# Phylogeny and divergence times of suckers (Cypriniformes: Catostomidae) inferred from Bayesian total-evidence analyses of molecules, morphology, and fossils

Justin C Bagley [Corresp., 1, 2, 3] , Richard L Mayden [4] , Phillip M Harris [5]

1 Department of Biology, Virginia Commonwealth University, Richmond, VA, United States

2 Departamento de Zoologia, Universidade de Brasília, Brasília, DF, Brazil

3 Departamento de Zoologia e Botânica, IBiLCE, Universidade Estadual Paulista, São José do Rio Preto, SP, Brazil

4 Department of Biology, Saint Louis University, St. Louis, MO, United States

5 Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, United States

Corresponding Author: Justin C Bagley
Email address: jcbagley@vcu.edu

Catostomidae ('suckers') is a diverse (76 species) and broadly distributed family of Holarctic freshwater fishes with a rich fossil record and a considerable number (~35%) of threatened and imperiled species. We integrate DNA sequences (three mitochondrial genes, three nuclear genes), morphological data, and fossil information to infer sucker phylogenetic relationships and divergence times using Bayesian 'total-evidence' methods, and then test hypotheses about the temporal diversification of the group. Our analyses resolved many nodes within subfamilies and clarified Catostominae relationships to be of the form ((Thoburniini, Moxostomatini), (Erimyzonini, Catostomini)). Patterns of subfamily relationships were incongruent, but mainly supported two placements of the Myxocyprininae; distinguishing these using Bayes factors lent strongest support to a model with Myxocyprininae sister to all remaining sucker lineages. We improved our Bayesian total-evidence dating analysis by excluding problematic characters, using a clock-partitioning scheme identified by Bayesian model selection, and employing a fossilized birth-death tree prior accommodating morphological data and fossils. The resulting chronogram showed that suckers evolved since the Late Cretaceous–Eocene, and that the Catostomini and Moxostomatini clades have accumulated species diversity since the early to mid-Miocene. These results agree with the fossil record and confirm previous hypotheses about dates for the origins of Catostomide and catostomine diversification, but reject previous molecular hypotheses about the timing of divergence of ictiobines, and between Asian–North American lineages. Overall, our findings from a synthesis of multiple data types enhance understanding of the phylogenetic relationships, taxonomic classification, and temporal diversification of suckers, while also highlighting practical methods for improving Bayesian divergence dating models by coupling phylogenetic

informativeness profiling with relaxed-clock partitioning.

1   *Research Article*

2   **Phylogeny and divergence times of suckers (Cypriniformes: Catostomidae)**

3   **inferred from Bayesian total-evidence analyses of molecules, morphology, and**

4   **fossils**

5   Justin C. Bagley[1,2,3,*], Richard L. Mayden[4], and Phillip M. Harris[5]

6   *Author Addresses:* [1]Department of Biology, Virginia Commonwealth University, Richmond,

7   VA, USA; [2]Departamento de Zoologia, Universidade de Brasília, Brasília, DF, Brazil;

8   [3]Departamento de Zoologia e Botânica, IBiLCE, Universidade Estadual Paulista, São José do

9   Rio Preto, SP, Brazil; [4]Department of Biology, Saint Louis University, St. Louis, MO, USA;

10   [5]Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, USA

11   *Author E-mail Addresses:* JCB: jcbagley@vcu.edu; RLM: richard.mayden@slu.edu; PMH:

12   pharris@ua.edu.

13   *Corresponding author*: Dr. Justin C. Bagley; Phone: +001 (804)-828-0820; Fax: +001 (804)-

14   828-0503; E-mail: jcbagley@vcu.edu.

15   *Running head:* Sucker phylogeny and divergence times (37 characters w/spaces)

16

17

18   *Contents:* ~8,431 words (Title, Abstract, main text, and Acknowledgements; excluding

19   References and table/figure captions); 4 tables; 5 figures.

## Abstract

20 **Abstract**

21 Catostomidae ('suckers') is a diverse (76 species) and broadly distributed family of Holarctic

22 freshwater fishes with a rich fossil record and a considerable number (~35%) of threatened and

23 imperiled species. We integrate DNA sequences (three mitochondrial genes, three nuclear

24 genes), morphological data, and fossil information to infer sucker phylogenetic relationships and

25 divergence times using Bayesian 'total-evidence' methods, and then test hypotheses about the

26 temporal diversification of the group. Our analyses resolved many nodes within subfamilies and

27 clarified Catostominae relationships to be of the form ((Thoburniini, Moxostomatini),

28 (Erimyzonini, Catostomini)). Patterns of subfamily relationships were incongruent, but mainly

29 supported two placements of the Myxocyprininae; distinguishing these using Bayes factors lent

30 strongest support to a model with Myxocyprininae sister to all remaining sucker lineages. We

31 improved our Bayesian total-evidence dating analysis by excluding problematic characters, using

32 a clock-partitioning scheme identified by Bayesian model selection, and employing a fossilized

33 birth-death tree prior accommodating morphological data and fossils. The resulting chronogram

34 showed that suckers evolved since the Late Cretaceous–Eocene, and that the Catostomini and

35 Moxostomatini clades have accumulated species diversity since the early to mid-Miocene. These

36 results agree with the fossil record and confirm previous hypotheses about dates for the origins

37 of Catostomidae and catostomine diversification, but reject previous molecular hypotheses about

38 the timing of divergence of ictiobines, and between Asian–North American lineages. Overall,

39 our findings from a synthesis of multiple data types enhance understanding of the phylogenetic

40 relationships, taxonomic classification, and temporal diversification of suckers, while also

41 highlighting practical methods for improving Bayesian divergence dating models by coupling

42 phylogenetic informativeness profiling with relaxed-clock partitioning.

43   **Introduction**

44   'Suckers' in the family Catostomidae (Cypriniformes) form a large family of Holarctic

45   freshwater fishes with 76 extant species in 14 genera native to North America and Asia (Harris et

46   al., 2014). Seventy-five species from 13 genera occur in North America (Harris et al., 2014;

47   Nelson et al., 2004; Unmack et al. 2014), where they constitute the third largest freshwater fish

48   clade, after darters (Etheostomatinae) and minnows (Cyprinidae), and ~8% of the continental

49   ichthyofauna (Warren et al., 2000). *Myxocyprinus asiaticus* is endemic to eastern China and

50   *Catostomus catostomus* is the only extant trans-Pacific species (Harris et al., 2014). Suckers are

51   an ancient group whose fossil record spans the Cenozoic, from the early Eocene onwards

52   (Cavender, 1986; Smith, 1992; Smith et al., 2002; Appendix S1). Around 35% of the taxa (26 to

53   >35 species or genetic lineages) are endangered, threatened, or of special conservation concern

54   (Harris et al., 2014; Nelson et al., 2004; Warren et al., 2000).

55          Geographical distributions of suckers and their taxonomy and relationships have attracted

56   the interest of systematists and biogeographers for over 150 years (e.g. Agassiz, 1854). Pre-1900

57   systematics and taxonomy studies dealt with species descriptions and higher-level classifications

58   of the group (reviewed by Harris et al., 2014; Smith, 1992). Subsequently, key papers on sucker

59   classification designated genera, subgenera, and tribes (Hubbs, 1930; Robins & Raney, 1956)

60   and contributed a pre-cladistics phylogeny (Miller, 1959), as well as the first phylogeny for

61   species in the tribe Moxostomatini (Jenkins, 1970). The findings of post-systematics studies of

62   sucker interrelationships (last 40 years), which were based on a variety of different data types,

63   are summarized in Table 1 and Figure 1. These studies supported the recognition of four

64   subfamilies (Myxocyprininae, Cycleptinae, Ictiobinae, and Catostominae) and tribes

65   (Catostomini, Erimyzonini, Moxostomatini, Thoburniini) plus the modern allocation of genera to

66   these groups (Harris & Mayden, 2001; Harris et al., 2002). Most phylogenetic analyses of

67    suckers to date have analyzed DNA sequence data from increasing numbers of mitochondrial

68    DNA (mtDNA) or nuclear DNA (nDNA) genes. Recent advances include Clements et al.'s

69    (2012) study of phylogenetic relationships of Moxostomatini using the first dataset based on two

70    loci—mtDNA cyt*b* and nuclear growth hormone intron (GHI) sequences. Also, Unmack et al.

71    (2014) studied the phylogeny and biogeography of *Pantosteus*, a former subgenus of *Catostomus*

72    (Harris & Mayden, 2001; Harris et al., 2002; Smith, 1966) using independent morphological and

73    mtDNA genetic analyses. Overall, while the inclusion of additional taxa and sequence data in

74    recent analyses has yielded novel insights on sucker relationships, a consensus among

75    hypotheses regarding relationships of higher-level species groups, i.e. subfamilies, has yet to

76    emerge (Table 1; Fig. 1). Moreover, although time-calibrated phylogenies based on

77    comprehensive tip sampling are necessary to understand the tempo and mode of diversification

78    of groups of freshwater fishes (e.g. Near et al., 2011, 2012), such a tool has yet to be inferred for

79    suckers. This leaves workers at a distinct disadvantage in considering the macroevolutionary,

80    ecological, or conservation trends of suckers in a phylogenetic context.

81        In this study, we use 'total-evidence' analyses of multilocus sequence data from three

82    mtDNA genes and three nuclear genes representing the most comprehensive sampling of suckers

83    to date, combined with available morphological and fossil data, to infer the phylogenetic

84    relationships and divergence times of suckers in a Bayesian framework. We discuss the

85    phylogenetic and taxonomic implications of our results in light of previous phylogenetic studies

86    of suckers. We apply Bayesian divergence time methods to the data (e.g. Drummond et al., 2012)

87    and take advantage of rich information on the age and distributions of sucker fossils (e.g.

88    Cavender, 1986; Smith et al., 2002; Appendix S1) to improve divergence time estimation by

89    incorporating extant and fossil sampling in a "fossilized birth-death process" tree prior (Stadler

90    2010; Gavryushkina et al., 2014; Heath et al. 2014). Moreover, by coupling assessments of the

91    phylogenetic signal of different data subsets with evaluation of clock-partitioning strategies, we

92    were able to avoid potentially confounding effects of problematic characters on our divergence

93    time inferences. The resulting time-calibrated phylogeny is then used to test several hypotheses

94    about the temporal diversification of suckers. Namely, we test the hypotheses ($H_1$) that Asian

95    Myxocyprininae and North American suckers (Cycleptinae) diverged since ~14 million years

96    ago (Ma) in the mid-Miocene, and ($H_2$) that the initial divergence of Ictiobinae lineages occurred

97    since ~10 Ma in the late Miocene, as indicated by mtDNA cyt*b* gene divergences (Sun et al.,

98    2007). We also tested the hypotheses ($H_3$) that the diversification of lineages within

99    Catostominae followed ~20 Ma in the Miocene (Smith, 1992; Sun et al., 2007), and ($H_4$) that

100   Catostomidae species have diversified from a tetraploid ancestor since ~50 Ma in the early

101   Eocene (Uyeno & Smith, 1972). Overall, by yielding a new phylogeny of suckers and divergence

102   dates for their most recent common ancestors (MRCAs), our study sheds light on the

103   interrelationships, taxonomic classification, and tempo of speciation in a diverse and threatened

104   clade of Holarctic freshwater fishes.

105   **Materials & Methods**

106   *Molecular taxon sampling, laboratory methods, and sequence alignment*

107   We obtained and sequenced mtDNA and nDNA genes from tissue samples of 121 sucker

108   specimens from throughout the geographical range of the family. Samples were provided by

109   ichthyological collections or colleagues (see Acknowledgements), or were already in-hand at the

110   beginning of the study. We sequenced taxa representing all sucker genera and all Catostominae

111   species except for †*Moxostoma lacerum*, an historically extinct taxon last sampled from the

112   Mississippi River Basin in 1893 (NatureServe, 2013), and *Chasmistes cujus*, an endangered

113    species (NatureServe, 2014) for which it was difficult to obtain samples (Table 2). To

114    complement our sampling and add data from more unlinked loci to our datasets, we obtained

115    additional sequences from GenBank (see section 2.2). Overall, we obtained genetic sequences

116    for the most comprehensive taxonomic sampling of the family to date, including 78

117    species/lineages representing all 14 extant genera, including four 'candidate species' within

118    Moxostomatini (Table 2). Based on recent hypotheses of relationships of suckers and

119    cypriniform fishes (e.g. Mayden et al., 2008; Saitoh et al., 2006; Smith, 1992), we included DNA

120    sequences from five outgroup taxa in our datasets: *Cyprinus carpio* (Cyprinidae); *Gyrinocheilus*

121    *aymonieri* (Gyrinocheilidae); and *Cobitis striata*, *Chromobotia macracanthus*, and *Leptobotia*

122    *mantschurica* (Cobitidae). We follow the taxonomy of subfamilies listed in Harris and Mayden

123    (2001) and Harris et al. (2002).

124        All molecular laboratory work for this project conducted in the R. L. Mayden laboratory

125    was approved under Saint Louis University Institutional Animal Care and Use Committee

126    (IACUC) protocol #2467. We extracted whole genomic DNA using QIAGEN DNeasy Tissue

127    kits (QIAGEN, Valencia, CA, Catalog No. 69506), or the CTAB method of Sagahi-Maroof et al.

128    (1984). We amplified and sequenced the mtDNA cyt*b* gene using PCR primers and parameters

129    in Harris et al. (2002). We amplified and sequenced the mtDNA NADH subunit 2 (ND2) gene

130    with primers 562 (5′-TAA GCT ATC GGG CCC ATA CC-3′) and 449 (5′-TGC TTA GGG CTT

131    TGA AGG CTC-3′) from LGL Genetics, Bryan, TX, using the same PCR amplification

132    parameters used for cyt*b*. We also sequenced the first part of the nuclear IRBP gene using PCR

133    primers and methods in Chen et al. (2008), and we sequenced nuclear ribosomal protein *S7*

134    intron 1 (*RPS7*) using PCR primers and parameters in Chow & Takeyama (1998). We purified

135    double-stranded PCR products using columns or gel extraction kits (QIAGEN, Valencia, CA).

136   Given that catostomids are tetraploids (Uyeno & Smith, 1972), we cloned nuclear PCR products

137   to ensure orthologous sequences were used in subsequent phylogenetic reconstructions. Purified

138   PCR products were cloned using the TOPO TA Cloning® Kit (Invitrogen Corp., Carlsbad, CA)

139   with TOP10 chemically competent or electrocompetent cells. Positive colonies were chosen

140   randomly and cultured, and then plasmids preps were purified using the QIAprep Spin Miniprep

141   Kit (QIAGEN, Valencia, CA) and sequenced in both directions using universal M13 primers. We

142   sequenced all genes in both directions on an Applied Biosystems 3100 Genetic Analyzer using

143   ABI PRISM BigDye Terminator v2.0 or v3.0 Cycle Sequencing Kits (Applied Biosystems,

144   Foster City, CA).

145        We edited sequence chromatographs, assembled sequence contigs, and created final DNA

146   alignments using Geneious v5.4 (Kearse et al., 2012). We translated all gene sequences into

147   amino acid sequences to check alignments for stop codons or elevated nonsynonymous

148   substitution numbers, because these signatures can indicate the presence of nuclear mtDNA gene

149   copies, or 'NUMTs'. Mitochondrial DNA and IRBP sequences aligned straightforwardly 'by-

150   eye' in Geneious. However, we aligned nuclear *RPS7* sequences, and nuclear genes from other

151   studies listed in section 2.2, in MAFFT v6.850 (Katoh & Toh, 2008) using the local-pair FFTS

152   algorithm with MAXITERATE = 50.

153   *Dataset construction, model selection, and Bayesian phylogenetic analyses*

154   We collated seven datasets for our analyses. (1) The 'concatenated mtDNA' dataset consisted of

155   mtDNA cyt*b* and ND2 sequences for 126 tips (121 sucker samples and 5 outgroup samples) plus

156   58 mtDNA cytochrome oxidase subunit 1 (*cox1*) gene sequences from GenBank. (2) A 'four-

157   locus' DNA dataset with the same 126 tips but adding the following nuclear data to the

158   concatenated mtDNA dataset: 44 GHI sequences from Clements et al. (2012) and GenBank, plus

159    113 IRBP sequences and 52 *RPS7* sequences from our study. (3) ==We reanalyzed a morphological==

160    ==dataset first presented by Smith (1992), containing 123 morphological characters for 64 taxa==

161    ==including 62 sucker taxa and two outgroup taxa.== This matrix contained two extinct taxa: the

162    historically extinct †*M. lacerum*, and †*Amyzon,* an Eocene–Oligocene genus of fossil suckers

163    composed of five valid species known from British Columbia, Washington, Nevada, and

164    Wyoming (four species), and Jilin province, China (Bruner, 1991; Chang et al., 2001; Smith,

165    1992; Appendix S1). (4) Our 'total-evidence' dataset contained all morphological and molecular

166    characters analyzed in this study, for 85 taxa. Two final nDNA datasets were: (6) a 'nuclear

167    IRBP' alignment for the only nuclear gene with complete sampling, and (7) a 'concatenated

168    nDNA' alignment containing all three nuclear loci.

169        We selected the most appropriate partitioning schemes and models of sequence evolution

170    for each of the DNA data 'subsets' or 'blocks' used in our phylogenetic analyses in

171    PartitionFinder v1.1.1 (Lanfear et al., 2012), which included codon positions of the mtDNA and

172    IRBP alignments, plus the GHI and *RPS7* genes. We ran PartitionFinder simultaneously on all

173    initial DNA subsets using the greedy heuristic search algorithm, which we set to conduct model

174    comparisons to determine the 'best-fit' partitioning scheme based on the Bayesian information

175    criterion (BIC). PartitionFinder relies heavily on PhyML, which we set to link branch lengths,

176    search 56 substitution models, and estimate the base frequencies, proportion of invariant sites ($I$),

177    and the gamma shape distribution ($\Gamma$) of each model using maximum-likelihood.

178        We performed partitioned Bayesian phylogenetic analyses on all seven datasets in

179    MrBayes 3.2.2 (Ronquist et al., 2012). Analyzing different taxon and character combinations

180    allowed us to evaluate the effect of different data types, and of including morphological data and

181    extinct taxa, on phylogenetic analyses of suckers. During molecular analyses, we specified

182    partitioning schemes and best-fit models of sequence evolution selected in PartitionFinder,

183    except where the selected model was not implemented in MrBayes we used the next most closely

184    related model in the GTR family of models. Given that rate variation among morphological

185    characters can confound phylogenetic branch lengths, applying $\Gamma$-distributed rates can greatly

186    improve models of morphological evolution (e.g. Clarke & Middleton, 2008). Thus, for the

187    morphological analysis, we specified Lewis's (2001) Markov variable (Mkv) model with $\Gamma$-

188    distributed rate heterogeneity and left characters unordered in state polarity (default). For each

189    dataset, we conducted three independent MrBayes runs of eight chains, each with a Markov

190    chain Monte Carlo (MCMC) chain length of 50 million generations. We diagnosed run

191    convergence using the potential scale reduction factor, which should approach values of 1 when

192    stationarity has been reached (Ronquist et al., 2012).

193    *Bayesian total-evidence dating and relaxed-clock partitioning*

194    We estimated divergence times as times to the most recent common ancestor ($t_{MRCA}$s) through

195    Bayesian relaxed molecular clock analyses of the total-evidence dataset in BEAST v2.4.5

196    (Bouckaert et al., 2014). For our tree prior, we employed the fossilized birth-death (FBD)

197    process model (Gavryushkina et al., 2014; Heath et al., 2014) as modified by Gavryushkina et al.

198    (2017), because this model accommodates total-evidence datasets, avoids the need for arbitrary

199    calibration densities, and accommodates all fossils available for a group, and not merely ad hoc

200    selections. In addition to 83 extant ingroup and outgroup samples, 19 extinct sucker taxa and

201    their available minimum ages from the fossil record were included in the BEAST analyses based

202    on evidence provided in Appendix S1. We calculated the approximate sampling proportion for

203    extant lineages ($\rho$) as 0.97 and set the FBD time of origin prior ($t_{or}$) to the Cypriniformes 'Root'

204    node calibration discussed in Appendix S1. Morphological characters were partitioned into

205 groups having the same number of states, and each partition was assigned an Mkv model (Lewis,

206 2001; conditioning on the use of only variable characters) with $\Gamma$-distributed rate variation

207 ("partitioned mode" in Gavryushkina et al., 2017). We partitioned the DNA data and set site

208 models according to the best scheme identified in PartitionFinder. Other authors have employed

209 FBD models on a fixed topology (e.g. Heath et al., 2014) or simultaneously estimated topology

210 and FBD parameters (e.g. Gavryushkina et al., 2017). We allowed most nodes, including

211 subfamily relationships, to change freely. However, we constrained the Catostomini and

212 Moxostomatini crown groups to be monophyletic, consistent with our MrBayes results and

213 hypotheses of previous studies (see Results and Discussion sections). Posterior nodal support can

214 be markedly weakened when fossils have unscored characters (e.g. Gavryushkina et al., 2017) or

215 lack data. To avoid spurious relationships arising from the large proportion of fossil Ictiobinae

216 taxa (57%) and Myxocyprininae taxa (50%) lacking data, samples of *Carpiodes*, *Ictiobus*,

217 Ictiobinae, and Myxocyprininae were each constrained to be monophyletic, consistent with our

218 other phylogenetic results.

219      According to Ho & Lanfear (2010), implementing multiple relaxed-clock models for

220 different data subsets in a data-partitioning scheme can provide a more biologically realistic way

221 to model among-lineage rate variation and improves the fit of relaxed-clock models to the data.

222 Using a relaxed-clock partitioning scheme may also yield more precise date estimates with

223 narrower credible intervals (Ho & Lanfear, 2010). We statistically tested whether allocating

224 separate uncorrelated lognormal relaxed clocks to different DNA data subsets, through 'relaxed-

225 clock partitioning', yielded divergence time estimation models that provided a better fit to the

226 data than assuming a single model of branch-specific rates across all data subsets. First, we

227 estimated divergence dates and marginal likelihood scores for each of 12 different relaxed-clock

228 partitioning models ($M_1$–$M_{12}$) with one to eight clocks (see Results and Discussion). Second, we

229 estimated Bayes factors and conducted Bayesian model selection to identify the best-supported

230 model.

231        We ran five replicate searches of each model in BEAST (MCMC = $2 \times 10^7$ generations,

232 sampling every 4000) using the 'BEASTRunner.sh' script in PIrANHA (Bagley, 2017). We then

233 estimated log-marginal likelihoods for each model by conducting path sampling (PS) (Baele et

234 al., 2012) for 100 steps ($10^6$ generations each), while specifying a ~B(0.3, 1) distribution for

235 spacing the path steps (Xie et al., 2011). We calculated $2\log_e(B_{10})$ Bayes factors from the log-

236 marginal likelihoods and evaluated 'weight of evidence' of the models according to criteria in

237 Kass & Raftery (1995). We took posterior distributions from the best-supported model as our

238 best estimates of the time tree and divergence dates for Catostomidae lineages. We summarized

239 parameters from the best-supported model, and ensured convergence and adequate effective

240 sample sizes (ESS >100–200), using Tracer v1.5 (Rambaut & Drummond, 2014). We calculated

241 a maximum clade credibility tree annotated with mean node ages from 5,000 post-burn-in trees

242 in TreeAnnotator v2.4.5.

243 *Using phylogenetic informativeness profiles to exclude problematic characters*

244 Among other factors influencing phylogenetic inference, such as sampling effects on branch

245 lengths or nodal support values (e.g. Heath et al., 2008; Pyron, 2011), the varying

246 informativeness of different character sets can substantially and adversely affect phylogenetic

247 divergence dating results (Dornburg et al., 2014). Fortunately, recent methods for quantifying

248 and visualizing 'phylogenetic informativeness' (PI) profiles of character sets through time

249 (Townsend, 2007; Townsend et al., 2012) provide a framework for identifying and excluding

250 problematic character sets (Dornburg et al., 2014). One common pitfall is the use of characters

251     whose profiles exhibit a decline in informativeness towards the root. As noted by Townsend &

252     Leuenberger (2011), this decline marks a "rain shadow of noise", with the corresponding dataset

253     losing phylogenetic informativeness due to an increase in predicted homoplasy. In turn,

254     homoplasious loci or character sets, but especially mtDNA datasets exhibiting high saturation or

255     rootward declines in PI, have been shown to mislead global branch length values during

256     divergence time estimation (e.g. Brandley et al., 2011; Dornburg et al., 2014).

257         To identify and exclude potentially problematic character sets in our database, and to

258     evaluate whether support versus instability of subfamily relationships correlated to statistical

259     power to revolve branching order, we evaluated the Townsend (2007) PI and resolution

260     probabilities of each character set through time using PhyDesign (López-Giráldez & Townsend,

261     2011). Estimating PI requires prior information on evolutionary-genetic rates and phylogeny.

262     Thus, we estimated rates for DNA characters and morphological characters using HyPhy (Pond

263     et al., 2005) and BayesTraits (Pagel & Meade, 2014), respectively, and ran analyses along the

264     BEAST time tree from the best-supported clock-partitioning model (see Results and Discussion).

265     We estimated net PI for all eight data subsets in Table 3 during the 'subfamily divergence epoch'

266     spanning branches leading to Myxocyprininae, Cycleptinae, Ictiobinae, and Catostominae. These

267     analyses permitted broad comparisons of signal in the mitochondrial versus nuclear data, and

268     molecular versus morphological character sets. We excluded character sets that exhibited steep

269     declines in PI towards the root of our phylogeny from the final divergence time results presented

270     below. We compared three resolution probabilities, including probability correct, probability

271     polytomy, and probability incorrect or 'phylogenetic noise' (equations 11–13 in Townsend et al.,

272     2012) for the excluded subsets versus other molecular subsets over the subfamily divergence

273     epoch. We also evaluated sensitivity of resolution probability approximations to varying the

274    internode time length ($t_0$) parameter, by recalculating over 10 $t_0$ values representing declining

275    fractions of the epoch.

276    **Results**

277    *Dataset characteristics and DNA substitution models*

278    Our final data matrices ranged in size from 123 characters in the morphological dataset to 6048

279    molecular and morphological characters in the total-evidence dataset (Table 3). Proportions of

280    missing data and parsimony informative characters ranged from 0.3–36.4% and from 21.9–

281    97.6%, respectively, across datasets (Table 3). PartitionFinder identified seven unique DNA

282    sequence subsets (scheme BIC = 141854.71523), and the best-fit DNA substitution model for

283    each subset is listed in Table S1. Morphological character subsets (in subset M) were assigned

284    Mkv+$\Gamma$ models, as described above. None of the mtDNA genes sequenced in this study showed

285    signs of NUMTs, and we found no indels in the IRBP sequences; however, GHI and *RPS7* genes

286    aligned with ~32 and ~16 ingroup indels/gaps, respectively. We archived our sequence

287    alignments and phylogenetic tree results in Mendeley Data (doi: 10.17632/trw6sb4v7w.1).

288    *Phylogenetic relationships among sucker subfamilies*

289    We placed subfamilies Myxocyprininae and Cycleptinae as sister lineages in most trees, but with

290    higher posterior support [Bayesian posterior probability (BPP) = 0.74–0.99] in runs based largely

291    on mtDNA-encoded genes (Figs. 2A, 2D, S1 and S2) and weak BPP support (≤0.69) in other

292    runs (Figs. 2C, 2E, 3, S3). Frequently, when this pattern was obtained, Myxocyprininae +

293    Cycleptinae was resolved as sister to a monophyletic Ictiobinae with variable posterior support

294    (BPP = 0.69–0.98; Figs. 2, 3, S1 and S2). This 'subfamily pattern 1', with ((Myxocyprininae,

295    Cycleptinae), Ictiobinae), departs markedly from the placement of Ictiobinae sister to

296    Cycleptinae (sometimes including *Myxocyprinus*) + Catostominae in previous analyses of

297    morphology (Smith, 1992) and molecular data (Doosey et al., 2010; Harris & Mayden, 2001;

298    Mayden et al., 2008; Saitoh et al., 2006). By contrast, several trees agreed in placing

299    Myxocyprininae as sister to all other sucker lineages, which agrees with previous mtDNA results

300    presented in Harris & Mayden (2001). This 'subfamily pattern 2' relationship was strongly

301    supported with BPP = 1 in our total-evidence consensus topology from BEAST (Fig. 4) and

302    resolved with low support in the four-locus topology (Figs. 2B and S4). Distinguishing between

303    these two conflicting sets of subfamily relationships is difficult, because each is supported by

304    molecular and total-evidence topologies herein and agrees with at least one previous molecular

305    study. To objectively determine the arrangement of these subfamilies with the greatest weight of

306    evidence conditional on our total-evidence dataset, we compared subfamily patterns 1 and 2

307    using Bayes factors. We ran MrBayes as described above, except employing topological

308    constraints set to subfamily pattern 2, and then used stepping-stone sampling (Xie et al., 2011;

309    Baele et al., 2012) to estimate the log-marginal likelihoods of the models, from which Bayes

310    factor tests were conducted through comparisons to the unconstrained model matching subfamily

311    pattern 1. Conducting 50,000 generations of stepping-stone sampling (sampling every 2500

312    generations) during each of 50 steps produced a total of 250,000 MCMC generations for

313    marginal likelihood estimation. The subfamily pattern 2 model constraining Myxocyprininae as

314    sister to all other suckers had a higher log-marginal likelihood score ($-68390.83$) than the

315    unconstrained subfamily pattern 1 model ($-68910.88$), and a $2\log_e(B_{10})$ Bayes factor of $-1040.10$

316    provided definitive weight of evidence against the unconstrained model.

317    *Relationships of early-diverging sucker genera*

318    We found that the early-diverging sucker genera *Myxocyprinus*, *Cycleptus*, *Carpiodes*, and

319    *Ictiobus* formed well-supported clades in most analyses. However, *Ictiobus* relationships were

320    resolved in a polytomy in our analysis of the concatenated mtDNA dataset (Fig. S1). In several

321    other trees, including those based on our total-evidence and mtDNA + morphology datasets

322    (Figs. 3 and S2), the *Ictiobus* clade received weak BPP support and eventually collapses into a

323    paraphyletic grade. In the analysis of the four-locus dataset with higher numerical sampling in

324    this clade, *I. bubalus* and *I. niger* relationships had high BPP support (BPP = 0.94–1) but were

325    para-/polyphyletic, leaving their relationship to *I. cyprinellus* uncertain (Fig. S4).

326    *Phylogenetic relationships among the Catostominae*

327    Within the largest sucker subfamily, Catostominae, we consistently resolved clades with the

328    tribes Thoburniini + Moxostomatini and Erimyzonini + Catostomini across analyses. Our more

329    robust, multilocus and total-evidence trees resolved these relationships with definitive support

330    values of BPP = 0.99–1. Within the Thoburniini + Moxostomatini clade, we consistently inferred

331    the genus *Thoburnia* to be paraphyletic, with *T. atripinnis* sister to a clade containing the three

332    *Hypentelium* species (mostly BPP = 1). The sole exception to this was that our morphology tree

333    resolved *Thoburnia* as monophyletic with *T. atripinnis* sister to all other *Thoburnia* with strong

334    support (BPP = 1). Within the *Hypentelium* clade, we inferred an identical and strongly

335    supported set of relationships of the form (*H. roanokense*, (*H. etowanum*, *H. nigricans*)) in the

336    mtDNA, four-locus, nDNA, and total-evidence gene trees (Figs. 3, S1, S2, S4 and S5). We

337    obtained the same set of relationships in our morphology analysis, but with weak (BPP =0.69)

338    support for the *H. etowanum–H. nigricans* node (Fig. S3). We resolved *Moxostoma* as

339    monophyletic with BPP = 0.89–1, except for a paraphyletic pattern in the morphology consensus

340    tree. Within the Erimyzonini, *Erimyzon* was monophyletic (BPP = 1) and sister to *Minytrema*

341    (e.g. Fig. 3). None of the molecular or total-evidence topologies we inferred resolved

342    *Catostomus* as monophyletic relative to *Chasmistes*, *Deltistes*, or *Xyrauchen*. Here, yet again,

343    results from the morphology tree departed from our other results, failing to resolve relationships

344    among these or virtually any other catostomine lineages with strong support (Figs. 1C and S3).

345    As a result, we do not discuss the morphology consensus topology further in this section.

346         Relationships within the Erimyzonini and Catostomini were similar to those in previous

347    molecular studies (e.g. Harris et al., 2002; Doosey et al., 2010). Within *Erimyzon*, our results

348    placed *E. sucetta* rather than *E. oblongus* as sister to *E. tenuis*. Within Catostomini, we

349    consistently resolved 9 well-supported major clades within Catostominae (e.g. Figs. 3, S1 and

350    S4). Although relationships among these clades received varying posterior support, the species

351    groups we identified were highly supported in multiple analyses and provide more tenable

352    phylogenetic hypotheses than previously proposed for this tribe (Smith, 1992; Smith et al.,

353    2002). 'Clade 4' corresponded to a monophyletic Erimyzonini, while clades 5 through 9 included

354    various Catostomini subclades composed largely of *Catostomus* samples. For conciseness, we

355    provide an in-depth assessment of relationships only within 'Clade 5', whose results have the

356    most important phylogenetic and taxonomic implications. More granular presentation and

357    discussion of relationships within and among Catostomini clades 1–3 and 5–9 is provided at the

358    end of Appendix S1. 'Clade 5' corresponded mostly to the genus *Pantosteus* (Unmack et al.

359    2014) and was sister to the remaining Catostomini. Within Clade 5, we consistently resolved *P.*

360    *nebuliferus* + *P. plebeius* as sister to a clade containing all remaining *Pantosteus*, with (*P.*

361    *platyrhynchus*, (*P. santaanae*, (*P. clarkii*, (*P. d. discobolus*, *P. d. yarrowi*)))) (e.g. Fig. 3).

362    Alternative topologies inferred for this clade involved rearrangements placing *P. santaanae* sister

363    to *P. clarkii*, but with non-significant posterior support (e.g. Figs. S1 and S4).

364    *Bayesian total-evidence dating and relaxed-clock partitioning*

365     Bayes factor comparisons of 12 clock-partitioning models showed that removing mtDNA 1st and

366     3rd codon sites deemed to be problematic during PI profiling (see below) progressively improved

367     model log-marginal likelihoods and posterior evidence (Table 4). For example, codon-partitioned

368     relaxed-clock models were overwhelmingly supported over simpler models allocating a single

369     relaxed clock to all data subsets, or all DNA subsets. The most complex model allocating relaxed

370     clocks to each data subset but including only mtDNA 2nd positions, $M_{12}$, was decisively

371     supported as the best model. Compared with this model, other subset and clock schemes

372     produced negative improvements to the model, indicated by negative log Bayes factors;

373     however, the second best model separated DNA from morphological data subsets, consistent

374     with a decoupling of molecular rates from morphological rates (Table 4). Independent runs of the

375     best-supported BEAST model achieved ESS scores of >100–200 for all parameters and

376     converged on similar phylogeny and parameter estimates including mean and 95% highest

377     posterior densities (HPDs; i.e. credible intervals) for sucker $t_{MRCA}$s. For example, the posterior

378     ESS of the best model was 463 and that for the FBD model was 134. The only exception to this

379     was the tMRCA for Moxostomatini, which received ESS scores between 72 and 89. Clocks on

380     different data subsets exhibited substantial among-lineage rate heterogeneity, with posterior

381     means and 95% HPD intervals of the 'ucldStdev' (uncorrelated lognormal relaxed clock standard

382     deviation) and 'coefficientOfVariation' (coefficient of variation of branch-specific rates)

383     statistics excluding zero (Fig. S7); thus, relaxed-clock models were warranted by the data.

384          In the final total-evidence dating analysis, the mean posterior age estimate for the $t_{MRCA}$

385     of all suckers was 63.16 Ma in the Late Cretaceous, with credible intervals ranging from Late

386     Cretaceous to the Paleocene–Eocene boundary (95% HPD: [54.02, 74.6]; Fig. 4). The four

387     sucker subfamilies had variable posterior age estimates ranging approximately an order of

388   magnitude. The Cycleptinae had the youngest posterior age estimate of 5.07 Ma in the Pliocene

389   (95% HPD: [0.87, 10.23]). Following their Late Cretaceous origin based on a stem age

390   corresponding to the MRCA of all suckers, the posterior $t_{MRCA}$ estimate for Myxocyprininae

391   dated their diversification to 42.27 Ma in the early to mid-Eocene (95% HPD: [39.67, 54.58]).

392   Subsequently, Catostominae species diversified since an intermediate posterior age of 34.37 Ma

393   near the Eocene–Oligocene boundary (95% HPD: [25.54, 42.77]), and Ictiobinae species had the

394   oldest posterior age estimate, dating to 49.69 Ma in the early Eocene (95% HPD: [48.88, 52.52]).

395   The catostomine tribes diverged approximately 29.87 Ma in the Oligocene (Catostomini +

396   Erimyzonini) and 20.78 Ma in the early Miocene (Moxostomatini + Thoburniini). The genera

397   *Catostomus* and *Moxostoma*, which correspond to tribes Catostomini and Moxostomatini,

398   diversified since 17.65 Ma and 15.25 Ma ago in the early-mid Miocene, respectively (Fig. 4).

399   *Phylogenetic informativeness profiles*

400   We evaluated potential impacts of phylogenetic signal on incongruent subfamily relationships by

401   estimating resolution probabilities, over the subfamily divergence epoch (~63.2–34.4 Ma)

402   spanning the divergence of sucker subfamilies. Overall, mtDNA 1st and 3rd codon position data

403   subsets exhibited among the highest PI values, but with distinct Miocene peaks followed by

404   declining PI towards the root (Fig. 5). This suggested a prominent loss of evolutionary

405   information due to homoplasy; therefore, we excluded these sites from final divergence dating

406   analyses, in order to avoid potentially negative effects on the topology and time-calibrated

407   branch lengths (Dornburg et al., 2014). All other molecular data subsets had substantial and

408   relatively constant predicted PI decaying over Paleocene or Eocene to present (recent spikes are

409   anomalies; Townsend et al. 2008). Morphological characters had slightly higher signal than

410   IRBP sites and exhibited stability before decaying 20 Ma to present (Fig. 5). In addition to nearly

411  constant net PI (Fig. 5), the retained character subsets also had low probabilities of phylogenetic

412  noise or polytomies, with notable increases in the probability of an incorrect topology only for

413  internode distances less than ~$0.35(t_0)$, or <10.0 million years (Myr) (Fig. S9).

414

## Discussion

416  *Sucker phylogeny and incongruence of subfamily lineages*

417  Our phylogenetic reconstructions of Catostomidae relationships are similar to several previous

418  morphological and molecular studies. For example, others have hypothesized that Catostomidae

419  is monophyletic in studies focusing on suckers (e.g. Doosey et al., 2010; Ferris & Whitt, 1978;

420  Harris & Mayden, 2001; Smith, 1992; Fig. 1) and taxonomically broader analyses (e.g. Mayden

421  et al., 2008; Saitoh et al., 2006). Our finding that the four currently recognized sucker

422  subfamilies are monophyletic with definitive support also agrees with earlier phylogenetic

423  studies based on morphology (Smith, 1992) and molecules (Chen & Mayden, 2012; Clements et

424  al., 2012; Doosey et al., 2010; Harris and Mayden, 2001; Harris et al., 2002; Sun et al., 2007).

425  This is perhaps unsurprising, as we reanalyzed previous morphological and molecular datasets

426  alongside new sequence data. Yet, ours are the first results definitively supporting patterns of

427  monophyly at the family and subfamily levels based on dense taxonomic sampling of mtDNA

428  and nuclear gene sequences for all sucker genera and most species, plus total-evidence analyses,

429  with BPP at or near 1 across datasets (Figs. 2, 3, and S1–S6).

430         Previous molecular phylogenetic studies of higher-level sucker relationships have often

431  encountered difficulty in resolving relationships among sucker subfamilies (Chen & Mayden,

432  2012; Doosey et al., 2010; Harris et al., 2002; Sun et al., 2007). Likewise, relationships among

433  the Myxocyprininae, Cycleptinae, and Ictiobinae lineages were incongruent across analyses of

434   different datasets (Figs. 2, 3, S1–S4), but with two main patterns that we deemed subfamily

435   pattern 1, with the form ((Myxocyprininae, Cycleptinae), Ictiobinae), and subfamily pattern 2,

436   with Myxocyprininae sister to all other sucker subfamilies. We distinguished between these two

437   alternative hypotheses using a topological constraint test based on Bayes factors. The result

438   yielded log-marginal likelihood estimates and Bayes factors giving definitive weight of evidence

439   against the unconstrained subfamily pattern 1 model. Given this result, a placement of

440   Myxocyprininae as sister to all other suckers seems most probable at this point, thus we favor the

441   patterns of subfamily relationships in our four-locus and Bayesian total-evidence dating

442   topologies that are consistent with this result. Nevertheless, the question still remains: What

443   factors have likely influenced the difficulty of our study and previous studies to resolve

444   phylogenetic relationships among sucker subfamily lineages? Overall, our phylogenetic

445   informativeness analyses highlight two potential explanations for the observed incongruence in

446   subfamily relationships across analyses. First, PI profiling identified the mtDNA 1$^{st}$ and 3$^{rd}$

447   codon position data subsets as problematic character sets likely compromised by homoplasy due

448   to nucleotide saturation (Fig. 5); hence, we felt justified in excluding these sites from our final

449   divergence dating analyses. Second, our results suggest that predicted phylogenetic noise of the

450   combined datasets over the subfamily divergence epoch (Fig. S9) was most likely a limiting

451   factor for resolving Cycleptinae as sister to Ictiobinae. Whereas internode distances for

452   Catostominae and Myxocyprininae crown clades were generally longer, being ~11 Myr to 18

453   Myr in length, and associated with significant posterior support, that for Cycleptinae + Ictiobinae

454   had a short internode distance of only 1 Myr (95% HPDs: [0.01,9.1]) and non-significant

455   posterior support in our time tree (Fig. 4). Together with the more frequent incongruence and

456   lower support for Cycleptinae compared to Ictiobinae across our MrBayes topologies, this

457    suggests that Cycleptinae acted as a "rogue taxon" switching positions on the tree (Aberer et al.,

458    2013). We hypothesize that our BEAST total-evidence tree inferred subfamily relationships that

459    were more consistent with Bayes factor tests, and had greater nodal support for early-diverging

460    nodes, by limiting the rogue movements of Cycleptinae.

461    *Monophyly of early-diverging sucker genera, and relationships within* Ictiobus

462    The genera *Myxocyprinus*, *Cycleptus*, *Carpiodes*, and *Ictiobus* formed well-supported clades in

463    our results. However, relationships among *Ictiobus* species were resolved in a polytomy or

464    paraphyletic grade in several cases, limiting our resolution of this clade. These *Ictiobus* results

465    disagree with previous mtDNA- or nDNA-based studies resolving relationships among *Ictiobus*

466    species with strong maximum-likelihood bootstrap support (Doosey et al., 2010), and Smith's

467    (1992) hypothesis of relationships among four *Ictiobus* species based mainly on morphology.

468    Given these findings, and that our current results fail to unquestionably place the Cycleptinae as

469    sister to the Ictiobinae, drawing phylogenetic or taxonomic conclusions about *Ictiobus* species

470    relationships would seem premature, and we recommend more in-depth analyses of these taxa.

471    *Phylogenetic relationships among the Catostominae*

472    Within the Catostominae, our multilocus and total-evidence results strongly supported sister

473    relationships between Thoburniini + Moxostomatini, and between Erimyzonini + Catostomini.

474    These findings agree well with previous molecular results (Harris et al., 2002; Clements et al.,

475    2012), except for mitochondrial trees in Doosey et al. (2010) showing the Erimyzonini as sister

476    to all other clades within Catostominae. Interestingly, however, our catostomine relationships

477    conflict with the analysis of Smith (1992), whose morphological data we re-analyzed. In Smith's

478    (1992) sucker phylogeny, the Erimyzonini is resolved as sister to a clade containing what are

479    currently regarded as the Moxostomatini and Thoburniini (Harris & Mayden, 2001; Harris et al.,

480    2002). Smith pointed out that this set of relationships was supported by >20 apomorphies that

481    changed at the node representing the MRCA of these lineages in his parsimony tree. But this

482    conclusion is only as sound as the phylogeny upon which character state transitions were mapped

483    by Smith (1992), which, at this node and several other key nodes, is rejected by our mtDNA,

484    nDNA, and multilocus trees, as well as total-evidence results from analyzing Smith's data

485    together with molecular datasets.

486        Regarding our Thoburniini + Moxostomatini clade, genus *Thoburnia* was inferred to be

487    paraphyletic based on the placement of *T. atripinnis* sister to the *Hypentelium* clade, mostly with

488    strong BPP support. Doosey et al. (2010) and Clements et al. (2012) obtained the same

489    relationship for *T. atripinnis*. However, our morphology tree resolved *Thoburnia* as

490    monophyletic, which is consistent with Smith's (1992) original analysis of the morphological

491    data we used, suggesting further data or analyses are needed to clarify these relationships. We

492    consistently inferred *Hypentelium roanokense* as sister to a clade of *H. etowanum* + *H. nigricans*,

493    across molecular, morphological, and total-evidence analyses, though with varying BPP (Figs. 3,

494    S1, S2, S4 and S5). These results match relationships inferred by Buth (1980) using isozyme data

495    reflecting variation at 40 putative loci. By contrast, our results conflict with Smith's (1992)

496    hypothesis, which resolved *H. roanokense* as sister to *H. nigricans*; however, this relationship

497    was based on a single morphological character, dermethmoid spine shape. Taking this into

498    consideration, the broad congruence between multiple data types, as well as our re-analysis of

499    Smith's (1992) data, suggests high confidence in the inference that *H. roanokense* is the earliest

500    diverging lineage in the genus. Within our Moxostomatini clade, *Moxostoma* was monophyletic

501    consistent with previous analyses (e.g. Clements et al. 2012).

502     Relationships within the Erimyzonini and Catostomini were very similar to those in

503 Harris et al. (2002) and consistent with Doosey et al. (2010), but they contradicted Smith (1992),

504 especially by resolving relationships within *Erimyzon* while placing *E. sucetta* sister to *E. tenuis*.

505 As in previous molecular results for Erimyzonini, *Erimyzon* was monophyletic and sister to

506 *Minytrema* in our results. However, Catostomini genera were not generally obtained as

507 monophyletic, and in no case was *Catostomus* monophyletic relative to *Chasmistes*, *Deltistes*, or

508 *Xyrauchen*. Our molecular and total-evidence analyses consistently resolved 9 well-supported

509 major clades within Catostominae (e.g. Figs. 3, S1 and S2). Here, we focus on relationships

510 within Clade 5, which corresponded to the former subgenus *Pantosteus*, which Unmack et al.

511 (2014) recently elevated to genus. Smith (1966) recognized six species within *Pantosteus*: *P.*

512 *clarkii*, *C. columbianus*, *P. discobolus*, *P. plebeius*, *P. platyrhynchus*, and *P. santaanae*. We

513 sampled all of these, including both subspecies of *P. discobolus*, but consistently inferred a

514 polyphyletic *Pantosteus*, with *P. nebuliferus* (recognized as distinct from *P. plebeius* by Miller et

515 al., 2005; Nelson et al., 2004) falling within Clade 5 but *C. columbianus* placed in Clade 8

516 (discussed below). A clade with *P. nebuliferus* + *P. plebeius* was frequently sister to all

517 remaining *Pantosteus* (e.g. Fig. 3). Notwithstanding incongruent results among analyses in the

518 two papers, the consensus of results from our study and those of Unmack et al. (2014) seems to

519 lend strongest support to the former relationship, with *P. santaanae* sister to a clade containing

520 *P. clarkii* and *P. discobolus* lineages. The polyphyly of *Pantosteus* and *nebuliferus–plebeius*

521 sister relationship are concordant with the results of Doosey et al.'s (2010) analyses using RY-

522 coding for third position mtDNA substitutions, although they inferred *P. santaanae* as sister to a

523 clade containing other members of *Pantosteus*. These results also agree with mitochondrial and

524 morphological analyses of Unmack et al. (2014). However, our results depart from Doosey et al.

525    (2010) and agree better with Unmack et al. (2014) in strongly supporting a sister relationship

526    between *C. columbianus* and *C. tahoensis*. That the morphological and molecular data analyzed

527    herein support the monophyly and diagnosability of *Pantosteus* relative to *Catostomus*, without

528    rendering *Catostomus* paraphyletic, strongly supports Unmack et al.'s (2014) decision to

529    redefine *Pantosteus* to exclude *C. columbianus*. We note that this taxonomic arrangement is also

530    consistent with studies on morphological and biochemical variation in western suckers (Koehn,

531    1969; Smith, 1992; Smith & Koehn, 1971). Also, *C. columbianus* has an open frontoparietal

532    fontanelle, a key diagnostic character of this clade, whereas other *Pantosteus* species have the

533    frontoparietal fontanelle closed or reduced to a narrow slit (Smith, 1966).

534    *"Catostomus" polyphyly and introgressive hybridization*

535    As noted above, *Catostomus* was never resolved in our study as monophyletic relative to

536    *Chasmistes*, *Deltistes*, or *Xyrauchen*, and this result is concordant with phylogenetic results of

537    Doosey et al. (2010) based on mtDNA *ND4/ND5* sequences. Hybridization of *Catostomus* with

538    *Chasmistes*, *Deltistes*, and *Xyrauchen* is well documented (Buth et al., 1987; Markle et al., 2005;

539    Mock et al., 2006; Tranah & May, 2006), and may be related to the non-monophyly of

540    *Catostomus* relative to *Chasmistes* and *Deltistes*. However, while *Xyrauchen texanus* has been

541    documented to hybridize with *C. latipinnis* and *C. insignis* (Buth et al., 1987; Hubbs & Miller,

542    1953), the majority of these reports evaluate hybridization between *X. texanus* and *C. latipinnis*

543    (Buth et al., 1987). Samples of *Xyrauchen* used in this study originated from the Dexter National

544    Fish Hatchery, which obtained the original hatchery stock of Razorback sucker from Lake

545    Mohave, Arizona, where hybridization with *C. latipinnis* has been documented but allozyme

546    evidence indicates only low levels of introgression of *C. latipinnis* with *X. texanus* (Buth et al.,

547    1987). Therefore, placement of *C. insignis* sister to *X. texanus* here and by Doosey et al. (2010)

548    suggests that introgression is not a factor in either study. As such, "*Xyrauchen*" embedded within

549    *Catostomus* renders the latter polyphyletic. Even if hybridization-mediated introgression were

550    considered as an ad hoc explanation of this pattern, this is difficult to distinguish from the more

551    parsimonious hypothesis of common ancestry, and the available data do not demonstrate that any

552    hybridization events among these taxa have corresponded to the Neogene–present timeframe of

553    their divergences inferred by our time tree. Thus, we advocate the tentative placement of

554    "*Xyrauchen*" into synonymy with *Catostomus* until additional fossil or molecular evidence

555    rejects an inference of common ancestry in favor of Neogene hybridization of these taxa.

556    *Bayesian total-evidence dating and relaxed-clock partitioning*

557    In showing that the best clock-partitioned BEAST models excluded sites identified as

558    problematic in our PI profiling analysis, the results of our Bayes factor clock-partitioning model

559    comparisons bolster Ho and Lanfear's (2010) recommendation that accounting for differences in

560    substitution rates among data partitions through clock-partitioning is not only feasible but also

561    improves phylogenetic divergence dating models. We believe that by employing a clock-

562    partitioning scheme objectively chosen in this way allowed our final BEAST FBD analysis to

563    more correctly estimate topology and rate variance among branches, and better handle rate

564    heterogeneity of the retained characters. However, while the inclusion of fossil taxa in an FBD

565    model in the final BEAST analysis certainly improved our divergence time estimates over what

566    might be obtained using node calibration or tip-dating methods (e.g. Arcila et al., 2015;

567    Gavryushkina et al., 2017), one limitation of this analysis was that nodal support was reduced

568    within the Catostomini and Ictiobinae. This pattern was caused by rogue placements of fossil

569    taxa lacking character data, which were constrained within these crown groups but made up

570    $\geq 50\%$ of tip sampling (Fig. 4). Still, this mainly caused misleading relationships and lowered

571 nodal support within the Catostomini; after removing fossil taxa, relationships within Ictiobinae

572 would be essentially identical to our preferred MrBayes topologies. After pruning extinct taxa,

573 our time tree will provide a suitable basis for interrogating the comparative biogeography and

574 evolution of all groups of suckers, except for patterns within Catostomini. One alternate way

575 forward for researchers interested in using our results for comparative phylogenetics would be to

576 convert one of our preferred topologies (Figs. 3 and S2) to an ultrametric tree while constraining

577 subfamily and tribal node ages to mean $t_{MRCA}$ estimates shown in Fig. 4.

578     A major goal of our study was to use our final total-evidence dating results to test

579 hypotheses on the temporal diversification of suckers. Our divergence dating results (Fig. 4)

580 generally agree with the fossil record but reject or confirm different molecular hypotheses about

581 the temporal diversification of sucker subfamilies. Unsurprisingly, given our incorporation of all

582 fossil sucker lineages in the paleontological literature under an FBD model (accounting for

583 extant and fossil sampling levels), our BEAST results strongly support hypothesis $H_4$ that

584 Catostomidae lineages have diversified since ~50 Ma in the Early Eocene, which is widely

585 accepted as the minimum age of the origin of suckers based on stratigraphic information for the

586 oldest sucker fossils (review and refs. in Appendix S1). Our results also support Sun et al.'s

587 (2007) proposal, or our hypothesis $H_3$, that catostomine lineages in the most speciose sucker

588 subfamily went on to diversify since ~20 Ma. Indeed, initial divergences and subsequent

589 diversification of all four catostomine tribes has proceeded since around ~34–17.6 Ma in the

590 Eocene–Miocene, with 95% credible intervals ranging from 43 to 11.04 Ma (Fig. 4), and the

591 $t_{MRCA}$s for ~81% (64/79) of extant species/lineages in our time tree (all catostomines) coincide

592 with the last 20 million years.

593     In contrast to hypotheses $H_3$ and $H_4$ discussed above, we reject two previous molecular

594   hypotheses about the tempo of sucker evolution advanced by Sun et al. (2007). First, we reject

595   hypothesis $H_1$ because we infer that the Asian Myxocyprininae diverged from North American

596   suckers during the Late Cretaceous, and the 95% credible intervals for this divergence do not

597   overlap with their proposed ~14 Ma Miocene date for the MRCA of *Myxocyprinus* and

598   *Cycleptus*. Second, we reject $H_2$ given that we infer an early Eocene origin for Ictiobinae,

599   including the extinct †*Amyzon* and †*Vasnetzovia* ictiobine lineages, and this vastly predates Sun

600   et al.'s (2007) proposed origin of the clade. Given Sun et al. (2007) produced divergence time

601   estimates using only cyt*b* divergences and a global molecular clock assuming a 2.0% $Myr^{-1}$

602   pairwise rate for vertebrate mtDNA, there are too many methodological distinctions between our

603   approach and theirs to pinpoint a single factor causing our results to contrast theirs so strongly.

604   However, our more comprehensive and nuanced approach using Bayesian total-evidence dating

605   not only allowed us to use a realistic FBD tree prior incorporating the speciation-extinction-

606   fossilization sampling process (Gavryushkina et al., 2017), but also permitted estimation of

607   evolutionary rates for each character subset analyzed. We inferred a slower rate of evolution for

608   mtDNA $2^{nd}$ position sites, $6.53 \times 10^{-4}$ substitutions $site^{-1}$ $Myr^{-1}$, and in fact all DNA subsets

609   (mean: 0.0055 substitutions $site^{-1}$ $Myr^{-1}$), than the global rate applied by Sun et al. (2007). This

610   is partly due to our best model excluding the most variable mtDNA codon sites, but it appears

611   that differences between our $t_{MRCA}$ estimates and theirs are not fully accounted for based on

612   substitution rates alone. Nevertheless, our more appropriate modeling of the evolutionary

613   processes producing variation in sucker DNA sequences and morphological characters, and

614   extant and fossil taxon sampling, has allowed us to estimate older and undoubtedly more

615   accurate divergence dates, especially for deeper nodes in the sucker phylogeny.

616       We infer divergence times for major sucker lineages that are conspicuously older than

617    those recently estimated from multilocus analyses of other North American freshwater fish

618    clades, including sunfishes and black basses (Centrarchidae; Near et al., 2005, 2011), bullhead

619    and madtom catfishes (Ictaluridae; Hardman & Hardman, 2008), and darters (Etheostomatinae;

620    Near et al., 2011). Whereas the diversification of these major lineages has occurred since around

621    the Eocene–Oligocene transition ~34 Ma, a time of global cooling (Zachos et al., 2001), we infer

622    an earlier Late Cretaceous–Eocene age for the onset of sucker subfamily divergences. While tip-

623    dating approaches can lead to older divergence dates (e.g. O'Reilly et al. 2015), we believe that

624    this result accurately reflects the relatively longer timeframe of sucker evolution captured by the

625    fossil record. This general timeframe for sucker evolution also correlates well to the Late

626    Paleocene Thermal Maximum, a period of greater ambient and sea-surface temperatures, higher

627    sea levels, and higher precipitation and humidity (Zachos et al., 2001). Sucker diversification

628    thus appears to have initiated during a period of climate change and sea level rise, which may

629    have facilitated the isolation of ancestral sucker populations. Our results also suggest that

630    approximately 4 to 7 sucker genera may have been present in North America by the Oligocene, a

631    period coinciding with the arrival of minnows in the family Cyprinidae on the continent based on

632    the broader fossil record of North American teleost fishes (e.g. Cavender 1986). Together with

633    the molecular results from other studies above, this implies that the subsequent diversification of

634    these genera, including at least *Cycleptus*, †*Amyzon*, *Ictiobus*, and *Carpiodes* as well as the

635    speciose Catostominae, would have coincided with the diversification of most other major

636    lineages of North American freshwater fishes.

637    **Conclusions**

638   We have presented the results of a phylogenetic analysis of Holarctic sucker fishes (family

639   Catostomidae) drawing on the most comprehensive dataset to date and inferring, separately and

640   jointly, the phylogeny and divergence times of suckers while including fossil taxa as tips. Our

641   molecular and total-evidence results corroborated relationships hypothesized in previous

642   molecular studies and yielded evidence in favor of some new hypotheses of relationships within

643   and among subfamilies, for example with Bayes factor support for Myxocyprininae sister to all

644   other sucker lineages. Our study also highlights how using PI profiling to identify problematic

645   character sets can subsequently improve or provide additional evidence for clock-partitioning

646   scheme choice during Bayesian relaxed-clock divergence dating. Our divergence-dating results

647   strongly supported the hypotheses that Catostomidae lineages have diversified since ~50 Ma in

648   the Early Eocene (Uyeno & Smith, 1972), and that tribes within the most speciose subfamily,

649   Catostominae, have diversified since ~20 Ma in the Eocene–Miocene (Smith, 1992; Sun et al.,

650   2007). Moreover, we hypothesized that incongruent subfamily relationships were driven in part

651   by problematic mtDNA 1st and 3rd codon sites, and by "rogue taxon" movements of Cycleptinae

652   and fossil taxa, for example in our FBD process time tree. Our analysis could be extended to test

653   this latter hypothesis using additional statistical analyses of rogue taxa (e.g. Aberer et al., 2014)

654   and internode uncertainty (Zhou et al., 2017), and by additional resolution analyses employing

655   Monte Carlo simulations and tests of their assumptions (Townsend et al., 2012), which were

656   beyond the scope of the present study. Nevertheless, our results suggest that future studies of

657   suckers will benefit from using PI profiles as a predictive tool to select loci for subsequent

658   phylogenetic analyses (Dornburg et al., 2014).

659   **Acknowledgements**

671    **References**

672    Aberer AJ, Krompass D, Stamatakis A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an
673           efficient algorithm and webservice. *Systematic Biology* **62:**162-166.
674    Agassiz L. 1854. Notice of a collection of fishes from the southern bend of the Tennessee River, in the
675           State of Alabama. *American Journal of Science and Arts 28 second series* **17:**297-308,353-365.
676    Arcila D, Pyron RA, Tyler JC, Ortí G, Betancur-RR. 2015. An evaluation of fossil tip-dating versus node-
677           age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics*
678           *and Evolution* **82:**131-145.
679    Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the
680           accuracy of demographic and molecular clock model comparison while accommodating
681           phylogenetic uncertainty. *Molecular Biology and Evolution* **29:**2157-2167.
682    Bagley JC. 2017. justincbagley/PIrANHA: PIrANHA version 0.1.4 [Data Set]. *Zenodo.* Available at
683           http://doi.org/10.5281/zenodo.571185.
684    Bouckaert R, Heled J, Künert D, Vaughan TG, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ.
685           2014. BEAST2: a software platform for Bayesian evolutionary analysis. *PLoS Computational*
686           *Biology* **10:**e1003537.

687     Brandley MC, Wang Y, Guo X, Montes de Oca AN, Fería-Ortíz M, Hikida T, Ota H. 2011.
688             Accommodating heterogeneous rates of evolution in molecular divergence dating methods: an
689             example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. *Systematic*
690             *Biology* **60:**3-15.
691     Bruner JC. 1991. Comments on the genus *Amyzon* (family Catostomidae). *Journal of Paleontology*
692             **65:**678-686.
693     Buth DG. 1980. Genetics and systematic relationships in the catostomid genus *Hypentelium*. *Copeia*
694             **1980:**280-290.
695     Buth DG, Murphy RW, Ulmer L. 1987. Population differentiation and introgresseve hybridization of the
696             flannelmouth sucker and of hatchery and native stocks of the razorback sucker. *Transactions of*
697             *the American Fisheries Society* **116:**103-110.
698     Cavender T. 1986. Review of the fossil history of North American freshwater fishes. In: Hocutt CH,
699             Wiley EO, eds. *The Zoogeography of North American Freshwater Fishes*. New York: John Wiley
700             and Sons. 699-724
701     Chang M, Miao D, Chen Y, Zhou J, Pingfu C. 2001. Suckers (Fish, Catostomidae) from the Eocene of
702             China account for the family's current disjunct distributions. *Science in China Series D-Earth*
703             *Sciences* **44:**577-586.
704     Chen W-J, Mayden RL. 2012. Phylogeny of suckers (Teleostei: Cypriniformes: Catostomidae): further
705             evidence of relationships provided by the single-copy nuclear gene IRBP2. *Zootaxa* **3586:**195-
706             210.
707     Chen W-J, Miya M, Saitoh K, Mayden RL. 2008. Phylogenetic utility of two existing and four novel
708             nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes
709             (Ostariophysi) as a case study. *Gene* **423:**125-134.
710     Chow S, Takeyama H. 1998. Intron length variation observed in the creatine kinase and ribosomal protein
711             genes of the swordfish *Xiphias gladius*. *Fisheries Science* **64:**397-402.
712     Clarke JA, Middleton K. 2008. Mosaicism, modules, and the evolution of birds: results from a Bayesian
713             approach to the study of morphological evolution using discrete character data. *Systematic*
714             *Biology* **57:**185-201.
715     Clements MD, Bart Jr. HL, Hurley DL. 2012. A different perspective on the phylogenetic relationships of
716             the Moxostomatini (Cypriniformes: Catostomidae) based on cytochrome-*b* and Growth Hormone
717             intron sequences. *Molecular Phylogenetics and Evolution* **63:** 159-167.
718     Doosey M, Bart H, Saitoh K, Miya M. 2010. Phylogenetic relationships of catostomid fishes
719             (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular*
720             *Phylogenetics and Evolution* **54:**1028-1034.

721     Dornburg A, Townsend JP, Friedman M, Near TJ. 2014. Phylogenetic informativeness reconciles ray-
722         finned fish molecular divergence times. *BMC Evolutionary Biology* **14:**169.

723     Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the
724         BEAST 1.7. *Molecular Biology and Evolution* **29:**1969-1973.

725     Ferris SD, Whitt GS. 1978. Phylogeny of tetraploid catostomid fishes based on the loss of duplicate gene
726         expression. *Systematic Zoology* **27:**189-206.

727     Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2017. Bayesian total-
728         evidence dating reveals the recent crown radiation of penguins. *Systematic Biology* **66:**57-73.

729     Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled ancestor trees
730         for epidemiology and fossil calibration. *PLoS Computational Biology* **10(12):**e1003919.

731     Hardman M, Hardman LM. 2008. The relative importance of body size and paleoclimatic change as
732         explanatory variables influencing lineage diversification rate: an evolutionary analysis of
733         bullhead catfishes (Siluriformes: Ictaluridae). *Systematic Biology* **57:**116-130.

734     Harris PM, Hubbard G, Sandel M. 2014. Catostomidae: Suckers. In: Warren ML, Burr BM, eds.
735         *Freshwater fishes of North America: Volume 1: Petromyzontidae to Catostomidae*. Baltimore,
736         MD: Johns Hopkins University Press. 451-502

737     Harris PM, Mayden RL. 2001. Phylogenetic relationships of major clades of Catostomidae (Teleostei:
738         Cypriniformes) as inferred from mitochondrial SSU and LSU rDNA sequences. *Molecular
739         Phylogenetics and Evolution* **20:**225-237.

740     Harris PM, Mayden RL, Perez HE, Garcia de Leon F. 2002. Phylogenetic relationships of Redhorse
741         (*Moxostoma*) and Jumprock (*Scartomyzon*) suckers (Cypriniformes: Catostomidae) based on
742         mitochondrial cytochrome *b* sequence data. *Journal of Fish Biology* **61:**1433-1452.

743     Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses.
744         *Journal of Systematics and Evolution* **46:**239-257.

745     Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth–death process for coherent calibration of
746         divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States
747         of America* **111:**E2957-E2966.

748     Ho SY, Lanfear R. 2010. Improved characterisation of among-lineage rate variation in cetacean
749         mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA* **21:**138-146.

750     Hubbs CL. 1930. Materials for a revision of the catostomid fishes of eastern North America.
751         *Miscellaneous Publications of the Museum of Zoology of the University of Michigan*, No. 20.

752     Hubbs CL, Miller RR. 1953. Hybridization in nature between the fish genera *Catostomus* and *Xyrauchen*.
753         *Papers of the Michigan Academy of Science, Arts and Letters* **38:**207-234.

754     Jenkins RE. 1970. *Systematic studies of the catostomid fish tribe Moxostomatini*. Ithaca, NY: Cornell

755        University.

756    Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* **90:**773-795.

757    Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program.
758        *Bioinformatics* **9:**286-298.

759    Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz
760        S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. 2012. Geneious Basic: an integrated
761        and extendable desktop software platform for the organization and analysis of sequence data.
762        *Bioinformatics* **28:**1647-1649.

763    Kettratad J, Markle DF. 2010. Redescription of the Tyee Sucker, *Catostomus tsiltcoosensis*
764        (Catostomidae). *Western North American Naturalist* **70:**273-287.

765    Koehn RK. 1969. Hemoglobins of fishes of the genus *Catostomus* from western North America. *Copeia*
766        **1969:**21-30.

767    Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning
768        schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29:**
769        1695-1701.

770    Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character
771        data. *Systematic Biology* **50:**913-925.

772    López-Giráldez F, Townsend JP. 2011. PhyDesign: an online application for profiling phylogenetic
773        informativeness. *BMC Evolutionary Biology* **11:**152.

774    Markle DF, Cavalluzzi MR, Simon DC. 2005. Morphology and taxonomy of Klamath Basin suckers
775        (Catostomidae). *Western North American Naturalist* **65:**473-489.

776    Mayden RL, Chen WJ, Bart Jr. HL, Doosey MH, Simons AM, Tang KL, Wood RM, Agnew MK, Yang
777        L, Hirt MV, Clements MD, Saitoh K, Sado T, Miya M, Nishida M. 2009. Reconstructing the
778        phylogenetic relationships of the earth's most diverse clade of freshwater fishes—order
779        Cypriniformes (Actinopterygii: Ostariophysi): a case study using multiple nuclear loci and the
780        mitochondrial genome. *Molecular Phylogenetics and Evolution* **51:**500-514.

781    Mayden RL, Tang KL, Wood RM, Chen W.J, Agnew MK, Conway KW, Yang L, Simons AM, Bart HL,
782        Harris PM, Li J, Wang X, Saitoh K, He S, Liu H, Chen Y, Mutsumi N, Miya M. 2008. Inferring
783        the Tree of Life of the order Cypriniformes, the earth's most diverse clade of freshwater fishes.
784        *Journal of Systematics and Evolution* **46:**424-438.

785    Miller RR. 1959. Origin and affinities of the freshwater fish fauna of western North America. In: Hubbs
786        CL, ed. Zoogeography. Washington, D.C.: American Association for the Advancement of
787        Science Publication 51. 187-222

788    Miller RR, Minckley WL, Norris SM. 2005. *Freshwater fishes of Mexico*. Chicago, IL: University of

789   Chicago Press.

790 Mock KE, Evans RP, Crawford M, Cardall BL, Janecke SU, Miller MP. 2006. Rangewide molecular

791   structuring in the Utah sucker (*Catostomus ardens*). *Molecular Ecology* **15:**2223-2238.

792 NatureServe, 2013. *Moxostoma lacerum*. The IUCN Red List of Threatened Species. Version 2015.2.

793   <<u>www.iucnredlist.org</u>>. Last accessed on 01 August 2015.

794 NatureServe, 2014. *Chasmistes cujus*. Red List of Threatened Species. Version 2015.2.

795   <<u>www.iucnredlist.org</u>>. Last accessed on 01 August 2015.

796 Near TJ, Bolnick DI, Wainwright PC. 2005. Fossil calibrations and molecular divergence time estimates

797   in centrarchid fishes (Teleostei: Centrarchidae). *Evolution* **59:**1768-1782.

798 Near TJ, Bossu CM, Bradburd GS, Carlson RL, Harrington RC, Hollingsworth Jr. PR, Keck BP, Etnier

799   DA. 2011. Phylogeny and temporal diversification of darters (Percidae: Etheostomatinae).

800   *Systematic Biology* **60:**565-595.

801 Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith

802   WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of*

803   *the National Academy of Sciences of the United States of America* **109:**13698-13703.

804 Nelson JS, Crossman EJ, Espinosa-Perez H, Findley LT, Gilbert CR, Lea RN, Williams JD. 2004.

805   *Common and Scientific Names of Fishes from the United States, Canada, and Mexico*. Bethesda,

806   MD: American Fisheries Society Special Publication 29.

807 O'Reilly JE, dos Reis M, Donoghue PC. 2015. Dating tips for divergence-time estimation. *Trends in*

808   *Genetics*, **31:**637-650.

809 Pagel M, Meade A. 2014. BayesTraits (1st edition). Reading, UK. Available at

810   http://www.evolution.rdg.ac.uk/BayesTraits.html.

811 Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*

812   **21:**676-679.

813 Pyron RA. 2011. Divergence time estimation using fossils as terminal taxa and the origins of

814   Lissamphibia. *Systematic Biology* **60:**466-481.

815 Rambaut A, Drummond AJ. 2014. Tracer v1.5. Available at http://beast.bio.edu.ac.uk/Tracer.

816 Robins CR, Raney EC. 1956. Studies of the catostomid fishes of the genus *Moxostoma*, with descriptions

817   of two new species. Cornell University Agricultural Experiment Station Memoir 343, 1-56.

818 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA,

819   Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice

820   across a large model space. *Systematic Biology* **61:**539-542.

821 Sagahi-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length

822   polymorphism in barley: Mendelian inheritance, chromosomal location, and population

823        dynamics. *Proceedings of the National Academy of Sciences of the United States of America*

824        **81:**8014-8019.

825    Saitoh K, Sado T, Mayden RL, Hanzawa N, Nakamura K, Nishida M, Miya M. 2006. Mitogenomic

826        evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first

827        evidence toward resolution of higher-level relationships of the world's largest freshwater fish

828        clade based on 59 whole mitogenome sequences. *Journal of Molecular Evolution* **63:**826-841.

829    Smith GR. 1966. Distribution and evolution of the North American catostomid fishes of the subgenus

830        *Pantosteus*, genus *Catostomus*. Miscellaneous Publications of the Museum of Zoology of the

831        University of Michigan **129:**1-132.

832    Smith GR. 1992. Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America

833        and Asia. In: Mayden RL, ed. *Systematics, historical ecology, and North American freshwater*

834        *fishes*. Stanford, CA: Stanford University Press. 778-813

835    Smith GR, Dowling TE, Gobalet KW, Lugaski T, Shiozawa DK, Evans RP. 2002. Biogeography and

836        timing of evolutionary events among Great Basin fishes. In: Hershler R, Madsen DB, Currey DR,

837        eds. *Great Basin aquatic systems history*. Washington, D.C.: Smithsonian Institution Press. 175-

838        234

839    Smith GR, Koehn RK. 1971. Phenetic and cladistic studies of biochemical and morphological

840        characteristics of *Catostomus*. *Systematic Zoology* **20:**282-297.

841    Sun YH, Xie CX, Wang WM, Liu SY, Treer T, Chang MM. 2007. The genetic variation and

842        biogeography of catostomid fishes based on mitochondrial and nucleic DNA sequences. *Journal*

843        *of Fish Biology* **70:**291-309.

844    Tranah GJ, May B. 2006. Patterns of intra- and interspecies genetic diversity in Klamath River Basin

845        suckers. *Transactions of the American Fisheries Society* **135:**306-316.

846    Townsend JP. 2007. Profiling phylogenetic informativeness. *Systematic Biology* **56:**222-231.

847    Townsend JP, Leuenberger C. 2011. Point of View: Taxon sampling and the optimal rates of evolution

848        for phylogenetic inference. *Systematic Biology* **60:**358-365.

849    Townsend JP, López-Giráldez F, Friedman R. 2008. The phylogenetic informativeness of nucleotide and

850        amino acid sequences for reconstructing the vertebrate tree. *Journal of Molecular Evolution*

851        **67:**437-447.

852    Townsend JP, Su Z, Tekle Y. 2012. Phylogenetic signal and noise, predicting the power of a dataset to

853        resolve phylogeny. *Systematic Biology* **5:**835-849.

854    Unmack PJ, Dowling TE, Laitinen NJ, Secor CL, Mayden RL, Shiozawa DK, Smith GR. 2014. Influence

855        of introgression and geological processes on phylogenetic relationships of western North

856        American mountain suckers (*Pantosteus*, Catostomidae). *PLoS One* **9:**e90061.

857   Uyeno T, Smith GR. 1972. Tetraploid origin of the karyotype of catostomid fishes. *Science* **175:**644-646.

858   Warren Jr. ML, Burr BM, Walsh SJ, Bart HL, Cashner RC, Etnier DA, Freeman BJ, Kuhajda BR,

859         Mayden RL, Robison HW, Ross ST. 2000. Diversity, distribution, and conservation status of the

860         native freshwater fishes of the southern United States. *Fisheries* **25:**7-31.

861   Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011. Improving marginal likelihood estimation for

862         Bayesian phylogenetic model selection. *Systematic Biology* **60:**150-160.

863   Zhou X, Lutteropp S, Czech L, Stamatakis A, von Looz M, Rokas A. 2017. Quartet-based computations

864         of internode certainty provide accurate and robust measures of phylogenetic incongruence.

865         *bioRxiv* 168526.

**Table 1**(on next page)

Summary of post-systematic phylogenetic and taxonomic studies of suckers in the family Catostomidae

Abbreviations: ML, maximum likelihood; MP, maximum parsimony; *n*, sample size, including ingroup and outgroup taxa or lineages; UPGMA, unweighted pair-group method with arithmetic mean; Wagner, Wagner method for inferring 'most parsimonious' tree.

1

| Study | *n* | Data type | Analysis | Supported clades/relationships |
|---|---|---|---|---|
| Ferris & Whitt (1978) | 30 | 20 isozyme loci (loss of duplicate gene expression) | Wagner | Ictiobinae + (Cycleptinae + Catostominae), three catostomin tribes |
| Smith (1992) | 64 | 157 morphological, biochemical and early life-history transformation series | MP | Ictiobinae + (Cycleptinae + Catostominae), *Moxostoma* in paraphyletic grade with "*Scartomyzon*" within Moxostomatini; also trichotomy of *M. ariommum* + *Thoburnia* + *Hypentelium* |
| Harris & Mayden (2001) | 16 | mtDNA SSU and LSU rDNA sequences | MP | four monophyletic subfamilies, Thoburniini resurrected and includes *Hypentelium*; Catostomini + Erimyzonini and Moxostomatini + Thoburniini |
| Harris & Mayden (2002) | 50 | mtDNA cytochrome *b* (cyt*b*) gene | MP and ML | four monophyletic subfamilies, provisional Thoburniini; "*Scartomyzon*" as a junior synonym of *Moxostoma* |
| Sun et al. (2007) | 17 | mtDNA cyt*b* gene | UPGMA | monophyletic Ictiobinae and Catostominae; Moxostomatini + Catostomini |
| Doosey et al. (2010) | 60 | mtDNA NADH subunit 4 and 5 (ND4/5) sequences | ML | ((Cycleptinae, (Myxocyprininae, Ictiobinae)), Catostominae); also among catostomins: Erimyzonini + (Moxostomatini + Thoburniini) |
| Chen & Mayden (2012) | 67 | interphotoreceptor retinoid-binding protein (IRBP) sequences | ML | supported previous relationships including monophyly of subfamilies and tribes except Thoburniini |
| Clements et al. (2012) | 45 | growth hormone intron (GHI) | ML | para-/polyphyletic *Moxostoma* and *Scartomyzon* |
| Unmack et al. (2014) | 24 | morphology, plus cyt*b* and other mtDNA gene sequences | MP, ML, and Bayesian | monophyletic *Pantosteus* within Catostomini, except for *Catostomus* (*P.*) *columbianus* (potential hybrid origin) |

2

## Table 2(on next page)

List of sequences used in the present study including museum/field numbers and GenBank accession numbers.

Sequences generated during this study are shown in bold. Dash symbol, '–', indicates no sequence was available.

[a] Taxonomic classification of suckers used here is based on Harris & Mayden (2001) and Harris et al. (2002).

[b] Institutional, project, and individual abbreviations: BRK, Bernard R. Kuhajda; CToL, Cypriniformes Tree of Life Project; DAN, David A. Neely; JEB, James E. Brooks; JFBM, (James Ford) Bell Museum of Natural History, University of Minnesota; HLB, Henry L. Bart, Jr.; MSB, Museum of Southwest Biology; PBB, Peter B. Berendzen; UAIC, University of Alabama Ichthyological Collection; USU, Utah State University; UNL, University of Nebraska Lincoln; OS, Oregon State University; BYU, Monte L. Bean Life Science Museum, Brigham Young University; Louis Bernatchez, Université Laval; NCSM, North Carolina State Museum of Natural Sciences; and TU, Tulane University Biodiversity Research Institute (formerly Tulane University Museum of Natural History).

| Taxon [a] | Museum / field numbers [b] | cyt*b* | ND2 | *cox1* | IRBP | *RPS7* | GHI |
|---|---|---|---|---|---|---|---|
| Order Cypriniformes | | | | | | | |
| Family Cyprinidae | | | | | | | |
| Subfamily Cyprininae | | | | | | | |
| *Cyprinus carpio* | CToL 1751 | AP009047 | AP009047 | EU524006 | FJ197101 | DQ163924 | FJ265047 |
| Family Gyrinocheilidae | | | | | | | |
| *Gyrinocheilus aymonieri* | CToL 612 | NC008672 | NC008672 | JF915620 | JX470019 | – | FJ265031 |
| Family Cobitidae | | | | | | | – | – |
| Subfamily Cobitinae | | | | | | | |
| *Cobitis striata* | CToL 230 | NC004695 | NC004695 | – | Mayden et al. (2009; K. Saitoh, unpublished) | – | – |
| Subfamily Botinae | | | | | | | |
| *Chromobotia macracanthus* | CToL 217 | AC024175 | AC024175 | KF738207 | FJ197086 | – | – |
| *Leptobotia mantschurica* | – | AB242170 | AB242170 | AB242170 | FJ197087 | – | FJ265035 |
| Family Catostomidae | | | | | | | |
| Subfamily Myxocyprininae | | | | | | | |
| *Myxocyprinus asiaticus* 1 | – | NC006401 | NC006401 | AY526869 | – | – | – |
| *M. asiaticus* 2 | – | AY526869 | AY526869 | – | – | – | – |
| **M. asiaticus 3** | UAIC 11698.01 | **JX488760** | **JX488826** | – | **JX488937** | **This study** | **FJ265052** |
| Subfamily Ictiobinae | | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Carpiodes carpio* 1 | – | AY366087 | AY366087 | – | – | – | – |
| **Car. carpio** 2 | UAIC 11219.08 | AF454867 | **JX488827** | JN024866 | – | – | JF837387 |
| **Car. cyprinus** | TU 157.01 | **JX488761** | **JX488828** | EU523924 | **JX488938** | – | GU937849 |
| **Car. velifer** | TU 108.16 | **JX488762** | **JX488829** | JN024878 | **JX488939** | – | JF837435 |
| **Ictiobus bubalus** 1 | TU 124.07 | **JX488763** | **JX488830** | KF929996 | **JX488940** | – | JF799533 |
| **I. bubalus** 2 | TU 120.01 | **JX488764** | **JX488831** | – | **JX488941** | – | – |
| **I. bubalus** 3 | TU 244.03 | **JX488765** | **JX488832** | – | **JX488942** | – | – |
| **I. cyprinellus** 1 | TU 101.05 | **JX488766** | **JX488833** | EU524687 | – | – | – |
| **I. cyprinellus** 2 | TU 107.15 | **JX488767** | **JX488834** | – | **JX488943** | – | GU937840 |
| **I. niger** 1 | TU 121.01 | **JX488768** | **JX488835** | EU524107 | **JX488944** | – | FJ226251 |
| **I. niger** 2 | TU 124.06 | **JX488769** | **JX488836** | – | **JX488945** | – | – |
| Subfamily Cycleptinae | | | | | | – | – |
| *Cycleptus elongatus* 1 | – | NC008645 | NC008645 | – | – | – | – |
| **Cyc. elongatus** 2 | UAIC 11371.01 | AF454868 | **JX488837** | KF929801 | **JX488946** | – | FJ265028 |
| **Cyc. meridionalis** | UNL 1-30 | **JX488770** | **JX488838** | – | **JX488947** | – | – |
| Subfamily Catostominae | | | | | | | |
| Tribe Catostomini | | | | | | | |
| **Catostomus ardens** 1 | USU-UBL | **JX488771** | **JX488839** | JN024886 | **JX488948** | – | – |
| **C. ardens** 2 | USU-USD | **JX488772** | **JX488840** | – | **JX488949** | – | – |
| **C. bernardini Río Papogochic** | BRK0507-2 | **JX488773** | **JX488841** | EU668462 | **JX488950** | – | – |
| **C. bernardini Río Batopilas 1** | JEB0504-2 | **JX488774** | **JX488842** | – | **JX488951** | – | – |
| **C. bernardini Río Batopilas 2** | JEB0505-7 | **JX488775** | **JX488843** | – | **JX488952** | – | – |
| **C. bernardini Río Yecora** | JEB0510-1 | **JX488776** | **JX488844** | – | **JX488953** | – | – |
| **C. cahita** 1 | BRK0511-6 | **JX488777** | **JX488845** | EU668375 | **JX488954** | **This study** | – |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *C. cahita* **2** | BRK0511-8 | **JX488778** | **JX488846** | – | **JX488955** | **This study** | – |
| *C. catostomus* | – | – | – | W.J. Chen | W.J. Chen | – | – |
| *C. catostomus* | UAIC 11237.04 | AF454871 | **JX488847** | EU523925 | – | – | GU937824 |
| *C. columbianus* | OS 17548 | **JX488780** | **JX488849** | – | **JX488957** | – | – |
| *C. commersoni* 1 | – | NC008647 | NC008647 | – | – | **This study** | – |
| *C. commersoni* **2** | UAIC 11156.03 | **JX488781** | **JX488850** | EU523931 | **JX488958** | **This study** | JF799535 |
| *C. fumeiventris* **1** | – | **JX488784** | **JX488853** | – | **JX488961** | – | – |
| *C. fumeiventris* **2** | – | **JX488785** | **JX488854** | – | **JX488962** | – | – |
| *C. insignis* | MSB 49603-2 | **JX488786** | **JX488855** | – | **JX488963** | – | – |
| *C. insignis* | MSB 49603-4 | **JX488787** | **JX488856** | HQ556974 | **JX488964** | **This study** | – |
| *C. latipinnis* **1** | MSB 49601 | **JX488788** | **JX488857** | – | **JX488965** | **This study** | – |
| *C. latipinnis* **2** | MSB 49602 | **JX488789** | **JX488858** | – | **JX488966** | – | – |
| *C. leopoldi* | JEB15-1 | **JX488790** | **JX488859** | EU668394 | **JX488967** | – | – |
| *C. macrocheilus* | OS 15886 | **JX488791** | **JX488860** | EU523932 | **JX488968** | – | – |
| *C. microps* | OS 17571-1 | **JX488792** | **JX488861** | HQ557219 | **JX488969** | **This study** | – |
| *C. microps* | OS 17571-2 | **JX488793** | **JX488862** | – | **JX488970** | – | – |
| *C. occidentalis humboldtianus* **1** | OS 15623-1 | **JX488795** | **JX488864** | – | **JX488972** | – | – |
| *C. occidentalis humboldtianus* **2** | OS 15623-2 | **JX488796** | **JX488865** | – | **JX488973** | – | – |
| *C. occidentalis lacusanserinus* | OS 1943BB | **JX488797** | **JX488866** | – | **JX488974** | – | – |
| *C. occidentalis mniotiltus* **1** | UAIC 13448.01-2 | **JX488798** | **JX488867** | – | **JX488975** | – | – |
| *C. occidentalis mniotiltus* **2** | UAIC 13448.01-4 | **JX488799** | **JX488868** | – | **JX488976** | – | – |
| *C. occidentalis occidentalis* | UAIC 11546.02-1 | **JX488800** | **JX488869** | KF558280 | **JX488977** | **This study** | – |
| *C. rimiculus* | OS 15908 | AF454875 | **JX488874** | – | **JX488982** | – | – |
| *C. snyderi* | OS 15900 | **JX488806** | **JX488876** | – | **JX488984** | – | – |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *C. tahoensis* 1 | BuffaloCrk4 | **JX488807** | **JX488877** | JN024954 | **JX488985** | **This study** | – |
| *C. tahoensis* 2 | SmokeCrk1 | **JX488808** | **JX488878** | – | **JX488986** | – | – |
| *C. tsiltcoosensis* 1 | OS X113 | **JX488809** | **JX488879** | – | **JX488987** | – | – |
| *C. tsiltcoosensis* 2 | OS X114 | **JX488810** | **JX488880** | – | **JX488988** | – | – |
| *C. warnerensis* 1 | OS 14251-2 | **JX488814** | **JX488884** | – | **JX488992** | **This study** | – |
| *C. warnerensis* 2 | OS 14251-3 | **JX488815** | **JX488885** | – | **JX488993** | – | – |
| *C. wigginsi* | JEB0511-1 | **JX488816** | **JX488886** | EU668440 | **JX488994** | – | – |
| *Catostomus* **sp. Coquille River** | OS X130 | **JX488811** | **JX488881** | – | **JX488989** | – | – |
| *Catostomus* **sp. Wall Canyon 1** | OS X61 | **JX488812** | **JX488882** | – | **JX488990** | – | – |
| *Catostomus* **sp. Wall Canyon 2** | OS X67 | **JX488813** | **JX488883** | – | **JX488991** | – | – |
| *Chasmistes brevirostris* | OS 15963 | **JX488817** | **JX488887** | – | **JX488995** | – | GU937825 |
| *Ch. liorus mictus* 1 | BYU 56945 | **JX488818** | **JX488888** | – | **JX488996** | – | – |
| *Ch. liorus mictus* 2 | BYU 56946 | **JX488819** | **JX488889** | – | **JX488997** | – | – |
| *Deltistes luxatus* | OS 15922 | AF454870 | **JX488890** | – | **JX488998** | – | GU937831 |
| *Pantosteus clarkii* | MSB 49600 | **JX488779** | **JX488848** | HQ556940 | **JX488956** | – | – |
| *P. discobolus discobolus* | BYU 57986 | **JX488782** | **JX488851** | – | **JX488959** | – | GU937830 |
| *P. discobolus yarrowi* | UAIC 12776.01-2 | **JX488783** | **JX488852** | – | **JX488960** | – | – |
| *P. nebuliferus* | HLB1484 | **JX488794** | **JX488863** | EU668538 | **JX488971** | – | – |
| *P. platyrhynchus* 1 | BYU 58618 | **JX488801** | **JX488870** | EU523934 | **JX488978** | – | – |
| *P. platyrhynchus* 2 | BYU 58621 | **JX488802** | **JX488871** | – | **JX488979** | – | – |
| *P. plebeius* 1 | DAN0521-10 | **JX488803** | **JX488872** | EU668409 | **JX488980** | – | GU937833 |
| *P. plebeius* 2 | DAN0526-01 | **JX488804** | **JX488873** | – | **JX488981** | – | – |
| *P. santaanae* | UAIC 12384.01 | **JX488805** | **JX488875** | JN024948 | **JX488983** | – | – |
| *Xyrauchen texanus* 1 | OS X97 | **JX488824** | **JX488935** | HQ556969 | **JX489042** | **This study** | – |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *X. texanus* **2** | OS X98 | **JX488825** | **JX488936** | – | **JX489043** | – | – |
| Tribe Erimyzonini | | | | | | – | – |
| ***Erimyzon oblongus* 1** | UAIC 11109.09 | AF454876 | **JX488891** | HQ579034 | – | **This study** | – |
| *Erimyzon oblongus* 2 | BRK05-52 | – | – | – | **JX488999** | – | GU937837 |
| ***E. sucetta* 1** | UAIC 12286.01 | AF45478 | **JX488892** | EU524567 | – | – | – |
| *E. sucetta* 2 | C306 | – | – | – | **JX489000** | – | – |
| ***E. tenuis*** | UAIC 12370.04 | AF454877 | **JX488893** | JN025452 | **JX489001** | **This study** | GU937838 |
| *Minytrema melanops* 1 | – | AB242116 | AB242116 | EU524839 | – | – | FJ265050 |
| *Minytrema melanops* 2 | – | DQ536432 | DQ536432 | – | – | – | – |
| ***Minytrema melanops* 3** | UAIC 11141.01 | AF454879 | **JX488897** | – | – | – | – |
| ***Minytrema melanops* 4** | BRK056 | – | – | – | **JX489005** | – | – |
| Tribe Moxostomatini | | | | | | – | – |
| ***Moxostoma albidum* 1** | UAIC 12365.01 | AF454901 | **JX488898** | EU751969 | **JX489006** | **This study** | JF799563 |
| ***M. albidum* 2** | UAIC 13446.01 | AF454902 | **JX488899** | – | **JX489007** | **This study** | – |
| ***M. anisurum*** | UAIC 11606.02 | AF454880 | **JX488900** | EU524146 | **JX489008** | **This study** | JF799537 |
| ***M. ariommum*** | UAIC 12071.01 | AF454903 | **JX488901** | JN027267 | **JX489009** | **This study** | JF799557 |
| ***M. austrinum*** | UAIC 12375.01 | AF454898 | **JX488902** | – | **JX489010** | **This study** | JF799565 |
| ***M. sp. cf. austrinum*** | UAIC 12373.01 | AF454904 | **JX488903** | – | **JX489011** | **This study** | JF799564 |
| ***M. breviceps*** | UAIC 11314.08 | AF454888 | **JX488904** | JN027271 | **JX489012** | **This study** | JF799542 |
| ***M. carinatum* 1** | UAIC 11005.03 | AF454883 | **JX488905** | EU524147 | **JX489013** | **This study** | JF799547 |
| ***M. carinatum* 2** | PBB0023 | **JX488820** | **JX488906** | – | **JX489014** | **This study** | – |
| ***M. cervinum*** | UAIC 11004.01 | AF454906 | **JX488907** | JN027274 | **JX489015** | **This study** | JF799556 |
| ***M. collapsum* 1** | UAIC 11007.03 | AF454882 | **JX488908** | JN027276 | **JX489016** | **This study** | – |
| ***M. collapsum* 2** | UAIC 12376.05 | AF454881 | **JX488909** | – | **JX489017** | **This study** | – |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *M. congestum* 1 | UAIC 13506.05 | AF522290 | **JX488910** | JN027282 | **JX489018** | **This study** | JF799560 |
| *M. congestum* 2 | UAIC 13508.05 | AF522291 | **JX488911** | – | **JX489019** | **This study** | – |
| *M. congestum* 3 | UAIC 13512.05 | AF522292 | **JX488912** | – | **JX489020** | **This study** | – |
| *M. duquesnei* 1 | UAIC 11310.09 | AF454894 | **JX488913** | EU524861 | – | – | – |
| *M. duquesnei* 2 | JFBM38581 | AF454895 | **JX488914** | – | **JX489021** | **This study** | JF799554 |
| *M. erythrurum* 1 | UAIC 12237.03 | AF454886 | **JX488915** | EU524867 | **JX489022** | **This study** | JF799551 |
| *M. erythrurum* 2 | JFBM37043 | AF454887 | **JX488916** | – | **JX489023** | **This study** | – |
| *M. hubbsi* 1 | NCSM 36440 | AF522289 | **JX488917** | EU524877 | **JX489024** | **This study** | JF799544 |
| *M. hubbsi* 2 | 06-19 | **JX488821** | **JX488918** | – | **JX489025** | **This study** | – |
| *M. lachneri* | UAIC 12370.02 | AF454900 | **JX488919** | JN027292 | **JX489026** | **This study** | JF799559 |
| *M. sp. cf. lachneri* | UAIC 12462.03 | AF454907 | **JX488920** | – | **JX489027** | **This study** | – |
| *M. macrolepidotum* | UAIC 11221.10 | AF454890 | **JX488921** | EU524149 | **JX489028** | **This study** | JF799539 |
| *M. sp. cf. macrolepidotum* | UAIC 11643.01 | AF454885 | **JX488922** | – | **JX489029** | **This study** | F799541 |
| *M. mascotae* | UAIC 12374.01 | AF454899 | **JX488923** | – | **JX489030** | **This study** | JF799566 |
| *M. pappillosum* | UAIC 13462.01 | AF454883 | **JX488924** | JN027303 | **JX489031** | **This study** | JF799538 |
| *M. poecilurum* 1 | – | NC008674 | NC008674 | – | – | – | – |
| *M. poecilurum* 2 | UAIC 11442.01 | AF454896 | **JX488925** | HQ579038 | **JX489032** | **This study** | JF799552 |
| *M. sp. cf. poecilurum* | UAIC 12746.13 | AF454897 | **JX488926** | – | **JX489033** | **This study** | – |
| *M. pisolabrum* | UAIC 11154.05 | AF454889 | **JX488927** | HQ557206 | **JX489034** | **This study** | – |
| *M. robustum* | UAIC 11916.01 | AF454891 | **JX488928** | JN027313 | **JX489035** | **This study** | JF799546 |
| *M. rupiscartes* | UAIC 12376.06 | AF454905 | **JX488929** | JN027315 | **JX489036** | **This study** | JF799558 |
| *M. valenciennesi* | JFBM36305 | AF454893 | **JX488930** | EU524150 | **JX489037** | **This study** | JF799543 |

Tribe Thoburniini

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Hypentelium etowanum* | UAIC 12523.08 | AF454908 | **JX488894** | JN026830 | **JX489002** | – | GU937836 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *H. nigricans* 1 | – | NC008676 | NC008676 | – | – | – | – |
| ***H. nigricans* 2** | UAIC 11138.02 | AF454909 | **JX488895** | EU524667 | **JX489003** | **This study** | JF799571 |
| ***H. roanokense*** | UAIC 13449.02 | AF454910 | **JX488896** | JN026848 | **JX489004** | **This study** | JF799570 |
| ***Thoburnia atripinnis*** | UAIC 13463.01 | AF454911 | **JX488931** | HQ937020 | **JX489038** | **This study** | JF799569 |
| ***T. hamiltoni* 1** | NCSM 45840-1 | **JX488822** | **JX488932** | – | **JX489039** | – | JF799567 |
| ***T. hamiltoni* 2** | NCSM 45840-3 | **JX488823** | **JX488933** | – | **JX489040** | **This study** | – |
| ***T. rhothoeca*** | UAIC 11009.05 | AF454912 | **JX488934** | JN028432 | **JX489041** | **This study** | JF799568 |

1

**Table 3**(on next page)

Characteristics of each of the seven morphological and molecular datasets analyzed in this study.

Data subsets: 1 = mtDNA $1^{st}$ codon sites; 2 = mtDNA $2^{nd}$ codon sites; 3 = mtDNA $3^{rd}$ codon sites; 4 = IRBP $1^{st}$ codon sites; 5 = IRBP $2^{nd}$ codon sites; 6 = IRBP $3^{rd}$ codon sites; 7 = GHI + *RPS7* sites, combined; M = morphological characters; '+' symbols indicate that a given subset was included in the dataset while '−' symbols indicate it was not.

[†] During BEAST analyses, but not MrBayes analyses, morphological characters were subdivided into various subsets described in the text and Mendeley-deposited input files.

| Dataset | n (no. ingroup tips) | No. characters (missing data % of total) | No. parsimony informative characters (% total) | Data subsets | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 5 | | 6 | 7 | † M |
| concatenated mtDNA | 126 (121) | 2836 bp (11.9%) | 1216 (42.9%) | + | + | + | − | − | − | − | − |
| four-locus | 85 (80) | 5925 bp (23.1%) | 1893 (31.9%) | + | + | + | + | + | + | + | − |
| morphology | 64 (62) | 123 (11.7) | 120 (97.6%) | − | − | − | − | − | − | − | + |
| mtDNA + morphology | 85 (80) | 2959 (11.9%) | 1336 (45.2%) | + | + | + | − | − | − | − | + |
| total-evidence | 85 (80) | 6048 (23.3%) | 2013 (33.3%) | + | + | + | + | + | + | + | + |
| nuclear IRBP | 113 (108) | 839 bp (0.3%) | 210 (25.0%) | − | − | − | + | + | + | − | − |
| concatenated nDNA | 113 (108) | 3089 bp (36.4%) | 677 (21.9%) | − | − | − | + | + | + | + | − |

1

## Table 4 (on next page)

Bayes factor tests comparing twelve Bayesian relaxed-clock partitioning models applied to the total-evidence dataset.

Models including only 3$^{rd}$ codon sites of the mtDNA matrix were ruled out, because those sites had the highest amounts of predicted homoplasy during PI profiling (Results and Discussion, Fig. 5), and excluding them yielded the single greatest improvement in log-marginal likelihoods (e.g. $M_1$ versus $M_4$). Numbers and letters in brackets correspond to data subsets defined in the text and Table 3. Log-marginal likelihood estimates were derived from path sampling, and Bayes factors were estimated for each model when compared against the best-supported model.

| Model | Mitochondrial DNA | Problematic sites (#) | Relaxed clocks (#) | Clock-partitioning | Log-marginal likelihood | $2\log_e(B_{10})$ Bayes factor |
|---|---|---|---|---|---|---|
| $M_1$ | 1st, 2nd, and 3rd codon sites | 2 | 1 | [1, 2, 3, 4, 5, 6, 7, M] | −69461.62 | −89832.36 |
| $M_2$ | 1st, 2nd, and 3rd codon sites | 2 | 2 | [1, 2, 3, 4, 5, 6, 7] [M] | −69078.47 | −89066.05 |
| $M_3$ | 1st, 2nd, and 3rd codon sites | 2 | 8 | [1] [2] [3] [4] [5] [6] [7] [M] | −67222.91 | −85354.94 |
| $M_4$ | 1st + 2nd codon sites | 1 | 1 | [1, 2, 4, 5, 6, 7, M] | −38876.51 | −28662.13 |
| $M_5$ | 1st + 2nd codon sites | 1 | 2 | [1, 2, 4, 5, 6, 7] [M] | −38385.92 | −27680.96 |
| $M_6$ | 1st + 2nd codon sites | 1 | 7 | [1] [2] [4] [5] [6] [7] [M] | −37616.86 | −26142.84 |
| $M_7$ | 1st codon sites | 1 | 1 | [1, 4, 5, 6, 7, M] | −34866.49 | −20625.92 |
| $M_8$ | 1st codon sites | 1 | 2 | [1, 4, 5, 6, 7] [M] | −34382.43 | −19673.97 |
| $M_9$ | 1st codon sites | 1 | 6 | [1] [4] [5] [6] [7] [M] | −33682.60 | −18237.36 |
| $M_{10}$ | 2nd codon sites | 0 | 1 | [2, 4, 5, 6, 7, M] | −25442.32 | −1793.75 |
| $M_{11}$ | 2nd codon sites | 0 | 2 | [2, 4, 5, 6, 7] [M] | −25138.71 | −1186.53 |
| $M_{12}$ | 2nd codon sites | 0 | 6 | [2] [4] [5] [6] [7] [M] | −24545.44 | 0 |

1

**Figure 1**(on next page)

Six alternative hypotheses of phylogenetic relationships among Catostomidae subfamilies and tribes.

These results were based on previous studies of electrophoretic data (A); morphological, behavioral, and developmental characters (B); and DNA sequence data (C–F).

**A. Ferris and Whitt (1978, Fig. 3)**

Cycleptinae — *Cycleptus*
Ictiobinae — *Ictiobus* / *Carpiodes*
Erimyzonini — *Minytrema* / *Erimyzon*
Moxostomatini — *M. erythrurum* / *Moxostoma* grade / *Hypentelium* / *Thoburnia*
Catostominae
*M. duquesnei*
*Catostomus* (*Pantosteus*)
*C. macrocheilus*
*Chasmistes*
*Catostomus*
*C. commersoni*
*Xyrauchen*
Catostomini

**B. Smith (1992, Fig. 6)**

Ictiobinae — †*Amyzon* / *Ictiobus* / *Carpiodes*
"Cycleptinae" — *Myxocyprinus* / *Cycleptus*
Catostomini — *Catostomus* spp. (+ *Pantosteus*) / *Xyrauchen* / *Deltistes* / *Chasmistes*
*Minytrema*
*Erimyzon*
*Moxostoma* grade
"*Scartomyzon*" (*Moxostoma*)
*Hypentelium*
*Thoburnia* grade
"*Scartomyzon*" (*M.*)
Catostominae
Moxostomatini

**C. Harris and Mayden (2001, Fig. 5)**

Myxocyprininae — *Myxocyprinus*
Ictiobinae — *Ictiobus* / *Carpiodes*
Cycleptinae — *Cycleptus*
Catostomini — *C. catostomus* / *C. commersoni* / *Xyrauchen*
*Minytrema*
*Erimyzon*
Thoburniini — *Thoburnia* / *Hypentelium*
*M. poecilurum*
"*S.*" (*M.*) *cervinus*
*M. carinatum*
"*S.*" (*M.*) *ariommus*
*M. anisurum*
Catostominae
Moxostomatini

**D. Harris et al. (2002, Fig. 4)**

Catostomini — *Xyrauchen* / *Deltistes* / *Catostomus* / *Catostomus* (+ *Pantosteus* grade)
*Minytrema*
*Erimyzon*
*T. rothoeca*
*T. atripinnis*
*Hypentelium*
'*Scartomyzon*' grade
*Moxostoma* grade
*Moxostoma* grade
*M. poecilurum*
"*Scartomyzon*" (*M.*)
*M.* sp. cf. *poecilurum*
"*Scartomyzon*" (*M.*)
Catostominae
Moxostomatini

**E. Sun et al. (2007, Fig. 1)**

*Minytrema*
Myxocyprininae — *Myxocyprinus*
Cycleptinae — *Cycleptus*
Ictiobinae — *Ictiobus* / *Carpiodes*
*Erimyzon*
Moxostomatini — *Hypentelium* / *Moxostoma* grade / "*Scartomyzon*" (*M.*)
*C. catostomus*
*Chasmistes*
*Deltistes*
*Xyrauchen*
*C. commersoni*
*T. rothoeca*
Catostominae
Catostomini

**F. Doosey et al. (2010, Fig. 1)**

Cycleptinae — *Cycleptus*
Myxocyprininae — *Myxocyprinus*
Ictiobinae — *Ictiobus* / *Carpiodes*
Erimyzonini — *Minytrema* / *Erimyzon*
Moxostomatini — *Moxostoma*
Thoburniini — *Thoburnia* / *Hypentelium*
*C. latipinnis*
*Catostomus* (*Pantosteus*) grade + *Xyrauchen*
*C. commersoni*
*Catostomus* grade
*Deltistes*
*Chasmistes*
*C. rimiculus*
Catostominae
Catostomini

## Figure 2(on next page)

Consensus topologies from MrBayes (Ronquist et al., 2012) analyses of the concatenated mtDNA (A), four-locus (B), morphology (C), mtDNA + morphology (D), total-evidence (E), concatenated nDNA (F), and nuclear IRBP (G) datasets.

Subfamily and tribe names and colors shown in panel A are followed throughout; in panels B–G, text labels highlight positions of different taxa, including subfamilies (Myx., Myxocyprininae), tribe Catostomini, and *Pantosteus*, and dagger symbols indicate extinct taxa. Filled black circles indicate nodes with significant Bayesian posterior probabilities (BPP ≥0.95), and gray circles indicate near-significant nodes (BPP = 0.80–0.94). Scale bars are given in units of substitutions/site.

**A. concatenated mtDNA**

Catostomini (genus "*Catostomus*")

*Pantosteus*

Erimyzonini

Moxostomatini (genus *Moxostoma*)

Myxocyprininae
Cycleptinae
Ictiobinae

Thoburniini

Outgroup taxa

0.2

**B. four-locus**

*Pantosteus*

Outgroup taxa

Myx. sister to all other suckers

0.4

**C. morphology**

paraphyletic Moxostomatini

†

Thoburniini

*Pantosteus*

Outgroup taxa

Myxocyprininae
Cycleptinae

†

Ictiobinae

Catostomini

genera *Deltistes*, *Chasmistes*, & *Xyrauchen*

0.4

**D. mtDNA + morphology**

*Pantosteus*

†

Outgroup taxa

†

0.2

**E. total-evidence**

*Pantosteus*

†

Outgroup taxa

†

0.4

**F. nuclear IRBP**

*Pantosteus*

paraphyletic Erimyzonini

Outgroup taxa

paraphyletic Ictiobinae

Cycleptinae

paraphyletic Thoburniini

0.4

**G. concatenated nDNA**

*Pantosteus*

Catostomini (genus "*Catostomus*")

Outgroup taxa

Myxocyprininae & Ictiobinae, in part
Cycleptinae

Ictiobinae, in part

Erimyzonini

Moxostomatini

Thoburniini, in part

Thoburniini, in part

0.06

# Figure 3(on next page)

Phylogeny of suckers inferred from Bayesian analysis of the combined mtDNA (cyt*b*, *cox1*, ND2), nDNA (IRBP, GHI, and *RPS7*), and morphological data (123 characters) in the total-evidence dataset.

Within tip labels, dagger symbols indicate extinct taxa, and museum (voucher) or field numbers are followed by the number of the individual sequenced (Table 1). Nodes are labeled with Bayesian posterior probability support values above 0.50. Scale bars are in units of substitutions/site.

**Figure 4**(on next page)

Time-calibrated phylogeny of Catostomidae derived from the best-supported Bayesian total-evidence dating model identified in Table 4.

In parentheses beside subfamily and tribe names, numbers of extinct or fossil tips are presented out of total sample size for the corresponding clade. Along nodes, filled black circles indicate significant posterior support (BPP ≥0.95), gray circles indicate near-significant support (BPP = 0.80–0.94), and horizontal node bars show 95% highest posterior densities of age estimates. The inset table shows Bayesian posterior age estimates given as mean $t_{MRCA}$s, with their 95% highest posterior densities (HPDs; credible intervals), in units of millions of years ago (Ma). Geological epoch abbreviations: Paleo., Paleocene; Oligo., Oligocene; Q., Quaternary; Pl., Pliocene; P., Pleistocene.
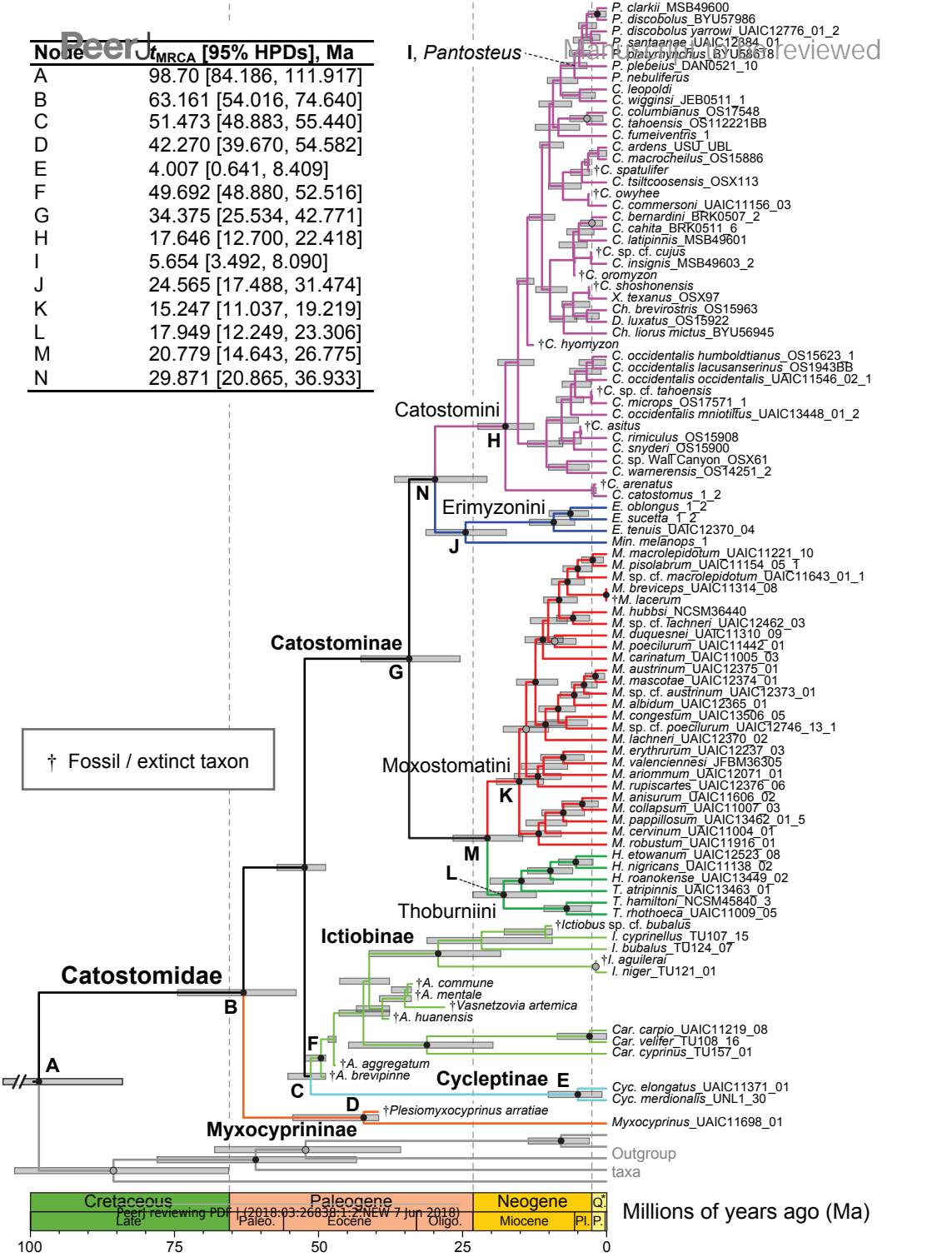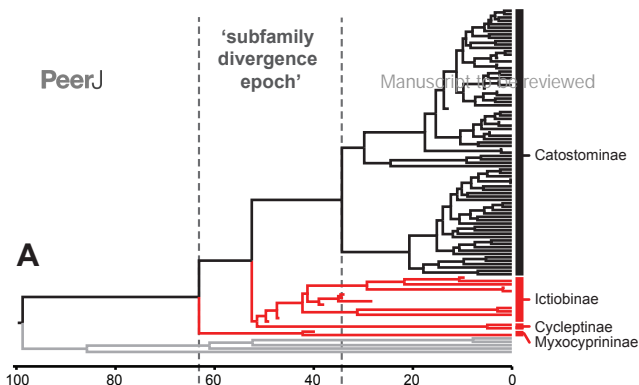
| Node | t_MRCA [95% HPDs], Ma |
|------|------------------------|
| A | 98.70 [84.186, 111.917] |
| B | 63.161 [54.016, 74.640] |
| C | 51.473 [48.883, 55.440] |
| D | 42.270 [39.670, 54.582] |
| E | 4.007 [0.641, 8.409] |
| F | 49.692 [48.880, 52.516] |
| G | 34.375 [25.534, 42.771] |
| H | 17.646 [12.700, 22.418] |
| I | 5.654 [3.492, 8.090] |
| J | 24.565 [17.488, 31.474] |
| K | 15.247 [11.037, 19.219] |
| L | 17.949 [12.249, 23.306] |
| M | 20.779 [14.643, 26.775] |
| N | 29.871 [20.865, 36.933] |

I, *Pantosteus*

*P. clarkii* MSB49600
*P. discobolus* BYU57986
*P. discobolus yarrowi* UAIC12776_01_2
*P. santaanae* UAIC12384_01
*P. platyrhynchus* BYU58618
*P. plebeius* DAN0521_10
*P. nebuliferus*
*C. leopoldi*
*C. wigginsi* JEB0511_1
*C. columbianus* OS17548
*C. tahoensis* OS112221BB
*C. fumeiventris*_1
*C. ardens* USU_UBL
*C. macrocheilus* OS15886
†*C. spatulifer*
†*C. tsiltcoosensis* OSX113
†*C. owyhee*
*C. commersoni* UAIC11156_03
*C. bernardini* BRK0507_2
*C. cahita* BRK0511_6
*C. latipinnis* MSB49601
†*C.* sp. cf. *cujus*
*C. insignis* MSB49603_2
†*C. oromyzon*
†*C. shoshonensis*
*X. texanus* OSX97
*Ch. brevirostris* OS15963
*D. luxatus* OS15922
*Ch. liorus mictus* BYU56945
†*C. hyomyzon*
*C. occidentalis humboldtianus* OS15623_1
*C. occidentalis lacusanserinus* OS1943BB
†*C. occidentalis occidentalis* UAIC11546_02_1
†*C.* sp. cf. *tahoensis*
*C. microps* OS17571_1
*C. occidentalis mniotiltus* UAIC13448_01_2
†*C. asitus*
*C. rimiculus* OS15908
*C. snyderi* OS15900
*C.* sp. Wall Canyon OSX61
*C. warnerensis* OS14251_2
†*C. arenatus*
*C. catostomus*_1_2
*E. oblongus*_1_2
*E. sucetta*_1_2
*E. tenuis* UAIC12370_04
*Min. melanops*_1
*M. macrolepidotum* UAIC11221_10
*M. pisolabrum* UAIC11154_05_1
*M.* sp. cf. *macrolepidotum* UAIC11643_01_1
*M. breviceps* UAIC11314_08
†*M. lacerum*
*M. hubbsi* NCSM36440
*M.* sp. cf. *lachneri* UAIC12462_03
*M. duquesnei* UAIC11310_09
*M. poecilurum* UAIC11442_01
*M. carinatum* UAIC11005_03
*M. austrinum* UAIC12375_01
*M. mascotae* UAIC12374_01
*M.* sp. cf. *austrinum* UAIC12373_01
*M. albidum* UAIC12365_01
*M. congestum* UAIC13506_05
*M.* sp. cf. *poecilurum* UAIC12746_13_1
*M. lachneri* UAIC12370_02
*M. erythrurum* UAIC12237_03
*M. valenciennesi* JFBM36305
*M. ariommum* UAIC12071_01
*M. rupiscartes* UAIC12376_06
*M. anisurum* UAIC11606_02
*M. collapsum* UAIC11007_03
*M. pappillosum* UAIC13462_01_5
*M. cervinum* UAIC11004_01
*M. robustum* UAIC11916_01
*H. etowanum* UAIC12523_08
*H. nigricans* UAIC11138_02
*H. roanokense* UAIC13449_02
*T. atripinnis* UAIC13463_01
*T. hamiltoni* NCSM45840_3
*T. rhothoeca* UAIC11009_05
†*Ictiobus* sp. cf. *bubalus*
*I. cyprinellus* TU107_15
*I. bubalus* TU124_07
†*I. aguilerai*
*I. niger* TU121_01
†*A. commune*
†*A. mentale*
†*Vasnetzovia artemica*
†*A. huanensis*
*Car. carpio* UAIC11219_08
*Car. velifer* TU108_16
*Car. cyprinus* TU157_01
†*A. aggregatum*
†*A. brevipinne*
*Cyc. elongatus* UAIC11371_01
*Cyc. merdionalis* UNL1_30
†*Plesiomyoxcyprinus arratiae*
*Myxocyprinus* UAIC11698_01
Outgroup
taxa

Catostomini

Catostominae

Erimyzonini

Moxostomatini

Thoburniini

Ictiobinae

Cycleptinae

Myxocyprininae

**Catostomidae**

† Fossil / extinct taxon

A B C D E F G H I J K L M N

Millions of years ago (Ma)

| Cretaceous | Paleogene | Neogene | Q. |
|---|---|---|---|
| Late | Paleo. | Eocene | Oligo. | Miocene | Pl. P. |

100 75 50 25 0

**Figure 5**(on next page)

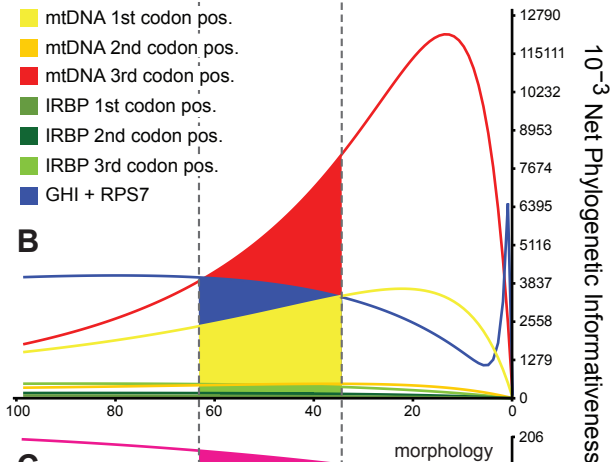Phylogenetic informativeness of the sucker datasets.

(A) BEAST chronogram from Fig. 4 used for site rate calculations, showing the three main subfamilies with incongruent relationships highlighted in red. Phylogenetic informativeness profiles matched to the chronogram time scale are shown for seven subsets of the four-locus dataset (B), and for the morphology dataset (C). Colored shading in (B) and (C) indicate areas integrated below each profile over the 'subfamily divergence epoch' (~63.2–34.4 Ma) containing branching relationships among the four subfamilies.