A peer-reviewed version of this preprint was published in PeerJ on 26 October 2018.

View the peer-reviewed version (peerj.com/articles/5762), which is the preferred citable publication unless you specifically need to cite this preprint.

https://doi.org/10.7717/peerj.5762
Evolution of African barbs from the Lake Victoria drainage system, Kenya

Violet M Ndenda 1, 2, Mariana Mateos Corresp., 1, Luis A Hurtado Corresp. 1

1 Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas, United States of America
2 Department of Zoology, Kenya, Maseno University, Maseno, Kenya

Corresponding Authors: Mariana Mateos, Luis A Hurtado
Email address: marianamateosh@gmail.com, lhurtado@tamu.edu

The Lake Victoria drainage basin (LVD) in Kenya is home to ten nominal species of small barbs (‘Barbus’) and one of large barbs (Labeobarbus altianalis). A recent molecular study genetically characterized small barbs in this region and found evidence of introgression between certain species, further complicating the taxonomy and species identification of these fishes. This study aimed to extend our understanding on the evolution of these fishes by: (1) examining the phylogenetic relationships of small barbs of the Kenyan LVD with those reported from other ichthyological provinces of Africa; (2) testing the sister relationship between ‘Barbus’ profundus, endemic to Lake Victoria, and ‘Barbus’ radiatus, also found in Lake Victoria, which had been previously synonymized; (3) determining whether putatively pure individuals of ‘Barbus’ cercops are found in the Kenyan LVD, as a previous study only found hybrid individuals of this species in this region; and (4) examining the phylogenetic relationships of Labeobarbus altianalis with other Labeobarbus species. To this end, we obtained mitochondrial Cytochrome b and nuclear Growth Hormone (GH) intron 2 gene sequences of nine ‘Barbus’ species from the LVD in Kenya, as well as cytochrome b sequences for L. altianalis. We conducted Maximum likelihood and Bayesian phylogenetic analyses to establish their evolutionary relationships in relation to many other barbs specimens from Africa. Phylogenetic analyses did not reveal instances of hybridization/introgression among the individuals sequenced by us. A sister relationship between ‘B’. profundus and ‘B’. radiatus was not found. This latter species shows instead a sister relationship with a lineage comprised of two species from West Africa. Other sister relationships between taxa from the East coast and other ecoregions from Africa are observed, suggesting that past drainage connections and vicariant events contributed to the diversification of this group. Finally, only a single haplotype was recovered among the L. altianalis individuals examined, which is most similar to a specimen from Lake Edward in Uganda.
EVOLUTION OF AFRICAN BARBS FROM THE LAKE VICTORIA DRAINAGE SYSTEM, KENYA

Violet M. Ndeda¹,², Mariana Mateos¹, Luis A. Hurtado¹* 

¹ Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, Texas
² Present address: Department of Zoology, Maseno University, Kenya

*Corresponding author:
Luis A. Hurtado, Department of Wildlife and Fisheries Sciences, Texas A&M University, 534 John Kimbrough Blvd., Wildlife Fisheries Ecological Sciences, Building #1537, Suite 278, College Station, TX 77843-2258, USA.
Tel: +1-979-458-9878; Fax: +1-979-845-4096; E-mail: lhurtado@tamu.edu
ABSTRACT

The Lake Victoria drainage basin (LVD) in Kenya is home to ten nominal species of small barbs (‘Barbus’) and one of large barbs (Labeobarbus altianalis). A recent molecular study genetically characterized small barbs in this region and found evidence of introgression between certain species, further complicating the taxonomy and species identification of these fishes. This study aimed to extend our understanding on the evolution of these fishes by: (1) examining the phylogenetic relationships of small barbs of the Kenyan LVD with those reported from other ichthyological provinces of Africa; (2) testing the sister relationship between ‘Barbus’ profundus, endemic to Lake Victoria, and ‘Barbus’ radiatus, also found in Lake Victoria, which had been previously synonymized; (3) determining whether putatively pure individuals of ‘Barbus’ cercops are found in the Kenyan LVD, as a previous study only found hybrid individuals of this species in this region; and (4) examining the phylogenetic relationships of Labeobarbus altianalis with other Labeobarbus species. To this end, we obtained mitochondrial Cytochrome b and nuclear Growth Hormone (GH) intron 2 gene sequences of nine ‘Barbus’ species from the LVD in Kenya, as well as cytochrome b sequences for L. altianalis. We conducted Maximum likelihood and Bayesian phylogenetic analyses to establish their evolutionary relationships in relation to many other barbs specimens from Africa. Phylogenetic analyses did not reveal instances of hybridization/introgression among the individuals sequenced by us. A sister relationship between ‘B’. profundus and ‘B’. radiatus was not found. This latter species shows instead a sister relationship with a lineage comprised of two species from West Africa. Other sister relationships between taxa from the East coast and other ecoregions from Africa are observed, suggesting that past drainage connections and vicariant events contributed to the diversification of this group. Finally, only a single haplotype was recovered among the L.
altianalis individuals examined, which is most similar to a specimen from Lake Edward in Uganda.
1. INTRODUCTION

Barbs constitute a significant component of the freshwater fish fauna of Africa, and represent the most species-rich group of cyprinids in this continent (Hayes and Armbruster, 2017; Leveque and Daget, 1984; Ren and Mayden, 2016; Skelton, 1988; Skelton, 1993; Skelton et al., 1991).

Molecular characterization of African barbs from different regions has greatly contributed to our understanding on the diversity and evolution of these fishes (Beshera et al., 2016; De Graaf et al., 2007; Hayes and Armbruster, 2017; Muwanika et al., 2012; Ren and Mayden, 2016; Schmidt et al., 2017; Yang et al., 2015). A large dataset of DNA sequences (particularly of the mitochondrial Cytochrome b gene) of African barbs from different regions has accrued, providing a resource for performing phylogenetic analyses across regions, which will enhance knowledge on the systematics, evolution, and biogeographic history of this important group.

Although they were treated as part of Barbus Cuvier and Cloquet, 1816, which included >800 species distributed across Eurasia and Africa (Berrebi et al., 1996; Skelton, 2012; Skelton et al., 1991), molecular phylogenetic studies have corroborated that this taxonomically complex and heterogeneous assemblage is a polyphyletic group (Ren and Mayden, 2016; Tsigenopoulos et al., 2002; Yang et al., 2015). The large hexaploid African barbs are now classified as Labeobarbus (tribe Torini; Tsigenopoulos et al., 2010; Yang et al., 2015), whereas the large tetraploid African barbs are classified as Pseudobarbus (tribe Smiliogastrini; Yang et al., 2015). The small, diploid, African barbs have also been assigned to the tribe Smiliogastrini, and Yang et al. (2015) proposed to include all of them within the genus Enteromius, the oldest available genus-group name for these fishes, even though they do not appear to correspond to a monophyletic group. This proposal is controversial, with some authors supporting it (Hayes and Armbruster, 2017), whereas others proposing that this group be referred to as ‘Barbus’ to reflect
its taxonomic uncertainty (Schmidt and Bart, 2015; Stiassny and Sakharova, 2016). To minimize confusion with a previous study that was conducted in the same area as the present study (Schmidt et al., 2017), hereafter we generally refer to them as ‘Barbus’ in the main text. Nonetheless, in figures and references to figures, we generally use the genus and species name that was used by the original contributor of the corresponding sequence.

Barbs are an important biodiversity component of the Lake Victoria drainage Basin (LVD) in Kenya, and play a significant role in food security and socioeconomic development of the local community (Ochumba and Many Ala, 1992; Okeyo, 2014). A recent multilocus study (Schmidt et al., 2017), molecularly characterized most of the small barbs nominal species present in this region. Phylogenetic relationships among them were analyzed, and high levels of genetic divergence within some recognized species were uncovered. Further complicating the taxonomy and species identification within this group, this study revealed evidence of introgression involving five small barbs species. Several important phylogenetic questions, however, still remain to be answered for Kenyan LVD barbs.

First, phylogenetic relationships of small barbs from this region, which belong to the East Coast province, with those from other ichthyological provinces of Africa have not been examined. Numerous sequences, mostly of the Cytb gene, are available for other small barbs from the East Coast, as well as the Nilo-Sudan, Upper Guinea, Lower Guinea, Congo, and Southern provinces (as defined by Levêque et al., 2008; Roberts, 1975). The African continent has had a complex and dynamic geological history, in which past hydrological connections may have enabled exchange of taxa from different regions (Salzburger et al., 2014; Stewart, 2001).

Second, Schmidt et al. (2017) did not include ‘B. profundus’ in their study, a species endemic to Lake Victoria, for which molecular analyses can help to clarify its evolutionary
Greenwood (1970) originally described ‘B.’ *profundus* as a subspecies of ‘B.’ *radiatus*, another species found in Lake Victoria and other localities in Kenya. He considered ‘B.’ *radiatus* was comprised of three subspecies: *B. radiatus profundus*, *B. radiatus radiatus* and *B. radiatus aurantiacus*. Stewart (1977), however, based on meristic and morphometric analyses concluded ‘B.’ *profundus* is a separate species from ‘B.’ *radiatus*; but did not find a basis for the separation of the other two subspecies. The two species occupy different depths in Lake Victoria; ‘B.’ *profundus* is distributed at depths between 16 and 65 m (Greenwood, 1970), whereas ‘B.’ *radiatus* occupies shallower waters (Stewart, 1977). Molecular analyses are thus needed to examine the relationship between ‘B.’ *profundus* and ‘B.’ *radiatus*.

Third, due to mitochondrial introgression from other species (i.e., ‘B.’ *neumayeri* or ‘B.’ c.f. *paludinosus* “Jipe”), Schmidt et al. (2017) were not able to obtain Cytb sequences that could be attributed to the ‘B.’ *cercops* lineage. It is thus unclear whether pure populations of this species exist in the Kenyan LVD, which is important for conservation. Lack of Cytb sequences also limits examination of the evolutionary relationships of ‘B.’ *cercops* with the other small African barbs for which Cytb sequences are available.

Finally, Cytb sequences for the large barb *Labeobarbus altianalis* in the Kenyan LVD have not been examined. This species has historically constituted an important fishery in this region (Whitehead, 1959), but overfishing has severely decimated its populations (Ochumba and Many Ala, 1992). Genetic diversity for this species in the Kenyan LVD has been studied with the mitochondrial control region, which revealed some population structure (Chemoiwa et al., 2013). Muwanika et al. (2012) examined partial Cytb sequences for *L. altianalis* from different localities in Uganda, including the Lake Victoria and Albertine basins. Therefore, obtaining Cytb sequences from this species in the Kenyan LVD will allow examination of differences
among the populations from both countries. In addition, a large dataset of Cytb sequences exists for *Labeobarbus* from different regions in Africa, which has not been analyzed with this species (Beshera et al., 2016).

Herein, obtained Cytb sequences from eight species of small barbs and the large barb *L. altianalis* from different localities in the Kenyan LVD, and conducted phylogenetic analyses of these with a large dataset of reported sequences of small and large African barbs. We also obtained sequences of the nuclear GH intron and conducted phylogenetic analyses to help determine whether individuals were pure or exhibited evidence of hybridization/introggression (Schmidt et al., 2017). Our main objectives were to: (1) examine the phylogenetic relationships of small barbs of the Kenyan LVD with those reported from other ichthyological provinces of Africa; (2) test the sister relationship between ‘B.’ *profundus* and ‘B.’ *radiatus*; (3) determine whether putatively pure individuals of ‘B.’ *cercops* are found in the Kenyan LVD; and (4) examine the phylogenetic relationships of *L. altianalis* with other *Labeobarbus* species.

### 2. MATERIALS AND METHODS

#### 2.1 Tissue source for DNA

We used ethanol-preserved fin clips from nine species (*Labeobarbus altianalis*, ‘B.’ *apleurogramma*, ‘B.’ *profundus*, ‘B.’ *cercops*, ‘B.’ *nyanzae*, ‘B.’ *kerstenii*, ‘B.’ *jacksoni*, ‘B.’ *neumayeri*, and ‘B.’ *paludinosus*) loaned by the Kenya Marine and Fisheries Institute (KMFRI). The remainder of the specimens is stored in formalin at KMFRI. These specimens were originally identified by fish taxonomists from KMFRI using morphological identification keys according to Greenwood (1962). They were obtained from sixteen localities in the Lake Victoria drainage area (LVD) in Kenya, which included Lake Victoria, rivers draining to the lake and
associated dams [represented by triangles in Fig. 1; circles indicate the approximate location of specimens from Schmidt et al. (2017) included in our analyses].

2.3 DNA isolation, PCR amplification and sequencing

Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN Inc.). The quality of extracted DNA was examined by visualization on a 1.5% agarose electrophoresis gel, and quantified with a NanoDrop® ND-1000 spectrophotometer. Fragments of one mitochondrial (Cytochrome $b$; Cytb; ~1140bp) and one nuclear (Growth Hormone Intron 2; GH; ~520bp) gene were PCR amplified from 1-4 individuals per locality. PCR was performed in a 25 μl reaction containing 19.9 μl ultrapure water, 0.5 μl dNTP mix (2.5 mM), 2.5 μl of 10X buffer, 0.5 μl of each 10μM primer, 0.1 μl Taq polymerase (OneTaq, New England Biolabs, Inc), and 1 μl of DNA template. Cytb was amplified with primers Cytb L15267 (5’AATGACTTGAAGAACCACCGT3’) and H16461 (5’CTTCGGATTACAAGACC3’), following Briolay et al. (1998). GH intron 2 was amplified using primers GH102F (5’TCGTGTACAACACCTGCACCAGC-3’), GH148R (5’ TCCTTTCCGGTGGGTGCCTCA-3’), from Mayden et al. (2009). PCR amplification included a denaturation step of 2 min at 95°C followed by 35 cycles of 1 min at 95°C, 30 s at 58–60°C (Cytb)/ 55°C (GH) and 1 min at 72°C followed in turn by a final extension of 6 min at 72°C. Successful amplification was verified by running the PCR amplicons alongside a standard Lambda ladder on a 1.5% agarose gel stained with GelRed™ (Biotium Inc., Hayward, CA, USA). Products were sequenced bi-directionally using the amplification primers in an ABI 3730 capillary sequencer.

2.5 Sequence assembly and alignment
Nucleotide sequences were assembled and edited with Sequencher 4.8 (Gene Codes, Ann Arbor, MI, USA). Newly generated Cytb of ‘Barbus’ were combined with publicly available sequences of African ‘Barbus’ and their allies (Systomus, Barboides, Clypeobarbus, Pseudobarbus, Labeobarbus), including sequences from Schmidt et al. (2017). Sequences were aligned with MAAFT v.6.0 (Katoh and Toh, 2008). Cytb aligned sequences were translated into amino acids to verify the alignments and to rule out the occurrence of frameshifts and early stop codons that could be indicative of pseudogenes or sequencing errors. Species from the family Catostomidae, which represent a group of tetraploids thought to have arisen due to a hybridization event early (60 million years) in the history of the cypriniform fishes, were initially used as outgroups (Uyeno and Smith, 1974). Following preliminary analyses, the dataset was pruned to retain only taxa relevant to this study: the newly generated sequences; small barbs closely related to the taxa in this study; close relatives of Labeobarbus altianalis, and four appropriate outgroup taxa (Pethia ticto, Hampala macrolepidota, Puntigrus tetrazona, and Systomus sarana; following Schmidt et al. (2017)). The GH dataset included the newly generated sequences, representatives of seven Barbus species from the same region, and sequences of Pethia and Garra were used as outgroups.

### 2.6 Phylogenetic analyses

Phylogenetic analyses were performed using maximum likelihood (Stamatakis, 2014), and Bayesian inference (Huelsenbeck et al., 2001). Appropriate models of sequence evolution for these analyses were determined using PARTITIONFINDER v2.7 (Guindon et al., 2010; Lanfear et al., 2012; Lanfear et al., 2017) and JModeltest 2.1.10 (Darriba et al., 2012) under the Akaike Information Criterion (AIC), corrected AIC(c), and Bayesian Information Criterion (BIC) (Table 1).
Bayesian analyses were performed in MrBayes 3.2.6 (Ronquist et al., 2012) via the CIPRES Science Gateway (Miller et al., 2010). We used the model indicated by the BIC criterion of JModeltest or the closest more complex model available in MrBayes. The analysis was run for 10,000,000 generations consisting of four independent Markov Chain Monte Carlo (MCMC) chains sampled at every 1000 generations. TRACER v1.6 was used to assess MCMC stationarity and to ensure adequate effective sampling size values (>200) were achieved. The first 25% of the sampled trees were discarded as burn-in, whereas the remaining sampled trees were summarized with “sumt” command implemented in MrBayes.

Maximum Likelihood (ML) analysis was implemented in RaxML v 8.2.6 (Stamatakis, 2014) using rapid bootstrap and GTRGAMMA model via the CIPRES Science Gateway (Miller et al., 2010) to generate a maximum likelihood tree. Clade support was examined by a nonparametric bootstrap analysis of 200 replicates and summarized with 50% majority rule consensus tree computed using the SUMTREES script (v.3.3.1) (Sukumaran and Holder, 2010).

3. RESULTS AND DISCUSSION

We obtained new Cytb sequences for 48 specimens and new GH sequences for 34 specimens (overlap = 26; see Supporting Table S1; GenBank Accession Nos. MH484522-MH484603). Alignments are available as nexus files under Datasets S1 and S2. Phylogenetic reconstructions using GH and Cytb DNA sequences are shown in Figures 2 and 3–8, respectively. In our trees, we generally retained the genus and species name given by the original contributors of the sequences. With the exception of ‘B.’ cercops Cytb sequences (see below), for the seven ‘Barbus’ species that overlapped between our study and that of Schmidt et al. (2017) within the
LVD region, the Cytb haplotypes and GH alleles that we obtained were identical or very similar to those reported by Schmidt et al. (2017) (Figs. 2–8).

Comparison of Cytb and GH trees based on our sequences alone does not suggest instances of introgression or hybridization. All of the eight putative ‘B.’ cercops specimens examined (localities 1, 5, 9; Fig. 1) had a GH sequence identical to the single allele reported by Schmidt et al. (2017) for 15 ‘B.’ cercops specimens in the same area (i.e., localities 5, 16, 21, 23–25). The ‘B.’ cercops GH allele is distinct from alleles found in specimens assigned to all other species examined to date. Based on GH, the ‘B.’ cercops lineage (red branches in Fig. 2) forms a monophyletic group with ‘B.’ sp. “Jipe” and a clade comprised of ‘B.’ jacksoni, ‘B.’ trispilopleura, and one specimen assigned to B. trimaculatus (Fig. 2).

The Cytb sequences of our ‘B.’ cercops specimens (n = 6) formed a distinct lineage (red branches in Fig. 4) that excluded the five ‘B.’ cercops specimens reported by Schmidt et al. (2017) and all other reported sequences to date [with the exception of GenBank record AF180841; identified as Barbus nyanzae from Kenya (Tsigenopoulos et al., 2002); discussed below]. Maximum Cytb divergence within this clade was 0.39% (K2P). This ‘B.’ cercops Cytb lineage was part of a larger clade (Clade A; Figs. 3 and 4) that included the closest relatives of ‘B.’ cercops according to the GH gene (see above), as well as additional lineages assigned to several other species. The Cytb sequence of the five putative ‘B.’ cercops specimens examined by Schmidt et al. (2017), including four for which they also obtained GH sequences, clustered with the Cytb sequences of other species. Four of their ‘B.’ cercops Cytb sequences clustered with a clade made up mostly of ‘B.’ neumayeri specimens (brown branches; Figs. 3 and 5; found in localities 5 and 25; Fig. 1), whereas one of their ‘B.’ cercops Cytb sequences (from locality 23) clustered with individuals belonging to a subclade (i.e., “Jipe”) of specimens assigned to ‘B.’
cf. *paludinosus* (Clade D; green branches; Figs. 3 and 5). Therefore, whereas Schmidt et al. (2017) detected evidence consistent with mitochondrial introgression from other species (i.e., *B.* *neumayeri* and *B.* cf. *paludinosus*) into all five of the *B.* *cercops* specimens that they characterized for Cytb as well as for morphology or GH, we found no evidence of introgression among our specimens. Whether GenBank record AF180841 (Fig. 4) is an error or a result of introgression of *B.* *cercops* mitochondria into *B.* *nyanzae* cannot be determined [although Schmidt et al. (2017) did not detect such introgression], because information about the nuclear genetic background is not available for this specimen. Similarly, the identity of *B.* *cercops* voucher 1550 [blue taxon label in Fig. 2; from Schmidt et al. (2017)] is questionable, as its Cytb sequence (KX178183) is very divergent (not shown). A MegaBLAST search (Morgulis et al., 2008; Zhang et al., 2000) against the non-redundant NCBI nucleotide database indicates that this sequence, along with records KX178138 and KX178111 also labeled as *B.* *cercops*, belong to a different fish order (i.e., Siluriformes), implying a labeling or contamination error. Schmidt et al. (2017) also detected a pattern suggestive of introgression in *B.* *kerstenii*. They assigned 30 specimens to *B.* *kerstenii* based on morphology, and characterized 16 of these for GH and 28 for Cytb (15 for both genes). The 16 GH sequences grouped into a distinct clade that excluded specimens assigned to other species (Fig. 2). In contrast, for Cytb, 11 grouped within the *B.* *neumayeri* clade (Fig. 5), one grouped with specimens assigned to *B.* *nyanzae* (purple in Fig. 6), and the remaining 17 formed a highly distinct clade that excluded specimens assigned to other species (magenta in Fig. 6). The latter presumably represent “pure” *B.* *kerstenii* specimens, as the phylogenetic position of this unique clade (i.e., as a close relative of *B.* *nyanzae*) is generally congruent between the two genes. Based on this criterion, the four
‘B.’ kerstenii specimens for which we obtained both GH and Cytb sequences represent “pure” individuals.

Our study is the first to report Cytb and GH DNA sequences from ‘B.’ profundus, a species endemic to Lake Victoria (Greenwood, 1970). Each of the four specimens examined had a different Cytb haplotype (max. within clade divergence = 0.59% K2P), and formed a well supported clade (Fig. 7). The ‘B.’ profundus lineage falls within Clade B, which contains mostly species distributed in the Nilo-Sudan (North) and Upper Guinea (West) provinces, as well as ‘B.’ radiatus. A sister relationship between B’. profundus and ‘B.’ radiatus, however, was not recovered in our Cytb or GH analyses, despite the fact that these species had been previously synonymized (Greenwood, 1970) and co-occur in Lake Victoria, albeit at different depths (Stewart, 1977). Instead, the Cytb phylogenetic reconstruction (Fig. 7) shows a sister relationship between ‘B.’ radiatus and a lineage comprised of two species from West Africa (in the Lower Guinea ichthyological province): ‘B.’ aspilus (from Cameroon) and ‘B.’ cf. guirali (the sample is from Gabon; ‘B.’ guirali is reported from Cameroon, Gabon and Congo). This is congruent with the results of Ren and Mayden (2016), who examined the same sequences for these taxa, but lacked ‘B.’ profundus. Therefore, ‘B.’ profundus and ‘B.’ radiatus do not appear to constitute sister taxa, but phylogenetic analyses with additional taxa and markers are necessary to identify their closest relatives.

By including sequences from independent studies, our Cytb analyses revealed previously unknown relationships involving LVD species. ‘B.’ yongei was sister (~12% K2P divergent) to a lineage comprised of two almost identical haplotypes from Guinea identified as Enteromius sp. and Enteromius stigmatopygus in Hayes and Armbruster (2017), which implies an East vs. West Africa divergence (Fig. 4). Another sister relationship involving LVD taxa uncovered by our
analyzes was that of ‘B.’ nyanzae (LVD) and ‘B.’ laticeps (from Tanzania) (Fig. 6), which were
~9% divergent (K2P).
A single haplotype was recovered among the six L. altianalis individuals examined (Fig. 8), representing five localities. This haplotype is identical to GenBank record KX178106, submitted as part of the Schmidt et al. (2017) study PopSet (Accession No. 1187422681), but erroneously labeled as ‘B.’ cercops (Fig. 8). The lack of Cytb diversity among our L. altianalis specimens sharply contrasts with a previous study of this species in this area that reports high haplotype diversity for the mitochondrial control region (Chemoiwa et al. (2013). Although our phylogenetic analyses provide little resolution within Labeobarbus, they suggest that our L. altianalis haplotype is most similar to GenBank record JN983691 (627 bp; K2P distance = 1.2%); a specimen from Uganda (Lake Edward; Albertine drainage; ~200 km West of Lake Victoria) contributed by Muwanika et al. (2012). The other Labeobarbus specimens examined by Muwanika et al. (2012) from LVD and the Albertine drainage in Uganda were 2.2–4.5% (K2P) divergent from our L. altianalis haplotype. Banister (1973) proposed, based on morphology, two groups within Labeobarbus: the Labeobarbus intermedius complex (L. intermedius, L. altianalis, ‘Barbus’ acuticeps, and ‘B.’ ruasae) and the Labeobarbus bynni complex (L. bynni, L. gananensis, L. oxyrhynchus, and ‘B.’ longifilis). Cytb phylogenetic reconstructions in this study, however, do not support the separation of these two groups, which is congruent with the findings of a previous phylogenetic analyses that lacked L. altianalis sequences (Beshera et al., 2016). Nonetheless, Cytb may be too conserved to adequately assess these relationships.
Several interesting broad-scale phylogeographic patterns are apparent with the available Cytb data on African small barbs. Continental Africa is divided into nine ichthyological
provinces (reviewed in Levêque et al., 2008). The Lake Victoria Drainage belongs to the East
Coast province. From our analyses, the following East Coast vs. West splits can be inferred:

1. ‘B.’ radiatus (LVD) vs. ‘B.’ cf. guiirali + ‘B.’ aspilus (West: Lower Guinea province); (2) ‘B.’
profundus (LVD) vs. one or more of the members of Clade B (all are from western Africa except
‘B.’ radiatus); and (3) ‘B.’ yongei (LVD) vs. E. stigmatopygus (West: Niger River). One or more
additional East vs. West splits will be identified within clades A and C once relationships within
these are resolved. Two East Coast vs. Southern ecoregion splits are inferred: (1) ‘B.’
trimaculatus (South Africa) vs. ‘B.’ jacksoni + ‘B.’ perince + ‘B.’ trispilopleura (including
specimens assigned to ‘B.’ tanapelagius and ‘B.’ humilis; Fig. 4); and (2) the basal split within
Clade D (‘B.’ cf. paludinosus). The multiple divergences between the East Coast and other
provinces suggest that the dynamic and complex geological history of Africa provided
opportunities, through hydrological connections, for exchange of lineages from different regions
(Salzburger et al., 2014; Stewart, 2001).

4. CONCLUSION

The taxonomy and evolutionary history of the African barbs, including the role of hybridization,
is far from resolved, and will require a much broader sampling of taxa, geographic locations, and
 genetic markers than what is presently available. Nonetheless, our analyses, which included
most (if not all) available Cytb and GH sequences for this group, revealed several key insights.
First, apparently pure ‘B.’ cercops individuals do occur at the three localities where we obtained
this species, including the Kendu Bay area (locality 5), where Schmidt et al. (2017) reported a
‘B.’ cercops specimen harboring a ‘B.’ neumayeri mitochondrion. Secondly, ‘B.’ radiatus does
not appear to be sister to ‘B.’ profundus, with which it was previously synonymized. Thirdly, we
found evidence of several sister relationships between taxa from the East Coast and other
coregions of Africa, suggesting that past drainage connections and vicariant events contributed
to the diversification of this group. Finally, only a single haplotype was recovered among the L.
altianalis individuals examined, which is most similar to a specimen from Lake Edward than to
specimens from other localities in Uganda.

ACKNOWLEDGEMENTS
The Kenya Marine and Fisheries Research Institute (KMFRI) provided the samples
analyzed in this study. James Woolley provided suggestions and comments regarding this work.
Cecilia Smith at Texas A&M University Libraries helped with the elaboration of the study area
map. M.Sc. studies of VN were supported by fellowships from Fulbright and the American
Association of University Women.
REFERENCES CITED


Beshera, K. A., P. M. Harris, and R. L. Mayden. 2016. Novel evolutionary lineages in Labeobarbus (Cypriniformes; Cyprinidae) based on phylogenetic analyses of mtDNA sequences. Zootaxa 4093:363-81. 10.11646/zootaxa.4093.3.4


Figure 1. Localities of specimens sampled in this study (triangles). Circles represent localities within the Lake Victoria Drainage (LVD) sampled by Schmidt et al. (2017) that were not sampled in our study (locations are approximate based on their description of locality, as coordinates were not reported). The map was developed with ArcMap version 10.3—a part of the ESRI ArcGIS® Desktop suite. Localities where each species was sampled for our study are as follows: ‘B.’ apleurogramma [2]; ‘B.’ cercops [1, 5, 7, 9]; ‘B.’ cf. paludinosus [13]; ‘B.’ jacksoni [7]; ‘B.’ kerstenii [1, 6, 14]; ‘B.’ neumayeri [3]; ‘B.’ nyanzae [7, 10]; ‘B.’ profundus [8]; and L. altianalis [4, 5, 10, 11, 12].

Figure 2. Inferred relationships based on the GH gene. RaxML bootstrap consensus (60% majority rule) tree. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate Cytb sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses. For visual clarity, several node support labels have been omitted. Red taxon labels indicate specimens identified by Schmidt et al. (2017) as introgressed. Blue taxon label indicates specimen whose cyt b sequence is falls within the order Siluriformes. Clade labels and colors correspond to those in other figures. The dashed line of Clade A indicates those lineages that were not found monophyletic with the rest of Clade A members (as defined by the cyt b tree; Figs. 3 and 4).
Figure 3. Inferred relationships among the major clades in this study based on the Cytb gene, and general distribution (East, West, or South Africa; where applicable). RaxML Bootstrap consensus (60% majority rule). Specific clades are expanded in Figs. 4–8). Numbers by nodes represent ML Bootstrap support values. For visual clarity, several node support labels have been omitted. Each taxon label contains the GenBank Accession Number and/or the citation and voucher ID.

Figure 4. Inferred relationships within Clade A (expanded from Fig. 3) based on the Cytb gene. RaxML bootstrap consensus (60% majority rule) tree. All taxa except those with grey shading are found in East Africa. Each taxon label contains the GenBank Accession Number and/or the citation and voucher ID, as well as locality label [if they were from the Lake Victoria Drainage (LVD); in bracket] corresponding to labels in Fig. 1. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate GH sequence was also generated in present study. Asterisks by nodes represent support values > 97% for all analyses.

Figure 5. Inferred relationships within the clades of ‘B.’ neumayeri, ‘B.’ apleurogramma, and ‘B.’ cf. paludinosus (Clade D) (expanded from Fig. 3) based on the Cytb gene. RaxML bootstrap consensus (60% majority rule) tree. All taxa except the one with grey shading are found in East Africa. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate GH sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses.
For visual clarity, several node support labels have been omitted. Red taxon label indicates putatively introgressed individuals from Schmidt et al. (2017).

**Figure 6.** Inferred relationships within Clade C (expanded from Fig. 3) based on the Cytb gene. RaxML bootstrap consensus (60% majority rule) tree. All taxa except those with grey shading are found in East Africa. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate GH sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses. For visual clarity, several node support labels have been omitted. Red taxon label indicates specimen from Schmidt et al. (2017) assigned to ‘B.’ kerstenii based on morphology and GH sequence, but with a Cytb sequence that falls within ‘B.’ nyanzae.

**Figure 7.** Inferred relationships within Clade B (expanded from Fig. 3) based on the Cytb gene. RaxML bootstrap consensus (60% majority rule) tree. All taxa except those ‘B.’ radiatus and ‘B.’ profundus (boldface taxon labels) are distributed in West Africa. Asterisks by taxon names indicate GH sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses. For visual clarity, several node support labels have been omitted.

**Figure 8.** Inferred relationships within the clades “Labeobarbus and allies” (expanded from Fig. 3) based on the Cytb gene. RaxML bootstrap consensus (60% majority rule) tree. Boldfaced
taxon labels correspond to our specimens of *L. altianalis*. Numbers in brackets correspond to localities in Fig. 1.
**SUPPORTING INFORMATION DESCRIPTIONS**

**Table S1.** Specimen information for samples sequenced in this study including geographic coordinates and GenBank Accession Numbers. Locality numbers match labels in Figures. Blue font indicates specimens for which both the Cytb and GH sequence were obtained.

**Dataset S1.** Cytochrome b gene alignment. Annotated nexus-formatted file of the Cytochrome b gene sequence alignment used in this study.

**Dataset S2.** GH intron gene alignment. Annotated Nexus-formatted file of the GH gene sequence alignment used in this study.
Localities of specimens sampled in this study (triangles).

Circles represent localities within the Lake Victoria Drainage (LVD) sampled by Schmidt et al. (2017) that were not sampled in our study (locations are approximate based on their description of locality, as coordinates were not reported). The map was developed with ArcMap version 10.3—a part of the ESRI ArcGIS® Desktop suite. Localities where each species was sampled for our study are as follows: ‘B.’ apleurogramma [2]; ‘B.’ cercops [1, 5, 7, 9]; ‘B.’ cf. paludinosus [13]; ‘B.’ jacksoni [7]; ‘B.’ kerstenii [1, 6, 14]; ‘B.’ neumayeri [3]; ‘B.’ nyanzae [7, 10]; ‘B.’ profundus [8]; and L. altianalis [4, 5, 10, 11, 12].
Inferred relationships based on the GH gene.

RaxML bootstrap consensus (60% majority rule) tree. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate Cytb sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses. For visual clarity, several node support labels have been omitted. Red taxon labels indicate specimens identified by Schmidt et al. (2017) as introgressed. Blue taxon label indicates specimen whose cytB sequence is falls within the order Siluriformes. Clade labels and colors correspond to those in other figures. The dashed line of Clade A indicates those lineages that were not found monophyletic with the rest of Clade A members (as defined by the cytB tree; Figs. 3 and 4).
Inferred relationships among the major clades in this study based on the Cytb gene, and general distribution (East, West, or South Africa; where applicable).

RaxML Bootstrap consensus (60% majority rule). Specific clades are expanded in Figs. 4–8). Numbers by nodes represent ML Bootstrap support values. For visual clarity, several node support labels have been omitted. Each taxon label contains the GenBank Accession Number and/or the citation and voucher ID.
Figure 4

Inferred relationships within Clade A (expanded from Fig. 3) based on the Cytb gene.

RaxML bootstrap consensus (60% majority rule) tree. All taxa except those with grey shading are found in East Africa. Each taxon label contains the GenBank Accession Number and/or the citation and voucher ID, as well as locality label [if they were from the Lake Victoria Drainage (LVD); in bracket] corresponding to labels in Fig. 1. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate GH sequence was also generated in present study. Asterisks by nodes represent support values > 97% for all analyses.
Figure 5

Inferred relationships within the clades of ‘B.’ neumayeri, ‘B.’ apleurogramma, and ‘B.’ cf. paludinosus (Clade D) (expanded from Fig. 3) based on the Cytb gene.

RaxML bootstrap consensus (60% majority rule) tree. All taxa except the one with grey shading are found in East Africa. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate GH sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses. For visual clarity, several node support labels have been omitted. Red taxon label indicates putatively introgressed individuals from Schmidt et al. (2017).
Inferred relationships within Clade C (expanded from Fig. 3) based on the Cytb gene.

RaxML bootstrap consensus (60% majority rule) tree. All taxa except those with grey shading are found in East Africa. Bold-faced taxon labels indicate sequences generated by the present study. Asterisks by taxon names indicate GH sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses. For visual clarity, several node support labels have been omitted. Red taxon label indicates specimen from Schmidt et al. (2017) assigned to ‘B.’ *kerstenii* based on morphology and GH sequence, but with a Cytb sequence that falls within ‘B.’ *nyanzae*.
Inferred relationships within Clade B (expanded from Fig. 3) based on the Cytb gene.

RaxML bootstrap consensus (60% majority rule) tree. All taxa except those ‘B.’ radiatus and ‘B.’ profundus (boldface taxon labels) are distributed in West Africa. Asterisks by taxon names indicate GH sequence was also generated in present study. Numbers in brackets correspond to localities in Fig. 1. Node support values from Bayesian (above) and ML below are given by nodes; an asterisk indicates support > 97% in all corresponding analyses. For visual clarity, several node support labels have been omitted.
Figure 8

Inferred relationships within the clades “Labeobarbus and allies” (expanded from Fig. 3) based on the Cytb gene.

RaxML bootstrap consensus (60% majority rule) tree. Boldfaced taxon labels correspond to our specimens of *L. altianalis*. Numbers in brackets correspond to localities in Fig. 1.
Labeobarbus and allies

L. altianalis
Table 1 (on next page)

Description of characters and substitution models identified by model selection analyses.

Best model selected by: (a) jModeltest 2.1.10 v20160303 (Darriba et al., 2012) according to each criterion (AICc, AIC, BIC) and its corresponding weight (on a fixed BioNJ tree); and (b) the best partitioning scheme according to the BIC implemented in PartitionFinder 2.7 (Guindon et al., 2010; Lanfear et al., 2012; Lanfear et al., 2017): branchlengths = linked; and search = greedy.
Table 1. Description of characters and substitution models identified by model selection analyses. Best model selected by: (a) jModeltest 2.1.10 v20160303 (Darriba et al., 2012) according to each criterion (AICc, AIC, BIC) and its corresponding weight (on a fixed BioNJ tree); and (b) the best partitioning scheme according to the BIC implemented in PartitionFinder 2.7 (Guindon et al., 2010; Lanfear et al., 2012; Lanfear et al., 2017): branchlengths = linked; and search = greedy.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Non-redundant taxa</th>
<th>Characters used</th>
<th>Parsimony informative</th>
<th>Partitioning Scheme</th>
<th>AICc (weight)</th>
<th>AIC (weight)</th>
<th>BIC (weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH</td>
<td>46</td>
<td>201</td>
<td>63</td>
<td>1</td>
<td>TPM2uf (0.36)</td>
<td>TVM (0.22)</td>
<td>TVM (0.24)</td>
</tr>
<tr>
<td>Cytb</td>
<td>291</td>
<td>1023</td>
<td>507</td>
<td>1</td>
<td>TIM2+I+G (0.99)</td>
<td>TIM2+I+G (0.61)</td>
<td>TIM2+I+G (0.99)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Codon 1</td>
<td>SYM+I+G</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Codon 2</td>
<td>HKY+I+G</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Codon 3</td>
<td>GTR+G</td>
<td></td>
</tr>
</tbody>
</table>