Multi-locus phylogeny reveals instances of mitochondrial introgression and unrecognized diversity in Kenyan barbs (Cyprininae: Smiliogastrini).

Ray C. Schmidt^{ab*}

Henry L. Bart Jr.^a

Wanja Dorothy Nyingi^c

^aDepartment of Ecology and Evolutionary Biology, Tulane University, 400 Boggs Hall, New Orleans, LA 70118, United States

^bPresent Affiliation: Smithsonian Mpala Postdoctoral Fellow, Mpala Research Centre, PO Box 555-10400- Nanyuki, Kenya

^cIchthyology Section, National Museums of Kenya, PO Box 40658-00100- Nairobi, Kenya

*Corresponding author: rschmidt@tulane.edu, Office: +254 020 259-7401

Abstract

The phylogenetics and taxonomic status of small African barbs (Cyprininae: Smiliogastrini) remains unresolved despite the recent decision to elevate the genus name *Enteromius* for the group. The main barrier to understanding the origin of African small barbs and evolutionary relationships within the group is the poor resolution of phylogenies published to date. These phylogenies usually rely on mitochondrial markers and have limited taxon sampling. Here we investigate the phylogenetic relationships of small barbs of Kenya utilizing cytochrome *b*, Growth Hormone intron 2, and RAG1 markers from multiple populations of many species in the region. This multi-locus study produced well supported phylogenies and revealed additional issues that complicate understanding the relationships among East African barbs. We observed widespread mtDNA introgression within the Kenyan barbs, highlighting the need to include nuclear markers in phylogenetic studies of the group. The GH intron 2 resolved heterospecific individuals and aided in inferring the species level phylogeny. The study reveals unrecognized diversity within the group, including within species reported to occur throughout East Africa, and it provides the groundwork for future taxonomic work in the region and across Africa.

Keywords: East Africa; '*Barbus*'; Growth Hormone; Hybridization

1. Introduction

The freshwater fauna in East Africa currently faces myriad threats. Climate change, human population expansion and development threaten freshwater ecosystems in the region, as well as, the humans that rely on these systems for survival (Thieme et al. 2010). Water shortages affect drinking water supplies, crop production, grazing, fisheries, and the needs of freshwater organisms generally. Surveys of freshwater systems across the African continent are allowing

scientists to re-evaluate the biodiversity present and re-assess conservation priorities (Daniels et al., 2015; Dijkstra et al., 2016; Morris et al. 2016; Schmidt et al. 2016). The present study summarizes the findings of field studies we conducted across southern Kenya since 2010 and attempts to infer phylogenetic relationships of small, presumably diploid, African barbs from Kenya.

Small barbs are ubiquitous across sub-Saharan Africa, with an estimated 200–500 species (Lévêque and Daget, 1984). However, the group has long been in a state of taxonomic confusion. Berrebi et al. (1996) recommended that small barbs in sub-Saharan Africa be referred to as ‘*Barbus*’, to distinguish these taxa from the tetraploid *Barbus* s. str. found in Europe and North Africa. Recently, members of this group were placed in the tribe Smiliogastrini, and the genus *Enteromius* was elevated to accommodate all species that were previously recognized as ‘*Barbus*’ (Yang et al., 2015). Yang et al.’s (2015) decision to elevate *Enteromius* was criticized because the phylogenies were poorly resolved and included limited taxon sampling (Schmidt and Bart, 2015). Moreover, grouping all sub-Saharan African barbs in *Enteromius* subsumed other morphologically-diagnosed, presumed monophyletic genera (e.g., *Barboides* and *Clypeobarbus*) within the group (Schmidt and Bart, 2015; Stiassny et al., 2016). While *Enteromius* has been used by some (Armbruster et al., 2015; Decru et al., 2016; Eschmeyer et al., 2016; Skelton, 2016), others have argued against the change based on Yang et al.’s (2015) analysis (Schmidt and Bart, 2015; Ren and Mayden, 2016; Stiassny and Sakharova, 2016; Stiassny et al., 2016). Stiassny et al. (2016) summarized this position in their recent description of new ‘*Barbus*’ species, commenting that, if taxonomy is to reflect phylogeny, it is logically inconsistent to recognize the non-monophyletic *Enteromius* as the valid genus for all African smiliogastrins currently within ‘*Barbus*’. Though it is evident that the African barbs are not closely related to

Barbus s. str; available evidence (Yang et al., 2015; Ren and Mayden, 2016) does not support recognizing *Enteromius* as the genus for all African smiliogastrin barbs. It is for these reasons that we continue to use '*Barbus*' for small African smiliogastrin barbs.

Many small barb species were collected during several expeditions across southern Kenya. Some of the populations sampled were not readily assigned to known species in Kenya (Seegers et al., 2003), and others that were Kenyan populations of wide-ranging species (e.g., '*B.*' *paludinosus* and '*B.*' *kerstenii*) exhibited morphological differences among river drainages. For some of the latter populations, specimens were abnormally large and displayed morphological characters intermediate between those of different described species, suggesting hybridization. By examining phylogenetic relationships of '*Barbus*' species collected across Kenya, and comparative material from neighboring countries, we aim to better understand the diversity and relationships of small barbs in the East African region and call attention to instances of inter-specific hybridization within the group. We accomplish the latter by using a combination of nuclear and mitochondrial markers. We show that mitochondrial markers, such as cytochrome *b* (cyt *b*) and cytochrome *c* oxidase subunit 1 (CO1), which are commonly used in phylogenetic studies, are problematic because they are prone to introgression within the group. Including samples from geographically widespread populations allows us to assess whether the current taxonomy accurately reflects the diversity within this important group.

2 Materials and Methods

2.1 Taxon sampling and protocols

'*Barbus*' specimens were collected across Kenya during three International Research Experience for Students (IRES) expeditions funded by the National Science Foundation (NSF) in

2010, 2011, and 2012 (Fig. 1). Specimens were collected with seine and electrofisher. Muscle tissue and fin clips were preserved in 95% ethanol, and voucher specimens were fixed in 10% formalin and stored in 75% ethanol. Specimens were identified using guides and studies from the region (Greenwood, 1962; Seegers, 1996; Seegers et al., 2003). Species descriptions and type material were also examined to properly identify the species. All specimens are archived in the National Museums of Kenya (NMK) Ichthyology Collection in Nairobi. Tissues samples are shared between the NMK and the Tulane University Biodiversity Research Institute (TUBRI). In addition to the specimens collected in Kenya, comparative material from Cornell University Museum of Vertebrates (CUMV), Kansas University (KU), and the Florida Museum of Natural History (FLMNH) were included in the analysis. All specimens included in the molecular analyses are listed in Supplemental Table 1.

Total DNA was extracted from preserved muscle and/or fin tissue using an Invitrogen PureLink Genomic DNA mini prep kit. The mitochondrial cytochrome *b* (cyt *b*) gene, intron 2 of the nuclear Growth Hormone (GH), and the nuclear RAG1 (recombination activating gene 1, exon 3), gene were used to infer phylogenetic relationships of the East African barbs. Cytochrome *b* was amplified with primers Glu-2 and Pro-R1 (Hardman and Page, 2003) with a touchdown protocol (Schmidt et al. 2014). RAG1 was amplified with 2533F and 4090R following published protocols (Lopez et al., 2004; Yang and Mayden, 2010). GH intron 2 was amplified using primers developed from published and unpublished GH sequences from African barbs (Mayden et al., 2009). These primers were GH23F (5'-TGTCGGTGGTGCTGGTCAGT-3'), GH102F (5'-TCGTGTACAACACCTGCACCAGC-3'), and GH148R (5'-TCCTTTCCGGTGGGTGCCTCA-3'). The primer pair GH102F/GH148R works for most of the East African barbs and the primer pair GH23F/GH148R amplified all African barbs, including

West African species. The protocol used to amplify GH intron 2 was (94°C for 60 secs – initial denaturation; 30 cycles of 94°C for 30 sec, 56°C for 30 sec, 72°C for 40 secs; final extension at 72°C for 5 minutes).

Amplified products were visually checked on an Ethidium Bromide stained gel. Fragments were purified with ExoSAP-IT (Amersham Biosciences) and sequenced with the amplification primers using the BigDye Terminator Reaction Mix from Applied Biosystems. Sequencing reactions were visualized on an ABI 3730XL DNA analyzer and contigs were aligned and visually edited in Sequeuncher 4.6 (Gene Codes Corporation). All products were submitted to GenBank and accession numbers from each product are listed in Supplementary Table 1. Specimens with heterospecific *cyt b* haplotypes were re-extracted and re-sequenced to ensure that human error or contamination did not contribute to falsely recognizing these specimens as putative hybrids.

2.2 Sequence alignment and phylogenetic analyses

The partial *cyt b* and RAG1 sequences were aligned with ClustalW v 1.4 (Thompson et al., 1994) and edited in BioEdit v. 7.0.9 (Hall, 1997). We tested for saturation in *cyt b* using Xia's test (Xia et al., 2003, Xia and Lemey, 2009) implemented in DAMBE 5 (Xia, 2013), with protein-coding genes separated by 1 + 2 and 3rd codon positions. We assessed stationarity for each codon position with the χ^2 test (Foster, 2004) implemented in PAUP* 4.0a151 (Swofford, 2002). GH Intron 2 sequences were aligned with Prankster (Löytynoja and Goldman, 2008). The indels (ambiguous areas within alignment) were coded in FastGap version 1.2 (Borchsenius, 2009) using the simple coding method (Simmons and Ochoterena, 2000). The characters derived from coding the indels were then concatenated to the end of the intron 2 alignment. The addition of these gap characters is useful in resolving species level phylogenies (Simmons et al., 2001;

Schmidt et al., 2014). From these alignments four analyses were completed. The first alignment comprised only the *cyt b* data; the second was GH intron 2 and the characters derived from gap coding. The third alignment was the RAG1 data, and the fourth was a concatenated dataset of *cyt b*, GH intron 2, characters from gap coding, and RAG1. Specimens with heterospecific *cyt b* haplotypes, as determined in this study, were coded as missing data for the *cyt b* gene in the concatenated alignment (Near et al., 2011). We selected several Asian *smiliogastrin* sequences from GenBank and unpublished data to include as outgroup taxa based on Yang et al. (2015).

PartitionFinder2 (Lanfear et al., 2016) was used to determine the best model of evolution for each codon position in the *cyt b* and RAG1 datasets. JModeltest 2.1.9 (Darriba et al., 2012) was used to determine the optimal model of evolution for GH intron 2. Maximum likelihood (ML) and Bayesian inferences analyses were performed to infer the phylogeny. GARLI 2.0 (Zwickl, 2006) via the CIPRES Science Gateway (Miller et al., 2010) was used to perform the ML analyses. Nodal support was inferred by bootstrap proportions after 1000 bootstrap replicates with 3 search reps each. Bootstrap values were obtained by producing majority rule consensus trees in Mesquite (Maddison and Maddison, 2011). Bayesian analysis was performed in MrBayes version 3.2 (Rohnquist et al., 2012) via the CIPRES Science Gateway (Miller et al., 2010). Posterior probabilities were assessed with 5 million generations, sampling trees every 100 generations. The first 25% of trees were discarded as burn-in. All trees were visualized in FigTree v 1.4.2 (Rambaut, 2007).

3 Results

3.1 Cytochrome b

Our analysis is based on 1098 base pair (bp) of *cyt b* data 146 individuals representing nominal small barb species from across Kenya, and representatives from Tanzania, Ethiopia, and South Africa. The Xia's test did not detect significant amounts of saturation within the 3rd codon position, assuming a symmetrical phylogeny; and the χ^2 test did not detect non-stationarity within the *cyt b* matrix. Uncorrected pairwise distances between ingroup and outgroup taxa ranged from 12.1% to 18.4% and from 8.4% to 20.9% among ingroup taxa (Supplemental Table 3). The models of evolution for each codon position of the *cyt b* matrix selected by AIC and implemented in the ML and Bayesian analyses are shown in Supplemental Table 2. The *cyt b* analyses revealed significant genetic diversity within several clades from Kenya and raise questions about sequences from specimens identified as '*B.*' cf. *paludinosus* and '*Barbus*' spp. from Tanzania (Fig. 2). '*Barbus*' *neumayeri* is sister to all other barbs in the analyses, although this relationship is not well-supported. The relationships among the different groups of barbs are likewise unresolved, and there is moderate support that our '*B.*' *apleurogramma* clade is sister to the '*B.*' *paludinosus* clade (Fig. 2). High levels of divergence are observed within '*B.*' *kerstenii*, '*B.*' *paludinosus*, and '*B.*' *apleurogramma* (Fig. 2, Supplemental Table 3). The likely undescribed '*Barbus*' sp. "Jipe" from Lake Jipe in southern Kenya is sister to the clade containing '*B.*' *cercops*, '*B.*' *jacksoni*, and '*B.*' *trispilopleura* (Fig. 2).

Our analysis revealed numerous cases in which the placement of the sequence within the *cyt b* phylogeny disagrees with the species' identification based on external morphology. These cases are found within the '*B.*' *nyanzae* and '*B.*' *paludinosus* clades, and extensively within the '*B.*' *neumayeri* clade, which contains numerous sequences from specimens that were identified morphologically as '*B.*' *kerstenii* or '*B.*' *cercops* (Fig. 2). Interestingly, we were unable to sequence any unique '*B.*' *cercops* *cyt b* haplotypes in our study. All specimens that we identified

as '*B.*' *cercops* that we were able to successfully sequence had '*B.*' *neumayeri* or '*B.*' *paludinosus* like *cyt b* haplotypes.

3.2 Growth Hormone Intron 2, RAG1, and total evidence.

Growth Hormone intron 2 matrix from 94 specimens included a PRANK alignment of 332 bp and 38 characters derived from gap coding. Uncorrected pairwise distance among ingroup taxa ranged from 0.8% to 11.7 % (Supplemental Table 3). The model of evolution selected for the GH intron 2 and RAG1 by AIC and implemented in the ML and Bayesian analyses are shown in Supplemental Table 2. The characters derived from gap coding the alignment were treated as standard characters. The analysis resulted in a phylogeny largely similar to the *cyt b* analysis, but the placement of '*B.*' *neumayeri* is different. Support for some of the deeper nodes are higher but the branching patterns are still equivocal. The phylogeny inferred from the GH intron 2 matrix resolved the placement of specimens with heterospecific *cyt b* haplotypes (Fig. 2), including a distinct clade of '*B.*' *cercops*, which is sister to '*B.*' *jacksoni* + '*B.*' *trispiloptera*. The analysis from the 1443 bp alignment of RAG1 from 39 specimens revealed a phylogeny similar to the GH intron 2, although it also did not resolve the relationship among the barb lineages (Supplemental Fig. 1). '*Barbus*' *apleurogramma* is not sister to the '*B.*' *paludinosus* complex in the RAG1 phylogeny. The uncorrected pairwise distance among ingroup taxa ranged from 0.7% to 3.2% (Supplemental Table 3).

The total evidence analysis included 2901 characters and resulted in a well resolved phylogeny (Fig. 3). The phylogeny resulting from the ML analysis is well supported, but lacks support at some nodes in the phylogeny from the Bayesian analysis. Both of the combined analyses revealed a group comprising '*B.*' *radiatus*, '*B.*' *nyanzae*, and '*B.*' *kerstenii* that is sister to all other ingroup taxa. The phylogeny revealed deep divergences within the '*B.*' *kerstenii* and

‘*B.*’ *paludinosus* complexes, suggesting unrecognized diversity within both groups. This analysis also suggests that certain taxa do not belong to the morphospecies as they were identified as. ‘*Barbus*’ cf. *paludinosus* from Ethiopia are distinct from the other ‘*B.*’ *paludinosus*. The placement of ‘*B.*’ *apleurogramma* is poorly resolved, but is not sister to the ‘*B.*’ *paludinosus* group (Fig 3).

4 Discussion

4.1 Introgression within East African barbs

Our analysis provides novel evidence for widespread introgression within the barbs sampled. While introgression is well known within North American cyprinids (Dowling and DeMarais, 1993; Broughton and Gold, 2000; Moyer et al., 2009), this situation has not been observed within the African barbs. It is interesting to note that introgression is occurring at various levels within the phylogeny. One specimen identified as ‘*B.*’ *kerstenii* is introgressed with mtDNA from the closely related ‘*B.*’ *nyanzae* (Fig. 1). Many specimens identified as ‘*B.*’ *kerstenii* and ‘*B.*’ *cercops*, from multiple populations are introgressed with mtDNA from the more distantly related ‘*B.*’ *neumayeri*. One specimen of ‘*B.*’ *cercops* is introgressed with mtDNA from a distinctive form of ‘*B.*’ cf. *paludinosus* collected at the same site in Lake Victoria. The putative ‘*B.*’ *kerstenii* x ‘*B.*’ *neumayeri* hybrids had intermediate counts and achieved much larger body sizes (> 90 mm SL, Bart, unpublished) than either parent species. The ranges of the parent species overlap in Kenya (Fig. 1). Morphological analyses of these populations are ongoing and will quantify how prevalent these heterospecific individuals are within the populations.

The occurrence of introgression within the East African barbs highlights the need for including nuclear markers in phylogenetic analyses of African cyprinids. The inclusion of GH

intron 2 effectively resolved the placement of the heterospecific individuals in the present study and has proven to be effective for resolving species level relationships in studies of other groups of fishes (Moyer et al., 2009; Clements et al., 2012; Schmidt et al., 2014; 2016). The primers developed for this study were designed to amplify all African barbs, and work equally well for barbs in East and West Africa. Future genetic studies of African barbs should include this marker.

4.2 Intraspecific variation and undescribed species.

Our study revealed unrecognized diversity within the East African barbs. ‘*Barbus*’ sp. “Jipe” is likely an undescribed barb species found in Lake Jipe on the Kenya-Tanzania border. Additionally, we were unable to assign morphospecies to rather divergent cytb, GH intron, and RAG1 sequences from barb specimens from Tanzania. Specimens from two different lots initially identified as ‘*B.*’ *apleurogramma*, these sequences (identified in our trees as ‘*Barbus*’ spp. Tanzania) are clearly distinct from our ‘*B.*’ *apleurogramma* clade, which includes topotypic ‘*B.*’ *apleurogramma*. The voucher specimens these sequences were derived from and other material from the area need to be more closely examined before we can determine whether or not they represent undescribed species.

The analyses also revealed diversity within specimens identified as ‘*B.*’ *kerstenii*. Three distinct groups, corresponding to the Tsavo River (lower Athi River); the upper and middle Athi, Tana, Pangani and Umba Rivers; and Lake Victoria and the Southern Ewaso Ng’iro, are evident in our phylogeny (Fig. 3). Described from the Pangani River and currently reported to occur throughout eastern and southern Africa (Fig. 1), ‘*B.*’ *kerstenii* is likely restricted to the eastern flowing rivers in Kenya and Tanzania (Seegers, 1996). A morphological study and taxonomic

revision of populations of '*B.*' *kerstenii* from Kenya and Tanzania is underway. A study of all populations presently considered '*B.*' *kerstenii* from central Tanzania, southward is needed.

The widespread straightfin barb ('*B.*' *paludinosus*), described from Mozambique, is reported to occur from South Africa to Ethiopia. Our study questions this assumption. Our phylogenies reveal three genetically distinct groups of populations of '*B.*' *paludinosus* within Kenya. Curiously, the most distinct group comprises specimens from broadly disjunct populations in southern Kenya (Lake Jipe and Lake Victoria). This group is sister to two groups of '*B.*' cf. *paludinosus* specimens, one comprising specimens from Lake Victoria and the other distributed across a broad area of Kenya and Tanzania. All of these populations form a group that is sister to '*B.*' *paludinosus* from South Africa (Fig. 3). Previous studies have documented morphological variation within the straightfin barb, but suggested that the variation is clinal and pigmentation patterns are plastic (Greenwood 1962; Seegers, 1996; Farm, 2001); though Greenwood (1962) suggested that the Ethiopian populations may be distinct. Morphological studies of '*B.*' cf. *paludinosus* specimens from all of the areas represented in our phylogenies are ongoing and should resolve the taxonomic status of these groups.

Specimens from Ethiopia initially identified as '*B.*' *paludinosus* are indeed distinct from the Kenyan and South African '*B.*' *paludinosus* populations. Cytochrome *b*, GH intron 2, and RAG1 sequences from these Ethiopian specimens are very divergent from the other straightfin barb sequences and, based on divergences within this cluster, likely represent several distinct species. The small barbs with serrated dorsal spines from Ethiopia have been variously referred to as '*B.*' *paludinosus*, '*B.*' *pleurogramma*, '*B.*' *akakianus*, '*B.*' *amphigramma*, and '*B.*' *kerstenii* (Greenwood, 1962; Golubstov and Berendzen, 2004; de Graaf et al., 2007). '*Barbus*' *pleurogramma* was described from Lake Tana and is a valid species. '*Barbus*' *akakianus* was

described from the Awash River basin in central Ethiopia and is regarded as a junior synonym of '*B.*' *paludinosus* (Lévêque and Daget 1984). '*Barbus*' *amphigramma*, described from the Athi River basin in Kenya, is also considered a junior synonym of '*B.*' *paludinosus* (Greenwood, 1962). A detailed morphological study of serrated barbs from Ethiopia did not support recognizing these taxa as distinct from '*B.*' *paludinosus* (Golubstov and Berendzen, 2005). However, our results clearly shows that Ethiopian specimens identified as '*B.*' cf. *paludinosus* are genetically distinct from populations of '*B.*' *paludinosus* from Kenya and South Africa. Voucher specimens and additional populations from the area need further examination to resolve the taxonomic status of small barbs with serrated dorsal spines from Ethiopia.

With the unrecognized diversity within '*B.*' *kerstenii*, '*B.*' cf. *paludinosus*, and the undescribed species from the area, our study finds that the current taxonomy of East African small barbs underestimates the diversity within the group. This situation is also observed in West (Bamba et al., 2011; Bamba. 2012; Schmidt, 2014; Pezold et al., 2016) and Central Africa (Decru et al., 2016; Stiassny et al., 2016) suggesting that the number of African barb species is likely in the upper range of the estimated 200–500 species (Leveque and Daget 1984). This study and recent work in the suckermouth catfishes (Schmidt et al., 2015) suggests that the number of endemic species in the region is underestimated by current taxonomy; underscoring the need to re-evaluate other freshwater taxa in the region (Thieme et al., 2005).

4.3 Interspecific relationships and taxon sampling.

The phylogenies produced in this study provide insights into the relationships among the Kenyan barb species. The relationships among these species and other African barbs were inferred and artificially grouped by morphological traits (Lévêque, 2003; De Weirdt et al., 2007). The utility of these traits as phylogenetic characters (e.g., presence of serrated dorsal spine, cephalic pit

lines, completeness of lateral line, spots versus longitudinal stripes, and number of barbels) is still unknown (Roberts, 2010; Armbruster et al., 2015). Our study did group spotted barbs with un-serrated dorsal spines (*B.* *cercops*, *B.* *trispilopleura*, and *B.* *jacksoni*) in a clade, although the large clade that includes these species also include species with serrated dorsal spines (e.g., *B.* *apleurogramma* and *B.* spp. Tanzania). These characters may be informative regionally, as is seen in the barbs from the Upper Guinean Forests (Schmidt, 2014), but more complete continental phylogenies are needed to fully understand the utility of these characters.

The relationships revealed in this study differ from recently published multi-locus phylogenies of African barbs (Ren and Mayden, 2016). The phylogeny produced by Ren and Mayden (2016) did not show the sister relationship of *B.* *nyanzae* and *B.* *kerstenii* that our multi-locus phylogeny revealed. This relationship hypothesized by Greenwood (1962) is found in each of our phylogenies. The *cyt b* sequence data included in their analyses is most similar to *B.* *yongei* in our analysis and is likely misidentified. Their analysis also did not show *B.* *radiatus* as sister to *B.* *nyanzae* and *B.* *kerstenii* (Ren and Mayden, 2016). These disparities could be due to misidentification of previously published sequences, introgression, or taxon sampling. Fishes of the East Coast ichthyo-province (Roberts, 1975) have affinities with Congolese, Nilo-Sudanic, and Zambesian fishes. As sampling within the region and across the continent improves, the relationship among these different groups of barbs should become better resolved. The placement of *B.* *apleurogramma* is not well resolved in our analysis. The placement of this taxon and *B.* *radiatus* are important moving forward as both taxa have subgeneric names (*Agrammobarbus* and *Beirabarbus* respectively) that could be applied to monophyletic clades within the larger African '*Barbus*' phylogeny.

5 Conclusions

317 The results of this study suggests that mitochondrial introgression, undescribed taxa, and
 318 misidentified specimens all confuse interpretations of phylogenetic relationships of Kenyan
 319 small barbs, and may be contributing to confusion within other phylogenetic studies of the
 320 African barbs. The study revealed widespread mitochondrial introgression in several Kenyan
 321 barb species. Including data from GH intron 2 resolved the placement of heterospecific taxa and
 322 proved informative for inferring species level relationships. The geographically wide-ranging
 323 ‘*B.*’ *kerstenii* and ‘*B.*’ *paludinosus* have several genetically distinct lineages in Kenya, and the
 324 current taxonomy of these groups and others underrepresents the diversity within the region.
 325 Increased sampling of these and other barbs from surrounding countries, including sequencing of
 326 nuclear genetic markers, will improve phylogenetic resolution, and support stable generic
 327 assignment within this problematic group.

328 **Acknowledgements**

329 Funding was provided by NSF OISE grants 0968727; 1215395 and 1355361 to HLB and WDN.
 330 Partial funding for this project was provided by the Smithsonian Mpala Research Fellowship to
 331 RCS. The following participants of the IRES program whose dedication and hard work allowed
 332 the project to succeed. Collecting team members denoted by superscript (2010^A, 2011^B, and
 333 2012^C) Principal investigators: D. Sigana^{ABC} (University of Nairobi), M. Ogada^{ABC}
 334 (Conservation Solutions Afrika), and G. Talarchek (Tulane University). IRES student
 335 participants: R. Anderson^C, D. Lach^B, L. Mathews^B, N. Moses^B, H. Strobel^C, and T. Woods^A
 336 (Tulane University); S. Abade^A, J. Abong’o^B, L. Asande^A, K. Laban Losili^A, C. Mburu^C, P.
 337 N’gang’a^C, M. Odhiambo^B, R. Onwong’a^A, W. Owako^C, M. Oyier^C, P. Tanui^B (University of
 338 Nairobi); A. Ewing^A, M. Patterson^C, and L. Spivey^A (Xavier University of New Orleans); and J.
 339 Gathua^{ABC}, G. Kosgei^{BC}, T. Ndiwa^{AB} (Kenya Wetlands Biodiversity Research Group). Special

thanks to D. Lach and the IRES students for their assistance with the molecular work. Comparative material was provided by A. Bentley (KU), J.P. Friel (CUMV; present affiliation Alabama Museum of Natural history), and R. Robins (FLMNH) who also provided meristics from voucher specimens. In addition, we thank the Kenyan Wildlife Service (KWS) particularly J. Nyunja (Wetlands program) and the Warden and staff at Meru National Park for allowing us to collect within the park boundaries. We thank the Mpala Research Centre and its staff for their hospitality. Peter Nyamenya (Kisumu Museum) and Dr. W. Ojwang (KEMFRI) were of great assistance in Nyanza County. Our drivers A. Mukiri and L. Nzangi Kioko (NMK), A. Mureithi and I. Mogoi (UoN) provided wonderful service. The staff of the NMK Ichthyology and Herpetology sections provided invaluable support while in the field and in the laboratory. M.H. Doosey assisted with acquiring unpublished Growth Hormone data for the cyprinids included as outgroup taxa.

References

- Armbruster, J.W., Stout, C.C., Hayes, M.H., 2015. An empirical test for convergence using African barbs (Cypriniformes: Cyprinidae). *Evol. Ecol.* 2015, 1–16.
- Bamba, M., Vreven, E.J., Snoeks, J., 2011. Description of *Barbus teugelsi* sp. Nov. (Cypriniformes: Cyprinidae) from the Little Scarcies basin in Guinea, Africa. *Zootaxa*. 2998, 48–65.
- Bamba, M., 2012. The ‘Barbus’ Species (Cypriniformes, Cyprinidae) from Côte d’Ivoire: A Systematic Revision of some West African Species Complexes. Ph.D. Dissertation, Katholieke Universiteit Leuven.

361 Berrebi, P., Kottelat, M., Skelton, P., Ráb, P., 1996. Systematics of *Barbus*: state of the art and
 362 heuristic comments. *Folia Zool.* 45 (Suppl. 1), 5–12.

363 Borchsenius, F., 2009. FastGap 1.2. Distributed by author Aarhus, Denmark: Department of
 364 Biological Sciences, University of Aarhus. http://192.38.46.42/aubot/fb/FastGap_home.htm
 365 (Downloaded May 2012).

366 Broughton, R.E., Gold, J.R., 2000. Phylogenetic relationships in the North American cyprinid
 367 genus *Cyprinella* (Actinopterygii: Cyprinidae) based on sequences of the Mitochondrial ND2
 368 and ND4L genes. *Copeia*. 2000 (1), 1–10.

369 Clements, M.D., Bart Jr., H.L., Hurley, D.L., 2012. A different perspective on the phylogenetic
 370 relationships of the Moxostomatini (Cypriniformes: Catostomidae) based on cytochrome-b and
 371 the Growth Hormone intron sequences. *Mol. Phylogenet. Evol.* 63, 159–167.

372 Daniels, S.R., Phiri, E.E., Klaus, S., Albrecht, C., Cumberlidge, N., 2015. Multi-locus phylogeny
 373 of the Afrotropical freshwater crab fauna reveals historic drainage connectivity and transoceanic
 374 dispersal since the Eocene. *Syst. Biol.* 64 (4), 549–567.

375 Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new
 376 heuristics and parallel computing. *Nat. Methods* 9 (8), 772.

377 de Graaf, M., Megens, H., Samallo, J., Sibbing, F.A., 2007. Evolutionary origin of the Lake
 378 Tana's (Ethiopia) small *Barbus* species: indications of rapid ecological divergence and
 379 speciation. *Anim. Biol.* 57 1, 39–48.

380 De Weirtdt, D., Getahun, A., Tshibwabwa, S., Teugels, G.G., 2007. Cyprinidae (p. 467–572). In:
 381 The fresh and brackish water fishes of Lower Guinea, West-Central Africa. Stiassny, M.L.J.,

382 Teugels, G.G., Hopkins, C.D., (eds). Publications scientifiques du Museum, MRAC. Vol. 1: 800
 383 pp.

384 Decru, E., Meolants, T., Degelas, K., Vreven, E., Verheyens, E., Snoeks, J., 2016. Taxonomic
 385 challenges in freshwater fishes: a mismatch between morphology and DNA barcoding in fish of
 386 the north-eastern part of the Congo basin. Mol. Ecol. 16 (1), 342–352.

387 Dijkstra, K.D.B., Kipping, J., Meziere, N., 2015. Sixty new dragonfly and damselfly species
 388 from Africa (Odonata). Odonatologica. 44 (4), 447–678.

389 Dowling, T.E., DeMarais, B.D., 1993. Evolutionary significance of introgressive hybridization in
 390 cyprinid fishes. Nature. 362, 444–446.

391 Eschmeyer, W.N., Fricke, R., van der Lann, R., (eds). CATALOG OF FISHES: GENERA,
 392 SPECIES, REFERENCES.
 393 (<http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>). Electronic
 394 version accessed 01 Nov 2016. [This version was edited by Bill Eschmeyer.]

395 Farm, B.P., 2001. Experimentally induced pigment changes in small African ‘*Barbus*’ (Teleostei:
 396 Cyprinidae): Synonymy of ‘*Barbus*’ *amphigramma* and ‘*Barbus*’ *taitensis* with ‘*Barbus*’
 397 *paludinosus*. Copeia. 2001 (1), 243–247.

398 Foster P. 2004. Modeling compositional heterogeneity. Syst. Biol. 53:485–95.

399 Golubstov, A.S., Berendzen, P.B., 2005. A review of the small barbs (*Barbus*, Cyprinidae) with a
 400 serrated dorsal spine from Ethiopia, with a discussion of the origin of forms with a reduced
 401 dorsal spine from the White Nile and Omo River tributaries. J. Ichthyol. 45 (1), S58–S88.

402 Greenwood, P.H., 1962. A revision of certain *Barbus* (Pisces, Cyprinidae) from east, central, and
 403 south Africa. Bull. br. Mus. nat. Hist. Zool. 8 (4), 153–208.

404 Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
 405 programs for Windows 95/98/NT. Nucleic Acids Res. 41, 95–98.

406 Hardman, M., Page, L.M., 2003. Phylogenetic relationships among bullhead catfishes of the
 407 Genus *Ameiurus* (Siluriformes: Ictaluridae). Copeia 2003 (1), 20–33.

408 Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., Calcott, B., 2016 PartitionFinder 2: new
 409 methods for selecting partitioned models of evolution for molecular and morphological
 410 phylogenetic analyses. Molecular biology and evolution. DOI:
 411 dx.doi.org/10.1093/molbev/msw260

412 Lévêque, C., Daget, J., 1984. Cyprinidae. In: In: Daget, J., Gosse, J.P., Thys van den
 413 Audenaerde, D.F.E., (Eds.), Check-list of the Freshwater Fishes of Africa. vol. 1. CLOFFA.
 414 ISNB Bruxelles, MRAC, Tervuren. ORSTOM, Paris. pp. 217–342.

415 Lévêque, C., 2003. Cyprinidae. In: Faune des poissons d’eaux douces et saumâtres de l’Afrique
 416 de l’Ouest, D. Paugy, C. Lévêque & G. G. Teugels (eds.), Collection faune tropicale 40. Vol. 1,
 417 IRD, Paris; MNHN, Paris ; MRAC, Tervuren, pp. 322–436.

418 Lopez, J.A., Chen, W-J., Orti, G., 2004. Esociform phylogeny. Copeia. 2004 (3), 449–464.

419 Löytynoja, A., Goldman, J., 2008. Phylogeny-aware gap placement prevents errors in sequence
 420 alignment and evolutionary analysis. Science 320, 1632–1635.

421 Maddison, W.P., Maddison, D.R., 2011. Mesquite: A Modular System for Evolutionary
 422 Analysis. Version 2.75 <<http://mesquiteproject.org>>.

423 Mayden, R.L., Chen, W.-J., Bart Jr., H.L., Doosey, M.H., Simons, A.M., Tang, K.L., Wood,
 424 R.M., Agnew, M.K., Yang, Y., Hirt, M.V., Clements, M.D., Saitoh, K., Sado, T., Miya, M.,
 425 Nishida, M., 2009. Reconstructing the phylogenetic relationships of the earth's most diverse
 426 clade of freshwater fishes—order Cypriniformes (Actinopterygii: Ostariophysi): A case study
 427 using multiple nuclear loci and the mitochondrial genome. *Mol. Phylogenet. Evol.* 51, 500–514.

 428 Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for
 429 inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments*
 430 *Workshop (GCE)*, 14 November 2010, New Orleans, LA. pp. 1–8.

 431 Morris, J., Ford, A.G.P., Ali, J.R., Peart, C.R., Bills, R., Day, J.J., 2016. High levels of genetic
 432 structure and striking phenotypic variability in a sexually dimorphic suckermouth catfish from
 433 the African Highveld. *Biol. J. Linn. Soc.* 117 (3), 528–546.

 434 Moyer, G.R., Remington, R.K., Turner, T.F., 2009. Incongruent gene trees, complex
 435 evolutionary processes, and the phylogeny of a group of North American minnows
 436 (*Hybognathus* Agassiz 1855). *Mol. Phylogenet. Evol.* 50, 514–525.

 437 Near, T.J., Bossu, C.M., Bradburd, G.S., Carlson, R.L., Harrington, R.C., Hollingsworth, P.R.,
 438 Keck, B.P., Etnier, D.A., 2001. Phylogeny and Temporal Diversification of Darters (Percidae:
 439 *Etheostomatinae*). *Syst. Biol.* 60 (5), 565–595.

 440 Pezold, F., Schmidt, R.C., Stiassny, M.L.J., 2016. A Survey of Fishes of the Geebo – Dugbe
 441 River Confluence, Sinoe County, Liberia, with an Emphasis on Tributary Creeks. *AQUA.* 22 (3),
 442 97–122.

 443 Rambault, A., 2007. FigTree v1.4.2 available at <http://tree.bio.ed.ac.uk/software/figtree/>
 444 (Downloaded January 2016)

445 Ren, Q., Mayden, R.L., 2016. Molecular phylogeny and biogeography of African diploid barbs,
 446 '*Barbus*', and allies in Africa and Asia (Teleostei: Cypriniformes). Zool. Scripta. 2016, 1–8.
 447 Roberts, T.R. 1975. Geographical distribution of African freshwater fishes. Zool. J. Linnean.
 448 Soc. 57, 249–319.
 449 Roberts, T.R., 2010. *Barbus sensitivus*, a new species with extensive pitlines from the Sanaga
 450 River (Teleostei: Cyprinidae). AQUA. 16 (4), 187–196.
 451 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B.,
 452 Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic
 453 inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542.
 454 Schmidt, R.C. 2014. Historical Biogeography of Fishes of the Fouta Djallon Highlands and
 455 Surrounding Areas. Ph.D. Dissertation, Tulane University.
 456 Schmidt, R.C., Bart Jr., H.L., Nyingi, W.D., Gichuki, N.N., 2014. Phylogeny of suckermouth
 457 catfishes (Mochokidae: *Chiloglanis*) from Kenya: the utility of Growth Hormone introns in
 458 species level phylogenies. Mol. Phylogenet. Evol. 79, 415–421.
 459 Schmidt, R.C., Bart Jr., H.L., Nyingi, W.D., 2015. Two new species of African suckermouth
 460 catfishes, genus *Chiloglanis* (Siluriformes: Mochokidae), from Kenya with remarks on other taxa
 461 from the area. Zootaxa. 4044 (1), 45–64.
 462 Schmidt, R.C., Bart Jr., H.L., 2015. Nomenclatural changes should not be based on equivocally
 463 supported phylogenies: Reply to Yang et al. 2015. Mol. Phylogenet. Evol. 90, 193–194.

464 Schmidt, R.C., Bart Jr., H.L., Pezold, F., 2016. High levels of endemism in suckermouth
 465 catfishes (Mochokidae: *Chiloglanis*) from the Upper Guinean forest of West Africa. Mol.
 466 Phylogenet. Evol. 100, 199–205.

467 Seegers, L., 1996. The Fishes of Lake Rukwa. Africa Museum.

468 Seegers, L., De Vos, L., Okeyo, D.O., 2003. Annotated checklist of the freshwater fishes of
 469 Kenya. J. East Afr. Nat. Hist. 92, 11–47.

470 Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic
 471 analysis. Syst. Biol. 49 (2), 369–381.

472 Simmons, M.P., Ochoterena, H., Carr, T.G., 2001. Incorporation, relative homoplasy, and effect
 473 of gap characters in sequence-based phylogenetic analyses. Syst. Biol. 50 (3), 454–462.

474 Skelton, P.H., 2016. Name changes and additions to the southern African freshwater fish fauna.
 475 Afr. J. Aquat. Sci. 2016, 1–7.

476 Stiassny, M.L.J., Liyandja, T.L.D., Iyaba, J.C.M., 2016. A new small barb (Cyprininae:
 477 Smiliogastrini) from the N'sele and Mayi Ndombe rivers in the lower reaches of the middle
 478 Congo basin (Democratic Republic of Congo, Central Africa). Am. Mus. Novit. 3848, 1–15.

479 Stiassny, M.L.J., Sakharova, H., 2016. Review of the smiliogastrin cyprinids of the Kwilu River
 480 (Kasai Basin, central Africa), revised diagnosis for *Clypeobarbus* (Cyprinidae: Cyprininae:
 481 Smiliogastrini) and description of a new species. J. Fish. Biol. 88 (4), 1394–1412.

482 Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods).
 483 Version 4.0a151. Sinauer Associates, Sunderland, Massachusetts.

484 Thieme, M.L. Abell, R., Burgess, N., Lehner, B., Dinerstein, E., Olson, D., 2005. Freshwater
 485 ecoregions of Africa and Madagascar: a conservation assessment. Island Press.

486 Thieme, M.L., Lehner, B., Abell, R., Matthews, J., 2010. Exposure of Africa's freshwater
 487 biodiversity to a changing climate. *Conserv. Lett.* 3, 324–331.

488 Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. ClustalW: improving the sensitivity of
 489 progressive multiple sequence alignment through sequence weighting, position specific gap
 490 penalties and weight matrix choice. *Nucl. Acids Res.* 22, 4673–4680.

491 Xia, X., Zheng, X., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation
 492 and its application. *Mol. Phylogenet. Evol.* 26, 1–7.

493 Xia, X., Lemey, P., 2009. Assessing substitution saturation with DAMBE. *The phylogenetic*
 494 *handbook: a practical approach to DNA and protein phylogeny* 2. 615–630.

495 Xia, X., 2013. DAMBE5: A comprehensive software package for data analysis in molecular
 496 biology and evolution. *Mol. Biol. Evol.* 30 (7), 1720–1728.

497 Yang, L., Mayden, R.L. 2010. Phylogenetic relationships, subdivision, and biogeography of the
 498 cyprinid tribe Labeonini (sensu Rainboth, 1991) (Teleostei: Cypriniformes), with comments on
 499 the implications of lips and associated structures in the labeonin classification. *Mol. Phylogenet.*
 500 *Evol.* 54 (1), 254–265.

501 Yang, L., Sado, T., Hirt, M.V., Pasco-Viel, E., Arunachalam, M., Li, J., Wang, X., Freyhol, J.,
 502 Saitob, K., Simons, A.M., Miya, M., He, S., Mayden, R.L., 2015. Phylogeny and polyploidy:
 503 resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). *Mol. Phylogenet.*
 504 *Evol.* 85, 97–116.

Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion. Ph.D. Dissertation, The University of Texas at Austin.

Figure Captions

Fig. 1 Collection localities of barbs included in this study. Localities of borrowed specimens from Tanzania and Ethiopia are solid triangles. White lines denote drainage boundaries. Ranges of parent species of heterospecific specimens in our study are also shown.

Fig 2. Phylogeny of East African barbs inferred from separate analyses of the partial *cyt b* data and GH intron 2 plus the gap characters. Posterior probabilities from BI analysis/bootstrap (bs) values for ML (1000 replicates) above branches (values above 95/.95 not shown). Values below .70/.50 are collapsed, and dashed branches collapse (<50 bs) in ML analysis. Species of interest denoted by color, heterospecific specimens identified with color of species determined by external morphology.

Fig 3. Phylogeny of East African barbs inferred from a concatenated dataset (partial *cyt b*, GH intron 2, characters derived from gap coding intron 2, and RAG1). Posterior probabilities from BI analysis/bootstrap (bs) values for ML (1000 replicates) above branches (values above 95/.95 not shown). Values below .70/.50 are collapsed, and dashed branches collapse (<50 bs) in ML analysis. Species of interest denoted by color, heterospecific specimens identified with color of species determined by external morphology.