The LINNEAN SOCIETY of London ZOOLOGICAL Journal of the Linnean Society

Zoological Journal of the Linnean Society, 2014, **172**, 556–615. With 15 figures

Miniatures, morphology and molecules: *Paedocypris* and its phylogenetic position (Teleostei, Cypriniformes)

RALF BRITZ[1]*, KEVIN W. CONWAY[2] and LUKAS RÜBER[3]

[1]*Department of Zoology, Natural History Museum, Cromwell Road, London SW7 5BD, UK*
[2]*Department of Wildlife and Fisheries Sciences and Biodiversity Research and Teaching Collections, Texas A&M University, College Station, TX 77843, USA*
[3]*Naturhistorisches Museum der Burgergemeinde Bern, Bernastrasse 15, 3005 Bern, Switzerland*

Received 23 September 2013; revised 13 June 2014; accepted for publication 26 June 2014

We review the morphological and molecular evidence that Mayden & Chen recently used to infer that the developmentally truncated fish genus *Paedocypris* is not a member of the teleost order Cypriniformes or carp-like fishes, but is 'the basal sister group to all Cypriniformes'. This hypothesis contradicts several previous studies that used molecular sequence data or morphological characters. A review of the morphological characters that Mayden & Chen discussed and mapped onto their 'simplified tree' shows that these, analysed alone, rather support a close relationship of the cyprinids *Sundadanio*, *Danionella*, and *Paedocypris*. We also present four additional analyses of morphological data, which all contradict Mayden & Chen's result. Despite its highly reductive skeleton, posing a serious problem when analysing its phylogenetic position with skeletal characters, the presence in *Paedocypris* of the basioccipital masticatory plate is compelling evidence that it is a member of the Cyprinoidei (Cyprinidae plus Psilorhynchidae). Our reanalysis and exploration of their molecular sequence data shows that only a single gene, *EGR3*, of the six nuclear genes analysed by Mayden & Chen, is responsible for the position of *Paedocypris* as 'the basal sister group to all Cypriniformes'. Three independent methods to visualize and analyse phylogenetic signal and conflict of data sets (phylogenetic networks, splits analysis methods or SAMS, and site-wise likelihood analyses) reveal a high level of character conflict and noise in Mayden & Chen's data set. The 'basal' position of *Paedocypris* seems to be the outcome of the interplay of two long-branch effects. We apply the same analytical methods to the data set from Rüber *et al.*'s molecular analysis of the phylogenetic position of *Paedocypris* and discuss our findings. We conclude that none of the molecular data sets compiled to date can establish the phylogenetic position of *Paedocypris* with confidence. Morphological data suggest that *Paedocypris* and *Danionella* are sister genera, and that their closest relative is *Sundadanio*, although the position of these three miniatures among cyprinoids is still unclear.

doi: 10.1111/zoj.12184

ADDITIONAL KEYWORDS: *a priori* data exploration – Cypriniformes – developmentally truncated miniatures – long-branch attraction – miniaturization – molecular data quality – paedomorphosis – progenesis – proportioned dwarfs – Teleostei – zebrafish.

INTRODUCTION

Developmental truncation or progenesis is a remarkable form of heterochrony, in which sexual maturation is greatly accelerated in relation to somatic development (see Gould, 1977 for a review). Progenesis commonly leads to sexually mature, dwarfed organisms with a large number of larval or juvenile features and, along with neoteny, is one of the two types of paedomorphosis (Gould, 1977; Hanken & Wake, 1993). Developmentally truncated, or progenetic, miniatures have been notoriously difficult to place phylogenetically because truncation may affect large parts of the morphological character set normally used to place a taxon in a phylogenetic context (Hanken & Wake, 1993). Such taxa are often considered systematic enigmata and are frequently classified in

*Corresponding author. E-mail: r.britz@nhm.ac.uk

high-level taxonomic categories, like the tiny invertebrates *Symbion* and *Xenoturbella*, each the sole members of their own phyla, the Cycliophora (Funch & Christensen, 1995) and Xenoturbellida (Bourlat *et al*., 2006), respectively.

Prime examples of enigmatic miniature taxa in the vertebrate world are the progenetic goby *Schindleria* (see Johnson & Brothers, 1993) or the developmentally truncated clupeomorph *Sundasalanx* (Siebert, 1997). Recently, *Paedocypris* (Kottelat *et al*., 2006), a remarkable genus of the teleost order Cypriniformes, was added to this group of progenetic teleosts.

With more than 4000 species, the Cypriniformes is one of the largest groups of bony fishes, dominating the freshwaters of North America, Africa, and Eurasia. The order Cypriniformes is a well-supported monophyletic group with numerous autapomorphies (see Fink & Fink, 1981, 1996), and includes both very large species, reaching up to 2 m in length (Howes, 1991), and very small forms, maturing at under 10 mm in length (Kottelat *et al*., 2006). Especially in this miniature size class, more than a dozen remarkable new species have been discovered over the last 10 years (Kottelat *et al*., 2006; Britz & Kottelat, 2008; Britz, 2009; Britz, Conway & Rüber, 2009; Conway, Kottelat & Tan, 2011; Britz, Kottelat & Tan, 2012). Among these, the genus *Paedocypris* has received considerable attention because it includes some of the smallest fishes and vertebrates (Kottelat *et al*., 2006). *Paedocypris* comprises three species of tiny fishes with a maximum standard length of 10–12 mm (Kottelat *et al*., 2006; Britz & Kottelat, 2008) that meet Weitzman & Vari's (1988) definition of miniature fishes as fishes not known to exceed 26 mm in standard length. Rüber *et al*. (2007) and Britz & Conway (2009) distinguished two types among miniature cypriniform fishes: proportioned dwarfs and developmentally truncated miniatures. Proportioned dwarfs are tiny but almost identical copies of their larger relatives, whereas developmentally truncated miniatures resemble the larval stages of their relatives. Britz & Conway (2009) studied the skeleton of *Paedocypris* in detail and found that 61 skeletal (bony and cartilaginous) structures that are commonly present in cypriniforms and other teleosts either fail to develop or are greatly reduced in size or complexity. *Paedocypris* is thus a developmentally truncated miniature, whereas *Boraras*, another miniature Asian cypriniform in the same size class, is an example of a proportioned dwarf (Rüber *et al*., 2007; Britz & Conway, 2009). The mostly larval skeleton in the miniature and developmentally truncated species of *Paedocypris* is associated with remarkable evolutionary morphological novelties, a highly unusual combination among bony fishes (Kottelat *et al*., 2006; Rüber *et al*., 2007; Britz & Conway, 2009).

Based on the analysis of cytochrome *b* (cyt *b*) gene sequence data from 227 cypriniform taxa, Rüber *et al*. (2007) hypothesized a sister-group relationship of *Paedocypris* and *Sundadanio*, another genus of miniature cyprinids from Southeast Asia (Conway *et al*., 2011). Both genera were recovered in the cyprinid subfamily Danioninae (Rasborinae clade A of Rüber *et al*., 2007), a large group composed predominantly of South and Southeast Asian minnows, including the zebrafish, *Danio rerio* (Hamilton, 1822). In their anatomical analysis of the skeleton of *Paedocypris*, Britz & Conway (2009) concluded that this genus is most closely related to *Danionella* (a genus of miniature cyprinids from Indo-Burma; Britz, 2009), with *Sundadanio* as the sister group of the two. These two independent sources of data, morphological and molecular, thus showed some level of consensus as to where *Paedocypris* belongs among Cyprinidae. Subsequent molecular analyses by Fang *et al*. (2009) and Tang *et al*. 2010, 2011: fig. 2) have supported the results of Rüber *et al*. (2007).

In a recent molecular paper based on the analysis of nucleotide sequences from six nuclear genes, Mayden & Chen (2010: 152) published the highly surprising and unexpected hypothesis that *Paedocypris* is not closely related to either *Danionella* or *Sundadanio*, and is actually not even a cypriniform, but 'the basal sister group to all Cypriniformes'. In addition to the result of their phylogenetic analysis, they argued that the position of *Paedocypris* (p. 152) 'as the basal sister group to all Cypriniformes' is 'also supported by previously proposed morphological characters', in reference to the characters listed in Britz & Conway (2009).

In the present paper we review the molecular and morphological evidence presented by Mayden & Chen (2010) in support of their hypothesis that *Paedocypris* is not a cypriniform, and discuss their claims in light of our findings.

MATERIAL AND METHODS

Morphological data and phylogenetic analyses

We used the morphological characters and taxon sampling of Britz & Conway (2009), Mayden & Chen (2010), and Conway (2011) to compile lists of characters and character states (see Appendix 1), from which five morphological data matrices were derived (matrices 1–5; see Appendices 2–6). The first data matrix (matrix 1, Appendix 2) mirrored the morphological characters and taxon sampling of Mayden & Chen (2010), including 68 characters for 81 cypriniform and four outgroup taxa. Characters in this matrix were coded exactly as presented in Mayden & Chen's table 1 (pp. 158–161) and figure 3 (p. 163). For example, Mayden & Chen's character 1 ('Weberian Apparatus') was coded as present (character state 1) for all taxa, excluding the

outgroup taxon *Gonorynchus greyi* Richardson, 1845 (which was coded with the character state 0). Mayden & Chen's character 2 ('Bony kinethmoid element') was coded as absent (character state 0) for all outgroup taxa, coded as present but cartilaginous (character state 1; equivalent to Mayden & Chen's character 2, state A) for *Paedocypris*, and coded as present and ossified (character state 2; equivalent to Mayden & Chen's character 2, state B) for all remaining cypriniform taxa in the data matrix, including *Danionella* (even though the kinethmoid element is absent in this taxon; Roberts, 1986; Britz *et al*., 2009), following Mayden & Chen's placement of their character 2, state B, at the node uniting all Cypriniformes but *Paedocypris* in their figure 3. Mayden & Chen's character 3 ('Fifth ceratobranchial enlarged extending further dorsally than other ceratobranchials') was coded as present (character state 1; equivalent to Mayden & Chen's character 3, state A) for all cypriniform taxa, and coded as absent (character state 0) for all outgroup taxa, following Mayden & Chen's placement of their character 3, state A, at the node uniting all Cypriniformes in their figure 3. Using this procedure, the final data matrix (Appendix 2) consisted of 60 binary characters and eight multistate characters (each with three character states; 0, 1, or 2) for 85 taxa. We continue to recognize Psilorhynchidae for *Psilorhynchus* (as in Mayden & Chen, 2010, fig. 3, but not fig. 2), because there are currently two contradictory hypotheses about its phylogenetic position (e.g. see Tang *et al*., 2010 with *Psilorhynchus* as the sister group to the Cyprinidae versus Tang *et al*., 2011, figs 2, 3, with *Psilorhynchus* as the sister group to cyprinine Cyprinidae).

The second data matrix (matrix 2, Appendix 3) that we compiled also mirrored the morphological characters and taxon sampling of Mayden & Chen (2010). Characters in this matrix were, however, coded from observations on actual specimens that were cleared and stained for bone and cartilage investigation following the protocol of Taylor & Van Dyke (1985), dissected following the protocol of Weitzman (1974), and examined using a Zeiss Discovery V20 stereomicroscope. Of the 85 taxa investigated by Mayden & Chen (2010) we were able to obtain specimens for only 45 (Table 1). For the remaining 40 taxa, we were able to obtain specimens of close relatives (members of the same genus or hypothesized sister genera) for 34 (which we 'substituted' for Mayden & Chen's taxa; Table 1), but could not locate appropriate substitutes for the remaining six, which were excluded from this matrix and its subsequent analysis. During the process of character coding we noticed several issues with Mayden & Chen's characters, including the inaccurate coding of characters (characters 9, 22, 24, and 27) and even the duplication of characters, including character 31 (identical to character 17), 37 (identical to character 30), 48 (identical to both characters 17 and 31), 51 (overlaps with character 32), 52 (identical to character 33), 53 (overlaps with character 32), and 55 (identical to character 35). To overcome these issues, we reverted the coding for characters 9 ['ectopterygoid overlapping/separated from the palatine (= autopalatine) anteriorly'], 24 ('infraorbital 5 in contact with/separated from the supraorbital'), and 27 ('anterior opening of the trigeminofacial chamber contained within prootic/positioned between prootic and pterosphenoid') to reflect their original usage, as either derived characteristics of Cypriniformes (*sensu* Fink & Fink, 1981) or Cyprinidae (*sensu* Cavender & Coburn, 1992), and removed all duplicate characters, which resulted in a final matrix of 61 characters for 79 taxa (Appendix 3). Only a single multistate character (character 2; with three character states: 0, 1, or 2) is present in this matrix (versus eight in matrix 1) as we saw no reason to arbitrarily inflate the number of states for characters 7 and 10, to follow the erroneous homology assumption for character 19 (see discussion below), or to use different coding for the same state of characters 42–45 in *Sundadanio* (a method adopted in Mayden & Chen's fig. 3).

The third data matrix (matrix 3, Appendix 4) combined the characters and taxon sampling schemes of Britz & Conway (2009) and Conway (2011). We first modified the character matrix of Conway (2011) by restricting the taxon sampling to include only a single species of the genus *Psilorhynchus*, *Psilorhynchus balitora* (Hamilton, 1822), and by restricting the character set to remove characters that were applicable only to these excluded species (character 116) or characters that became autapomorphies of *P. balitora* (characters 9, 33, 85, 103, 105, and 106). We modified this matrix further by adding taxa from Britz & Conway (2009), including all miniature taxa [except for *Rasbora spilocerca* Rainboth & Kottelat, 1987, for which material was unavailable] and non-miniature taxa listed in their table 1. This resulted in a final data matrix of 121 characters for 65 taxa (Appendix 4).

The fourth data matrix (matrix 4, Appendix 5) that we compiled followed the taxon sampling of matrix 3 with the addition of 23 characters from Britz & Conway (2009), including 18 'absence' and five 'progressive' characters. This resulted in a final data matrix of 144 characters for 65 taxa (Appendix 5).

The fifth and final morphological data matrix (matrix 5, Appendix 6) mirrored the taxon sampling of matrices 3 and 4. In this matrix, we excluded 11 absence characters included in matrices 3 and 4 that were shown by Britz & Conway (2009) to be either linked to miniaturization or linked to organism-wide developmental truncation to assess their effect on the phylogenetic placement of the highly miniaturized

Table 1. List of taxa included in the molecular phylogenetic investigation of Mayden & Chen (2010) and material examined herein and used to construct matrix 2 (Appendix 3)

Order	Suborder	Family	Original taxon	Substitute taxon	Specimen voucher number(s)
Gonorynchiformes		Gonorynchidae	*Gonorynchus greyi* outgroup	–	BMNH 1984.11.30.3
Characiformes		Characidae	*Chalceus macrolepidotus* outgroup	–	BMNH 1972.7.27.922 AMNH 38183
Characiformes		Alestidae	*Phenacogrammus interruptus* outgroup	–	AMNH 227592; TU 200482
Characiformes		Bagridae	*Pseudobagrus tokiensis* outgroup	–	BMNH uncat.
Cypriniformes	Cobitoidei	Catostomidae	*Catostomus commersoni*	–	UAIC 14172.01
	Cobitoidei	Catostomidae	*Cycleptus elongatus*	–	UAIC 14922.01
	Cobitoidei	Gyrinocheilidae	*Gyrinocheilus aymonieri*	–	BMNH 2000.6.11.3518–3555
	Cobitoidei	Gyrinocheilidae	*Gyrinocheilus pennocki*	–	UAIC 14180.51
	Cobitoidei	Botiidae	*Botia dario*	*Botia almorhae*	BMNH 1872.4.17.27
	Cobitoidei	Botiidae	*Leptobotia pellegrini*	*Leptobotia fasciata*	BMNH 1889.6.8.64–75
	Cobitoidei	Botiidae	*Syncrossus beauforti*	*Syncrossus hymenophysa*	BMNH 13490
	Cobitoidei	Botiidae	*Yasuhikotakia morleti*	–	UAIC 14181.05
	Cobitoidei	Cobitidae	*Acantopsis* sp.	*Acantopsis dialuzona*	CAS 27927
	Cobitoidei	Cobitidae	*Cobitis takatsuensis*	*Cobitis dalmatica*	BMNH uncat.
	Cobitoidei	Cobitidae	*Niwaella multifasciata*	–	UAIC 14183.15
	Cobitoidei	Cobitidae	*Pangio oblonga*	*Pangio muraeniformes*	BMNH 1970.9.3.138–167
	Cobitoidei	Cobitidae	*Canthophrys gongota*	–	BMNH 1872.4.17.4
	Cobitoidei	Balitoridae	*Homaloptera parclitella*	*Homaloptera* sp.	BMNH uncat.
	Cobitoidei	Balitoridae	*Sewellia lineolata*	–	UAIC 14169.43
	Cobitoidei	Nemacheilidae	*Acanthocobitis* sp.	–	UAIC 14166.01
	Cobitoidei	Nemacheilidae	*Barbatula barbatula*	–	AMNH 37608
	Cobitoidei	Nemacheilidae	*Lefua costata*	–	UAIC 14183.18
	Cobitoidei	Nemacheilidae	*Oreonectes platycephalus*	*Oreonectes* sp.	BMNH uncat.
	Cobitoidei	Nemacheilidae	*Schistura savona*	*Schistura balteata*	TCWC uncat.
	Cobitoidei	Nemacheilidae	*Traccatichthys pulcher*	–	TCWC 16419.01
	Cobitoidei	Nemacheilidae	*Triplophysa gundriseri*	*Triplophysa microps*	BMNH 1889.2.1.1709–1713
	Cobitoidei	Vaillantellidae	*Vaillantella maassi*	*Vaillantella euepiptera*	BMNH 2000.10.25.1301–1332
	Cyprinoidei	Cyprinidae	*Acheilognathus tabira*	*Acheilognathus typus*	CBM-ZF 11423
	Cyprinoidei	Cyprinidae	*Acrossocheilus paradoxus*	–	CBM-ZF 11426
	Cyprinoidei	Cyprinidae	*Aphyocypris chinensis*	–	CBM-ZF 11424
	Cyprinoidei	Cyprinidae	*Aspidoparia morar*	Not available	–
	Cyprinoidei	Cyprinidae	*Barbonymus gonionotus*	*Barbonymus schwanfeldii*	TU 199857
	Cyprinoidei	Cyprinidae	*Barbus callipterus*	*Barbus leonensis*	BMNH 1974.9.18.77–117
	Cyprinoidei	Cyprinidae	*Barilius bendelisis*	–	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Biwia zezera*	Not available	–
	Cyprinoidei	Cyprinidae	*Boraras merah*	–	BMNH 2004.4.26.22–29
	Cyprinoidei	Cyprinidae	*Chela dadiburjori*	*Chela laubuca*	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Danio albolineatus*	–	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Danio dangila*	–	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Danio erythromicron*	–	UAIC 14166.23
	Cyprinoidei	Cyprinidae	*Danio margaritatus*	–	BMNH 2007.10.9.15–16
	Cyprinoidei	Cyprinidae	*Danio rerio*	–	BMNH 2001.3.12.76–92
	Cyprinoidei	Cyprinidae	*Danionella* sp.	*Danionella dracula*	BMNH 2008.1.1.100–119
	Cyprinoidei	Cyprinidae	*Devario regina*	*Devario devario*	BMNH 1889.2.1.1235-8
	Cyprinoidei	Cyprinidae	*Esomus longimanus*	*Esomus caudiocellatus*	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Garra spilota*	*Garra flavatra*	BMNH uncat.

Table 1. *Continued*

Order	Suborder	Family	Original taxon	Substitute taxon	Specimen voucher number(s)
	Cyprinoidei	Cyprinidae	*Gobio gobio*	–	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Gymnocypris przewalskii*	*Schizothorax grahami*	BMNH 1914.1.28:26–35
	Cyprinoidei	Cyprinidae	*Hampala macrolepidota*	–	BMNH 2001.3.13.297–317
	Cyprinoidei	Cyprinidae	*Hemibarbus barbus*	–	USNM 45234
	Cyprinoidei	Cyprinidae	*Horadandia atukorali*	–	USNM 271483
	Cyprinoidei	Cyprinidae	*Ischikauia steenackeri*	–	CBM-ZF 11396
	Cyprinoidei	Cyprinidae	*Labeo chrysophekadion*	*Labeo* cf. *longipinnis*	UAIC 14180.55
	Cyprinoidei	Cyprinidae	*Leptobarbus hoevenii*	–	TCWC 16417.01
	Cyprinoidei	Cyprinidae	*Luciosoma setigerum*	*Luciosoma* sp.	BMNH 2001.3.13.109–121
	Cyprinoidei	Cyprinidae	*Macrochirichthys macrochirus*	–	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Megalobrama amblycephala*	Not available	–
	Cyprinoidei	Cyprinidae	*Microdevario kubotai*	–	BMNH 2004.6.25.6–10
	Cyprinoidei	Cyprinidae	*Microdevario nana*	–	BMNH 2004.6.25.1–5
	Cyprinoidei	Cyprinidae	*Microrasbora rubescens*	–	BMNH 2004.6.25.11–13
	Cyprinoidei	Cyprinidae	*Notropis baileyi*	*Notropis stramineus*	TCWC 7509.09
	Cyprinoidei	Cyprinidae	*Opsariichthys uncirostris*	–	BMNH 1902.5.30.53–54
	Cyprinoidei	Cyprinidae	*Paedocypris* sp. 1	*Paedocypris* sp.	BMNH 2008.4.11.2–11
	Cyprinoidei	Cyprinidae	*Paedocypris* sp. 2	*Paedocypris* sp.	BMNH 2008.4.11.2–11
	Cyprinoidei	Cyprinidae	*Paracheilognathus himantegus*	–	USD(CTOL)2685
	Cyprinoidei	Cyprinidae	*Paralaubuca typus*	–	CMK 12292
	Cyprinoidei	Cyprinidae	*Pelecus cultratus*	–	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Phoxinus percnurus sachalinensis*	*Phoxinus* sp.	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Puntius titteya*	–	TCWC 16418.01
	Cyprinoidei	Cyprinidae	*Rasbora bankanensis*	*Rasbora pauciperforata*	BMNH 1995.5.17.58–97
	Cyprinoidei	Cyprinidae	*Rasbora steineri*	*Rasbora trilineata*	BMNH 2000.10.25.72–116
	Cyprinoidei	Cyprinidae	*Rhodeus ocellatus kurumeus*	*Rhodeus ocellatus*	ROM 69344; USNM 191394
	Cyprinoidei	Cyprinidae	*Romanogobio ciscaucasicus*	Not available	–
	Cyprinoidei	Cyprinidae	*Sarcocheilichthys parvus*	Not available	–
	Cyprinoidei	Cyprinidae	*Scardinius erythropthalmus*	*Leuciscus cephalus*	BMNH 1970.10.14:16–28
	Cyprinoidei	Cyprinidae	*Semotilus altromaculatus*	–	TU 67150
	Cyprinoidei	Cyprinidae	*Sqaulidus chankaensis*	*Squalidus gracilis*	CMK 21104
	Cyprinoidei	Cyprinidae	*Sundadanio axelrodi*	*Sundadanio echinus*	BMNH 1982.3.29.50–54
	Cyprinoidei	Cyprinidae	*Tanichthys albonubes*	–	BMNH uncat.
	Cyprinoidei	Cyprinidae	*Tinca tinca*	–	BMNH 1983.4.20.21–40
	Cyprinoidei	Cyprinidae	*Trigonostigma heteromorpha*	*Trigonostigma hengeli*	BMNH 2004.6.25.18.21
	Cyprinoidei	Cyprinidae	*Yaoshanicus arcus*	Not available	–
	Cyprinoidei	Cyprinidae	*Zacco sieboldii*	–	CBM-ZF 11165
	Cyprinoidei	Psilorhynchidae	*Psilorhynchus sucatio*	–	CAS 50289
	Cyprinoidei	Psilorhynchidae	*Psilorhynchus homaloptera*	*Psilorhynchus balitora*	UMMZ 244782

Taxa listed under 'Substitute Taxon' include those that served as substitute taxa for Mayden & Chen's original taxa in our investigation. Voucher numbers refer to cleared and double-stained material. Museum abbreviations: AMNH, American Museum of Natural History, New York, USA; BMNH, Natural History Museum, London, UK; CAS, California Academy of Sciences, San Francisco, California, USA; CBM-ZF, Natural History Museum and Institute, Chiba, Japan; CMK, Collection of Maurice Kottelat, Cornol, Switzerland; ROM, Royal Ontario Museum, Toronto, Canada; TCWC, Biodiversity Research and Teaching Collections of Texas A&M University, College Station, Texas, USA; TU, Tulane University Biodiversity Research Institute, New Orleans, Louisiana, USA; UAIC, University of Alabama Ichthyology Collection, Tuscaloosa, Alabama, USA; UMMZ, University of Michigan Museum of Zoology, Ann Arbor, Michigan, USA; USD(CTOL), Collection of Paula Mabee, University of South Dakota, Vermillion, South Dakota, USA; USNM, National Museum of Natural History, Smithsonian Institution, Washington D.C., USA.

and developmentally truncated *Paedocypris*. These comprised the following characters, included in both matrices 3 and/or 4: absence of intercalar (character 20 in matrices 3 and 4), absence of hypobranchial 2 (73 in matrices 3 and 4), absence of interhyal cartilage and bone (87 in matrices 3 and 4), absence of hypural 6 (116 in matrices 3 and 4), absence of nasal (122 in matrix 4), absence of extrascapular (127 in matrix 4), absence of sclerotics (128 in matrix 4), absence of infraorbital bones posterior to first (129 in matrix 4), absence of hypobranchial 1 (132 in matrix 4), absence of mesocoracoid (133 in matrix 4), and absence of uroneural 2 (135 in matrix 4). This resulted in a final matrix of 133 characters for 65 taxa (Appendix 6).

All five morphological data matrices were assembled using TEXTWRANGLER 2.3 (Bare Bones Software) and finalized in MacClade 4.05 (Maddison & Maddison, 2002). Question marks (?) in matrices 2–5 indicate that a particular character could not be scored because it was not observable in the material examined, and a dash (–) indicates that a particular character could not be scored because of inapplicability. Each matrix was analyzed using the principles of parsimony analysis, as implemented in PAUP* (Swofford, 2000). Equally parsimonious cladograms were obtained using a heuristic search with a random-addition sequence (ten replicates) and tree bisection and reconnection (TBR) branch swapping. All characters were assigned equal weight and were left unordered. The maximum number of trees saved (Maxtrees) was allowed to automatically increase by 100 during each analysis. Branches were collapsed (creating polytomies) when branch lengths were equal to zero and the 'Multrees' option was 'in effect'. Searches that continued to recover multiple islands of trees after ten random additions were repeated using a parsimony ratchet (Nixon, 1999), as implemented in PAUPRat (Sikes & Lewis, 2001; via the CIPRES science gateway portal, http://www.phylo.org), using the following search parameters: 200 ratchet 'reps', '-pct' set to 20, weighting mode set to 'uniform', output set to 'terse', and run for ten replications ('nreps' = 10). Recovered cladograms were rooted on *Elops* sp. cf. *senegalensis* Regan, 1909 and a strict consensus method was used to summarize common information across the equally parsimonious cladograms recovered. Character polarity was determined by outgroup comparison (Maddison, Donoghue & Maddison, 1984) with *Elops*, and both the ACCTRAN and DELTRAN modes of optimization were used to investigate character optimization in the resulting cladograms. Lists of character-state changes along branches for both optimization modes and the different trees are provided in Figures S1–S4.

Bootstrapping was employed to obtain statistical support for recovered clades. Bootstrap values are based on 1000 replicates, conducted using a fast heuristic search, and with starting trees obtained via stepwise addition. Bootstrap and strict consensus trees were viewed and converted into graphic files using FigTree 1.3.1 (http://tree.bio.ed.ac.uk/).

Reductive characters as synapomorphies

Members of the genera *Paedocypris* and *Danionella*, and to some extent those of *Sundadanio*, lack an unusually high number of skeletal characters that are otherwise present at the level of cypriniforms and/or teleosts. When establishing the phylogenetic position of *Paedocypris* among cypriniforms, we were faced with the problem of having a taxon with an exceptionally large number of absence characters contrasted by few skeletal presence characters in this miniature fish. In terms of hypothesizing a 'one to one relationship' (Rieppel & Kearney, 2002; Britz & Johnson, 2011) between two characters as a prerequisite to establish their homology, absence characters pose a real problem, as their similarity cannot be evaluated in the same way as that for a structure that is present. For example, the absence of the sixth hypural in *Paedocypris* and *Danionella* is not open to the application of the 'homology criteria' of Remane (1952). Nelson (1978) argued that 'the absence of a character is not a character'. Nevertheless, reductive (absence) characters have generally been considered admissible as putative synapomorphies since Hennig (1966: 95), citing the absence of wings in fleas, stated that '. . . a "character" may also be the absence of an organ'. And Hennig (1966: 95) admitted absence characters as evidence for relationship when he noted that '. . . the absence of wings in the Anoplura and Mallophaga is a synapomorphous character . . .'. Previous phylogenetic studies that dealt specifically with reductive characters in fishes acknowledged the potential problems surrounding the use of reductive characters, but stressed that absences are characters, and thus putative synapomorphies (Weitzman & Fink, 1985: 10, 'We also use simple loss characters . . . as a synapomorphy . . .'; Schaefer, Weitzman & Britski, 1989: 204–205, 'The central problem in all cases is to determine at what level an apomorphic reductive character can be interpreted as synapomorphic'; Begle, 1991: 35, 'In an important sense, losses are merely transformations, and like gains, may be unique at a particular level of analysis, that is, perfectly congruent with the other characters . . .'; Winterbottom, 1991: 256, 'Six of these synapomorphies involve the lack of ossification in various structures in the latter two species').

Although we agree in principle with Weitzman & Fink's (1985), Schaefer, Weitzman & Britski's (1989), Begle's (1991), and Winterbottom's (1991) conclusions, we were still concerned that the unusually high number of absence characters in *Paedocypris* would

influence our phylogenetic analysis and favour a monophyletic grouping of taxa that share the largest number of absences, independent of their actual relationships. In order not to fall into this 'homoplasy trap' we needed to evaluate the absence characters. We concluded that absences that occur in miniature but not closely related cyprinids (as for example the shared absence of hypural 6 in the rasborin *Boraras*, the danionin *Danionella*, and the cyprinin *Sawbwa*, see Britz & Conway, 2009) would be of little use, as they would be the result of independent events of miniaturization. We also decided that absences that could be easily explained as straightforward cases of developmental truncation would also be of little significance (see also Winterbottom, 1991). Characters in this category are the absence of those bones in miniatures that ossify late in the developmental sequence of *D. rerio* (see Fig. 2). The procedure to evaluate the different absence characters in this way is explained further in Britz & Conway (2009), and in the Results section below.

Molecular data, phylogenetic analyses, and data exploration

Mayden & Chen (2010) analysed nucleotide sequence data of six nuclear genes for 81 cypriniform and four outgroup taxa using partitioned maximum-likelihood (ML) and partitioned Bayesian inference. To further explore their data, sequences from the Mayden & Chen (2010) data set were retrieved from GenBank. We noted that the six sequences (EU292700, FJ531261, FJ531290, FJ531319, FJ531347, and FJ531367) for Mayden & Chen's (2010) '*Danionella* sp.' are listed in GenBank as belonging to '*Danionella mirifica* Britz, 2003'. Tang *et al*. (2010: 191) mentioned that 'Reexamining the material used in Conway, Chen & Mayden (2008) showed that the voucher specimen for the taxon identified as "*Danionella mirifica*" (CBM-ZF-11312) is currently an undescribed species of *Danionella*, labelled as *Danionella* sp. "India" in this study. Sequences from this individual are represented by GenBank numbers EU292700, FJ531261, FJ531290, FJ531319, FJ531347, and FJ531367'. The only *Danionella* species known from India is *Danionella priapus* Britz, 2009. Because we are uncertain of the species identity of this sample, we refer to it as *Danionella* sp. in our study. Each of the six nuclear gene data sets was assembled individually in GENEIOUS 5.4 (Drummond *et al*., 2011). In addition, the data set of Rüber *et al*. (2007) that is based on complete cyt *b* sequences for 227 cypriniform taxa, including 213 cyprinids, was also further explored.

Phylogenetic analyses were performed under the ML criterion, as implemented in RAxML 7.0.4 (Stamatakis, 2006) using the GTRGAMMA model, and conducting both rapid bootstrap analysis (1000 replicates) and search for the best-scoring ML tree in one single run (option –fa). For the Mayden & Chen (2010) data, each gene was analysed separately with (by codon position) and without data partitioning (50% majority rule consensus trees are provided in Figs S5 and S6). In addition a concatenated data set (six genes) was analysed, again with [six partitions (by gene) and 18 partitions (by gene and codon position)] and without data partition. By excluding *EGR3* from Mayden & Chen's (2010) dataset, we also analysed a restricted five-gene data set with [five partitions (by gene) and 15 partitions (by gene and codon position)] and without data partition, so that we could check independently its influence on the concatenated six-gene data set (Fig. S7). The Rüber *et al*. (2007) data set was analysed with (by codon position) and without data partition. We conducted three independent RAxML runs for each data set and partition strategy to check for the repeatability of results. As results for each data set and partition strategy were highly congruent, we only report those with the highest log-likelihood scores.

Three alternative tree topologies were statistically evaluated for each data set using the approximately unbiased (AU; Shimodaira, 2002), Kishino–Hasegawa (KH; Kishino & Hasegawa, 1989), and Shimodaira–Hasegawa (SH; Shimodaira & Hasegawa, 1999) tests, as implemented in CONSEL 0.1k (Shimodaira & Hasegawa, 2001) and one million multiscale bootstrap replicates. Site-wise log-likelihood values were calculated in RAxML 7.0.4. The three alternative tree topologies evaluated were: (1) *Paedocypris* as sister group to the remaining Cypriniformes (hypothesis in Mayden & Chen 2010); (2) *Paedocypris* as sister group to the remaining Cyprinidae plus Psilorhynchidae; and (3) monophyly of Danioninae with inclusion of *Paedocypris* (hypothesis in Rüber *et al*. 2007).

To assess the influence of particular sites on competing hypotheses we conducted a per-site likelihood analysis (Evans *et al*., 2010). The topologies compared were: (1) *Paedocypris* as sister group to the remaining Cypriniformes (hypothesis in Mayden & Chen, 2010) versus *Paedocypris* as sister group to the remaining Cyprinidae plus Psilorhynchidae; and (2) *Paedocypris* as sister group to the remaining Cypriniformes versus monophyly of Danioninae with the inclusion of *Paedocypris* (hypothesis in Rüber *et al*., 2007). For each topology, the per-site log-likelihood (psln L) values were recovered in RAxML. Then, for each comparison (a and b; see above), the difference in psln L (= Δpsln L) for the two topologies was calculated and visualized. Again, these analyses were conducted on the concatenated data sets of both Mayden & Chen (2010) and Rüber *et al*. (2007).

We used SplitTrees 4.11.3 (Huson & Bryant, 2006) to calculate phylogenetic networks based on the

neighbor-net algorithm. Both logDet (corrects for bias in base composition) and p-distance transformations were evaluated. The resulting network graphs give a first indication of signal-like patterns and conflict present in the alignments, and hence aid in analysing and visualizing information content in a nucleotide alignment (Huson & Bryant, 2006). The concatenated data sets of Mayden & Chen (2010) and Rüber *et al*. (2007) were used for these analyses.

Split-supporting nucleotide patterns with putative synapomorphies were visualized with SAMS 1.4 beta (splits analysis methods), which allows the identification of conserved split-supporting positions without reference to a tree (Wägele & Mayer, 2007). This method is particularly helpful in exploratory analyses of data sets, especially for visualizing support and conflict (signal-to-noise ratio) in an alignment. Wägele & Mayer (2007) mainly used this approach to address potential long-branch effects that are known to obscure phylogenetic relationships in molecular data sets. We used the default settings in all analyses and focused on the 120 best-supported splits. Both the concatenated Mayden & Chen (2010) and the Rüber *et al*. (2007) data sets were used for the SAMS analyses.

RESULTS AND DISCUSSION

MORPHOLOGICAL EVIDENCE

Mayden & Chen (2010) severely criticized Britz & Conway's (2009) results and interpretations, and accused them (p. 155) of 'a reliance upon the idea of progressive evolution in miniaturization via a step by step process and idealistic morphology, only exacerbated by their limited selection of species assumed to be related to *Paedocypris*, the conflation of miniaturization as a causative explanation for the developmental admixture of synapomorphic and plesiomorphic traits not characteristic of either Cypriniformes or the formerly recognized family Cyprinidae and the lack of some of these characteristics because the species is neither a cyprinid nor a cyprinoid'. Their criticism touches on several different issues, and we will discuss them in sequence.

Is Paedocypris a developmentally truncated organism?

One of the controversial points in Mayden & Chen (2010) and one of their main criticisms is centred on the question of *Paedocypris* being a developmentally truncated fish. This is summarized in their statement (p. 164–165): 'What may appear as a developmental truncation or loss of one or more terminal conditions in a developmental pathway in a taxon may be equally or more parsimoniously interpreted as a taxon that never possessed the terminal character state . . . This is the case for some of the character transformations and character states identified for *Paedocypris* by Britz & Conway (2009)'.

Britz & Conway (2009) listed 42 bones that are present in most other cyprinids, but are absent in *Paedocypris*, and an additional 18 bony structures the sizes of which are reduced in *Paedocypris* and their complexity simplified, compared with other cypriniforms. Mayden & Chen (2010) interpreted the lack and/or simplification of several bony structures in *Paedocypris* as plesiomorphies, rather than secondary absences or reductions resulting from developmental truncation in this tiny cyprinid, the explanation favoured by Britz & Conway (2009). Among these structures are the kinethmoid and preethmoid, two bones that characterize Cypriniformes and are two strong synapomorphies supporting their monophyly (Fink & Fink, 1981). As described by Britz & Conway (2009), the kinethmoid is present, but only cartilaginous, in *Paedocypris*, and Mayden & Chen (2010: 165) considered the 'cartilaginous condition . . . a precursor to the ossification', which 'under this interpretation serves as the synapomorphy of Cypriniformes' and 'the ossification of this element . . . a further derived state . . . uniting all Cypriniformes, exclusive of *Paedocypris*'. Mayden & Chen (2010: 165) similarly interpret the absence of the preethmoid in *Paedocypris* as a plesiomorphy when they write that the 'preethmoid is interpreted to have evolved in the ancestor to all Cypriniformes, exclusive of *Paedocypris*'. They thus assumed that the absence of kinethmoid and preethmoid, two of the 42 bones that are missing in *Paedocypris*, represents a primary absence, a plesiomorphy, at the level of cypriniforms.

Referring to four additional reductions (characters 46–49 in their table 2), however, Mayden & Chen (2010: 167) state that these shared characters of *Paedocypris*, *Danionella*, and *Sundadanio* are 'independently derived in each of these three taxa, likely through what Britz & Conway (2009) determine to be developmental truncation or convergence'. Mayden & Chen (2010) thus assume that developmental truncation is a likely explanation for at least some of the bone absences cited by Britz & Conway (2009). This assumption, however, is in stark contradiction to their earlier statement that the absence of the kinethmoid and preethmoid are plesiomorphies of *Paedocypris*, and not the result of developmental truncation.

Developmental truncation of the skeleton of a bony fish (or truncation in general) may happen at the level of a character, in that a single bone fails to ossify (as in the case of the interhyal in *Danio rerio*; Cubbage & Mabee, 1996), or it may occur at the level of the organism, in that a large number of bones, known to ossify at later stages of development in close

relatives, fail to ossify. The result is that the adult truncated fish looks very much like the larval stage of a close relative (see Fig. 1).

Britz & Conway (2009) presented an elegant way to determine whether *Paedocypris* is developmentally truncated, and which of the absences in this tiny fish can be explained by an event of developmental truncation at the organism level and which cannot. To achieve this, Britz & Conway (2009) mapped the absent bones in *Paedocypris* onto a diagram compiled from published data by Cubbage & Mabee (1996) and Bird & Mabee (2003) on the skeletal development of the zebrafish *Danio rerio*. The zebrafish is a species considered to be closely related to *Paedocypris* according to Rüber *et al*. (2007) and a cypriniform for which we have the most comprehensive set of data on its sequence of ossification. Britz & Conway (2009) demonstrated that the majority of bones absent in *Paedocypris*

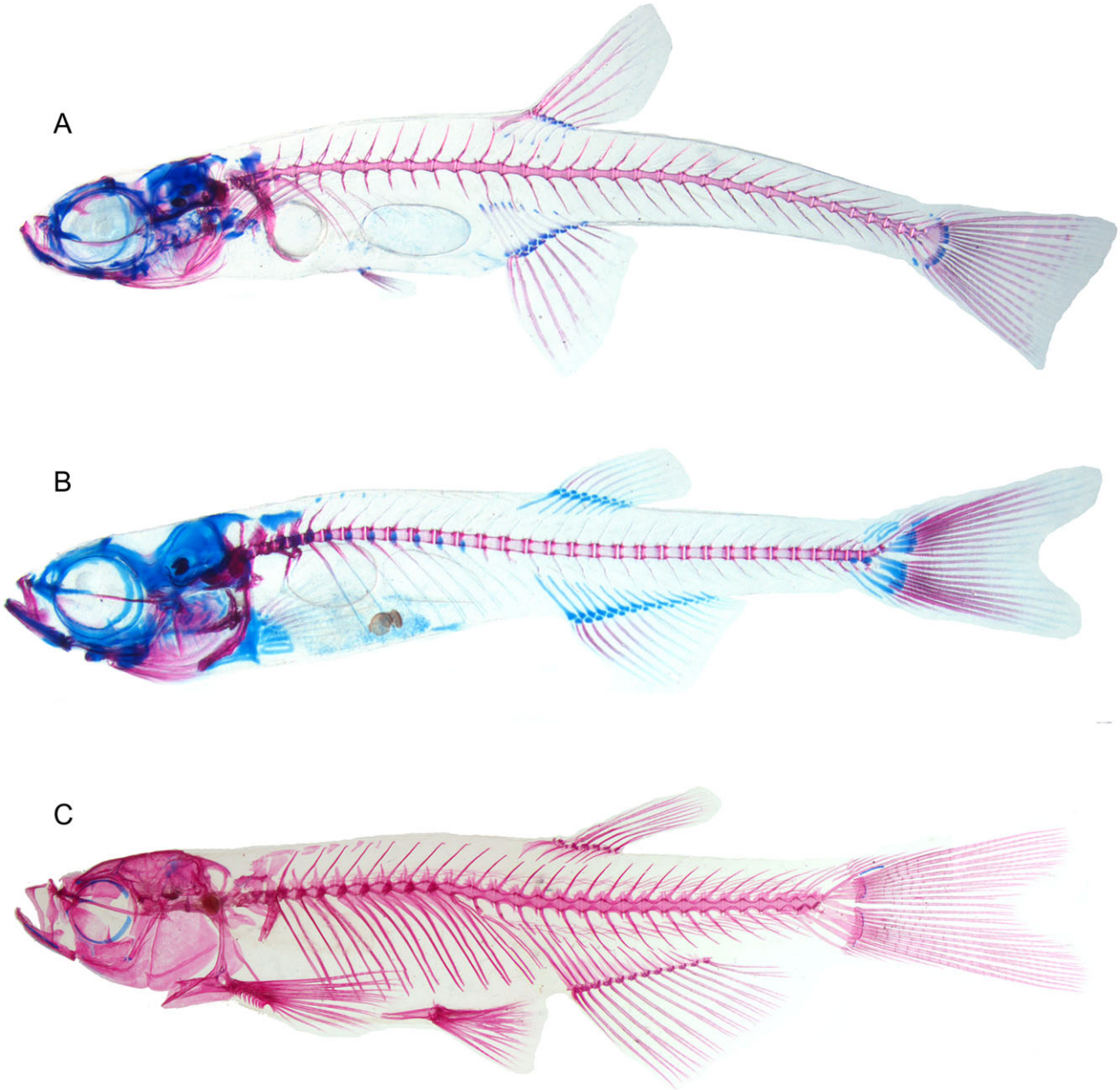

Figure 1. Cleared and double-stained specimens of (A) adult *Paedocypris* sp. of 11.2 mm standard length (SL) (BMNH 2008.4.11.2–11), (B) larval *Danio rerio* of 7.5 mm SL (BMNH uncatalogued), and (C) adult *D. rerio* of 26 mm SL (BMNH 2001.3.12.76–92). Note the large quantity of cartilage (blue, bone in red) in the cranium in (A) and (B) compared with (C), clearly emphasizing the developmentally truncated, larval-like skeleton of *Paedocypris*.

are those that appear late in the osteological development of *Danio rerio* (see Fig. 2). They argued that the absence of most of these bones in *Paedocypris* can therefore be explained by a relatively simple scenario of developmental truncation. The developmental stage of an adult *Paedocypris* thus corresponds well with that of a 7.5-mm larva of *Danio rerio* (see Fig. 1). The extent of developmental truncation, as calculated by the number of absent bones, is remarkable in *Paedocypris*. and is paralleled only by species of the miniature goby genus *Schindleria* (Johnson & Brothers, 1993) and the miniature cyprinids of the genus *Danionella* (see Britz *et al*., 2009). Although miniaturization and developmental truncation go hand in hand in *Paedocypris*, these two evolutionary phenomena are two independent processes, because miniature species do not have to be developmentally truncated, as in the case of the proportioned dwarfs belonging to the genera *Boraras* and *Danio* (Conway, 2005; Conway *et al*., 2008). In these proportioned dwarfs only very few bones are absent compared with their larger cyprinid relatives.

Of the bones that fail to ossify in *Paedocypris*, 29 occur towards the end of the development of the zebrafish skeleton, and can thus be explained by an organism-wide truncation. Among these are bones typically present in most teleosts, like the parietal, postcleithrum, vomer, nasal, infraorbitals, and supraneurals. Also among them are the kinethmoid and preethmoid, two of the bones for which Mayden & Chen (2010) claim that their absence cannot be explained by truncation, but needs to be interpreted as a primary absence. If this were the case, however, their argument would have to be expanded to include all the other 27 bones ossifying along with the kinethmoid and preethmoid at the end of the ossification sequence, like the bones cited above (Fig. 2). This would mean that the absence of bones like the parietal, postcleithrum, vomer, and nasal in *Paedocypris* would also have to be interpreted as a primary absence. In consequence, one would need to conclude that *Paedocypris* possesses more than ten plesiomorphic absences that would put it at the base not only of teleosts, but of osteognathostomes (= osteichthyans of authors), for which the presence of these bones is characteristic. This, of course, is an absurd assumption. Mayden & Chen (2010) seem to have ignored the fact that *Paedocypris* shares with *Danionella* the absence of 59 of the bones and structures present in other cypriniforms, and one would expect a similar argument for *Danionella* as put forth by Mayden & Chen (2010) for *Paedocypris*. The same absences in *Danionella*, however, were explained by Mayden & Chen (2010: 160) as 'independently derived' in this taxon.

It is thus clear that Mayden & Chen (2010) confused support with non-conflict of the morphological characters that they listed and discussed in relation to their molecular hypothesis of the phylogenetic placement of *Paedocypris*, as pointed out previously by Britz & Conway (2011). All Mayden & Chen (2010) did was map the characters onto their 'simplified tree', which then determined their polarity, contradicting their own statement (p. 152) that 'outgroup comparisons (were) used to determine synapomorphies'. Mayden & Chen's (2010) mapping approach, of course, excludes any character conflict *a priori*, although the way they plotted the characters and the chosen tree have several serious problems associated with them, which are discussed further below.

The confusion created by Mayden & Chen's (2010) interpretation of absences does not end here, however, and it seems that they misunderstood Britz & Conway (2009). Some of the bones for which Mayden & Chen (2010: 167) do accept developmental truncation as an explanation for their absence ('The remaining characters identified in Table 2 . . .'), cannot actually be interpreted as the result of a simple truncation, as in the case of the ectopterygoid and posttemporal. The ectopterygoid and posttemporal develop relatively early in *Danio rerio*, and are not part of the late-phase ossifications (Cubbage & Mabee, 1996). Their absence in *Paedocypris* therefore cannot result from a straightforward organism-wide truncation, as demonstrated and explicitly discussed by Britz & Conway (2009), and as is evident in our Figure 2.

We conclude, as did Britz & Conway (2009), that *Paedocypris* is a highly developmentally truncated fish. We thus reject Mayden & Chen's claim that it is not, which is based on a misunderstanding and misinterpretation of Britz & Conway's (2009) results.

Paedocypris*: a taxonomic enigma*

Hanken & Wake (1993: 510) noted that a 'challenge to phylogenetic analysis is posed when the miniaturized, adult morphology is so specialized . . . as to obscure affinities with any other known taxon; many such groups are taxonomic enigmas'. To some extent *Paedocypris* is such an 'enigma', even though the presence of a Weberian apparatus identifies it as a member of the Otophysi (Britz & Conway, 2009). We want to use this opportunity to explain in more detail the approach of Britz & Conway (2009), and add additional information from several analyses to better understand the phylogenetic position of *Paedocypris*, a fish with a predominantly larval skeleton characterized by the absence of numerous bones. To face the 'enigma' issue raised by Hanken & Wake (1993), Britz & Conway (2009) first checked all synapomorphies for Cypriniformes previously proposed by Fink & Fink (1981, 1996) for their presence in *Paedocypris*. Of Fink & Fink's (1981, 1996) nine unique synapomorphies of Cypriniformes, four are present in *Paedocypris* (kinethmoid element, fifth

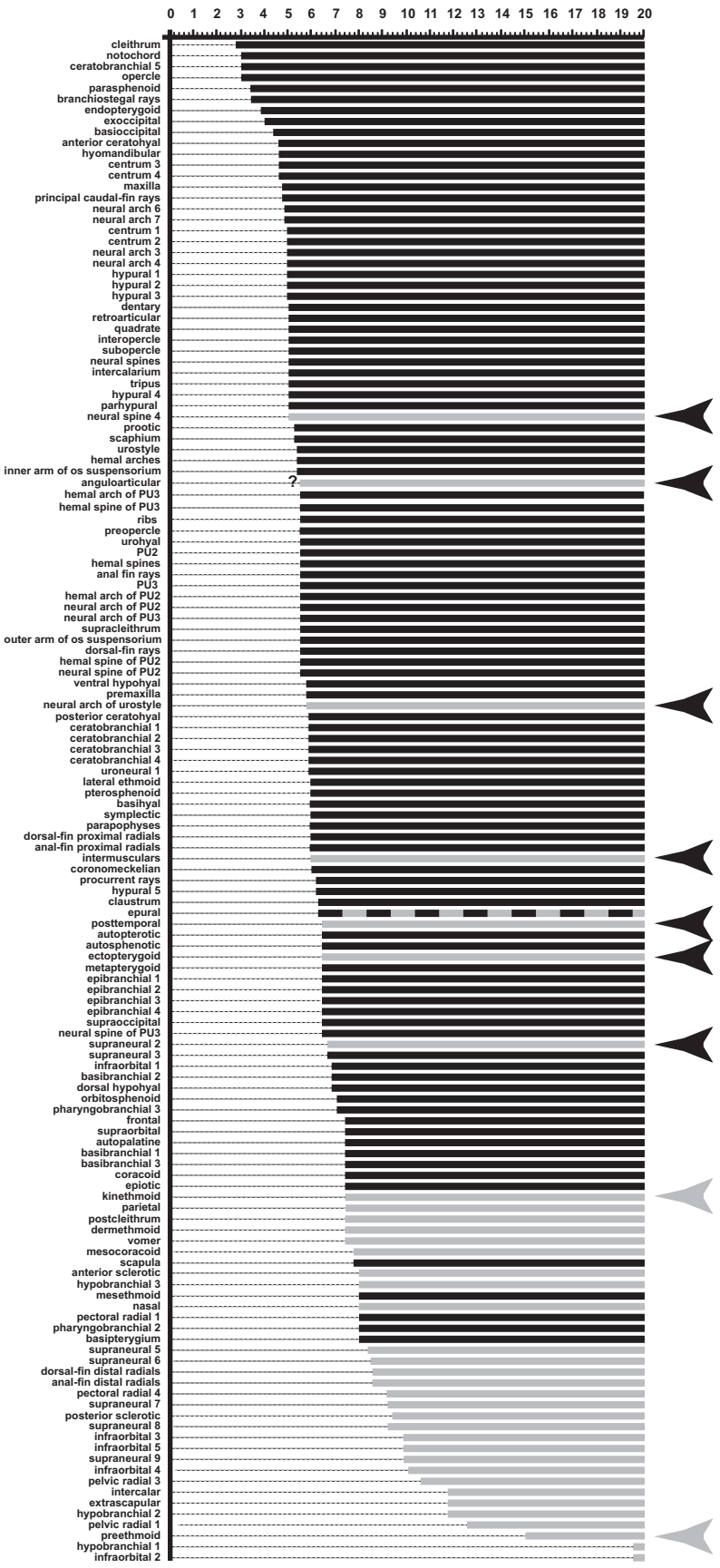

Figure 2. Schematic representation to illustrate relationships between losses and reductions in the skeleton of *Paedocypris* and sequence of ossification of *Danio rerio*, modified from Britz & Conway (2009: fig. 13). First appearance of ossifications shown as horizontal bars at given lengths. Black bars represent ossifications present and grey bars represent ossifications absent in *Paedocypris*. Black-and-grey bar represents ossification with polymorphic character state in *Paedocypris*. Question mark at bar of anguloarticular indicates lack of information for its two components: angular and articular. Grey bars marked with a grey arrowhead represent bones absent in *Paedocypris*, which cannot be explained by organism-wide developmental truncation. The two grey bars marked with grey arrowheads represent two bones among late-phase ossifications, for which Mayden & Chen (2010) rejected the hypothesis that their absence in *Paedocypris* was a result of developmental truncation.

ceratobranchial enlarged, extending much further dorsally than other ceratobranchials, teeth on the fifth ceratobranchial ankylosed to the bone, and lateral process of the second centrum elongate and extending well into the body musculature). Of the additional eight synapomorphies of Cypriniformes, which are homoplastic among Ostariophysi according to Fink & Fink (1981), *Paedocypris* has six (jaw teeth absent, teeth on second and third pharyngobranchials and basihyal absent, two posterior pharyngobranchial toothplates absent, toothplate associated with basibranchials 1–3 absent, epurals two or fewer, adipose fin absent). *Paedocypris* thus lacks five unique and two homoplastic cypriniform synapomorphies from Fink & Fink's (1981) list, and these were interpreted by Britz & Conway (2009) as secondary absences. Britz & Conway (2009) also discussed the previously proposed synapomorphies of the family Cyprinidae and concluded that *Paedocypris* has two of the four characters proposed by Siebert [1987; absence of a uncinate process on epibranchial 1 (EB1) and 2; absence of pharyngobranchial 1 (PB1)], and none of the six characters proposed by Cavender & Coburn (1992). Britz & Conway (2009: 403) hypothesized that of the eight cyprinid synapomorphies of Siebert (1987) and Cavender & Coburn (1992) that *Paedocypris* lacks, 'three are inapplicable because the structures in question are absent . . . and lack of the remaining five might be due to the developmental truncation'.

In addition to the two cyprinid synapomorphies taken from Siebert (1987; absence of an uncinate process on EB1 and 2; absence of PB1), Britz & Conway (2009) proposed a third synapomorphy for *Paedocypris* and Cyprinidae: the presence of a masticatory plate on the basioccipital process. Britz & Conway (2009) thought that this character showed *Paedocypris* to be a cyprinid. The masticatory plate is an anvil-shaped bony process of the basioccipital covered by a highly keratinized tissue pad and situated at the anterior end of the basioccipital process (Howes, 1981). The lower pharyngeal jaws work against this pad to break down food (Sibbing, 1982). This masticatory plate is present in all cyprinids, including *Paedocypris*, and is also present in *Psilorhynchus* (Conway, 2011). Conway (2011) interpreted this character as support for the Cyprinoidei, including the Cyprinidae and Psilorhynchidae. Mayden & Chen (2010: 158, table 2), however, reinterpreted the presence of a masticatory plate 'to be synapomorphic for Cypriniformes', without any explanation. Referring to the masticatory plate, Mayden & Chen (2010: 158, table 2) stated in addition that there was 'further modification of this structure in Catostomidae as the palatal organ in all species and horny plate in some (Eastman, 1977)' and that '. . . this plate is reduced in size in the common ancestor to Cobitoidea and lost above the Gyrinocheilidae plus other Cobitoidea', again without explanation of how they arrived at their interpretations. As it turns out, the palatal organ is a soft tissue structure that sits in the roof of the oral cavity far anterior to the actual masticatory process, with its overlying keratinized pad (see Reid, 1982), and is an evolutionary novelty and a synapomorphy of Cypriniformes as a whole (Hernandez *et al*., 2007). As a soft-tissue structure it cannot be homologous to the masticatory plate, which is a bony structure. The fact that cyprinids have both structures (the palatal organ and the masticatory process, plus associated pad) shows that they are not homologous, as they fail Patterson's (1982) test of conjunction. Mayden & Chen's (2010) erroneous claim that catostomids have a masticatory plate that has been modified into a palatal organ is thus based on the confusion of two structures. Mayden & Chen (2010) went to great lengths to try and argue away this unique derived similarity between *Paedocypris* and other cyprinids. To fit their incorrect level of inclusiveness for the masticatory plate as a synapomorphy of cypriniforms they not only hypothesized erroneous homology assumptions but also had to assume a secondary loss in all cobitoids above *Gyrinocheilus*; however, *Gyrinocheilus* does not even have a masticatory plate with a keratinized pad covering it (see Conway, 2011; Doosey & Bart, 2011). The presence of a masticatory plate is thus still a valid, putative synapomorphy of *Paedocypris* and Cyprinidae, as originally suggested by Britz & Conway (2009), a monophyletic group that also includes Psilorhynchidae (see Conway, 2011).

We have demonstrated previously (Britz & Conway, 2009), and here above, that *Paedocypris* is without doubt one of the most developmentally truncated vertebrates. Mayden & Chen's (2010) claim that this is

not the case is erroneous, and is based on a complete misunderstanding and misinterpretation of Britz & Conway (2009), and on the confusion of primary and secondary absences of a dozen characters. Mayden & Chen (2010) should have heeded Hanken & Wake's (1993: 510) warning about the potential pitfalls developmentally truncated organisms like *Paedocypris* may provide: 'Particularly confounding are the many instances of reversal in which plesiomorphic states are restored; when organismal-wide, or global, paedomorphosis is involved, there may be profound difficulty in determining whether taxa are basal or derived'. This applies to *Paedocypris* and the highly contrasting interpretation of its characters by Mayden & Chen (2010) and Britz & Conway (2009).

Paedocypris*, developmental truncation, and homoplasy*

Another complicating issue when dealing with miniaturized taxa, which may be termed the 'homoplasy trap', was noted by Hanken & Wake (1993: 503): 'Because of the homoplasy that frequently accompanies size decrease in closely related groups, miniaturized taxa may be regarded incorrectly as monophyletic, representing a single instance of reduction'. To address this issue of a greater potential for homoplastic absences and reductions to occur in unrelated miniatures, Britz & Conway (2009) studied the osteology of 21 additional closely and more remotely related miniaturized and non-miniaturized cyprinids, and tabulated the presence and absence of certain bones. They noticed that some of these bones were almost universally absent in miniatures, even in those that are not closely related, like uroneural 2 or hypural 6. As the absence of these bones seems to be closely linked to miniaturization itself, their absence as a character state is of little phylogenetic value when comparing miniature fishes (e.g. see Conway, 2005). In contrast to the bones frequently absent in miniature cyprinids, other bones were lacking only in a few of the miniatures, and the distribution of these absences was thus much more restricted systematically. A few bones, like the vomer, ectopterygoid, posttemporal, and postcleithrum, were only absent in three of the 19 miniatures that Britz & Conway (2009) studied: *Paedocypris*, *Sundadanio*, and *Danionella*. To avoid the problem of increased levels of homoplasy using absences and reductions that are frequently associated with miniaturization, Britz & Conway (2009) excluded all absences of bones that have a wide distribution among miniatures as evidence in their search for the phylogenetic placement of *Paedocypris*, and only allowed those absences with a restricted distribution amongst miniatures as potential synapomorphies.

In addition, Britz & Conway (2009) used as phylogenetic evidence the absence in *Paedocypris* of those bones that were not the result of simple organism-wide truncation. These bones are not part of the large number of late-phase ossifications of *Danio rerio* (bones ossifying at standard lengths greater than 7 mm; Cubbage & Mabee, 1996), but are among the early-phase ossifications, and their absence in *Paedocypris* is thus not caused by developmental truncation. Interestingly, two of these absent bones in *Paedocypris* (the posttemporal and ectopterygoid) are also among the bones that are absent in only a small number of taxa among miniature cyprinids, and thus have a restricted distribution. Among all of the bone absences that Britz & Conway (2009) recorded for miniature cyprinids, such systematically restricted absences and those that are not the result of organism-wide truncation qualify, in our opinion, as stronger, more convincing putative synapomorphies.

In addition to the absences of bones with a restricted distribution and those that are not the result of an organism-wide developmental truncation, Britz & Conway (2009) also used a third type of character as support for their phylogenetic hypothesis. This included those characters not affected by either miniaturization or developmental truncation, characters they termed 'progressive' ('innovate characters' in Weitzman & Fink, 1985). Among these were modifications of the Weberian apparatus and gill arches, which seem to be at least highly unusual, if not unique, among cypriniforms.

To avoid the homoplasy trap discussed by Hanken & Wake (1993), Britz & Conway (2009) used these three different types of characters detailed in the last few paragraphs. Britz & Conway (2009) concluded that *Paedocypris*, *Sundadanio*, and *Danionella* form a monophyletic group because the three taxa share eight derived characters. Four of these eight characters, or putative synapomorphies, have a restricted distribution (absence of postcleithrum, absence of vomer, absence of ectopterygoid, and absence of posttemporal) among miniatures. Two of these four characters with a restricted distribution are not the result of developmental truncation (absence of ectopterygoid and absence of posttemporal; see discussion above and Fig. 2). The remaining four of the total of eight are progressive characters (concha scaphii reaches to the end of the processus ascendens of the scaphium; inner arms of ossa suspensoria are fused in their proximal portion, greatly elongated, following the curvature of the anterior chamber of the swimbladder; gap between enlarged neural arches 3 and 4 filled by extensive development of cartilage; and tripus is hypertrophied and unusually elongate). *Paedocypris* and *Danionella* share, in addition, seven derived characters or putative synapomorphies. Four of these seven have a systematically restricted distribution among miniature cyprinids (absence of parietal, absence of HB3, absence of supraneural 5, and absence of supraneural 6), one is not the result of an organism-wide truncation

(absence of neural spine of fourth vertebra), and the remaining two are progressive characters (EB5 articulates with the levator process of EB4 and with the tip of CB5, PB3 is fused with the distal tip of EB4). The proposed close relationship of *Paedocypris* with *Danionella* and *Sundadanio* hypothesized by Britz & Conway (2009) is thus supported by three classes of characters: absences that were not the result of truncation; absences with a restricted distribution among miniatures; and progressive characters.

Morphological characters from actual specimens and actual trees

Mayden & Chen (2010) accused Britz & Conway (2009) of failing 'to provide any type of matrix of character data for homologies of morphological characters, metrics as evidence of any type of phylogenetic analysis was [*sic*] conducted, or that their hypotheses of homology had been corroborated or falsified; thus, one cannot replicate their study to evaluate the evidence supporting their multiple hypothses [*sic*].' Although Britz & Conway (2009) explained their approach in detail and provided numerous photographic illustrations and a comprehensive discussion of the different characters of *Paedocypris*, which makes it very easy to 'replicate' or critically evaluate their study, they did not provide a character matrix or an extensive phylogenetic analysis with numerous cypriniforms. Even though not explicitly stated in their paper, Britz & Conway's (2009) main reason for avoiding the large-scale 'matrix' approach was simply to avoid the pitfalls associated with miniaturized taxa, as discussed by Hanken & Wake (1993; i.e. the available character set is restricted because of developmental truncation, characters affected by truncation may be wrongly interpreted as plesiomorphies, and there is a danger of scoring characters that are homoplastic because of independent miniaturization, which may group all miniatures based on shared non-homologous reductions).

Whereas Mayden & Chen (2010) criticized Britz & Conway (2009) for the lack of a matrix, and discussed and reinterpreted Britz & Conway's (2009) morphological characters, they themselves failed to provide a data matrix and phylogenetic analysis. Instead, Mayden & Chen (2010: 163) presented a phylogenetic tree in their figure 3 that they called 'a simplified tree depicting our current understanding of phylogenetic relationships of the major clades of the Cypriniformes'. This 'simplified tree', however, differs from the actual tree Mayden & Chen (2010) obtained from their molecular analyses, and contains taxa not even included in their original molecular tree (e.g. *Ellopostoma*). The origin of this 'simplified tree' and the factual evidence supporting it is not explained or revealed anywhere in their paper. This 'simplified tree' is then used by Mayden & Chen (2010) to 'optimize' the 'hypothesized morphological synapomorphies discussed in the text', but again the authors fail to reveal how they optimized the characters, as their distribution along the tree appears to be anything but plausible.

To perform a repeatable analysis of the characters listed by Mayden & Chen (2010), we coded all species-level taxa used in their analysis with the character state assigned by Mayden & Chen in their 'simplified tree' (our matrix 1 in Appendix 2). Our intent was to test Mayden & Chen's claim that their molecular topology is also supported by morphology, something that they did not attempt in their study. The single tree (87 steps; consistency index, CI = 0.87; retention index, RI = 0.96) obtained by parsimony analysis of this morphological data set with PAUP differs dramatically from Mayden & Chen's 'simplified tree' and is illustrated in Figure 3A. The tree is mostly unresolved and shows most cobitoids and all cyprinoids each in a polytomy. The only resolved cypriniform clades within the entire tree, although with low or no bootstrap support, are the two species of catostomids, which are recovered as sister taxa, and a clade consisting of *Sundadanio*, *Danionella*, and *Paedocypris*, with the same relationships as hypothesized by Britz & Conway (2009). This means that even the limited number of the morphological characters Mayden & Chen (2010) 'optimized' onto their 'simplified tree' actually support Britz & Conway's (2009) conclusions, at least in part, rather than their own.

Mayden & Chen (2010) make no reference to any vouchered morphological specimens in their study, and it is unclear how they actually mapped morphological characters with such precision onto their 'simplified' tree. Because of this, we were interested to see what the result would be if we scored all 68 characters listed in table 2 in Mayden & Chen (2010) from specimens for their taxa. Of the 85 species that Mayden & Chen (2010) used in their molecular analysis, we were able to obtain specimens of all but six, and had to rely on closely related substitute species for another 35 taxa (see Table 1).

As we scored the different taxa for the different characters, we noticed a number of inconsistencies in Mayden & Chen's (2010) character list that needed to be corrected first, and are discussed here. Their character 22 reads 'Absence of PB1' in their table 2 (p. 159), but reads 'Presence of PB1 is hypothesized to be a synapomorphy of Cyprinoidea' in the text (p. 166). This character was proposed by Siebert (1987), who recorded PB1 to be absent in Cyprinidae and in cobitoids except Gyrinocheilidae and Catostomidae. Siebert (1987) interpreted the absence of PB1 as a synapomorphy of Cyprinidae and as an independently evolved synapomorphy of more derived cobitoids. Mayden & Chen's (2010) view concerning the presence of PB1 as

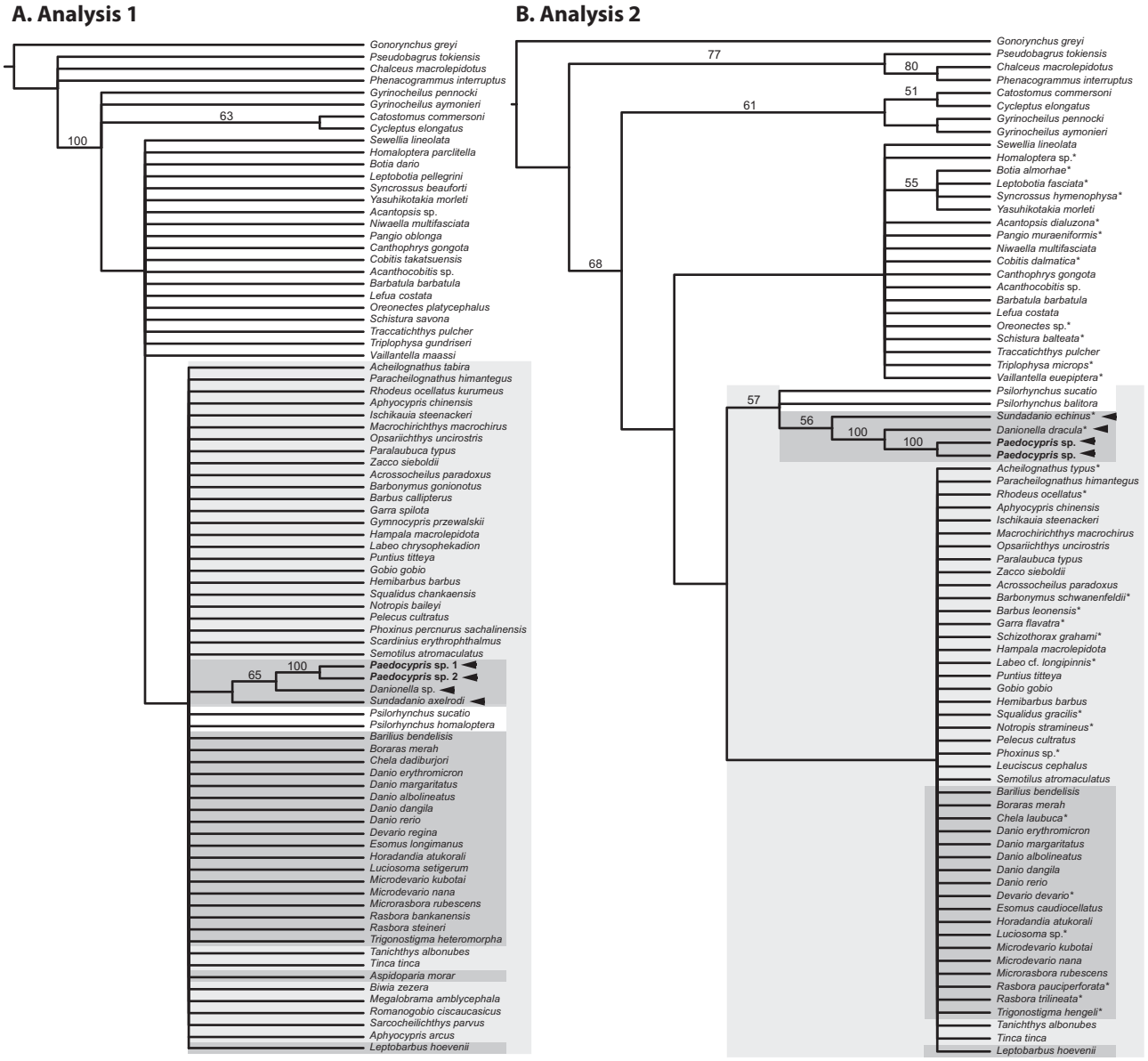

Figure 3. Phylogenetic analyses of morphological characters listed in table 2 of Mayden & Chen (2010). A, single most-parsimonious cladogram (87 steps; consistency index, CI = 0.87; retention index, RI = 0.96) resulting from parsimony analysis of 68 characters, coded following the character distribution from Mayden & Chen's figure 3 (matrix 1 in Appendix 2). B, strict consensus of 209 equally parsimonious cladograms (114 steps; CI = 0.53; RI = 0.87) obtained from parsimony ratchet analysis of 61 characters, coded based on an investigation of actual specimens (matrix 2 in Appendix 3). Numbers above branches indicate bootstrap support. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The miniature taxa *Paedocypris* (in bold), *Danionella*, and *Sundadanio* are highlighted with a black arrowhead. Substitute taxa (see text and Table 1 for an explanation) are highlighted with an asterisk.

a synapomorphy of cyprinoids is thus erroneous. The description of this character was thus corrected to 'absence of PB1' in our matrix 2 (Appendix 3).

In Mayden & Chen's (2010) table 2 (p. 159), their character 21 'Absence of an uncinate process on EB1 and EB2' is attributed to Siebert (1987). Mayden & Chen (2010) noted that this character is a 'Proposed synapomorphy for Cyprinidae *sensu* Fink & Fink (1981)'. In their interpretation in their table 2, however, they stated: 'Presence of uncinate process on EB1 and EB2 synapomorphic for Cyprinoidea'. To add to the confusion they mentioned in the text (p. 165) that the 'lack of the uncinate processes on EB1 and EB2 is considered a synapomorphy for Cyprinidae by Fink & Fink

(1981)'. This character has nothing to do with Fink & Fink (1981), who looked only at synapomorphies of ordinal-level Ostariophysi, but originated with Siebert (1987). Siebert's (1987: 158) character 2 reads 'absence of EB I and II uncinate processes', and he scored Cyprinidae and higher loaches with absence of uncinate processes on EB1 and 2, but Gyrinocheilidae and Catostomidae with their presence. He interpreted the absence of the processes as a synapomorphy of cyprinids on the one hand and more derived loaches on the other hand, and the presence of the processes in Gyrinocheilidae and Catostomidae as a plesiomorphy. Siebert's (1987) conclusion is thus counter to that of Mayden & Chen (2010), and no explanation is provided by Mayden & Chen for the change of level of inclusiveness.

Mayden & Chen (2010) also appear to have misinterpreted one of Fink & Fink's ordinal-level characters for the Cypriniformes (the absence of overlap between the ectopterygoid and palatine, i.e. autopalatine). Because the ectopterygoid is absent in *Paedocypris* (Britz & Conway, 2009), this character is logically interpreted as inapplicable to this taxon, yet in their table 2 and figure 3, Mayden & Chen (2010) reinterpreted this character as an autapomorphy of *Paedocypris*. In doing so, Mayden & Chen changed the focus of Fink & Fink's original character from the relationship between two bones (the autopalatine and ectopterygoid) to the absence of a bone (the ectopterygoid), which also happens to be the focus of their character 47 (loss of ectopterygoid). The same can be said for Mayden & Chen's interpretation of two characters from Chen, Yue & Lin (1984), one relating to the relationship between infraorbital 5 and the supraorbital (Mayden & Chen's character 24), and the other relating to the position of the anterior opening of the trigeminofacial chamber (Mayden & Chen's character 27). We reverted the description and coding for character 9 to reflect its original usage as a derived characteristic of Cypriniformes (*sensu* Fink & Fink, 1981), and to avoid overlap with character 47 in our matrix 2. We also reverted the description and coding for characters 24 and 27 in our matrix 2 to reflect their original usage as derived characteristics of the Cyprinidae by Chen *et al*. (1984) and Cavender & Coburn (1992).

We also discovered that several characters are listed in Mayden & Chen's (2010) table 2 more than once. For example, their character 32 ('Three uppermost pectoral-fin rays hypertrophied in males and articulating with likewise hypertrophied dorsal-most pectoral radial, which is very tightly associated with the second pectoral radial') overlaps completely with their characters 51 ('Hypertrophied first pectoral radial and three uppermost pectoral fin radials') and 53 ('Hypertrophied first pectoral-fin ray'). Their character 33 ('Basipterygia of pelvic girdle in males hypertrophied with a large flange of membrane bone directed anterodorsally toward the outer arm of the os suspensorium') is identical to their character 52 ('Hypertrophied anterodorsally directed basipterygium'). Their character 35 ('Presence of hemal spines in the abdominal region starting with vertebra 7') is identical to their character 55 ('Haemal spines in abdominal region'). Although Mayden & Chen (2010: 159, table 2) consider characters 32, 33 and 35 'Synapomorphies of Paedocyprididae', characters 51–53 and 55 are 'Features found only in *Paedocypris*'. We therefore removed characters 51–53 and 55 from the list. In addition, their character 30 ('PB2 and 3 at the same level and confluent with each other and EB4') is identical to their character 37 ('PB3 and EB4 connected via continuous cartilage'). Although Mayden & Chen (2010: 159, table 2) interpreted character 30 as an 'Autapomorphy for *Paedocypris*, Paedocyprididae, Paedocypridoidea (28A)', they concluded that character 37 is 'Independently derived in miniature ancestral species, one for *Paedocypris* (37B) and an ancestral species of *Danionella* (37B*)' (p. 160). When we checked how these characters (28A, 37B, and 37B*) were optimized onto their 'simplified tree' we noticed that 28A does not appear on the branch leading to *Paedocypris*, but on the branch leading to Cyprinoidea. It seems that Mayden & Chen (2010) confused 28A with 30A, as they cite state 28A for their character 30, but even taking that into account, their confusion cannot be resolved, as 30A is mapped as a synapomorphy of 'Cypriniformes less *Paedocypris*'. A character state 30A does not appear anywhere in their table 2. We therefore removed character 37 from the list.

One of Mayden & Chen's (2010) characters is listed three times in their table 2. Their character 17 ('Number of postcleithra reduced to one'), character 31 ('Postcleithra absent'), and character 48 ('Postcleithrum') are identical for *Paedocypris*, in which postcleithra are absent (Britz & Conway, 2009). Nevertheless, all three characters are 'optimized' onto their 'simplified tree', and each one appears only on the branch leading to *Paedocypris*. We therefore removed characters 31 and 48 from our matrix 2, which thus included a total of 61 characters.

This second data matrix (matrix 2, Appendix 3) was analysed with a parsimony ratchet and resulted in the consensus tree shown in Figure 3B (114 steps; CI = 0.53; RI = 0.87). Again this tree is drastically different from Mayden & Chen's (2010) 'simplified tree', and is for the most part unresolved with catostomids plus gyrinocheilids as the sister group to the remaining cobitoids plus cyprinoids, but with no bootstrap support. The three different miniature species in question, *Sundadanio*, *Danionella*, and *Paedocypris*, are recovered as a monophyletic group, although with low

bootstrap support. This clade of miniatures together with the two species of *Psilorhynchus* forms the sister group to the remainder of the Cyprinidae, again with no bootstrap support. Within the miniature clade, *Danionella* and *Paedocypris* are well-supported sister taxa with *Sundadanio* as their sister group, although with low bootstrap support. It is thus evident that even morphological characters coded from specimens for all the characters and taxa in Mayden & Chen's (2010) molecular analysis show support for the hypothesis published by Britz & Conway (2009).

The difference in coding between matrix 1 (see Appendix 2), derived from Mayden & Chen's (2010) tree, and matrix 2 (see Appendix 3), derived from coding characters from specimens, is quite remarkable. Only considering the 45 species identical in both data sets, we found a difference in coding of 13.3% between the two data sets (2835 cells in total, with 377 differing cell entries between data matrices 1 and 2). Some taxa stood out as having an even higher number of discrepancies between matrix 1 and 2, particularly the loaches. For example, there are 14 characters coded differently (23%) between matrix 1 and 2 for *Yasuhikotakia morleti* (Tirant, 1885) and *Lefua costata* (Kessler, 1876), and as many as 16 (26%) for *Niwaella multifasciata* (Wakiya & Mori, 1929) and *Canthophrys gongota* (Hamilton, 1822). This is because Mayden & Chen (2010) underestimated, ignored, or were perhaps unaware of the independent losses and reductions that have occurred in the cobitoid lineage, and have been documented in the literature previously by Sawada (1982) and Siebert (1987) (see also Conway, 2011).

To overcome the restriction in the number of characters available from matrix 1 and 2, and to evaluate the influence of introducing additional miniature taxa to the data matrix, we performed a third analysis using a modification of Conway's (2011) recent cypriniform data set, and added the miniature and non-miniature taxa listed in Table 1 in Britz & Conway (2009), so that the final data set included 120 morphological characters from 65 taxa (matrix 3 in Appendix 4). The consensus tree of the 4522 most-parsimonious trees (316 steps; CI = 0.42; RI = 0.84) is shown in Figure 4A. Again the topology of this tree is drastically different from the 'simplified tree' in Mayden & Chen (2010) or from their trees resulting from their molecular analyses. This tree resolves cobitoids and cyprinoids each as a monophyletic group, although with no bootstrap support. The three miniature genera *Sundadanio*, *Danionella*, and *Paedocypris* plus *Psilorhynchus balitora* are recovered as a monophyletic group within the large cyprinoid polytomy, but without bootstrap support, and with only *Danionella* and *Paedocypris* as well-supported sister taxa.

We performed a fourth analysis, in which we added not only the miniatures and non-miniatures from Britz & Conway's table 1, but also the various bone absences and progressive skeletal characters of *Paedocypris* listed in Britz & Conway (2009). This data set thus included 65 taxa and 144 characters (matrix 4 in Appendix 5). The resulting consensus tree of 9568 equally parsimonious cladograms (362 steps; CI = 0.43; RI = 0.84) is shown in Figure 4B. This tree also recovers cobitoid and cyprinoid monophyly, as in Conway (2011), but again with no bootstrap support, but has a better resolution among cyprinoids than the tree resulting from the third analysis (matrix 3). All miniatures and non-miniature taxa from table 1 in Britz & Conway (2009) plus *Psilorhynchus balitora* end up in a monophyletic group within cyprinoids, not supported by bootstrapping, however, and mostly unresolved, with *D. rerio* at its base. The three miniatures in question, *Sundadanio*, *Danionella*, and *Paedocypris*, form a well-supported monophyletic group with the same relationships to each other as presented in Britz & Conway (2009). Twelve characters support this monophyletic unit (for details, see Fig. S3), independent of the optimization mode ACCTRAN or DELTRAN. Six of these 12 characters have a consistency index of 1.0 and are thus unique synapomorphies of the three miniatures (character 117, dermethmoid absent; character 118, vomer absent; character 122, ectopterygoid absent; character 131, inner arms of ossa suspensoria fused; character 132, gap between enlarged neural arches 3 and 4 filled by extensive development of cartilage; and character 133, tripus hypertrophied and unusually elongate). The monophyly of *Danionella* plus *Paedocypris* is supported by a further seven unique synapomorphies (character 119, parietal absent; character 121, angular absent; character 123, pelvic radial cartilages 2 and 3 absent; character 124, rudimentary neural arch on urostyle absent; character 125, middle radials of dorsal and anal fins absent; character 126, intermuscular bones absent; and character 128, epibranchial 5 articulates not only with the levator process of epibranchial 4, but also with the tip of ceratobranchial 5) and an additional three (DELTRAN) or four (ACCTRAN) synapomorphies, with some level of homoplasy (for details, see Fig. S3).

Finally, we were interested to see what the effect on the tree topology was if we removed from matrix 4 all bone absences as characters that can either be linked to miniaturization itself or to organism-wide developmental truncation, following the rationale explained in Britz & Conway (2009) and above. This resulted in a matrix of 133 characters for 65 taxa (matrix 5 in Appendix 6). The strict consensus tree of 176 equally parsimonious cladograms (310 steps; CI = 0.46; RI = 0.85) resulting from parsimony ratchet analysis of matrix 5 (Appendix 6) shows a monophyletic cyprinoid and cobitoid lineage, but again with no bootstrap support (Fig. 5). Monophyly of the miniature clade plus

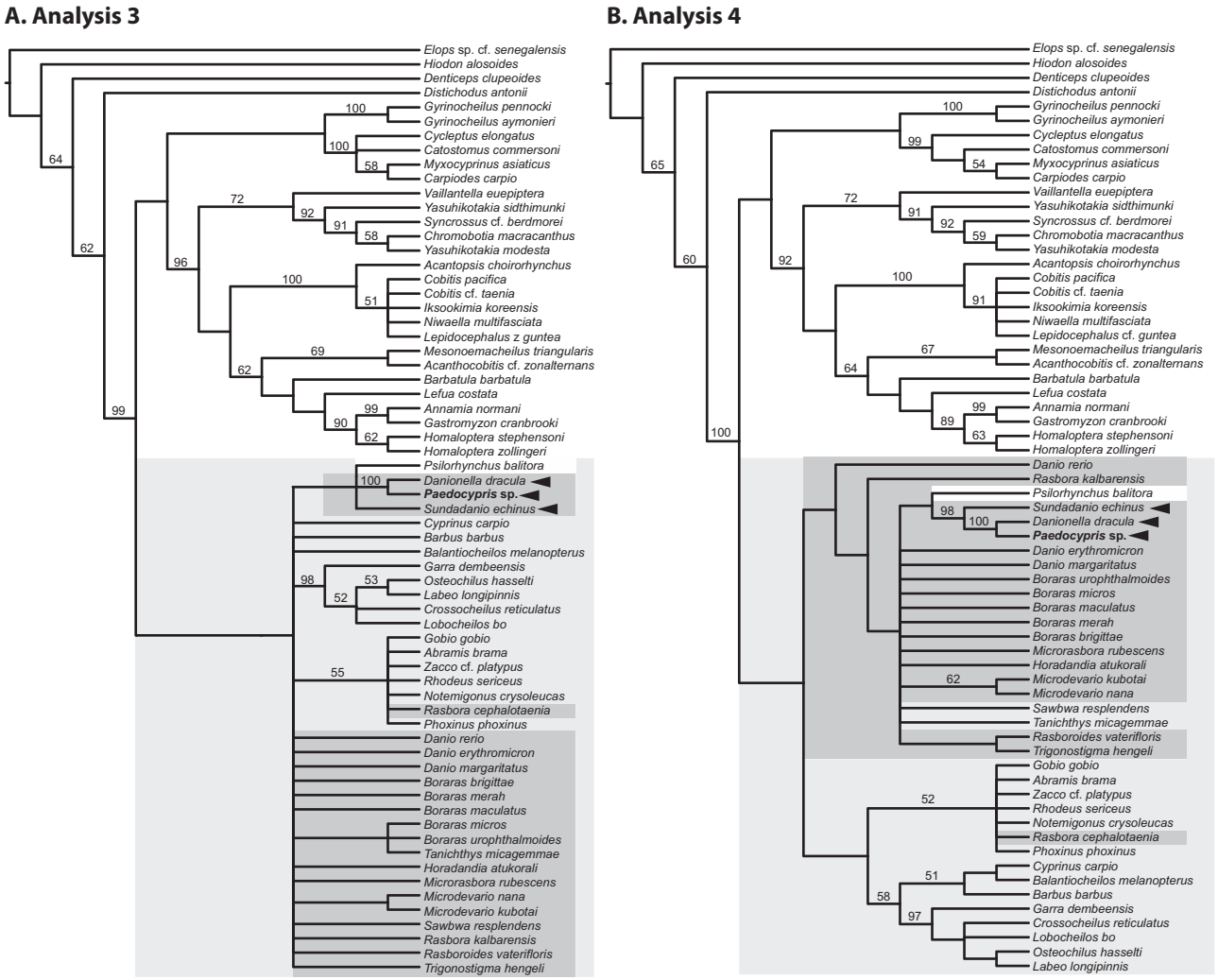

Figure 4. Phylogenetic analyses of morphological characters in Conway (2011) and Britz & Conway (2009). A, strict consensus of 4522 equally parsimonious cladograms (316 steps; consistency index, CI = 0.42; retention index, RI = 0.84) obtained from parsimony analysis of 120 characters (from Conway, 2011) for 65 taxa (matrix 3 in Appendix 4). B, strict consensus of 9568 equally parsimonious cladograms (362 steps; CI = 0.43; RI = 0.84) obtained from parsimony analysis of 144 characters [including 120 from Conway (2011) and 24 from Britz & Conway (2009)] for 65 taxa (matrix 4 in Appendix 5). Numbers above branches indicate bootstrap support. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The miniature taxa *Paedocypris* (in bold), *Danionella*, and *Sundadanio* are highlighted with a black arrowhead.

Psilorhynchus, as in analysis 4 (Fig. 4B), is no longer supported and that clade is broken up into a number of poorly supported or unsupported clades with a single exception: the three genera *Sundadanio*, *Paedocypris*, and *Danionella* form a well-supported monophyletic unit and show the same relationships as proposed by Britz & Conway (2009). Their monophyly and that of *Danionella* plus *Paedocypris* are supported by the same six and seven unique synapomorphies, respectively, as in analysis 4 (see lists above), in addition to three (for *Sundadanio*, *Danionella*, plus *Paedocypris*, independent of DELTRAN or ACCTRAN) and nine (*Danionella* plus *Paedocypris* ACCTRAN) or 11 (*Danionella* plus *Paedocypris* DELTRAN) homoplasious synapomorphies, respectively (for details, see Fig. S4).

All of our analyses of morphological characters support a sister-group relationship of *Paedocypris* and *Danionella*, and also to some extent a closer relationship of these with *Sundadanio* (Figs 4B, 5). In most analyses *Paedocypris* is also recovered within, or closely associated with, Cyprinidae and never at the base of Cypriniformes. We therefore strongly reject Mayden & Chen's (2010) unfounded claim that the position of *Paedocypris* (p. 152) 'as the basal sister group of all

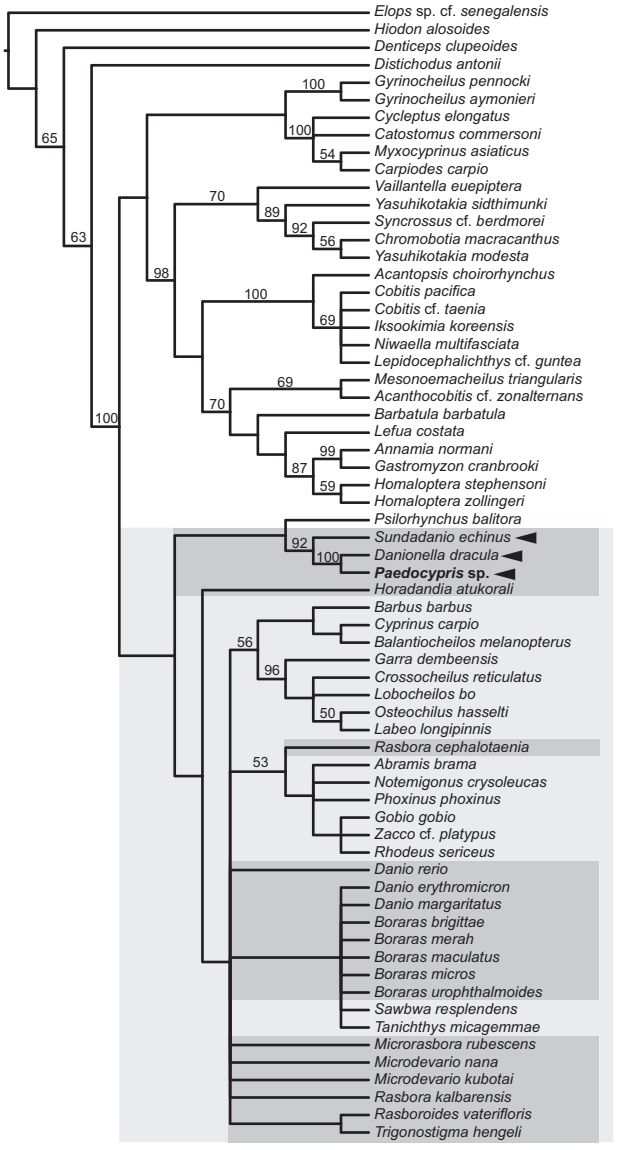

Figure 5. Strict consensus of 176 equally parsimonious cladograms (310 steps; consistency index, CI = 0.46; retention index, RI = 0.85) obtained from parsimony ratchet analysis of 133 characters [including 117 from Conway (2011) and 16 from Britz & Conway (2009)] for 65 taxa (matrix 5 in Appendix 6). Numbers above branches indicate bootstrap support. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The miniature taxa *Paedocypris* (in bold), *Danionella*, and *Sundadanio* are highlighted with a black arrowhead.

Cypriniformes' is 'also supported by previously proposed morphological characters'. The presence of a masticatory plate on the basioccipital covered by a keratinized epithelium, against which the lower pharyngeal jaws bite, is strong character evidence that *Paedocypris* is a member of the Cyprinoidei (Cyprinidae plus Psilorhynchidae) (or of the Cyprinidae, if Psilorhynchidae is treated as a subfamily Psilorhynchinae, of the Cyprinidae; see, e.g. Liao & Kullander, 2013). The precise position among cyprinoids of the monophyletic group consisting of *Sundadanio*, *Danionella*, and *Paedocypris* is, however, still unclear. *Danionella* has been considered a member of the Danionini based on the presence of the danionin notch in the lower jaw (Roberts, 1986), which is not developed, however, in *Sundadanio* and *Paedocypris*. The morphological data sets we used in analyses 3–5 comprised exclusively skeletal characters. This may have been the reason for our lack of a better-resolved phylogenetic position of these three miniatures because of their highly reductive skeleton combined with a high number of skeletal autapomorphies in each of the three taxa. Future morphological studies looking at soft-tissue characters and additional character systems not commonly used, such as egg surface or larval structures (see e.g. Britz, 1997; Britz & Cambray, 2001), are needed to better understand the phylogenetic position of these tiny, but fascinating fishes.

Molecular evidence

Previous molecular phylogenetic analyses were inconsistent regarding the placement of *Paedocypris*. Although Rüber *et al*. (2007) and subsequent studies (Fang *et al*., 2009; Tang *et al*., 2010; Tang *et al*., 2011, figs 1, 2, but not fig. 3) placed *Paedocypris* within the cyprinid subfamily Danioninae, Mayden & Chen (2010, see also Chen, Lavoué & Mayden, 2013, who used sequences from the same data set; Tang *et al*., 2011, fig. 3) placed it as the sister group to Cypriniformes. By reassessing the Rüber *et al*. (2007) and Mayden & Chen (2010) data sets we aimed to explore these inconsistencies. As reported in Mayden & Chen (2010), the concatenated data set from the six nuclear genes consisted of 5733 bp for 85 taxa, including four outgroups, with the following gene partitions: *EGR1* (846 bp), *EGR2B* (816 bp), *EGR3* (906 bp), *IRBP* (849 bp), *RAG1* (1497 bp), and *RH* (819 bp). The Rüber *et al*. (2007) data set based on the complete cyt *b* gene sequence consisted of 1131 bp for 228 taxa (for more details, see Rüber *et al*., (2007).

Reanalysis of Mayden & Chen's and Rüber et al.*'s data sets, and tests of alternative hypotheses*

Our ML reanalyses of the Mayden & Chen (2010) data set, with and without data partition, gave similar results, and no differences were observed regarding interrelationships pertinent to the focus of this paper (data not shown), hence only data from the unpartitioned analyses are provided. With the exception of the phylogenetic placement of *Tinca tinca* (Linnaeus, 1758), albeit not supported by bootstrapping, our ML tree of the concatenated data (Fig. 6) is identical to that presented by Mayden & Chen (2010: fig. 2A, B). The ML

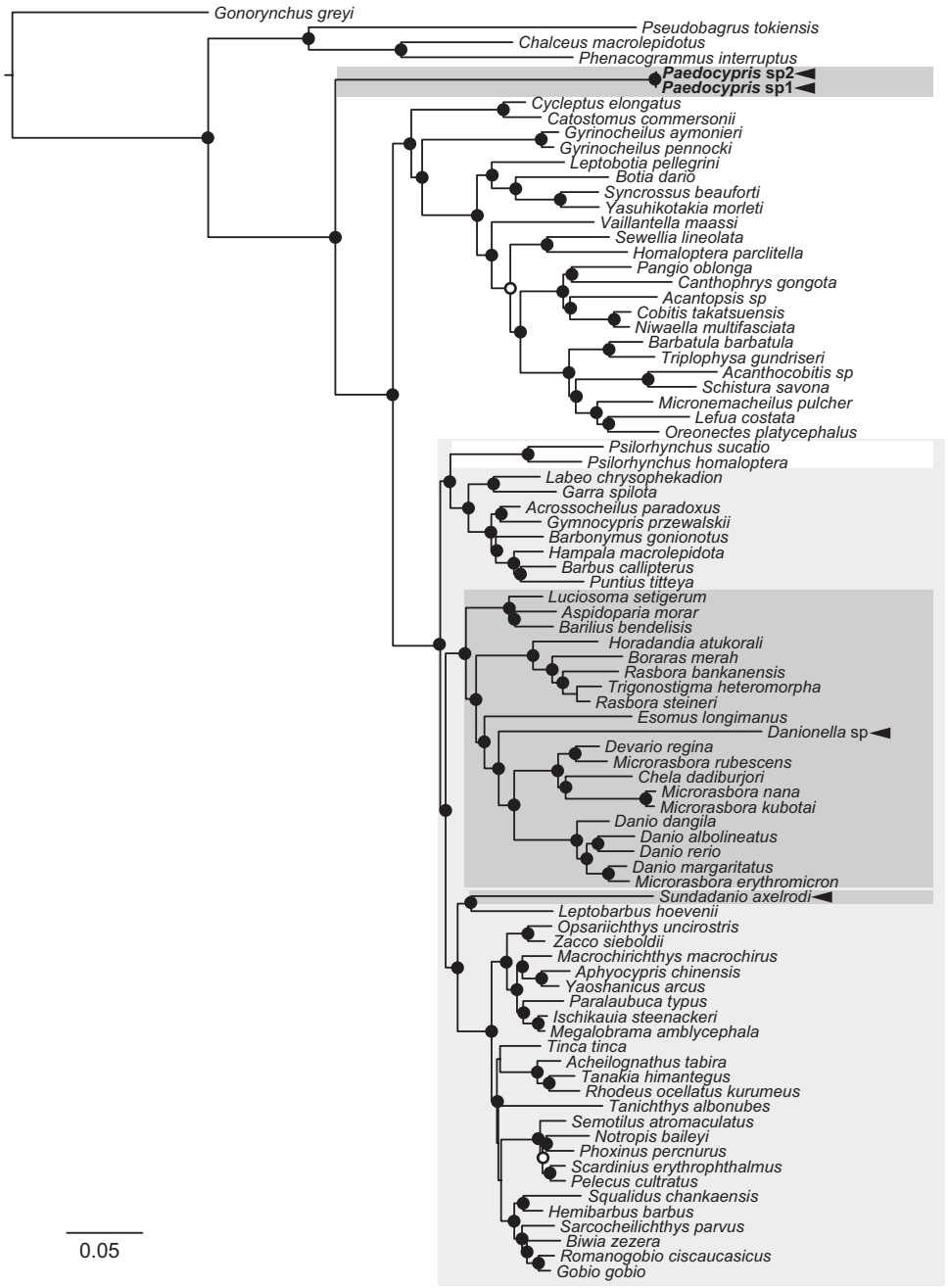

Figure 6. Maximum-likelihood (ML) tree of the combined Mayden & Chen (2010) data set comprising 5733 bp of six nuclear genes (*EGR1*, *EGR2B*, *EGR3*, *IRBP*, *RAG1*, and *RH*). Filled circles on nodes denote support for ML (bootstrap value, BP ≥ 70%). Open circles on nodes denote support for ML (BP ≥ 50 and < 70). Absence of circles on a node denotes no support for ML (BP < 50%). Scale bar indicates substitutions per site. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The miniature taxa *Paedocypris* (in bold), *Danionella*, and *Sundadanio* are highlighted with a black arrowhead.

analyses of each of the six nuclear genes are shown separately in Figures 7 and 8 (50% majority rule consensus trees are provided in Figs S5 and S6). Bootstrap values for the main branches, especially those relating to the placement of *Paedocypris*, were generally low for each single gene analysis. Our analysis of *EGR1* recovered *Paedocypris* deep inside the family Cyprinidae as the closest relative of *Leptobarbus*, with *Sundadanio* as the sister group of both, but with no boostrap support (Fig. 7A). *Danionella* sits within the

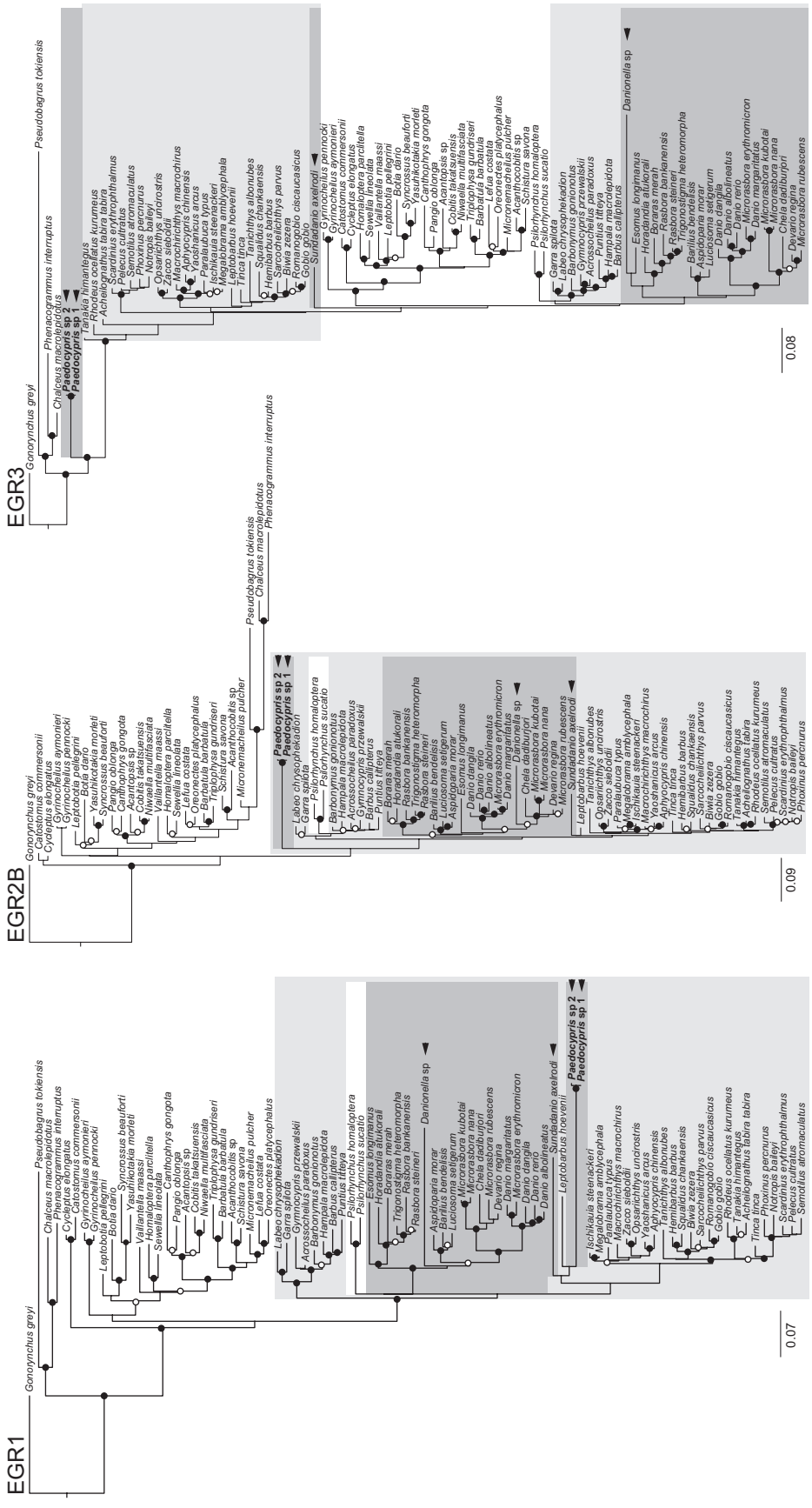

Figure 7. Single-gene maximum-likelihood analyses of the Mayden & Chen (2010) data set for the three nuclear markers *EGR1*, *EGR2B*, and *EGR3* are shown. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The miniature taxa *Paedocypris* (in bold), *Danionella*, and *Sundadanio* are highlighted with a black arrowhead.

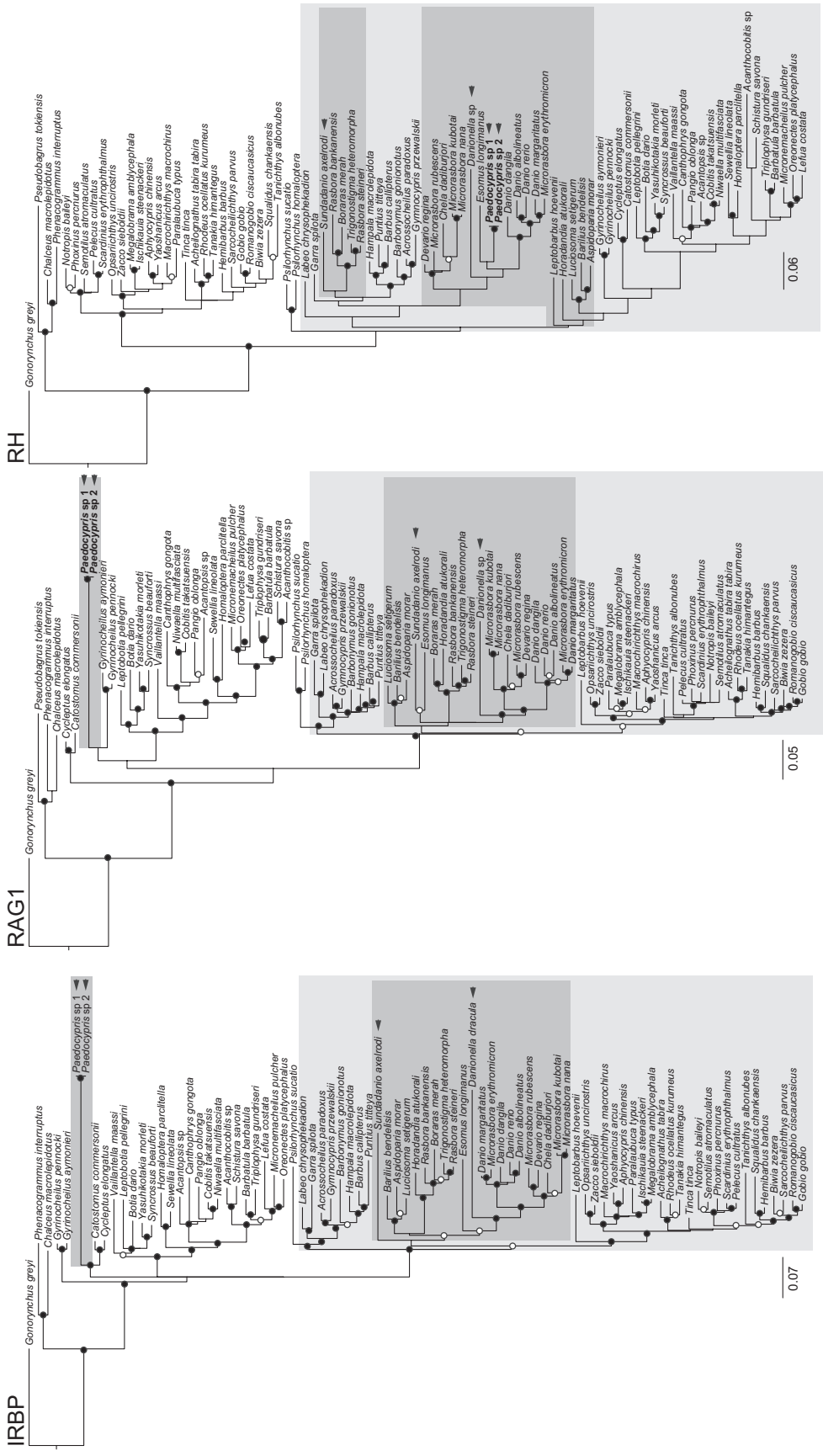

Figure 8. Single-gene maximum-likelihood analyses of the Mayden & Chen (2010) data set for the three nuclear markers *IRBP*, *RAG1*, and *RH* are shown. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The miniature taxa *Paedocypris* (in bold), *Danionella*, and *Sundadanio* are highlighted with a black arrowhead.

cyprinid subfamily Danioninae. In our analysis of the *EGR2B* sequence, *Paedocypris* forms the sister group of all other Cyprinidae, including *Psilorhynchus* (Fig. 7B), but again with no bootstrap support. Analysis of the gene *EGR3* resolved *Paedocypris* as sister group to the remaining Cypriniformes (Fig. 7C) with high support, with the same topology as recovered with the concatenated data set (Fig. 6) and in Mayden & Chen's (2010) combined analysis. The fourth gene, *IRBP*, analysed separately, yielded *Paedocypris* as the well-supported sister group of the two catostomid genera, with this group in a trichotomy with all other Cyprinoidei and all remaining Cobitoidei, except Gyrinocheilidae, the sister group of this trichotomy (Fig. 8A), but with no bootstrap support. Analysing *RAG1*, the fifth gene, we obtained *Paedocypris* as the sister group of the two *Gyrinocheilus* species within a monophyletic Cobitoidei (Fig. 8B), but without any bootstrap support. And finally, analysis of the sixth gene, *RH*, recovered *Paedocypris* deep inside danionine cyprinids, as the sister group of *Esomus* (Fig. 8C), but again with no boostrap support. Our separate analyses thus reveal that each gene recovers *Paedocypris* at a different position in the phylogeny, mostly with little or no bootstrap support, and that these positions also differ dramatically from each other. It is important to note that only a single gene, *EGR3*, recovers *Paedocypris* in the same well-supported position as the ML analysis of the concatenated data set of Mayden & Chen (2010). We also note that Mayden & Chen's (2010) partial RY-coding yielded the same topology as Mayden & Chen's (2010) analysis of all nucleotide sequences, but the branch that determines the sister group position of *Paedocypris* to the remaining cypriniforms has no bootstrap support (bootstrap value of less than 50%), contrary to what Mayden & Chen (2010: 156) claimed when they stated that 'Maximum likelihood analyses of the six nuclear genes . . . implementing equal weighting and partial RY-coding procedures . . . yielded identical and *strongly supported* [our emphasis] topologies . . . with *Paedocypris* forming a monophyletic group sister to all remaining members of the Cypriniformes, a result that *received excellent bootstrap support* [our emphasis]'. To evaluate further the influence of *EGR3* on the concatenated data set, we also ran three analyses (with 5, 15, and no partitions) with only five genes of the Mayden & Chen (2010) data set, by excluding *EGR3*. The three resulting trees were almost identical (only results from the data set without partition are shown in Fig. S7). The only differences were in the position of the catostomids as sister group to the remaining Cypriniformes (no partition, five partitions) or as sister group to the other Cobtioidei (15 partitions), and in the relative phylogenetic position of *Tinca* and *Tanichthys*, with none of these alternative placements having bootstrap support of > 70%. All three analyses recover *Paedocypris* as the sister group to all other cyprinids including *Psilorhynchus* (see Fig. S7) with low bootstrap support (no partition 73%, five partitions 78%, 15 partitions 82%), and not as the sister group of all other cypriniforms. These analyses independantly confirm the driving influence of *EGR3* in the six-gene concatenated data set.

The results of the AU, SH, and KH tests are shown in Table 2. For the single-gene analyses of the Mayden & Chen (2010) data, the three alternative hypotheses tested (*Paedocypris* sister group to Cypriniformes, *Paedocypris* sister group to Cyprinidae including Psilorhynchidae, and Danioninae monophyly) could not be rejected with *EGR1*, *EGR2B*, and *RAG1*. The topology of *Paedocypris* as sister group to Cyprinidae including Psilorhynchidae was rejected by *EGR3* and *RH*, whereas the monophyly of Danioninae was rejected by *IRBP* (AU only), *EGR3*, and *RH* (AU and KH). Interestingly, the rejection of the monophyly of Danioninae (including *Paedocypris*) by *RH* is not because of *Paedocypris*, which in the ML tree clusters with the majority of the Danioninae, but because of *Boraras*, two *Rasbora* species, and *Trigonostigma*, which cluster outside the remaining Danioninae. In the concatenated Mayden & Chen (2010) data set the monophyly of Danioninae was rejected; however, the hypothesis that *Paedocypris* is the sister group to Cyprinidae was not rejected with SH. It is clear from the data presented in Table 2, and from the single-gene and from the five-gene analyses that *EGR3* plays the most significant role in contributing towards the rejection of Danioninae monophyly in Mayden & Chen's (2010) concatenated data set.

Our ML analysis of the Rüber *et al*. (2007) data set, which comprises a single mitochondrial locus, but with an extensive taxon coverage, supports these authors' previous conclusion that *Paedocypris* is the sister group to *Sundadanio*, and both in return are the sister group to the remaining Danioninae (Fig. 9). Although Rüber *et al*. (2007) used Bayesian inference, our ML reanalysis of their data set did not find significant bootstrap support for the sister-group relationship between *Paedocypris* + *Sundadanio* and the remaining Danioninae (Fig. 9).

None of the three alternative hypotheses (*Paedocypris* as sister group to Cypriniformes, *Paedocypris* as sister group to Cyprinidae, and Danioninae monophyly) could be rejected by the Rüber *et al*. (2007) cyt *b* data set.

Site-wise likelihood analyses of Mayden & Chen's and Rüber et al.*'s data sets*

To further explore the contribution of individual genes in the Mayden & Chen (2010) data set, and positions within a gene in the Mayden & Chen (2010) and Rüber *et al*. (2007) data sets, to the overall phylogenetic signal in favour of a particular hypothesis, we conducted

Table 2. Log-likelihoods and *P* values of AU, KH, and SH tests for alternative topologies evaluated

Data set	ML score	AU	KH	SH
EGR1 (Mayden & Chen, 2010)				
ML tree	−13125.68125	0.597	0.567	0.709
Paedocypris sister group to Cypriniformes	−13127.78084	0.508	0.433	0.626
Paedocypris sister group to Cyprinidae plus Psilorhynchidae	−13128.48288	0.430	0.402	0.674
Danioninae monophyly	−13134.40247	0.123	0.193	0.338
EGR 2 (Mayden & Chen, 2010)				
Paedocypris sister group to Cyprinidae plus Psilorhynchidae (= ML tree)	−11040.24439	0.900	0.899	0.974
Paedocypris sister group to Cypriniformes	−11051.53318	0.162	0.140	0.218
Danioninae monophyly	−11050.38495	0.124	0.101	0.247
EGR3 (Mayden & Chen, 2010)				
Paedocypris sister group to Cypriniformes (= ML tree)	−12467.56589	1.000	1.000	1.000
Paedocypris sister group to Cyprinidae plus Psilorhynchidae	−12526.881	**<0.001**	**<0.001**	**<0.001**
Danioninae monophyly	−12534.01928	**<0.001**	**<0.001**	**<0.001**
IRBP (Mayden & Chen, 2010)				
ML tree	−17965.42361	0.772	0.679	0.898
Paedocypris sister group to Cypriniformes	−17969.5947	0.258	0.198	0.509
Paedocypris sister group to Cyprinidae plus Psilorhynchidae	−17967.7222	0.445	0.321	0.712
Danioninae monophyly	−17984.500	**0.043**	0.059	0.085
RAG1 (Mayden & Chen, 2010)				
ML tree	−26957.90366	0.615	0.585	0.780
Paedocypris sister group to Cypriniformes	−26961.78093	0.421	0.399	0.604
Paedocypris sister group to Cyprinidae plus Psilorhynchidae	−26960.50144	0.527	0.415	0.776
Danioninae monophyly	−26975.81833	0.085	0.121	0.216
RH (Mayden & Chen, 2010)				
ML tree	−15373.45489	0.748	0.685	0.903
Paedocypris sister group to Cypriniformes	−15381.19347	0.364	0.315	0.623
Paedocypris sister group to Cyprinidae plus Psilorhynchidae	−15441.65795	**0.004**	**0.003**	**0.006**
Danioninae monophyly	−15410.87908	**0.042**	**0.040**	0.133
Combined (Mayden & Chen, 2010)				
Paedocypris sister group to Cypriniformes (= ML tree)	−99007.16684	0.994	0.990	0.993
Paedocypris sister group to Cyprinidae plus Psilorhynchidae	−99033.19067	**0.008**	**0.010**	0.111
Danioninae monophyly	−99089.08194	**0.001**	**<0.001**	**<0.001**
BMC partition one (Rüber *et al*., 2007)				
Danioninae monophyly (= ML tree)	−76969.22416	0.891	0.863	0.943
Paedocypris sister group to Cypriniformes	−77167.90495	0.097	0.076	0.123
Paedocypris sister group to Cyprinidae plus Psilorhynchidae	−77136.36058	0.181	0.137	0.198

P values < 0.05 (shown in bold) indicate that the data allow rejecting the respective alternative topology. Abbreviations: AU, approximately unbiased; KH, Kishino-Hasegawa; SH, Shimodaira-Hasegawa.

site-wise likelihood analyses (Evans *et al*., 2010). The results from the concatenated Mayden & Chen (2010) data set are shown in Figure 10A and B, and those from the Rüber *et al*. (2007) data set are shown in Figure 11A and B. In Figure 10A positive Δpsln *L* values on the *y*-axis correspond to positions in the nucleotide alignment that exhibit a higher likelihood for the hypothesis that *Paedocypris* is sister group to the Cypriniformes, and negative Δpsln *L* values correspond to positions that exhibit a higher likelihood for the hypothesis that *Paedocypris* is sister group to the Cyprinidae. Similarly, in Figure 10B positive Δpsln *L* values correspond to positions that exhibit a higher likelihood for the hypothesis that *Paedocypris* is the sister group to the Cypriniformes, and negative Δpsln *L* values correspond to positions that exhibit a higher likelihood for the hypothesis that *Paedocypris* is a member of the Danioninae. None of the Δpsln *L* distributions show a strong pattern, i.e. the majority of nucleotide positions do not favour a particular phylogenetic hypothesis. Most Δpsln *L* values are around zero, thus indicating a near equal fit to both alternative hypotheses in each pairwise comparison, and do not favour one topology over the other. For *EGR3* we found no or only very few strongly negative (less than −1) Δpsln *L* values, compared with a large

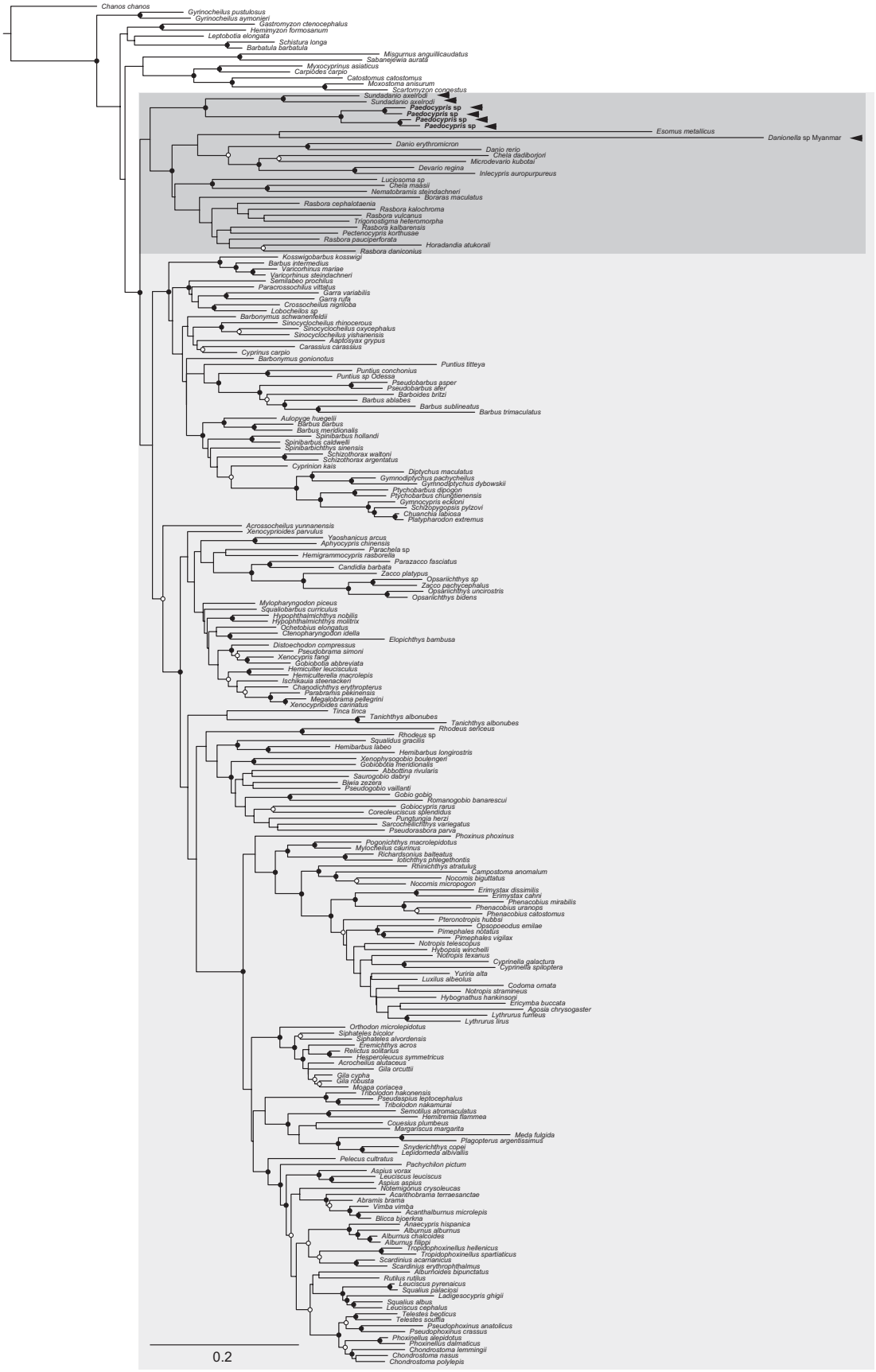

Figure 9. Maximum-likelihood (ML) tree of the Rüber *et al*. (2007) data set comprising 1131 bp of mitochondrial gene cyt *b* with bootstrap support values added to branches. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The miniature taxa *Paedocypris* (in bold), *Danionella*, and *Sundadanio* are highlighted with a black arrowhead. The scale bar indicates substitutions per site.

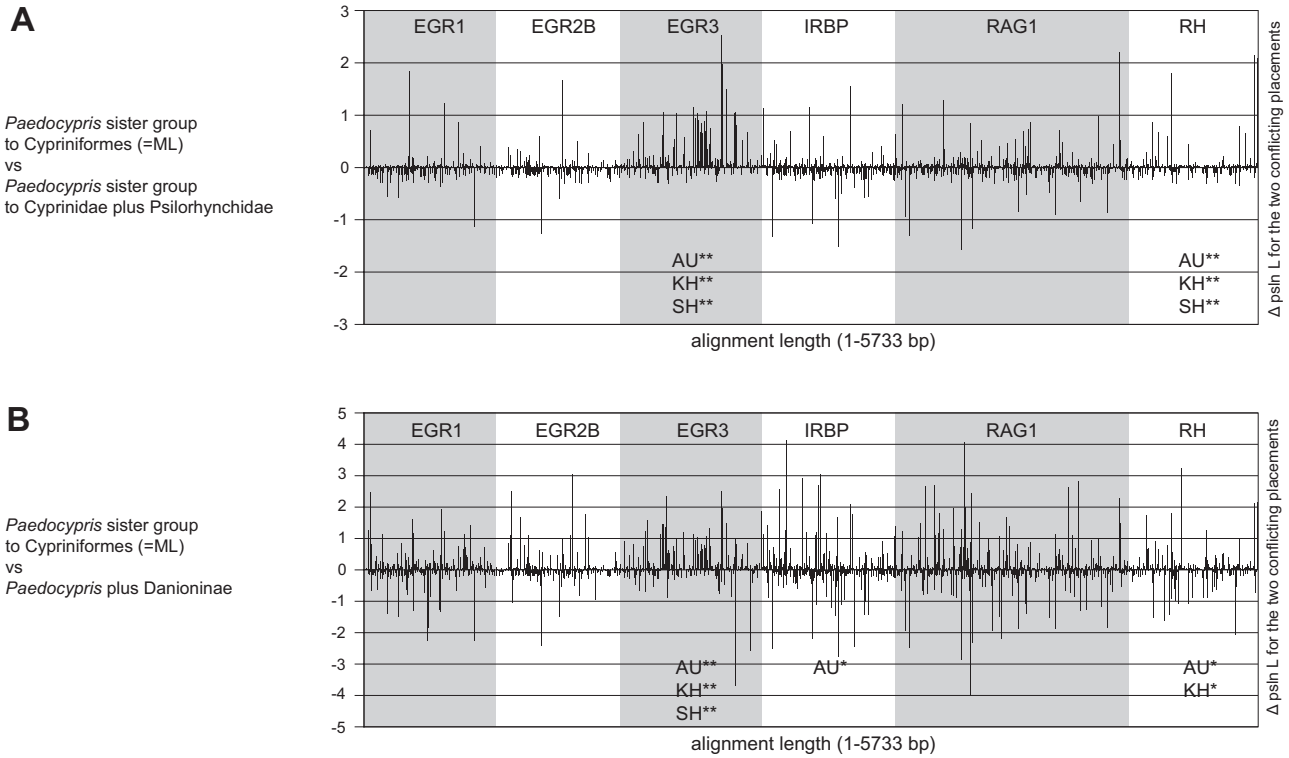

Figure 10. Differences in the per-site log-likelihood (Δpsln *L*) for the different hypotheses regarding the phylogenetic placement of *Paedocypris* based on the concatenated Mayden & Chen (2010) data set consisting of six nuclear markers and a total alignment length of 5733 bp (*x*-axis). In (A), the hypotheses that *Paedocypris* is sister group to the Cypriniformes (ML tree) versus *Paedocypris* being sister group to Cyprinidae plus Psilorhynchidae are compared, and in (B) the hypotheses that *Paedocypris* is sister group to the Cypriniformes (ML tree) versus *Paedocypris* being a member of the Danioninae are compared. A positive value on the *y*-axis indicates a position in the alignment with a higher per-site log-likelihood value than the alternative hypothesis, and a negative value indicates a position in the alignment with a higher per-site log-likelihood value for the alternative hypothesis than for the ML topology. Significant results for the AU, KH, and SH tests using single-gene analyses are also indicated (see also Table 2). Abbreviations: AU, approximately unbiased; KH, Kishino-Hasegawa; SH, Shimodaira-Hasegawa.

accumulation of positive Δpsln *L* values in the middle of the gene. Similarly for *RH*, positive Δpsln *L* values dominate in the comparison between *Paedocypris* as sister group to Cypriniformes and *Paedocypris* as sister group to Cyprinidae (Fig. 10A). These results are thus in good agreement with the AU, KH, and SH tests, namely that *EGR3*, and to a lesser extent *RH*, are responsible for the rejection of the hypotheses that *Paedocypris* is either sister group to Cyprinidae or a member of the Danioninae.

In the sitewise likelihood analysis of the Rüber *et al*. (2007) data set shown in Figure 11A, positive Δpsln *L* values on the *y*-axis correspond to positions in the nucleotide alignment that exhibit a higher likelihood for the hypothesis that *Paedocypris* is a member of the Danioninae, and negative Δpsln *L* values correspond to positions that exhibit a higher likelihood for the hypothesis that *Paedocypris* is the sister group to the Cypriniformes. Similarly, in Figure 11B positive Δpsln *L* values on the *y*-axis correspond to positions in the nucleotide alignment that exhibit a higher likelihood for the hypothesis that *Paedocypris* is a member of the Danioninae, and negative Δpsln *L* values correspond to positions that exhibit a higher likelihood for the

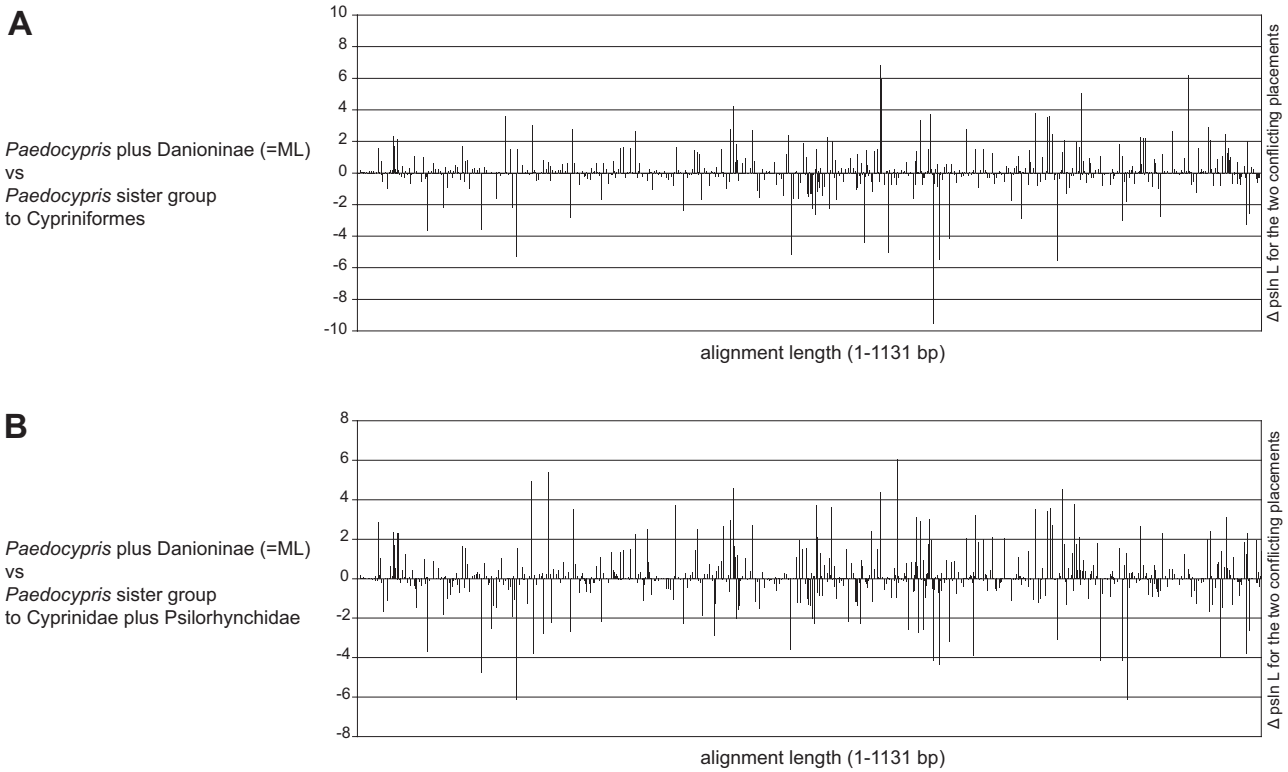

Figure 11. Differences in the per-site log-likelihood (Δpsln *L*) for the different hypotheses regarding the phylogenetic placement of *Paedocypris* based on the Rüber *et al*. (2007) data set consisting of the mitochondrial cyt *b* gene and a total alignment length of 1131 bp (*x*-axis). In (A) the hypotheses that *Paedocypris* is a member of the Danioninae (ML tree) versus *Paedocypris* being sister group to the Cypriniformes are compared, and in (B) the hypotheses that *Paedocypris* is a member of the Danioninae (ML tree) versus *Paedocypris* being sister group to Cyprinidae plus Psilorhynchidae are compared. A positive value on the *y*-axis indicates a position in the alignment with a higher per-site log-likelihood value than the alternative hypothesis, and a negative value indicates a position in the alignment with a higher per-site log-likelihood value for the alternative hypothesis than for the ML topology.

hypothesis that *Paedocypris* is the sister group to the Cyprinidae. The sitewise likelihood analysis of the Rüber *et al*. (2007) data set thus resembles that of the Mayden & Chen (2010) data set in that both positive and negative Δpsln *L* values are nearly evenly distributed along the cyt *b* alignment, and in that no specific topology is favoured over another. Again, these results are in good agreement with the AU, KH, and SH tests, showing that with the cyt *b* data set alone the different alternative hypotheses regarding the phylogenetic position of *Paedocypris* cannot be rejected.

Phylogenetic networks of Mayden & Chen's and Rüber et al.*'s data sets*

Phylogenetic networks are increasingly used in cases in which the evolutionary history of a set of taxa is poorly described by a strictly bifurcating phylogenetic tree (Huson & Bryant, 2006). This is particularly the case when reticulate events such as hybridization, horizontal gene transfer, recombination, or gene duplication and loss are suspected; however, one class of phylogenetic networks, termed splits networks, can be used to represent incompatible and ambiguous signals in a data set. Here, parallel edges, rather than single branches, are used to represent incompatible splits in the data.

Phylogenetic networks based on either logDet or p-distance transformations gave very similar results, and hence we only present those based on logDet distances. The split tree of the Mayden & Chen (2010) data set is shown in Figure 12, and the split tree from the Rüber *et al*. (2007) data set is shown in Figure 13. In the Mayden & Chen (2010) split tree the Danioninae taxa are not all grouped together. An assemblage of long branches includes *Paedocypris* and the four outgroup taxa. In addition, the miniature taxa *Danionella* and *Sundadanio* are placed apart from the remaining Danioninae. The group-supporting signal seems generally low and the central part of the graph is dominated by many contradicting edges.

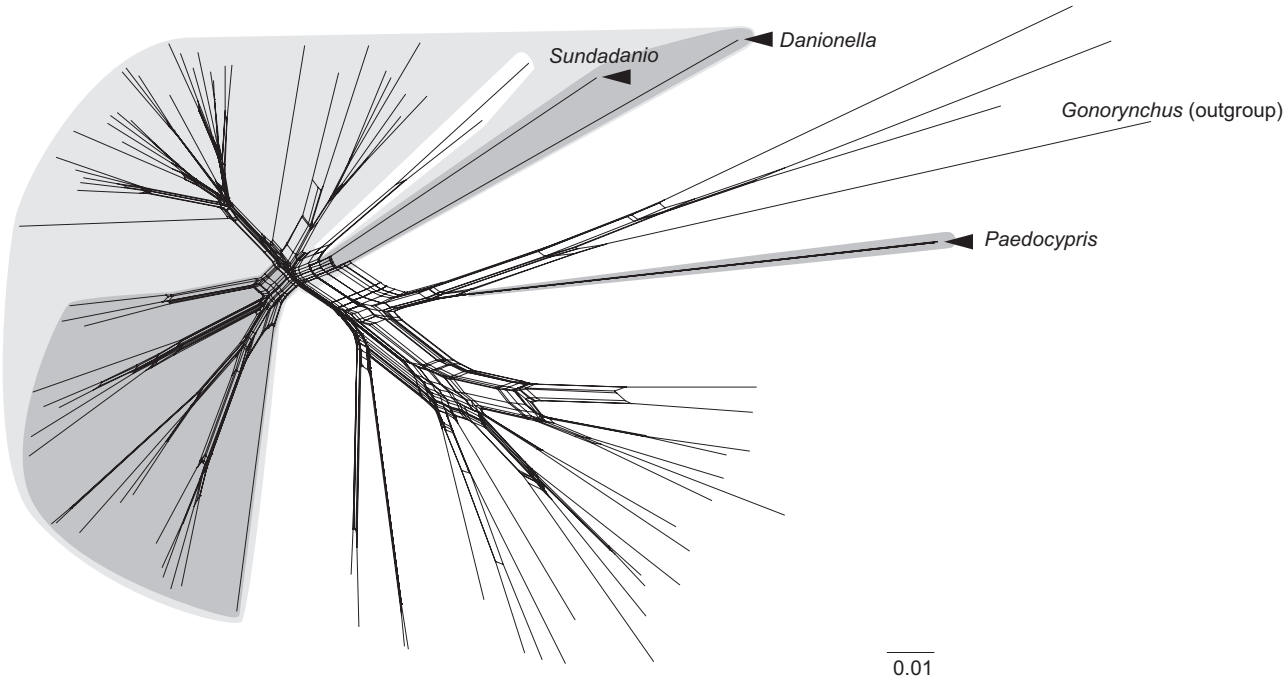

Figure 12. The split-tree phylogenetic networks based on logDet distance transformations of the Mayden & Chen (2010) data set. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The outgroup (*Gonorynchus*) and the danionine genera *Danionella*, *Paedocypris*, and *Sundadanio* are labeled, and the three miniatures are further highlighted with arrowheads.

In the split tree of the Rüber *et al*. (2007) data set (Fig. 13) all Danioninae group together and the three developmentally truncated miniatures (*Paedocypris*, *Sundadanio*, and *Danionella*) do not show extremely long branches, in contrast to the split tree resulting from the Mayden & Chen (2010) data set. As in the Mayden & Chen (2010) data set, the network derived from the Rüber *et al*. (2007) data set also shows that the central part of the graph is dominated by many contradicting edges.

Splits analysis methods (SAMS) of Mayden & Chen's and Rüber et al.*'s data sets*

It is often challenging to discern the phylogenetic signal from noise in molecular phylogenetics (e.g. Wägele & Rödding, 1998; Jeffroy *et al*., 2006; Wägele *et al*., 2009; Philippe *et al*., 2011; Betancur-R. *et al*., 2013a). In the presence of sufficient phylogenetic signal, random noise should not affect the recovery of the true relationships; however, non-random noise will introduce conflicting signal and might obscure true relationships. Hence, exploratory analyses using split decomposition methods are a useful way to look at the strength of the phylogenetic signal in an alignment. We looked at the signal-to-noise ratio in the concatenated Mayden & Chen (2010) data set and in the Rüber *et al*. (2007) data set using SAMS (Wägele & Mayer, 2007) that allowed us to identify conserved split-supporting positions without reference to a tree, and to compare these with the actual splits obtained in the ML analyses. Figure 14 shows the splits-support spectrum analyses of the Mayden & Chen (2010) data set. The first four splits, which are compatible with the ML tree, show markedly higher support than the remaining splits. The best-supported split separates the two *Paedocypris* species from the rest, including all outgroup taxa, and is thus evidence for the highly autapomorphic DNA sequences of their analysed genes. Regarding the phylogenetic position of *Paedocypris*, however, none of these four best splits is relevant. The remaining splits show low support, a slow decrease in support, and numerous incompatible splits are interspersed with compatible splits. It is important to note that split number 6 (*Paedocypris* plus non-cypriniform outgroups versus all other cypriniforms), the one responsible for the basal position of *Paedocypris*, is not supported by a single binary position, but by a few asymmetrical ones and a larger number of noisy positions. In addition, there are several mutually incompatible splits, including *Paedocypris* or with *Paedocypris* and one or more of the outgroups. Thus the Mayden & Chen (2010) data set does not seem informative regarding the phylogenetic position of *Paedocypris*. It is also interesting that the existence of this larger number of mutually

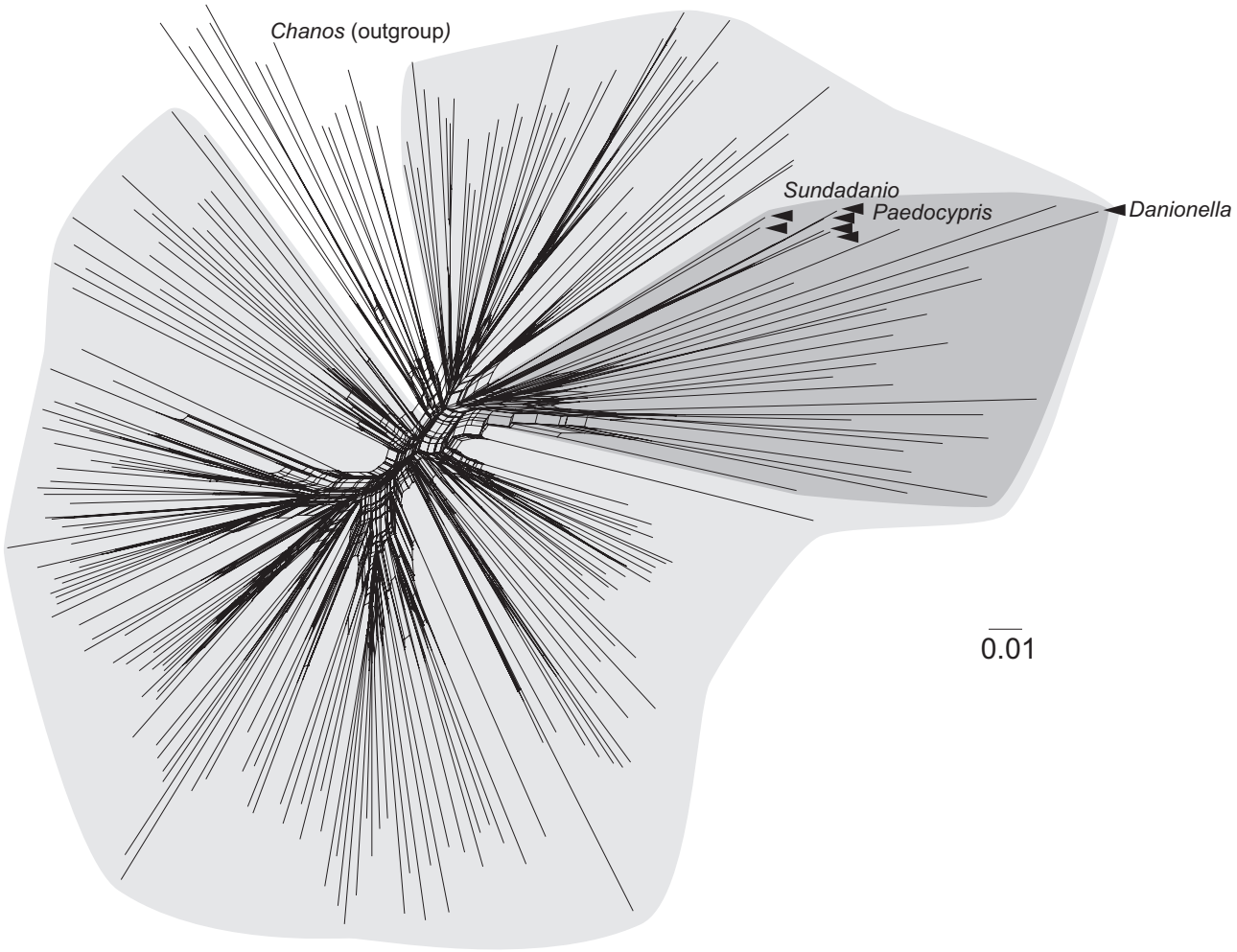

Figure 13. The split-tree phylogenetic networks based on logDet distance transformations of the Rüber *et al*. (2007) data set. Members of Cyprinidae are highlighted in light grey; members of Danioninae are highlighted in dark grey. The outgroup (*Gonorynchus*) and the danionine genera *Danionella*, *Paedocypris*, and *Sundadanio* are labeled, and the three miniatures are further highlighted with arrowheads.

incompatible splits was not apparent from the published tree topology. To a lesser degree mutually incompatible splits are also identified involving *Danionella* or *Sundadanio* (Fig. 14). The many mutually incompatible splits with the long-branch species *Paedocypris* are indicative of a long-branch effect, as explained by Wägele & Mayer (2007; for a discussion, see below). The splits-support spectrum analyses of the Rüber *et al*. (2007) data set is shown in Figure 15. Here, among the first ten best splits, five were compatible with the ML tree, although none are relevant regarding the phylogenetic position of *Paedocypris*. Support was low, and again, numerous incompatible splits are interspersed with compatible splits.

Highly incongruent phylogenetic tree topologies between different phylogenetic studies using one or a few genes, like the dramatic differences in the phylogenetic position of *Paedocypris* between Mayden & Chen (2010) and Rüber *et al*. (2007), are not uncommon (Jeffroy *et al*., 2006; Philippe *et al*., 2011; Salichos & Rokas, 2013). Multigene phylogenies are thought to be superior to single-gene phylogenies because of an increase in phylogenetically informative positions, which reduce sampling error and hence increase resolution. Because at the same time the potential for systematic errors increases with increasing alignment length, however, the accuracy of the obtained results does not necessarily increase. Hence, multigene phylogenies do not necessarily lead to correct tree topologies, and on the contrary may result in wrong, yet statistically highly supported trees. Systematic errors are the result of misspecifications in the model of sequence evolution and include: across-site rate variation, heterotachy (shifts in site-specific rates over

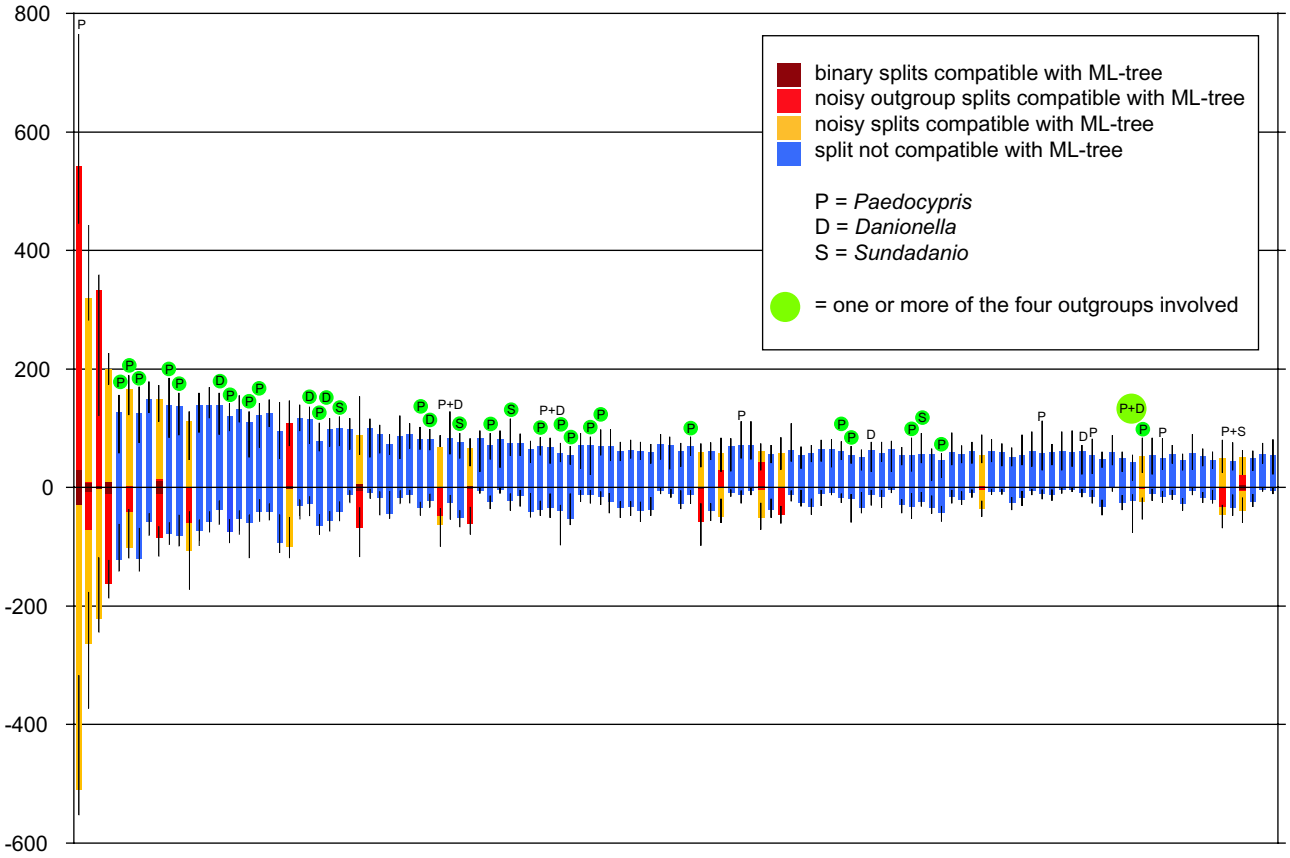

Figure 14. Splits-support spectrum for the concatenated Mayden & Chen (2010) data set. The first 120 best-supported splits are shown on the horizontal axis, sorted according to column heights. The number of sequence positions supporting a particular split is indicated on the vertical axis by the height of each column, and each partition of a split is indicated separately above and below the horizontal axis. Blue columns indicate splits that are not compatible with the maximum-likelihood (ML) topology shown in Figure 6. The sequence positions supporting a split compatible with the ML topology are composed of binary splits (dark red), noisy outgroup splits (red), and noisy splits (orange), following Wägele & Mayer (2007). Splits in which one or more of the four outgroups were involved are highlighted with a green filled circle, and those in which one of the danionine genera *Paedocypris*, *Danionella*, or *Sundadanio* were involved are highlighted with the letter P, D, or S, respectively. The first ten splits are: 1, *Paedocypris* (2 spp.) versus the rest; 2, Gonorynchiformes (*Gonorynchus*) + Siluriformes (*Pseudobagrus*) + Characiformes (*Phenacogrammus* and *Chalceus*) versus the rest; 3, *Gyrinocheilus* (2 spp.) versus the rest; 4, Characiformes (2 spp.) versus the rest; 5, Gonorynchiformes + Characiformes (*Phenacogrammus*) + *Paedocypris* (2 spp.) versus the rest; 6, Gonorynchiformes + Siluriformes + Characiformes (2 spp.) + *Paedocypris* (2 spp.) versus the rest; 7, Gonorynchiformes + *Siluriformes* + *Paedocypris* (2 spp.) versus the rest; 8, Gonorynchiformes + *Phenacogrammus* versus the rest; 9, Siluriformes + Characiformes (2 spp.) versus the rest; 10, Characiformes (2 spp.) + *Paedocypris* (2 spp.) versus the rest.

time), site-interdependent evolution, compositional heterogeneity, and site-heterogeneous nucleotide/amino acid replacement. The apparent signal arising from these model misspecifications is often referred to as 'non-phylogenetic' or 'misleading' phylogenetic signal, which becomes particularly problematic when dealing with mutational saturation in long branches. Long branches caused by fast evolutionary rates are affected by long-branch attraction (clustering of long-branch taxa irrespective of their true phylogenetic relationship), causing systematic errors (Felsenstein, 1978). Long-branch attraction is a common, but often neglected, phenomenon in phylogenetics. Wägele & Mayer (2007) distinguished three classes of long-branch effects (LBEs): class I, the symplesiomorphy trap; class II, erosion of phylogenetic signal; class III, misleading and invisible attraction as a result of non-homologous similarities (parallel substitutions). All three classes are known to obscure phylogenetic relationships in existing molecular data sets.

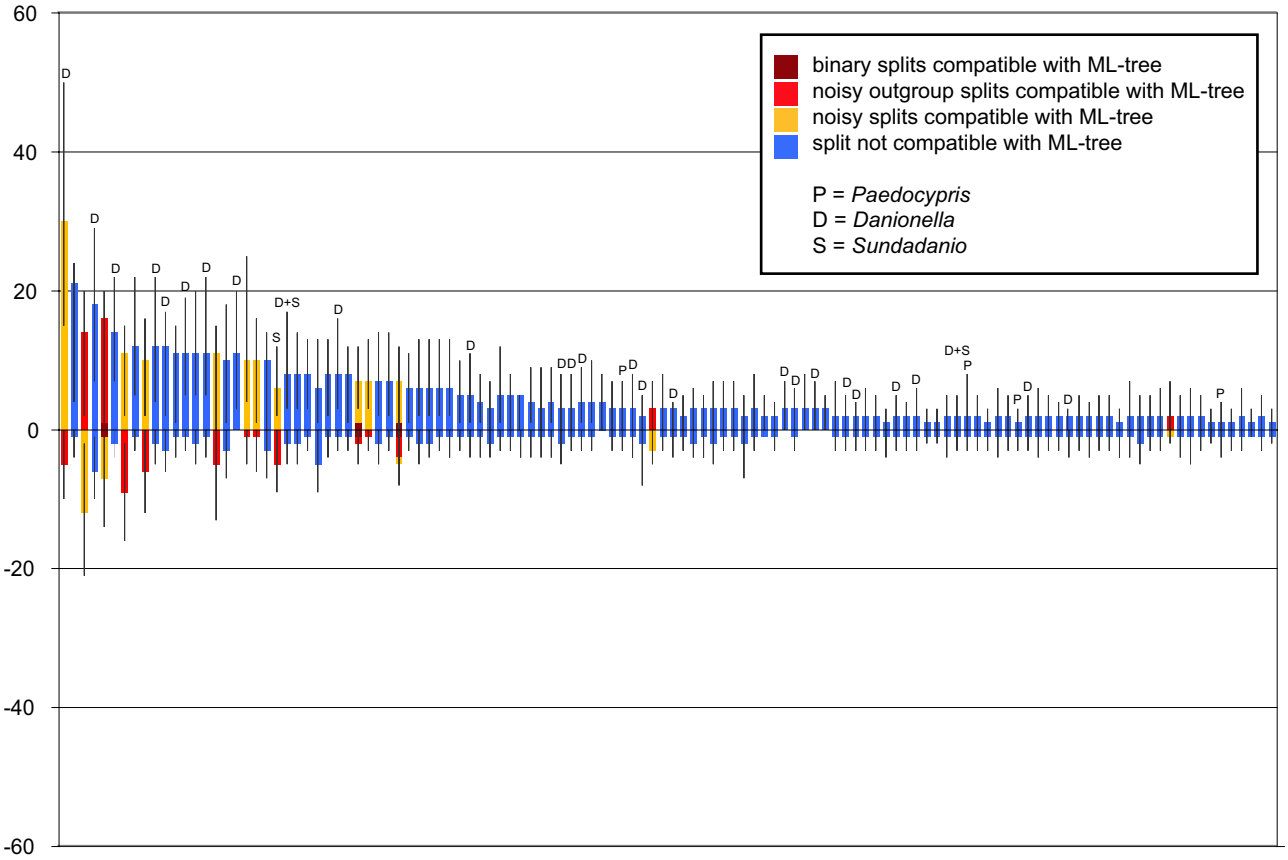

Figure 15. Splits-support spectrum for the Rüber *et al*. (2007) data set. The first 120 best-supported splits are shown on the horizontal axis, sorted according to column heights. The number of sequence positions supporting a particular split is indicated on the vertical axis by the heights of each column, and each partition of a split is indicated separately above and below the horizontal axis. Blue columns indicate splits that are not compatible with the ML topology shown in Figure 6. The sequence positions supporting a split compatible with the ML topology are composed of binary splits (dark red), noisy outgroup splits (red), and noisy splits (orange), following Wägele & Mayer (2007). Splits in which one of the danionine genera *Paedocypris*, *Danionella*, or *Sundadanio* were involved are highlighted with the letter P, D, or S, respectively. The first ten splits are: 1, *Danionella* + *Esomus* versus the rest; 2, *Luciosoma* + *Esomus* versus the rest; 3, *Meda* + *Plagopterus* versus the rest; 4, *Danionella* + *Puntius titteya* versus the rest; 5, *Scartomyzon* + *Moxostoma* versus the rest; 6, *Danio rerio* + *Danionella* versus the rest; 7, *Inlecypris* + *Devario* versus the rest; 8, *Inlecypris* + *Puntius titteya* versus the rest; 9, *Barbatula* + *Schistura* versus the rest; 10, *Danionella* + *Chanos* versus the rest.

We have shown that the phylogenetic position of *Paedocypris* as sister group of Cypriniformes in the Mayden & Chen (2010) analyses is solely driven by one gene, *EGR3*, in the concatenated data set of six nuclear genes. When analysed separately each of the six genes resulted in highly incongruent topologies, providing indication that either class-II LBE alone or in combination with a class-III LBE might be responsible for the placement of *Paedocypris* as sister group to the Cypriniformes in the Mayden & Chen (2010) data set. We hypothesize that erosion of phylogenetic signal, and/or misleading and invisible attraction because of non-homologous similarities along the long branch leading to *Paedocypris*, has 'dragged' this taxon towards the root of the tree. There are two ways in which saturation arising from multiple hits can cause LBE: multiple hits can either destroy phylogenetic signal when too few conserved synapomorphies remain after some time (class-II effect) or when substitutions produce non-homologous similarities (class-III effects) (Wägele & Mayer, 2007). Whereas Wägele & Mayer (2007: 5) acknowledged that class-II LBEs are difficult to detect, they noted that 'a conflict between morphology and molecular data in combination with the occurrence of long branches should be alarming'.

Three nuclear markers (*EGR1*, *EGR2B*, and *EGR3*) employed by Mayden & Chen (2010) belong to the early growth response (*EGR*) gene family. Chen *et al*. (2008) identified four *EGR* copies in teleosts (*EGR1*, *EGR2A*,

EGR2B, and *EGR3*), but did not find a homologous (or significantly similar) copy to the mammalian *EGR4*. The presence of the two *EGR2* copies (*EGR2A* and *EGR2B*) was attributed to a fish-specific genome duplication. It can be speculated that the surprising results obtained with *EGR3* regarding the phylogenetic position of *Paedocypris* as the sister group of Cypriniformes might be the result of strong selection pressure in *Paedocypris* because *EGR* functions as an early growth response factor. Alternatively, it could be possible that the *EGR3* of *Paedocypris* is not homologous with *EGR3* of the remaining Cypriniformes, but rather a paralog. These possibilities need to be explored in the future.

The cyt *b* data set employed by Rüber *et al*. (2007) is also very limited in confidently placing *Paedocypris* among Cypriniformes (Figs 11, 13, 15; Table 2), as our reanalyses show.

We conclude from the reassessment of the molecular evidence regarding the phylogenetic position of *Paedocypris* that the molecular markers thus far employed may not be efficient/sufficient in confidently addressing the phylogenetic position of this remarkable cyprinid. In contrast the detailed morphological analysis and character evaluation we have presented above seems to provide a better-supported hypothesis of where *Paedocypris* belongs.

Improved models of sequence evolution, more appropriate to handle real-world data, are continuing to be developed and will certainly help reduce non-phylogenetic signal in systematic studies in the future. In the meantime, one of the most effective ways to avoid non-phylogenetic signal in an alignment is data exclusion. It has become clear in recent studies that the emphasis in phylogenetic studies should not be merely on the quantity of data, but also on its quality (e.g. issues of non-stationarity and saturation), an issue particularly important in the area of phylogenomics where increasingly large quantities of genomic data for non-model taxa are becoming available for phylogeny inference (Jeffroy *et al*., 2006; Wägele & Mayer, 2007; Jenner & Littlewood, 2010; Philippe *et al*., 2011). Selecting data in such a way as to minimize non-phylogenetic signal is an important step away from phylogenetic incongruence towards more accurate phylogenetic trees (Jeffroy *et al*., 2006; Philippe *et al*., 2011).

MAYDEN & CHEN'S NEW FAMILY GROUP NAMES AND COMMENTS ON RECENT STUDIES ON *PAEDOCYPRIS*

In addition to their aim to determine the phylogenetic position of *Paedocypris*, Mayden & Chen (2010) also created some level of confusion relating to their nomenclatural actions. Mayden & Chen (2010) erected new families for three cyprinid genera, continuing the approach started by Chen & Mayden (2009), who aimed to introduce the names Leptobarbidae and Tanichthyidae, but also used Acheilognathidae, Tincidae, Leuciscidae, and Gobionidae in their figure 2. As family group names are regulated by the International Code of Zoological Nomenclature, Tanichthyidae is not an available name from Chen & Mayden (2009), as it fails to meet the criteria set out by the code. Mayden & Chen (2010) erected a new family, Paedocyprididae, and a new superfamily Paedocypridoidea, for the genus *Paedocypris*, and introduced a new family, Sundadanionidae, for the genus *Sundadanio*. They also provided a diagnosis for Tanichthyidae to correct their earlier *lapsus* (Chen & Mayden, 2009), and to make that name available. The extensive diagnosis for Paedocyprididae in Mayden & Chen (2010) was copied verbatim from Kottelat *et al*.'s (2006) diagnosis, and even contains all the figure references to Kottelat *et al*.'s (2006) original paper and several incorrectly converted symbols. Mayden & Chen's (2010) erection of the family Paedocyprididae on page 172 can probably be considered the shortest-lived family group name ever proposed in ichthyological history, as their proposal was rejected by Tang *et al*. (2010), who synonymized Paedocyprididae with Danioninae on p. 208 in the same issue of *Molecular Phylogenetics and Evolution*. We further note here that the three family diagnoses for Paedocyprididae, Sundadanionidae, and Tanichthyidae were added to the paper after it was accepted, and after the accepted version of the article was published on the journal's webpage. This is just one of the many significant differences between the accepted draft of Mayden & Chen (2010) and the printed version. In addition to the diagnoses of the three family group names, large parts of the text of the accepted manuscript were changed after this was made available online, including the addition of Mayden & Chen's (2010: 155) heavy criticism of Britz & Conway's (2009) study as 'idealistic morphology'. A total of 15 new references were also added after Mayden & Chen's (2010) accepted manuscript was made available online. We find the entire procedure of how an accepted manuscript that had gone through peer review was changed significantly in the proof stage highly unusual and disturbing. It is deeply unsettling that Mayden & Chen's (2010) manuscript needed and received these significant changes after peer review and after it was accepted.

Mayden & Chen's (2010) various new family group names and new classification have been mostly ignored (Tang *et al*., 2011, 2013) or even rejected outright (Tang *et al*., 2010; Wang *et al*., 2012) by members of the same research group. In other cases the new names

have been applied inconsistently, again by the same research group. For example, Sulaiman & Mayden (2012) used Leptobarbidae, but used Paedocyprinidae instead of Paedocypprididae and failed to use Sundadanionidae for *Sundadanio*, which they classified in Cyprinidae. And although Bufalino & Mayden (2010a, b, c) used Leuciscidae for the North American minnows, Schönhuth *et al*. (2012) continued to use Cyprinidae. It seems that the issues raised by Britz & Conway (2011) continue.

The scientific interest in the highly miniaturized fishes of the genus *Paedocypris* has continued, and additional studies have highlighted the puzzling combination of characters influenced by developmental truncation and highly derived novel characters. Recently, Liu *et al*. (2012) studied the chromosome structure in two species of the genus *Paedocypris* (*Paedocypris carbunculus* Kottelat & Britz, 2008 and *Paedocypris* sp. Singkep). They found a diploid number of 30 chromosomes in *P. carbunculus* and 34 in *P*. sp. Singkep. The very low DNA content per cell of 0.36 pg, indicating a genome size of approximately 350 Mb in *Paedocypris* [compared with 0.34 pg in *Sphoeroides spengleri* (Bloch, 1785), the fish with the smallest genome reported to date] was interpreted by Liu *et al*. (2012) as a case of genome miniaturization. Surprisingly, Liu *et al*. (2012: 5) concluded that 'Cytogenetically, *Paedocypris* cannot be placed in any cypriniform taxa at subfamily, family and superfamily levels'. As the reason why this was impossible, they argued that: 'Extensive cytogenetic studies in Cypriniformes have revealed chromosome numbers in more than 200 species belonging to all the six families, most of them have ≥ 48 chromosomes. With 30–34 chromosomes in diploid cells, *Paedocypris* clearly distinguishes itself from the three described superfamilies. If *Paedocypris* indeed belongs to the Cypriniformes, it forms a new superfamily, Paedocyproidea'. This argument that difference means non-membership is flawed when looked at in a phylogenetic context. Kottelat *et al*. (2006) and Britz & Conway (2009) have demonstrated that *Paedocypris* has numerous characters that are unique among cypriniforms or even teleosts. These characters are, of course, autapomorphies of this genus, and do not in any way exclude them from being cypriniforms or teleosts. In the same sense, the low DNA content and chromosome number of *Paedocypris* compared with other cypriniforms only reiterate the common theme highlighted previously by Kottelat *et al*. (2006), Rüber *et al*. (2007), Britz & Kottelat (2008), and Britz & Conway (2009): *Paedocypris* shows a puzzling combination of larval characters, because of its high degree of developmental truncation and highly unusual, often unique, novelties. And the small number of chromosomes and low DNA content of its cells are no exceptions.

In their short review of metamorphosis in teleosts, McMenamin & Parichy (2012) commented briefly on *Paedocypris* and *Danionella* as examples of paedomorphic freshwater fishes. Inexplicably, they cite Mayden & Chen (2010) for the fact that *Paedocypris* and *Danionella* are 'failing to undergo normal metamorphosis and becoming reproductively mature while maintaining a larva-like overall morphology', whereas Mayden & Chen (2010) actually argued against Britz & Conway's (2009) interpretation that the many bone absences in *Paedocypris* are the result of developmental truncation.

Recently, *Paedocypris* has also been included in an analysis on the biogeography of the teleost order Otophysi by Chen *et al*. (2013). The authors used essentially the same gene set as Mayden & Chen (2010) but excluded the *IRBP* gene. Consequently, *Paedocypris* was resolved as the sister group to all other Cypriniformes, as in Mayden & Chen (2010). The position of *Paedocypris* in their tree influences, of course, the age estimates for the order Cypriniformes and pushes their age from 96.1 back to 117.7 Myr, or from 126.7 to 158.9 Myr, respectively, depending on the analysis (Chen *et al*., 2013: fig. 3 versus fig. S6). Its position also influences the ancestral area reconstruction for Cypriniformes and its subgroups. If, as we have shown above, the phylogenetic position of *Paedocypris* at the base is an artifact and a consequence of interacting long-branch effects, then the age estimates in Chen *et al*. (2013) as well as their area reconstruction in their historical biogeographic analyses will have to be revised.

CONCLUDING REMARKS

Species of the genus *Paedocypris* are among the most unusual fishes described in the last few decades. Their interesting characters, be they anatomical, physiological, behavioural, or genetic, will continue to fascinate and puzzle a broad range of researchers. Our studies of *Paedocypris* have highlighted several important aspects, of which one is the difficulty to work with developmentally truncated animals and how to analyse their confusing character combinations. We are convinced that an evaluation of reductive characters in highly developmentally truncated organisms is an important step before any phylogenetic analysis can be performed.

Trying to understand the dramatic differences in the phylogenetic placement between the molecular analyses of Rüber *et al*. (2007) and Mayden & Chen (2010) has also forced us to look beyond the number of taxa, genes, and nucleotides, and to try to find ways to evaluate the quality of data sets and not rely on quantity only. This led us to apply three different methods that help to visualize and assess phylogenetic signal and

noise in molecular data sets, and that are independent of the phylogenetic analysis itself. The current practice of assuming that the latest analysis with more genes and more taxa produces the 'best hypothesis' will need to be replaced by the thorough evaluation of the quality of molecular datasets (see Mooi & Gill, 2010; *Zootaxa* special issue by Carvalho & Craig, 2011; Wägele *et al*., 2009; Wägele & Mayer, 2007; Jenner & Littlewood, 2010; Philippe *et al*., 2011). This will also finally enable systematists to actually compare molecular trees with each other, which is badly needed, especially because contradictory phylogenetic trees seem to get published within ever-shorter periods of time (see the latest example of flatfish monophyly in Betancur-R. *et al*., 2013a and Betancur-R. & Ortí, 2014 versus flatfish paraphyly in Campbell, Chen & Lopez, 2013; Campbell, Chen & Lopez, 2014; and Betancur-R. *et al*., 2013b, 2013c, 2014). Because of the prevalence of non-phylogenetic signal in phylogenomics, applying molecular data sets to difficult phylogenetic problems does not necessarily mean that the most extensive taxon and nucleotide sampling will provide the most accurate phylogenetic hypothesis (Philippe *et al*., 2011). The resolution of the tree of life largely depends on efficient ways to prevent deleterious effects of non-phylogenetic signal, such as better procedures for the selection of orthologous genes suitable for a certain phylogenetic problem, as well as novel models of sequence evolution that can better account for systematic errors in phylogenetics.

ACKNOWLEDGEMENTS

We thank Richard Mayden and Wei-Jen Chen for publishing their 2010 paper that pushed us to embark on this study. We are grateful to Paula Mabee and Maurice Kottelat for sending us specimens at short notice. We also thank Gavin Naylor for discussions on the molecular aspects of this project. Five reviewers (Randall Mooi and four anonymous reviewers) provided useful comments, critical remarks, and thoughtful suggestions that made us go back to our analyses, and that helped to improve the paper.

REFERENCES

Begle DP. 1991. Relationships of the osmeroid fishes and the use of reductive characters in phylogenetic analysis. *Systematic Zoology* **40:** 33–53.

Betancur-R. R, Broughton RE, Wiley EO, Carpenter K, Lopez JA, Li C, Holcroft NI, Arcila D, Sanciangco M, Cureton J, Zhang F, Buser T, Campbell M, Rowley T, Ballesteros JA, Lu G, Grande T, Arratia G, Ortí G. 2013b. The tree of life and a new classification of bony fishes. *PLoS Currents Tree of Life*. 2013 Apr 18.

Betancur-R. R, Li C, Munroe TA, Ballesteros JA, Orti G. 2013a. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Systematic Biology* **62:** 763–785. doi: 10.1093/sysbio/syt039.

Betancur-R. R, Ortí G. 2014. Molecular evidence for the monophyly of flatfishes (Carangimorpharia: Pleuronectiformes). *Molecular Phylogenetics and Evolution* **73:** 18–22.

Betancur-R. R, Wiley EO, Miya M, Lecointre G, Bailly N, Ortí G. 2013c. *New and revised classification of bony fishes version 2*. Version date: 27 Nov 2013. Available at: http://www.deepfin.org/Classification_v2.htm

Betancur-R. R, Wiley E, Bailly N, Miya M, Lecointre G, Ortí G. 2014. *Phylogenetic Classification of Bony Fishes, Version 3*. Available at: http://www.deepfin.org/Classification_v3.htm

Bird NC, Mabee PM. 2003. Developmental morphology of the axial skeleton of the zebrafish, *Danio rerio* (Ostariophysi: Cyprinidae). *Developmental Dynamics* **228:** 337–357.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444:** 85–88.

Britz R. 1997. Egg surface structure and larval cement glands in nandid and badid fishes (Teleostei, Percomorpha), with remarks on phylogeny and zoogeography. *American Museum Novitates* **3195:** 1–17.

Britz R. 2009. *Danionella priapus*, a new species of miniature cyprinid fish from West Bengal, India (Teleostei: Cypriniformes: Cyprinidae). *Zootaxa* **2277:** 53–60.

Britz R, Cambray J. 2001. Structure of egg surfaces and attachment organs in anabantoids. *Ichthyological Exploration of Freshwaters* **12:** 267–288.

Britz R, Conway KW. 2009. Osteology of *Paedocypris*, a miniature and highly developmentally truncated fish (Teleostei: Ostariophysi: Cyprinidae). *Journal of Morphology* **270:** 389–412.

Britz R, Conway KW. 2011. The cypriniformes tree of confusion. Pg. 73–78. In: M. R. de Carvalho & M. T. Craig (eds.), *Morphological and molecular approaches to the phylogeny of fishes: integration or conflict? Zootaxa* **2946:** 1–142.

Britz R, Conway KW, Rüber L. 2009. Spectacular morphological novelty in a miniature cyprinid fish, *Danionella dracula* n. sp. *Proceedings of the Royal Society London B* **276:** 2179–2186.

Britz R, Johnson GD. 2011. Comments on the establishment of the one to one relationship between characters as a prerequisite for homology assessment in phylogenetic studies. *Zootaxa* **2946:** 65–72.

Britz R, Kottelat M. 2008. *Paedocypris carbunculus*, a new species of miniature fish from Borneo (Teleostei: Cypriniformes: Cyprinidae). *Raffles Bulletin of Zoology* **56:** 415–422.

Britz R, Kottelat M, Tan HH. 2012. *Fangfangia spinocleithralis*, a new genus and species of miniature cyprinid from Kalimantan Tengah, Borneo, Indonesia (Teleostei: Cypriniformes: Cyprinidae). *Ichthyological Exploration of Freshwaters* **22:** 327–335.

Bufalino AP, Mayden RL. 2010a. Phylogenetic evaluation of North American Leuciscidae (Actinopterygii: Cypriniformes: Cyprinoidea) as inferred from analyses of mitochondrial and nuclear DNA sequences. *Systematics and Biodiversity* **8:** 493–505.

Bufalino AP, Mayden RL. 2010b. Phylogenetic relationships of North American phoxinins (Actinopterygii: Cypriniformes: Leuciscidae) as inferred from S7 nuclear DNA sequences. *Molecular Phylogenetics and Evolution* **55:** 143–152.

Bufalino AP, Mayden RL. 2010c. Molecular phylogenetics of North American phoxinins (Actinopterygii: Cypriniformes: Leuciscidae) based on RAG1 and S7 nuclear DNA sequence data. *Molecular Phylogenetics and Evolution* **55:** 274–283.

Campbell MA, Chen WJ, Lopez JA. 2013. Are flatfishes (Pleuronectiformes) monophyletic? *Molecular Phylogenetics and Evolution* **69:** 763–785. Available online 19 July 2013.

Campbell MA, Chen WJ, Lopez JA. 2014. Molecular data do not provide unambiguous support for the monophyly of flatfishes (Pleuronectiformes): a reply to Betancur-R and Ortí. *Molecular Phylogenetics and Evolution* **75:** 149–153.

Carvalho MR de, Craig MT, eds. 2011. Morphological and molecular approaches to the phylogeny of fishes: integration or conflict? *Zootaxa* **2946:** 1–142.

Cavender TM, Coburn M. 1992. Phylogenetic relationships of North American Cyprinidae. In: Mayden R, ed. *Systematics, historical ecology and North American freshwater fishes*. Stanford: Stanford University Press, 293–327.

Chen WJ, Lavoué S, Mayden RL. 2013. Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). *Evolution* **67:** 2218–2239.

Chen WJ, Mayden RL. 2009. Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the world's largest clade of freshwater fishes: further evidence from six nuclear genes. *Molecular Phylogenetics and Evolution* **52:** 544–549.

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

Chen XL, Yue PQ, Lin RD. 1984. [Major groups within the family Cyprinidae and their phylogenetic relationships]. *Acta Zootaxonomica Sinica* **9:** 424–440. [In Chinese with English summary].

Conway KW. 2005. Monophyly of the genus *Boraras* (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters* **16:** 249–264.

Conway KW. 2011. Osteology of the South Asian Genus *Psilorhynchus* McClelland, 1839 (Teleostei: Ostariophysi: Psilorhynchidae) with investigation of its phylogenetic relationships within the Order Cypriniformes. *Zoological Journal of the Linnean Society* **163:** 50–154.

Conway KW, Chen WJ, Mayden RL. 2008. The 'Celestial Pearl danio' is a miniature *Danio* (s.s.) (Ostariophysi: Cyprinidae): evidence from morphology and molecules. *Zootaxa* **1686:** 1–28.

Conway KW, Kottelat M, Tan HH. 2011. Review of the Southeast Asian miniature cyprinid genus *Sundadanio* (Ostariophysi: Cyprinidae) with descriptions of seven new species from Indonesia and Malaysia. *Ichthyological Exploration of Freshwaters* **22:** 251–288.

Cubbage CC, Mabee PM. 1996. Development of the cranium and paired fins in the zebrafish Danio rerio (Ostariophysi: Cyprinidae). *Journal of Morphology* **229:** 121–160.

Doosey MH, Bart HL. 2011. Morphological variation of the palatal organ and chewing pad of Catostomidae (Teleostei: Cypriniformes). *Journal of Morphology* **272:** 1092–1108.

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. 2011. *Geneious v5.4*, Available at http://www.geneious.com

Eastman JT. 1977. The pharyngeal bones and teeth of catostomid fishes. *American Midland Naturalist* **97:** 68–88.

Evans NM, Holder MT, Barbeitos MS, Okamura B, Cartwright P. 2010. The phylogenetic position of Myxozoa: exploring conflicting signals in phylogenomic and ribosomal data sets. *Molecular Biology and Evolution* **27:** 2733–2746.

Fang F, Norén M, Liao TY, Källersjö M, Kullander SO. 2009. Molecular phylogenetic interrelationships of the south Asian cyprinid genera *Danio*, *Devario* and *Microrasbora* (Teleostei, Cyprinidae, Danioninae). *Zoologica Scripta* **38:** 237–256.

Felsenstein J. 1978. Cases in which parsimony or compatibilty methods will be positively misleading. *Systematic Zoology* **27:** 401–410.

Fink SV, Fink WL. 1981. Interrelationships of the ostariophysan fishes (Teleostei). *Zoological Journal of the Linnean Society* **72:** 297–353.

Fink SV, Fink WL. 1996. Interrelationships of the Ostariophysi. In: Stiassny M, Parenti L, Johnson GD, eds. *Interrelationships of fishes*. San Diego: Academic Press, 209–249.

Funch P, Christensen RM. 1995. Cyclophora is a new phylum with affinities to Entoprocta and Ectoprocta. *Nature* **378:** 711–714.

Gould SJ. 1977. *Ontogeny and phylogeny*. Cambridge, MA: The Belknap Press of Harvard University Press.

Hanken J, Wake DB. 1993. Miniaturization of body size: organismal consequences and evolutionary significance. *Annual Review of Ecology and Systematics* **24:** 501–519.

Hennig W. 1966. *Phylogenetic systematics*. Urbana, USA: University of Illinois Press.

Hernandez LP, Bird NC, Staab KL. 2007. Using zebrafish to investigate cypriniform evolutionary novelties: functional development and evolutionary diversification of the kinethmoid. *Journal of Experimental Zoology (Mol Dev Evol)* **308B:** 625–641.

Howes GJ. 1981. Anatomy and phylogeny of the Chinese Major Carps *Ctenopharyngodon* Steind., 1866 and *Hypophthalmichthys* Blkr., 1860. *Bulletin of the British Museum of Natural History, Zoology* **41:** 1–52.

Howes GJ. 1991. Systematics and biogeography: an overview. In: Winfield IJ, Nelson JS, eds. *Cyprinid fishes systematics, biology and exploitation*. London: Chapman and Hall, 1–33.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–267. Software available at http://www.splitstree.org

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends in Genetics* **22:** 225–231.

Jenner RA, Littlewood DTJ. 2010. Invertebrate Problematica: kinds, causes, and solutions. In: Telford MJ, Littlewood DTJ, eds. *Animal evolution – genomes, fossils and trees*. Oxford: Oxford University Press, 107–126.

Johnson GD, Brothers EB. 1993. *Schindleria*: a paedomorphic goby (Teleostei: Gobioidei). *Bulletin of Marine Sciences* **52:** 441–471.

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* **29:** 170–179.

Kottelat M, Britz R, Tan HH, Witte KE. 2006. *Paedocypris*, a new genus of Southeast Asian cyprinid fish with a remarkable sexual dimorphism, comprises the world's smallest vertebrate. *Proceedings of the Royal Society London B* **273:** 895–899.

Liao T, Kullander SO. 2013. Phylogenetic significance of the kinethmoid-associated Y-shaped ligament and long intercostal ligaments in the Cypriniformes (Actinopterygii: Ostariophysi). *Zoologica Scripta* **42:** 71–87.

Liu S, Tan HH, Tan SL, Hong Y. 2012. Chromosome evolution and genome miniaturization in minifish. *PLoS ONE* **7:** e37305. doi:10.1371/journal.pone.0037305.

Maddison DR, Maddison WP. 2002. *MacClade 4.05 analysis of phylogeny and character evolution [computer software manual]*. Sunderland, MA: Sinauer Associates.

Maddison WP, Donoghue MJ, Maddison DR. 1984. Outgroup analysis and parsimony. *Systematic Zoology* **33:** 83–103.

Mayden RL, Chen WJ. 2010. The world's smallest vertebrate species of the genus *Paedocypris*: a new family of freshwater fishes and the sister group to the world's most diverse clade of freshwater fishes (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution* **57:** 152–175.

McMenamin SK, Parichy DM. 2012. Metamorphosis in teleosts. *Current topics in Developmental Biology* **103:** 127–165.

Mooi RD, Gill AC. 2010. Phylogenies without synapomorphies – A crisis in fish systematics: time to show some character. *Zootaxa* **2450:** 26–40.

Nelson G. 1978. Ontogeny, phylogeny, paleontology, and the biogenetic law. *Systematic Zoology* **27:** 324–345.

Nixon KC. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15:** 407–414.

Patterson C. 1982. Morphological characters and homology. In: Joysey KA, Friday AE, eds. *Problems of phylogenetic reconstruction*. London: Academic Press, 21–74.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* **9:** e1000602. doi:10.1371/journal.pbio.1000602.

Reid GM. 1982. The form, function and phylogenetic significance of the vomero-palatine organ in cyprinid fishes. *Journal of Natural History* **16:** 497–510.

Remane A. 1952. *Die Grundlagen des natürlichen Systems der vergleichenden Anatomie und der Phylogenetik*. Leipzig: Akademische Verlagsgesellschaft.

Rieppel O, Kearney M. 2002. Similarity. *Biological Journal of the Linnean Society* **75:** 59–82.

Roberts TR. 1986. *Danionella translucida*, a new genus and species of cyprinid fish from Burma, one of the smallest living vertebrates. *Environmental Biology of Fishes* **16:** 231–241.

Rüber L, Kottelat M, Tan HH, Ng PKL, Britz R. 2007. Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. *BMC Evolutionary Biology* **7:** 38–47.

Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497:** 327–331.

Sawada Y. 1982. Phylogeny and Zoogeography of the Superfamily Cobitoidea (Cyprinoidei, Cypriniformes). *Memoirs of the Faculty of Fisheries, Hokkaido University* **28:** 65–223.

Schaefer SA, Weitzman SH, Britski HA. 1989. Review of the Neotropical catfish genus *Scoloplax* (Pisces: Loricarioidei: Scoloplacidae) with comments on reductive characters in phylogenetic analysis. *Proceedings of the Academy of Natural Sciences of Philadelphia* **141:** 181–211.

Schönhuth S, Shiozawa DK, Dowling TE, Mayden RL. 2012. Molecular systematics of western North American cyprinids (Cypriniformes: Cyprinidae). *Zootaxa* **3586:** 281–303.

Shimodaira H. 2002. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and inference* **90:** 227–244.

Shimodaira H, Hasegawa M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution* **16:** 1114–1116.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17:** 1246–1247.

Sibbing FA. 1982. Pharyngeal mastication and food transport in the carp (*Cyprinus carpio* L.): a cineradiographic and electromyographic study. *Journal of Morphology* **172:** 223–258.

Siebert DJ. 1987. Interrelationships among families of the order Cypriniformes (Teleostei). Unpublished D. Phil. Thesis. City University of New York.

Siebert DJ. 1997. Notes on the anatomy and relationships of *Sundasalanx* Roberts (Teleostei, Clupeidae), with descriptions of four new species from Borneo. *Bulletin. of the Natural History Museum London (Zoology)* **63:** 13–26.

Sikes DS, Lewis PO. 2001. PAUPRat: PAUP* implementation of the parsimony ratchet. Beta software, version 1. (Distributed by the Authors, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, USA.).

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Sulaiman ZH, Mayden RL. 2012. Cypriniformes of Borneo (Actinopterygii, Otophysi): an extraordinary fauna for integrated studies on diversity, systematics, evolution, ecology, and conservation. *Zootaxa* **3586:** 359–376.

Swofford DL. 2000. *PAUP*: phylogenetic analysis using parsimony (*and other methods) [computer software manual].* Sunderland, MA: Sinauer Associates.

Tang KL, Agnew MK, Chen WJ, Hirt MV, Raley ME, Sado T, Schneider LM, Yang L, Bart HL, He S, Liu H, Miya M, Saitoh K, Simons AM, Wood RM, Mayden RL. 2011. Phylogeny of the gudgeons (Teleostei: Cyprinidae: Gobioninae). *Molecular Phylogenetics and Evolution* **61:** 103–124.

Tang KL, Agnew MK, Hirt MV, Sado T, Schneider LM, Freyhof J, Sulaiman Z, Swartz E, Vidthayanon C, Miya M, Saitoh K, Simons AM, Wood RM, Mayden RL. 2010. Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). *Molecular Phylogenetics and Evolution* **57:** 189–214.

Tang KL, Lumbantobing DN, Mayden RL. 2013. The phylogenetic placement of Oxygaster van Hasselt, 1823 (Teleostei: Cypriniformes: Cyprinidae) and the taxonomic status of the family-group name Oxygastrinae Bleeker, 1860. *Copeia* **2013:** 13–22.

Taylor WR, Van Dyke GC. 1985. Revised procedures for staining and clearing small fishes and other vertebrates for bone and cartilage study. *Cybium* **9:** 107–119.

Wägele JW, Letsch H, Klussmann-Kolb A, Mayer C, Misof B, Wägele H. 2009. Phylogenetic support values are not necessarily informative: the case of the Serialia hypothesis (a mollusk phylogeny). *Frontiers in Zoology* **6:** 12.

Wägele JW, Mayer C. 2007. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology* **7:** 147.

Wägele JW, Rödding F. 1998. A Priori Estimation of Phylogenetic Information Conserved in Aligned Sequences. *Molecular Phylogenetics and Evolution* **9:** 358–365.

Wang XZ, Gan XN, Li JB, Mayden RL, He SP. 2012. Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. *Science China Life Sciences* **55:** 761–773.

Weitzman SH. 1974. Osteology and evolutionary relationships of the Sternoptychidae, with a new classification of stomiatoid families. *Bulletin of the American Museum of Natural History* **153:** 327–478.

Weitzman SH, Fink SV. 1985. Xenurobryconin phylogeny and putative pheromone pumps in glandulocaudine fishes (Teleostei: Characidae). *Smithsonian Contributions to Zoology* **421:** 1–121.

Weitzman SH, Vari RP. 1988. Miniaturization in South American freshwater fishes; an overview and discussion. *Proceedings of the Biological Society of Washington* **101:** 444–465.

Winterbottom R. 1991. The *Trimmatom nanus* species complex (Actinopterygii, Gobiidae): phylogeny and progenetic heterochrony. *Systematic Zoology* **39:** 253–265.

APPENDIX 1

List of characters and character states contained in morphological character matrices 1–5 (Appendices 2–6). See text for source of characters. Characters in our matrix 1 that are equivalent to those used by Mayden & Chen (2010) include reference to their character number and state (e.g. M&C 2A refers to character 2, state A, in Table 2 of Mayden & Chen, 2010).

MATRIX-1 CHARACTERS

1. Weberian apparatus: absent (0); present (1).
2. Kinethmoid element: absent (0); cartilage (1; M&C 2A); bone (2; M&C 2B).
3. Fifth ceratobranchial: similar to other ceratobranchial elements (0); enlarged, extending much further dorsally than other ceratobranchial elements (1; M&C 3A).
4. Teeth on ceratobranchial 5: not ankylosed to bone (0); ankylosed to bone (1; M&C 4A).
5. Lateral process of the second vertebral centrum: short, not extending far into body musculature (0); elongate, extending far into body musculature (1; M&C 5A).
6. Preethmoid: absent (0); present (1; M&C 6A).
7. Autopalatine process that extends to abut mesethmoid: absent (0); present and poorly developed (1; M&C 7A); present and well developed (2; M&C 7B).
8. Autopalatine/endopterygoid articulation: absent (0); present (1; M&C 8A).
9. Ectopterygoid: present (0); absent (1; M&C 9A). Note: Mayden & Chen changed the focus of this character from the interaction between two ossifications (the ectopterygoid and the autopalatine) to the absence of a single bone (the ectopterygoid). In doing so, they have created duplication between this character and their character 47.
10. Premaxilla: ascending process absent, extending furthest dorsally lateral to midline (0); ascending process poorly developed, extending furthest dorsally along midline (1; M&C 10A); ascending process well developed, extending furthest dorsally along midline (2; M&C 10B).
11. Jaw teeth: present (0); absent (1; M&C 11A).
12. Teeth on second and third pharyngobranchials and basihyal: present (0); absent (1; M&C 12A).
13. Pharyngobranchial toothplates: present (0); absent (1; M&C 13A).
14. Basibranchial 1–3 toothplate: present (0); absent (1; M&C 14A).
15. Epurals: two or more (0); fewer than two (1; M&C 15A).
16. Adipose fin: present (0); absent (1; M&C 16A).

17. Postcleithra: present (0); absent (1; M&C 17B). Note: Mayden & Chen change the focus of this character from the number of postcleithra (*sensu* Fink & Fink, 1981) to the presence/absence of postcleithra. In doing so, they have created duplication between this character and characters 31 and 48. 17A is not mentioned by Mayden & Chen.
18. Maxillary barbels: absent (0); present (1; M&C 18A).
19. Masticatory plate on basioccipital: absent (0; M&C 19D); present (1; M&C 19A); further modified as palatal organ (2; M&C 19B). Note: Mayden & Chen's character 19C is mapped illogically on their tree because all taxa that have 19B and 19D also exhibit 19C. 19C is excluded here. Information provided for this character in Mayden & Chen's table 2 and figure 3 conflict, and we have followed the coding presented in their figure for their character-19 states A, B, and D.
20. Pharyngeal process of the basioccipital: absent (0); present (1; M&C 20A).
21. Uncinate processes on EB1 and EB2: present (0); absent (1; M&C 21A).
22. Pharyngobranchial 1: present (0); absent (1; M&C 22A). Note: information provided for this character in Mayden & Chen's table 2 conflicts with that provided in the text (p. 166).
23. Interorbital septum formed by orbitosphenoid and parasphenoid: absent (0); present (1; M&C 23A).
24. Infraorbital 5: present (0); absent (1; M&C 24A).
25. Opercular canal: absent (0); present (1; M&C 25A).
26. Deep, well-developed subtemporal fossae: absent (0); present (1; M&C 26A).
27. Anterior opening of the trigemino-facial chamber: absent (0); positioned between prootic and pterosphenoid, or completely within the prootic (1; M&C 27A).
28. Second and third centra: separate (0); fused (1; M&C 28A).
29. Overlap of pharyngobranchial 2 by pharyngobranchial 3: absent (0); present (1; M&C 29A).
30. Pharyngobranchial 2 and pharyngobranchial 3: not at the same level and not confluent with epibranchial 4 cartilage (0); at the same level and confluent with epibranchial 4 cartilage (1). Note: information provided for this character in Mayden & Chen's table 2 and figure 3 conflict, and we have followed the coding presented in their figure for their character. This character duplicates character 37 below.
31. Postcleithra: present (0); absent (1; M&C 31A). Note: this character duplicates characters 17 and 48.
32. Three uppermost pectoral-fin rays and pectoral radial 1: not exhibiting sexual dimorphism (0); hypertrophied in males (1; M&C 32A). Note: this character duplicates characters 51 and 53.
33. Basipterygia of pelvic girdle: not exhibiting sexual dimorphism (0); hypertrophied in males (1; M&C 33A). Note: this character duplicates character 52.
34. Hemitrichs of first pelvic-fin ray: not exhibiting sexual dimorphism (0); hypertrophied in males (1; M&C 34A). Note: this character duplicates character 53.
35. Hemal spines: present only in caudal region (0); present in abdominal region (1; M&C 35A). Note: this character duplicates character 55.
36. Epibranchial 5: not articulating with tip of ceratobranchial 5 (0); articulating with tip of ceratobranchial 5 (1; M&C 36B). Note: 36A is not mentioned by Mayden & Chen.
37. Pharyngobranchial 3 and epibranchial 4: not connected via a continuous cartilage (0); connected via a continuous cartilage (1; M&C 37B). Note: this character duplicates character 30. 37A is not mentioned by Mayden & Chen.
38. Neural spine on fourth neural arch: present (0); absent (1; M&C 38B). Note: 38A is not mentioned by Mayden & Chen.
39. Parietal: present (0); absent (1; M&C 39B). Note: 39A is not mentioned by Mayden & Chen.
40. Hypobranchial 3: present (0); absent (1; M&C 40B). Note: 40A is not mentioned by Mayden & Chen.
41. Supraneurals posterior to supraneural 3: present (0); absent (1; M&C 41B). Note: 41A is not mentioned by Mayden & Chen.
42. Processus ascendens of the scaphium: present (0); absent (1; M&C 42A); absent (2; M&C 42B). Note: Mayden & Chen list only character 42B in their table 2, but map both character 42A (for *Sundadanio*) and 42B (for *Paedocypris* and *Danionella*) in their figure 3. We have followed the coding presented in their figure for their character.
43. Inner arms of ossa suspensoria: separate (0); fused proximally (1; M&C 43A); fused proximally (2; M&C 43B). Note: Mayden & Chen list only character 43B in their table 2, but map both character 43A (for *Sundadanio*) and 43B (for *Paedocypris* and *Danionella*) in their figure 3. We have followed the coding presented in their figure for their character.
44. Gap between neural arches 3 and 4: with little cartilage (0); with extensive cartilage (1; M&C 44A); with extensive cartilage (2; M&C 44B). Note: Mayden & Chen list only character 44B in their table 2, but map both character 44A (for *Sundadanio*) and 44B (for *Paedocypris* and *Danionella*) in their figure 3. We have followed the coding presented in their figure for their character.

45. Tripus: small (0); hypertrophied and unusually elongate (1; M&C 45A); hypertrophied and unusually elongate (1; M&C 45B). Note: Mayden & Chen list only character 45B in their table 2, but map both character 45A (for *Sundadanio*) and 45B (for *Paedocypris* and *Danionella*) in their figure 3. We have followed the coding presented in their figure for their character.
46. Vomer: present (0); absent (1; M&C 46B). Note: 46A is not mentioned by Mayden & Chen.
47. Ectopterygoid: present (0); absent (1; M&C 47B). Note: 47A is not mentioned by Mayden & Chen.
48. Postcleithra: present (0); absent (1; M&C 48B). Note: 48A is not mentioned by Mayden & Chen.
49. Posttemporal: present (0); absent (1; M&C 49B or 49b). Note: Mayden & Chen list only character 49B in their table, but map both 49b (for *Sundadanio*) and 49B (for *Paedocypris* and *Danionella*) in their figure 2. 49A is not mentioned by Mayden & Chen.
50. Intercalarium: with contact to associated centrum (0); reduced, represented by a small interossicular ligament, without contact to associated centrum (1; M&C 50A).
51. Three uppermost pectoral radials: not exhibiting sexual dimorphism (0); hypertrophied in males (1; M&C 51A). Note: this character duplicates character 32.
52. Basipterygia of pelvic girdle: not exhibiting sexual dimorphism (0); hypertrophied in males (1; M&C 52A). Note: this character duplicates character 33.
53. First pectoral-fin ray: not exhibiting sexual dimorphism (0); hypertrophied in males (1; M&C 53A). Note: this character duplicates character 32.
54. Large dorsally directed triangular process on the lateral face of the outer arm of the os suspensorium: absent (0); present (1; M&C 54A).
55. Hemal spines: present only in caudal region (0); present in abdominal region (1; M&C 55A). Note: this character duplicates character 35.
56. Soft-tissue structures as a prepelvic keratinized pad: absent (0); present (1; M&C 56A).
57. Keratinized flanges of skin covering enlarged first pelvic ray: absent (0); present (1; M&C 57A).
58. Enlarged genital opening forming a bag around the first anal-fin rays: absent (0); present (1; M&C 58A).
59. Males (in reference to *Danionella*) with additional flanges to *Paedocypris* on the os suspensorium confluent with the later process on the second vertebra: absent (0); present (1; M&C 59A).
60. Males (in reference to *Danionella*) with cartilaginous nodule associated with the anterior face of rib 5 and anterior swim bladder chamber: absent (0); present (1; M&C 60A).
61. Males (in reference to *Danionella*) with anterior shift of genital opening and anus: absent (0); present (1; M&C 61A).
62. Maxillo–mandibular cartilage: absent (0); present (1; M&C 62A).
63. Scapulocoracoid cartilage: with two ossifications (0); with single ossification in males (1; M&C 63A).
64. Fifth pectoral-fin ray: not exhibiting sexual dimorphism (0); exhibiting a sexually dimorphic flange along dorsal edge (1; M&C 64A).
65. Cleithrum: without large sexually dimorphic flange (0); with large sexually dimorphic flange (1; M&C 65A).
66. Outer arm of the os suspensorium: not exhibiting sexual dimorphism (0); hypertrophied in males (1; M&C 66A).
67. Fifth rib: not exhibiting sexual dimorphism (0); with large flange on inner face (1; M&C 67A).
68. Enlarged, likely drumming muscle present between the outer arm of the os suspensorium and fifth rib: absent (0); present in males (1; M&C 68A).

MATRIX-2 CHARACTERS

1. Weberian apparatus: absent (0); present (1).
2. Kinethmoid element: absent (0); cartilage (1); bone (2).
3. Fifth ceratobranchial: similar to other ceratobranchial elements (0); enlarged, extending much further dorsally than other ceratobranchial elements (1).
4. Teeth on ceratobranchial 5: not ankylosed to bone (0); ankylosed to bone (1).
5. Lateral process of the second vertebral centrum: short, not extending far into body musculature (0); elongate, extending far into body musculature (1).
6. Preethmoid: absent (0); present (1).
7. Autopalatine process that extends to abut mesethmoid: absent (0); present (1).
8. Autopalatine/endopterygoid articulation: absent (0); present (1).
9. Ectopterygoid–autopalatine overlap: present (0); absent (1).
10. Premaxilla: extending furthest dorsally lateral to midline (0); extending furthest dorsally along midline (1).
11. Jaw teeth: present (0); absent (1).
12. Teeth on second and third pharyngobranchials and basihyal: present (0); absent (1).
13. Pharyngobranchial toothplates: present (0); absent (1).
14. Basibranchial 1–3 toothplate: present (0); absent (1).
15. Epurals: two or more (0); fewer than two (1).
16. Adipose fin: present (0); absent (1).
17. Postcleithra: present (0); absent (1).

18. Maxillary barbels: absent (0); present (1).
19. Masticatory plate on basioccipital: absent (0); present (1).
20. Pharyngeal process of the basioccipital: absent (0); present (1).
21. Uncinate processes on EB1 and EB2: present (0); absent (1).
22. Pharyngobranchial 1: present (0); absent (1).
23. Interorbital septum formed by orbitosphenoid and parasphenoid: absent (0); present (1).
24. Infraorbital 5: present (0); absent (1).
25. Opercular canal: absent (0); present (1).
26. Deep, well-developed subtemporal fossae: absent (0); present (1).
27. Anterior opening of the trigemino-facial chamber: contained within prootic (0); positioned between prootic and pterosphenoid (1).
28. Second and third centra: separate (0); fused (1).
29. Overlap of pharyngobranchial 2 by pharyngobranchial 3: absent (0); present (1).
30. Pharyngobranchial 2 and pharyngobranchial 3: not at the same level and not confluent with epibranchial-4 cartilage (0); at the same level and confluent with epibranchial-4 cartilage (1).
31. Three uppermost pectoral-fin rays and pectoral radial 1: not exhibiting sexual dimorphism (0); hypertrophied in males (1).
32. Basipterygia of pelvic girdle: not exhibiting sexual dimorphism (0); hypertrophied in males (1).
33. Hemitrichs of first pelvic-fin ray: not exhibiting sexual dimorphism (0); hypertrophied in males (1).
34. Hemal spines: present only in caudal region (0); present in abdominal region (1).
35. Epibranchial 5: not articulating with tip of ceratobranchial 5 (0); articulating with tip of ceratobranchial 5 (1).
36. Neural spine on fourth neural arch: present (0); absent (1).
37. Parietal: present (0); absent (1).
38. Hypobranchial 3: present (0); absent (1).
39. Supraneurals posterior to supraneural 3: present (0); absent (1).
40. Processus ascendens of the scaphium: present (0); absent (1).
41. Inner arms of ossa suspensoria: separate (0); fused proximally (1).
42. Gap between neural arches 3 and 4: with little cartilage (0); with extensive cartilage (1).
43. Tripus: small (0); hypertrophied and unusually elongate (1).
44. Vomer: present (0); absent (1).
45. Ectopterygoid: present (0); absent (1).
46. Posttemporal: present (0); absent (1).
47. Intercalarium: with contact to associated centrum (0); reduced, represented by a small interossicular ligament, without contact to associated centrum (1).
48. Large dorsally directed triangular process on the lateral face of the outer arm of the os suspensorium: absent (0); present (1).
49. Soft-tissue structures as a prepelvic keratinized pad: absent (0); present (1).
50. Keratinized flanges of skin covering enlarged first pelvic ray: absent (0); present (1).
51. Enlarged genital opening forming a bag around the first anal-fin rays: absent (0); present (1).
52. Males (in reference to *Danionella*) with additional flanges to *Paedocypris* on the os suspensorium confluent with the later process on the second vertebra: absent (0); present (1).
53. Males (in reference to *Danionella*) with cartilaginous nodule associated with the anterior face of rib 5 and anterior swim bladder chamber: absent (0); present (1).
54. Males (in reference to *Danionella*) with anterior shift of genital opening and anus: absent (0); present (1).
55. Maxillo–mandibular cartilage: absent (0); present (1).
56. Scapulocoracoid cartilage: with two ossifications (0); with single ossification in males (1).
57. Fifth pectoral-fin ray: not exhibiting sexual dimorphism (0); exhibiting a sexually dimorphic flange along dorsal edge (1).
58. Cleithrum: without large sexually dimorphic flange (0); with large sexually dimorphic flange (1).
59. Outer arm of the os suspensorium: not exhibiting sexual dimorphism (0); hypertrophied in males (1).
60. Fifth rib: not exhibiting sexual dimorphism (0); with large flange on inner face (1).
61. Enlarged, likely drumming muscle present between the outer arm of the os suspensorium and fifth rib: absent (0); present in males (1).

MATRIX-3 CHARACTERS

1. Kinethmoid: absent (0); present (1).
2. Preethmoid: absent (0); present (1).
3. Second preethmoid element: absent (0); second preethmoid cartilage present (1); second preethmoid present (2).
4. Articulation between ethmoid complex and frontals: absent, posterodorsal edge of ethmoid complex or supraethmoid firmly sutured or tightly abutting anterior edge of frontals (0); present, posterodorsal edge of ethmoid complex articulating in a shallow facet along anterior edge of frontals (1).
5. Relationship between ethmoid complex and vomer: separate (0); fused (1).

6. Contact between orbitosphenoid and ethmoid complex: absent (0); present (1).
7. Anterolateral process of the lateral ethmoid: absent (0); present (1).
8. Erectile suborbital spine, formed by enlargement of lateral ethmoid: absent (0); present (1).
9. Interorbital septum: interorbital septum formed only by the orbitosphenoid (0); interorbital septum formed by the orbitosphenoid and a dorsal component of the parasphenoid (1).
10. Ventral keel on parasphenoid: absent (0); present (1).
11. Contact between orbitosphenoid and pterosphenoid: present (0); absent (1).
12. Basisphenoid: present (0); absent (1).
13. Anterior opening of trigemino-facial chamber: anterior opening of trigemino-facial chamber contained within prootic (0); anterior opening of trigemino-facial chamber on anterior edge of prootic, its anterior edge bordered by the posterior edge of the pterosphenoid or lateral wing of the parasphenoid (1).
14. Postepiphysial fontanelle: absent (0); present (1).
15. Connection between supraorbital sensory canal and otic sensory canal: present, supraorbital sensory canal connected to otic sensory canal (0); absent, supraorbital sensory canal disjunct from otic sensory canal (1).
16. Fenestration of the dilatator fossa: absent (0); present (1).
17. Pterotic fossa: absent (0); present (1).
18. Supratemporal commissure: located on parietal or extrascapular (0); located on supraoccipital (1).
19. Exoccipital: contacting antimere along midline (0); separate from antimere (1).
20. Intercalar: present (0); absent (1).
21. Subtemporal fossa: absent or represented only by a shallow concavity on the ventral surface of the neurocranium (0); present, represented by a deep concavity on the ventral surface of the neurocranium (1).
22. Basioccipital process: absent (0); present as long parallel processes, not united to form a canal around the dorsal aorta (1); present as a large bony structure, forming a canal around the dorsal aorta (2).
23. Fenestration of the basioccipital process: absent (0); present (1).
24. Pharyngeal process of the basioccipital process: absent (0); present, horizontal in cross section (1); present, terete in cross section (2).
25. Masticatory plate: absent (0); present (1).
26. Supraorbital sensory canal: present (0); absent (1).
27. Anterolateral process of the frontal: absent (0); present (1).
28. Lateral occipital foramen: absent (0); present (1).
29. Antorbital: present (0); absent (1).
30. Relationship between anteriormost portion of infraorbital sensory canal and infraorbital 1: anteriormost portion of infraorbital sensory canal completely enclosed on infraorbital 1 (0); anteriormost portion of infraorbital sensory canal completely or partially disjunct from infraorbital 1 (1).
31. Nature of infraorbital series posterior to infraorbital 1: infraorbital series posterior to infraorbital 1 represented by four, roughly plate-like bones (0); infraorbital series posterior to infraorbital 1 represented by a variable number of small tubular ossifications around the infraorbital canal (0).
32. Spatial arrangement of infraorbital 5 and supraorbital: in contact or separated only by a short distance (0); widely separate (1).
33. Dermopalatine: present (0); absent (1).
34. Dorsomedial autopalatine process: absent (0); present (1).
35. Autopalatine–endopterygoid articulation: absent (0); present (1).
36. Ectopterygoid–autopalatine overlap: present (0); absent (1).
37. Length of autopalatine: short, terminating level with or slightly anterior to area of articulation with ethmoid region (0); autopalatine greatly extended past area of articulation with ethmoid region, terminating dorsal to maxilla (1).
38. Preautopalatine element: absent (0); present and cartilaginous (1); present and ossified (2).
39. Mesiopreautopalatine: absent (0); present, cartilaginous (1); present, ossified (2).
40. Endopterygoid: expansive and sheet-like (0); reduced and rod-like posteriorly (1).
41. Metapterygoid–quadrate fenestra: absent (0); present (1).
42. Metapterygoid: simple posterior lamina near hyomandibular (0); with membrane bone flanges forming suture to anterior edge of the hyomandibular (1).
43. Symplectic: simple dorsal lamina near metapterygoid (0); with membrane bone flanges on dorsal edge forming suture to ventral edge of metapterygoid (1).
44. Lateral face of hyomandibular: without membrane bone flange (0); with large membrane bone flange overlapping preopercle (1).
45. Ventral margin of the opercle: straight or convex (0); concave (1).
46. Opercular sensory canal: absent (0); present (1).
47. Connection between preopercular sensory canal and infraorbital sensory canal: absent (0); present (1).

48. Interopercle: anteriormost point directed anteriorly (0); anteriormost point directed anteromedially towards ventral midline (1).
49. Interopercular–preopercular articulation: absent (0); present (1).
50. Anterodorsal head of opercle: absent, or if present short (0); greatly elongate (1).
51. Premaxilla: extending furthest dorsally at a point lateral to the dorsal midline (0); extending furthest dorsally adjacent to the dorsal midline (1).
52. Oral teeth: present (0); absent (1).
53. Coronomeckelian: present (0); absent (1).
54. Anguloarticular: similar in width to the posteriormost point of the dentary in ventral view, not extended laterally as a sharp point (0); much wider than the posteriormost point of the dentary in ventral view, extending laterally as a sharp point (1).
55. Maxilla: without large spine-like process overlapping dorsal surface of premaxillae (0); with large spine-like process overlapping dorsal surface of premaxillae (1).
56. Mandibular sensory canal: present (0); absent (1).
57. Basihyal toothplate: present (0); absent (1).
58. Pharyngobranchial tooth plates: present (0); absent (1).
59. Anterior basibranchial copula toothplate: present (0); absent (0).
60. Posterior basibranchial copula: with dermal toothplates associated with dorsal surface (0); with dermal (toothless) plates associated with dorsal surface (1); without dermal plates or toothplates (2).
61. Gill-raker teeth: present (0); absent (1).
62. Gill rakers: small-based elements, base narrower than total height (0); wide-based elements, base wider than total height (1).
63. Gill rakers: without comb-like projections along dorsal edge (0); with comb-like projections along dorsal edge (1).
64. Gill-filament ossifications: absent (0); present (1).
65. Basihyal: anterior edge straight or slightly rounded, with a singular cartilaginous tip (0); anterior edge bifurcated, with two prominent cartilaginous tips (1).
66. Sublingual ossifications: absent (0); present (1).
67. Posterior basibranchial copulae: complete (0); segmented (1).
68. Ventral keel on posterior basibranchial copula or derivative elements: present (0); absent (1).
69. Basibranchial 1: present (0); absent (1).
70. Basibranchial 2: anterior edge similar in width or only slightly wider than posterior edge (0); anterior edge markedly wider than posterior edge (1).
71. Posterior tip of basibranchial 3: on same plane as anterior tip, not intimately associated with the tip of hypobranchial 3 (0); lower than anterior tip, articulating in a concave facet on the anterior tip of hypobranchial 3 (1).
72. Basibranchial 4: absent (0); present (1).
73. Hypobranchial 2: present (0); absent, only hypobranchial-3 cartilage present (1).
74. Hypobranchial 3: present (0); absent, only hypobranchial-3 cartilage present (1).
75. Ceratobranchial 5: extending no further dorsally than other ceratobranchials (0); extending much further dorsally than other ceratobranchials (1).
76. Transversus ventralis V process on ceratobranchial 5: absent (0); present (1).
77. Ceratobranchial-5 teeth: teeth bound to ceratobranchial 5 by collagenous fibers (0); teeth ankylosed to ceratobranchial 5 (1).
78. Epibranchial 1: not dorsally overlapped by epibranchial 2 (0); dorsally overlapped by epibranchial 2 (1).
79. Anterior head of epibranchial 1: not twisted ventrad (0); twisted ventrad (1).
80. Anterior edge of epibranchial 1: without large membranous flange on anterior edge (0); with large membranous flange on anteroventral edge (1).
81. Pharyngobranchial 1 element: present and ossified (0); present but cartilaginous (1); absent (2).
82. Pharyngobranchial 3: not overlapping pharyngobranchial 2 (0); overlapping pharyngobranchial 2 (1).
83. Foramen between pharyngobranchials 2 and 3: absent (0); present (1).
84. Pharyngobranchial uncinate processes: present (0); absent (1).
85. Number of branchiostegal rays: more than three (0); three (1).
86. Interhyal element: present and ossified (0); present and cartilaginous (1); interhyal element absent (2).
87. Lateral face of the posterior ceratohyal: flat (0); with knob-like process posteriorly (1).
88. Lateral process of the second vertebral centrum: short (0); elongate, projecting well into somatic musculature (1).
89. Shape of the tripus: triangular (0); Y-shaped (1).
90. Neural spine of fourth vertebral centrum: similar in size or smaller than more posterior spines (0); larger than more posterior spines (1); absent (2).
91. Connection between lateral process of the second vertebral centrum and the outer arm of the os suspensorium: no contact between lateral process of the second vertebral centrum and the outer arm of the os suspensorium or contact without fusion of the two structures (0); ventral arm of the lateral process of the second

vertebral centrum fused with the outer arm of the os suspensorium (1); both dorsal and ventral arms of the lateral process of the second vertebral centrum fused to the outer arm of the os suspensorium (2).
92. Supraneural 2: present (0); absent (0).
93. Supraneural 3: with simple median crest, without division (0); median crest with modified dorsal surface (1).
94. Supraneural 3/supraneural 5 contact: absent (0); present (1).
95. Claustrum: posterodorsal edge not sutured to anterior edge of neural complex (0); posterodorsal edge sutured to anterior edge of neural complex (1).
96. Dorsal prezygapophysis of fifth vertebral centrum: dorsal prezygapophysis of fifth vertebral centrum widely separate from neural arch of fourth vertebral centrum (0); dorsal prezygapophysis of fifth vertebral centrum abutting neural arch of fourth vertebral centrum (1).
97. Parapophyses: autogenous (0); fused to centra (1).
98. Ventral prezygapophyses: ventral prezygapophyses separate from ventral postzygapophyses on caudal vertebrae (0); ventral prezygapophyses strongly interdigitating with ventral postzygapophyses on caudal vertebrae (1).
99. Contact between fifth parapophysis and outer arm of the os suspensorium: absent (0); present (1).
100. Lateral process of the first vertebral centrum: absent or, when present, short (0); elongate, projecting well into somatic musculature (1).
101. Intercalarium: base with contact to vertebral column (1); base without contact to vertebral column (1).
102. Os suspensorium: not surrounding anterior swim-bladder chamber (0); completely surrounding anterior swim-bladder chamber (1).
103. Bony contact between anteriormost point of coracoid and cleithrum: present (0); absent (1).
104. Posttemporal: present (0); absent (1).
105. Postcleithrum: more than one (0); one (1); absent (2).
106. Cleithrum: well separated from vertebral elements (0); closely associated with lateral process of second vertebral centrum (1).
107. Number of pectoral radials: four (0); three (1).
108. Baudelot's ligament: ligamentous (0); with ligament ossification (1).
109. Cleithral–occipital ligament: absent (0); present (1).
110. Rib associated with pelvic splint: similar in shape to other ribs (0); greatly thickened and robust (1).
111. Third unbranched dorsal-fin ray: segments of each hemitrich free (0); proximal segments of each hemitrich fused, forming a 'spine'-like ray (1).
112. Terminal compound centrum: preural centrum 1, ural centrum 1, and ural centrum 2 separate (0); preural centrum 1, ural centrum 1, and ural centrum 2 fused, forming a compound centrum (1).
113. Number of epurals: two or more (0); one (1).
114. Proximal tip of hypural 1: separate from surrounding hypural and parhypural elements, and autogenous with ural centrum 1 or compound centrum (0); separate from surrounding hypural and parhypural elements, and separated from compound centrum by a short hiatus (1); fused to proximal tip of parhypural (2); fused to proximal tip of parhypural + hypural 2 + compound centrum (3).
115. Hypural 6: present (0); absent (1).
116. Hypural 3: proximal tip not fused with associated centrum (0); proximal tip fused with compound centrum (1).
117. Lateral line-bearing scale size: moderate to large, height of lateral line-bearing scale much greater than height of lateral-line canal ossification (0); small, height of lateral line-bearing scale equal to or shorter than height of lateral-line canal ossification (1).
118. Anterior radii: present (0); absent (1).
119. Lateral-line canal ossifications: similar in size and shape along entire length of canal (0); third and fourth lateral-line canal ossifications much larger than more anterior and posterior ossifications (1).
120. Anterior swim-bladder chamber: single (0); divided (1).

MATRIX-4 CHARACTERS

1. Kinethmoid: absent (0); present (1).
2. Preethmoid: absent (0); present (1).
3. Second preethmoid element: absent (0); second preethmoid cartilage present (1); second preethmoid present (2).
4. Articulation between ethmoid complex and frontals: absent, posterodorsal edge of ethmoid complex or supraethmoid firmly sutured or tightly abutting anterior edge of frontals (0); present, posterodorsal edge of ethmoid complex articulating in a shallow facet along anterior edge of frontals (1).
5. Relationship between ethmoid complex and vomer: separate (0); fused (1).
6. Contact between orbitosphenoid and ethmoid complex: absent (0); present (1).
7. Anterolateral process of the lateral ethmoid: absent (0); present (1).
8. Erectile suborbital spine, formed by enlargement of lateral ethmoid: absent (0); present (1).
9. Interorbital septum: interorbital septum formed only by the orbitosphenoid (0); interorbital septum

formed by the orbitosphenoid and a dorsal component of the parasphenoid (1).
10. Ventral keel on parasphenoid: absent (0); present (1).
11. Contact between orbitosphenoid and pterosphenoid: present (0); absent (1).
12. Basisphenoid: present (0); absent (1).
13. Anterior opening of trigemino-facial chamber: anterior opening of trigemino-facial chamber contained within prootic (0); anterior opening of trigemino-facial chamber on anterior edge of prootic, its anterior edge bordered by the posterior edge of the pterosphenoid or lateral wing of the parasphenoid (1).
14. Postepiphysial fontanelle: absent (0); present (1).
15. Connection between supraorbital sensory canal and otic sensory canal: present, supraorbital sensory canal connected to otic sensory canal (0); absent, supraorbital sensory canal disjunct from otic sensory canal (1).
16. Fenestration of the dilatator fossa: absent (0); present (1).
17. Pterotic fossa: absent (0); present (1).
18. Supratemporal commissure: located on parietal or extrascapular (0); located on supraoccipital (1).
19. Exoccipital: contacting antimere along midline (0); separate from antimere (1).
20. Intercalar: present (0); absent (1).
21. Subtemporal fossa: absent or represented only by a shallow concavity on the ventral surface of the neurocranium (0); present, represented by a deep concavity on the ventral surface of the neurocranium (1).
22. Basioccipital process: absent (0); present as long parallel processes, not united to form a canal around the dorsal aorta (1); present as a large bony structure, forming a canal around the dorsal aorta (2).
23. Fenestration of the basioccipital process: absent (0); present (1).
24. Pharyngeal process of the basioccipital process: absent (0); present, horizontal in cross section (1); present, terete in cross section (2).
25. Masticatory plate: absent (0); present (1).
26. Supraorbital sensory canal: present (0); absent (1).
27. Anterolateral process of the frontal: absent (0); present (1).
28. Lateral occipital foramen: absent (0); present (1).
29. Antorbital: present (0); absent (1).
30. Relationship between anteriormost portion of infraorbital sensory canal and infraorbital 1: anteriormost portion of infraorbital sensory canal completely enclosed on infraorbital 1 (0); anteriormost portion of infraorbital sensory canal completely or partially disjunct from infraorbital 1 (1).
31. Nature of infraorbital series posterior to infraorbital 1: infraorbital series posterior to infraorbital 1 represented by four, roughly plate-like bones (0); infraorbital series posterior to infraorbital 1 represented by a variable number of small tubular ossifications around the infraorbital canal (0).
32. Spatial arrangement of infraorbital 5 and supraorbital: in contact or separated only by a short distance (0); widely separate (1).
33. Dermopalatine: present (0); absent (1).
34. Dorsomedial autopalatine process: absent (0); present (1).
35. Autopalatine–endopterygoid articulation: absent (0); present (1).
36. Ectopterygoid–autopalatine overlap: present (0); absent (1).
37. Length of autopalatine: short, terminating level with or slightly anterior to area of articulation with ethmoid region (0); autopalatine greatly extended past area of articulation with ethmoid region, terminating dorsal to maxilla (1).
38. Preautopalatine element: absent (0); present and cartilaginous (1); present and ossified (2).
39. Mesiopreautopalatine: absent (0); present, cartilaginous (1); present, ossified (2).
40. Endopterygoid: expansive and sheet-like (0); reduced and rod-like posteriorly (1).
41. Metapterygoid–quadrate fenestra: absent (0); present (1).
42. Metapterygoid: simple posterior lamina near hyomandibular (0); with membrane bone flanges forming suture to anterior edge of the hyomandibular (1).
43. Symplectic: simple dorsal lamina near metapterygoid (0); with membrane bone flanges on dorsal edge forming suture to ventral edge of metapterygoid (1).
44. Lateral face of hyomandibular: without membrane bone flange (0); with large membrane bone flange overlapping preopercle (1).
45. Ventral margin of the opercle: straight or convex (0); concave (1).
46. Opercular sensory canal: absent (0); present (1).
47. Connection between preopercular sensory canal and infraorbital sensory canal: absent (0); present (1).
48. Interopercle: anteriormost point directed anteriorly (0); anteriormost point directed anteromedially towards ventral midline (1).
49. Interopercular–preopercular articulation: absent (0); present (1).
50. Anterodorsal head of opercle: absent, or if present short (0); greatly elongate (1).

51. Premaxilla: extending furthest dorsally at a point lateral to the dorsal midline (0); extending furthest dorsally adjacent to the dorsal midline (1).
52. Oral teeth: present (0); absent (1).
53. Coronomeckelian: present (0); absent (1).
54. Anguloarticular: similar in width to the posteriormost point of the dentary in ventral view, not extended laterally as a sharp point (0); much wider than the posteriormost point of the dentary in ventral view, extending laterally as a sharp point (1).
55. Maxilla: without large spine-like process overlapping dorsal surface of premaxillae (0); with large spine-like process overlapping dorsal surface of premaxillae (1).
56. Mandibular sensory canal: present (0); absent (1).
57. Basihyal toothplate: present (0); absent (1).
58. Pharyngobranchial tooth plates: present (0); absent (1).
59. Anterior basibranchial copula toothplate: present (0); absent (0).
60. Posterior basibranchial copula: with dermal toothplates associated with dorsal surface (0); with dermal (toothless) plates associated with dorsal surface (1); without dermal plates or toothplates (2).
61. Gill-raker teeth: present (0); absent (1).
62. Gill rakers: small-based elements, base narrower than total height (0); wide-based elements, base wider than total height (1).
63. Gill rakers: without comb-like projections along dorsal edge (0); with comb-like projections along dorsal edge (1).
64. Gill-filament ossifications: absent (0); present (1).
65. Basihyal: anterior edge straight or slightly rounded, with a singular cartilaginous tip (0); anterior edge bifurcated, with two prominent cartilaginous tips (1).
66. Sublingual ossifications: absent (0); present (1).
67. Posterior basibranchial copulae: complete (0); segmented (1).
68. Ventral keel on posterior basibranchial copula or derivative elements: present (0); absent (1).
69. Basibranchial 1: present (0); absent (1).
70. Basibranchial 2: anterior edge similar in width or only slightly wider than posterior edge (0); anterior edge markedly wider than posterior edge (1).
71. Posterior tip of basibranchial 3: on same plane as anterior tip, not intimately associated with the tip of hypobranchial 3 (0); lower than anterior tip, articulating in a concave facet on the anterior tip of hypobranchial 3 (1).
72. Basibranchial 4: absent (0); present (1).
73. Hypobranchial 2: present (0); absent, only hypobranchial-3 cartilage present (1).
74. Hypobranchial 3: present (0); absent, only hypobranchial-3 cartilage present (1).
75. Ceratobranchial 5: extending no further dorsally than other ceratobranchials (0); extending much further dorsally than other ceratobranchials (1).
76. Transversus ventralis V process on ceratobranchial 5: absent (0); present (1).
77. Ceratobranchial-5 teeth: teeth bound to ceratobranchial 5 by collagenous fibers (0); teeth ankylosed to ceratobranchial 5 (1).
78. Epibranchial 1: not dorsally overlapped by epibranchial 2 (0); dorsally overlapped by epibranchial 2 (1).
79. Anterior head of epibranchial 1: not twisted ventrad (0); twisted ventrad (1).
80. Anterior edge of epibranchial 1: without large membranous flange on anterior edge (0); with large membranous flange on anteroventral edge (1).
81. Pharyngobranchial-1 element: present and ossified (0); present but cartilaginous (1); absent (2).
82. Pharyngobranchial 3: not overlapping pharyngobranchial 2 (0); overlapping pharyngobranchial 2 (1).
83. Foramen between pharyngobranchials 2 and 3: absent (0); present (1).
84. Pharyngobranchial uncinate processes: present (0); absent (1).
85. Number of branchiostegal rays: more than three (0); three (1).
86. Interhyal element: present and ossified (0); present and cartilaginous (1); interhyal element absent (2).
87. Lateral face of the posterior ceratohyal: flat (0); with knob-like process posteriorly (1).
88. Lateral process of the second vertebral centrum: short (0); elongate, projecting well into somatic musculature (1).
89. Shape of the tripus: triangular (0); Y-shaped (1).
90. Neural spine of fourth vertebral centrum: similar in size or smaller than more posterior spines (0); larger than more posterior spines (1); absent (2).
91. Connection between lateral process of the second vertebral centrum and the outer arm of the os suspensorium: no contact between lateral process of the second vertebral centrum and the outer arm of the os suspensorium or contact without fusion of the two structures (0); ventral arm of the lateral process of the second vertebral centrum fused with the outer arm of the os suspensorium (1); both dorsal and ventral arms of the lateral process of the second vertebral centrum fused to the outer arm of the os suspensorium (2).
92. Supraneural 2: present (0); absent (0).

93. Supraneural 3: with simple median crest, without division (0); median crest with modified dorsal surface (1).
94. Supraneural 3/supraneural 5 contact: absent (0); present (1).
95. Claustrum: posterodorsal edge not sutured to anterior edge of neural complex (0); posterodorsal edge sutured to anterior edge of neural complex (1).
96. Dorsal prezygapophysis of fifth vertebral centrum: dorsal prezygapophysis of fifth vertebral centrum widely separate from neural arch of fourth vertebral centrum (0); dorsal prezygapophysis of fifth vertebral centrum abutting neural arch of fourth vertebral centrum (1).
97. Parapophyses: autogenous (0); fused to centra (1).
98. Ventral prezygapophyses: ventral prezygapophyses separate from ventral postzygapophyses on caudal vertebrae (0); ventral prezygapophyses strongly interdigitating with ventral postzygapophyses on caudal vertebrae (1).
99. Contact between fifth parapophysis and outer arm of the os suspensorium: absent (0); present (1).
100. Lateral process of the first vertebral centrum: absent or, when present, short (0); elongate, projecting well into somatic musculature (1).
101. Intercalarium: base with contact to vertebral column (1); base without contact to vertebral column (1).
102. Os suspensorium: not surrounding anterior swim-bladder chamber (0); completely surrounding anterior swim-bladder chamber (1).
103. Bony contact between anteriormost point of coracoid and cleithrum: present (0); absent (1).
104. Posttemporal: present (0); absent (1).
105. Postcleithrum: more than one (0); one (1); absent (2).
106. Cleithrum: well separated from vertebral elements (0); closely associated with lateral process of second vertebral centrum (1).
107. Number of pectoral radials: four (0); three (1).
108. Baudelot's ligament: ligamentous (0); with ligament ossification (1).
109. Cleithral–occipital ligament: absent (0); present (1).
110. Rib associated with pelvic splint: similar in shape to other ribs (0); greatly thickened and robust (1).
111. Third unbranched dorsal fin ray: segments of each hemitrich free (0); proximal segments of each hemitrich fused, forming a 'spine'-like ray (1).
112. Terminal compound centrum: preural centrum 1, ural centrum 1, and ural centrum 2 separate (0); preural centrum 1, ural centrum 1, and ural centrum 2 fused, forming a compound centrum (1).
113. Number of epurals: two or more (0); one (1).
114. Proximal tip of hypural 1: separate from surrounding hypural and parhypural elements, and autogenous with ural centrum 1 or compound centrum (0); separate from surrounding hypural and parhypural elements, and separated from compound centrum by a short hiatus (1); fused to proximal tip of parhypural (2); fused to proximal tip of parhypural + hypural 2 + compound centrum (3).
115. Hypural 6: present (0); absent (1).
116. Hypural 3: proximal tip not fused with associated centrum (0); proximal tip fused with compound centrum (1).
117. Lateral line-bearing scale size: moderate to large, height of lateral line-bearing scale much greater than height of lateral-line canal ossification (0); small, height of lateral line-bearing scale equal to or shorter than height of lateral-line canal ossification (1).
118. Anterior radii: present (0); absent (1).
119. Lateral-line canal ossifications: similar in size and shape along entire length of canal (0); third and fourth lateral-line canal ossifications much larger than more anterior and posterior ossifications (1).
120. Anterior swim-bladder chamber: single (0); divided (1).
121. Nasal: present (0); absent (1).
122. Dermethmoid: present (0); absent (1).
123. Vomer: present (0); absent (1).
124. Parietal: present (0); absent (1).
125. Dermopterotic: present (0); absent (1).
126. Extrascapular: present (0); absent (1).
127. Anterior and posterior sclerotic bones: present (0); absent (1).
128. Infraorbitals 2–5: present (0); absent (1).
129. Angular: present (0); absent (1).
130. Ectopterygoid: present (0); absent (1).
131. Hypobranchial-1 cartilage: present (0); absent (1).
132. Mesocoracoid: present (0); absent (1).
133. Pelvic radial-2 and -3 cartilage: present (0); absent (1).
134. Uroneural 2: present (0); absent (1).
135. Rudimentary neural arch on urostyle: present (0); absent (1).
136. Middle radials of dorsal and anal fins: present (0); absent (1).
137. Intermuscular bones: present (0); absent (1).
138. Scales: present (0); absent (1).
139. Epibranchial-5 cartilage: articulates only with epibranchial 4 (0); articulates not only with the levator process of epibranchial 4, but also with the tip of ceratobranchial 5 (1).
140. Pharyngobranchial 3: cartilage along posterior edge of pharyngobranchial 3 separate from cartilaginous head of epibranchial 4 (0); cartilage along

posterior edge of pharyngobranchial 3 confluent with the cartilaginous head of epibranchial 4 (1).
141. Scaphium: concha scaphii not extending to the end of the processus ascendens of the scaphium (0); concha scaphii reaches all the way up to the end of the processus ascendens of the scaphium (1).
142. Inner arms of ossa suspensoria: separate (0); fused in their proximal portion, greatly elongated, following the curvature of the anterior chamber of the swimbladder (1).
143. Gap between enlarged neural arches 3 and 4: open (0); filled by extensive development of cartilage (1).
144. Tripus: small (0); hypertrophied and unusually elongate (1).

Matrix-5 characters

1. Kinethmoid: absent (0); present (1).
2. Preethmoid: absent (0); present (1).
3. Second preethmoid element: absent (0); second preethmoid cartilage present (1); second preethmoid present (2).
4. Articulation between ethmoid complex and frontals: absent, posterodorsal edge of ethmoid complex or supraethmoid firmly sutured or tightly abutting anterior edge of frontals (0); present, posterodorsal edge of ethmoid complex articulating in a shallow facet along anterior edge of frontals (1).
5. Relationship between ethmoid complex and vomer: separate (0); fused (1).
6. Contact between orbitosphenoid and ethmoid complex: absent (0); present (1).
7. Anterolateral process of the lateral ethmoid: absent (0); present (1).
8. Erectile suborbital spine, formed by enlargement of lateral ethmoid: absent (0); present (1).
9. Interorbital septum: interorbital septum formed only by the orbitosphenoid (0); interorbital septum formed by the orbitosphenoid and a dorsal component of the parasphenoid (1).
10. Ventral keel on parasphenoid: absent (0); present (1).
11. Contact between orbitosphenoid and pterosphenoid: present (0); absent (1).
12. Basisphenoid: present (0); absent (1).
13. Anterior opening of trigemino-facial chamber: anterior opening of trigemino-facial chamber contained within prootic (0); anterior opening of trigemino-facial chamber on anterior edge of prootic, its anterior edge bordered by the posterior edge of the pterosphenoid or lateral wing of the parasphenoid (1).
14. Postepiphysial fontanelle: absent (0); present (1).
15. Connection between supraorbital sensory canal and otic sensory canal: present, supraorbital sensory canal connected to otic sensory canal (0); absent, supraorbital sensory canal disjunct from otic sensory canal (1).
16. Fenestration of the dilatator fossa: absent (0); present (1).
17. Pterotic fossa: absent (0); present (1).
18. Supratemporal commissure: located on parietal or extrascapular (0); located on supraoccipital (1).
19. Exoccipital: contacting antimere along midline (0); separate from antimere (1).
20. Subtemporal fossa: absent or represented only by a shallow concavity on the ventral surface of the neurocranium (0); present, represented by a deep concavity on the ventral surface of the neurocranium (1).
21. Basioccipital process: absent (0); present as long parallel processes, not united to form a canal around the dorsal aorta (1); present as a large bony structure, forming a canal around the dorsal aorta (2).
22. Fenestration of the basioccipital process: absent (0); present (1).
23. Pharyngeal process of the basioccipital process: absent (0); present, horizontal in cross section (1); present, terete in cross section (2).
24. Masticatory plate: absent (0); present (1).
25. Supraorbital sensory canal: present (0); absent (1).
26. Anterolateral process of the frontal: absent (0); present (1).
27. Lateral occipital foramen: absent (0); present (1).
28. Antorbital: present (0); absent (1).
29. Relationship between anteriormost portion of infraorbital sensory canal and infraorbital 1: anteriormost portion of infraorbital sensory canal completely enclosed on infraorbital 1 (0); anteriormost portion of infraorbital sensory canal completely or partially disjunct from infraorbital 1 (1).
30. Nature of infraorbital series posterior to infraorbital 1: infraorbital series posterior to infraorbital 1 represented by four, roughly plate-like bones (0); infraorbital series posterior to infraorbital 1 represented by a variable number of small tubular ossifications around the infraorbital canal (1).
31. Spatial arrangement of infraorbital 5 and supraorbital: in contact or separated only by a short distance (0); widely separate (1).
32. Dermopalatine: present (0); absent (1).
33. Dorsomedial autopalatine process: absent (0); present (1).
34. Autopalatine–endopterygoid articulation: absent (0); present (1).

35. Ectopterygoid–autopalatine overlap: present (0); absent (1).
36. Length of autopalatine: short, terminating level with or slightly anterior to area of articulation with ethmoid region (0); autopalatine greatly extended past area of articulation with ethmoid region, terminating dorsal to maxilla (1).
37. Preautopalatine element: absent (0); present and cartilaginous (1); present and ossified (2).
38. Mesiopreautopalatine: absent (0); present, cartilaginous (1); present, ossified (2).
39. Endopterygoid: expansive and sheet-like (0); reduced and rod-like posteriorly (1).
40. Metapterygoid–quadrate fenestra: absent (0); present (1).
41. Metapterygoid: simple posterior lamina near hyomandibular (0); with membrane bone flanges forming suture to anterior edge of the hyomandibular (1).
42. Symplectic: simple dorsal lamina near metapterygoid (0); with membrane bone flanges on dorsal edge forming suture to ventral edge of metapterygoid (1).
43. Lateral face of hyomandibular: without membrane bone flange (0); with large membrane bone flange overlapping preopercle (1).
44. Ventral margin of the opercle: straight or convex (0); concave (1).
45. Opercular sensory canal: absent (0); present (1).
46. Connection between preopercular sensory canal and infraorbital sensory canal: absent (0); present (1).
47. Interopercle: anteriormost point directed anteriorly (0); anteriormost point directed anteromedially towards ventral midline (1).
48. Interopercular–preopercular articulation: absent (0); present (1).
49. Anterodorsal head of opercle: absent, or if present short (0); greatly elongate (1).
50. Premaxilla: extending furthest dorsally at a point lateral to the dorsal midline (0); extending furthest dorsally adjacent to the dorsal midline (1).
51. Oral teeth: present (0); absent (1).
52. Coronomeckelian: present (0); absent (1).
53. Anguloarticular: similar in width to the posteriormost point of the dentary in ventral view, not extended laterally as a sharp point (0); much wider than the posteriormost point of the dentary in ventral view, extending laterally as a sharp point (1).
54. Maxilla: without large spine-like process overlapping dorsal surface of premaxillae (0); with large spine-like process overlapping dorsal surface of premaxillae (1).
55. Mandibular sensory canal: present (0); absent (1).
56. Basihyal toothplate: present (0); absent (1).
57. Pharyngobranchial tooth plates: present (0); absent (1).
58. Anterior basibranchial copula toothplate: present (0); absent (0).
59. Posterior basibranchial copula: with dermal toothplates associated with dorsal surface (0); with dermal (toothless) plates associated with dorsal surface (1); without dermal plates or toothplates (2).
60. Gill-raker teeth: present (0); absent (1).
61. Gill rakers: small-based elements, base narrower than total height (0); wide-based elements, base wider than total height (1).
62. Gill rakers: without comb-like projections along dorsal edge (0); with comb-like projections along dorsal edge (1).
63. Gill-filament ossifications: absent (0); present (1).
64. Basihyal: anterior edge straight or slightly rounded, with a singular cartilaginous tip (0); anterior edge bifurcated, with two prominent cartilaginous tips (1).
65. Sublingual ossifications: absent (0); present (1).
66. Posterior basibranchial copulae: complete (0); segmented (1).
67. Ventral keel on posterior basibranchial copula or derivative elements: present (0); absent (1).
68. Basibranchial 1: present (0); absent (1).
69. Basibranchial 2: anterior edge similar in width or only slightly wider than posterior edge (0); anterior edge markedly wider than posterior edge (1).
70. Posterior tip of basibranchial 3: on same plane as anterior tip, not intimately associated with the tip of hypobranchial 3 (0); lower than anterior tip, articulating in a concave facet on the anterior tip of hypobranchial 3 (1).
71. Basibranchial 4: absent (0); present (1).
72. Hypobranchial 3: present (0); absent, only hypobranchial-3 cartilage present (1).
73. Ceratobranchial 5: extending no further dorsally than other ceratobranchials (0); extending much further dorsally than other ceratobranchials (1).
74. Transversus ventralis V process on ceratobranchial 5: absent (0); present (1).
75. Ceratobranchial-5 teeth: teeth bound to ceratobranchial 5 by collagenous fibres (0); teeth ankylosed to ceratobranchial 5 (1).
76. Epibranchial 1: not dorsally overlapped by epibranchial 2 (0); dorsally overlapped by epibranchial 2 (1).
77. Anterior head of epibranchial 1: not twisted ventrad (0); twisted ventrad (1).
78. Anterior edge of epibranchial 1: without large membranous flange on anterior edge (0); with large membranous flange on anteroventral edge (1).

79. Pharyngobranchial-1 element: present and ossified (0); present but cartilaginous (1); absent (2).
80. Pharyngobranchial 3: not overlapping pharyngobranchial 2 (0); overlapping pharyngobranchial 2 (1).
81. Foramen between pharyngobranchials 2 and 3: absent (0); present (1).
82. Pharyngobranchial uncinate processes: present (0); absent (1).
83. Number of branchiostegal rays: more than three (0); three (1).
84. Lateral face of the posterior ceratohyal: flat (0); with knob-like process posteriorly (1).
85. Lateral process of the second vertebral centrum: short (0); elongate, projecting well into somatic musculature (1).
86. Shape of the tripus: triangular (0); Y-shaped (1).
87. Neural spine of fourth vertebral centrum: similar in size or smaller than more posterior spines (0); larger than more posterior spines (1); absent (2).
88. Connection between lateral process of the second vertebral centrum and the outer arm of the os suspensorium: no contact between lateral process of the second vertebral centrum and the outer arm of the os suspensorium or contact without fusion of the two structures (0); ventral arm of the lateral process of the second vertebral centrum fused with the outer arm of the os suspensorium (1); both dorsal and ventral arms of the lateral process of the second vertebral centrum fused to the outer arm of the os suspensorium (2).
89. Supraneural 2: present (0); absent (0).
90. Supraneural 3: with simple median crest, without division (0); median crest with modified dorsal surface (1).
91. Supraneural 3/supraneural 5 contact: absent (0); present (1).
92. Claustrum: posterodorsal edge not sutured to anterior edge of neural complex (0); posterodorsal edge sutured to anterior edge of neural complex (1).
93. Dorsal prezygapophysis of fifth vertebral centrum: dorsal prezygapophysis of fifth vertebral centrum widely separate from neural arch of fourth vertebral centrum (0); dorsal prezygapophysis of fifth vertebral centrum abutting neural arch of fourth vertebral centrum (1).
94. Parapophyses: autogenous (0); fused to centra (1).
95. Ventral prezygapophyses: ventral prezygapophyses separate from ventral postzygapophyses on caudal vertebrae (0); ventral prezygapophyses strongly interdigitating with ventral postzygapophyses on caudal vertebrae (1).
96. Contact between fifth parapophysis and outer arm of the os suspensorium: absent (0); present (1).
97. Lateral process of the first vertebral centrum: absent or, when present, short (0); elongate, projecting well into somatic musculature (1).
98. Intercalarium: base with contact to vertebral column (1); base without contact to vertebral column (1).
99. Os suspensorium: not surrounding anterior swim-bladder chamber (0); completely surrounding anterior swim-bladder chamber (1).
100. Bony contact between anteriormost point of coracoid and cleithrum: present (0); absent (1).
101. Posttemporal: present (0); absent (1).
102. Postcleithrum: more than one (0); one (1); absent (2).
103. Cleithrum: well separated from vertebral elements (0); closely associated with lateral process of second vertebral centrum (1).
104. Number of pectoral radials: four (0); three (1).
105. Baudelot's ligament: ligamentous (0); with ligament ossification (1).
106. Cleithral–occipital ligament: absent (0); present (1).
107. Rib associated with pelvic splint: similar in shape to other ribs (0); greatly thickened and robust (1).
108. Third unbranched dorsal fin ray: segments of each hemitrich free (0); proximal segments of each hemitrich fused, forming a 'spine'-like ray (1).
109. Terminal compound centrum: preural centrum 1, ural centrum 1, and ural centrum 2 separate (0); preural centrum 1, ural centrum 1, and ural centrum 2 fused, forming a compound centrum (1).
110. Number of epurals: two or more (0); one (1).
111. Proximal tip of hypural 1: separate from surrounding hypural and parhypural elements, and autogenous with ural centrum 1 or compound centrum (0); separate from surrounding hypural and parhypural elements, and separated from compound centrum by a short hiatus (1); fused to proximal tip of parhypural (2); fused to proximal tip of parhypural + hypural 2 + compound centrum (3).
112. Hypural 3: proximal tip not fused with associated centrum (0); proximal tip fused with compound centrum (1).
113. Lateral line-bearing scale size: moderate to large, height of lateral line-bearing scale much greater than height of lateral-line canal ossification (0); small, height of lateral line-bearing scale equal to or shorter than height of lateral-line canal ossification (1).
114. Anterior radii: present (0); absent (1).
115. Lateral-line canal ossifications: similar in size and shape along entire length of canal (0); third and fourth lateral-line canal ossifications much larger than more anterior and posterior ossifications (1).

116. Anterior swim-bladder chamber: single (0); divided (1).
117. Dermethmoid: present (0); absent (1).
118. Vomer: present (0); absent (1).
119. Parietal: present (0); absent (1).
120. Dermopterotic: present (0); absent (1).
121. Angular: present (0); absent (1).
122. Ectopterygoid: present (0); absent (1).
123. Pelvic radial-2 and -3 cartilage: present (0); absent (1).
124. Rudimentary neural arch on urostyle: present (0); absent (1).
125. Middle radials of dorsal and anal fins: present (0); absent (1).
126. Intermuscular bones: present (0); absent (1).
127. Scales: present (0); absent (1).
128. Epibranchial-5 cartilage: articulates only with epibranchial 4 (0); articulates not only with the levator process of epibranchial 4, but also with the tip of ceratobranchial 5 (1).
129. Pharyngobranchial 3: cartilage along posterior edge of pharyngobranchial 3 separate from cartilaginous head of epibranchial 4 (0); cartilage along posterior edge of pharyngobranchial 3 confluent with the cartilaginous head of epibranchial 4 (1).
130. Scaphium: concha scaphii not extending to the end of the processus ascendens of the scaphium (0); concha scaphii reaches all the way up to the end of the processus ascendens of the scaphium (1).
131. Inner arms of ossa suspensoria: separate (0); fused in their proximal portion, greatly elongated, following the curvature of the anterior chamber of the swimbladder (1).
132. Gap between enlarged neural arches 3 and 4: open (0); filled by extensive development of cartilage (1).
133. Tripus: small (0); hypertrophied and unusually elongate (1).

APPENDIX 2

CHARACTER MATRIX 1

68 characters from Mayden & Chen (2010) for 85 taxa coded based on information in Mayden & Chen (2010). See text for further explanation.

	1	2	3	4	5	6	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	12345678
Gonorynchus greyi	0000000000	0000000000	0000000000	0000000000	0000000000	0000000000	00000000
Pseudobagrus tokiensis	1000000000	0000000000	0000000000	0000000000	0000000000	0000000000	00000000
Chalceus macrolepidotus	1000000000	0000000000	0000000000	0000000000	0000000000	0000000000	00000000
Phenacogrammus interruptus	1000000000	0000000000	0000000000	0000000000	0000000000	0000000000	00000000
Sewellia lineolata	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Homaloptera parclitella	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Catostomus commersoni	1211112102	1111110021	0000000001	0000000000	0000000000	0000000000	00000000
Cycleptus elongatus	1211112102	1111110021	0000000001	0000000000	0000000000	0000000000	00000000
Botia Dario	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Leptobotia pellegrini	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Syncrossus beauforti	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Yasuhikotakia morleti	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Acantopsis sp.	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Niwaella multifasciata	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Pangio oblonga	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Cobitis takatsuensis	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Canthophrys gongota	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Gyrinocheilus aymonieri	1211112102	1111110001	0000000001	0000000000	0000000000	0000000000	00000000
Gyrinocheilus pennocki	1211112102	1111110001	0000000001	0000000000	0000000000	0000000000	00000000
Acanthocobitis sp.	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Barbatula barbatula	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Lefua costata	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Oreonectes platycephalus	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Schistura savona	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Traccatichthys pulcher	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Triplophysa gundriseri	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Vaillantella maassi	1211112102	1111110101	0000000001	0000000000	0000000000	0000000000	00000000
Acheilognathus tabira	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Paracheilognathus himantegus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000

APPENDIX 2 *Continued*

	1	2	3	4	5	6	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	12345678
Rhodeus o. kurumeus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Aphyocypris chinensis	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Ischikauia steenackeri	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Macrochirichthys macrochirus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Opsariichthys uncirostris	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Paralaubuca typus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Zacco sieboldii	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Acrossocheilus paradoxus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Barbonymus gonionotus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Barbus callipterus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Garra spilota	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Gymnocypris przewalskii	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Hampala macrolepidota	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Labeo chrysophekadion	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Puntius titteya	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Gobio gobio	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Hemibarbus barbus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Squalidus chankaensis	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Notropis baileyi	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Pelecus cultratus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Phoxinus p. sachalinensis	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Scardinius erythropthalmus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Semotilus atromaculatus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Paedocypris sp. 1	1111101011	1111111010	0101100000	1111111111	1111111111	1111111100	00000000
Paedocypris sp. 2	1111101011	1111111010	0101100000	1111111111	1111111111	1111111100	00000000
Psilorhynchus sucatio	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Psilorhynchus homaloptera	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Barilius bendelisis	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Boraras merah	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Chela dadiburjori	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Danio erythromicron	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Danio margaritatus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Danio albolineatus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Danio dangila	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Danio rerio	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Danionella sp.	1211112102	1111110111	1110011111	0000011111	1111111110	0000000011	11000000
Devario regina	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Esomus longimanus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Horadandia atukorali	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Luciosoma setigerum	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Microdevario kubotai	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Microdevario nana	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Microrasbora rubescens	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Rasbora bankanensis	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Rasbora steineri	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Trigonostigma heteromorpha	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Sundadanio axelrodi	1211112102	1111110111	1110011111	0000000000	0222211110	0000000000	00111111
Tanichthys albonubes	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Tinca tinca	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Aspidoparia morar	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Biwia zezera	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Megalobrama amblycephala	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Romanogobio ciscaucasicus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Sarcocheilichthys parvus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Yaoshanicus arcus	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000
Leptobarbus hoevenii	1211112102	1111110111	1110011111	0000000000	0000000000	0000000000	00000000

APPENDIX 3

CHARACTER MATRIX 2

61 characters from Mayden & Chen (2010) for 79 taxa coded based on examination of vouchered material listed in Table 1. "?" represents missing data. "-" represents inapplicable character coding. "X" represents polymorphic character coding. See text for further explanation.

	1	2	3	4	5	6	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1
Gonorynchus greyi	000--00001	111011-000	10--001000	0000?0000-	-0-000--00	0-000000-0	0
Pseudobagrus tokiensis	1000-00-00	0?0110-100	101-00-100	0000000111	0000000000	000000000-	0
Chalceus macrolepidotus	1000000000	0?11000000	0001001000	0000000000	0000000000	0000000000	0
Phenacogrammus interruptus	1000100000	0101000000	0001001000	0000000000	0000000000	0000000000	0
Sewellia lineolata	1201111111	111111-100	110-001100	0000000011	0000001000	0000000000	0
Homaloptera sp.	1201111111	1111111100	110-011100	0000000001	0000001000	0000000000	0
Catostomus commersoni	1211111111	1111111000	000-001000	0000000100	0000000000	0000000000	0
Cycleptus elongatus	1211111111	1111111000	0000001000	0000000100	0000000000	0000000000	0
Botia almorhae	1201101111	1111111100	111-001100	000000000?	0000001000	0000000000	0
Leptobotia pellegrini	1201101111	1111111100	111-001100	0000000000	0000001000	0000000000	0
Syncrossus hymenophysa	1201101111	1111111100	111-001100	0000000000	0000001000	0000000000	0
Yasuhikotakia morleti	1201101111	1111111100	111-001100	0000000000	0000001000	0000000000	0
Acantopsis dialuzona	1201101111	111111-100	110--01100	0000000011	0000001000	0000000000	0
Pangio muraeniformis	1201101111	111111-100	11---01100	0000000011	0000001000	0000000000	0
Niwaella multifasciata	1201101111	111111-100	11---01100	0000000011	0000001000	0000000000	0
Cobitis dalmatina	1201101111	111111-100	11---01100	0000000011	0000001000	0000000000	0
Canthophrys gongota	1201101111	111111-100	11---01100	0000000011	0000001000	0000000000	0
Gyrinocheilus aymonieri	120-111111	1111111000	0001-01000	0000000100	0000001000	0000000000	0
Gyrinocheilus pennocki	120-111111	1111111000	0001-00000	0000000100	0000001000	0000000000	0
Acanthocobitis sp.	1201101111	1111110100	110--01100	0000000000	0000001000	0000000000	0
Barbatula barbatula	1201111111	111111-100	110--01100	0000000010	0000001000	0000000000	0
Lefua costata	1201111111	1111-1-100	110--01100	0000000010	0000001000	0000000000	0
Oreonectes sp.	1201111111	111111-100	110--01100	0000000010	0000001000	0000000000	0
Schistura balteata	1201101111	111111-100	110--01100	0000000001	0000001000	0000000000	0
Traccatichthys pulcher	1201111111	1111111100	110--01100	0000000000	0000001000	0000000000	0
Triplophysa microps	1201111111	111111-100	110--01100	0000000011	0000001000	0000000000	0
Vaillantella euepiptera	1201101111	111111-100	110-001100	0000000010	0000001000	0000000000	0
Acheilognathus typus	1211111111	1111111011	1111011010	0000000000	0000000000	0000000000	0
Paracheilognathus himantegus	1211111111	1111111111	1111011010	0000000000	0000000000	0000000000	0
Rhodeus ocellatus	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Aphyocypris chinensis	1211111111	1111111011	110?011010	0000000000	0000000000	0000000000	0
Ischikauia steenackeri	1211111111	1111111011	1100111000	0000000000	0000000000	0000000000	0
Macrochirichthys macrochirus	1211111111	1111111011	111?111010	0000000000	0000000000	0000000000	0
Opsariichthys uncirostris	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Paralaubuca typus	1211101111	1111111011	1101111010	0000000000	0000000000	0000000000	0
Zacco sieboldii	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Acrossocheilus paradoxus	1211111111	1111111111	1111111010	0000000000	0000000000	0000000000	0
Barbonymus schwanenfeldii	1211111111	1111111111	1111111010	0000000000	0000000000	0000000000	0
Barbus leonensis	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Garra flavatra	1211111111	111111-111	1100111110	0000000000	0000000000	0000000000	0
Schizothorax graham	1211111111	111111?111	1101111110	0000000000	0000000000	0000000000	0
Hampala macrolepidota	1211111111	1111111111	1111111010	0000000000	0000000000	0000000000	0
Labeo cf. *longipinnis*	1211111111	1111111111	1111111010	0000000000	0000000000	0000000000	0
Puntius titteya	1211111111	1111111111	111-011010	0000000000	0000000000	0000000000	0
Gobio gobio	1211111111	1111111111	1101111010	0000000000	0000000000	0000000000	0
Hemibarbus barbus	1211111111	1111111011	1101111010	0000000000	0000000000	0000000000	0
Squalidus majimae	1211111111	1111111111	1101111010	0000000000	0000000000	0000000000	0
Notropis stramineus	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Pelecus cultratus	1211111111	1111111011	1101111010	0000000000	0000000000	0000000000	0
Phoxinus sp.	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Leuciscus leuciscus	1211111111	1111111011	1101111?10	0000000000	0000000000	0000000000	0

APPENDIX 3 *Continued*

	1	2	3	4	5	6	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1
Semotilus atromaculatus	1211111111	1111111011	1101?10010	0000000000	0000000000	0000000000	0
Paedocypris sp.	11111000-0	111111-011	010--0-001	1111111111	1111111111	1000000000	0
Paedocypris sp.	11111000-0	111111-011	010--0-001	1111111111	1111111111	1000000000	0
Psilorhynchus sucatio	1211111111	111111-011	110-011101	0000010011	0000010000	0000000000	0
Psilorhynchus balitora	1211111111	111111-011	1101011101	0000010011	0000010000	0000000000	0
Barilius bendelisis	1211111111	1111111111	1100111110	0000000000	0000000000	0000000000	0
Boraras merah	1211111111	1111111011	110-011110	0000000000	0000000000	0000000000	0
Chela laubuca	1211111111	1111111011	1100011010	0000000000	0000000000	0000000000	0
Danio erythromicron	1211111111	1111111011	1100011010	0000000000	0000000000	0000000000	0
Danio margaritatus	1211111111	1111111011	1100011010	0000000000	0000000000	0000000000	0
Danio albolineatus	1211111111	1111111111	1100011010	0000000000	0000000000	0000000000	0
Danio dangila	1211111111	1111111111	1100011010	0000000000	0000000000	0000000000	0
Danio rerio	1211111111	1111111111	1100011010	0000000000	0000000000	0000000000	0
Danionella dracula	10111000-1	1111111011	110--0-001	0100111111	1111110000	0111100000	0
Devario devario	1211111111	1111111111	1100111010	0000000000	0000000000	0000000000	0
Esomus caudiocellatus	1211111111	111111-111	1100111010	0000000000	0000000000	0000000000	0
Horadandia atukorali	1211111111	1111111011	111-011110	0000000000	0000000000	0000000000	0
Luciosoma sp.	1211111111	1111111111	1100110110	0000-00000	0000000000	0000000000	0
Microdevario kubotai	1211111111	1111111011	1100011010	0000000000	0000000000	0000000000	0
Microdevario nana	1211101111	1111111011	110-011010	0000000000	0000000000	0000000000	0
Microrasbora rubescens	1211111111	1111111011	1100011010	0000000000	0000000000	0000000000	0
Rasbora pauciperforata	1211111111	1111111011	1101011110	0000000000	0000000000	0000000000	0
Rasbora trilineata	1211111111	1111111011	1101111110	0000000000	0000000000	0000000000	0
Trigonostigma hengeli	1211111111	1111111011	1100011110	0000000000	0000000000	0000000000	0
Sundadanio echinus	12111111-1	111111-011	110-001100	0000000001	111X110000	0000011111	1
Tanichthys albonubes	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Tinca tinca	1211111111	1111111011	1101011010	0000000000	0000000000	0000000000	0
Leptobarbus hoevenii	1211111111	1111111111	1111111110	0000000000	0000000000	0000000000	0

APPENDIX 4

CHARACTER MATRIX 3

120 characters from Conway (2011) for 65 taxa coded based on examination of vouchered material listed in Conway (2011) and Britz & Conway (2009). "?" represents missing data. "-" represents inapplicable character coding. "X" represents polymorphic character coding. *Lepidocephalichthys* cf. *guntea* is abbreviated as *Lepidoceph*. cf. *guntea*; *Acanthocobitis* cf. *zonalternans* is abbreviated as *Acantho*. cf. *zonalternans*.

	0	0	0	0	0	0	0	0	0	1	1	1
	1	2	3	4	5	6	7	8	9	0	1	2
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
*Elops*_sp._cf._*senegalensis*	0000001000	0000000100	00--000000	0000000000	0000000000	0000000000	0000000000	0000000000	0000000--0	-000-00000	--00000000	000000000-
Hiodon alosoides	0000000000	0000100010	00--000010	0-0----000	0000000000	0000000000	0000000000	0000000000	0000010--0	-000-01000	--00100000	001000000-
Denticeps clupeoides	0000000000	0010000001	00--000000	0100000000	0000000000	0000001012	1000000000	0000000000	0000000--0	-000-01000	--00101000	000010?0-
Distichodus antonii	0000000001	0101000000	00--000100	0110000000	1001000000	0000001002	0000000000	0000000000	0000000000	0100-00000	0000000000	0101000100
Psilorhynchus balitora	1100000000	0111000001	1201100110	0111111000	00010-1010	1100011112	1000010100	0110101000	2001100102	200-000010	0001200?01	0112100100
Cyprinus carpio	1120000011	0111010000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0000000010	0000100000	1112000000
Barbus barbus	1100000010	0110010000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0000000010	0000100000	1112000000
Balantiocheilos melanopterus	1100000010	0111010000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0000000010	0000100000	1112000000
Garra dembeensis	1100001011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000101	0000101100	2101100100	000-000010	0000200001	0112000000
Osteochilus hasselti	1100000011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000111	0000101101	2101100100	000-000010	0000100000	0112000000
Labeo longipinnis	1100001011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000111	0000101101	2101100100	000-000010	0000100000	0112000000
Lobocheilos bo	1100000011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000101	0000101101	2101100100	0000000010	0000100000	0112000000
Crossocheilus reticulatus	1100000011	0111010001	1202100110	0111110000	0011010010	1100001112	1000000101	0000101101	2101100100	000-000010	0000100000	0112000000
Gobio gobio	1100000000	0110100000	1201100110	0111110000	0000010000	1100001112	1000010100	0000101000	2101100100	0011000010	0000100000	0112000100
Abramis brama	1100000000	0110100000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000000
Zacco cf. *platypus*	1100000000	0110100000	1201100110	0011110000	0000000000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000100
Rhodeus sericeus	1100000001	0110100-01	1201100110	0111110000	00000-0000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000100
Notemigonus crysoleucas	1100000000	0110100001	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000000
Rasbora cephalotaenia	1100000000	0110000001	1201100110	0011110000	0000010000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000000
Phoxinus phoxinus	1100000000	0110100000	1201100110	0111110000	00000-0000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112001000
Gyrinocheilus pennocki	1100001001	0100001100	0100000110	1110111000	00000-1100	0100011111	1000000110	000100-000	0011110100	0000110001	1000100100	0112000000
Gyrinocheilus aymonieri	1100001001	0110001100	0100000110	1110111000	00000-1100	0100011111	1000000110	000100-000	0011110100	0000110001	1000100100	0112000000
Myxocyprinus asiaticus	1110001010	0101001100	0210000111	1111110110	01100-1000	1100011112	1010011000	0001101000	10011?1100	1000010000	0000100000	0112010000
Cycleptus elongatus	1110001000	0110101100	0210000111	1111110110	01100-1000	1100011112	1010011000	0001101000	1001101100	1000010000	0000100000	0112010000
Catostomus commersoni	1110001000	0111001-00	021001011-	1-11110110	01100--001	1100011112	1010011000	0001101000	1001101100	1000010000	0000100000	0112010000
Carpiodes carpio	1110001010	0101101100	0210000111	1111110110	01100-1001	1100011112	1010011000	0001101000	1001101100	1000010001	0000100100	0112010000
Syncrossus cf. *berdmorei*	1021110110	0111000100	0100001111	1-10110220	0001000000	1100001112	1000011011	0000111000	2001101101	0100110100	1000100010	0112001010
Chromobotia macracanthus	1021110110	0111001100	0100001111	1-10110220	0001000000	1100001112	1000011011	0100111000	2101101101	0100110100	1000100010	0112001010
Yasuhikotakia modesta	1021111111	0111000100	0100001111	1-10110220	0001000000	1100001112	---0011011	0000111000	2101101101	0100110100	1000100010	0112001010
Yasuhikotakia sidthimunki	1021110110	0111000100	0100000111	1-10110220	0001000000	1100001112	---0011010	0000111000	2001101100	0100100100	1000100010	0112001?10
Vaillantella euepiptera	1021111000	0110000100	0100000011	1-10110200	0001000000	1100001112	---0011110	0000111000	2001100100	0100100100	1000200010	0112001?10
Cobitis pacifica	1021110100	1111-00-01	010001011-	--10110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	0113101010

APPENDIX 4 *Continued*

	0	0	0	0	0	0	0	0	0	1	1	1
	1	2	3	4	5	6	7	8	9	0	1	2
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890
Cobitis cf. *taenia*	1021110100	1111-00-01	0100011011-	--10110001	10011--000	1110011112	1001011011	1100111000	2001100100	000-101000	1110201010	0113101010
Iksookimia koreensis	1021110100	1111-00-01	0100011011-	--10110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	0113101010
Niwaella multifasciata	1021110100	1111-00-01	0100011011-	--10110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	0113101010
Acantopsis choirorhynchus	1021110100	1111-00100	0100011011-	1110110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1100201010	0113101010
Lepidoceph. cf. *guntea*	1021110100	1111-00-01	0100011011-	-1101110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	0113101010
Barbatula barbatula	1120010000	0111100101	0200000111	1-11110200	00010-0000	1100011112	--11100011	0100111000	2001100110	200-100001	1000201010	0112101011
Mesonoemacheilus triangularis	1020111000	0111000100	0200000111	1-11110220	00010-0000	1100001112	1001111011	0100111000	2001100110	2000100101	1000200010	0112001011
Lefua costata	1120011000	0111-00-01	0200011011-	--11110200	00010--000	1100011112	1001011011	0100111000	2001100110	210-100001	1000200010	01-2101?11
Acantho. cf. *zonalternans*	1020110000	0111000100	0200000111	1-11110200	00010-0000	1100001112	1001111011	0100111000	2001100110	2000100101	1000100010	0112101011
Annamia normani	1120011001	0111000111	0202000011	1-11110000	00010-0000	1101111112	1101011011	0100011000	2001110110	201-100000	1000210011	0113101?11
Gastromyzon cranbrooki	1120011000	0111000111	0201000011	1-11110000	00010-1000	1101111112	1101-11011	0100011000	2001110110	201-100000	1000110011	0112101011
Homaloptera stephensoni	1120111000	0111000111	00--000011	1111110200	00010-0000	1100011112	1001011011	0100111000	2001120110	200-100000	1000210011	0111101011
Homaloptera zollingeri	1120011000	0110000111	00--000011	1111110200	00010-0000	1100011112	1001111011	0100111000	2001120110	200-100000	1000210011	0112100011
Paedocypris sp.	1000-00000	11-1-00-11	020011011-	--100-0000	00000--000	110-011112	---0000000	0011101000	2001110102	010-000-10	10-1201000	011010---0
Danionella dracula	0000-00000	11-1-00-11	020011011-	--100-0000	0--00--000	?10-011112	---0000000	0011101000	2001110102	010-000-10	00-120-00-	011010---0
Sundadanio echinus	1000-00000	0111-00-11	020110011-	01111-0000	00000--000	1100011112	1000000000	0000101000	2001100100	0000000010	0001201000	0112X0---0
Danio rerio	1100000000	0110000-00	1201100110	0011110000	0000000000	1100001112	1000000000	0000101000	2101110100	0000000010	0000100000	0112100000
Danio erythromicron	1100000000	0110-00-01	120111011-	0011110000	000000-000	1100011112	1000000000	0000101000	2101110100	0000000010	0000100000	011210-0-0
Danio margaritatus	1100000000	0110-00-01	120111011-	0011110000	000000-000	1100011112	1000000000	0000101000	2101110100	0000000010	0000100000	011210-0-0
Boraras brigittae	1100000000	0110-00-01	1201110110	0-11110000	00000--000	1100011112	1000000000	0000101000	2101100100	0000000010	0000100000	011210-0-0
Boraras merah	1100000000	0110-00-01	1201110110	0-11110000	00000--000	1100011112	1000000000	0000101000	2101100100	0000000010	0000100000	011210-0-0
Boraras maculatus	1100000000	0110-00-01	1201110110	0-11110000	00000--000	1100011112	1000000000	0000101000	2101100100	0000000010	0000100000	011210-0-0
Boraras micros	1000000000	0110-00-01	120111011-	0-11110000	00000--000	1100011112	1000000000	0010101000	2101100100	0000000010	0000100000	011210-0-0
Boraras urophthalmoides	1100000000	0110-00-01	1201110110	0-11110000	00000--000	1100011112	1000000000	0010101000	2101101000	0000000010	0000100000	011210-0-0
Horadandia atukorali	1000000000	0111-10-00	120110011-	0-11110000	00000--000	1100001112	1000000000	0000101000	2101101000	0000000010	0000100000	011210-0-0
Microrasbora rubescens	1100000000	0110-00-01	120110011-	0111110000	00000--000	1100001112	1000000000	0000101000	2101111000	0010000010	0000100000	011210-0-0
Microdevario kubotai	1100000000	0110100-01	1201100110	0111110000	000000-000	1100001112	1000000000	0010101000	2101111000	0000000011	0000100000	011210-0-0
Microdevario nana	1000000000	0110-00-01	120110011-	0-11110000	00000--000	1100001112	1000000000	0000101000	2101111000	0000000011	0000101000	011210-0-0
Sawbwa resplendens	1120000000	0111-00-00	120111011-	--11110000	00000--000	1100011112	1000000000	0000101000	2101100100	0000000010	0000101000	111200---0
Tanichthys micagemmae	1100000000	0110-00-01	120111011-	--11110000	00000--000	1100001112	1000000000	0010101000	2101100100	0010000010	0000100000	0112100000
Rasbora kalbarensis	1100000000	0110-00-00	120110011-	--11110000	00000--000	1100001112	1000000000	0000101000	2101100100	0000000010	0000100000	011210---0
Rasboroides vaterifloris	1100000010	0111100-01	1201100110	0-11110000	00000000000	1100001112	1000000000	0000101000	2101101000	0000000010	0000100000	0112000000
Trigonostigma hengeli	1100000010	0110100-01	1201100110	0011110000	00000000000	1100001112	1000000000	0000101000	2101101000	0000000010	0000100000	0112000000

APPENDIX 5

CHARACTER MATRIX 4

144 characters from Conway (2011; characters 1–120) and Britz & Conway (2009; 121–144) for 65 taxa coded based on examination of vouchered material listed in Conway (2011) and Britz & Conway (2009). "?" represents missing data. "-" represents inapplicable character coding. "x" represents polymorphic character coding. *Lepidocephalichthys* cf. *guntea* is abbreviated as *Lepidoceph.* cf. *guntea*; *Acanthocobitis* cf. *zonalternans* is abbreviated as *Acantho.* cf. *zonalternans*.

	0	0	0	0	0	0	0	0	0	1	1	1	1	1	
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234
Elops sp. cf. *senegalensis*	0000001000	0000000100	00--000000	0000000000	0000000000	0000000000	0000000000	0000000000	0000000--0	-000-00000	--00000000	000000000-	0000000000	0000-00000	--0-
Hiodon alosoides	0000000000	0000100010	00--000010	0-0----000	0000000000	0000000000	0000000000	0000000000	0000010--0	-000-01000	--00100000	001000000-	0000000000	0000-00000	--0-
Denticeps clupeoides	0000000000	0010000001	00--000000	0100000000	0000000000	0000001012	1000000000	0000000000	0000000--0	-000-01000	--00101000	000010?0-	0000000000	0000-00000	--0-
Distichodus antonii	0000000001	0101000000	00--000100	0110000000	1001000000	0000001002	0000000000	0000000000	0000000000	0100-00000	0000000000	0101000100	0000000000	0000000000	0000
Psilorhynchus balitora	1100000000	0111000001	1201100110	0111111000	00010-1010	1100011112	1000010100	0110101000	2001100102	200-000010	0001200?01	0112100100	0000000000	0001000001	1000
Cyprinus carpio	1120000011	0111010000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0000000010	0000100000	1112000000	0000000000	0000000000	0000
Barbus barbus	1100000010	0110010000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0000000010	0000100000	1112000000	0000000000	0000000000	0000
Balantiocheilos melanopterus	1100000010	0111010000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0000000010	0000100000	1112000000	0000000000	0000000000	0000
Garra dembeensis	1100001011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000101	0000101100	2101100100	000-000010	0000200001	0112000000	0000000000	0000000000	0000
Osteochilus hasselti	1100000011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000111	0000101101	2101100100	000-000010	0000100000	0112000000	0000000000	0000000000	0000
Labeo longipinnis	1100001011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000111	0000101101	2101100100	000-000010	0000100000	0112000000	0000000000	0000000000	0000
Lobocheilos bo	1100000011	0110010001	1202100110	0111110000	0010010010	1100001112	1000000101	0000101101	2101100100	0000000010	0000100000	0112000000	0000000000	0000000000	0000
Crossocheilus reticulatus	1100000011	0111010001	1202100110	0111110000	0011010010	1100001112	1000000101	0000101101	2101100100	000-000010	0000100000	0112000000	0000000000	0000000000	0000
Gobio gobio	1100000000	0110100000	1201100110	0111110000	0000010000	1100001112	1000010100	0000101000	2101100100	0011000010	0000100000	0112000100	0000000000	0000000000	0000
Abramis brama	1100000000	0110100000	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000000	0000000000	0000000000	0000
Zacco cf. *platypus*	1100000000	0110100000	1201100110	0011110000	0000000000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000100	0000000000	0000000000	0000
Rhodeus sericeus	1100000001	0110100-01	1201100110	0111110000	00000-0000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000100	0000000000	0000000000	0000
Notemigonus crysoleucas	1100000000	0110100001	1201100110	0111110000	0000010000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000000	0000000000	0000000000	0000
Rasbora cephalotaenia	1100000000	0110000001	1201100110	0011110000	0000010000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112000000	0000000000	0000000000	0000
Phoxinus phoxinus	1100000000	0110100000	1201100110	0111110000	00000-0000	1100001112	1000000100	0000101000	2101100100	0011000010	0000100000	0112001000	0000000000	0000000000	0000
Gyrinocheilus pennocki	1100001001	0100001100	0100000110	1110111000	00000-1100	0100011111	1000000110	000100-000	0011110100	0000110001	1000100100	0112000000	0000000000	0000000000	1000
Gyrinocheilus aymonieri	1100001001	0110001100	0100000110	1110111000	00000-1100	0100011111	1000000110	000100-000	0011110100	0000110001	1000100100	0112000000	0000000000	0000000000	1000
Myxocyprinus asiaticus	1110001010	0101001100	0210000111	1111110110	01100-1000	1100011112	1010011000	0001101000	1001101100	10011?1100	1000010000	0000100000	0112010000	0000000000	0000
Cycleptus elongates	1110001000	0110101100	0210000111	1111110110	01100-1000	1100011112	1010011000	0001101000	1001101100	1000010000	0000100000	0112010000	0000000000	0000000000	0000
Catostomus commersoni	1110001000	0111001-00	021001011-	1-11110110	01100--001	1100011112	1010011000	0001101000	1001101100	1000010000	0000100000	0112010000	0000000000	0000000000	0000
Carpiodes carpio	1110001010	0101101100	0210000111	1111110110	01100-1001	1100011112	1010011000	0001101000	1001101100	1000010001	0000100100	0112010000	0000000000	0000000000	0000
Syncrossus cf. *berdmorei*	1021110110	0111000100	0100001111	1-10110220	0001000000	1100001112	1000011011	0000111000	2001101101	0100110100	1000100010	0112001010	0000000000	0001000000	0000
Chromobotia macracanthus	1021110110	0111001100	0100001111	1-10110220	0001000000	1100001112	1000011011	0100111000	2101101101	0100110100	1000100010	0112001010	0000000000	0001000000	0000
Yasuhikotakia modesta	1021111111	0111000100	0100001111	1-10110220	0001000000	1100001112	---0011011	0000111000	2101101101	0100110100	1000100010	0112001010	0000000000	0001000000	0000
Yasuhikotakia sidthimunki	1021110110	0111000100	0100000111	1-10110220	0001000000	1100001112	---0011010	0000111000	2001101100	0100100100	1000100010	0112001?10	0000000000	0001000000	0000
Vaillantella euepiptera	1021111000	0110000100	0100000011	1-10110200	0001000000	1100001112	---0011110	0000111000	2001100100	0100100100	1000200010	0112001?10	0000000000	0001000000	0000

APPENDIX 5 *Continued*

	0		0		0	0	0	0	0	0		1		1		1		1	
	1		2		3	4	5	6	7	8		9		0		1		2		3		4	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234
Cobitis pacifica	1021110100	1111-00-01	010001011-	--10110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	0113101010	1000110100	0001000000	1000
Cobitis cf. *taenia*	1021110100	1111-00-01	010001011-	--10110001	10011--000	1110011112	1001011011	1100111000	2001100100	000-101000	1110201010	0113101010	1000110100	0001000000	1000
Iksookimia koreensis	1021110100	1111-00-01	010001011-	--10110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	0113101010	1000110100	0001000000	1000
Niwaella multifasciata	1021110100	1111-00-01	010001011-	--10110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	0113101010	1000110100	0001000000	1000
Acantopsis choirorhynchus	1021110100	1111-00100	010001011-	11101110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1100201010	0113101010	0000010000	0001000000	1000
Lepidoceph. cf. *guntea*	1021110100	1111-00-01	010001011-	-1101110001	10011--000	1110011112	1001011011	1100111010	2001100100	000-101000	1110201010	01131010-0	1000110100	0001000000	1000
Barbatula barbatula	1120010000	0111100101	0200000111	1-11110200	00010-0000	1100011112	---1110011	0100111000	2001100110	200-100001	1000201010	0112101011	0000000000	0001000000	1000
Mesonoemacheilus triangularis	1020111000	0111000100	0200000111	1-11110220	00010-0000	1100001112	1001111011	0100111000	2001100110	2000100101	1000200010	0112001011	0000000000	0001000000	1000
Lefua costata	1120011000	0111-00-01	020001011-	--11110200	00010--000	1100011112	1001011011	0100111000	2001100110	210-100001	1000200010	01-2101?11	0000000000	0001000000	1000
Acantho. cf. *zonalternans*	1020110000	0111000100	0200000111	1-11110200	00010-0000	1100001112	1001111011	0100111000	2001100110	2000100101	1000100010	0112101011	0000000000	0001000000	1000
Annamia normani	1120011001	0111000111	0202000011	1-11110000	00010-0000	1101111112	1101011011	0100011000	2001110110	201-100000	1000210011	01131012?11	0000000000	0001000000	1000
Gastromyzon cranbrooki	1120011000	0111000111	0201000011	1-11110000	00010-1000	1101111112	1101-11011	0100011000	2001110110	201-100000	1000110011	0112101011	0000000000	0001000000	1000
Homaloptera stephensoni	1120111000	0111000111	00--000011	11111110200	00010-0000	1100011112	1001011011	0100111000	2001120110	200-100000	1000210011	0111101011	0000000000	0001000000	1000
Homaloptera zollingeri	1120011000	0110000111	00--000011	11111110200	00010-0000	1100011112	1001111011	0100111000	2001120110	200-100000	1000210011	0112100011	0000000000	0001000000	1000
Paedocypris sp.	1000-00000	11-1-00-11	020011011-	--100-0000	00000--000	110-011112	---0000000	0011101000	2001111102	010-000-10	10-1201000	011010---0	1111111111	1111111111	1111
Danionella dracula	0000-00000	11-1-00-11	020011011-	--100-0000	0--00--000	?10-011112	---0000000	0011101000	2001111102	010-000-10	00-120-00-	011010---0	1111111111	1111111111	1111
Sundadanio echinus	1000-00000	0111-00-11	020110011-	01111-0000	00000--000	1100001112	1000000000	0000101000	2001100100	0000000010	0001201000	0112X0---0	1110111001	1101000000	1111
Danio rerio	1100000000	0110000-00	1201100110	0011110000	0000000000	1100001112	1000000000	0000101000	2101110100	0000000010	0000100000	0112100000	0000000000	0000000000	0000
Danio erythromicron	1100000000	0110-00-01	120111101-	0011110000	000000-000	1100001112	1000000000	0000101000	2101111000	0000000010	0000100000	011210-0-0	0000010000	1001000000	0000
Danio margaritatus	1100000000	0110-00-01	120111101-	0011110000	000000-000	1100001112	1000000000	0000101000	2101111000	0000000010	0000100000	011210-0-0	1000010000	1001000000	0000
Boraras brigittae	1100000000	0110-00-01	120111101-	0-11110000	00000--000	1100001112	1000000000	0000101000	2101100100	0000000010	0000100000	011210-0-0	1000011000	1001000000	0000
Boraras merah	1100000000	0110-00-01	120111101-	0-11110000	00000--000	1100001112	1000000000	0000101000	2101100100	0000000010	0000100000	011210-0-0	1000011000	0001000000	0000
Boraras maculatus	1100000000	0110-00-01	120111101-	0-11110000	00000-000	1100001112	1000000000	0000101000	2101100100	0000000010	0000100000	011210-0-0	1000010000	0001000000	0000
Boraras micros	1000000000	0110-00-01	120111101-	0-11110000	00000--000	1100001112	1000000000	0010101000	2101101000	0000000010	0000100000	011210-0-0	1000011000	1101000000	0000
Boraras urophthalmoides	1100000000	0110-00-01	120111101-	0-11110000	00000--000	1100001112	1000000000	0010101000	2101101000	0000000010	0000100000	011210-0-0	1000011000	1001000000	0000
Horadandia atukorali	1000000000	0111-10-00	120110011-	0-11110000	00000--000	1100001112	1000000000	0000101000	2101101000	0000000010	0000100000	011210-0-0	1000010000	0001000000	0000
Microrasbora rubescens	1100000000	0110-00-01	120110011-	0111110000	00000--000	1100001112	1000000000	0000101000	2101111100	0010000010	0000100000	011210-0-0	0000010000	0001000000	0000
Microdevario kubotai	1100000000	0110100-01	1201100110	0111110000	000000-000	1100001112	1000000000	0010101000	2101111000	0000000011	0000100000	011210-0-0	0000010000	1001000000	0000
Microdevario nana	1000000000	0110-00-01	120111011-	0-11110000	00000--000	1100001112	1000000000	0000101000	2101111000	0000000011	0000101000	011210-0-0	0000010000	1001000000	0000
Sawbwa resplendens	1120000000	0111-00-00	120111011-	--11110000	00000--000	1100001112	1000000000	0000101000	2101100100	0000000010	0000101000	111200---0	1000010100	0101000100	0000
Tanichthys micagemmae	1100000000	0110-00-01	120111011-	--11110000	00000--000	1100001112	1000000000	0010101000	2101100100	0010000010	0000100000	0112100000	1000011000	1001000000	0000
Rasbora kalbarensis	1100000000	0110-00-00	120110011-	--11110000	00000--000	1100001112	1000000000	0000101000	2101100100	0000000010	0000100000	011210---0	0000010000	0000000000	0000
Rasboroides vaterifloris	1100000010	0111100-01	1201100110	0-11110000	0000000000	1100001112	1000000000	0000101000	2101100100	0000000010	0000100000	0112000000	0000010000	0001000000	0000
Trigonostigma hengeli	1100000010	0110100-01	1201100110	0011110000	0000000000	1100001112	1000000000	0000101000	2101100100	0000000010	0000100000	0112000000	0000010000	0000000000	0000

APPENDIX 6

CHARACTER MATRIX 5

133 characters from Conway (2011; characters 1–117) and Britz & Conway (2009; 118–133) for 65 taxa coded based on examination of vouchered material listed in Conway (2011) and Britz & Conway (2009). "?" represents missing data. "-" represents inapplicable character coding. "x" represents polymorphic character coding. *Lepidocephalichthys* cf. *guntea* is abbreviated as *Lepidoceph*. cf. *guntea*; *Acanthocobitis* cf. *zonalternans* is abbreviated as *Acantho*. cf. *zonalternans*.

	0	0	0	0	0	0	0	0	0	1	1	1	1	
	1	2	3	4	5	6	7	8	9	0	1	2	3	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	123
Elops sp. cf. *senegalensis*	0000001000	0000000100	0--0000000	0000000000	0000000000	0000000000	0000000000	0000000000	0000--0-00	0-00000--0	0000000000	00000-0000	000-00000-	-0-
Hiodon alosoides	0000000000	0000100010	0--0000100	-0----0000	0000000000	0000000000	0000000000	0000000000	0000--0-00	0-01000--0	0100000001	00000-0000	000-00000-	-0-
Denticeps clupeoides	0000000000	0010000000	0--0000000	1000000000	0000000000	0000010121	0000000000	0000000000	0000--0-00	0-01000--0	0101000000	000?0-0000	000-00000-	-0-
Distichodus antonii	0000000001	0101000000	0--0001000	1100000001	0010000000	0000010020	0000000000	0000000000	0000000010	0-00000000	0000000010	1001000000	0000000000	000
Psilorhynchus balitora	1100000000	0111000001	2011001100	1111110000	0010-10101	1000111121	0000101000	1010100020	0110102200	-000010000	1200?01011	2001000000	0000000011	000
Cyprinus carpio	1120000011	0111010001	2011001100	1111100000	0000100001	1000011121	0000001000	0010100021	0110100000	0000010000	0100000111	2000000000	0000000000	000
Barbus barbus	1100000010	0110010001	2011001100	1111100000	0000100001	1000011121	0000001000	0010100021	0110100000	0000010000	0100000111	2000000000	0000000000	000
Balantiocheilos melanopterus	1100000010	0111010001	2011001100	1111100000	0000100001	1000011121	0000001000	0010100021	0110100000	0000010000	0100000111	2000000000	0000000000	000
Garra dembeensis	1100001011	0110010001	2021001100	1111100000	0100100101	1000011121	0000001010	0010110021	0110100000	-000010000	0200001011	2000000000	0000000000	000
Osteochilus hasselti	1100000011	0110010001	2021001100	1111100000	0100100101	1000011121	0000001110	0010110121	0110100000	-000010000	0100000011	2000000000	0000000000	000
Labeo longipinnis	1100001011	0110010001	2021001100	1111100000	0100100101	1000011121	0000001110	0010110121	0110100000	-000010000	0100000011	2000000000	0000000000	000
Lobocheilos bo	1100000011	0110010001	2021001100	1111100000	0100100101	1000011121	0000001010	0010110121	0110100000	0000010000	0100000011	2000000000	0000000000	000
Crossocheilus reticulatus	1100000011	0111010001	2021001100	1111100000	0110100101	1000011121	0000001010	0010110121	0110100000	-000010000	0100000011	2000000000	0000000000	000
Gobio gobio	1100000000	0110100001	2011001100	1111100000	0000100001	1000011121	0000101000	0010100021	0110100001	1000010000	0100000011	2001000000	0000000000	000
Abramis brama	1100000000	0110100001	2011001100	1111100000	0000100001	1000011121	0000001000	0010100021	0110100001	1000010000	0100000011	2000000000	0000000000	000
Zacco cf. *platypus*	1100000000	0110100001	2011001100	0111100000	0000000001	1000011121	0000001000	0010100021	0110100001	1000010000	0100000011	2001000000	0000000000	000
Rhodeus sericeus	1100000001	0110100-01	2011001100	1111100000	0000-00001	1000011121	0000001000	0010100021	0110100001	1000010000	0100000011	2001000000	0000000000	000
Notemigonus crysoleucas	1100000000	0110100001	2011001100	1111100000	0000100001	1000011121	0000001000	0010100021	0110100001	1000010000	0100000011	2000000000	0000000000	000
Rasbora cephalotaenia	1100000000	0110000001	2011001100	0111100000	0000100001	1000011121	0000001000	0010100021	0110100001	1000010000	0100000011	2000000000	0000000000	000
Phoxinus phoxinus	1100000000	0110100001	2011001100	1111100000	0000-00001	1000011121	0000001000	0010100021	0110100001	1000010000	0100000011	2010000000	0000000000	000
Gyrinocheilus pennocki	1100001001	0100001100	1000001101	1101110000	0000-11000	1000111111	0000001100	0100-00000	1110100000	0110001100	0100100011	2000000000	0000000001	000
Gyrinocheilus aymonieri	1100001001	0110001100	1000001101	1101110000	0000-11000	1000111111	0000001100	0100-00000	1110100000	0110001100	0100100011	2000000000	0000000001	000
Myxocyprinus asiaticus	1110001010	0101001100	2100001111	1111101100	1100-10001	1000111121	0100110000	0110100010	0111100100	0010000000	0100000011	2100000000	0000000000	000
Cycleptus elongates	1110001000	0110101100	2100001111	1111101100	1100-10001	1000111121	0100110000	0110100010	0111100100	0010000000	0100000011	2100000000	0000000000	000
Catostomus commersoni	1110001000	0111001-00	21001011-1	-111101100	1100--0011	1000111121	0100110000	0110100010	0111100100	0010000000	0100000011	2100000000	0000000000	000
Carpiodes carpio	1110001010	0101101100	2100001111	1111101100	1100-10011	1000111121	0100110000	0110100010	0111100100	0010001000	0100100011	2100000000	0000000000	000
Syncrossus cf. berdmorei	1021110110	0111000100	1000011111	-101102200	0010000001	1000011121	0000110110	0011100020	0111101010	0110100100	0100010011	2010100000	0000000000	000
Chromobotia macracanthus	1021110110	0111001100	1000011111	-101102200	0010000001	1000011121	0000110110	1011100021	0111101010	0110100100	0100010011	2010100000	0000000000	000
Yasuhikotakia modesta	1021111111	0111000100	1000011111	-101102200	0010000001	100001112-	--00110110	0011100021	0111101010	0110100100	0100010011	2010100000	0000000000	000
Yasuhikotakia sidthimunki	1021110110	0111000100	1000001111	-101102200	0010000001	100001112-	--00110100	0011100020	0111100010	0100100100	0100010011	201?100000	0000000000	000

APPENDIX 6 *Continued*

	0	0	0	0	0	0	0	0	0	1	1	1	1
	1	2	3	4	5	6	7	8	9	0	1	2	3
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	123
Vaillantella euepiptera	1021111000	0110000100	1000000111	-101102000	0010000001	1000011122-	--00111100	0011100020	0110100010	0100100100	0200010011	201?100000	000
Cobitis pacifica	1021110100	1111-00-00	10001011--	-101100011	0011--0001	1100111121	0010110111	1011101020	0110100000	-101000111	0201010011	3010100001	000
Cobitis cf. *taenia*	1021110100	1111-00-00	10001011--	-101100011	0011--0001	1100111121	0010110111	1011100020	0110100000	-101000111	0201010011	3010100001	000
Iksookimia koreensis	1021110100	1111-00-00	10001011--	-101100011	0011--0001	1100111121	0010110111	1011101020	0110100000	-101000111	0201010011	3010100001	000
Niwaella multifasciata	1021110100	1111-00-00	10001011--	-101100011	0011--0001	1100111121	0010110111	1011101020	0110100000	-101000111	0201010011	3010100001	000
Acantopsis choirorhynchus	1021110100	1111-00100	10001011-1	1101100011	0011--0001	1100111121	0010110111	1011101020	0110100000	-101000110	0201010011	3010100000	000
Lepidoceph. cf. *guntea*	1021110100	1111-00-00	10001011--	1101100011	0011--0001	1100111121	0010110111	1011101020	0110100000	-101000111	0201010011	3010-00001	000
Barbatula barbatula	1120010000	0111100100	2000001111	-111102000	0010-00001	1000111122-	--11100110	1011100020	0110110200	-100001100	0201010011	2010110000	000
Mesonoemacheilus triangularis	1020111000	0111000100	2000001111	-111102200	0010-00001	1000011121	0011110110	1011100020	0110110200	0100101100	0200010011	2010110000	000
Lefua costata	1120011000	0111-00-00	20001011--	-111102000	0010--0001	1000111121	0010110110	1011100020	0110110210	-100001100	0200011001-	201?110000	000
Acantho. cf. *zonalternans*	1020110000	0111000100	2000001111	-111102000	0010-00001	1000011121	0011110110	1011100020	0110110200	0100101100	0100010011	2010110000	000
Annamia normani	1120011001	0111000110	2020000111	-111100000	0010-00001	1011111121	1010110110	1001100020	0110110201	-100000100	0210011011	301?110000	000
Gastromyzon cranbrooki	1120011000	0111000110	2010000111	-111100000	0010-10001	1011111121	101-110110	1001100020	0110110201	-100000100	0110011011	2010110000	000
Homaloptera stephensoni	1120111000	0111000110	0--00000111	1111102000	0010-00001	1000111121	0010110110	1011100020	0110110200	?100000100	0210011011	1010110000	000
Homaloptera zollingeri	1120011000	0110000110	0--00000111	1111102000	0010-00001	1000111121	0011110110	1011100020	0110110200	?100000100	0210011011	2000110000	000
Paedocypris sp.	1000-00000	11-1-00-10	20011011--	-100-00000	0000--0001	10-0111122-	--00000000	0110100020	0110102010	-000-1010-	1201000011	00---01111	111
Danionella dracula	0000-00000	11-1-00-10	20011011--	-100-00000	--00--000?	10-0111122-	--00000000	0110100020	0110102010	-000-1000-	120-00-011	00---01111	111
Sundadanio echinus	1000-00000	0111-00-10	20110011-0	1111-00000	0000--0001	1000111121	0000000000	0010100020	0110100000	0000010000	1201000011	20---01101	111
Danio rerio	1100000000	0110000-01	20110011--	0111100000	0000000001	1000011121	0000000000	0010100021	0110100000	0000010000	0100000011	2000000000	000
Danio erythromicron	1100000000	0110-00-01	20111011-0	0111100000	00000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Danio margaritatus	1100000000	0110-00-01	20111011-0	0111100000	00000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Boraras brigittae	1100000000	0110-00-01	20111011--	-111100000	0000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Boraras merah	1100000000	0110-00-01	20111011--	-111100000	0000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Boraras maculatus	1100000000	0110-00-01	20111011--	-111100000	0000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Boraras micros	1000000000	0110-00-01	20111011-0	-111100000	0000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Boraras urophthalmoides	1100000000	0110-00-01	20111011--	-111100000	0000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Horadandia atukorali	1000000000	0111-10-01	20110011-0	-111100000	0000--0001	1000011121	0000000000	0010100021	0110100000	0000010000	0100000011	20-0-00000	000
Microrasbora rubescens	1100000000	0110-00-01	20110011-0	1111100000	0000--0001	1000011121	0000000000	0010100021	0110100001	0000010000	0100000011	20-0-00000	000
Microdevario kubotai	1100000000	0110100-01	20110011--	1111100000	00000--0001	1000011121	0000000000	0010100021	0110100000	0000011000	0100000011	20-0-00000	000
Microdevario nana	1000000000	0110-00-01	20110011-0	-111100000	0000--0001	1000011121	0000000000	0010100021	0110100000	0000011000	0101000011	20-0-00000	000
Sawbwa resplendens	1120000000	0111-00-01	20111011--	-111100000	0000--0001	1000111121	0000000000	0010100021	0110100000	0000010000	0101000111	20---00000	000
Tanichthys micagemmae	1100000000	0110-00-01	20111011--	-111100000	0000--0001	1000111121	0000000000	0010100021	0110100001	0000010000	0100000011	2000000000	000
Rasbora kalbarensis	1100000000	0110-00-01	20110011--	-111100000	0000--0001	1000011121	0000000000	0010100021	0110100000	0000010000	0100000011	20---00000	000
Rasboroides vaterifloris	1100000010	0111100-01	2011001100	-111100000	0000000001	1000011121	0000000000	0010100021	0110100000	0000010000	0100000011	2000000000	000
Trigonostigma hengeli	1100000010	0110100-01	2011001100	0111100000	0000000001	1000011121	0000000000	0010100021	0110100000	0000010000	0100000011	2000000000	000

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Figure S1. Morphological character distributions for analysis 2.
Figure S2. Morphological character distributions for analysis 3.
Figure S3. Morphological character distributions for analysis 4.
Figure S4. Morphological character distributions for analysis 5.
Figure S5. Single-gene maximum-likelihood analyses of the Mayden and Chen (2010) data set, 50% majority rule bootstrap tree shown.
Figure S6. Single-gene maximum-likelihood analyses of the Mayden and Chen (2010) data set, 50% majority rule bootstrap tree shown.
Figure S7. Maximum likelihood (ML) tree of the combined Mayden & Chen (2010) data set comprising 4827 bp of five nuclear genes (*EGR1*, *EGR2B*, *IRBP*, *RAG1*, and *RH*).