Research Article

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

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

Author Addresses: [1]Department of Biology, Virginia Commonwealth University, Richmond, VA, USA; [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, USA; [5]Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, USA

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

Corresponding author: Dr. Justin C. Bagley; Phone: +001 (804)-828-0820; Fax: +001 (804)-828-0503; E-mail: jcbagley@vcu.edu.

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

Keywords: Catostomidae; divergence time estimation; fossilized birth-death process; molecular phylogenetics; phylogenetic informativeness profiles; relaxed molecular clock

Contents: ~8,304 words (Title, Abstract, main text, and Acknowledgements; excluding References and table/figure captions); 4 tables; 5 figures.

Abstract

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 but reject or confirm different molecular hypotheses about the timing of divergence within and between sucker subfamilies. 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.

Introduction

'Suckers' in the family Catostomidae (Cypriniformes) form a large family of Holarctic freshwater fishes with 76 extant species in 13 genera native to North America and Asia (Harris et al., 2014). Seventy-five species from 12 genera occur in North America (Harris et al., 2014; Nelson et al., 2004), where they constitute the third largest freshwater fish clade, after darters (Etheostomatinae) and minnows (Cyprinidae), and ~8% of the continental ichthyofauna (Warren et al., 2000). *Myxocyprinus asiaticus* is endemic to eastern China and *Catostomus catostomus* is the only extant trans-Pacific species (Harris et al., 2014). Suckers are an ancient group whose fossil record spans the Cenozoic, from the early Eocene onwards (Cavender, 1986; Smith, 1992; Smith et al., 2002; Appendix S1). Around 35% of the taxa (26 to >35 species or operational taxonomic units, 'OTUs') are endangered, threatened, or of special conservation concern (Harris et al., 2014; Nelson et al., 2004; Warren et al., 2000).

Geographical distributions of suckers and their taxonomy and relationships have attracted the interest of systematists and biogeographers for over 150 years (e.g. Agassiz, 1854). Pre-1900 systematics and taxonomy studies dealt with species descriptions and higher-level classifications of the group (reviewed by Harris et al., 2014; Smith, 1992). Subsequently, key papers on sucker classification designated genera, subgenera, and tribes (Hubbs, 1930; Robins & Raney, 1956) and contributed a pre-cladistics phylogeny (Miller, 1959), as well as the first phylogeny for species in the tribe Moxostomatini (Jenkins, 1970). The findings of post-systematics studies of sucker interrelationships (last 40 years), which were based on a variety of different data types, are summarized in Table 1 and Figure 1. These studies supported the recognition of four subfamilies (Myxocyprininae, Cycleptinae, Ictiobinae, and Catostominae) and tribes (Catostomini, Erimyzonini, Moxostomatini, Thoburniini) plus the modern allocation of genera to these groups (Harris & Mayden, 2001; Harris et al., 2002). Most phylogenetic analyses of

66 suckers to date have analyzed DNA sequence data from increasing numbers of mitochondrial
67 DNA (mtDNA) or nuclear DNA (nDNA) genes. Recent advances include Clements et al.'s
68 (2012) study of phylogenetic relationships of Moxostomatini using the first multilocus dataset of
69 mtDNA cyt*b* and nuclear growth hormone intron (GHI) sequences. Also, Unmack et al. (2014)
70 studied the phylogeny and biogeography of *Pantosteus*, a subgenus of *Catostomus* (Harris &
71 Mayden, 2001; Harris et al., 2002; Smith, 1966) using independent morphological and
72 multilocus genetic analyses. Overall, while the inclusion of additional taxa and sequence data in
73 recent analyses has yielded novel insights on sucker relationships, a consensus among
74 hypotheses regarding relationships of higher-level species groups, i.e. subfamilies, has yet to
75 emerge (Table 1; Fig. 1). Moreover, although time-calibrated phylogenies based on
76 comprehensive tip sampling are necessary to understand the tempo and mode of diversification
77 of groups of organisms (e.g. Near et al., 2011, 2012), such a tool has yet to be inferred for
78 suckers. This leaves workers at a distinct disadvantage in considering the macroevolutionary,
79 ecological, or conservation trends of suckers in a phylogenetic context.
80 In this study, we use 'total-evidence' analyses of multilocus sequence data from three
81 mtDNA genes and three nuclear genes representing the most comprehensive sampling of suckers
82 to date, combined with available morphological and fossil data, to infer the phylogenetic
83 relationships and divergence times of suckers in a Bayesian framework. We discuss the
84 phylogenetic and taxonomic implications of our results in light of previous phylogenetic studies
85 of suckers. We apply Bayesian relaxed-clock models to the data (Drummond et al., 2012) and
86 take advantage of rich information on the age and distributions of sucker fossils (e.g. Cavender,
87 1986; Smith et al., 2002; Appendix S1) to improve divergence time estimation by incorporating
88 extant and fossil sampling in a "fossilized birth-death process" tree prior (Gavryushkina et al.,

2014). Moreover, by coupling assessments of the phylogenetic signal of different data subsets with evaluation of clock-partitioning strategies, we were able to avoid potentially confounding effects of problematic characters on our divergence time inferences. The resulting time-calibrated phylogeny is then used to test several hypotheses about the temporal diversification of suckers. Namely, we test the hypotheses (H_1) that Asian Myxocyprininae and North American suckers (Cycleptinae) diverged since ~14 million years ago (Ma) in the mid-Miocene, and (H_2) that the initial divergence of Ictiobinae lineages occurred since ~10 Ma in the late Miocene, as indicated by mtDNA cyt*b* gene divergences (Sun et al., 2007). We also tested the hypotheses (H_3) that the diversification of lineages within Catostominae followed ~20 Ma in the Miocene (Smith, 1992; Sun et al., 2007), and (H_4) that Catostomidae species have diversified from a tetraploid ancestor since ~50 Ma in the early Eocene (Uyeno & Smith, 1972). Overall, by yielding a new phylogeny of suckers and divergence dates for their most recent common ancestors (MRCAs), our study sheds light on the interrelationships, taxonomic classification, and tempo of speciation in a diverse and threatened clade of Holarctic freshwater fishes.

Materials & Methods

Molecular taxon sampling, laboratory methods, and sequence alignment

We obtained and sequenced mtDNA and nDNA genes from tissue samples of 121 sucker specimens from throughout the geographical range of the family. Samples were provided by ichthyological collections or colleagues (see Acknowledgements), or were already in-hand at the beginning of the study. We sequenced taxa representing all sucker genera and all Catostominae species except for †*Moxostoma lacerum*, an historically extinct taxon last sampled from the Mississippi River Basin in 1893 (NatureServe, 2013), and *Chasmistes cujus*, an endangered species (NatureServe, 2014) for which it was difficult to obtain samples (Table 2). To

112 complement our sampling and add data from more unlinked loci to our datasets, we obtained
113 additional sequences from GenBank (see section 2.2). Overall, we obtained genetic sequences
114 for the most comprehensive taxonomic sampling of the family to date, including 78
115 species/lineages representing all 13 extant genera, including four 'candidate species' within
116 Moxostomatini (Table 2). Based on recent hypotheses of relationships of suckers and
117 cypriniform fishes (e.g. Mayden et al., 2008; Saitoh et al., 2006; Smith, 1992), we included DNA
118 sequences from five outgroup taxa in our datasets: *Cyprinus carpio* (Cyprinidae); *Gyrinocheilus*
119 *aymonieri* (Gyrinocheilidae); and *Cobitis striata*, *Chromobotia macracanthus*, and *Leptobotia*
120 *mantschurica* (Cobitidae). We follow the taxonomy of subfamilies listed in Harris and Mayden
121 (2001) and Harris et al. (2002).
122 All molecular laboratory work for this project conducted in the R. L. Mayden laboratory
123 was approved under Saint Louis University Institutional Animal Care and Use Committee
124 (IACUC) protocol #2467. We extracted whole genomic DNA using QIAGEN DNeasy Tissue
125 kits (QIAGEN, Valencia, CA, Catalog No. 69506), or the CTAB method of Sagahi-Maroof et al.
126 (1984). We amplified and sequenced the mtDNA cyt*b* gene using PCR primers and parameters
127 in Harris et al. (2002). We amplified and sequenced the mtDNA NADH subunit 2 (ND2) gene
128 with primers 562 (5′-TAA GCT ATC GGG CCC ATA CC-3′) and 449 (5′-TGC TTA GGG CTT
129 TGA AGG CTC-3′) from LGL Genetics, Bryan, TX, using the same PCR amplification
130 parameters used for cyt*b*. We also sequenced the first part of the nuclear IRBP gene using PCR
131 primers and methods in Chen et al. (2008), and we sequenced nuclear ribosomal protein *S7*
132 intron 1 (*RPS7*) using PCR primers and parameters in Chow & Takeyama (1998). We purified
133 double-stranded PCR products using columns or gel extraction kits (QIAGEN, Valencia, CA).
134 Given that catostomids are tetraploids (Uyeno & Smith, 1972), we cloned nuclear PCR products

to ensure orthologous sequences were used in subsequent phylogenetic reconstructions. Purified PCR products were cloned using the TOPO TA Cloning® Kit (Invitrogen Corp., Carlsbad, CA) with TOP10 chemically competent or electrocompetent cells. Positive colonies were chosen randomly and cultured, and then plasmids preps were purified using the QIAprep Spin Miniprep Kit (QIAGEN, Valencia, CA) and sequenced in both directions using universal M13 primers. We sequenced all genes in both directions on an Applied Biosystems 3100 Genetic Analyzer using ABI PRISM BigDye Terminator v2.0 or v3.0 Cycle Sequencing Kits (Applied Biosystems, Foster City, CA).

We edited sequence chromatographs, assembled sequence contigs, and created final DNA alignments using Geneious v5.4 (Kearse et al., 2012). We translated all gene sequences into amino acid sequences to check alignments for stop codons or elevated nonsynonymous substitution numbers, because these signatures can indicate the presence of nuclear mtDNA gene copies, or 'NUMTs'. Mitochondrial DNA and IRBP sequences aligned straightforwardly 'by-eye' in Geneious. However, we aligned nuclear *RPS7* sequences, and nuclear genes from other studies listed in section 2.2, in MAFFT v6.850 (Katoh & Toh, 2008) using the local-pair FFTS algorithm with MAXITERATE = 50.

Dataset construction, model selection, and Bayesian phylogenetic analyses

We collated seven datasets for our analyses. (1) The 'concatenated mtDNA' dataset consisted of mtDNA cyt*b* and ND2 sequences for 126 tips (121 sucker samples and 5 outgroup samples) plus 58 mtDNA cytochrome oxidase subunit 1 (*cox1*) gene sequences from GenBank. (2) A 'four-locus' DNA dataset with the same 126 tips but adding the following nuclear data to the concatenated mtDNA dataset: 44 GHI sequences from Clements et al. (2012) and GenBank, plus 113 IRBP sequences and 52 *RPS7* sequences from our study. (3) We reanalyzed a morphological

Comentado [UdMO1]: Are they really 7? Or 6 data sets; in according with Table 3, the partition mtDNA + morphology are no mentioned in this section

158 dataset first presented by Smith (1992), containing 123 morphological characters for 64 taxa
159 including 62 sucker taxa and two outgroup taxa. This matrix contained two extinct taxa: the
160 historically extinct †*M. lacerum*, and †*Amyzon*, an Eocene–Oligocene genus of fossil suckers
161 composed of five valid species known from British Columbia, Washington, Nevada, and
162 Wyoming (four species), and Jilin province, China (Bruner, 1991; Chang et al., 2001; Smith,
163 1992; Appendix S1). (4) Our 'total-evidence' dataset contained all morphological and molecular
164 characters analyzed in this study, for 85 taxa. Two final nDNA datasets were: (6) a 'nuclear
165 IRBP' alignment for the only nuclear gene with complete sampling, and (7) a 'concatenated
166 nDNA' alignment containing all three nuclear loci.

Comentado [UdMO2]: Where is the fifth data set?

167 We selected the most appropriate partitioning schemes and models of sequence evolution
168 for each of the DNA data 'subsets' or 'blocks' used in our phylogenetic analyses in
169 PartitionFinder v1.1.1 (Lanfear et al., 2012), which included codon positions of the mtDNA and
170 IRBP alignments, plus the GHI and *RPS7* genes. We ran PartitionFinder simultaneously on all
171 initial DNA subsets using the greedy heuristic search algorithm, which we set to conduct model
172 comparisons to determine the 'best-fit' partitioning scheme based on the Bayesian information
173 criterion (BIC). PartitionFinder relies heavily on PhyML, which we set to link branch lengths,
174 search 56 substitution models, and estimate the base frequencies, proportion of invariant sites (I),
175 and the gamma shape distribution (Γ) of each model using maximum-likelihood.
176 We performed partitioned Bayesian phylogenetic analyses on all seven datasets in
177 MrBayes 3.2.2 (Ronquist et al., 2012). Analyzing different taxon and character combinations
178 allowed us to evaluate the effect of different data types, and of including morphological data and
179 extinct taxa, on phylogenetic analyses of suckers. During molecular analyses, we specified
180 partitioning schemes and best-fit models of sequence evolution selected in PartitionFinder,

except where the selected model was not implemented in MrBayes we used the next most closely related model in the GTR family of models. Given that rate variation among morphological characters can confound phylogenetic branch lengths, applying Γ-distributed rates can greatly improve models of morphological evolution (e.g. Clarke & Middleton, 2008). Thus, for the morphological analysis, we specified Lewis's (2001) Markov variable (Mkv) model with Γ-distributed rate heterogeneity and left characters unordered in state polarity (default). For each dataset, we conducted three independent MrBayes runs of eight chains, each with a Markov chain Monte Carlo (MCMC) chain length of 50 million generations. We diagnosed run convergence using the potential scale reduction factor, which should approach values of 1 when stationarity has been reached (Ronquist et al., 2012).

Bayesian total-evidence dating and relaxed-clock partitioning

We estimated divergence times as times to the most recent common ancestor (t_{MRCA}s) through Bayesian relaxed molecular clock analyses of the total-evidence dataset in BEAST v2.4.5 (Bouckaert et al., 2014). For our tree prior, we employed the fossilized birth-death (FBD) process model (Gavryushkina et al., 2014; Heath et al., 2014) as modified by Gavryushkina et al. (2017), because this model accommodates total-evidence datasets, avoids the need for arbitrary calibration densities, and accommodates all fossils available for a group, and not merely ad hoc selections. In addition to 83 extant ingroup and outgroup samples, 19 extinct sucker taxa and their available minimum ages from the fossil record were included in the BEAST analyses based on evidence provided in Appendix S1. We calculated the approximate sampling proportion for extant lineages (ρ) as 0.97 and set the FBD time of origin prior (t_{or}) to the Cypriniformes 'Root' node calibration discussed in Appendix S1. Morphological characters were partitioned into groups having the same number of states, and each partition was assigned an Mkv model (Lewis,

2001; conditioning on the use of only variable characters) with Γ-distributed rate variation ("partitioned mode" in Gavryushkina et al., 2017). We partitioned the DNA data and set site models according to the best scheme identified in PartitionFinder. Other authors have employed FBD models on a fixed topology (e.g. Heath et al., 2014) or simultaneously estimated topology and FBD parameters (e.g. Gavryushkina et al., 2017). We allowed most nodes, including subfamily relationships, to change freely. However, we constrained the Catostomini and Moxostomatini crown groups to be monophyletic, consistent with our MrBayes results and hypotheses of previous studies (see Results and Discussion). Posterior nodal support can be markedly weakened when fossils have unscored characters (e.g. Gavryushkina et al., 2017) or lack data. To avoid spurious relationships arising from the large proportion of fossil Ictiobinae taxa (57%) and Myxocyprininae taxa (50%) lacking data, samples of *Carpiodes*, *Ictiobus*, Ictiobinae, and Myxocyprininae were each constrained to be monophyletic, consistent with our other phylogenetic results.

According to Ho & Lanfear (2010), implementing multiple relaxed-clock models for different data subsets in a data-partitioning scheme can provide a more biologically realistic way to model among-lineage rate variation and improves the fit of relaxed-clock models to the data. Using a relaxed-clock partitioning scheme may also yield more precise date estimates with narrower credible intervals (Ho & Lanfear, 2010). We statistically tested whether allocating separate uncorrelated lognormal relaxed clocks to different DNA data subsets, through 'relaxed-clock partitioning', yielded BEAST models that provided a better fit to the data than assuming a single model of branch-specific rates across all data subsets. First, we estimated divergence dates and marginal likelihood scores for each of 12 different relaxed-clock partitioning models (M_1–

226 M_{12}) with one to eight clocks (see Results and Discussion). Second, we estimated Bayes factors
227 and conducted Bayesian model selection to identify the best-supported model.
228 We ran five replicate searches of each model in BEAST (MCMC = 2×10^7 generations,
229 sampling every 4000) using the 'BEASTRunner.sh' script in PIrANHA (Bagley, 2017). We then
230 estimated log-marginal likelihoods for each model by conducting path sampling (PS) (Baele et
231 al., 2012) for 100 steps (10^6 generations each), while specifying a ~B(0.3, 1) distribution for
232 spacing the path steps (Xie et al., 2011). We calculated $2\log_e(B_{10})$ Bayes factors from the log-
233 marginal likelihoods and evaluated 'weight of evidence' of the models according to criteria in
234 Kass & Raftery (1995). We took posterior distributions from the best-supported model as our
235 best estimates of the time tree and divergence dates for Catostomidae lineages. We summarized
236 parameters from the best-supported model, and ensured convergence and adequate effective
237 sample sizes (ESS >100–200), using Tracer v1.5 (Rambaut & Drummond, 2014). We calculated
238 a maximum clade credibility tree annotated with mean node ages from 5,000 post-burn-in trees
239 in TreeAnnotator v2.4.5.

240 *Using phylogenetic informativeness profiles to exclude problematic characters*
241 Among other factors influencing phylogenetic inference, such as sampling effects on branch
242 lengths or nodal support values (e.g. Heath et al., 2008; Pyron, 2011), the varying
243 informativeness of different character sets can substantially and adversely affect phylogenetic
244 divergence dating results (Dornburg et al., 2014). Fortunately, recent methods for quantifying
245 and visualizing 'phylogenetic informativeness' (PI) profiles of character sets through time
246 (Townsend, 2007; Townsend et al., 2012) provide a framework for identifying and excluding
247 problematic character sets (Dornburg et al., 2014). One common pitfall is the use of characters
248 whose profiles exhibit a decline in informativeness towards the root. As noted by Townsend &

249 Leuenberger (2011), this decline marks a “rain shadow of noise”, with the corresponding dataset
250 losing phylogenetic informativeness due to an increase in predicted homoplasy. In turn,
251 homoplasious loci or character sets, but especially mtDNA datasets exhibiting high saturation or
252 rootward declines in PI, have been shown to mislead global branch length values during
253 divergence time estimation (e.g. Brandley et al., 2011; Dornburg et al., 2014).
254 To identify and exclude potentially problematic character sets in our database, and to
255 evaluate whether support versus instability of subfamily relationships correlated to statistical
256 power to revolve branching order, we evaluated the Townsend (2007) PI and resolution
257 probabilities of each character set through time using PhyDesign (López-Giráldez & Townsend,
258 2011). Estimating PI requires prior information on evolutionary-genetic rates and phylogeny.
259 Thus, we estimated rates for DNA characters and morphological characters using HyPhy (Pond
260 et al., 2005) and BayesTraits (Pagel & Meade, 2014), respectively, and ran analyses along the
261 BEAST time tree from the best-supported clock-partitioning model (see Results and Discussion).
262 We estimated net PI for all eight data subsets in Table 3 during the ‘subfamily divergence epoch’
263 spanning branches leading to Myxocyprininae, Cycleptinae, Ictiobinae, and Catostominae. These
264 analyses permitted broad comparisons of signal in the mitochondrial versus nuclear data, and
265 molecular versus morphological character sets. We excluded character sets that exhibited steep
266 declines in PI towards the root of our phylogeny from the final divergence time results presented
267 below. We compared three resolution probabilities, including probability correct, probability
268 polytomy, and probability incorrect or ‘phylogenetic noise’ (equations 11–13 in Townsend et al.,
269 2012) for the excluded subsets versus other molecular subsets over the subfamily divergence
270 epoch. We also evaluated sensitivity of resolution probability approximations to varying the

internode time length (t_0) parameter, by recalculating over 10 t_0 values representing declining fractions of the epoch.

Results

Dataset characteristics and DNA substitution models

Our final data matrices ranged in size from 123 characters in the morphological dataset to 6048 molecular and morphological characters in the total-evidence dataset (Table 3). Proportions of missing data and parsimony informative characters ranged from 0.3–36.4% to 21.9–97.6%, respectively, across datasets (Table 3). PartitionFinder identified seven unique DNA sequence subsets (scheme BIC = 141854.71523), and the best-fit DNA substitution model for each subset is listed in Table S1. Morphological character subsets (in subset M) were assigned Mkv+Γ models, as described above. None of the mtDNA genes sequenced in this study showed signs of NUMTs, and we found no indels in the IRBP sequences; however, GHI and *RPS7* genes aligned with ~32 and ~16 ingroup indels/gaps, respectively. We archived our sequence alignments and phylogenetic tree results in Mendeley Data (doi: 10.17632/trw6sb4v7w.1).

Phylogenetic relationships among sucker subfamilies

We placed subfamilies Myxocyprininae and Cycleptinae as sister lineages in most trees, but with higher posterior support (BPP = 0.74–0.99) in runs based largely on mtDNA-encoded genes (Figs. 2A, 2D, S1 and S2) and weak BPP support (≤0.69) in other runs (Figs. 2C, 2E, 3, S3). Frequently, when this pattern was obtained, Myxocyprininae + Cycleptinae was resolved as sister to a monophyletic Ictiobinae with variable posterior support (BPP = 0.69–0.98; Figs. 2, 3, S1 and S2). This ‘subfamily pattern 1’, with ((Myxocyprininae, Cycleptinae), Ictiobinae), departs markedly from the placement of Ictiobinae sister to Cycleptinae (sometimes including *Myxocyprinus*) + Catostominae in previous analyses of morphology (Smith, 1992) and molecular

294 data (Doosey et al., 2010; Harris & Mayden, 2001; Mayden et al., 2008; Saitoh et al., 2006). By
295 contrast, several trees agreed in placing Myxocyprininae as sister to all other sucker lineages,
296 which agrees with previous mtDNA results presented in Harris & Mayden (2001). This
297 'subfamily pattern 2' relationship was strongly supported with BPP = 1 in our total-evidence
298 consensus topology from BEAST (Fig. 4) and resolved with low support in the four-locus
299 topology (Figs. 2B and S4). Distinguishing between these two conflicting sets of subfamily
300 relationships is difficult, because each is supported by molecular and total-evidence topologies
301 herein and agrees with at least one previous molecular study. To objectively determine the
302 arrangement of these subfamilies with the greatest weight of evidence conditional on our total-
303 evidence dataset, we compared subfamily patterns 1 and 2 using Bayes factors. We ran MrBayes
304 as described above, except employing topological constraints set to subfamily pattern 2, and then
305 used stepping-stone sampling (Xie et al., 2011; Baele et al., 2012) to estimate the log-marginal
306 likelihoods of the models, from which Bayes factor tests were conducted through comparisons to
307 the unconstrained model matching subfamily pattern 1. Conducting 50,000 generations of
308 stepping-stone sampling (sampling every 2500 generations) during each of 50 steps produced a
309 total of 250,000 MCMC generations for marginal likelihood estimation. The subfamily pattern 2
310 model constraining Myxocyprininae as sister to all other suckers had a higher log-marginal
311 likelihood score (−68390.83) than the unconstrained subfamily pattern 1 model (−68910.88), and
312 a $2\log_e(B_{10})$ Bayes factor of −1040.10 provided definitive weight of evidence against the
313 unconstrained model.

314 *Relationships of early-diverging sucker genera*

315 We found that the early-diverging sucker genera *Myxocyprinus*, *Cycleptus*, *Carpiodes*, and
316 *Ictiobus* formed well-supported clades in most analyses. However, *Ictiobus* relationships were

317 resolved in a polytomy in our analysis of the concatenated mtDNA dataset (Fig. S1). In several
318 other trees, including those based on our total-evidence and mtDNA + morphology datasets
319 (Figs. 3 and S2), the *Ictiobus* clade received weak BPP support and eventually collapses into a
320 paraphyletic grade. In the analysis of the four-locus dataset with higher numerical sampling in
321 this clade, *I. bubalus* and *I. niger* relationships had high BPP support (BPP = 0.94–1) but were
322 para-/polyphyletic, leaving their relationship to *I. cyprinellus* uncertain (Fig. S4).

323 *Phylogenetic relationships among the Catostominae*

324 Within the largest sucker subfamily, Catostominae, we consistently resolved clades with the
325 tribes Thoburniini + Moxostomatini and Erimyzonini + Catostomini across analyses. Our more
326 robust, multilocus and total-evidence trees resolved these relationships with definitive support
327 values of BPP = 0.99–1. Within the Thoburniini + Moxostomatini clade, we consistently inferred
328 the genus *Thoburnia* to be paraphyletic, with *T. atripinnis* sister to a clade containing the three
329 *Hypentelium* species (mostly BPP = 1). The sole exception to this was that our morphology tree
330 resolved *Thoburnia* as monophyletic with *T. atripinnis* sister to all other *Thoburnia* with strong
331 support (BPP = 1). Within the *Hypentelium* clade, we inferred an identical and strongly
332 supported set of relationships of the form (*H. roanokense*, (*H. etowanum*, *H. nigricans*)) in the
333 mtDNA, four-locus, nDNA, and total-evidence gene trees (Figs. 3, S1, S2, S4 and S5). We
334 obtained the same set of relationships in our morphology analysis, but with weak (BPP =0.69)
335 support for the *H. etowanum–H. nigricans* node (Fig. S3). We resolved *Moxostoma* as
336 monophyletic with BPP = 0.89–1, except for a paraphyletic pattern in the morphology consensus
337 tree. Within the Erimyzonini, *Erimyzon* was monophyletic (BPP = 1) and sister to *Minytrema*
338 (e.g. Fig. 3). None of the molecular or total-evidence topologies we inferred resolved
339 *Catostomus* as monophyletic relative to *Chasmistes*, *Deltistes*, or *Xyrauchen*. Here, yet again,

results from the morphology tree departed from our other results, failing to resolve relationships among these or virtually any other catostomine lineages with strong support (Figs. 1C and S3). As a result, we do not discuss the morphology consensus topology further in this section.

Relationships within the Erimyzonini and Catostomini were similar to those in previous molecular studies (e.g. Harris et al., 2002; Doosey et al., 2010). Within *Erimyzon*, our results placed *E. sucetta* rather than *E. oblongus* as sister to *E. tenuis*. Within Catostomini, we consistently resolved 9 well-supported major clades within Catostominae (e.g. Figs. 3, S1 and S4). Although relationships among these clades received varying posterior support, the species groups we identified were highly supported in multiple analyses and provide more tenable phylogenetic hypotheses than previously proposed for this tribe (Smith, 1992; Smith et al., 2002). 'Clade 4' corresponded to a monophyletic Erimyzonini, while clades 5 through 9 included various Catostomini subclades composed largely of *Catostomus* samples. For conciseness, we provide an in-depth assessment of relationships only within one clade, 'Clade 5', having the most important phylogenetic and taxonomic implications; more granular presentation and discussion of relationships within and among Catostomini clades 1–3 and 5–9 is provided at the end of Appendix S1. 'Clade 5' corresponded mostly to the subgenus *Pantosteus* (Unmack et al. 2014) and was sister to the remaining Catostomini. Within Clade 5, we consistently resolved *C. nebuliferus* + *C. plebeius* as sister to a clade containing all remaining *Pantosteus*, with (*C. platyrhynchus*, (*C. santaanae*, (*C. clarkii*, (*C. d. discobolus*, *C. d. yarrowi*)))) (e.g. Fig. 3). Alternative topologies inferred for this clade involved rearrangements placing *C. santaanae* sister to *C. clarkii*, but with non-significant posterior support (e.g. Figs. S1 and S4).

Bayesian total-evidence dating and relaxed-clock partitioning

Bayes factor comparisons of 12 clock-partitioning models showed that removing mtDNA 1st and 3rd codon sites deemed to be problematic during PI profiling (see below) progressively improved model log-marginal likelihoods and posterior evidence (Table 4). For example, codon-partitioned relaxed-clock models were overwhelmingly supported over simpler models allocating a single relaxed clock to all data subsets, or all DNA subsets. The most complex model allocating relaxed clocks to each data subset but including only mtDNA 2nd positions, M_{12}, was decisively supported as the best model; compared with this model, other subset and clock schemes produced negative improvements to the model, indicated by negative log Bayes factors. Independent runs of the best-supported BEAST model achieved ESS scores over 100 for nearly all parameters and converged on similar phylogeny and parameter estimates including mean and 95% highest posterior densities (HPDs; i.e. credible intervals) for sucker t_{MRCA}s. Also, clocks on different data subsets exhibited substantial among-lineage rate heterogeneity, with posterior means and 95% HPD intervals of the 'ucldStdev' (uncorrelated lognormal relaxed clock standard deviation) and 'coefficientOfVariation' (coefficient of variation of branch-specific rates) statistics excluding zero (Fig. S7); thus, relaxed-clock models were warranted by the data.

In the final total-evidence dating analysis, the mean posterior age estimate for the t_{MRCA} of all suckers was 63.16 Ma in the Late Cretaceous, with credible intervals ranging from Late Cretaceous to the Paleocene–Eocene boundary (95% HPD: [54.02, 74.6]; Fig. 4). The four sucker subfamilies had variable posterior age estimates ranging approximately an order of magnitude. The Cycleptinae had the youngest posterior age estimate of 5.07 Ma in the Pliocene (95% HPD: [0.87, 10.23]). Following their Late Cretaceous origin based on a stem age corresponding to the MRCA of all suckers, the posterior t_{MRCA} estimate for Myxocyprininae dated their diversification to 42.27 Ma in the early to mid-Eocene (95% HPD: [39.67, 54.58]).

Subsequently, Catostominae species diversified since an intermediate posterior age of 34.37 Ma near the Eocene–Oligocene boundary (95% HPD: [25.54, 42.77]), and Ictiobinae species had the oldest posterior age estimate, dating to 49.69 Ma in the early Eocene (95% HPD: [48.88, 52.52]). The catostomine tribes diverged approximately 29.87 Ma in the Oligocene (Catostomini + Erimyzonini) and 20.78 Ma in the early Miocene (Moxostomatini + Thoburniini). The genera *Catostomus* and *Moxostoma*, which correspond to tribes Catostomini and Moxostomatini, diversified since 17.65 Ma and 15.25 Ma ago in the early-mid Miocene, respectively (Fig. 4).

Phylogenetic informativeness profiles

We evaluated potential impacts of phylogenetic signal on incongruent subfamily relationships by estimating resolution probabilities, over the subfamily divergence epoch (~63.2–34.4 Ma) spanning the divergence of sucker subfamilies. Overall, mtDNA 1st and 3rd codon position data subsets exhibited among the highest PI values, but with distinct Miocene peaks followed by declining PI towards the root (Fig. 5). This suggested a prominent loss of evolutionary information due to homoplasy; therefore, we excluded these sites from final divergence dating analyses, in order to avoid potentially negative effects on the topology and time-calibrated branch lengths (Dornburg et al., 2014). All other molecular data subsets had substantial and relatively constant predicted PI decaying over Paleocene or Eocene to present (recent spikes are anomalies). Morphological characters had slightly higher signal than IRBP sites and exhibited stability before decaying 20 Ma to present (Fig. 5). In addition to nearly constant net PI (Fig. 5), the retained character subsets also had low probabilities of phylogenetic noise or polytomies, with notable increases in the probability of an incorrect topology only for internode distances less than ~0.35(t_0), or <10.0 million years (Myr) (Fig. S9).

Discussion

Sucker phylogeny and incongruence of subfamily lineages

Our phylogenetic reconstructions of Catostomidae relationships are similar to several previous morphological and molecular studies. For example, others have hypothesized that Catostomidae is monophyletic in studies focusing on suckers (e.g. Doosey et al., 2010; Ferris & Whitt, 1978; Harris & Mayden, 2001; Smith, 1992; Fig. 1) and taxonomically broader analyses (e.g. Mayden et al., 2008; Saitoh et al., 2006). Our finding that the four currently recognized sucker subfamilies are monophyletic with definitive support also agrees with earlier phylogenetic studies based on morphology (Smith, 1992) and molecules (Chen & Mayden, 2012; Clements et al., 2012; Doosey et al., 2010; Harris and Mayden, 2001; Harris et al., 2002; Sun et al., 2007). This is perhaps unsurprising, as we reanalyzed previous morphological and molecular datasets alongside new sequence data. Yet, ours are the first results definitively supporting patterns of monophyly at the family and subfamily levels based on dense taxonomic sampling of mtDNA and nuclear gene sequences for all sucker genera and most species, plus total-evidence analyses, with Bayesian posterior probabilities (BPP) at or near 1 across datasets (Figs. 2, 3, and S1–S6).

Previous molecular phylogenetic studies of higher-level sucker relationships have often encountered difficulty in resolving relationships among sucker subfamilies (Chen & Mayden, 2012; Doosey et al., 2010; Harris et al., 2002; Sun et al., 2007). Likewise, relationships among the Myxocyprininae, Cycleptinae, and Ictiobinae lineages were incongruent across analyses of different datasets (Figs. 2, 3, S1–S4), but with two main patterns that we deemed subfamily pattern 1, with the form ((Myxocyprininae, Cycleptinae), Ictiobinae), and subfamily pattern 2, with Myxocyprininae sister to all other sucker subfamilies. We distinguished between these two alternative hypotheses using a topological constraint test based on Bayes factors. The result yielded log-marginal likelihood estimates and Bayes factors giving definitive weight of evidence

against the unconstrained subfamily pattern 1 model. Given this result, a placement of Myxocyprininae as sister to all other suckers seems most probable at this point, thus we favor the patterns of subfamily relationships in our four-locus and Bayesian total-evidence dating topologies that are consistent with this result. Nevertheless, the question still remains: What factors have likely influenced the difficulty of our study and previous studies to resolve phylogenetic relationships among sucker subfamily lineages? Overall, our phylogenetic informativeness analyses highlight two potential explanations for the observed incongruence in subfamily relationships across analyses. First, PI profiling identified the mtDNA 1st and 3rd codon position data subsets as problematic character sets likely compromised by homoplasy due to nucleotide saturation (Fig. 5); hence, we felt justified in excluding these sites from our final divergence dating analyses. Second, our results suggest that predicted phylogenetic noise of the combined datasets over the subfamily divergence epoch (Fig. S9) was most likely a limiting factor for resolving Cycleptinae as sister to Ictiobinae. Whereas internode distances for Catostominae and Myxocyprininae crown clades were generally longer, being ~11 Myr to 18 Myr in length, and associated with significant posterior support, that for Cycleptinae + Ictiobinae had a short internode distance of only 1 Myr (95% HPDs: [0.01,9.1]) and non-significant posterior support in our time tree (Fig. 4). Together with the more frequent incongruence and lower support for Cycleptinae compared to Ictiobinae across our MrBayes topologies, this suggests that Cycleptinae acted as a "rogue taxon" switching positions on the tree (Aberer et al., 2013). We hypothesize that our BEAST total-evidence tree inferred subfamily relationships that were more consistent with Bayes factor tests, and had greater nodal support for early-diverging nodes, by limiting the rogue movements of Cycleptinae.

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

455 The genera *Myxocyprinus*, *Cycleptus*, *Carpiodes*, and *Ictiobus* formed well-supported clades in
456 our results. However, relationships among *Ictiobus* species were resolved in a polytomy or
457 paraphyletic grade in several cases, limiting our resolution of this clade. These *Ictiobus* results
458 disagree with previous mtDNA- or nDNA-based studies resolving relationships among *Ictiobus*
459 species with strong maximum-likelihood bootstrap support (Doosey et al., 2010), and Smith's
460 (1992) hypothesis of relationships among four *Ictiobus* species based mainly on morphology.
461 Given these findings, and that our current results fail to unquestionably place the Cycleptinae as
462 sister to the Ictiobinae, drawing phylogenetic or taxonomic conclusions about *Ictiobus* species
463 relationships would seem premature, and we recommend more in-depth analyses of these taxa.

464 *Phylogenetic relationships among the Catostominae*

465 Within the Catostominae, our multilocus and total-evidence results strongly supported sister
466 relationships between Thoburniini + Moxostomatini, and between Erimyzonini + Catostomini.
467 These findings agree well with previous molecular results (Harris et al., 2002; Clements et al.,
468 2012), except for mitochondrial trees in Doosey et al. (2010) showing the Erimyzonini as sister
469 to all other clades within Catostominae. Interestingly, however, our catostomine relationships
470 conflict with the analysis of Smith (1992), whose morphological data we re-analyzed. In Smith's
471 (1992) sucker phylogeny, the Erimyzonini is resolved as sister to a clade containing what are
472 currently regarded as the Moxostomatini and Thoburniini (Harris & Mayden, 2001; Harris et al.,
473 2002). Smith pointed out that this set of relationships was supported by >20 apomorphies that
474 changed at the node representing the MRCA of these lineages in his parsimony tree. But this
475 conclusion is only as sound as the phylogeny upon which character state transitions were mapped
476 by Smith (1992), which, at this node and several other key nodes, is rejected by our mtDNA,

nDNA, and multilocus trees, as well as total-evidence results from analyzing Smith's data together with molecular datasets.

Regarding our Thoburniini + Moxostomatini clade, genus *Thoburnia* was inferred to be paraphyletic based on the placement of *T. atripinnis* sister to the *Hypentelium* clade, mostly with strong BPP support. Doosey et al. (2010) and Clements et al. (2012) obtained the same relationship for *T. atripinnis*. However, our morphology tree resolved *Thoburnia* as monophyletic, which is consistent with Smith's (1992) original analysis of the morphological data we used, suggesting further data or analyses are needed to clarify these relationships. We consistently inferred *Hypentelium roanokense* as sister to a clade of *H. etowanum* + *H. nigricans*, across molecular, morphological, and total-evidence analyses, though with varying BPP (Figs. 3, S1, S2, S4 and S5). These results match relationships inferred by Buth (1980) using isozyme data reflecting variation at 40 putative loci. By contrast, our results conflict with Smith's (1992) hypothesis, which resolved *H. roanokense* as sister to *H. nigricans*; however, this relationship was based on a single morphological character, dermethmoid spine shape. Taking this into consideration, the broad congruence between multiple data types, as well as our re-analysis of Smith's (1992) data, suggests high confidence in the inference that *H. roanokense* is the earliest diverging lineage in the genus. Within our Moxostomatini clade, *Moxostoma* was monophyletic consistent with previous analyses (e.g. Clements et al. 2012).

Relationships within the Erimyzonini and Catostomini were very similar to those in Harris et al. (2002) and consistent with Doosey et al. (2010), but they contradicted Smith (1992), especially by resolving relationships within *Erimyzon* while placing *E. sucetta* sister to *E. tenuis*. As in previous molecular results for Erimyzonini, *Erimyzon* was monophyletic and sister to *Minytrema* in our results. However, Catostomini genera were not generally obtained as

monophyletic, and in no case was *Catostomus* monophyletic relative to *Chasmistes*, *Deltistes*, or *Xyrauchen*. Our molecular and total-evidence analyses consistently resolved 9 well-supported major clades within Catostominae (e.g. Figs. 3, S1 and S2). Here, we focus on relationships within Clade 5, which corresponded to the former subgenus *Pantosteus*, which Unmack et al. (2014) recently elevated to genus. Smith (1966) recognized six species within *Pantosteus*: *C. clarkii*, *C. columbianus*, *C. discobolus*, *C. plebeius*, *C. platyrhynchus*, and *C. santaanae*. We sampled all of these, including both subspecies of *C. discobolus*, but consistently inferred a polyphyletic *Pantosteus*, with *C. nebuliferus* (recognized as distinct from *C. plebeius* by Miller et al., 2005; Nelson et al., 2004) falling within Clade 5 but *C. columbianus* placed in Clade 8 (discussed below). A clade with *C. nebuliferus* + *C. plebeius* was frequently sister to all remaining *Pantosteus* (e.g. Fig. 3). Notwithstanding incongruent results among analyses in the two papers, the consensus of results from our study and that of Unmack et al. (2014) seems to lend strongest support to the former relationship, with *C. santaanae* sister to a clade containing *C. clarkii* and *C. discobolus* lineages. The polyphyly of *Pantosteus* and *nebuliferus–plebeius* sister relationship are concordant with the results of Doosey et al.'s (2010) analyses using RY-coding for third position mtDNA substitutions, although they inferred *C. santaanae* as sister to a clade containing other members of *Pantosteus*. These results also agree with molecular and morphological analyses of Unmack et al. (2014). However, our results depart from Doosey et al. (2010) and agree better with Unmack et al. (2014) in strongly supporting a sister relationship between *C. columbianus* and *C. tahoensis*. Given the morphological and molecular data analyzed herein support the monophyly and diagnosability of *Pantosteus* relative to *Catostomus*, without rendering *Catostomus* paraphyletic, strongly supports Unmack et al.'s (2014) decision to redefine *Pantosteus* to exclude *C. columbianus*. We note that this taxonomic arrangement is also

consistent with studies on morphological and biochemical variation in western suckers (Koehn, 1969; Smith, 1992; Smith & Koehn, 1971). Also, *C. columbianus* has an open frontoparietal fontanelle, a key diagnostic character of this subgenus, whereas other *Pantosteus* species have the frontoparietal fontanelle closed or reduced to a narrow slit (Smith, 1966).

*"*Catostomus*" polyphyly and introgressive hybridization*

As noted above, *Catostomus* was never resolved in our study as monophyletic relative to *Chasmistes*, *Deltistes*, or *Xyrauchen*, and this result is concordant with phylogenetic results of Doosey et al. (2010) based on mtDNA *ND4/ND5* sequences. Hybridization of *Catostomus* with *Chasmistes*, *Deltistes*, and *Xyrauchen* is well documented (Buth et al., 1987; Markle et al., 2005; Mock et al., 2006; Tranah & May, 2006), and may be related to the non-monophyly of *Catostomus* relative to *Chasmistes* and *Deltistes*. However, while *Xyrauchen texanus* has been documented to hybridize with *C. latippinis* and *C. insiginis* (Buth et al., 1987; Hubbs & Miller, 1953), the majority of these reports evaluate hybridization between *X. texanus* and *C. latipinnis* (Buth et al., 1987). Samples of *Xyrauchen* used in this study originated from the Dexter National Fish Hatchery, which obtained the original hatchery stock of Razorback sucker from Lake Mohave, Arizona, where hybridization with *C. latipinnis* has been documented but allozyme evidence indicates only low levels of introgression of *C. latipinnis* with *X. texanus* (Buth et al., 1987). Therefore, placement of *C. insignis* sister to *X. texanus* here and by Doosey et al. (2010) suggests that introgression is not a factor in either study. As such, "*Xyrauchen*" embedded within *Catostomus* renders the latter polyphyletic. Even if hybridization-mediated introgression were considered as an ad hoc explanation of this pattern, this is difficult to distinguish from the more parsimonious hypothesis of common ancestry, and the available data do not demonstrate that any hybridization events among these taxa have corresponded to the Neogene–present timeframe of

546 their divergences inferred by our time tree. Thus, we advocate the tentative placement of
547 "*Xyrauchen*" into synonymy with *Catostomus* until additional fossil or molecular evidence
548 rejects an inference of common ancestry in favor of Neogene hybridization of these taxa.

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

550 In showing that the best clock-partitioned BEAST models excluded sites identified as
551 problematic in our PI profiling analysis, the results of our Bayes factor clock-partitioning model
552 comparisons bolster Ho and Lanfear's (2010) recommendation that accounting for differences in
553 substitution rates among data partitions through clock-partitioning is not only feasible but also
554 improves phylogenetic divergence dating models. We believe that by employing a clock-
555 partitioning scheme objectively chosen in this way allowed our final BEAST FBD analysis to
556 more correctly estimate topology and rate variance among branches, and better handle rate
557 heterogeneity of the retained characters. However, while the inclusion of fossil taxa in an FBD
558 model in the final BEAST analysis certainly improved our divergence time estimates over what
559 might be obtained using node calibration or tip-dating methods (e.g. Arcila et al., 2015;
560 Gavryushkina et al., 2017), one limitation of this analysis was that nodal support was reduced
561 within the Catostomini and Ictiobinae. This pattern was caused by rogue placements of fossil
562 taxa lacking character data, which were constrained within these crown groups but made up
563 ≥50% of tip sampling (Fig. 4). Still, this mainly caused misleading relationships and lowered
564 nodal support within the Catostomini; after removing fossil taxa, relationships within Ictiobinae
565 would be essentially identical to our preferred MrBayes topologies. After pruning extinct taxa,
566 our time tree will provide a suitable basis for interrogating the comparative biogeography and
567 evolution of all groups of suckers, except for patterns within Catostomini. One alternate way
568 forward for researchers interested in using our results for comparative phylogenetics would be to

convert one of our preferred topologies (Figs. 3 and S2) to an ultrametric tree while constraining subfamily and tribal node ages to mean t_{MRCA} estimates shown in Fig. 4.

A major goal of our study was to use our final total-evidence dating results to test hypotheses on the temporal diversification of suckers. Our divergence dating results (Fig. 4) generally agree with the fossil record but reject or confirm different molecular hypotheses about the temporal diversification of sucker subfamilies. Unsurprisingly, given our incorporation of all fossil sucker lineages in the paleontological literature under an FBD model (accounting for extant and fossil sampling levels), our BEAST results strongly support hypothesis H_4 that Catostomidae lineages have diversified since ~50 Ma in the Early Eocene, which is widely accepted as the minimum age of the origin of suckers based on stratigraphic information for the oldest sucker fossils (review and refs. in Appendix S1). Our results also support Sun et al.'s (2007) proposal, or our hypothesis H_3, that catostomine lineages (in the most speciose sucker subfamily) went on to diversify since ~20 Ma. Indeed, initial divergences and subsequent diversification of all four catostomine tribes has proceeded since around ~34–17.6 Ma in the Eocene–Miocene, with 95% credible intervals ranging from 43 to 11.04 Ma (Fig. 4), and the t_{MRCA}s for ~81% (64/79) of extant species/lineages in our time tree (all catostomines) coincide with the last 20 million years.

In contrast to hypotheses H_3 and H_4 discussed above, we reject two previous molecular hypotheses about the tempo of sucker evolution advanced by Sun et al. (2007). First, we reject hypothesis H_1 because we infer that the Asian Myxocyprininae diverged from North American suckers during the Late Cretaceous, and the 95% credible intervals for this divergence do not overlap with their proposed ~14 Ma Miocene date for the MRCA of *Myxocyprinus* and *Cycleptus*. Second, we reject H_2 given that we infer an early Eocene origin for Ictiobinae,

592 including the extinct †*Amyzon* and †*Vasnetzovia* ictiobine lineages, and this vastly predates Sun
593 et al.'s (2007) proposed origin of the clade. Given Sun et al. (2007) produced divergence time
594 estimates using only cyt*b* divergences and a global molecular clock assuming a 2.0% Myr^{-1}
595 pairwise rate for vertebrate mtDNA, there are too many methodological distinctions between our
596 approach and theirs to pinpoint a single factor causing our results to contrast theirs so starkly.
597 However, our more comprehensive and nuanced approach using Bayesian total-evidence dating
598 not only allowed us to use a realistic FBD tree prior incorporating the speciation-extinction-
599 fossilization sampling process (Gavryushkina et al., 2017), but also permitted estimation of
600 evolutionary rates for each character subset analyzed. Our much slower inferred rate for mtDNA
601 2^{nd} position sites, 6.53×10^{-4} substitutions $site^{-1}$ Myr^{-1}, and in fact all DNA subsets (mean:
602 0.0055 substitutions $site^{-1}$ Myr^{-1}), suggests that Sun et al.'s (2007) age estimates were probably
603 inflated because they used a substitution rate that is unrealistic for suckers. Still, differences
604 between our t_{MRCA} estimates and theirs are not fully accounted for based on substitution rates
605 alone—our rate estimates differ from theirs by roughly half to 1.5-fold, but our dates differ by up
606 to 4.5-fold. Nevertheless, our more appropriate modeling of the evolutionary processes
607 producing variation in sucker DNA sequences and morphological characters, and extant and
608 fossil taxon sampling, has allowed us to estimate older and undoubtedly more accurate
609 divergence dates, especially for deeper nodes in the sucker phylogeny.
610 We infer divergence times for major sucker lineages that are conspicuously older than
611 those recently estimated from multilocus analyses of other North American freshwater fish
612 clades, including sunfishes and black basses (Centrarchidae; Near et al., 2005, 2011), bullhead
613 and madtom catfishes (Ictaluridae; Hardman & Hardman, 2008), and darters (Etheostomatinae;
614 Near et al., 2011). Whereas the diversification of these major lineages has occurred since around

the Eocene–Oligocene transition ~34 Ma, a time of global cooling (Zachos et al., 2001), we infer an earlier Late Cretaceous–Eocene age for the onset of sucker subfamily divergences. This general timeframe for sucker evolution correlates well to the Late Paleocene Thermal Maximum, a period of greater ambient and sea-surface temperatures, higher sea levels, and higher precipitation and humidity (Zachos et al., 2001). Sucker diversification thus appears to have initiated during a period of climate change and sea level rise, which may have facilitated the isolation of ancestral sucker populations. Our results also suggest that approximately 4 to 7 sucker genera may have been present in North America by the Oligocene, a period coinciding with the arrival of minnows in the family Cyprinidae on the continent based on the broader fossil record of North American teleost fishes (e.g. Cavender 1986). Together with the molecular results from other studies above, this implies that the subsequent diversification of these genera, including at least *Cycleptus*, †*Amyzon*, *Ictiobus*, and *Carpiodes* as well as the speciose Catostominae, would have coincided with the diversification of most other major lineages of North American freshwater fishes.

Conclusions

We have presented the results of a phylogenetic analysis of Holarctic sucker fishes (family Catostomidae) drawing on the most comprehensive dataset to date and inferring, separately and jointly, the phylogeny and divergence times of suckers while including fossil taxa as tips. Our molecular and total-evidence results corroborated relationships hypothesized in previous molecular studies and yielded evidence in favor of some new hypotheses of relationships within and among subfamilies, for example with Bayes factor support for Myxocyprininae sister to all other sucker lineages. Our study also highlights how using PI profiling to identify problematic character sets can subsequently improve or provide additional evidence for clock-partitioning

scheme choice during Bayesian relaxed-clock divergence dating. Our divergence-dating results strongly supported the hypotheses that Catostomidae lineages have diversified since ~50 Ma in the Early Eocene (Uyeno & Smith, 1972), and that tribes within the most speciose subfamily, Catostominae, have diversified since ~20 Ma in the Eocene–Miocene (Smith, 1992; Sun et al., 2007). Moreover, we hypothesized that incongruent subfamily relationships were driven in part by problematic mtDNA 1st and 3rd codon sites, and by "rogue taxon" movements of Cycleptinae and fossil taxa, for example in our FBD process time tree. Our analysis could be extended to test this latter hypothesis using additional statistical analyses of rogue taxa (e.g. Aberer et al., 2014) and internode uncertainty (Zhou et al., 2017), and by additional resolution analyses employing Monte Carlo simulations and tests of their assumptions (Townsend et al., 2012), which were beyond the scope of the present study. Nevertheless, our results suggest that future studies of suckers will benefit from using PI profiles as a predictive tool to select loci for subsequent phylogenetic analyses (Dornburg et al., 2014).

Acknowledgements

We are grateful to the following colleagues for providing specimens used in this study: Tom Turner and Alexandra Snyder, Museum of Southwest Biology; Karen Mock and Brian Cardall, Utah State University; Mike Bessert, University of Nebraska Lincoln; Douglas Markle, Oregon State University; Dennis Shiozawa, Monte L. Bean Life Science Museum, Brigham Young University; Vicky Albert and Louis Bernatchez, Université Laval; Andrew Simons, Bell Museum of Natural History, University of Minnesota; and Morgan Raley, North Carolina State Museum of Natural Sciences. We thank Guillermo Ortí and two anonymous reviewers for valuable comments on an earlier version of this manuscript, and we thank Andrew Eckert for

useful discussions of analyses and presentation. We also thank the Brigham Young University Fulton Supercomputing Lab for providing computational resources.

References

Aberer AJ, Krompass D, Stamatakis A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Systematic Biology* **62:**162-166.

Agassiz L. 1854. Notice of a collection of fishes from the southern bend of the Tennessee River, in the State of Alabama. *American Journal of Science and Arts 28 second series* **17:**297-308,353-365.

Arcila D, Pyron RA, Tyler JC, Ortí G, Betancur-RR. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution* **82:**131-145.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29:**2157-2167.

Bagley JC. 2017. justincbagley/PIrANHA: PIrANHA version 0.1.4 [Data Set]. *Zenodo*. Available from <http://doi.org/10.5281/zenodo.571185>.

Bouckaert R, Heled J, Künert D, Vaughan TG, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:**e1003537.

Brandley MC, Wang Y, Guo X, Montes de Oca AN, Fería-Ortíz M, Hikida T, Ota H. 2011. Accommodating heterogeneous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. *Systematic Biology* **60:**3-15.

Bruner JC. 1991. Comments on the genus *Amyzon* (family Catostomidae). *Journal of Paleontology* **65:**678-686.

Buth DG. 1980. Genetics and systematic relationships in the catostomid genus *Hypentelium*. *Copeia* **1980:**280-290.

Buth DG, Murphy RW, Ulmer L. 1987. Population differentiation and introgresseve hybridization of the flannelmouth sucker and of hatchery and native stocks of the razorback sucker. *Transactions of the American Fisheries Society* **116:**103-110.

Cavender T. 1986. Review of the fossil history of North American freshwater fishes. In: Hocutt, C.H, Wiley, E.O., eds. *The zoogeography of North American freshwater fishes*. New York: John Wiley and Sons. 699-724

692 Chang M, Miao D, Chen Y, Zhou J, Pingfu C. 2001. Suckers (Fish, Catostomidae) from the Eocene of
693 China account for the family's current disjunct distributions. *Science in China Series D-Earth*
694 *Sciences* **44:**577-586.
695 Chen W-J, Mayden RL. 2012. Phylogeny of suckers (Teleostei: Cypriniformes: Catostomidae): further
696 evidence of relationships provided by the single-copy nuclear gene IRBP2. *Zootaxa* **3586:**195-
697 210.
698 Chen W-J, Miya M, Saitoh K, Mayden RL. 2008. Phylogenetic utility of two existing and four novel
699 nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: the order Cypriniformes
700 (Ostariophysi) as a case study. *Gene* **423:**125-134.
701 Chow S, Takeyama H. 1998. Intron length variation observed in the creatine kinase and ribosomal protein
702 genes of the swordfish *Xiphias gladius*. *Fisheries Science* **64:**397-402.
703 Clarke JA, Middleton K. 2008. Mosaicism, modules, and the evolution of birds: results from a Bayesian
704 approach to the study of morphological evolution using discrete character data. *Systematic*
705 *Biology* **57:**185-201.
706 Clements MD, Bart Jr. HL, Hurley DL. 2012. A different perspective on the phylogenetic relationships of
707 the Moxostomatini (Cypriniformes: Catostomidae) based on cytochrome-*b* and Growth Hormone
708 intron sequences. *Molecular Phylogenetics and Evolution* **63:** 159-167.
709 Doosey M, Bart H, Saitoh K, Miya M. 2010. Phylogenetic relationships of catostomid fishes
710 (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular*
711 *Phylogenetics and Evolution* **54:**1028-1034.
712 Dornburg A, Townsend JP, Friedman M, Near TJ. 2014. Phylogenetic informativeness reconciles ray-
713 finned fish molecular divergence times. *BMC Evolutionary Biology* **14:**169.
714 Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the
715 BEAST 1.7. *Molecular Biology and Evolution* **29:**1969-1973.
716 Ferris SD, Whitt GS. 1978. Phylogeny of tetraploid catostomid fishes based on the loss of duplicate gene
717 expression. *Systematic Zoology* **27:**189-206.
718 Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2017. Bayesian total-
719 evidence dating reveals the recent crown radiation of penguins. *Systematic Biology* **66:**57-73.
720 Gavryushkina A, Welch D, Stadler T, Drummond AJ. 2014. Bayesian inference of sampled ancestor trees
721 for epidemiology and fossil calibration. *PLoS Computational Biology* **10(12):**e1003919.
722 Hardman M, Hardman LM. 2008. The relative importance of body size and paleoclimatic change as
723 explanatory variables influencing lineage diversification rate: an evolutionary analysis of
724 bullhead catfishes (Siluriformes: Ictaluridae). *Systematic Biology* **57:**116-130.
725 Harris PM, Hubbard G, Sandel M. 2014. Catostomidae: Suckers. In: Warren, M.L, Burr, B.M., eds.

726 *Freshwater Fishes of North America: Volume 1: Petromyzontidae to Catostomidae*. Baltimore,
727 MD: Johns Hopkins University Press. 451-502
728 Harris PM, Mayden RL. 2001. Phylogenetic relationships of major clades of Catostomidae (Teleostei:
729 Cypriniformes) as inferred from mitochondrial SSU and LSU rDNA sequences. *Molecular*
730 *Phylogenetics and Evolution* **20:**225-237.
731 Harris PM, Mayden RL, Perez HE, Garcia de Leon F. 2002. Phylogenetic relationships of Redhorse
732 (*Moxostoma*) and Jumprock (*Scartomyzon*) suckers (Cypriniformes: Catostomidae) based on
733 mitochondrial cytochrome *b* sequence data. *Journal of Fish Biology* **61:**1433-1452.
734 Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses.
735 *Journal of Systematics and Evolution* **46:**239-257.
736 Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth–death process for coherent calibration of
737 divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States*
738 *of America* **111:**E2957-E2966.
739 Hubbs CL. 1930. Materials for a revision of the catostomid fishes of eastern North America.
740 *Miscellaneous Publications of the Museum of Zoology of the University of Michigan*, No. 20.
741 Hubbs CL, Miller RR. 1953. Hybridization in nature between the fish genera *Catostomus* and *Xyrauchen*.
742 *Papers of the Michigan Academy of Science, Arts and Letters* **38:**207-234.
743 Jenkins RE. 1970. *Systematic studies of the catostomid fish tribe Moxostomatini*. Ithaca, NY: Cornell
744 University.
745 Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* **90:**773-795.
746 Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program.
747 *Bioinformatics* **9:**286-298.
748 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz
749 S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. 2012. Geneious Basic: an integrated
750 and extendable desktop software platform for the organization and analysis of sequence data.
751 *Bioinformatics* **28:**1647-1649.
752 Kettratad J, Markle DF. 2010. Redescription of the Tyee Sucker, *Catostomus tsiltcoosensis*
753 (Catostomidae). *Western North American Naturalist* **70:**273-287.
754 Koehn RK. 1969. Hemoglobins of fishes of the genus *Catostomus* from western North America. *Copeia*
755 **1969:**21-30.
756 Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning
757 schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29:**
758 1695-1701.
759 Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character

760 data. *Systematic Biology* **50:**913-925.
761 López-Giráldez F, Townsend JP. 2011. PhyDesign: an online application for profiling phylogenetic
762 informativeness. *BMC Evolutionary Biology* **11:**152.
763 Markle DF, Cavalluzzi MR, Simon DC. 2005. Morphology and taxonomy of Klamath Basin suckers
764 (Catostomidae). *Western North American Naturalist* **65:**473-489.
765 Mayden RL, Chen WJ, Bart Jr. HL, Doosey MH, Simons AM, Tang KL, Wood RM, Agnew MK, Yang
766 L, Hirt MV, Clements MD, Saitoh K, Sado T, Miya M, Nishida M. 2009. Reconstructing the
767 phylogenetic relationships of the earth's most diverse clade of freshwater fishes—order
768 Cypriniformes (Actinopterygii: Ostariophysi): a case study using multiple nuclear loci and the
769 mitochondrial genome. *Molecular Phylogenetics and Evolution* **51:**500-514.
770 Mayden RL, Tang KL, Wood RM, Chen W.J, Agnew MK, Conway KW, Yang L, Simons AM, Bart HL,
771 Harris PM, Li J, Wang X, Saitoh K, He S, Liu H, Chen Y, Mutsumi N, Mia M. 2008. Inferring
772 the Tree of Life of the order Cypriniformes, the earth's most diverse clade of freshwater fishes.
773 *Journal of Systematics and Evolution* **46:**424-438.
774 Miller RR. 1959. Origin and affinities of the freshwater fish fauna of western North America. In: Hubbs
775 CL, ed. Zoogeography. Washington, D.C.: American Association for the Advancement of
776 Science Publication 51. 187-222
777 Miller RR, Minckley WL, Norris SM. 2005. *Freshwater fishes of Mexico*. Chicago, IL: University of
778 Chicago Press.
779 Mock KE, Evans RP, Crawford M, Cardall BL, Janecke SU, Miller MP. 2006. Rangewide molecular
780 structuring in the Utah sucker (*Catostomus ardens*). *Molecular Ecology* **15:**2223-2238.
781 NatureServe, 2013. *Moxostoma lacerum*. The IUCN Red List of Threatened Species. Version 2015.2.
782 <www.iucnredlist.org>. Last accessed on 01 August 2015.
783 NatureServe, 2014. *Chasmistes cujus*. Red List of Threatened Species. Version 2015.2.
784 <www.iucnredlist.org>. Last accessed on 01 August 2015.
785 Near TJ, Bolnick DI, Wainwright PC. 2005. Fossil calibrations and molecular divergence time estimates
786 in centrarchid fishes (Teleostei: Centrarchidae). *Evolution* **59:**1768-1782.
787 Near TJ, Bossu CM, Bradburd GS, Carlson RL, Harrington RC, Hollingsworth Jr., et al. 2011. Phylogeny
788 and temporal diversification of darters (Percidae: Etheostomatinae). *Systematic Biology* **60:**565-
789 595.
790 Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, et al. 2012. Resolution of ray-finned
791 fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences of*
792 *the United States of America* **109:**13698-13703.
793 Nelson JS, Crossman EJ, Espinosa-Perez H, Findley LT, Gilbert CR, Lea RN, et al. 2004. *Common and*

794 *Scientific Names of Fishes from the United States, Canada, and Mexico*. Bethesda, MD:
795 American Fisheries Society Special Publication 29.
796 Pagel M, Meade A. 2014. BayesTraits (1st edition). Reading, UK. Available from
797 <http://www.evolution.rdg.ac.uk/BayesTraits.html>.
798 Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*
799 **21:**676-679.
800 Pyron RA. 2011. Divergence time estimation using fossils as terminal taxa and the origins of
801 Lissamphibia. *Systematic Biology* **60:**466-481.
802 Rambaut A, Drummond AJ. 2014. Tracer v1.5. Available from <http://beast.bio.edu.ac.uk/Tracer>.
803 Robins CR, Raney EC. 1956. Studies of the catostomid fishes of the genus *Moxostoma*, with descriptions
804 of two new species. Cornell University Agricultural Experiment Station Memoir 343, 1-56.
805 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. 2012. MrBayes 3.2:
806 efficient Bayesian phylogenetic inference and model choice across a large model space.
807 *Systematic Biology* **61:**539-542.
808 Sagahi-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length
809 polymorphism in barley: Mendelian inheritance, chromosomal location, and population
810 dynamics. *Proceedings of the National Academy of Sciences of the United States of America*
811 **81:**8014-8019.
812 Saitoh K, Sado T, Mayden RL, Hanzawa N, Nakamura K, Nishida M, et al. 2006. Mitogenomic evolution
813 and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first evidence
814 toward resolution of higher-level relationships of the world's largest freshwater fish clade based
815 on 59 whole mitogenome sequences. *Journal of Molecular Evolution* **63:**826-841.
816 Smith GR. 1966. Distribution and evolution of the North American catostomid fishes of the subgenus
817 *Pantosteus*, genus *Catostomus*. Miscellaneous Publications of the Museum of Zoology of the
818 University of Michigan **129:**1-132.
819 Smith GR. 1992. Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America
820 and Asia. In: Mayden RL, ed. *Systematics, historical ecology, and North American freshwater*
821 *fishes*. Stanford, CA: Stanford University Press. 778-813
822 Smith GR, Dowling TE, Gobalet KW, Lugaski T, Shiozawa DK, Evans RP. 2002. Biogeography and
823 timing of evolutionary events among Great Basin fishes. In: Hershler R. et al., eds. *Great Basin*
824 *aquatic systems history*. Washington, D.C.: Smithsonian Institution Press. 175-234
825 Smith GR, Koehn RK. 1971. Phenetic and cladistic studies of biochemical and morphological
826 characteristics of *Catostomus*. *Systematic Zoology* **20:**282-297.
827 Sun YH, Xie CX, Wang WM, Liu SY, Treer T, Chang MM. 2007. The genetic variation and

biogeography of catostomid fishes based on mitochondrial and nucleic DNA sequences. *Journal of Fish Biology* **70:**291-309.

Tranah GJ, May B. 2006. Patterns of intra- and interspecies genetic diversity in Klamath River Basin suckers. *Transactions of the American Fisheries Society* **135:**306-316.

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

Townsend JP, Leuenberger C. 2011. Point of View: Taxon sampling and the optimal rates of evolution for phylogenetic inference. *Systematic Biology* **60:**358-365.

Townsend JP, López-Giráldez F, Friedman R. 2008. The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. *Journal of Molecular Evolution* **67:**437-447.

Townsend JP, Su Z, Tekle Y. 2012. Phylogenetic signal and noise, predicting the power of a dataset to resolve phylogeny. *Systematic Biology* **5:**835-849.

Unmack PJ, Dowling TE, Laitinen NJ, Secor CL, Mayden RL, Shiozawa DK, Smith GR. 2014. Influence of introgression and geological processes on phylogenetic relationships of western North American mountain suckers (*Pantosteus*, Catostomidae). *PLoS One* **9:**e90061.

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

Warren Jr. ML, Burr BM, Walsh SJ, Bart HL, Cashner RC, Etnier DA, et al. 2000. Diversity, distribution, and conservation status of the native freshwater fishes of the southern United States. *Fisheries* **25:**7-31.

Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* **60:**150-160.

Zhou X, Lutteropp S, Czech L, Stamatakis A, von Looz M, Rokas A. 2017. Quartet-based computations of internode certainty provide accurate and robust measures of phylogenetic incongruence. *bioRxiv* 168526. <https://doi.org/10.1101/168526>.