Evolutionary biology

Geographical ancestry of Lake Malawi's cichlid fish diversity

Martin J. Genner1, Benjamin P. Ngatunga2, Semvua Mzighani2, Alan Smith3 and George F. Turner4

1 School of Biological Sciences, University of Bristol, Life Sciences Building, Bristol BS8 1TQ, UK
2 Tanzania Fisheries Research Institute (TAFIRI), PO Box 9750, Dar-es-Salaam, Tanzania
3 School of Biological, Biomedical and Environmental Science, University of Hull, Hull HU6 7RX, UK
4 School of Biological Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, UK

The Lake Malawi haplochromine cichlid flock is one of the largest vertebrate adaptive radiations. The geographical source of the radiation has been assumed to be rivers to the south and east of Lake Malawi, where extant representatives of the flock are now present. Here, we provide mitochondrial DNA evidence suggesting the sister taxon to the Lake Malawi radiation is within the Great Ruaha river in Tanzania, north of Lake Malawi. Estimates of the time of divergence between the Lake Malawi flock and this riverine sister taxon range from 2.13 to 6.76 Ma, prior to origins of the current radiation 1.20–4.06 Ma. These results are congruent with evaluations of 2–3.75 Ma fossil material that suggest past faunal connections between Lake Malawi and the Ruaha. We propose that ancestors of the Malawi radiation became isolated within the catchment during Pliocene rifting that formed both Lake Malawi and the Kipengere/Livingstone mountain range, before colonizing rivers to the south and east of the lake region and radiating within the lake basin. Identification of this sister taxon allows tests of whether standing genetic diversity has predisposed Lake Malawi cichlids to rapid speciation and adaptive radiation.

1. Introduction

Adaptive radiations make up a high proportion of biodiversity. In many cases, ancestors or sister species of these flocks have been identified, as with Galapagos finches [1], Hawaiian silverswords [2] and Canadian three-spined sticklebacks [3]. Identification of their origins has enabled discussion of events that initiated adaptive radiation, and allowed tests of whether diversification has been promoted by novel mutations that have arisen since colonization, or instead whether adaptation is based primarily on pre-existing genetic variation [4]. This is an important issue to resolve, because it can explain why only some colonizing lineages radiate when provided with ecological opportunity, and how parallel adaptive radiation can take place rapidly in geographically separated habitats.

The evolutionary origins of cichlid fishes radiations in East African lakes are largely elusive or speculative [5–8]. This is partly because of incomplete geographical and genomic sampling of riverine species within and surrounding lake basins. However, it is also due to intrinsic complexity of cichlid evolutionary relationships, as radiations may have been seeded by multiple riverine ancestors [8,9], and rivers can be recolonized by species with lacustrine ancestry [10]. A greater understanding of geographical and phylogenetic ancestry of cichlids is required to test whether functional genetic variation under divergent selection within lake radiations is present within riverine ancestors, and whether this variation has been shared among riverine cichlids through intraspecific gene flow and interspecific hybridization [9].
Lake Malawi contains a radiating flock of at least 450 haplochromine species [11]. Early phylogenetic reconstructions suggested that the lake radiation was monophyletic [12,13]. More recent phylogenies show two species outside the Lake Malawi catchment also fall within the flock, namely Astatotilapia calliptera and Astatotilapia sp. ‘Ruaha’ from this region which place it as a sister taxon to the Lake Malawi flock.

2. Material and methods

Genetic samples (fin clips) were collected from riverine haplochromines (electronic supplementary material, table S1; figure 1) and preserved in 95% ethanol. DNA was isolated using the Promega Wizard kit. Sequences of the mitochondrial gene NADH2 [7] were generated in riverine haplochromines and outgroup taxa, using CLUSTALW in DAMBE [19]. This resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two sets of calibrations [24,25] employed independently (electronic supplementary material, table S2). Bayesian phylogenies were generated with BEAST v. 1.8.0 [23] using Partitionfinder [21] time-calibrated trees resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two runs of 10 million generations. Resultant trees were generated and aligned with sequences of other haplochromines and outgroup taxa, using CLUSTALW in DAMBE [19]. This resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two sets of calibrations [24,25] employed independently (electronic supplementary material, table S2). Bayesian phylogenies were generated with BEAST v. 1.8.0 [23] using Partitionfinder [21] time-calibrated trees resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two runs of 10 million generations. Resultant trees were generated and aligned with sequences of other haplochromines and outgroup taxa, using CLUSTALW in DAMBE [19]. This resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two sets of calibrations [24,25] employed independently (electronic supplementary material, table S2). Bayesian phylogenies were generated with BEAST v. 1.8.0 [23] using Partitionfinder [21] time-calibrated trees resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two sets of calibrations [24,25] employed independently (electronic supplementary material, table S2). Bayesian phylogenies were generated with BEAST v. 1.8.0 [23] using Partitionfinder [21] time-calibrated trees resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two sets of calibrations [24,25] employed independently (electronic supplementary material, table S2). Bayesian phylogenies were generated with BEAST v. 1.8.0 [23] using Partitionfinder [21] time-calibrated trees resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two sets of calibrations [24,25] employed independently (electronic supplementary material, table S2). Bayesian phylogenies were generated with BEAST v. 1.8.0 [23] using Partitionfinder [21] time-calibrated trees resulted in an alignment of 1047 bp with 544 sequences (electronic supplementary material, table S1; figure 1) and 1/100 random models and two sets of calibrations [24,25] employed independently (electronic supplementary material, table S2).
3. Results and discussion
An as yet undescribed representative of the ‘modern haplochromine’ group, *Astatotilapia* sp. ‘Ruaha’, was present at three Great Ruaha sites (figure 1). On the basis of mitochondrial NADH2 DNA sequences, the species was resolved as an immediate sister taxon to the radiating flock (figure 1; electronic supplementary material, figure S1). The results of the analyses suggest they diverged between 2.13 Ma (95% highest posterior density (HPD) 1.52–2.84 Ma; using non-cichlid fossil derived calibrations from Friedman et al. [24]) and 6.76 Ma (95% HPD 3.76–10.12 Ma; using non-cichlid fossil derived calibrations from Schwarzer et al. [25]). This divergence took place before initial divergence of extant representatives of the Lake Malawi flock estimated at 1.2 Ma (95% HPD 1.52–2.84 Ma) or 4.06 Ma (95% HPD 2.02–6.59 Ma), from Friedman et al. [24] and Schwarzer et al. [25] calibrations, respectively. The *Astatotilapia* sp. ‘Ruaha’ lineage is geographically separated from the Malawi catchment by the Livingstone/Kipengere mountain range. This comprises steep mountainous areas and high altitude plateau, and it is plausible that both geography and low temperatures impose barriers to habitat occupancy and dispersal across the boundary [16]. The range was formed during Pliocene rifting that initiated formation of Lake Malawi [26], perhaps driving simultaneous population division and ecological opportunity for species flock formation.

Close evolutionary relationships between Malawi and upper Ruaha haplochromines are mirrored by recent observations from fish fossils of fluvialite deposits of the Chiwondo beds dated to between 2 and 3.75 Ma [18]. The Chiwondo fauna includes claredoite cichlids and tigerfish (*Hydrocynus*) [18], but geographically the nearest system containing extant representatives of these non-cichlid families is the Ruaha. It has been proposed on the basis of these fossils that rivers currently in Lake Malawi catchment were once extensions of the Great Ruaha system in pre-rift times [18]. Our results are compatible with this concept and imply further molecular studies may identify this region as a source of genetic diversity of other elements of the Malawi fauna. Notably, although the Chiwondo fauna includes representatives of Cichilidae, it has not been possible to identify remains to a lower taxonomic level [18].

It has been proposed that the ancestor of the Lake Malawi haplochromine flock is a riverine haplochromine similar to *Astatotilapia bloyeti* or *A. calliptera* [27]. Our study places specimens assigned to *A. bloyeti* in a sister clade to *Astatotilapia twillei*, consistent with previous analyses of both nuclear and mitochondrial markers [8], and our results suggest both taxa are more distantly related to Malawi cichlids than *Astatotilapia* sp. ‘Ruaha’. Our results also show that *A. calliptera* outside the Lake Malawi catchment are part of a geographically broader ‘Lake Malawi region’ flock. It remains equivocal whether the species secondarily colonized external rivers from Lake Malawi, or instead whether there have been multiple colonizations of *A. calliptera* from outside the catchment along with maintenance of the ancestral riverine phenotype [5,8,14]. In either case, given mitochondrial DNA evidence suggesting that *Astatotilapia* sp. ‘Ruaha’ is a sister species to the flock, and fossil evidence of historic connectivity of the Ruaha and Lake Malawi, it seems plausible that extant representatives of the Malawi flock are biogeographically derived from a species with a former distribution that encompassed both the Ruaha and Lake Malawi catchments. Further phylogenetic analyses based on nuclear genome data will help to provide further resolution of the relationship between *Astatotilapia* sp. ‘Ruaha’ and Malawi cichlids. Genome-wide data will also help to resolve whether *A. calliptera* occupy a basal, sister or derived position in the flock, which may force reconsideration of the biogeographic scenario suggested here.

Recent results show a high proportion of genomic diversity present within Lake Malawi cichlids is also present in riverine cichlids [9]. It has been proposed that riverine species may be active transporters of genomic material enabling rapid adaptation within lacustrine flocks. However, such situations require introgression among riverine taxa at contact zones, and gene flow across catchment boundaries. There is support for the concept of intraspecific gene flow across watersheds within Africa [16], but currently only indirect evidence of interspecific hybridization among river cichlids [27], and there is no evidence of interspecific hybridization among riverine haplochromines in the region surrounding Lake Malawi. A greater understanding of taxonomic and spatial patterns of genetic diversity within and among potentially ancestral riverine cichlids is required, including *Astatotilapia* sp. ‘Ruaha’. This would enable tests of the importance of active transport of genes through via riverine species and hybridization events for explaining shared genomic diversity among lacustrine radiations [9].

**Ethics.** Tanzania Commission for Research and Technology (COSTECH) issued a permit for this study (no. 2011-205-NA-2011-103).

**Data accessibility.** DNA sequences can be accessed from Genbank (KR010448–KR010461).

**Authors’ contributions.** M.J.G., B.P.N. and G.F.T. conceived the study and wrote the manuscript. M.J.G., B.P.N., S.M., A.S. and M.J.G. collected (KR010448–KR010461). This study was financially supported by the Royal Society-Leverhulme Trust Africa Award AA100023.

**Acknowledgements.** We thank TAFIRI staff and J. Swanstrom for field assistance.

**References.**


