2,4,6,8), RLIKU 1900.11–13, RLIKU 1901.15–16 (15–16)	AB673283 – 296	AB682964 – 969
Put	Putao, Kachin, Myanmar	Upper Irrawaddy ^b	5(4)	RLIKU 1902.1–5 (1–2,4–5)	AB673297 – 301	AB682970 – 973
Bag	Bago Division, Myanmar	Sittoung	5(3)	RLIKU 1903.1–5 (2–3,5)	AB673302 – 306	AB682974 – 976
Rak	Rakhine State, Myanmar	Thandwe ^b	5(2)	RLIKU 1904.1–5 (2–3)	AB673307 – 311	AB682977 – 978
Yan	Yangon Division, Myanmar	Lower Irrawaddy	5(4)	RLIKU 1905.1–5 (1–2,3,5)	AB673312 – 316	AB682979 – 982
Ayu	Ayutthaya, Thailand	Chao Phraya	5(5)	RLIKU 989.1–5 (1–5)	AB673317 – 321	AB682983 – 987
Pra	Prachinburi, Thailand	Bang Pakong	2(2)	RLIKU 983.1–2 (1–2)	AB673322 – 323	AB682988 – 989
Sur	Suratthani , Thailand	Tapi	4(3)	RLIKU 974.1–4 (1–3)	AB673324 – 328	AB682990 – 992
<i>C. batrachus</i> (outgroup)	Kyank Tan, Yangon, Myanmar	Lower Irrawaddy	1(1)	RLIKU 1907	AB673329	AB682993

^aLocality codes corresponding to those shown in Fig. 1

^bSamples collected from local markets

^cNumbers of specimens used for the ND2 and RAG1 gene sequencing are shown, respectively, outside and inside parentheses

^dVoucher numbers at the Research Laboratory of Ichthyology, Faculty of Fisheries, Kasetsart University are shown for specimens used for the ND2 gene sequencing. In parentheses, voucher numbers for the RAG1 gene sequencing are also shown

Table 2. Summary of gene sequence data for *Heteropneustes* individuals

Genes	No. of individuals	No. of haplotypes	Alignment length	Codon variable sites				Codon parsimony informative sites			
				1st	2nd	3rd	all	1st	2nd	3rd	all
ND2	73	48	1,044	69	26	138	233	63	24	131	218
RAG1	47	10	1,466	7	3	46	56	4	1	24	29

Table 3. Comparison of four morphological characters between the “*H. fossilis*-like” and “*H. kemratensis*-like” groups and statistical tests for the morphological distinction between the two groups

Morphological characters	<i>H. fossilis</i> -like group			<i>H. kemratensis</i> -like group			<i>F</i> -value	<i>p</i> -value
	range	(mean±SD)	n	range	(mean±SD)	n		
Standard length (mm)	74.1–242.2	(131.1±31.0)	47	80.5–225.4	(146.8±34.0)	42		
Meristic counts*								
Number of anal fin rays	59–74	(68.0±3.9)	47	75–85	(79.1±2.7)	42	237.890	0.000
Number of total vertebrae	56–62	(59.2±1.3)	47	61–65	(63.1±1.2)	42	207.990	0.000
Morphometric measurements**								
Pectoral spine length%SL	8.0–10.9	(9.6±0.6)	41	9.2–12.1	(10.9±0.9)	40	45.084	0.000
Caudal peduncle depth%SL	5.0–7.2	(6.3±0.6)	47	4.2–5.8	(5.2±0.4)	42	76.087	0.000

Morphological data shown in supplementary tables S2-S4 were compared. See Material and methods for details of statistical tests

*Tested by Student’s t-test

**Tested by ANCOVA

Fig. 1 (Ratmuangkhwang et al.)

Fig. 2 (Ratmuangkhwang et al.)

Fig. 3 (Ratmuangkhwang et al.)

Fig. 4 (Ratmuangkhwang et al.)

Fig. 5 (Ratmuangkhwang et al.)

— = “*H. fossilis*”-like traits
— = “*H. kemratensis*”-like traits

Clade A
Southeast Asia

Clade B
Northeastern parts of India

Clade C
Southwestern parts of India