Data Article

Data on the multilocus molecular phylogenies of the Neotropical fish family Prochilodontidae (Teleostei: Characiformes)

Benjamin W. Frable a,b,*, Bruno F. Melo c,d, Brian L. Sidlauskas a,d, Kendra Hoekzema a, Richard P. Vari d, Claudio Oliveira c

a Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR, USA
b Marine Vertebrate Collection, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA
c Departamento de Morfologia, Instituto de Biociências, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil
d Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

ARTICLE INFO

Article history:
Received 2 June 2016
Received in revised form 1 August 2016
Accepted 8 August 2016
Available online 18 August 2016

Keywords:
Phylogenetics
Prochilodontidae
Characiformes
*BEAST
Shimodaira–Hasegawa test

ABSTRACT

The data presented herein support the article "Molecular phylogenetics of the Neotropical fish family Prochilodontidae (Teleostei: Characiformes)" (B.F. Melo, B.L. Sidlauskas, B.W. Frable, K. Hoekzema, R.P. Vari, C. Oliveira, 2016) [1], which inferred phylogenetic relationships of the prochilodontids from an alignment of three mitochondrial and three nuclear loci (5279 bp) for all 21 recognized prochilodontid species and 22 related species. Herein, we provide primer sequences, museum voucher information and GenBank accession numbers. Additionally, we more fully describe the maximum-likelihood and Bayesian phylogenetic analyses of the concatenated dataset, detail the Bayesian species tree analysis, and provide the maximum likelihood topologies congruent with
prior morphological hypotheses that were compared with the unconstrained tree using Shimodaira–Hasegawa tests.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

**Specifications Table**

| Subject area      | Biology, Genetics and Genomics                          |
|-------------------|---------------------------------------------------------|
| More specific subject area | Phylogenetics and Phylogenomics                       |
| Type of data      | Tables, figures, primers, sequence alignment, museum voucher information, phylogenetic trees |
| How data was acquired | DNA extraction from tissue samples, gene amplification, Sanger sequencing |
| Data format       | Raw, filtered, analyzed                                 |
| Experimental factors | DNA extraction from muscle or fin tissue using Quiagen DNeasy kit or modified NaCl protocol |
| Experimental features | Sequences concatenated and aligned in Geneious (v.7.1.7), phylogenies generated using unconstrained and constrained maximum-likelihood (RAxML), concatenated Bayesian (MrBayes), and Bayesian species tree (*BEAST) methods. |
| Data source location | South America                                           |
| Data accessibility | Data provided with this article and in the GenBank public repository, GenBank: KX086740 through GenBank: KX087100 (see Table 2) and 16S: http://www.ncbi.nlm.nih.gov/popset/1021206184  
COI: http://www.ncbi.nlm.nih.gov/popset/1021205438  
Cytb: http://www.ncbi.nlm.nih.gov/popset/1021205579  
Myh6: http://www.ncbi.nlm.nih.gov/popset/1021205738  
Rag1: http://www.ncbi.nlm.nih.gov/popset/1021205893  
Rag2: http://www.ncbi.nlm.nih.gov/popset/1021206027 |

**Value of the data**

- New sequence data were used to infer the first complete molecular phylogenetic analysis of family Prochilodontidae.
- Dataset includes DNA sequences for all 21 valid prochilodontid species and 22 related characiform species, many of which are not otherwise represented in Genbank.
- These data facilitate synthesis with previously published sequences and can be reused in other studies because the loci are commonly used in fish phylogenetics.
- Constrained phylogenies permit statistical comparison of new molecular results with prior morphological hypotheses.

**1. Data**

We provide: 1) A table documenting the deposition of museum voucher specimens, 2) aa file containing concatenated alignments for all six loci, 3) a table containing GenBank accession numbers, 4) procedures, parameters and configuration scripts used to estimate phylogenetic relationships, 5) Newick-formatted treefiles inferred with maximum likelihood, concatenated Bayesian, and species tree methods, 6) Newick-formatted treefiles and PDF images of maximum likelihood phylogenies.
inferred under four topological constraints matching the morphological phylogeny of Castro and Vari [2], and 7) procedures used in Shimodaira–Hasegawa tests of alternative topologies.

2. Experimental design, materials and methods

2.1. Taxon sampling

This dataset included samples from 77 individuals: 55 individuals representing all 21 species of the three prochilodontid genera, and samples from 22 related taxa from the other three anostomoid families (Anostomidae, Chilodontidae, Curimatidae), three families previously hypothesized to be closely related to Anostomoidea (Hemiodontidae, Parodontidae and Serrasalmidae), and Brycon pesu (Bryconidae), as an outgroup. Nine of the samples were derived from previous studies [3–5], and thus 88% of these data are new to science. We used tissue samples stored in 95% ethanol or a saturated DMSO/NaCl solution, primarily from specimens deposited in museum and university collections (see Table 1 in Melo et al. [1]). We included multiple individuals for each prochilodontid species except Ichthyoloelephas longirostris, which is exceedingly rare in tissue collections. The authors BFM, BLS and RPV confirmed the taxonomic identity of most voucher specimens using morphological features.

2.2. Molecular dataset

We extracted genomic DNA using DNeasy Tissue kits (Qiagen Inc.) or a modified NaCl protocol from Lopera-Barrero et al. [6]. For this dataset, we amplified partial sequences of the mitochondrial

| Locus | Bp after alignment | PCR | Primer sequence (5’–3’) | ΠA | ΠC | ΠG | ΠT | Reference |
|-------|-------------------|-----|------------------------|----|----|----|----|-----------|
| 16S   | 510 bp            | 1   | 16Sa-L – ACCGTCTGTTTATCAAAAACAT 16Sb-H – CCGGTCTGAATCTCAGATCGT | 0.296 | 0.239 | 0.236 | 0.229 | [22] |
| COI   | 656 bp            | 1   | L6252-Asn – AAGCCGGGGAAAAGGCCCCGCAG H7271-COXI – TCTATATAGCCGAATGTTCTTT | 0.242 | 0.278 | 0.187 | 0.293 | [23] |
| Cytb  | 990 bp            | 1   | LNF – GACTTGAAAAACAYCGTTGT H08R2– GCTTTGGAATGAGGTGGAGTTAAAT | 0.269 | 0.310 | 0.146 | 0.275 | [4] |
| Myh6  | 710 bp            | 1st | F329 – CCGCMTGGCGATGTCTACACA 3R1 – ATTTCCACCACTCGATGTGA | 0.302 | 0.229 | 0.231 | 0.239 | [24] |
|       |                   | 2nd | A3F2 – GGAGATCATCTCGTCTCACA A3R2 – GTCACCACCACCGTGAAGAT | | | | |
| Rag1  | 1378 bp           | 1st | Rag1CF1 – ACCCTCCTGTACTGCTGAGAA 164F – AGCTCAAGCGGCCGCGCCAT | 0.250 | 0.239 | 0.287 | 0.224 | [4] |
|       |                   | 2nd | Rag1CR1 – CTGCGCGAGGACGCTGTTGCC Rag1Ri – GGGGATTAGTTGGGAGTTGAGATTAGAAT | 0.242 | 0.259 | 0.273 | 0.225 | [4,25] |
| Rag2  | 1029 bp           | 1st | 164F – AGCTCAAGCGGCCGCGCCAT 176F – GYGGCATTCTCATTCTCAACA | | | | |
Table 2
Specimens and loci used in Melo et al. [1]. For each individual, its taxonomic designation, collection catalog number of voucher, tissue specimen number, and GenBank accession numbers are given (GenBank:KX086740 through GenBank:KX087100).

| Taxon                  | Voucher   | Specimen | 16S | Co1 | Cytb | Myh6 | Rag1 | Rag2 |
|------------------------|-----------|----------|-----|-----|------|------|------|------|
| Ichthyoelephas humeralis | LBP 19,326 | 76,121   |     |     |      |      |     |      |
| Ichthyoelephas humeralis | ANSP 192,865 | 76,122   |     |     |      |      |     |      |
| Ichthyoelephas longirostris | LBP 251   | 4216     | KX087085 | KX086742 | KX086841 | KX086866 | KX086949 | KX087006 |
| Prochilodus argenteus    | LBP 251   | 4217     | KX087086 | KX086743 | KX086842 | KX086867 | KX086949 | KX087006 |
| Prochilodus argenteus    | LBP 2496  | 16,385   | KX087087 | KX086759 | KX086829 | KX086885 | KX086937 | KX086995 |
| Prochilodus brevis       | LBP 20,269 | 79,757   | KX087071 | KX086763 | KX086832 | KX086886 | KX086938 | KX087015 |
| Prochilodus costatus     | LBP 252   | 4222     | KX087079 | KX086744 | KX086821 | KX086868 | KX086950 | KX087009 |
| Prochilodus harttii      | LBP 7211  | 33,175   | KX087098 | KX086785 | KX086843 | KX086892 | KX086944 | KX087005 |
| Prochilodus lacustris    | LBP 9104  | 42,732   | KX087096 | KX086831 | KX086898 | KX086906 | KX087017 |
| Prochilodus lineatus     | LBP 45    | 3611     | KX087081 | KX086741 | KX086819 | KX086865 | KX087007 |
| Prochilodus nigricans    | LBP 2188  | 15,561   | KX087077 | KX086755 | KX086839 | KX086881 | KX086931 | KX087001 |
| Prochilodus nigricans    | LBP 1690  | 15,562   | KX087078 | KX086756 | KX086840 | KX086842 | KX086932 | KX087005 |
| Prochilodus nigricans    | LBP 7841  | 36,858   | KX087088 | KX086767 | KX086835 | KX086945 | KX087016 |
| Prochilodus nigricans    | LBP 8589  | 43,397   | KX087084 | KX086771 | KX086837 | KX086999 | KX087003 |
| Prochilodus nigricans    | LBP 12,865 | 53,496   | KX087090 | KX086774 | KX086836 | KX086902 | KX086955 | KX087014 |
| Prochilodus nigricans    | OS 18,792 | PE10045  | KX087093 | KX086787 | KX086827 | KX086913 | KX086966 | KX087000 |
| Prochilodus nigricans    | OS 18,792 | PE10058  | KX087094 | KX086788 | KX086824 | KX086914 | KX087001 |
| Prochilodus nigricans    | FMNH 113,534 | 29,513   | KX087095 | KX086797 | KX086828 | KX086925 | KX086974 | KX087020 |
| Prochilodus nigricans    | LBP 6127  | 29,513   | KX087099 | KX086764 | KX086816 | KX086891 | KX086943 | KX087021 |
| Prochilodus nigricans    | LBP 6127  | 29,514   | KX087102 | KX086764 | KX086816 | KX086891 | KX086943 | KX087021 |
| Prochilodus nigricans    | ANSP 40,692 | 43,398   | KX087090 | KX086784 | KX086834 | KX086910 | KX086963 | KX087002 |
| Prochilodus rubrotaeniatus | MHNG     | 2705.008 | KX087091 | KX086775 | KX086825 | KX086903 | KX086933 | KX087010 |
| Taxon                        | Voucher   | Specimen | 16S       | Co1       | Cytb      | Myh6      | Rag1      | Rag2      |
|-----------------------------|-----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Prochilodus rubrotaeniatus  | MHNG 2717.017 | SU08776  | KX087097  | KX086776  | KX086826  | KX086904  | KX086934  | KX087011  |
| Prochilodus rubrotaeniatus  | USNM 403.693 | GY11461  | KX087083  | KX086782  | KX086833  | KX086908  | KX086935  | KX087008  |
| Prochilodus vimboides       | LBP 2349  | 16,011   | KX087075  | KX086757  | KX086814  | KX086883  | KX086936  | KX086997  |
| Prochilodus vimboides       | LBP 10,180 | 47,662   | KX087074  | KX086773  | KX086815  | KX086901  | KX086954  | KX086998  |
| Semaprochilodus brama       | LBP 12,776 | 41,019   | KX087069  | KX086769  | KX086856  | KX086895  | KX086947  | KX087029  |
| Semaprochilodus brama       | LBP 12,807 | 41,171   | KX087070  | KX086770  | KX086857  | KX086896  | KX086948  | KX087031  |
| Semaprochilodus insignis    | LBP 1692  | 12,761   | KX087063  | KX086753  | KX086850  | KX086879  | –          | KX087032  |
| Semaprochilodus insignis    | LBP 1692  | 12,762   | KX087064  | KX086754  | KX086849  | KX086880  | KX086929  | –          |
| Semaprochilodus insignis    | OS 18,380 | PE10001  | KX087067  | KX086785  | KX086851  | KX086911  | KX086964  | KX087033  |
| Semaprochilodus insignis    | ANSP 180,205 | T43   | KX087061  | KX086796  | KX086852  | KX086923  | KX086973  | KX087034  |
| Semaprochilodus kneri       | LBP 1384  | 12,734   | KX087062  | –         | KX086845  | KX086874  | KX086928  | KX087035  |
| Semaprochilodus kneri       | LBP 3041  | 19,140   | KX087066  | –         | KX086846  | KX086888  | KX086941  | KX087036  |
| Semaprochilodus kneri       | LBP 3041  | 19,140   | KX087066  | KX086762  | KX086848  | KX086889  | –          | –          |
| Semaprochilodus kneri       | ANSP 187,277 | P4298 | KX087060  | KX086783  | KX086847  | KX086909  | KX086962  | KX087037  |
| Semaprochilodus laticeps    | LBP 1383  | 12,727   | KX087059  | KX086748  | KX086861  | KX086873  | KX086927  | –          |
| Semaprochilodus laticeps    | LBP 1383  | 12,728   | HQ171245  | KF562436  | HQ289536  | HQ288955  | HQ289152  | HQ289343  |
| Semaprochilodus laticeps    | FMNH 113,712 | 2004BSAQ01 | KX087068  | KX086778  | KX086860  | KX086906  | KX086942  | KX087030  |
| Semaprochilodus taeniurus   | LBP 1691  | 12,757   | KX087051  | KX086750  | KX086854  | KX086876  | –          | KX087025  |
| Semaprochilodus taeniurus   | LBP 1691  | 12,758   | KX087050  | KX086751  | KX086853  | KX086877  | –          | KX087024  |
| Semaprochilodus taeniurus   | LBP 1691  | 12,759   | KX087052  | KX086752  | KX086855  | KX086878  | –          | KX087026  |
| Semaprochilodus varii       | MHNG uncatalogued | 15,729 | KX087058  | KX086777  | KX086859  | KX086905  | KX086930  | KX087027  |
| Semaprochilodus varii       | ANSP 187,435 | 6929 | KX087057  | KX086746  | KX086858  | KX086871  | KX086957  | KX087028  |
| Leporellus cf. vittatus     | AUM 54,212 | T09912   | –         | KX086795  | KX086801  | KX086921  | KX086972  | KX086987  |
| Leporinus desmotes          | AUM 43,700 | V5274    | KX087040  | KX086798  | KX086813  | KX086926  | KX086975  | KX086986  |
| Leporinus friderici         | ANSP 189,264 | 7015 | KX087039  | KX086747  | KX086812  | KX086872  | KX086958  | KX086985  |
| Leporinus striatus          | LBP 3180  | 16,871   | KX087045  | KX086793  | KX086808  | KX086919  | KX086970  | KX086981  |
| Abramites hypselonotus      | AUM 53,775 | T08985   | KX087045  | KX086793  | KX086808  | KX086919  | KX086970  | KX086981  |
| Schizodon scotorhabdotus    | AUM 53,654 | T09707   | KX087047  | KX086794  | KX086810  | KX086920  | KX086971  | KX086984  |
| Chilodus fritillus          | AUM 51,355 | T10201   | KF562391  | KF562418  | KX086863  | KX086922  | KF562495  | KX086988  |
| Caenotropus mes tomorgmatos | ANSP 180,516 | T48    | KF562384  | KF562412  | KF562442  | KX086924  | KF562490  | KX086991  |
| Curimatopsis macrolepis     | ANSP 178,188 | 1697 | KX087053  | KX086740  | KX086800  | KX086864  | KX086940  | KX086977  |
| Curimata cyprinoides        | USNM 402,471 | 411-1-03 | KX087054  | KX086781  | KX086803  | KX086907  | KX086961  | KX086978  |
genes 16S rRNA (16S, 510 bp), cytochrome oxidase C subunit 1 (COI, 658 bp) and cytochrome B (Cytb, 991 bp) using one round of polymerase chain reaction (PCR). Additionally, we acquired sequences of the nuclear myosin heavy chain 6 gene (Myh6, 711 bp), recombination activating gene 1 (Rag1, 1379 bp), and recombination activating gene 2 (Rag2, 1030 bp) using nested-PCR following Oliveira et al. [3].

Table 2 (continued)

| Taxon                         | Voucher | Specimen | 16S         | COI         | Cytb       | Myh6       | Rag1        | Rag2        |
|-------------------------------|---------|----------|-------------|-------------|------------|------------|-------------|-------------|
| Psectrogaster amazonica      | OS 18,313 | PE10113 | KX087049    | KX086792    | KX086802   | KX086918   | KX086969    | KX086990    |
| Cyphocharax gilbert          | LBP 8343 | 40,130   | KX087056    | KX086768    | KX086805   | KX086894   | KX086946    | KX086989    |
| Cyphocharax spliatus         | LBP 4747 | 25,521   | KX087055    | KX086763    | KX086804   | –          | –           | –           |
| Anodus elongatus             | OS 18,724 | PE10110 | KX087043    | KX086791    | KX086806   | KX086917   | –           | KX086983    |
| Hemiodus unimaculatus        | OS18345  | PE10076  | KX087042    | KX086790    | KX086807   | KX086916   | KX086968    | KX086980    |
| Apareiodon affinis           | LBP 4591 | 24,665   | HQ171328    | –           | HQ289617   | HQ289837   | HQ289230    | HQ289424    |
| Parodon nasus                | LBP 1135 | 5635     | HQ171429    | –           | HQ289714   | HQ289137   | HQ289328    | HQ289521    |
| Colossoma macropomum         | LBP 5173 | 26,648   | HQ171343    | –           | HQ289632   | HQ289052   | HQ289245    | HQ289438    |
| Catoprion mento              | LBP 7556 | 35,624   | HQ171392    | –           | HQ289679   | HQ289100   | HQ289293    | –           |
| Metyynis lippincottianus     | LBP 6282 | 29,688   | HQ087041    | –           | HQ289651   | HQ289072   | HQ289265    | HQ289458    |
| Myleus schomburgii           | OS 18,990 | PE10044 | KX087046    | KX086786    | KX086862   | KX086912   | KX086965    | KX086979    |
| Brycon pesu                  | OS 18,361 | PE10072 | KX087038    | KX086789    | KX086799   | KX086915   | KX086967    | KX086976    |

Table 3

Position of each gene and codon within the alignment, with their partitions and best models of nucleotide evolution as determined by PartitionFinder.

| Gene               | Position | Partition | Best BIC model for MrBayes |
|--------------------|----------|-----------|----------------------------|
| 16S                | 1–510    | 1         | SYM+I+G                    |
| COI 1st position   | 511–1167/3 | 2         | GTR+G                      |
| COI 2nd position   | 512–1167/3 | 1         | SYM+I+G                    |
| COI 3rd position   | 513–1167/3 | 3         | HKY+I+G                    |
| Cytb 1st position  | 1169–2158/3 | 4         | GTR+G                      |
| Cytb 2nd position  | 1170–2158/3 | 1         | SYM+I+G                    |
| Cytb 3rd position  | 1168–2158/3 | 3         | HKY+I+G                    |
| Myh6 1st position  | 2160–2869/3 | 6         | HKY+I+G                    |
| Myh6 2nd position  | 2161–2869/3 | 6         | HKY+I+G                    |
| Myh6 3rd position  | 2159–2869/3 | 5         | SYM+G                      |
| Rag1 1st position  | 2871–4248/3 | 6         | HKY+I+G                    |
| Rag1 2nd position  | 2872–4248/3 | 6         | HKY+I+G                    |
| Rag1 3rd position  | 2870–4248/3 | 5         | SYM+G                      |
| Rag2 1st position  | 4249–5278/3 | 6         | HKY+I+G                    |
| Rag2 2nd position  | 4250–5278/3 | 6         | HKY+I+G                    |
| Rag2 3rd position  | 4251–5278/3 | 5         | SYM+G                      |

Amplification techniques and sequencing reactions are detailed in Melo et al. [1]. We amplified and included all six loci for 42 (of 77) individuals. In the rest of the matrix, we are missing one locus for 22 individuals, two loci for nine individuals, four for one individual and five for three specimens (both specimens of Ichthyoelephas humeralis and one of Prochilodus britskii; see Table 2). New sequences generated in this analysis were deposited in GenBank with accession numbers KX086740.
through KX087100. The precise matches of sequence accession numbers to gene and voucher appear in Table 2.

2.3. Alignment, partitioning, and model selection

We aligned and edited sequences using Geneious 7.1.7 ([7]; www.geneious.com). We assigned IUPAC ambiguity codes where we detected uncertainty of nucleotide identity. We performed the alignment of consensus sequences for each gene with the Muscle algorithm [8] implemented in Geneious using default parameters and inspected the sequences visually for obvious misalignments. We estimated the index of substitution saturation (Iss) using Dambe 5.3.38 [9] to evaluate the occurrence of substitution saturation. We found no indication of substitution saturation in transitions or transversions in any topologies. Initial examination of the complete 16S data revealed many uncertain alignments from length polymorphism in loop regions. We excluded these hypervariable regions in a reduced 16S submatrix that was in turn concatenated with the other five genes. The final concatenated dataset for all the sampled taxa is 5279 bp long with 8.9% missing data, 944 (17.9%) identical sites and 1463 of 1970 variable sites being parsimony-informative (matrix file Prochilodontidae_matrix.nex). Nucleotide frequencies are presented in Table 1.

We used PartitionFinder 1.1.0 [10] to select the partitioning scheme and the model molecular evolution for each partition in the scheme using the Bayesian information criterion (BIC). For this analysis, we assumed 16 possible partitions (Table 3), one for each codon position in the five coding genes (COI, Cytb, Myh6, Rag1 and Rag2), plus the 16S stems. Results identified six partitions with models summarized in Table 3.

2.4. Concatenated analyses

We analyzed the partitioned matrix using the Bayesian methods in MrBayes 3.2 [11] with substitution models identified by PartitionFinder (Table 3). We performed two Monte Carlo runs of four independent Markov chains (MCMC) for 20 million generations each, sampling every two thousand replicates. Methods for identifying the maximum-clade credibility (MCC) tree are discussed in Melo et al. [1]. We visualized and edited the final MCC phylogeny with FigTree v1.4.2 (tree file max_cred_tree_newick.nwk).

We inferred a maximum likelihood (ML) topology using RAxML HPC v.8 on XSEDE [12] on CIPRES Scientific Gateway v.3.3 [13]. Partitioning schemes were identified using PartitionFinder; however, substitution models were restricted to GTR due to the limitations of RAxML. Additional information on the ML analysis is provided in Melo et al. [1]. The final maximum likelihood phylogeny is provided here in tree file RAxML_bipartitions.unconstrained_result (Fig. 1).

2.5. Species tree analyses

We implemented the sequence-based species tree ancestral reconstruction method *BEAST [14]. This method estimates the posterior probability of all gene trees and species tree simultaneously from the alignment with informed priors on substitutions and rates of evolution. *BEAST requires a priori designation of individuals into species or OTUs (not individual organisms or sequences). Due to the non-monophyletic reconstructions of Prochilodus nigricans and P. rubrotaeniatus in concatenated analysis (see Melo et al. [1]), we assigned those species to two separate species units, denoted by 1 and 2 following the species name (see Fig. 5 in Melo et al. [1]). The final analysis included 77 individuals in 41 nominal species and four taxonomic units. We constrained Prochilodontidae to monophyly based on exceptionally evidence strong from morphology [2], and the concatenated molecular analyses [1]. Brycon pesu served as the outgroup.

We hypothesized six possible partitions (one for each gene), and used the BIC in PartitionFinder 1.1.4 [10] to estimate the best partitioning scheme and to select the best-fit model for each gene (Table 4). We implemented the uncorrelated lognormal distribution (UCLN) rate variation model to estimate trees in BEAST v 1.8.3 because previous empirical and simulation studies have demonstrated that the UCLN model is usually the most accurate and robust [15,16] when local clocks are not expected [17]. A lognormal prior
Fig. 1. Phylogenetic relationships of Prochilodontidae based on maximum likelihood analysis of the concatenated dataset. Numbers near nodes represent bootstrap support. Colored symbols correspond to those in Figs. 3 and 4 of Melo et al. [1]. (F1_RAxML_bestTree.unconstrained_result.nwk).
was set on the mean clock rate for each gene (Table 5; BEAST file StarBeast_Prochilodontidae_250Mgen.xml). A birth-death tree prior was chosen for node time estimation; this models the distribution under a birth-death stochastic branching process model (i.e., speciation and extinction rates can affect a lineage at 0.05

Fig. 2. Maximum likelihood topology with Ichthyoelephas constrained to be sister to Semaprochilodus. (F2_con-straint4_Ichthyoelephas_constrained_RAxML_bestTree.result.nwk).
any time) and is considered the most appropriate when extinction is known or suspected to have occurred in the group [15]. Priors and parameters were set in BEAUti 1.8.3 [18]. We ran four independent MCMC chains for 250 million generations, sampling data every 25,000 generations. The concatenation of the four

Fig. 3. Maximum likelihood topology with _Semaprochilodus taeniurus_ constrained to be sister to a clade containing _S. kneri_ and _S. insignis_. (F3_constraint1_Semaprochilodus_taeniurus_constrained_RAxML_bestTree.result.nwk).
independent runs attained sufficient coverage after 250 million generations with ESS > 200 for most statistics except for some of the root height priors, which are not as relevant to *BEAST analyses as are divergence time estimates in BEAST. The final maximum clade credibility tree was identified from 32,000

Fig. 4. Maximum likelihood topology with Semaprochilodus taeniurus constrained to be sister to a clade containing S. kneri and S. insignis, and S. insignis constrained to monophyly. (F4_constraint2_Semaprochilodus_constrained_RAxML_bestTree.result.nwk).
Fig. 5. Maximum likelihood topology with intrageneric relationships within Prochilodus constrained to those hypothesized by Castro and Vari [2]. (F5_constraint3_Prochilodus_constrained_RAxML_bestTree.result.nwk).
sampled trees with a log clade credibility of −8.56 (Fig. 5 in Melo et al. [1]; treefile StarBeast_MCC- Prochilodontidae_concatenation.nwk).

### 2.6. Shimodaira–Hasegawa tests

In order to compare support for the most likely molecular topology (Fig. 1; treefile F1_RAxML- bestTree.unconstrained_result.nwk) to support for the morphological hypothesis of Castro and Vari [2], we inferred ML trees in RAxML under four morphology-based constraints discussed in Melo et al. [1]. Constraint trees were created in Mesquite 3.04 [19], and results inferred under those constraints appear in Figs. 2–5. (treefiles F2_constraint4_Ichthyoelephas_constrained_RAxML_bestTree. result.nwk F3_constraint1_Semaprochilodus_taeniurus_constrained_RAxML_bestTree.result.nwk, F4constraint2_Semaprochilodus_constrained_RAxML_bestTree.result.nwk, F5_constraint3_Prochilodus_constrained_RAxML_bestTree.result.nwk). The best tree inferred under constraint four (Fig. 2) contains an extremely short branch subtending the Semaprochilodus + Prochilodus clade, effectively creating a genus-level polytomy. This topology likely results from the much poorer probability of the sequence data given any of the tree models available under constraint four. The maximum likelihood tree under constraint four essentially makes the best of a poor region of parameter space by setting the evolutionary history shared by Semaprochilodus and Ichthyoelephas, but not Prochilodus, to the minimum possible value. Branch length shortening under the other three constraints is substantially more subtle.

We compared the ML unconstrained phylogeny with the four constrained phylogenies using the Shimodaira–Hasegawa (SH) test [20] as implemented in phangorn v2.0.1 [21]. The script for
performing these analysis appears here as SHtest.r, and depends upon the FASTA alignment in pro-
chilodontidae.fasta.

Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at: http://dx.doi.org/10.1016/j.dib.2016.08.015.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.08.015.

References

[1] B.F. Melo, B.L. Sidlauskas, B.W. Frable, K. Hoekzema, R.P. Vari, C. Oliveira, Molecular phylogenetics of the Neotropical fish family Prochilodontidae (Teleostei: Characiformes), Mol. Phylogenet. Evol. 102 (2016) 189–201. http://dx.doi.org/10.1016/j.ympev.2016.05.037.
[2] R.M.C. Castro, R.P. Vari, Detritivores of the South American fish family Prochilodontidae (Teleostei: Ostariophysi: Characiformes): a phylogenetic and revisionary study, Smithsonian Contrib. Zool. 622 (2004) 1–189. http://dx.doi.org/10.5479/si.00810282.622.
[3] C. Oliveira, G.S. Avelino, K.T. Abe, T.C. Mariguela, R.C. Benine, G. Ortí, R.P. Vari, R.M. Corrêa e Castro, Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive ingroup sampling, BMC Evol. Biol. 11 (2011) 1–25. http://dx.doi.org/10.1186/1471-2148-11-275.
[4] K.T. Abe, T.C. Mariguela, G.S. Avelino, R.M.C. Castro, C. Oliveira, Multilocus molecular phylogeny of Gasteropelecidae (Ostariophysi: Characiformes) reveals the existence of an unsuspected diversity, Mol. Phylogenet. Evol. 69 (2013) 1209–1214. http://dx.doi.org/10.1016/j.ympev.2013.07.005.
[5] B.F. Melo, B.L. Sidlauskas, K. Hoekzema, R.P. Vari, C. Oliveira, The first molecular phylogeny of Chilodontidae (Teleostei: Ostariophysi: Characiformes) reveals cryptic biodiversity and taxonomic uncertainty, Mol. Phylogen. Evol. 70 (2014) 286–295. http://dx.doi.org/10.1016/j.ympev.2013.09.025.
[6] N.M. Lopera-Barrero, J.A. Povh, R.P. Ribeiro, P.C. Gomes, C.B. Jacometo, T. Lopes, S. da, Comparison of DNA extraction protocols of fish fin and larvae samples: modified salt (NaCl) extraction, Cienc. e Investig. Agrar (2008), http://dx.doi.org/10.4067/rcia.v35i1.374.
[7] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinform. 28 (2012) 1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.
[8] R. Edgar, Muscle: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinform. 5 (2004) 1–19. http://dx.doi.org/10.1186/1471-2105-5-13.
[9] X. Xia, DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution, Mol. Biol. Evol. 30 (2013) 1720–1728. http://dx.doi.org/10.1093/molbev/mst064.
[10] R. Lanfear, B. Calcott, S.Y.W. Ho, S. Guindon, PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses, Mol. Biol. Evol. 29 (2012) 1695–1701. http://dx.doi.org/10.1093/molbev/mss020.
[11] F. Ronquist, M. Teslenko, P. van der Mark, D.L. Ayres, A. Larget, B. Liu, M. Suchard, J.P. Huelsenbeck, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, Syst. Biol. 61 (2012) 539–542. http://dx.doi.org/10.1093/sysbio/sys029.
[12] A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, Bioinform. 22 (2006) 2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446.
[13] M.A. Miller, W. Pfeiffer, T. Schwarz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: Proceedings of the Gateway Computing Environments Workshop, GCE, 2010.
[14] J. Heled, A.J. Drummond, Bayesian inference of species trees from multilocus data, Mol. Biol. Evol. 27 (2010) 570–580. http://dx.doi.org/10.1093/molbev/msp274.
[15] A.J. Drummond, S.Y. Ho, M.J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence, PLoS Biol. 4 (5) (2006) e88. http://dx.doi.org/10.1371/journal.pbio.0040088.
[16] J. Arroyave, M.L.J. Stiassny, Phylogenetic relationships and the temporal context for the diversification of African characins of the family Alestidae (Ostariophysi: Characiformes): evidence from DNA sequence data, Mol. Phylogenet. Evol. 60 (2011) 285–397. http://dx.doi.org/10.1016/j.ympev.2011.04.016.
[17] M.D. Crisp, N.B. Hardy, L.G. Cook, Clock model makes a large difference to age estimates of long-stemmed clades with no internal calibration: a test using Australian grasstrees, BMC Evol. Biol. 14 (2014) 1–17. http://dx.doi.org/10.1186/s12862-014-0263-3.
[18] A.J. Drummond, M.A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7, Mol. Biol. Evol. 29 (2012) 1969–1973. http://dx.doi.org/10.1093/molbev/mss075.

[19] W.P. Maddison, D.R. Maddison, Mesquite: A modular system for evolutionary analysis, version 2.75, 2013. Published by the authors.

[20] H. Shimodaira, M. Hasegawa, Multiple comparisons of log-likelihoods with applications to phylogenetic inference, Mol. Biol. Evol. 16 (1999) 1114–1116.

[21] K.P. Schliep, Phangorn: phylogenetic analysis in R, Bioinform. 27 (2011) 592–593. http://dx.doi.org/10.1093/bioinformatics/btq706.

[22] S. Palumbi, Nucleic acids II: the polymerase chain reaction, in: D. Hillis, C. Moritz, B. Mable (Eds.), Molecular Systematics, Sinauer Associates Inc., Massachusetts, 1996, pp. 205–247.

[23] B.F. Melo, R.C. Benine, T.C. Mariguela, C. Oliveira, A new species of Tetragonopterus Cuvier, 1816 (Characiformes: Characidae: Tetragonopterinae) from the rio Jari, Amapá, northern Brazil, Ichthyol. 9 (2011) 49–56. http://dx.doi.org/10.1590/S1679-62252011000100002.

[24] C. Li, G. Ortí, G. Zhang, G. Lu, A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study, BMC Evol. Biol. 7 (2007) 44. http://dx.doi.org/10.1186/1471-2148-7-44.

[25] N.R. Lovejoy, B.B. Collette, Phylogenetic relationships of new world needlefishes (Teleostei: Belonidae) and the biogeography of transitions between marine and freshwater habitats, Copeia 1 (2001) 324–338. http://dx.doi.org/10.1643/0045-8511(2001)001[0324:PRONWN]2.0.CO;2.