Exon-based phylogenomics strengthens the phylogeny of Neotropical cichlids and identifies remaining conflicting clades (Cichlomorphae: Cichlidae: Cichlinae)

Katriina L. Ilves[1,2*], Dax Torti[3,4], and Hernán López-Fernández[1]

[1] Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON M5S 2C6 Canada

[2] Current address: Biology Department, Pace University, 1 Pace Plaza, New York, NY 10038

[3] Donnelly Sequencing Center, University of Toronto, 160 College Street, Toronto, ON M5S 3E1 Canada

[4] Current address: Ontario Institute for Cancer Research, MaRS Center, 661 University Avenue, Toronto, ON M5G 0A3 Canada

*Corresponding author at: Biology Department, Pace University, 1 Pace Plaza, New York, NY 10038; email: katriina.ilves@gmail.com

Abstract

The phenotypic, geographic, and species diversity of cichlid fishes have made them a group of great interest for studying evolutionary processes. Here we present a targeted-exon next-generation sequencing approach for investigating the evolutionary relationships of cichlid fishes (Cichlidae), with a particular focus on the Neotropical subfamily Cichlinae using a set of 923 primarily single-copy exons designed through mining of the Nile tilapia (*Oreochromis niloticus*) genome. Sequence capture and assembly were robust, leading to a complete dataset of 415 exons for 139 species (147

terminals) that consisted of 128 Neotropical species, six African taxa, and five Indo-Malagasy cichlids. Gene and species trees were calculated using alternative partitioning schemes and reconstruction methods. In general, all methods yielded similar topologies to previously hypothesized relationships within the Cichlinae and clarified several relationships that were previously poorly supported or in conflict. Additional work will be needed to fully resolve all aspects of Cichlinae phylogeny. Overall, this approach yielded a well-resolved phylogeny of Neotropical cichlids that will be of utility for future assessments of the evolutionary and ecological processes within this diverse group of fishes. Furthermore, the general methodology employed here of exon targeting and capture should be applicable to any group of organisms with the availability of a reference genome.

1. Introduction

Neotropical cichlid fishes are rapidly becoming a model to understand the evolutionary history and biogeography of the exceptionally diverse Neotropical freshwater fish fauna (e.g. Hulsey and García de León 2005, 2006; López-Fernández et al. 2013; Arbour and López-Fernández 2014, 2016; McMahan et al. 2013; Říčan et al. 2013; Astudillo-Clavijo et al. 2015; Tagliacollo et al. 2015; Burress 2016). Likewise, the emergence of Neotropical cichlids as models of adaptive diversification in riverine environments is starting to provide a meaningful complement to the long-established studies of adaptive radiation of lacustrine cichlids in Africa (e.g. Fryer and Iles 1972; Kocher 2004; Wagner et al. 2012; Brawand et al. 2014; Seehausen 2015). Nevertheless, continuing the evolutionary study of Neotropical cichlids depends on the availability of a

robust phylogenetic framework that allows reliable reconstruction of divergence times, supports comparative analysis of lineage and phenotype divergence, and clarifies our understanding of biogeographic history.

Numerous studies have addressed the intergeneric and higher level relationships of Neotropical cichlids or some of their clades, and a relatively clear phylogenetic structure for the subfamily has emerged over the last two decades (e.g. Kullander 1998, Farias et al. 1999, 2000; Hulsey et al. 2004; Concheiro-Pérez et al. 2007; Musilová et al. 2008, 2009; Smith et al. 2008; López-Fernández et al. 2010; Hulsey and Hollingsworth 2010; McMahan et al. 2013; Říčan et al, 2013, 2016). While these analyses have resulted in an increasingly stable understanding of Cichlinae relationships, a well-established taxonomy at the tribe level, and a relatively robust set of relationships among genera, a fully resolved and unambiguously supported phylogeny of Neotropical cichlids has yet to be achieved. This is particularly true of several basal relationships among genera or groups of genera within the three main tribes, Geophagini, Cichlasomatini and Heroini that remain poorly resolved or supported (e.g. López-Fernández et al. 2010; McMahan et al. 2013; Říčan et al. 2016).

Most analyses of Neotropical cichlid phylogeny have been based on relatively few loci (usually 10 or less) and often have been heavily informed by mitochondrial data (e.g. Říčan et al. 2008, Smith et al, 2008, López-Fernández et al. 2010, McMahan et al. 2013; Friedman et al. 2013). These studies are therefore limited in their ability to provide robust phylogenetic analyses, especially in the light of sequence saturation (e.g. Farias et al. 2001; López-Fernández et al. 2005), conflicting signal between nuclear and mitochondrial data (Dornburg et al. 2014; Říčan et al. 2016), and extensive basal short

branches that often receive poor statistical support (López-Fernández et al. 2005, 2010). Beyond these well-known limitations, the small size of these datasets does not allow the incorporation of species tree approaches to phylogenetic analyses, and thus, all these studies are susceptible to producing misleading relationships due to conflict between gene trees and species trees (e.g. Maddison 2007; Edwards 2009). A recent study by Říčan et al. (2016) attempted to circumvent some of these potential limitations by using a large dataset of concatenated Single Nucleotide Polymorphisms (SNPs) derived from Restriction Enzyme Associated DNA (ddRAD). They obtained a largely well-supported tree with the largest taxon sampling of Central American cichlids to date, but their analysis was not able to unambiguously resolve some relationships and was limited to only one clade within Cichlinae. Moreover, Říčan et al.'s (2016) dataset is not amenable to be analyzed under coalescent-based methods and therefore is not suitable to identify potential conflicts in phylogenetic relationships derived from the effects of deep coalescence (e.g. Edwards 2009, Heled and Drummond 2010).

This latter point is relevant because the Neotropical cichlid phylogeny is plagued by short basal branches that, in previous work, have been interpreted as evidence of rapid diversification (López-Fernández et al. 2005, 2010). Concordant with this interpretation, comparative studies are revealing patterns of rapid early lineage and phenotype diversification through habitat and diet-related morphological diversification (e.g. López-Fernández et al. 2013; Arbour and López-Fernández 2014.). These short branches, however, also represent a problem in phylogenetic reconstruction because rapid divergence can result in marked incongruence between gene divergence patterns and species divergence patterns due to population-level phenomena such as deep coalescence

(e.g. Kubatko and Degnan 2007; Edwards 2009). Traditional phylogenetic methods such as those generally used in Neotropical cichlid analyses assume that loci in the genome diverge in concert with the species-level divergence of their corresponding species and that the phylogenetic signal is additive across loci; however, it is widely understood that genes have their own phylogenetic histories (gene trees) and that those histories do not always coincide node-to-node with the phylogenetic history of evolving lineages (species trees) (e.g. Maddison 2007). These incongruences between gene trees and species trees can be caused by several mechanisms, including hybridization and introgression, but for phylogenetic purposes the most widespread and potentially problematic is incomplete lineage sorting (ILS) or deep coalescence (Edwards 2009). From a phylogenetic reconstruction point of view, the challenge is how to identify the signal of individual gene divergence that is congruent with species divergence (Heled & Drummond 2010). It has been repeatedly shown that concatenation of sequences from genes with incongruent gene trees can produce erroneous topologies with high statistical support (Edwards 2009, Heled & Drummond 2010, Mendes 2017). In the case of several Neotropical cichlid clades, short, poorly supported basal branches could result from poor sampling of that portion of the tree or represent times of divergence during which ILS resulted from fast species divergence. In the latter case, concatenated phylogenetic analyses would result in misleading topological reconstructions. Clearly separating these two scenarios, lack of data versus deep coalescence, may not be entirely possible, but the use of many independent nuclear loci (to increase gene tree sampling) and comparing concatenation methods and gene tree methods should help clarifying whether the pattern observed is due to a real evolutionary process or is the result of lack of data, improper analysis or

116 both. Until recently, both the availability of methods to analyze species trees and the
117 technical difficulties associated with identifying and sequencing many nuclear loci made
118 this type of analysis very difficult or plainly inaccessible. With the advent of massively
119 parallel sequencing and the development of methods that allow reconstructing
120 phylogenies using models based on the coalescent, it is now possible to re-evaluate the
121 phylogeny of Neotropical cichlids using phylogenomic datasets of hundreds of loci along
122 with methods that correct for the effects of ILS. Despite these promising developments,
123 however, computational limitations still restrict the options available to analyze truly
124 phylogenomic datasets (e.g. hundreds or thousands of loci) with relatively large numbers
125 of terminals and explicit coalescent analyses of sequence alignments.
126 In this study we used massively parallel sequencing and a custom-designed exon
127 target-capture toolkit (Ilves and López-Fernández 2014) to generate a phylogenomic
128 analysis of Neotropical cichlids under both concatenation and summary coalescent
129 approaches. The main goals of the study were to update currently available hypotheses of
130 Cichlinae relationships with a much larger dataset than previously available, and to
131 leverage the phylogenomic dataset to perform coalescent-based analyses that should
132 account for the possible effects of incomplete lineage sorting on our ability to accurately
133 reconstruct Cichlinae relationships.
134
135 **2. Materials and Methods**
136 *2.1. Taxon selection*
137 Selection of taxa aimed at including as many lineages of Neotropical cichlids as
138 possible. Representatives of all recognized genera were included for all tribes except the

139 Heroini. A recent proliferation of new generic names for taxa in the clade has
140 taxonomically separated several monophyletic lineages that were previously considered
141 congenerics (McMahan et al. 2015, Říčan et al. 2016). More importantly, many
142 previously "orphan" clades variously referred to as '*Cichlasoma' sensu lato* or '*Heros'*
143 have been formally assigned names. Given that these changes occurred after the dataset
144 for this study was generated, representatives for some of these new genera are absent, but
145 our dataset includes members of every clade within Heroini. Moreover, the study revises
146 the identification of a few lineages of questionable identity in a previous analysis led by
147 the senior author of this paper (López-Fernández et al. 2010, and see Říčan et al. 2013).
148 Although those changes did not greatly affect the overall results of those analyses, they
149 did incorporate additional uncertainty to the resolution of relationships within a clade of
150 Heroini and, therefore, it is pertinent to correct them.
151 The dataset used herein closely resembles that of López-Fernández et al. (2010),
152 and focuses on improving sampling of portions of the tree that were most problematic in
153 that study in hopes of improving resolution and support for uncertain groupings such as
154 those among the Cichlasomatini *Acaronia* and *Laetacara* ,as well as some basal
155 relationships among Geophagini and the heroin clade "amphilophines" sensu López-
156 Fernández et al. (2010). Taxon sampling also was designed to test the depth of
157 phylogenetic divergence that could be resolved using the targeted capture probes used in
158 this study (and see Ilves and López-Fernández 2014). At the most recent variation end of
159 the spectrum, we included two individuals of several species to test whether the dataset
160 had enough phylogenetic signal to identify species as monophyletic. We also tested the
161 ability of the dataset to resolve species-level relationships within genera by including a

162 comparatively large number of species within the large *Crenicichla-Teleocichla* clade of
163 Geophagini.
164 Altogether, the dataset included 147 terminals representing 139 species. Of these,
165 128 were Neotropical species and six were African taxa, which represent the sister group
166 of Cichlinae. These taxa include *Heterochromis multidens*, the sister to all African
167 cichlids, and both the reference sequence and a newly generated set of sequences for the
168 Nile Tilapia, *Oreochromis niloticus*, the species from which the probes for target capture
169 were developed (Ilves and López-Fernández 2014). Finally, we included five Indo-
170 Malagasy cichlids which are sister to the African-Neotropical clade. The Sri Lankan
171 cichlid *Pseudetroplus suratensis* was used as the outgroup in all analyses. *Pseudetroplus*
172 and its sister genus, *Etroplus*, are part of the Etroplinae subfamily which forms the sister
173 group to the rest of the family (e.g. Stiassny 1991, Sparks and Smith 2004, Sparks 2008,
174 López-Fernández et al. 2010). The complete list of species, museum catalog numbers,
175 accession numbers (when available) and general locality data are given in Appendix A.
176
177 *2.2. Library preparation for exon capture and sequencing*
178 DNA extraction procedures followed those from Ilves and López-Fernández
179 (2014). Exon capture of 923 targets and sequencing was performed at the Donnelly
180 Sequencing Centre (DSC) at the University of Toronto
181 (http://dsc.utoronto.ca/dsc/index.html) led by D. Torti, using the probes and general
182 protocol described in Ilves and López-Fernández (2014). A double-hybridization
183 procedure of probes to templates was performed, as this was found to significantly
184 increase yield in previous work (Ilves and López-Fernández 2014). Details about the

probe clean-up, library preparation, and exon capture procedure can be found in Appendix B. Paired-end sequencing was conducted on an Illumina HiSeq platform.

2.3. Automation of data processing and analysis

Multiple custom scripts were used to automate data processing and analysis from the initial step of read quality control to the creation of maximum likelihood phylogenetic gene trees. An overview file that describes each step can be found in Appendix C and subsequent appendices contain the specific script files used to perform each particular task or set of tasks.

2.4. Sequence quality control, read assembly and consensus sequence generation

The general procedures and parameters for the quality control of the sequencing reads, assembly of the reads into contigs, and generation of consensus sequences for each exon and species follow those from Ilves and López-Fernández (2014). A custom script was run that automated read quality control, contig assembly, and consensus sequence generation (Appendix D). Briefly, the standalone version of PRINSEQ (Schmieder and Edwards 2011) was used to retain only high-quality reads based on read length and base quality, bowtie2 version 2.1.0 (Langmead and Salzberg 2012) was used to map the contigs to a set of reference sequences from the Nile tilapia (*Oreochromis niloticus*) genome (Appendix E), and SAMtools version 0.1.19-44428cd (Li et al. 2009) was used to generate consensus sequences of the assembled contigs. Custom scripts were used to convert the FASTQ files to FASTA files for alignment (Appendix F) and convert all low quality base calls to “N” and trim terminal “N”s (Appendix G). FASTA sequences were

208 imported into Geneious version 7.1.8 (http://www.geneious.com, Kearse et al. 2012),
209 from which only sequences with a minimum of 100 bp were exported for subsequent use
210 in alignment and phylogenetic analyses.
211
212 *2.5. Sequence alignment and gene tree and species tree phylogenetic analyses*
213 A custom script was used to combine all sequences for each exon into a single file
214 (Appendix H). Sequence alignment for each exon was performed with muscle version
215 3.8.31 (Edgar 2004) using default parameters and was automated using a custom script
216 (Appendix I - MuscleAlignedScript_144taxa_X600_q20_d10_100bp.txt). Each
217 alignment was then manually inspected for quality and completeness in Geneious version
218 7.1.8 (http://www.geneious.com, Kearse et al. 2012). Only alignments that included a
219 sequence of at least 100 bp for each taxon were retained for phylogenetic analyses. The
220 32 opsin exon alignments in the target kit were excluded from all analyses because they
221 represent a family of genes for which duplication and pseudogenization events have been
222 documented, complicating their use in phylogenetics due to paralogy (e.g. Bowmaker
223 2008, Weadick et al. 2012).
224 Coding and non-coding regions of each alignment were inferred from the
225 annotated reference genome. Although the probes, as originally designed were intended
226 to target exon-coding regions, analyses of early alignments during the design of the
227 protocol revealed non-open reading frames for some sequences (see Ilves and López-
228 Fernández 2014). Additionally, subsequent iterations of annotation in the Nile Tilapia
229 genome revealed that some regions originally annotated as coding exonic regions actually
230 corresponded to non-coding fragments. Coding and non-coding regions were identified

231 and correspondingly separated for analyses in this paper. Bootstrapped gene trees (1000
232 replicates) for each exon alignment were generated using RAxML version 8.0.10
233 (Stamatakis 2008) with a GTRGAMMA model and a corresponding partition file with
234 the codon positions and non-coding regions of the exon. This procedure was automated
235 with a custom script (Appendix J).
236 ASTRAL-II version 4.8.0 (Mirarab and Warnow 2015) and STAR (Liu et al.
237 2009) species tree methods with multi-locus bootstrapping (Seo 2008) were run on the
238 resulting bootstrapped RAxML datasets. ASTRAL-II analyses with 500 bootstrap
239 replicates were run locally on a desktop Apple® iMac whereas STAR analyses were run
240 on the STRAW server (Shaw et al. 2013). Because ASTRAL analyses require a
241 maximum likelihood (ML) tree for each gene in addition to the set of bootstrapped tree
242 files, RAxML was used to conduct 40 ML searches on each exon alignment (custom
243 script Appendix K). To assess possible conflicts between species-tree and total-evidence
244 concatenated analyses, we also performed a Maximum Likelihood analysis of
245 concatenated 415 loci in RAxML with one thousand bootstrap replicates.
246
247 *2.6. Computing resources used:*
248 Read quality control, mapping, contig assembly, consensus sequence generation,
249 and sequence alignment and some individual gene alignment RAxML analyses were
250 performed locally on a desktop Apple iMac (3.5 GHz Intel Quad Core i7 with 32GB
251 RAM and 3TB hard drive). Some individual gene alignment RAxML analyses were
252 performed on the GPC supercomputer at the SciNet HPC Consortium (Loken et al.
253 2010). SciNet is funded by the Canada Foundation for Innovation under the auspices of

Compute Canada, the Government of Ontario, Ontario Research Fund - Research Excellence, and the University of Toronto.

2.7. Data availability

Alignments for all exons with complete taxon representation of 100 bp or greater as well as all bootstrapped phylogenetic trees, will be available on Dryad and all other data files, including raw fastq files, are available from corresponding author K.L. Ilves, following publication.

3. Results

3.1. Target capture

923 exons are included in the probe set first developed by Ilves and López-Fernández (2014). 32 exons correspond to opsins, which were excluded from this study *a priori* due to their histories of gene duplication, which left a total of 891 exons as potential targets. Enforcing the restrictions of a complete dataset where every species must have a sequence of at least 100 bp, resulted in a dataset of 428 captured exons. 13 of these alignments were deemed to be 'poor' after visual inspection due to large blocks of ambiguous base pair calls (N) present in multiple species, often comprising over 50% of the total sequence length, thereby resulting in a dataset of 415 exons (~47% of target set) for phylogenetic analysis. Although the minimum sequence length was set at 100 bp, the average minimum sequence length relative to the average reference sequence length of 1136 bp was ~62% (SD 23%), indicating that most species had sequence for most of the

length of each target. The final dataset comprised 147 taxa representing 139 species for 415 exons with a total length of 471,448 bp.

3.2. Depth of phylogenetic resolution in the exon-capture dataset

Analyses of multiple individuals within a species consistently resulted in monophyly at the species level with unambiguous support, as revealed by the concatenated analyses (Appendix L). These results were consistent at various levels of divergence, ranging from resequencing of the African reference *Oreochromis niloticus* to newly sequenced Neotropical species in all major tribes (e.g. Geophagini: *Crenicichla sveni*, Cichlasomatini: *Krobia petitella*, Heroini: *Symphysodon aequifasciatus*). Both concatenated and consensus coalescence analyses of species-level divergence within genera also revealed a large phylogenetic signal that resulted in generally well-resolved and supported relationships among species within genera (e.g. *Crenicichla*). These results generally support the notion that the target capture probes used in this study are adequate to resolve phylogenetic relationships within Cichlidae spanning family to species levels of divergence.

3.3. Phylogenetic analyses

All analyses, regardless of methods, recovered the expected monophyletic Neotropical subfamily Cichlinae as sister to a monophyletic African Pseudocrenilabrinae, both of which in turn are sister to a paraphyletic arrangement of Indian and Malagasy lineages in the subfamilies Etroplinae and Ptychochrominae (Fig. 1). These relationships have long been well established through numerous studies of molecular and

299 morphological datasets and will not be discussed further (Stiassny 1991, Sparks 2004,
300 2008, Smith et al. 2008, López-Fernández et al. 2010, McMahan et al. 2013, Friedman et
301 al. 2013, Matschiner et al. 2016). Relationships among the seven recognized tribes of
302 Neotropical cichlids (Cichlini, Retroculini, Astronotini, Chaetobranchini, Geophagini,
303 Cichlasomatini and Heroini), were likewise recovered unambiguously across analyses.
304 Results herein coincide with previous work in placing the Retroculini and Cichlini clade
305 as sister to the remainder of Cichlinae. The tribes Cichlasomatini and Heroini form a
306 monophyletic clade sister to a clade formed by Astronotini, Chaetobranchini and
307 Geophagini. Interestingly, all analyses recovered Astronotini (genus *Astronotus*) as sister
308 to the Chaetobranchini plus Geophagini clade. This relationship was also found by
309 Matschiner et al. (2016), but has not been universally recovered in previous analyses:
310 Smith et al. (2008) found *Astronotus* as sister to all Cichlinae except Cichlini and
311 Retroculini, López-Fernández et al. (2010) found it as sister to the Cichlasomatini and
312 Heroini clade, and McMahan et al. (2013) recovered the genus as sister to a clade of
313 Chaetobranchini and Geophagini. All these studies were based on a limited set of genes
314 and a large amount of mitochondrial data, suggesting that the much larger dataset used
315 herein resolves the previously unstable placement of the Astronotini.
316 In general, intergeneric relationships within each of the main Neotropical tribes, --
317 Geophagini, Cichlasomatini and Heroini--, were recovered with unambiguous support.
318 Nevertheless, considerable ambiguity was observed across methods in the placement of
319 some lineages within these tribes (see Fig. 1). Given the widespread similarity among
320 topologies, for the remainder of the paper we use the ASTRAL-II coalescent species-tree
321 (Fig. 1) as a reference because of the superior performance of Astral-II as a tool for

322 generating coalescent-based consensus species trees from gene trees (e.g. Arcila et al.
323 2016). Nevertheless, disagreements between this topology and those derived from the
324 STAR species-tree method and the concatenated super-matrix topology are highlighted
325 when pertinent (Fig. 2). Later we discuss the potential impact of topological uncertainty
326 on macroevolutionary analyses and historical biogeographic studies of Neotropical
327 cichlids.

328 All analyses coincide in placing the South American lineages as a paraphyletic
329 arrangement at the base of Heroini, with the genus *Pterophyllum* as the earliest diverging
330 lineage, followed by a clade of *Hoplarchus* and *Hypselecara* and then by a clade in which
331 *Mesonauta* is sister to *Heros*, which is in turn sister to *Symphysodon* and *Uaru*. A similar
332 arrangement has been found by other studies (e.g. McMahan et al. 2013), but the position
333 of *Pterophyllum* can be flipped with that of *Hypselecara* and *Hoplarchus* clade (e.g.
334 López-Fernández et al. 2010). Likewise, all analyses resulted in less than complete
335 support for the position of *Heros*, but the placement does not vary across trees in this
336 study. Support for the placement of *Heros* in previous studies (e.g. López-Fernández et
337 al. 2010) was relatively weak, and despite orders of magnitude increase in the number of
338 loci used herein, the phylogenomic analyses still result in less than 100% bootstrap
339 support for the position of the genus. Contrastingly, the current analyses removed all
340 ambiguity from the previously uncertain placement of the genus *Uaru* (López-Fernández
341 et al. 2010).

342 As in previous studies, the remainder of Heroini is comprised by a geographically
343 non-monophyletic arrangement of South and Central American lineages. Among these, a
344 clade including the genera *Rocio*, *Tomocichla*, *Herotilapia* and *Astatheros* is sister to the

345 rest of Heroini. This arrangement is equivalent to and in the same position of the
346 astatheroines clade sensu Říčan et al. (2016), but it does not include their newly
347 described genus *Cribroheros*, which was not represented in our dataset. Říčan et al.
348 (2016) recently split *Astatheros* sensu López-Fernández et al.'s (2010) into *Astatheros*,
349 *Rocio* and *Cribroheros*, but phylogenetically there is no incongruence among the groups
350 in both studies. The astatheroine clade is sequentially followed by the genus
351 *Australoheros* as sister to all other heroins, an arrangement identical with that found by
352 Říčan et al's (2016) analysis based on concatenated single nucleotide polymorphisms
353 from a restriction site associated DNA (RAD) dataset. This result differs from that of
354 López-Fernández et al. 2010 because the only species of *Australoheros* in that study was
355 inadvertently switched with that of the unrelated *Cryptoheros nanoluteus* (see Appendix
356 A for details, and Říčan et al. 2013). McMahan et al. (2013) also find a different
357 placement for *Australoheros*.

358 Our analyses unambiguously find a monophyletic clade of purely Central
359 American taxa that corresponds with the amphilophines of Říčan et al. (2016), even
360 though our analysis did not include their genera *Cryptoheros*, *Talamancaheros* and
361 *Isthmoheros*. Despite this correspondence, however, our analyses produced different
362 internal relationships among the included genera. For example, while Říčan et al. found
363 *Parachromis* in a subclade with *Amatitlania*, we find it in a different subclade that
364 includes *Amphilophus* (Fig. 1). Likewise, our grouping corresponds roughly with the
365 amphilophines of López-Fernández et al. 2010, but excludes *Trichromis ('Cichlasoma'*
366 *salvini* in their Fig. 1). Even though the results of our phylogenomic analyses have
367 stronger statistical support that those of López-Fernández et al. (2010) and of Říčan et al.

368 (2016, see their Fig. 5), the amphilophine clade found herein contains the largest number
369 of weakly supported nodes in our analyses. It is interesting to note that, among
370 amphilophines, the sequential position of *Petenia* and *Chortiheros* as sister to a clade of
371 *Amatitlania*, *Hypsophrys* and *Neetroplus* was consistently recovered in all analyses, but
372 support for this arrangement was ambiguous (Fig. 1, Appendices L & M). Říčan et al.
373 (2016) recovered these two genera as sister to each other, but none of our analyses
374 supported that relationship.

375 The largest disagreement among topologies obtained in this study, as well as with
376 those from previously published analyses, involves the genera in the informal clade
377 herichthyines, the Caribbean genus *Nandopsis*, and the caquetaines clade containing the
378 genera *Caquetaia*, *Heroina* and the newly named genus *Kronoheros* (Fig. 2A). The
379 relationship between these lineages and the amphilophines also varies among our three
380 analyses, and often differs from relationships found in other studies. It is interesting that
381 our concatenated and STAR coalescent analyses are more similar with each other than
382 either is to the ASTRAL-II topology. In Říčan et al.'s (2016) analysis, the caquetaines
383 were sister to a clade of amphilophines and herichthyins with *Nandopsis* sister to the
384 latter. This is a similar arrangement to that found in our STAR and concatenated
385 analyses, but in our ASTRAL-II topology (Fig. 2A) *Nandopsis* was sister to
386 amphilophines, and in turn, the two were sister to the caquetaines. In all analyses, at least
387 some of these relationships are inconclusively supported, although the concatenated
388 analysis obtained the highest bootstrap values of the three. In Říčan et al's (2016)
389 analysis, *Chiapaheros* remained in a polytomy. In our analysis, it is part of a well-
390 supported clade along with two other well-supported clades: *Thorichthys* and *Trichromis*

391 and a clade comprising *Herichthys*, *Vieja*, *Wajpamheros*, *Chuco* and *Theraps*, but the
392 relationships among these three lineages remains unclear (Fig. 2A).
393 Relationships among Geophagini in all analyses were generally compatible with
394 those previously described by López-Fernández et al. (2012) and comprised by two major
395 clades compatible with those described by López-Fernández et al. (2010, 2012). In the
396 first of these, a clade of *Guianacara* and *Mazarunia* is sister to a clade of
397 "apistogrammines" and "crenicichlines" sensu López-Fernández et al. (2010). Except for
398 a few nodes within the *Crenicichla-Teleocichla* group and another one within *Mazarunia*,
399 relationships within this large clade are largely congruent with previous results and well
400 supported. Additionally, the sister relationship between *Acarichthys* and *Biotoecus* was
401 unambiguously recovered, but support of its sister relationship to *Crenicichla* was always
402 below 100%.
403 A second clade of Geophagini comprised the genus *Biotodoma*, the
404 "geophagines", "mikrogeophagines" and "crenicaratines" of López-Fernández et al.
405 (2010). Monophyly of the "geophagines" (genera *Geophagus*, *Gymnogeophagus* and the
406 *'Geophagus' steindachneri* group) was unambiguously supported by all analyses, but
407 relationships within the group were different in the three topologies (Fig. 2B). Moreover,
408 the relative position of the genus *Biotodoma* and the "mikrogeophagines" (genera
409 *Mikrogeophagus* and *'Geophagus' brasiliensis*) was different in the coalescent-based
410 analyses compared to the concatenated topology. As discussed below, ambiguity in the
411 placement of the "geophagines" genera could have implications for estimating the age of
412 cichlids.

Relationships within Cichlasomatini were generally identical among analyses. In al cases, a clade including *Nannacara*, *Ivanacara* and *Cleithracara* was recovered as the unambiguously supported sister group to the rest of Cichlasomatini. This clade is equivalent to the "nannacarines" sensu López-Fernández et al. (2010), but its position with respect to the rest of the tribe is novel with respect to previous studies. Musilová et al. (2009) found the "nannacarines" as sister to *Laetacara* but with low support, and López-Fernández et al. (2010) found it as sister to a clade of *Laetacara*, *Acaronia* and their "andinoacarines", but also with low support. The genera *Cichlasoma* and *Aequidens* were found as sister to *Krobia* in all studies and as expected from previous work (Musilová et al. 2009, López-Fernández et al. 2010). Our analyses also unambiguously recovered a well-supported monophyly of *Andinoacara* and *Bujurquina* which in turn are sister to *Tahuantinsuyoa*. *Acaronia* and *Laetacara* were recovered as sister to the "andinoacarines" but the relative placement of the genera with respect to each other was not unambiguously supported in any of the analyses. Placement of these two genera has varied across studies: Musilová et al. (2009) found *Acaronia* to be sister to all Cichlasomatini, with *Laetacara* either sister to the remainder of the tribe or to the *Cichlasoma*, *Aequidens* and *Krobia* clade; López-Fernández et al. (2010) recovered a weakly supported sister clade that in turn was sister to "andinoacarines", but support was low.

4. Discussion

Even though higher-level relationships among clades and genera of Neotropical cichlids have become increasingly resolved and supported by recent work, the position of

436 several groups remains uncertain or weakly supported. In this study, we used a recently
437 developed set of exon-targeting probes and massive parallel next generation sequencing
438 to generate a large dataset aimed at resolving Cichlinae relationships that remain poorly
439 supported. Our phylogenomic analyses confirm many relationships previously found
440 among Neotropical cichlids, and provide unprecedented resolution and support for many
441 relationships that were previously weakly supported, especially near the base of the tree.
442 This is particularly true among some genera of Central American lineages in the
443 amphilophines clade, the unambiguous clarification of the relationship among
444 Geophagini clades crenicichlines, apistogrammines and guianacarines (sensu López-
445 Fernández et al. 2010), and the strongly supported position of the Cichlasomatini clade
446 nannacarines as the sister group to the rest of the tribe (Fig. 1 and see Appendices L &
447 M). Nevertheless, despite an increase of two orders of magnitude in data when compared
448 with other sequencing studies (e.g. López-Fernández et al. 2010; Říčan et al. 2013), some
449 relationships remain unclearly established or poorly supported. The three main regions of
450 the tree that continue to resist clear resolution include the Central American
451 herichthyines, the position of the Cichlasomatini genera *Laetacara* and *Acaronia*, and the
452 order of divergence among genera in the Geophagini clade including *Geophagus*,
453 *Gymnogeophagus* and the *'Geophagus' steindachneri* clade (Fig. 2).
454 Previous studies have repeatedly found that Neotropical cichlids diversified over a
455 relatively short time period, as evidenced by short branches and frequently weakly
456 supported relationships near the base of the tree (e.g. Farias et al. 1999, López-Fernández
457 et al. 2005, 2010). Most of the previously unresolved or weakly supported relationships
458 in the López-Fernández et al. (2010) study, which has the most comparable taxon

459 sampling to this study, are resolved with unambiguous support by the phylogenomic
460 analyses presented here regardless of the method used (compare Fig. 1 of both studies).
461 Thus, in combination, the generalized stability of higher-level phylogenetic hypotheses
462 obtained through both summary coalescent and concatenation methods in our analyses,
463 suggests that conflict between gene trees and the species tree is relatively rare among
464 Neotropical cichlids. It appears that incomplete lineage sorting and other sources of
465 misleading phylogenetic signal, such as introgression or hybridization, do not frequently
466 disrupt our ability to reconstruct relationships. We make this assertion with caution,
467 however, because detailed phylogeographic studies of Neotropical cichlids have shown
468 that at least some taxa may be affected by these problems, particularly at more recent
469 levels of divergence (e.g. Willis et al. 2013, 2017). Even more germane to our study, the
470 persistence of unresolved or poorly supported "deep" relationships within a few clades of
471 Cichlinae may indicate a role for deep coalescence or other confounding effects in some
472 early events of divergence among Neotropical cichlids (Fig. 2).
473 This point is particularly relevant because the tribes containing the remaining
474 unresolved or conflictive clades underwent relatively quick adaptive diversification
475 giving origin to a variety of lineages (López-Fernández et al. 2013, Arbour and López-
476 Fernández 2014). It is suggestive that the largest remaining conflicts within the tree
477 involve the Central American herichthyines, including some South American genera with
478 South American distribution and the Caribbean genus *Nandopsis* (Fig. 2A). Recent work
479 has shown that invasion of Central America by the tribe Heroini provided renewed
480 ecological opportunity that allowed this clade to rapidly diversify into a broad variety of
481 ecologically specialized forms (Arbour and López-Fernández 2016). It is conceivable that

482 such rapid adaptive divergence produced incomplete lineage sorting among some heroine
483 lineages, resulting in reduced phylogenetic resolution (e.g. Kubatko and Degnan 2007,
484 Edwards 2009). In fact, the early radiation of the Neotropical cichlid tribes in South
485 America has been similarly shown to have occurred quickly, potentially leaving a
486 similarly conflictive gene tree divergence patterns in other regions of the tree, particularly
487 within Geophagini, which dominates the lineage and functional diversity of South
488 American Neotropical cichlids (López-Fernández et al. 2013, Arbour and López-
489 Fernández 2014, Astudillo-Clavijo et al. 2015). Even if early adaptive radiation in
490 Neotropical cichlids resulted in incomplete lineage sorting in some clades, our results
491 suggest that its effects may not be extensive because, with the exceptions pointed out
492 above, both our concatenated and coalescent analyses recover largely congruent
493 phylogenies. Moreover, most of the relationships recovered herein are congruent with
494 those founds in most previous studies based on much smaller, concatenated datasets (e.g.
495 Říčan et al. 2008, Musilová et al. 2009, López-Fernández et al. 2010, McMahan et al.
496 2013) and with the ddRAD-based analysis of Central American heroines of Říčan et al.
497 (2016).

498 The ability to generate a reliable phylogeny has important consequences beyond
499 the mere systematic applications of the study. Uncertainty about the order of divergence
500 and relationships among genera and higher clades can affect our ability to reconstruct the
501 history of evolutionary divergence in cichlids. Three types of studies could be affected by
502 diminishing but still present uncertainty in Neotropical cichlid relationships. Robust
503 topologies are critical for using the fossil record to calibrate molecular phylogenies and
504 reconstruct the timeline of diversification of a clade. The fossil record of cichlids is

relatively scarce, and considerable debate has ensued regarding the identity and placement of fossils, particularly the recently described Eocene fossils from the Lumbrera formation in Argentina (e.g. Malabarba et al. 2010, 2014, Friedman et al. 2013, López-Fernández et al. 2013). On the one hand, the strong support received by most nodes in our analyses should provide a solid scaffold for time calibration. Unfortunately, one of the most unstable relationships, that involving the geophagine genera *Gymnogeophagus*, *Geophagus* and *'Geophagus' steindachneri*, affects the certainty with which the Eocene fossil †*Gymnogeophagus eocenicus* can be placed on the tree. As can be gleaned from Fig. 2B, the position of *Gymnogeophagus* and its putative fossil taxon in the phylogeny is unstable. As a consequence, calibration of *Gymnogeophagus* remains uncertain because it is unclear whether *Gymnogeophagus* is sister to the broadly distributed *Geophagus* sensus stricto or to the northern Andes clade including *'Geophagus' steindachneri* (Fig. 2B). Therefore, the inconclusively established position of a fossil-bearing clade can have important consequences in both reconstruction of the timeline of divergence and the historical biogeography of South American cichlids.

From a biogeographic point of view, the uncertainty observed among Central American cichlids is likely to have even more dramatic consequences. The incongruent placement of the Caribbean genus *Nandopsis* and of the South American caquetaines limits our ability to accurately recreate the history of heroine invasion of Central America and for understanding the events driving the potential recolonization of South America by some of the Mesoamerican lineages such as *Mesoheros* and the caquetaines. Recent studies that have addressed the historical biogeography of Central American cichlids and their relationships to South America have relied on single reconstructions of the

528 phylogeny relying on either a small number of concatenated loci (e.g. Říčan et al. 2013,
529 Tagliacollo et al. 2015) or on phylogenomic approaches based on concatenation of single
530 nucleotide polymorphisms (Říčan et al. 2016). These studies focus on interpretations of
531 the particular trees found in each of their analyses, but our analyses suggest that the
532 historical biogeography of Central America may require either a more exhaustive
533 analysis of phylogenetic relationships or the reconstruction of competing, alternative
534 scenarios that reflect the current uncertainty in the phylogeny. Based purely on
535 methodological arguments of performance and accuracy (e.g. Mirarab and Warnow 2015,
536 Arcila et al. 2017), it is possible that our ASTRAL-II topology provides a more stable
537 framework for analysis, but the weak support received by nodes in that arrangement
538 suggest that any interpretations of historical biogeography should be done cautiously.
539 Phylogenies are also becoming increasingly important as the framework to
540 perform macroevolutionary analyses of lineage and phenotypic divergence. With a
541 proliferation of comparative methods and the increased availability of well supported and
542 increasingly better dated trees, our ability to infer patterns and process of divergence
543 continues to improve. With several studies recently addressing the evolution of
544 Neotropical cichlids (e.g. Hulsey 2006, López-Fernández et al. 2013, Arbour and López-
545 Fernández 2013, 2014, 2016, Astudillo-Cavijo et al. 2015, Burress 2016), it is pertinent
546 to ask whether changes in the topology may require modification of our emerging
547 understanding of Cichlinae macroevolution. The phylogeny obtained herein is
548 remarkably similar to the López-Fernández et al. (2010) tree used in most of the
549 macroevolutionary analyses listed above. When analyses are performed on a sample of
550 the posterior distribution of the dated phylogeny and not just on the Maximum Clade

Credibility (MCC) tree, topological uncertainty should be reflected in the results such that observed macroevolutionary patterns should be robust to moderate topological changes. This should be particularly true for analyses of phenotypic divergence based on the construction of morphospaces, such as disparity through time and adaptive landscape inferences (e.g. Ingram and Mahler 2013, Slater and Pennel 2014). Likely more sensitive to changes in the tree are lineage through time and rate analyses because they depend on branch lengths in ultrametric trees, which in turn depend on the accuracy of both the topology and of age estimates. Because coalescent summary analyses such as ASTRAL-II and STAR do not provide branch lengths, the species trees estimated here cannot be used directly in comparative analyses that require ultrametric topologies derived from estimates of absolute time.

Finally, generating an ultrametric topology by dating phylogenomic datasets remains a challenge. In principle, there is no reason that phylogenomic species-tree hypotheses cannot be dated, but in practice it is not clear if the incorporation of large phylogenomic datasets and fossil data into actual analyses is computationally feasible (e.g. Bouckaert et al. 2014, Matschiner et al. 2016). Moreover, among other problems, traditional node dating methods require a priori placement of fossils on topologies, further complicating the use of fossils in unresolved clades, such as *†Gymnogeophagus eocenicus*. Alternatively, emerging total evidence dating methods that simultaneously generate phylogenies and age estimates combining molecular, morphological and fossil data may provide more flexibility and accuracy *vis a vis* uncertainty in the molecular phylogenies (e.g. Heath et al. 2014). Whether these methods can be employed in truly

phylogenomic contexts with hundreds of loci and taxa is not yet clear (e.g. Ronquist et al. 2012, Gavryushkina et al. 2016).

5. Conclusion

Phylogenomic analyses of Neotropical cichlids (subfamily Cichlinae) using 415 exons for 139 species in both concatenated and summary statistic coalescent frameworks resulted in generally well-resolved, strongly-supported and broadly-congruent topologies. The topologies obtained are similar to previous hypotheses of relationships among Cichlinae but, in general, provide stronger support for many relationships that previously had weak or conflicting support. The results of our analyses also suggest that the targeted genomic regions contain phylogenetic signal capable of resolving relationships at all levels of divergence within the clade. Nevertheless, we identified several regions of the tree in which relatively early divergence events cannot be reconstructed with certainty because different methods provide conflicting results, none of them with conclusive support. We suggest this incongruence may result from incompletely lineage sorting associated with the early adaptive divergence of the clades in which incongruence is observed. We argue that, despite some disagreements in the placement of these lineages both in this study and in previous analyses, most studies provide a broadly common signal of evolutionary relationships among Neotropical cichlids. A generally better supported phylogeny derived from our phylogenomic analyses should continue to provide a solid framework for the evolutionary analysis of lineage and phenotypic divergence in Neotropical cichlids. Future work will focus in further clarifying relationships within the few recalcitrant clades identified herein, and in leveraging the current dataset as a

596 uniquely strong molecular framework for clarifying the timeline of evolutionary
597 divergence among Neotropical cichlids.
598
599 **Acknowledgements**
600 The authors are grateful to M. Burridge, M. Zur and E. Holm (Royal Ontario Musem
601 [ROM]), M, Stiassny and B. Brown (American Museum of Natural History, New York),
602 M. Sabaj (Academy of Natural Sciences of Drexel University, Philadelphia), R. Rodíles-
603 Hernández (ECOSUR), L.R. Malabarba (Universidade Federal do Rio Grande do Sul,
604 Porro Alegre, Brazil), J. Armbruster (Auburn University Museum, Auburn, Alabama)
605 and Yves Fermon (Muséum National d'Histoire Naturelle, Paris) for curatorial assistance
606 and the loan of tissue samples not available at the ROM collection. We are grateful to the
607 people who assisted during fieldwork related to this work: J.H. Arbour, D. Bloom, R.R.
608 Hernández, F.E. Hauser, N.K. Lujan, N. Meliciano, E.A. Liverpool, C.G. Montaña, G.
609 Ortí, S. Refvik, M. A.P. Röepke, Soria, D.C. Taphorn, M. Tobler, S. C. Willis, K.O.
610 Winemiller, Funding for this work, including fieldwork and laboratory analyses was
611 provided by a Discovery Grant from the Natural Sciences and Engineering Research
612 Council of Canada, ROM Governors and the National Geographic Society grants to HLF.
613 Additional fieldwork was supported by a Coypu Foundation Grant to N.K. Lujan and
614 ROM Trust Funds through HLF. KLI was supported by a Rebanks Postdoctoral
615 Fellowship from the ROM Governors granted to HLF to develop a phylogenomic toolkit
616 for the phylogenetic analysis of cichlids.
617
618 **References**

619 Arbour JH, López-Fernández H. 2014. Adaptive landscape and functional diversity of
620 Neotropical cichlids: implications for the ecology and evolution of Cichlinae
621 (Cichlidae; Cichliformes). J Evol Biol 27:2431–42. doi: 10.1111/jeb.12486
622 Arbour JH, López-Fernández H. 2016. Continental cichlid radiations: functional diversity
623 reveals the role of changing ecological opportunity in the Neotropics. Proc R Soc
624 B Biol Sci 283:20160556. doi: 10.1098/rspb.2016.0556
625 Arcila D, Ortí G, Vari R, Armbruster JW, Stiassny MLJ, Ko KD, Sabaj MH, Lundberg J,
626 Revell LJ, Betancur-R. R, Much. 2017. Genome-wide interrogation advances
627 resolution of recalcitrant groups in the Tree of Life. Nat Ecol Evol. doi:
628 10.1038/s41559-016-0020
629 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard M a., Rambaut
630 A, Drummond AJ. 2014. BEAST 2: A Software Platform for Bayesian
631 Evolutionary Analysis. PLoS Comput Biol 10:1–6. doi:
632 10.1371/journal.pcbi.1003537
633 Bowmaker JK. 2008. Evolution of vertebrate visual pigments. Vision Res 48:2022–41.
634 doi: 10.1016/j.visres.2008.03.025
635 Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim
636 ZW, Bezault E, Turner-Maier J, Johnson J, Alcazar R, Noh HJ, Russell P, Aken
637 B, Alföldi J, Amemiya C, Azzouzi N, Baroiller J-F, Barloy-Hubler F, Berlin A,
638 Bloomquist R, Carleton KL, Conte M a., D'Cotta H, Eshel O, Gaffney L, Galibert
639 F, Gante HF, Gnerre S, Greuter L, Guyon R, Haddad NS, Haerty W, Harris RM,
640 Hofmann H a., Hourlier T, Hulata G, Jaffe DB, Lara M, Lee AP, MacCallum I,
641 Mwaiko S, Nikaido M, Nishihara H, Ozouf-Costaz C, Penman DJ, Przybylski D,

642 Rakotomanga M, Renn SCP, Ribeiro FJ, Ron M, Salzburger W, Sanchez-Pulido
643 L, Santos ME, Searle S, Sharpe T, Swofford R, Tan FJ, Williams L, Young S, Yin
644 S, Okada N, Kocher TD, Miska E a., Lander ES, Venkatesh B, Fernald RD,
645 Meyer A, Ponting CP, Streelman JT, Lindblad-Toh K, Seehausen O, Di Palma F.
646 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature
647 513:375–381. doi: 10.1038/nature13726
648 Dornburg A, Townsend JP, Friedman M, Near TJ. 2014. Phylogenetic informativeness
649 reconciles ray-finned fish molecular divergence times. BMC Evol Biol 14:169.
650 doi: 10.1186/s12862-014-0169-0
651 Edwards SV. 2009. Is a new and general theory of molecular systematics emerging?
652 Evolution 63:1–19. doi: 10.1111/j.1558-5646.2008.00549.x
653 Farias, I.P., Ortí, G., Sampaio, I., Schneider, H., Meyer, A., 1999. Mitochondrial DNA
654 phylogeny of the family Cichlidae: monophyly and fast molecular evolution of
655 the Neotropical assemblage. J. Mol. Evol. 48, 703–711.
656 Farias IP, Ortí G, Meyer A. 2000. Total evidence: molecules, morphology, and the
657 phylogenetics of cichlid fishes. J Exp Zool 288:76–92.
658 Friedman M, Keck BP, Dornburg A, Eytan RI, Martin CH, Darrin C, Wainwright PC,
659 Near TJ, Hulsey CD. 2013. Molecular and fossil evidence place the origin of
660 cichlid fishes long after Gondwanan rifting. Proc R Soc B Biol Sci 280:20131733.
661 doi: http://dx.doi.org/10.1098/rspb.2013.1733
662 Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ,
663 Avryushkina ALG, Eath TRAH, Sepka DATK, Tadler TAS, Elch DAW. 2016.

Bayesian Total-Evidence Dating Reveals the Recent Crown Radiation of Penguins. Syst Biol 0:syw060. doi: 10.1093/sysbio/syw060

Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc Natl Acad Sci. doi: 10.1073/pnas.1319091111

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. Mol Biol Evol 27:570–80. doi: 10.1093/molbev/msp274

Hulsey CD, García de León FJ, Rodiles-Hernández R. 2006. Micro- and macroevolutionary decoupling of cichlid jaws: a test of Liem's key innovation hypothesis. Evolution 60:2096–109.

Hulsey CD, García De León FJ. 2005. Cichlid jaw mechanics: Linking morphology to feeding specialization. Funct Ecol 19:487–494. doi: 10.1111/j.1365-2435.2005.00987.x

Ilves, K.L., López-Fernández, H. 2014. A targeted next-generation sequencing toolkit for exon-based cichlid phylogenomics. Mol. Ecol. Res. 14, 802–811.

Ingram T, Mahler DL. 2013. SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods Ecol Evol 4:416–425. doi: 10.1111/2041-210X.12034

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop

686 software platform for the organization and analysis of sequence data.
687 Bioinformatics 28, 1647-1649.
688 Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model.
689 Nat Rev Genet 5:288–298. doi: 10.1038/nrg1316
690 Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from
691 concatenated data under coalescence. Syst Biol 56:17–24. doi:
692 10.1080/10635150601146041
693 Kullander, S.O., 1998. A Phylogeny and Classification of the Neotropical Cichlidae
694 (Teleostei: Perciformes). In: Malabarba, L.R., Reis, R.E., Vari, R.P., Lucena,
695 Z.M., Lucena, C.A.S. (Eds.), Phylogeny and classification of Neotropical fishes.
696 EDIPUCRS, Porto Alegre, pp. 461–498.
697 Liu, L., Yu, L., Pearl, D.Kk, Edwards, S.V., 2009. Estimating species phylogenies using
698 coalescence times among sequences. Syst. Biol. 58, 468–477.
699 Loken, C., Gruner, D., Groer, L., Peltier, R., Bunn, N., Craig, M., Henriques, T.,
700 Dempsey, J., Yu. C.-H., Chen, Dursi, J.L., Chong, J., Northrup, S., Pinto, J., Knecht, N.,
701 Van Zon, R., 2010. SciNet: lessons learned from building a power-efficient top-
702 20 system and data centre. J. Phys.: Conf. Ser. 256, 012026. doi: 10.1088/1742-
703 6596/256/1/012026.
704 López-Fernández H, Winemiller KO, Honeycutt RL. 2010. Multilocus phylogeny and
705 rapid radiations in Neotropical cichlid fishes (Perciformes: Cichlidae: Cichlinae).
706 Mol Phylogenet Evol 55:1070–1086. doi: 10.1016/j.ympev.2010.02.020
707 López-Fernández H, Taphorn DC, Liverpool EA. 2012. Phylogenetic diagnosis and
708 expanded description of the genus Mazarunia Kullander,1990 (Teleostei :

Cichlidae) from the upper Mazaruni River , Guyana , with description of two new species. 10:465–486.

López-Fernández H, Arbour JH, Winemiller KO, Honeycutt RL. 2013. Testing for ancient adaptive radiations in neotropical cichlid fishes. Evolution 67:1321–37. doi: 10.1111/evo.12038

Maddison W. 1997. Gene trees in species trees. Syst Biol 46:523–536.

Mendes F, Hahn M. 2017. Why Concatenation Fails in the Anomaly Zone. BioRxiv. https://doi.org/10.1101/116509

Matschiner M, Musilová Z, Barth JMI, Starostová Z, Salzburger W, Steel M, Bouckaert R (2016) Bayesian Phylogenetic Estimation of Clade Ages Supports Trans-Atlantic Dispersal of Cichlid Fishes. Syst Biol syw076. doi: 10.1093/sysbio/syw076

McMahan CD, Chakrabarty P, Sparks JS, Smith WL, Davis MP. 2013. Temporal Patterns of Diversification across Global Cichlid Biodiversity (Acanthomorpha: Cichlidae). PLoS One 8:e71162. doi: 10.1371/journal.pone.0071162

McMahan CD, Matamoros WA, Piller KR, Chakrabarty P. 2015. Taxonomy and systematics of the herichthyins (Cichlidae: Tribe Heroini), with the description of eight new Middle American Genera. Zootaxa 3999:211–234. doi: 10.11646/zootaxa.3999.2.3

Mirarab, S., Warnow, T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–i52. doi: 10.1093/bioinformatics/btv234

Musilová Z, Říčan O, Novák J. 2009. Phylogeny of the Neotropical cichlid fish tribe Cichlasomatini (Teleostei: Cichlidae) based on morphological and molecular data, with the description of a new genus. J Zool Syst Evol Res 47:234–247. doi: 10.1111/j.1439-0469.2009.00528.x

Říčan O, Piálek L, Dragová K, Novák J. 2016. Diversity and evolution of the Middle American cichlid fishes (Teleostei: Cichlidae) with revised classificatio. Vertebr Zool 66:1–102.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. Syst Biol 61:973–999. doi: 10.1093/sysbio/sys058

Seehausen O. 2015. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. New Phytol 207:304–312. doi: 10.1111/nph.13450

Seo, T.K. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25, 960–971.

Shaw, T.I, Ruan, Z., Glenn, T.C., Liu, L. 2013. STRAW: Species TRee Analysis Web server. Nucleic Acids Res. 41, W238–W241. doi: 10.1093/nar/gkt377

Slater GJ, Pennell MW. 2014. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. Syst Biol 63:293–308. doi: 10.1093/sysbio/syt066

Smith L, Chakrabarty P, Sparks J. 2008. Phylogeny, taxonomy, and evolution of Neotropical cichlids (Teleostei: Cichlidae: Cichlinae). Cladistics 24:625–641.

Sparks JS. 2008. Phylogeny of the Cichlid Subfamily Etroplinae and Taxonomic Revision of the Malagasy Cichlid Genus Paretroplus (Teleostei: Cichlidae). Bull Am Museum Nat Hist 314:1. doi: 10.1206/314.1

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Thompson AW, Betancur-R R, López-Fernández H, Ortí G. 2014. A Time-Calibrated, Multi-locus Phylogeny of Piranhas, Pacus, and Allies (Characiformes: Serrasalmidae) and a Comparison of Species Tree Methods. Mol Phylogenet Evol In review:242–257. doi: 10.1016/j.ympev.2014.06.018

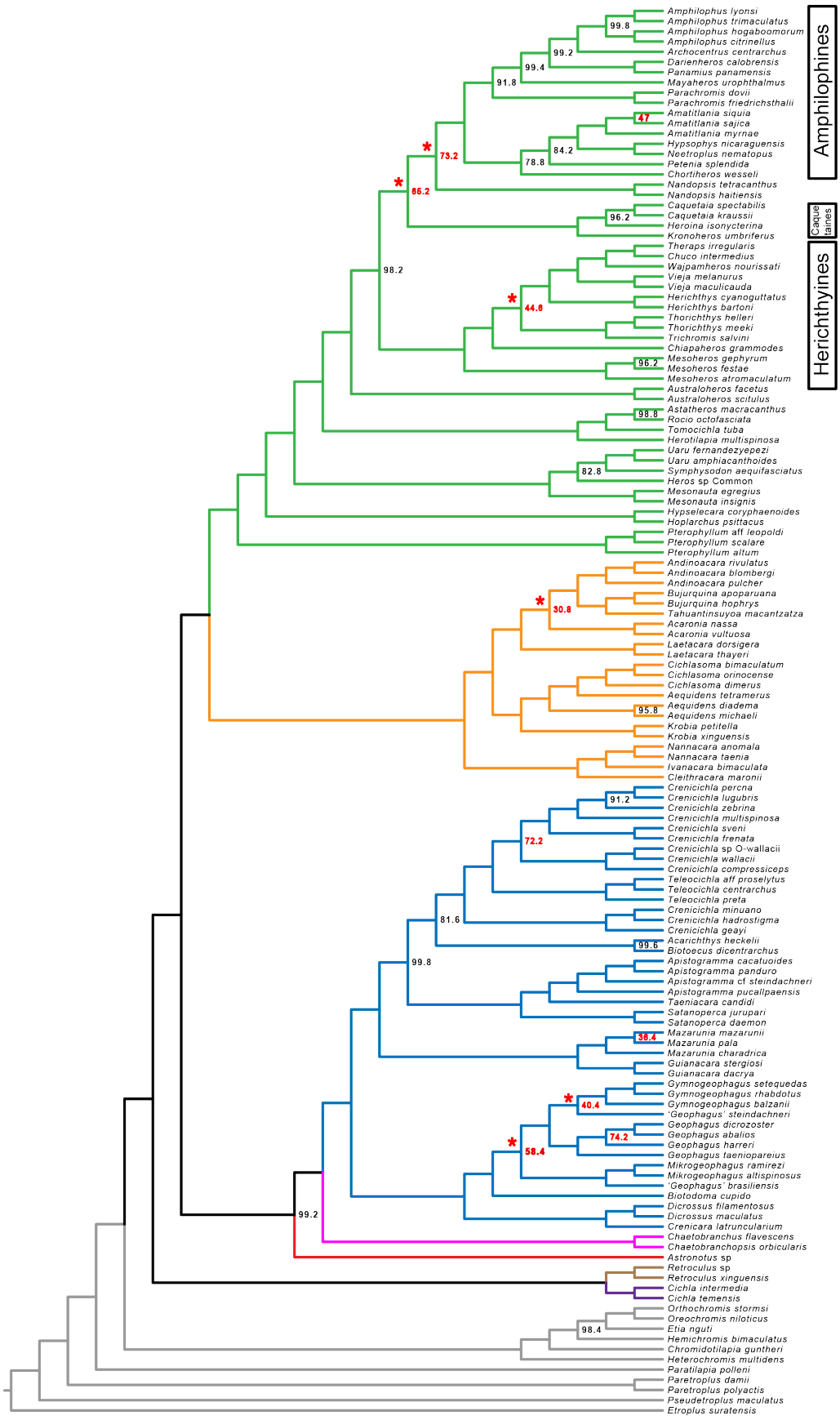
Tonini J, Moore A, Stern D, Shcheglovitova M, Ortí G. 2015. Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions. 1–15. doi: 10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be.Abstract

Wagner CE, Harmon LJ, Seehausen O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366–369. doi: 10.1038/nature11144

Weadick CJ, Loew ER, Rodd FH, Chang BSW. 2012. Visual pigment molecular evolution in the Trinidadian pike cichlid (*Crenicichla frenata*): a less colorful world for neotropical cichlids? Mol Biol Evol 29:3045–60. doi: 10.1093/molbev/mss115

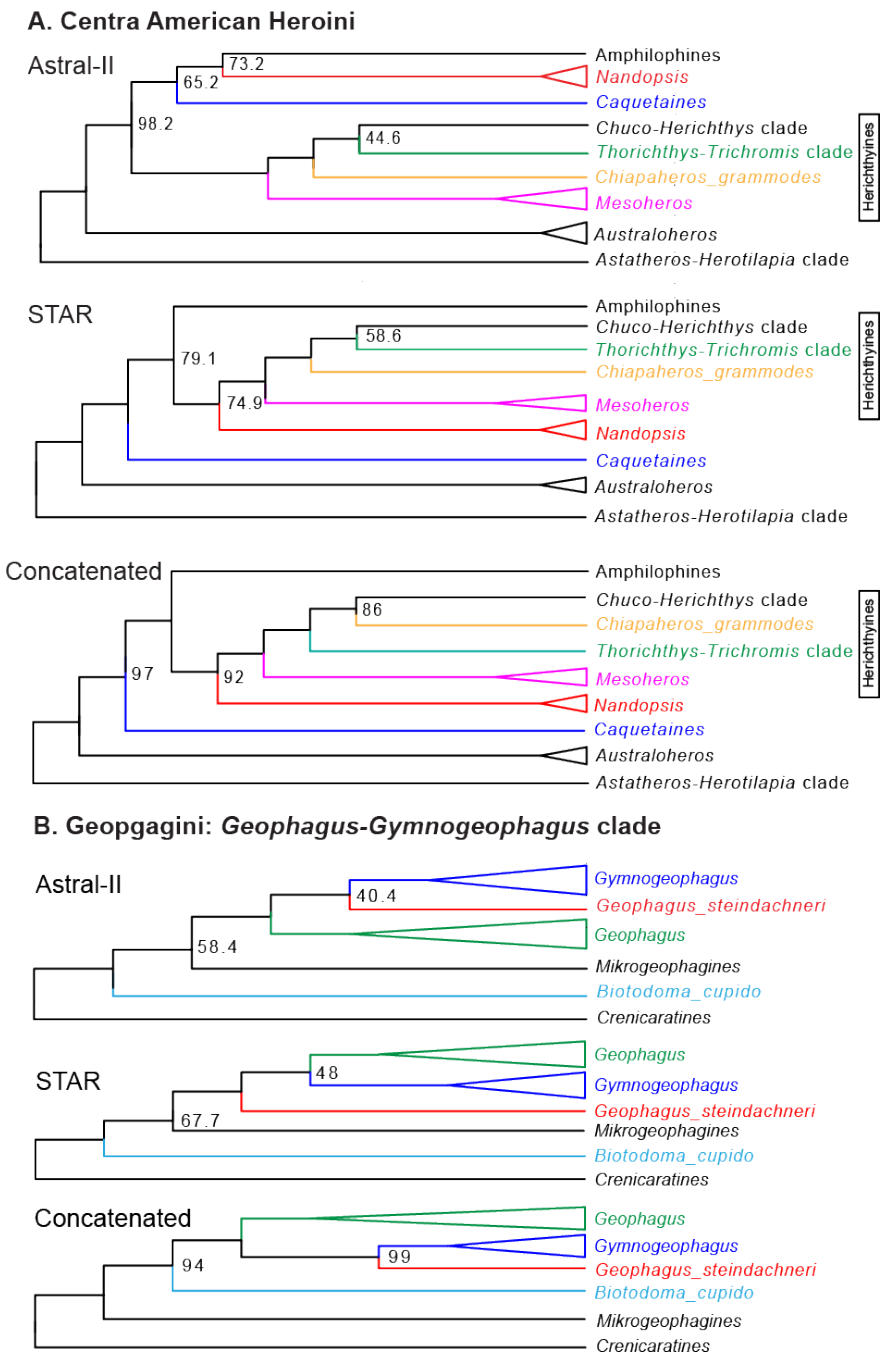
Willis SC, Farias IP, Ortí G. 2013. Testing mitochondrial capture and deep coalescence in Amazonian cichlid fishes (Cichlidae: Cichla). Evolution 68:256–68. doi: 10.1111/evo.12230

Willis SC. 2017. One species or four? Yes!...and, no. Or, arbitrary assignment of lineages to species obscures the diversification processes of Neotropical fishes. PLoS One 12:e0172349. doi: 10.1371/journal.pone.0172349


Amphilophus lyonsi
Amphilophus trimaculatus
Amphilophus hogaboomorum
Amphilophus citrinellus
Archocentrus centrarchus
Darienheros calobrensis
Panamius panamensis
Mayaheros urophthalmus
Parachromis dovii
Parachromis friedrichsthalii
Amatitlania siquia
Amatitlania sajica
Amatitlania myrnae
Hypsophys nicaraguensis
Neetroplus nematopus
Petenia splendida
Chortiheros wesseli
Nandopsis tetracanthus
Nandopsis haitiensis
Caquetaia spectabilis
Caquetaia kraussii
Heroina isonycterina
Kronoheros umbriferus
Theraps irregularis
Chuco intermedius
Wajpamheros nourissati
Vieja melanurus
Vieja maculicauda
Herichthys cyanoguttatus
Herichthys bartoni
Thorichthys helleri
Thorichthys meeki
Trichromis salvini
Chiapaheros grammodes
Mesoheros gephyrum
Mesoheros festae
Mesoheros atromaculatum
Australoheros facetus
Australoheros scitulus
Astatheros macracanthus
Rocio octofasciata
Tomocichla tuba
Herotilapia multispinosa
Uaru fernandezyepezi
Uaru amphiacanthoides
Symphysodon aequifasciatus
Heros sp Common
Mesonauta egregius
Mesonauta insignis
Hypselecara coryphaenoides
Hoplarchus psittacus
Pterophyllum aff leopoldi
Pterophyllum scalare
Pterophyllum altum
Andinoacara rivulatus
Andinoacara blombergi
Andinoacara pulcher
Bujurquina apoparuana
Bujurquina hophrys
Tahuantinsuyoa macantzatza
Acaronia nassa
Acaronia vultuosa
Laetacara dorsigera
Laetacara thayeri
Cichlasoma bimaculatum
Cichlasoma orinocense
Cichlasoma dimerus
Aequidens tetramerus
Aequidens diadema
Aequidens michaeli
Krobia petitella
Krobia xinguensis
Nannacara anomala
Nannacara taenia
Ivanacara bimaculata
Cleithracara maronii
Crenicichla percna
Crenicichla lugubris
Crenicichla zebrina
Crenicichla multispinosa
Crenicichla sveni
Crenicichla frenata
Crenicichla sp O-wallacii
Crenicichla wallacii
Crenicichla compressiceps
Teleocichla aff proselytus
Teleocichla centrarchus
Teleocichla preta
Crenicichla minuano
Crenicichla hadrostigma
Crenicichla geayi
Acarichthys heckelii
Biotoecus dicentrarchus
Apistogramma cacatuoides
Apistogramma panduro
Apistogramma cf steindachneri
Apistogramma pucallpaensis
Taeniacara candidi
Satanoperca jurupari
Satanoperca daemon
Mazarunia mazarunii
Mazarunia pala
Mazarunia charadrica
Guianacara stergiosi
Guianacara dacrya
Gymnogeophagus setequedas
Gymnogeophagus rhabdotus
Gymnogeophagus balzanii
'Geophagus' steindachneri
Geophagus dicrozoster
Geophagus abalios
Geophagus harreri
Geophagus taeniopareius
Mikrogeophagus ramirezi
Mikrogeophagus altispinosus
'Geophagus' brasiliensis
Biotodoma cupido
Dicrossus filamentosus
Dicrossus maculatus
Crenicara latruncularium
Chaetobranchus flavescens
Chaetobranchopsis orbicularis
Astronotus sp
Retroculus sp
Retroculus xinguensis
Cichla intermedia
Cichla temensis
Orthochromis stormsi
Oreochromis niloticus
Etia nguti
Hemichromis bimaculatus
Chromidotilapia guntheri
Heterochromis multidens
Paratilapia polleni
Paretroplus damii
Paretroplus polyactis
Pseudetroplus maculatus
Etroplus suratensis
Amphilophines
Caquetaines
Herichthyines
99.8
99.2
99.4
91.8
47
84.2
78.8
73.2
86.2
96.2
98.2
44.8
96.2
98.8
82.8
30.8
95.8
91.2
72.2
81.6
99.6
99.8
38.4
40.4
74.2
58.4
99.2
98.4

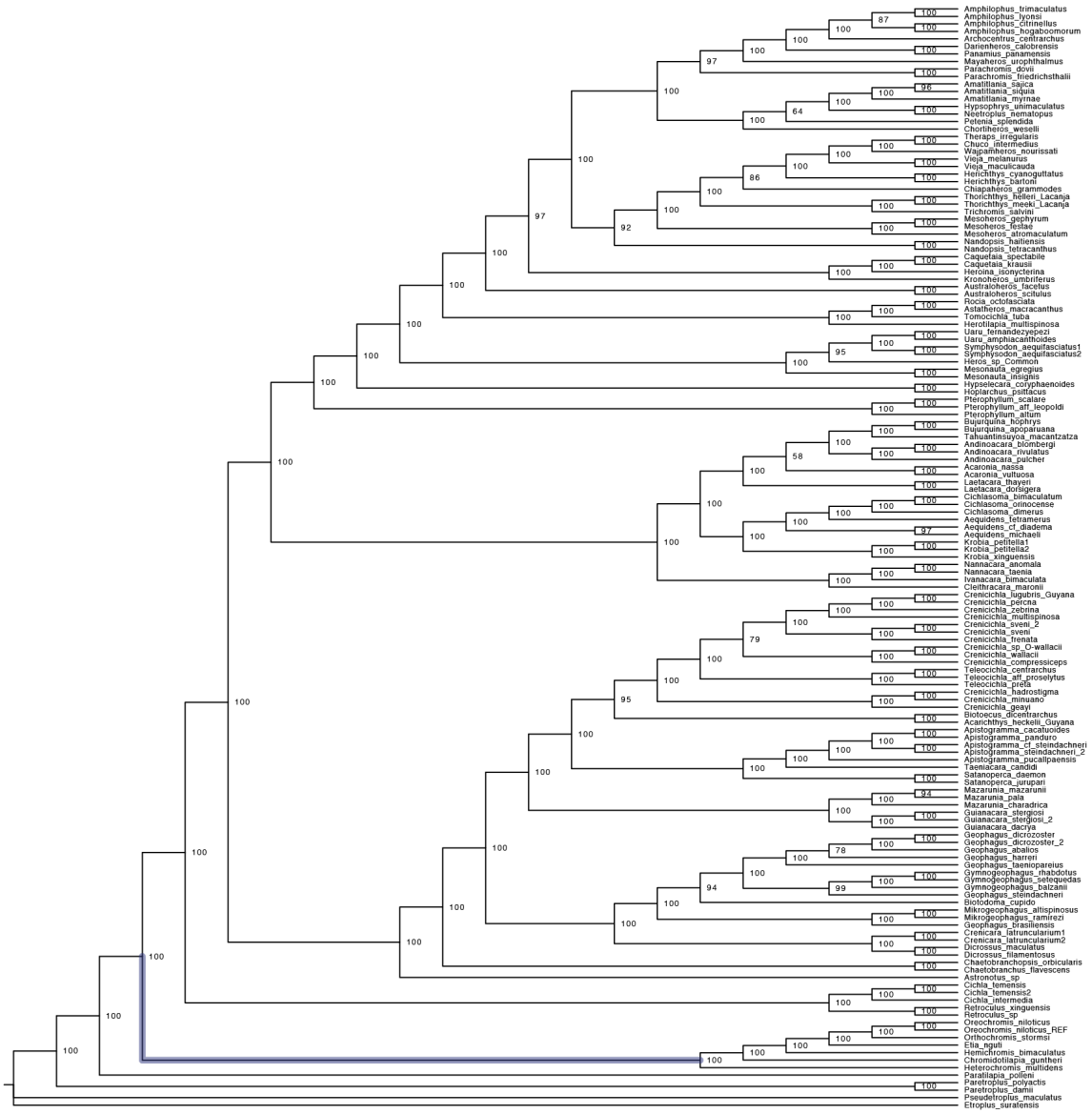
779

Figure 1. Species tree generated with ASTRAL-II for 415 loci comprising 471,448 bp. Colored clades depict formally recognized Neotropical cichlid tribes Heroini (Green), Cichlasomatini (Orange), Geophagini (Blue), Chaetobranchini (Magenta), Astronotini (Red), Retroculini (Brown), and Cichlini (Purple). See text for further discussion of relationships among and within tribes. Node bootstrap support is indicated when pertinent; nodes without labels received 100% support in this analysis. Nodes labeled in red received <75% bootstrap support. Nodes marked with an asterisk (*) represent weakly supported intergeneric relationships with incongruent resolution among two different species-tree and one concatenated phylogenetic analyses. See Figure 2 and Discussion for further analyses of these results. Complete topologies not shown here are provided in Appendices L and M along with node support.

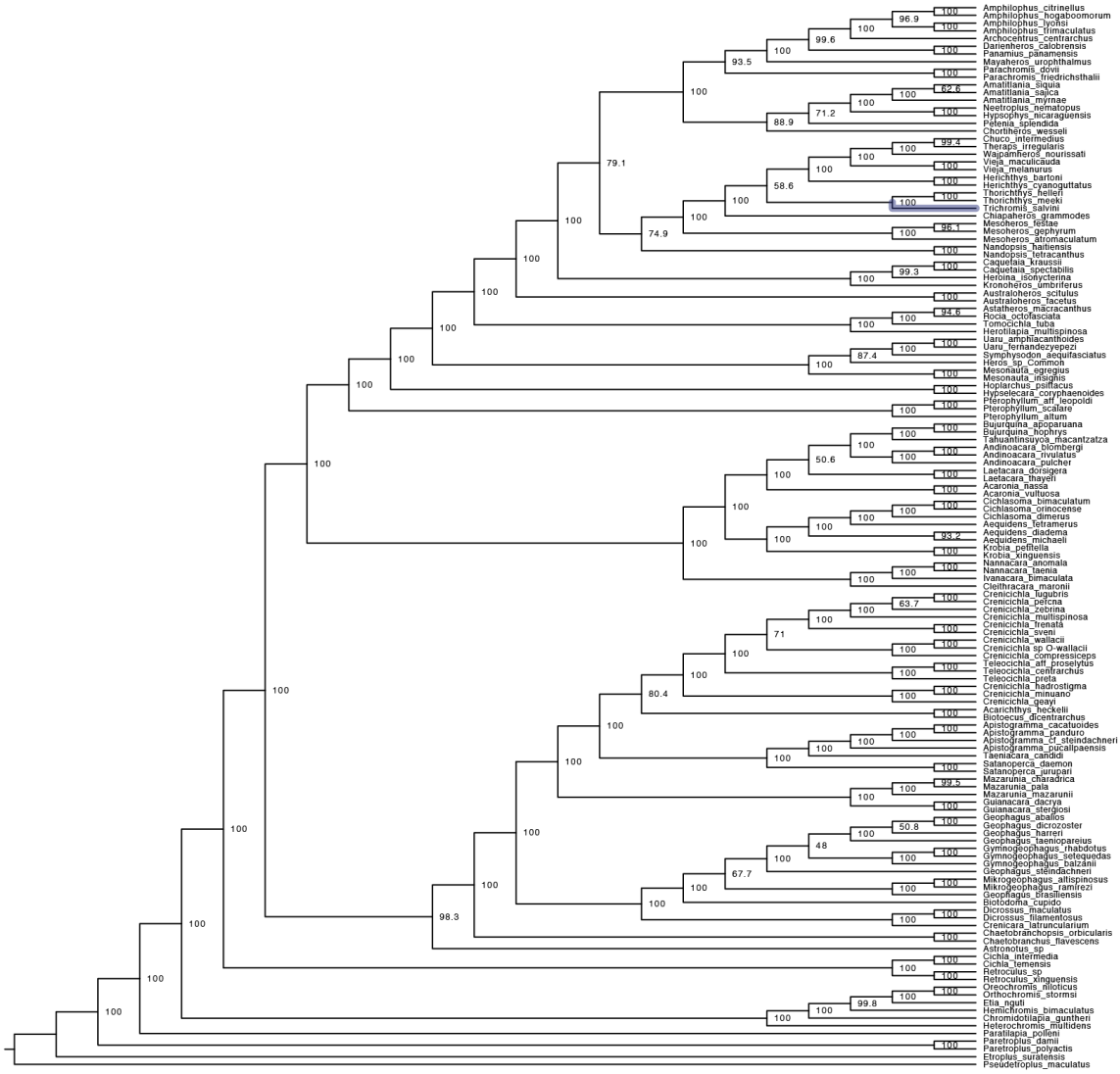

A. Centra American Heroini
Astral-II
73.2
65.2
98.2
44.6
Amphilophines
Nandopsis
Caquetaines
Chuco-Herichthys clade
Thorichthys-Trichromis clade
Chiapaheros_grammodes
Mesoheros
Australoheros
Astatheros-Herotilapia clade
Herichthyines
STAR
79.1
58.6
74.9
Amphilophines
Chuco-Herichthys clade
Thorichthys-Trichromis clade
Chiapaheros_grammodes
Mesoheros
Nandopsis
Caquetaines
Australoheros
Astatheros-Herotilapia clade
Herichthyines
Concatenated
86
97
92
Amphilophines
Chuco-Herichthys clade
Chiapaheros_grammodes
Thorichthys-Trichromis clade
Mesoheros
Nandopsis
Caquetaines
Australoheros
Astatheros-Herotilapia clade
Herichthyines
B. Geopgagini: Geophagus-Gymnogeophagus clade
Astral-II
40.4
58.4
Gymnogeophagus
Geophagus_steindachneri
Geophagus
Mikrogeophagines
Biotodoma_cupido
Crenicaratines
STAR
48
67.7
Geophagus
Gymnogeophagus
Geophagus_steindachneri
Mikrogeophagines
Biotodoma_cupido
Crenicaratines
Concatenated
94
99
Geophagus
Gymnogeophagus
Geophagus_steindachneri
Biotodoma_cupido
Mikrogeophagines
Crenicaratines
C. Cichlasomatini: Laetacara-Acaronia
Astral-II
30.8
Andinoacarines
Acaronia
Laetacara
STAR
50.6
Andinoacarines
Laetacara
Acaronia
Concatenated
58
Andinoacarines
Acaronia
Laetacara

791

Figure 2. Conflicting results among analyses. Each panel depicts the alternative topological arrangements found in the clades highlighted with asterisks in Fig. 1. Colors are meant to represent lineages within each panel and not to be used as comparison among clades in different panels. Numbers by nodes represent bootstrap support in each case; nodes without numbers received 100% bootstrap support in their respective analyses. See Fig. 1 for the ASTRAL-II species tree and Appendices L and M to see the concatenated and STAR topologies.



Appendix L – Maximum Likelihood concatenated tree produced from analysis of 415 loci with no partitioning. Numbers in each node denote bootstrap support based on 1000 replicates in RAxML. See text and Figures 1 and 2 for details.



Appendix M. Summary coalescent tree generated with the program STAR. Numbers in each node denote bootstrap support. See text and Figures 1 and 2 for details.