Molecular Ecology (1995) 4, 331–346

# Phylogeographic structure in mitochondrial DNA of a South-east Asian freshwater fish, *Hemibagrus nemurus* (Siluroidei; Bagridae) and Pleistocene sea-level changes on the Sunda shelf

J. J. DODSON, F. COLOMBANI and P. K. L. NG
*Département de biologie, Université Laval, Sainte-Foy, Québec, Canada G1K 7P4, Department of Zoology, National University of Singapore, Kent Ridge, Singapore 0511*

## Abstract

**We postulated that the biogeographical history of South-east Asia contributed to extensive admixture during Pleistocene low sea levels of genetic groups of an obligate freshwater fish (the river catfish, *Hemibagrus nemurus*) isolated during periods of high sea levels. During Pleistocene glacial maxima, the sea level was lower than at present and the islands of the Sunda shelf (Sumatra, Borneo and Java) and the Asian mainland were connected by lowlands traversed by rivers. Restriction fragment length polymorphisms in mitochondrial DNA were documented for 140 putative *H. nemurus* analysed from 13 sampling sites resulting in the definition of 35 haplotypes. The high level of haplotype differentiation (mean $P \times 100 = 2.22$, SD = 1.33) indicates that the subdivision of the ancestral *H. nemurus* group was extensive and probably occurred early in the Pleistocene. The occurrence of some genetically divergent groups of the *H. nemurus* complex occurring in sympatry in widely separated locations supports the proposition that low sea levels aided the dispersion and mingling of genetic groups. Based on both genetic and morphological evidence, the main *H. nemurus* line gave rise to three regional groups: (1) a morphologically distinct 'Indochinese' group composed of two mtDNA clades overlapping in east peninsular Malaysia; (2) a 'Sundaic' group composed of various lineages of differing morphology and genetic identity; (3) a genetically distinct 'Sarawak' group in west Borneo, similar in morphology to the 'Sundaic' and 'Indochinese' groups, but including a small, golden colour morph as a distinct clade. The morphologically similar Sundaic forms from west Java, Sumatra and west Borneo show some degree of genetic divergence, but their phylogenetic relationships are poorly resolved. The most genetically and morphologically distinct Sundaic clade, assigned to *H. hoevenii*, colonized the Kapuas river (west Borneo), east Sumatra and south peninsular Malaysia. Contrary to our original hypothesis and present biogeographical theory, little exchange of genetic groups has apparently occurred between the mainland and the Sunda Islands during recent glaciations.**

*Keywords*: population genetics, mtDNA, restriction fragment length polymorphisms, freshwater catfish, molecular phylogenetics, South-east Asia.

*Received 15 August 1994; revision received 12 December 1994; accepted 16 December 1994*

Contribution to the tropical aquatic biology program of SEAMEO-BIOTROP (South-east Asian Ministers of Education Organization-Regional Centre for Tropical Biology, Bogor, Indonesia).

Correspondence: Julian Dodson. Tel. +1 418 656 3289. Fax +1 418 656 2339. E-mail c034@music.ulaval.ca



## Introduction

The magnitude and geographic pattern of population-genetic structure of a species is a result of historical as well as contemporary events. The importance of historical biogeography in shaping intraspecific genetic structure is

well established and the ability of mitochondrial DNA (mtDNA) to retain a history of past isolation is well demonstrated (Avise *et al.* 1987; Avise 1989; Billington & Hebert 1991), even in the event of the contemporary admixture of groups that evolved allopatrically. The analysis of mtDNA has been particularly effective in demonstrating the influence of Pleistocene glaciation events in shaping the intraspecific genetic structure of continental fauna such as freshwater fishes (e.g. Billington & Hebert 1991; Bernatchez & Dodson 1994), birds (e.g. Zink & Dittman 1993; Avise *et al.* 1992) and mammals (e.g. Riddle *et al.* 1993). These studies have demonstrated the evolution of distinct intraspecific groups in separate glacial refugia during periods of glacial advances followed by postglacial expansion and the colonization of newly available habitat by ancestral populations. Many of these species are characterized by low levels of mtDNA haplotype differentiation, with closely related haplotypes dispersed over wide areas (e.g. Zink & Dittman 1993; Bernatchez & Dodson 1994). The reduced genetic variation in these species is most probably due to large-scale destruction of the fauna due to glaciation and severe population bottlenecks in glacial refugia (Bernatchez *et al.* 1989; Billington & Hebert 1991).

In contrast, some species sampled in more southerly localities have retained mtDNA haplotypes which diverged many millions of years ago. The divergent mtDNA groups of these species are associated with distinct geographic ranges reflecting long-term zoogeographic barriers to gene flow that are largely independent of glaciation events (reviewed by Avise *et al.* 1987). In some northern species affected by glaciation, older mtDNA haplotypes survive from isolation events that occurred during early Pleistocene glacial advances (e.g. Bernatchez & Dodson 1990). Yet even in these situations, divergent groups of related mtDNA haplotypes are most commonly spatially separated with geographically constrained contact zones between neighbouring groups. Only in rare cases do divergent groups separated by more than 1% nucleotide sequence divergence not occupy distinct ranges (category II in the classification of Avise *et al.* 1987), being widely spread within and between populations (Avise *et al.* 1992 and references therein).

The purpose of this study was to examine a unique biogeographical situation in which the widespread occurrence of distinct mtDNA clades was expected based on the geological history of the region. This region is South-east Asia and involves the world-wide variations in sea level that occurred during the glaciations of the Pleistocene. The Sunda continental shelf unites mainland Asia and peninsular Malaysia with the greater Sunda Islands (Sumatra, Java and Borneo). The shelf is no more than 100 meters in depth and most of it is much shallower. There is some discussion as to the exact amplitude of sea-level changes over the Sunda shelf, but it is generally accepted that sea-level during the last glaciation was between 100 and 200 meters lower than today (Verstappen 1975; Morley & Flenley 1987). Two, possibly three, older Pleistocene periods of low sea level have been reported in the order of 45–70 m (Verstappen 1975). Whatever the exact amount of lowering of sea levels, great areas of the shallow Sunda shelf were converted to land during periods of glacial maxima.

Extensive interconnections during periods of glacial maxima have probably substantially influenced the distribution of the flora and fauna. Verstappen (1975) illustrated five major drainage systems proposed to have existed on the Sunda lowlands during low sea-levels (Fig. 1). Three rivers emptied into the present day South China Sea; the Mekong, the Chao Phraya and tributaries collected south to the tip of peninsular Malaysia, and the North Sunda river and tributaries collected from eastern Sumatra, western Borneo and possibly from peninsular Malaysia. The drainage of the present day Java sea was eastward, but a central ridge is proposed to have largely separated the drainage of south-east Sumatra and Java

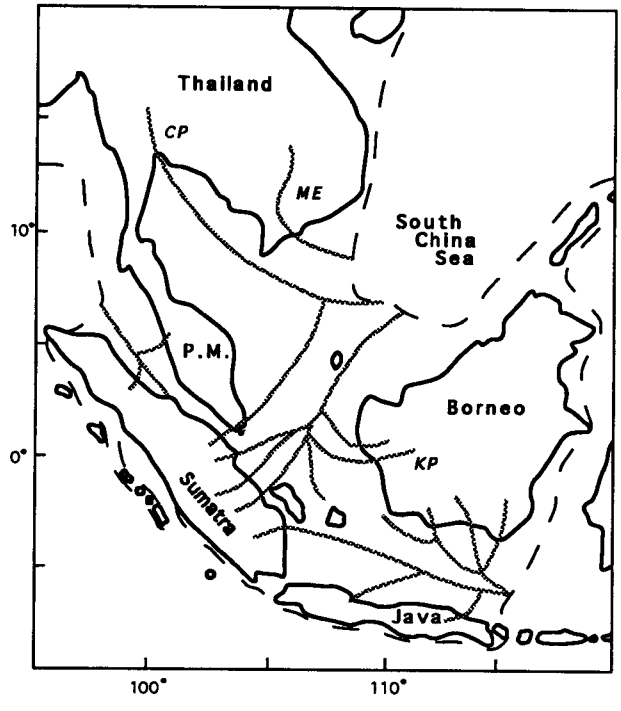


**Fig. 1** Location map of the various place names mentioned in the text. P.M., peninsular Malaysia. Contemporary river systems: *KP*, Kapuas; *ME*, Mekong; *CP*, Chao Phraya. Dashed lines indicate the edge of the Sunda shelf and the maximum extent of the Sunda lowlands during Pleistocene glacial maxima. Hatched lines indicate the major drainage systems proposed to have existed during the most recent glaciation. Adapted from Verstappen (1975).

from the drainages of southern Borneo. Finally, a smaller river drained north-westward between northern Sumatra and western peninsular Malaysia. Such wide-ranging interconnections during low sea-level phases may have facilitated the exchange of faunal elements. Yet during high sea levels [sea levels slightly above present ones are recorded during the Holocene (Tjia *et al.* 1984)], the isolation of Peninsular Malaysia, Sumatra, Borneo and Java from one another and from the continent may have favoured the evolution of geographical races (e.g. Zeuner 1941). Pleistocene sea-level changes on the Sunda shelf may have facilitated the admixture of intraspecific groups of freshwater fishes that evolved in allopatry during periods of high sea level.

Early this century, Molengraaff & Weber (1919) noted that the similarity of the freshwater fish fauna of west Borneo and south-east Sumatra supported their hypothesis that during Pleistocene low sea levels, the rivers of both of these areas formed part of the upper reaches of the North Sunda river. Of the 263 species of fish known from the Malay peninsula, 44% occur in the Mekong, 47% in the Chao Phraya and 66% in Borneo, Sumatra and Java (Kottelat 1989). The area cladogram for phallostethid species (priapus fishes) prepared by Parenti (1991) also illustrates the close relationship of Thailand, peninsular Malaysia and north-west Borneo. The faunal similarity of these regions has long been associated with dispersal during Pleistocene low sea levels (de Beaufort 1951).

We tested the proposition that the faunal similarity of the region observed at the level of species and genera should also be reflected at the intraspecific level, as evaluated by sequence divergence in mtDNA. During glacial phases, the islands of the Sunda shelf and the mainland were connected by lowlands traversed by rivers. Each time this happened, the faunas of peninsular Malaysia, Sumatra, Borneo and Java may have mingled in these lowlands with faunal exchanges among present-day land masses. In this case, we would expect the evolution of distinct intraspecific groups in allopatry during periods of glacial minima and their admixture during periods of glacial maxima. We thus predicted the presence of large mtDNA differences within populations in the absence of spatial segregation.

The bagrid catfish, *Hemibagrus nemurus* Valenciennes, 1839 (Mo 1991) was chosen to test the hypothesis. It is a typical component of the fish community inhabiting large rivers of the tropical rain forest. The species is usually abundant and not extensively cultured so that large-scale introductions by man were not expected to alter the natural distribution patterns of intraspecific genetic groups. It is large enough (maximum total length of ~ 60 cm) to be of interest as a food fish and thus available from local fishermen. The species is readily identified by its short adipose-fin base (relative to the anal-fin base), a thin dark lateral line running the length of the body and the maxillary barbels extending to the anal fin (Inger & Chin 1962). The taxonomic status of the species is, however, unclear.

Mo (1991) revised the generic classification of the Bagridae and referred all the south-east Asian bagrids previously assigned to *Mystus* Scopoli, 1777, to *Hemibagrus* Bleeker, 1862. The taxonomy of the genus is unstable and several species described by early taxonomists have been synonymized under *H. nemurus*. Roberts (1979) synonymized *H. johorensis* Herre, 1940 and *H. pahangensis* Herre, 1940, under *H. nemurus*. Mo (1991), however, recognized both species as valid taxa based on morphological criteria. Two species established by Bleeker (1847), namely *H. sieboldii* and *H. hoevenii*, had previously been synonymized under *H. nemurus* by Weber and de Beaufort (1913). However, Kottelat *et al.* (in press) and Kottelat and Lim (in press) recently recognized *H. hoevenii* as a valid taxon. Thus, a second objective of the study was to relate mtDNA identity to variations in morphology in an attempt to clarify the taxonomic status of *H. nemurus*.

## Materials and methods

### *Sampling*

Approximately 12 fish were sampled from each of 13 sites situated around the Sunda shelf (Fig. 2).

1 The northern-most sample (a) was obtained in north-east Thailand from the Sirinthorn reservoir, located on a tributary of the Moon river located just upstream of the confluence of the Moon and Mekong Rivers.

2 In peninsular Malaysia, six sites were sampled: (b) Terengganu (Kenyir Reservoir, eastern Malaysia), (c) Kelantan (north-east peninsular Malaysia, on the border with Thailand), (d) Pahang (Pahang river, east peninsular Malaysia), (e) Perak (Chenderoh Reservoir, west peninsular Malaysia), (f) Endau (Endau-Rompin Park, Johor Province, east peninsular Malaysia) and (g) Johor (Johor river at Kota Tingi, south peninsular Malaysia).

3 In western Borneo, three sites were sampled; (h) Sadong (Sadong river, at Serian, the Malaysian state of Sarawak, west Borneo), (i) Rajang (Rajang River at Kapit and Sibu, Sarawak), (j) Kapuas (Kapuas river at Pontianak, in the Indonesian state of west Kalimantan).

4 In eastern Sumatra, two sites were sampled; (k) Jambi (Batanghari river at Jambi, central Sumatra) (l) and Palembang (Musi river at Palembang, south Sumatra).

5 One sample (m) was obtained from West Java (Cirata reservoir).

Fish were either sampled by the authors or purchased from local fishermen or markets. They were immediately dissected to remove liver and eggs which were subsequently frozen or kept on wet ice prior to mtDNA extraction. Most carcasses were kept for morphological analysis.

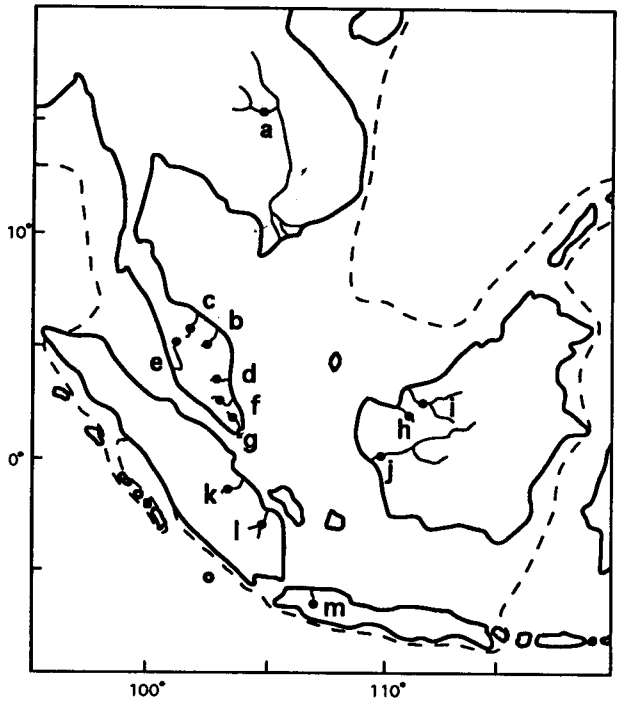


Fig. 2 Location map of *H. nemurus* sampling sites named in Table 1.

**Table 1** Sample locations, sample sizes and number of mtDNA haplotypes for populations of *Hemibagrus nemurus* sampled in South-east Asia. Letters (a–m) correspond to sampling sites illustrated in Fig. 1

| Sampling site | Sample size | Number of haplotypes |
|---|---|---|
| (a) North-east Thailand | 13 | 2 |
| (b) Terengganu | 9 | 3 |
| (c) Kelantan | 12 | 5 |
| (d) Pahang | 9 | 4 |
| (e) Perak | 13 | 2 |
| (f) Endau | 7 | 2 |
| (g) Johor | 2 | 1 |
| (h) Sadong | 14 | 3 |
| (i) Rajang | 10 | 4 |
| (j) Kapuas | 10 | 5 |
| (k) Jambi | 14 | 5 |
| (l) Palembang | 15 | 3 |
| (m) West Java | 12 | 3 |

### *Mitochondrial DNA extraction and analysis*

Mitochondrial DNA was successfully extracted and analysed for 140 fish (Table 1). Tissues were extracted using a modification of the rapid extraction method of Bernatchez *et al.* (1988); the phenol–phenol/chloroform–chloroform extraction was replaced with two phenol/chloroform–chloroform extractions. Aliquots of mtDNA were digested separately with five hexameric (*Apa*I, *Bst*EII, *Dra*I, *Eco*RV, *Pst*I), three multihexameric (*Acc*I, *Ava*I, *Hin*cII) and two multipentameric (*Ava*II, *Ban*II) restriction endonucleases. Mitochondrial DNA fragments were electrophoretically separated on 0.8% or 1.2% agarose gels for 4 h at 75 V (with the exception of *Ava*I fragments that were separated overnight at 20 V to improve the resolution of the large fragments).

Ethidium bromide staining revealed restriction fragments when mtDNA was obtained in sufficient quality and quantity. In cases of low or poor-quality yield, mtDNA fragments were transferred to nylon membranes under alkaline conditions and hybridized with a pure radiolabelled total *H. nemurus* probe prepared with the multiprime (Amersham) DNA labelling reaction. Membranes were autoradiographed using an intensifying screen for 5–16 h at –70 °C.

Size estimates of fragments were made by running simultaneously into the agarose gel digests of phage lambda DNA with *Hin*dIII and *Eco*RI–*Hin*dIII double digest. Fragments less than 200 base pairs in length could not be scored. Distinct single endonuclease patterns were identified by a specific letter. Each fish was assigned a multiletter code which described its composite mtDNA genotype.

Both distance and character-based analyses were used to define genetic groups and phylogenetic relationships. Restriction-site differences between genotypes were inferred from changes in fragment patterns since such changes could be readily explained by specific site gains or losses. The only exception was one of the two haplotypes characteristic of the putative *H. nemurus* population of the Endau river, peninsular Malaysia, whose fragment pattern could not be explained by specific site gains or losses from neighbouring haplotypes. Therefore, sequence divergence between haplotypes was first estimated according to Upholt's (1977) fragment method. The resulting distance matrix served to construct a phenogram using the unconstrained branch-length clustering method of Fitch and Margoliash (1967) (program FITCH in the PHYLIP 3.4 computer package, Felsenstein 1992) to estimate the divergence of the Endau haplotype from all others and to identify the next closest haplotype to serve as an outgroup for further analyses based on site methods. Estimates of nucleotide sequence divergence ($p$) among mtDNA haplotypes calculated by the site approach of Nei and Li (1979) were used to estimate genetic distance among all other haplotypes. The resulting site presence-absence matrix was used to generate phylogenetic trees according to Wagner parsimony criteria using the MIX program from PHYLIP. Symplesiomorphic and autapomorphic characters were omitted from the site matrix. As many trees of equal length were found, a consensus tree and bootstrapping estimates on branches were estimated

by running 10 000 bootstrap replicates with the BOOT program of PHYLIP. Phylogenetic trees were rooted using haplotype 35 as it was closest to the divergent Endau haplotype as revealed by the fragment method (see above). In addition, estimates of genetic distance using all restriction sites served to construct a phenogram using FITCH. As a measure of inter and intrapopulation genetic diversity, we computed the maximum likelihood estimation of the average number of nucleotide substitutions per site within populations (nucleotide diversity) and between populations (nucleotide divergence) (Nei & Tajima 1981, 1983; Nei 1987). The Neighbour-Joining program of PHYLIP was used to cluster populations according to nucleotide divergence.

### *Morphological analysis*

One hundred and twenty-five of the fish analysed for mtDNA were analysed for morphometrics (Fig. 3). Measurements included: standard length (SL); dorsal fin length (DL) measured from the base of the first ray to the tip of the longest ray; adipose fin length (AL) measured along the base of the fin where the fin joins the body; distance between the base of the last dorsal fin ray and the beginning of the base of the adipose fin (DA); body height measured at the anus (HA); the distance from the end of the adipose fin to the end of the hypural complex (PA); the maximum height of the caudal peduncle (CP); head length (HL) measured from the tip of the snout to the tip of the occipital process; the distance between the tip of the occipital process and the base of the first dorsal fin ray (OD); head width (HW), measured along the widest part with the opercula closed and tightly pressed against the body; eye diameter (ED) measured along its widest diameter; snout length (SN) measured along the longest distance between the tip of the upper jaw and the anterior margin of the eye and the interorbital distance (IO) measured along the shortest transverse distance between the orbits. The tip of the occipital process in the preserved specimens is usually clearly discerned, but in some specimens, the skin has to be pressed tightly against the skull to determine the position of the tip.

The 13 morphological variables were standardized and a principle components analysis was used to correct for size effects. The analysis revealed a strong size effect on the first principal axis due to the sampling of fish of different lengths. The variables were thus projected in the space orthogonal to the direction of the first axis to eliminate the size effect (Joliffe 1986). This data set was used in all further analyses. To avoid subjectivity in the clustering of fish based solely on the visual analysis of the graphical representation of the principal components analysis, the number of clusters in the data set was objectively determined by applying the cubic clustering criterion to Ward's hierarchical clustering method (Milligan & Cooper 1985). Fish were assigned to clusters using the nearest-centroid-sorting, nonhierarchical classification procedure (Lebart *et al.* 1984). Although a hierarchical method was used to generate the clustering solution, the criterion can be extended for use with nonhierarchical procedures (Milligan & Cooper 1985). A discriminant analysis was used to determine which morphological characteristics (corrected for

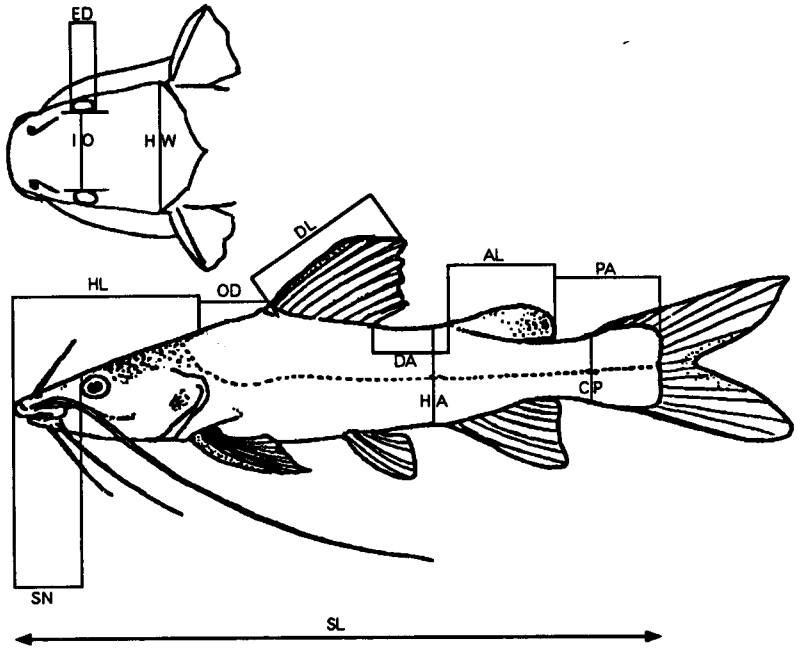


Fig. 3 Dorsal and lateral views of *H. nemurus* illustrating the morphometric measurements. See text for definitions of abbreviations.

the size effect) were most responsible for discriminating the clusters. Finally, the morphological clusters were directly compared to mtDNA-haplotype groups.

The 125 specimens for which both morphological measurements and mtDNA haplotype identity are available are stored in the Zoological Reference Collection (ZRC) of the National University of Singapore. The ZRC specimen codes, their respective mtDNA haplotype numbers and collection sites are provided in Appendix 1.

## Results

The 10 endonucleases generated a total of 197 restriction fragments with a mean of 68 fragments per individual. All enzymes were polymorphic and discriminated a total of 35 mtDNA haplotypes among the 140 *H. nemurus* analysed from 13 sampling sites (Tables 1 and 2). Fragment patterns generated by each enzyme and restriction-site differences between genotypes inferred from changes in fragment patterns are presented in Figs 4 and 5. The 102 restriction sites so identified comprised 54 synapomorphic sites. The mean size of the mtDNA genome, as estimated by averaging the sums of all digestion patterns, was 17921 bp. (SD = 525), similar to the length estimates obtained for a variety of other fish species (Billington & Hebert 1991).

An analysis of sequence divergence between haplotypes using the fragment method followed by clustering with FITCH (Fig. 6) revealed that one of the two haplotypes sampled at Endau (designated haplotype 1) was separated from all other haplotypes by an average sequence divergence of 5.86%. This highly divergent haplotype characterized six of the seven fish sampled at Endau and was found nowhere else. Lim *et al.* (1990) reported specimens from Endau-Rompin in peninsular Malaysia that were assigned to an undescribed species close to *H. nemurus* but morphologically different from this species in several re-

| | Composite haplotypes | | | | | | | | | | *n* | Locality |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | A | A | A | A | A | A | A | A | A | A | 6 | a |
| 2 | D | D | B | B | B | B | C | D | B | D | 1 | a |
| 3 | B | C | B | C | B | C | B | B | C | C | 14 | b,c |
| 4 | D | B | B | B | B | D | C | C | B | B | 3 | b,c |
| 5 | B | C | B | C | C | C | B | B | C | C | 2 | b,c |
| 6 | C | C | B | C | B | C | B | B | C | C | 1 | c |
| 7 | B | B | B | C | B | D | C | C | B | B | 1 | c |
| 8 | D | B | B | B | B | B | C | D | B | D | 3 | d |
| 9 | D | B | B | B | B | B | C | D | B | E | 3 | d |
| 10 | D | B | B | F | B | B | C | D | B | D | 1 | d |
| 11 | B | C | B | D | B | C | B | B | C | C | 2 | d |
| 12 | E | E | C | B | B | E | B | E | D | F | 9 | e |
| 13 | E | E | C | B | B | E | B | E | D | G | 4 | e |
| 14 | G | F | D | E | D | F | B | F | D | H | 1 | h |
| 15 | F | F | D | E | D | F | B | F | D | H | 8 | h |
| 16 | F | F | D | E | D | F | B | F | D | F | 11 | h,i |
| 17 | E | F | D | E | B | G | B | E | D | F | 2 | i |
| 18 | F | F | D | E | B | H | B | F | D | F | 1 | i |
| 19 | F | F | D | E | B | H | B | F | D | F | 1 | i |
| 20 | H | E | E | F | B | I | C | B | D | I | 10 | g,j,l |
| 21 | E | E | C | F | B | E | B | E | D | J | 1 | j |
| 22 | E | E | C | F | B | E | B | E | D | F | 3 | j |
| 23 | E | E | C | B | B | E | B | E | D | J | 1 | j |
| 24 | H | E | C | F | B | I | C | E | D | I | 1 | j |
| 25 | E | E | B | B | B | J | B | G | D | K | 5 | m |
| 26 | E | E | B | B | B | J | B | H | D | K | 6 | m |
| 27 | E | E | B | B | B | J | B | E | D | K | 1 | m |
| 28 | E | E | C | B | B | J | B | E | D | F | 3 | k |
| 29 | E | E | C | B | B | N | B | E | D | F | 17 | k,l |
| 30 | E | E | C | B | C | J | B | E | D | F | 1 | k |
| 31 | E | E | C | B | C | O | B | E | D | F | 1 | k |
| 32 | H | E | E | F | B | I | C | E | D | I | 2 | l |
| 33 | B | C | B | G | F | P | F | L | D | C | 10 | a |
| 34 | K | E | B | F | B | E | B | M | H | E | 3 | a |
| 35 | E | B | B | B | B | B | B | B | B | F | 1 | k |

**Table 2** Definitions of composite haplotypes, absolute frequency (*n*) and distribution of the mtDNA haplotypes resolved among *Hemibagrus nemurus*. Restriction enzymes are in order: *Apa*I, *Bst*EII, *Dra*I, *Eco*RV, *Pst*I, *Acc*I, *Ava*I, *Ava*II, *Ban*II, *Hinc*II. Capital letters identify fragment patterns presented in Fig. 3. Small letters refer to sample locations illustrated in Fig. 1

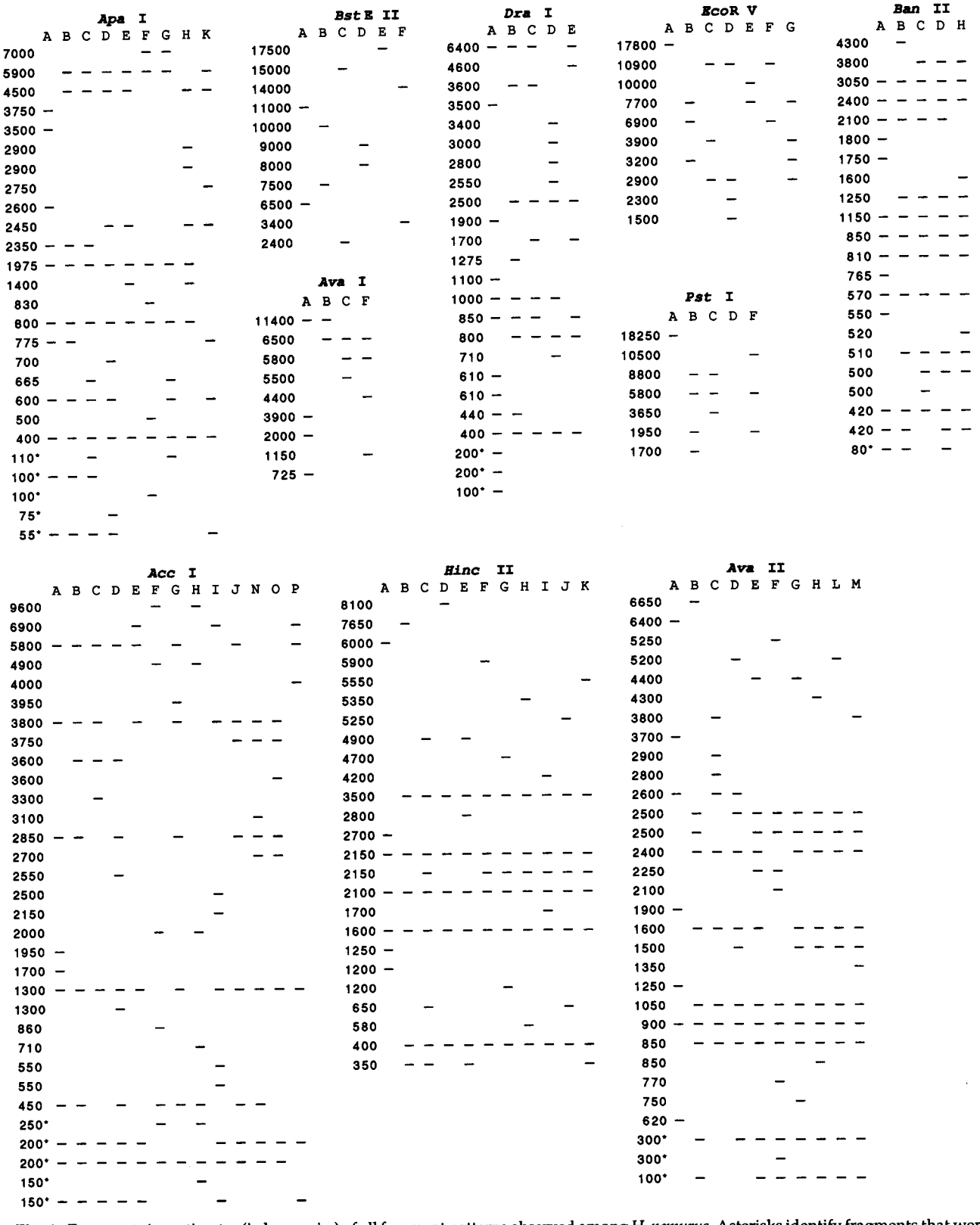


Fig. 4 Fragment size estimates (in base pairs) of all fragment patterns observed among *H. nemurus*. Asterisks identify fragments that were not observed but assumed under the criterion of minimum mutational steps involved in fragment changes.

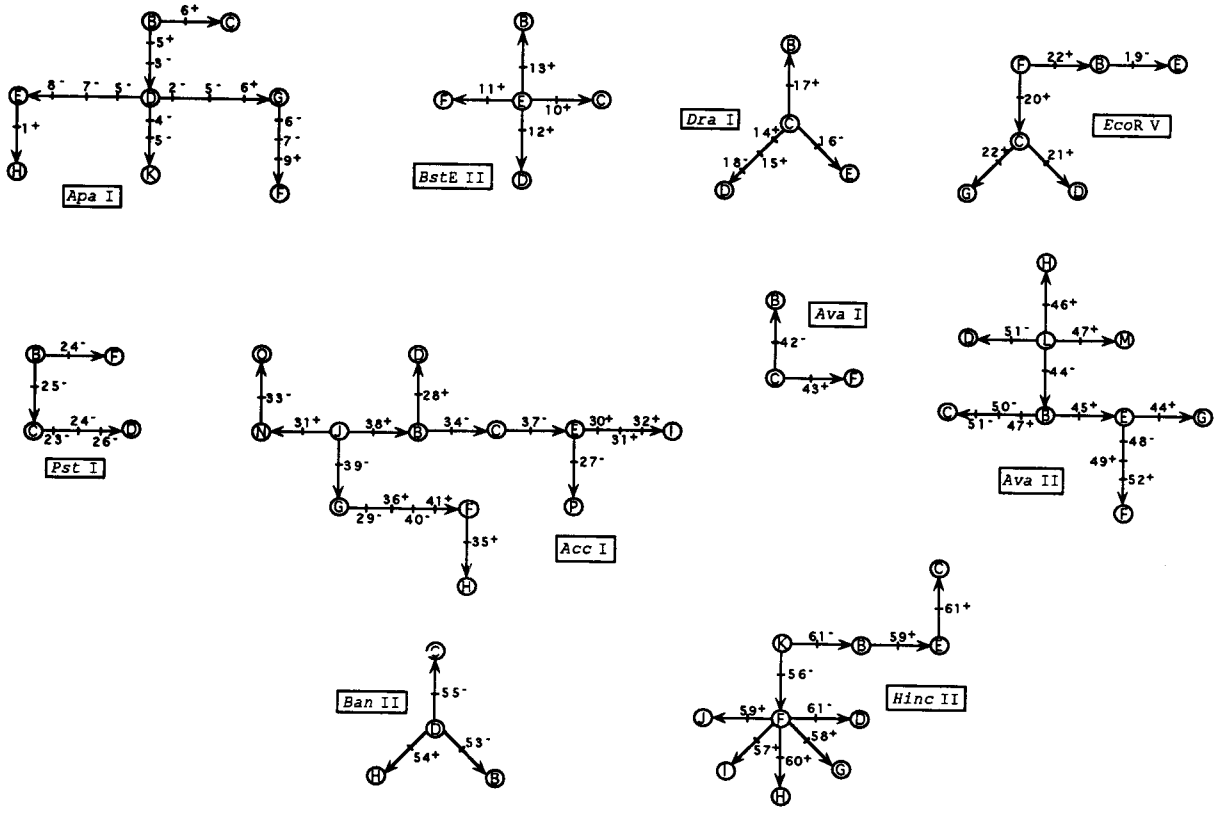


**Fig. 5** Parsimonious, unrooted networks illustrating relationships among fragment patterns observed for all restriction enzymes described in Fig. 4. Site changes [losses (–) or additions (+)] involved in moving between fragment patterns (letters) are numbered along branches (arrows indicate direction of changes) and were parsimoniously deduced from observed fragment patterns.

spects. These fish are one and the same and are considered to belong to a distinct species whose description will appear elsewhere (Ng & Ng, in press).

The phenogram constructed with FITCH clustered the remaining 34 haplotypes based on restriction sites (Fig. 7), and the majority-rule consensus tree (Fig. 8) revealed a number of discrete mtDNA clades, closely related haplotypes and a number of isolated haplotypes.

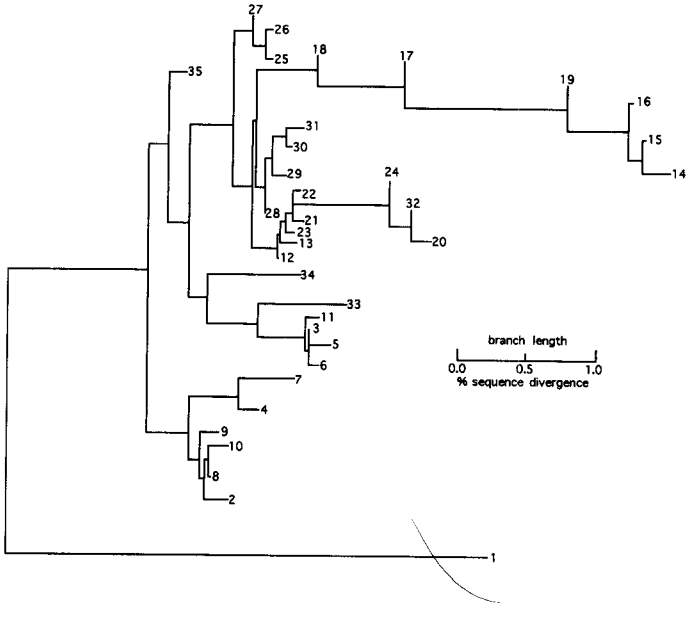


**Fig. 6** FITCH phenogram clustering the distance matrix, based on restriction fragments, of percent sequence divergence among the 35 mtDNA haplotypes described for *H. nemurus*. Numbers refer to haplotypes defined in Table 2. The tree was rooted with haplotype 1, an undescribed *nemurus*-like species from Endau river, Peninsular Malaysia.

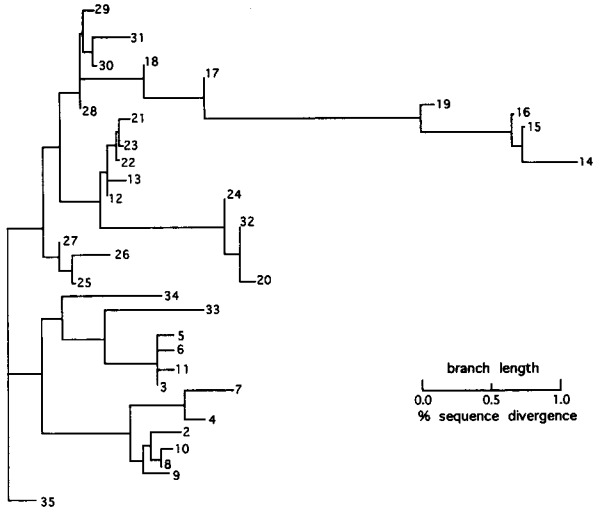


**Fig. 7** FITCH phenogram clustering the distance matrix, based on restriction sites, of percent sequence divergence among 34 mtDNA haplotypes described for *H. nemurus*. Numbers refer to haplotypes defined in Table 2. The tree was rooted with haplotype 35.

*Group I.* This group comprises six haplotypes from east peninsular Malaysia (2, 9, 4, 7, 8 and 10) and was supported at the 82% level in the majority-rule consensus tree. Restriction site 5 (*Apa*I fragment pattern D) was diagnostic for this group (Table 2).

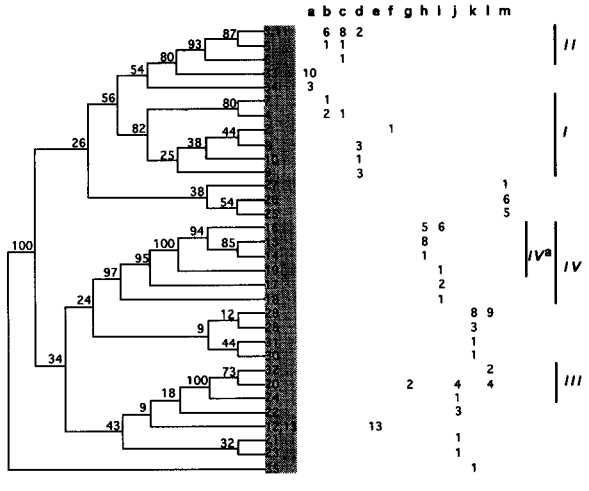


**Fig. 8** Majority rule consensus tree clustering 34 mtDNA haplotypes of *H. nemurus* and their geographical distribution. Numbers at the forks indicate the percent of times the group consisting of the haplotypes located to the right of the fork occurred among the trees, out of 10 000 trees. The tree was rooted with haplotype 35. Numbers in shaded area refer to haplotypes defined in Table 2. Letters a–m refer to sampling sites identified in Table 1 and numbers below sampling sites are absolute abundances. The principal mtDNA clades are identified with vertical lines.

*Group II.* This group comprises four haplotypes from east Peninsular Malaysia (3, 5, 6 and 11) and was supported at the 93% level in the majority rule consensus tree. *Bst*EII, *Acc*I, *Ban*II and *Hin*cII restriction sites were diagnostic for the group. Haplotype 33, from Thailand, clustered with group II in 80% of the bootstrap estimates and haplotype 34, also from Thailand, clustered with group II in 54% of the bootstrap estimates.

*Group III.* This group comprises haplotypes 20, 32 and 24 from the Kapuas river, west Borneo, Sumatra and peninsular Malaysia and was supported in 100% of the bootstrap estimates. *Apa*I, *Acc*I and *Hin*cII restriction sites were diagnostic for the group. Group III haplotypes were clustered with a larger group of haplotypes including 12 and 13, from Perak in west peninsular Malaysia, and 23, 21 and 22 from the Kapuas river. However, this larger cluster of haplotypes was supported in only 42% of the bootstrap estimates.

*Group IV.* This group comprises haplotypes 14, 15, 16, 17, 18 and 19 from the Sadong and Rajang Rivers of Sarawak, west Borneo, and was supported in 97% of the bootstrap estimates. *Bst*EII, *Dra*I and *Eco*RV restriction sites were diagnostic of the group. Haplotypes 14, 15, 16 and 19 formed a distinct clade (group IV$^a$) within this group supported in 100% of the bootstrap estimates.

The relationships of the remaining haplotypes were poorly defined in the majority rule consensus tree. Haplotypes 25, 26 and 27 from west Java were clustered in only 37.5% of the bootstrap estimates, although a diagnostic *Hin*cII restriction site discriminated the group from all others. Haplotypes 28, 29, 30 and 31 from east Sumatra were clustered in only 9% of the bootstrap estimates. No diagnostic restriction site was associated with this group. Finally, haplotype 35, also from Sumatra, is a distinct haplotype rather distant from all neighbouring haplotypes.

Divergent mtDNA clades and haplotypes were found together at several geographical locations (Fig. 9). Groups I and II intermingle throughout east peninsular Malaysia, with group I dominating in the south and group II dominating in the north {intergroup genetic distance; $P \times 100$ [mean (SD)] = 2.08 (0.21)}. Group III illustrates the widest distribution, intermingling with haplotypes 21,22, and 23 in the Kapuas river {intergroup genetic distance; $P \times 100$ [mean (SD)] = 1.17 (0.18)} and haplotype 29 at Palembang {$P \times 100$ [mean (SD)] = 1.43 (0.15)}. In Jambi, east Sumatra, the divergent haplotype 35 was found with haplotypes 28, 29, 30 and 31 {$P \times 100$ [mean (SD)] = 2.4 (0.21)}. In the Rajang river of Sarawak, west Borneo, group IV$^a$ haplotypes were found with haplotypes 17 and 18 ($P \times 100 = 2.39$, SD = 0.45). In many of these cases, genetic differences between sympatric groups and haplotypes were accompanied by significant morphological differences (see below).

The maximum likelihood estimation of the average

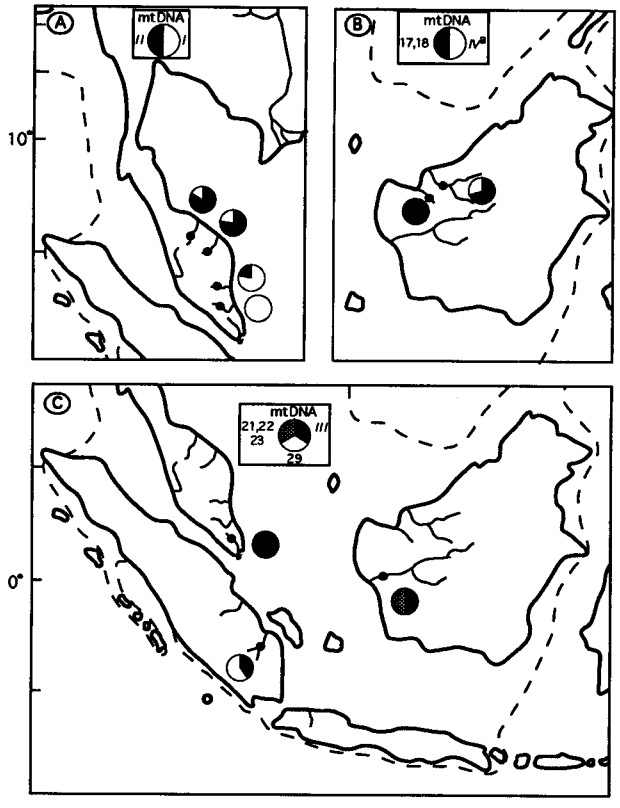


**Fig. 9** The geographic distribution of mtDNA phylogenetic groups occurring in sympatry and occupying more than one geographical location. A, mtDNA clades I and II; B, mtDNA clade IV$^a$ and haplotypes 17 and 18; C, mtDNA clade III and haplotypes 21, 22, 23 and 29.

number of nucleotide substitutions per site within populations (nucleotide diversity) and between populations (nucleotide divergence) revealed two groupings of populations and three divergent populations (Fig. 10). The populations of west and south peninsular Malaysia [Perak (e), Johor (g)], Palembang (l) in east Sumatra, west Java (m) and the Kapuas river (j) of west Borneo are genetically more similar than any are to either the rivers of Sarawak, west Borneo [Sadong river (h) and the Rajang river (i)], Jambi (k) in east Sumatra, or the rivers of east peninsular Malaysia and Thailand (sampling sites a, b, c and d). Intrapopulation genetic diversity was greatest at sites where divergent mtDNA clades intermingle. The Endau population (f) was excluded from this analysis as only one fish exhibited a haplotype (2) assigned to *H. nemurus*.

## *Morphometrics*

The cubic clustering criterion using Ward's method revealed five clusters of fish based on morphometric characteristics (Fig. 11). Two of the five groups subsequently defined by the nearest-centroid-sorting procedure were composed of one fish each. These two fish were reclassified in the group corresponding to its closest neighbour, producing three morphological groups. The discriminant analysis revealed that the three groups defined by the nearest-centroid-sorting procedure were significantly different ($F_{24,222} = 18.26$, $P < 0.0001$) with over 93% of observations correctly classified in their group of origin (Table 3). The first function of the discriminant analysis explained

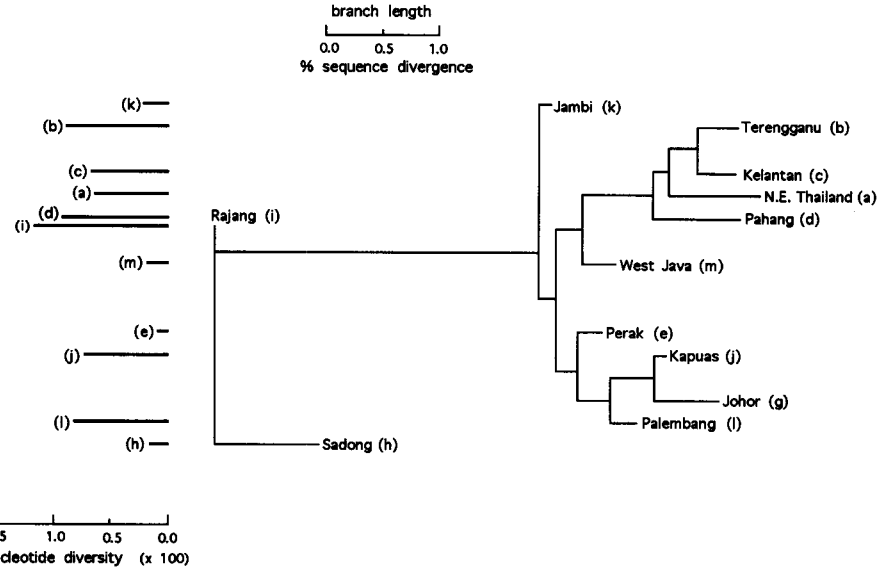


**Fig. 10** Neighbour-joining phenogram clustering populations according to the distance matrix resulting from maximum likelihood estimation of the net average number of nucleotide substitutions per site between populations (nucleotide divergence). Intrapopulation nucleotide diversity is presented to the left of each population. The tree was rooted with the Sadong population. Letters refer to populations identified in Table 1.

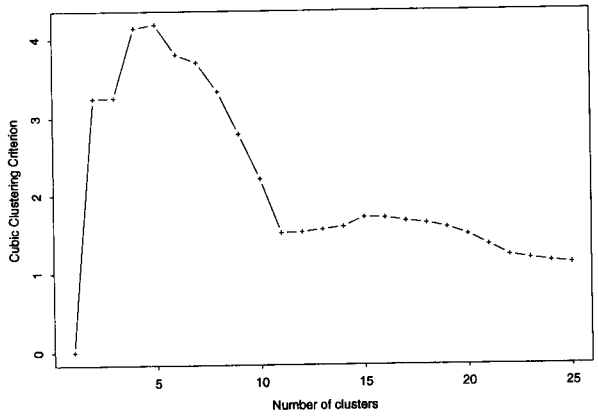


**Fig. 11** The cubic clustering criterion applied to the cluster analysis by Ward's method. The maximum value of the criterion across the hierarchy levels (number of clusters) is used to indicate the optimal number of clusters (5) in the data.

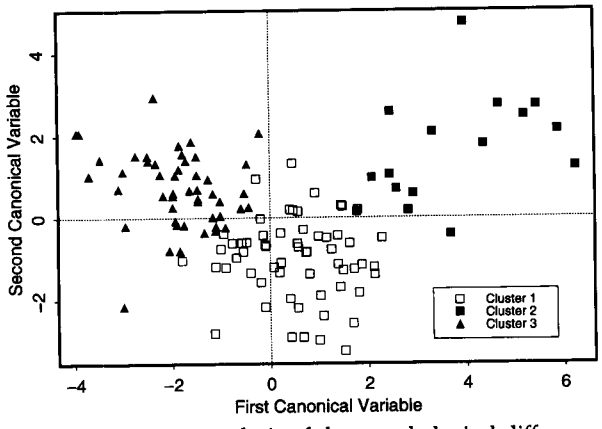


**Fig. 12** Discriminant analysis of the morphological differences among 125 *Mystus nemurus* based on 13 morphometric variables corrected for size differences among fish. The three clusters are plotted in the reduced plane of discriminant axes 1 and 2.

76.6% of the intergroup variance and separated the three groups (Fig. 12). The second function further discriminated group one from groups 2 and 3. The total canonical structure revealed that the first function was mainly correlated with AL (–0.78), DA (0.65), DL (0.62), OD (–0.62), PA (0.53) and HA (–0.51). The second function was mainly correlated with DL (0.67), HW (–0.59) and IO (–0.56). Fish of group 2 had a relatively smaller adipose fin, a greater distance between the dorsal and adipose fins, a larger dorsal fin situated closer to the occipital process and a slimmer body than fish of group 3. Fish of group one exhibited larger dorsal fins in conjunction with narrower heads relative to fish of groups 2 and 3.

### *The geographical distribution and genetic identity of morphotypes*

Morphological group 1 was mainly composed of fish from Perak and the Kapuas river loosely related to the mtDNA clade III, and fish from east Sumatra and west Java (Table 4). The haplotypes of all of these fish were poorly resolved in the majority-rule consensus tree. Morphological group 2 was composed mainly of fish characterized by haplotypes 20, 24 and 32, the widely dispersed mtDNA clade III. Morphological group 3 was mainly composed of fish from Thailand (haplotypes 33 and 34) and east peninsular Malaysia (mtDNA clades I and II). Fish from Sarawak, west Borneo (clade IV) contributed equally to the formation of morphological groups 1 and 3.

The majority of the most morphologically distinct fish (group 2) was characterized by mtDNA haplotypes of clade III, the most widely distributed of the genetic groups identified in this study. In addition to the morphometric

**Table 3** The grouping of 125 fish, each characterized by 13 morphometric variables corrected for size differences between fish, into three groups by discriminant analysis. The number of fish (with percentages in parentheses) correctly classified *a posteriori* in their group of origin and the totals are presented

| | To group | | | |
|---|---|---|---|---|
| From group | 1 | 2 | 3 | Total |
| 1 | 57 (95) | 0 (0) | 3 (5) | 60 |
| 2 | 1 (6) | 15 (94) | 0 (0) | 16 |
| 3 | 1 (2) | 0 (0) | 48 (98) | 49 |
| Total | 59 | 15 | 51 | |

**Table 4** The number of *H. nemurus* (percentage) classified in clusters formed according to morphological and genetic characteristics. mtDNA clusters are identified as clades defined in the majority rule consensus tree (I, II, III, IV). Haplotypes whose relationships were not clearly resolved in the majority rule consensus tree were pooled according to their clustering in the FITCH phenograms (Figs 6 and 7). Haplotype 35 was represented by only one fish and was thus pooled with the other, poorly resolved, Sumatran haplotypes. The three morphological clusters were formed according to the classification procedure of a discriminant analysis (Table 3)

| | Morphological clusters | | |
|---|---|---|---|
| mtDNA haplotype clusters | 1 | 2 | 3 |
| group I (2, 9, 4, 7, 8, 10) | 2 (18) | 0 (0) | 9 (82) |
| group II (3, 5, 6, 11) | 4 (25) | 0 (0) | 12 (75) |
| 33, 34 (Thailand) | 0 (0) | 0 (0) | 13(100) |
| group III (20, 32, 24) | 3 (30) | 10 (70) | 0 (0) |
| 12, 13, 23, 21, 22 (Perak, Kapuas) | 12 (75) | 1 (6) | 3 (19) |
| group IV (14, 15, 16, 17, 18 19) | 11 (46) | 1 (4) | 12 (50) |
| 25, 26, 27 (West Java) | 10 (100) | 0 (0) | 0 (0) |
| 28, 29, 30, 31, 35 (Sumatra) | 17 (77) | 3 (14) | 2 (9) |
| total number of fish | 59 | 15 | 51 |

characteristics included in the present analysis, these fish were all characterized by a distinctive black border on the caudal fin. Thus, the morphologically distinct forms of *H. nemurus* found in the Kapuas river, the Johor river and at Palembang appear to be of monophyletic origin. Most other fish of Sumatra, the Kapuas river, west peninsular Malaysia (Perak) and west Java exhibit a similar morphology (morphological group 1) regardless of mtDNA identity.

Morphological group 3 was composed of fish exhibiting a body form characterized by relatively small dorsal fins and large adipose fins. This group was composed mainly of fish from Thailand and the two mtDNA clades (I and II) identified in east peninsular Malaysia. Fish from west Borneo (clade IV) were distributed evenly between morphological groups 1 and 3. Despite their variable morphometric characteristics, differences in reproductive characteristics and coloration distinguished fish of haplotypes 14, 15, 16 and 19 (mtDNA clade IV$^a$) from all others. These fish were golden-brown in colour and sexually mature at 15 cm (SL). All other fish sampled in this study were slate-brown or black in colour and sexually mature at approximately 30 cm (SL).

## Discussion

The high level of haplotype differentiation (mean $P \times 100 = 2.22$, SD = 1.33) indicates extensive subdivision of the ancestral *H. nemurus* group. Assuming that the sequence divergence rate for mtDNA determined for mammals (Brown 1983) of 2% per million years is a maximum estimate for fish, the divergence of clades identified in this study probably occurred early in the Pleistocene. The sympatric occurrence of genetically and morphologically divergent groups of *H. nemurus* in locations presently separated by sea supports the proposition that low sea levels permitted the dispersion and mingling of genetic groups most probably evolved in allopatry. The historical interconnection of the Kapuas river in west Borneo with south-east Sumatra and peninsular Malaysia is clearly demonstrated by the distribution of the morphologically distinct mtDNA group III (haplotypes 20, 32 and 24) (Fig. 7). In particular, the occurrence of haplotype 20 in the Kapuas river, Johor river and at Palembang is striking evidence of recent faunal exchanges among these areas. Group III haplotypes were weakly clustered with a larger group of haplotypes from Perak in west peninsular Malaysia and the Kapuas river in west Borneo. Within this group, the occurrence of endemic haplotypes at Perak (12 and 13) and in the Kapuas river (21, 22 and 23) that differ by only 0.25% sequence divergence is also indicative of dispersal events occurring recently in the Pleistocene.

The sympatric occurrence of two well-defined mtDNA clades (I and II) in east peninsular Malaysia is further evidence of admixture of clades evolved in allopatry. Group I is endemic to the peninsula, possibly isolated during a high sea-level period when the peninsula was isolated from the mainland. Haplotypes 33 and 34 from the Mekong are most closely related to mtDNA group II, suggesting that group II may have been derived from an Indochinese group that invaded peninsular Malaysia as sea levels declined. As the average genetic distance between haplotypes 33 and 34 and group II is 1.3%, faunal exchanges between the Mekong and east peninsular Malaysia probably occurred during the Pleistocene.

The zoogeographic reconstruction based on the distribution of intraspecific genetic groups described in this study indicates that west peninsular Malaysia may have greater faunal affinities with Sumatra whereas east peninsular Malaysia may have greater affinities with the Indochinese fauna. West peninsular Malaysia and east Sumatra were in contact during low sea levels since the end of the Pliocene (de Beaufort 1951). East and west Malaysia are separated by a central range of mountains such that striking faunal differences should exist between the two sides of the Malaysian peninsula. This scenario, however, is not supported by the study of Johnson (1967) who detected no east–west differentiation in the freshwater fish fauna at higher taxonomic levels. As in the case of *H. nemurus*, it may be that east–west differentiation has occurred during the Pleistocene such that differences may only be manifest at the intraspecific level.

The degree of gene flow between sympatric clades, and thus their status as species, is unknown as so little is known of the ecology of the species. Given the morphological similarity of mtDNA groups I and II, found intermingled over a 250-km stretch of east peninsular Malaysia, there is no reason to believe that they are reproductively isolated sibling species. The possibility exists that these two genetic groups may represent the two species previously identified as *H. johorensis* and *H. pahangensis*, but more detailed morphometric and meristic studies will be required to test this proposition.

In contrast, group III, the most widely distributed mtDNA group, is morphologically distinct. There is therefore the possibility that this group is reproductively isolated from the sympatric forms of *H. nemurus* and may represent an undescribed sibling species. Indeed, the black-tailed forms characterized in this study by haplotypes 20, 32 and 24 have recently been proposed as a valid species, *Hemibagrus hoevenii* (Kottelat & Lim, in press). *H. hoevenii* was accepted as valid until Weber and de Beaufort (1913: 341) synonymized it under *H. nemurus*.

The genetic distinctiveness of fish sampled in Sarawak rivers (clade IV) (Fig. 10) indicates that they have long been isolated from other *nemurus* lineages. The earlier age of maturity of the small golden forms belonging to clade IV$^a$ may be indicative of some degree of reproductive iso-

lation from sympatric haplotypes 17 and 18. In the Rajang river, the two forms were sampled in different regions, with the golden forms occurring closer to the estuary and the more morphologically typical *H. nemurus* forms further upstream. Thus, the small golden forms of Sarawak rivers may represent an undescribed sibling species.

The phylogenetic relationships among the groups described in this paper cannot be clearly discerned. The most plausible scenario based on both genetic and morphological evidence is that the main *H. nemurus* line gave rise to three regional groups:

1 an 'Indochinese' group characterized by relatively large adipose fins and small dorsal fins, composed of mtDNA groups I and II overlapping in east peninsular Malaysia and haplotypes 33 and 34 occurring in Thailand;

2 a 'Sundaic' group composed of various lineages that are more or less well-defined morphologically and genetically. The morphologically similar forms from west Java, Sumatra and west Borneo (morphological group 1) show some degree of genetic divergence, but their phylogenetic relationships are poorly resolved in the present analysis. The most distinct Sundaic group is composed of the slender-bodied, black-tailed forms tentatively identified as *H. hoevenii* that has succeeded in colonizing the Kapuas, eastern Sumatra and south peninsular Malaysia;

3 a 'Sarawak' group, mtDNA clade IV composed of fish similar in morphometrics to both the 'Sundaic' and 'Indochinese' groups, but including the golden colour morphs characterized by a small size at sexual maturity (clade $N^a$).

The degree of dispersion of intraspecific genetic groups documented here was less than anticipated based on the widespread similarity of the freshwater fish fauna described at higher taxonomic levels. The observation that 15 of the 21 haplotypes recorded more than once were each collected at only one site illustrates that population isolation has been sufficiently long to permit the evolution of endemic haplotypes. In particular, the absence of clonal lines characteristic of any of the Sunda-island mtDNA groups among the 44 fish sampled in Thailand and east peninsular Malaysia indicates no faunal exchange between these areas during Pleistocene low sea levels. Although more extensive sampling of fish and locations may reveal more extensive admixture, the intraspecific genetic structure of *H. nemurus* and its closely related sibling species indicates very little exchange of genetic groups between the mainland and the Sunda islands since the fragmentation of the main *nemurus* line sometime early in the Pleistocene. The widespread occurrence of *nemurus*-like forms in South-east Asia appears to be the result of colonization events occurring before or early in the Pleistocene rather than the result of dispersal during recent Pleistocene low sea levels.

## Acknowledgements

This study was conducted while the senior author was on sabbatical leave in the Department of Zoology, National University of Singapore. Many thanks to Prof. T.J. Lam, Department Head, for providing access to the facilities of the Department of Zoology; Dr Victor Wong, Zoology, for use of his laboratory for mtDNA extractions and analyses; Dr K. K. Ho for access to facilities in the Department of Botany, NUS. We thank the many people who assisted in obtaining samples: Tang Fook Leong, Wilderness Experience, Petaling Jaya, Malaysia; Thalathiah Saidin, Freshwater Fisheries Research Centre, Malacca, Malaysia; Dr Zakaria Ismail, Department of Zoology, University of Malaya; Gilles Ouellette, Kuala Lumpur, Malaysia; Dr Sutrisno Sukimin, SEAMEO-BIOTROP, Bogor, Indonesia; Dr Kadarwan Soewardi, Fakultas Perikanan, Institut Pertanian Bogor, Bogor, Indonesia; Sim Yau Tong and Dominic Yeo, Kutching, Sarawak, Malaysia; Shawn Lum and Daniel Metcalfe, Botany, NUS; Hady Wirawan, Pontianak, West Kalimantan, Indonesia: Ben Malik and Dra. Asnawati, Dinas Perikanan, Jambi, Sumatra, Indonesia: Zainal Arifin, Dinas Perikanan, Palembang, Sumatra, Indonesia. Many thanks to Fernande Couturier who helped me to collect and dissect specimens throughout the region; Sharon Thomas and Mei Fun Choong at NUS for morphological measurements; Bun Tuck Weng and Teih Ling at NUS and Edith Bonnneau at Université Laval for help with mtDNA extractions and analysis; M. C. Baby at Université Laval for help with phylogenetic reconstruction; Kelvin Lim and the staff of the Zoological Reference Collection at NUS; Gaétan Daigle of the Service de consultation statistique, Université Laval. Many thanks to Drs Maurice Kottelat and Tyson Roberts for informative discussions concerning the freshwater fishes of South-east Asia and to Drs L. Bernatchez, M. Kottelat and three anonymous referees for providing critical reviews of the paper. This study was funded by Université Laval sabbatical funds and the Natural Sciences and Engineering Research Council of Canada.

## References

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Reviews in Ecology and Systematics*, **18**, 489–522.

Avise JT (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**, 1192–1208.

Avise JC, Alisauskas RT, Nelson WS, Ankney CD (1992) Matriarchal population genetic structure in an avian species with female natal philopatry. *Evolution*, **46**, 1084–1096.

Bernatchez L, Dodson JJ (1994) Phylogenetic relationships among palearctic and nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **51** (Suppl 1), 240–250.

Bernatchez L, Dodson JJ (1990) Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) revealed by mitochondrial DNA restriction analysis. *Evolution*, **44**, 1263–1271.

Bernatchez L, Savard L, Dodson JJ, Pallotta D (1988) Mitochondrial DNA sequence heterogeneity among James-Hudson Bay anadromous coregonines. *Finnish Fisheries Research*, **9**, 17–26.

Bernatchez L, Dodson JJ, Boivin S (1989) Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. *Journal of Fish Biology*, **35** (Supplement A), 233–244.