
This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Elsevier at http://dx.doi.org/10.1016/j.ympev.2016.09.027. Please refer to any applicable terms of use of the publisher.
Multiple colonisations of the Lake Malawi catchment by the genus *Opsaridium* (Teleostei: Cyprinidae)

Harold Sungani a,b,* , Benjamin P. Ngatunga c, Stephan Koblmüller d, Tuuli Mäkinen e, Paul H. Skelton e and Martin J. Genner a

a School of Biological Sciences, Life Sciences Building, 24 Tyndall Avenue, Bristol. BS8 1TQ. United Kingdom.

b Fisheries Research Unit, PO Box 27, Monkey Bay, Malawi.

c Tanzania Fisheries Research Institute, PO. Box 9750, Dar es Salaam, Tanzania.

d Institute of Zoology, University of Graz, Universitätsplatz 2, A-8010 Graz, Austria.

e South African Institute for Aquatic Biodiversity, Private Bag 1015, Grahamstown 6140, South Africa.

*Corresponding author at. School of Biological Sciences, Life Sciences Building, 24 Tyndall Avenue, Bristol. BS8 1TQ. United Kingdom.

Email address: harold.sungani@bristol.ac.uk (H. Sungani).
Abstract

It has been proposed that the fish faunas of African rivers assemble through multiple colonisation events, while lake faunas form additionally through intralacustine speciation. While this pattern has been established for many lineages, most notably cichlids, there are opportunities to further investigate the concept using phylogenies of congeneric endemic species within ancient lake catchments. The Lake Malawi catchment contains three river-spawning cyprinids of the genus *Opsaridium*, two of which are endemic. These species differ in body size, migratory behaviour and habitat use, but it has never previously been tested if these represent a monophyletic radiation, or have instead colonised the lake independently. We placed these species in a broader phylogeny of *Opsaridium* and the related genus *Raiamas*, including all known species from the river systems surrounding Lake Malawi. Our results suggest that each of the species has independently colonised the lake catchment, with all three taxa having well-defined sister taxa outside of the lake, and all sharing a common ancestor ~14.9 million years ago, before the Lake Malawi basin started to form ~8.6 million years ago. Additionally, the results strongly support previous observations that *Opsaridium* is not a monophyletic group, but instead contains *Raiamas* from the Congo drainage. Together these results are supportive of the concept that river fish faunas within African catchments are primarily assembled through a process of accumulation from independent origins, rather than within-catchment speciation and adaptive radiation. In light of these results we also suggest there is scope for a re-evaluation of systematics of both *Opsaridium* and *Raiamas*.

Key words. Molecular phylogeny, Chedrina, phylogeography, river fishes, East Africa
1. Introduction

It is estimated that the African freshwater fish fauna contains approximately 3000 species (Lévêque et al., 2008). A substantial component of this diversity can be found in the adaptive radiations of the East African Rift Valley lakes, while neighbouring river systems typically have lower diversity over equivalent spatial scales. Given that rivers and large lakes can both have a broad range of ecological niches, it is possible that the distribution of species diversity is not exclusively governed by niche availability, and instead the contrasting patterns of species diversity may have arisen as a result of a different community assembly processes operating within these different habitats (Seehausen, 2015). Lakes tend to be relatively stable environments over evolutionary timescales, potentially allowing high species diversity to arise through intralacustrine speciation and adaptive niche evolution. By contrast riverine environments tend to be more dynamic and unstable over evolutionary timescales, potentially giving a relatively greater role to immigration and between-catchment allopatric divergence (Seehausen, 2015).

Evidence in support of different assembly processes operating in lakes and rivers is most prominent in African cichlid fishes. Cichlids dominate the faunas of the East African lakes, and there is extensive evidence of intralacustrine diversification in most of the lake-restricted species groups (Seehausen, 1996; Snoeks, 2004; Koblmüller et al., 2008a; Salzburger et al., 2014). By contrast riverine cichlid assemblages in systems surrounding the Great Lakes of East Africa often are comprised of only a handful of geographically widespread species, albeit with some phenotypic divergence in allopatry (Katongo et al., 2007; Koblmüller et al., 2012; Banyankimbona et al., 2013; Egger et al., 2015; Meyer et al., 2015; Nichols et al., 2015). In some cases this has required description of distinct species, for example *Orthochromis* from the Lake Tanganyika drainage (De Vos and Seegers 1998). Further evidence that speciation tends to be constrained in rivers relative to lakes is present in the weakly radiating non-cichlid lineages from Lake Malawi [clariid catfishes (Agnèse and Teugels, 2001)] and Lake Tanganyika [*Synodontis* catfishes (Day et al., 2013; Pinton et al., 2013); claroteine catfishes (Peart et al., 2014); mastacembelid eels (Brown et al., 2010)]. Most African fish families, however, have failed to radiate in lakes to any degree and their diversity tends to be greatest in riverine or still-water pool habitats. In these lineages sister species often have allopatric distributions [e.g. Dorn et al. (2011) for *Nothobranchius* killifishes, Schmidt et al. (2014) for...
Chiloglanis catfishes], implying a strong role for geographical separation in the speciation of non-lacustrine freshwater fish lineages across the continent.

Lake Malawi (=Nyasa) is estimated to have formed between ~4.5 and 8 Ma (Danley et al., 2012). The catchment contains at least 450 species of cichlid fish of which only five are frequently found in multiple river systems (Snoeks, 2004). In addition to this diversity ~50 other non-cichlid fish species are known from the catchment, of which most are non-endemics (Snoeks, 2004). One of the less well-studied species groups, from a phylogenetic perspective, is that of three cyprinids in the genus Opsaridium. These species differ extensively in body size, migratory behaviour and habitat use. Two are large-bodied species endemic to the catchment and migrate from the lake to rivers to breed [mpasa, Opsaridium microlepis (Günther, 1864); sanjika, Opsaridium microcephalum (Günther, 1864)]. A third non-endemic species, dwarf sanjika, Opsaridium tweddleorum Skelton 1996, is small-bodied in comparison, and only found in streams and rivers. It has never previously been tested if these represent a monophyletic in-situ radiation, or have instead colonised the lake independently.

Opsaridium comprises part of the “chedrin” group. This group has been defined as the Chedrini, one of three tribes within the subfamily Danioninae alongside Danionini and Rasborini (Tang et al., 2010; Ahnelt et al., 2015). It has also been defined as the Chedrina, a subtribe of the tribe Danionini, within the subfamily Danioninae (Liao et al., 2011). The chedrin group has a broad distribution across Africa and Asia, and a combination of molecular and morphological evidence (Tang et al., 2010; Liao et al., 2011, Liao et al., 2012) suggests it comprises 17 genera. Eight of these genera are from Africa (Chelaethiops, Engraulicypris, Leptocypris, Mesobola, Neobola, Opsaridium, Raiamas and Rastrineobol a), while 10 are from Asia (Barilius, Bengala, Cabdio, Luciosoma, Malayochela, Nematobramis, Opsarius, Raiamas, Salmostoma and Securicula; generic names following Eschmeyer et al., 2016). Thus, only one genus, Raiamas, is found in both continents, but notably no species are shared between Africa and Asia. Morphological evidence has resolved monophyly of eight genera (Nematobramis, Luciostoma, Bengala, Opsarius, Raiamas, Salmostoma, Neobola and Chelaethiops) with respect to other representatives of the Danioninae (Liao et al., 2011). Molecular phylogenetic evidence has also repeatedly resolved monophyly of these 17 chedrin genera (Rüber et al., 2007; He et al., 2008; Tang et al., 2010; Liao et al., 2012), with the exception of the placement of Raiamas guttatus in a study by Fang et al. (2009). The chedrin group has also been proposed to include Esomus based on morphological evidence (Liao et al.,
2011, Liao et al., 2012), but the placement is not supported by molecular phylogenetic evidence (Tang et al., 2010).

Notably, molecular phylogenies have resolved African chedrins as monophyletic with respect to the Asiatic chedrins, demonstrating that *Raiamas* is a non-monophyletic group (Tang et al., 2010). Additionally, there have been indications that *Opsaridium* is non-monophyletic (Tang et al., 2010). Here we place the three *Opsaridium* from the Lake Malawi catchment into a time-calibrated phylogeny of African and Asiatic chedrins, including *Opsaridium* from neighbouring catchments. Our results provide further evidence that river faunas assemble through independent immigration events, rather than within-catchment radiation. The results are considered in light of the inferred age of Lake Malawi, and the current taxonomy of *Opsaridium* and *Raiamas*.

2. Methods

2.1 Taxonomic sampling

Our goal was to reconstruct a phylogeny of Lake Malawi catchment *Opsaridium* species within the broader context of the African and Asiatic chedrin species. Thus, we included data for nine of the 12 valid species in *Opsaridium* (Table 1). We used published data for four species, *Opsaridium boweni* (Fowler, 1930), *Opsaridium ubangiense* (Pellegrin, 1901), *Opsaridium zambezense* (Peters, 1852) and *Opsaridium peringueyi* (Gilchrist and Thompson, 1913). We generated new data for a further five species, *Opsaridium leleupi* (Matthes, 1965), *Opsaridium loveridgii* (Norman, 1922), *Opsaridium microlepis*, *Opsaridium microcephalum* and *Opsaridium tweedleorum*. We also generated new sequences for *Opsaridium zambezense*. Samples for these six species were collected from scientific surveys and artisanal fisheries. Three species were not available for study, namely *Opsaridium engrauloides* (Nichols, 1923), *Opsaridium maculicauda* (Pellegrin, 1926) and *Opsaridium splendens* Taverne and De Vos, 1997. Given evidence that some *Raiamas* are closely related to *Opsaridium* (Liao et al., 2012), we included data from nine of the 17 valid species, including the two Asiatic species *Raiamas guttatus* (Day 1870) and *Raiamas bola* (Hamilton, 1822), and seven African species *Raiamas moorii* (Boulenger 1900), *R. salmolucius* (Nichols and Griscom 1917), *R. christyi* (Boulenger 1920), *R. kheeli* Stiassny, Schelly and Schliewen 2006, *R. buchholzi* (Peter 1876), *R. batesii* (Boulenger 1914) and *R. senegalensis* (Steindachner 1870). Eight species were not available.
for study, namely *R. ansorgii* (Boulenger 1910), *R. intermedius* (Boulenger 1915), *R. levequei* Howes and Teugels 1989, *R. longirostris* (Boulenger 1902), *R. nigeriensis* (Daget 1959), *R. scarciensis* Howes and Teugels 1989, *R. shariensis* (Fowler 1949) and *R. steindachneri* (Pellegrin 1908). To place these *Opsaridium* and *Raiamas* data in a broader perspective, we included representatives of other chedrin genera from Africa (*Engraulicypris, Mesobola, Leptocypris, Neobola, Chelaethiops*), and Asia (*Opsarius, Barilius, Salmostoma, Securicula, Cabdio, Lucisoma, Malayochela and Nematobramis*). Two more distantly related cyprinids were employed as outgroups (*Danio* and *Barbus*). Finally, we included two catostomid genera (*Hypentelium* and *Ictiobus*), due to the close relationship of the Cyprinidae and Catostomidae (Near et al., 2012). This enabled the use of a fossil calibration of the earliest representative of the Catostomidae subfamily Ictiobinae to date the divergence between the Ictiobinae and the Catostominae, following Near et al. (2012).

### 2.2 DNA extraction and sequencing

Total genomic DNA was extracted from tissues (muscle and fin clips preserved in 95% EtOH and stored frozen) using the Wizard genomic DNA purification 106 kit 119 (Promega), following the protocol of the manufacturer. We amplified the nuclear recombination-activating gene (*RAG1*) gene using the forward primer 2598F (5’-CCA ACC CCT GCA CAC TCT ACG T-3’) and reverse primer 4067R (5’-TCA AAC GTT TTG GAC TGC C TT GCA TT-3’) from Liao et al. (2012). PCR consisted of an initial denaturation of 95°C for 5 min for *RAG1*, followed by 35 cycles of 95°C for 1 min, annealing at 58°C for 1 min, and extension at 72°C for 2 min, before a final extension step of 72°C for 7 min. We amplified the mitochondrial cytochrome oxidase subunit 1 (*COI*) gene using the forward primer LCO1490 (5’-GGT CAACAA ATC ATA AAG ATA TTG G-3’) and reverse primer HCO2198 (5’-TAA ACT TCA GGG TGA CCA AAA AAT CA-3’) from Folmer et al. (1994). PCR consisted of an initial denaturation of 95°C for 4 min, followed by 35 cycles of 95°C for 30s, annealing at 58°C for 30s, and extension at 72°C for 30s for *COI*, before a final extension step of 72°C for 7 min. Novel *RAG1* and *COI* sequences are listed in Table 1. Purified PCR products were sequenced on ABI3730XL sequencers at Macrogen (Amsterdam, NL). We also used previously published cytochrome b (*cytb*) and rhodopsin (*RH1*) sequences in our phylogenetic analyses (Table 2).

### 2.3 Phylogenetic analyses
Sequences for each gene were aligned using ClustalW in DAMBE 5.3.32 (Xia et al., 2013). Individual alignments for COI, cytb, RAG1 and RH1 consisted of 511, 826, 1308 and 704 nucleotides respectively. PartitionFinder (Lanfear et al., 2012) was used to identify the most likely data partitions for subsequent maximum likelihood (ML) and Bayesian inference (BI) analyses. ML analyses of the concatenated dataset were conducted with RAxML 8.2.0 (Stamatakis et al., 2008) with 100 bootstrap replicates, while BI analysis the concatenated dataset was conducted using a Markov Chain Monte Carlo (MCMC) algorithm in BEAST 1.8.2 (Drummond et al., 2012), with a lognormal relaxed clock model, and a run of 30 million generations recording every 10000 trees. The first 50% of these trees were removed as burn-in. Tracer 1.6 (Rambaut et al., 2013) was used to assess chain convergence to stationarity. Additionally we conducted BI analyses of the COI, cytb, RAG1 and RH1 genes independently, and the COI + RAG1 together (Figs. S1-S5).

For divergence dating in the framework of the BI analysis in BEAST we employed two calibrations from a comprehensive molecular phylogeny of ray-finned fishes by Near et al. (2012), informed by evidence that the family Catostomidae is closely related to the family Cyprinidae. First we calibrated divergence of the catostomid subfamily Ictiobinae (represented by Ictobus) from the subfamily Catostominae (represented by Hypentalium). Following Near et al. (2012) we used the evidence of the first occurrence of stem lineage Ictiobinae dated to 49.4 Ma, employing a lognormal prior with a mean of 0.764 and a standard deviation of 0.8 with 49.4 Ma as the minimum age offset. We set 57.0 Ma as a soft upper bound, based on fossil crown lineage Cypriniformes (see evidence in Near et al. 2012). Second, using the results of Near et al. (2012) we constrained the time of divergence between Cyprinidae and Catostomidae to a normal distribution with a mean divergence time of 98 Ma with a standard deviation of 11 Ma, giving a 95% probability range from 80 to 116 Ma, closely matching the range determined by Near et al. (2012). Note that the derived estimates should be considered approximations, as no internal calibrations are available and the application of calibrations points only at or close to the root is known to pull recent splits towards more ancient age estimates (Renner and Zhang, 2004).

The phylogenetic hypotheses of monophyly of the Lake Malawi Opsaridium, and monophyly of the genus Opsaridium, were tested using Shimodaira-Hasegawa (SH) tests. Maximum likelihood phylogenies with phylogenetic constraints imposed were constructed using RaxML
8.2.0. Site-wise log likelihoods for these trees and the maximum likelihood tree without constraints were then generated in RaxML. SH tests were then conducted in Consel (Shimodaira and Hasegawa, 2001).

2.4 Biogeography and ancestral area reconstruction

To reconstruct ancestral areas and the likelihood of dispersal and vicariance events we used Bayes-Lagrange (Statistical dispersal, extinction-cladogenesis, S-DEC) in RASP 3.2 (Yu et al., 2015), which account for topology, branch lengths and phylogenetic uncertainty. It was selected over other methods as it allowed us to assume dispersal events were only possible between adjacent catchments. We pruned the time-calibrated phylogeny (Fig. 1) to include one representative of each of the species in the “Africa clade”. We then assigned distributions to each of the species using available distribution information (http://www.gbif.org/; Fig. 2). *Opsaridium zambezense* records listed in GBIF (http://www.gbif.org/) from south of the Zambezi, Pungwe and Buzi rivers were not included as these likely represent *O. peringueyi* (Skelton, 1996; Skelton, 2001).

Eight biogeographic areas were determined as: Lake Malawi catchment, Zambezi, Congo, Nile, East Africa, South-East Africa, West Africa and Lower Guinea (Fig. 3). These closely follow the African ichthyofaunal regions, delimited by Roberts (1975), which in turn are strongly influenced by catchment boundaries. Lake Malawi has a single outflow, into the Zambezi, but its fauna was treated as distinct to i) allow an investigation of whether the catchment has been colonised multiple times, and ii) due to the separation of the two systems by a series of rapids and waterfalls that has resulted in two distinct faunas (Tweddle et al., 1979, Tweddle & Willoughby, 1979). The Lake Malawi catchment is adjacent, but not directly connected to, the Congo and East African biogeographic areas, and is geographically distant from the other four areas. The Congo area also included the Lake Rukwa catchment, given evidence of a shared fauna over the last 25,000 years (Cohen et al., 2013). The South East Africa area (Limpopo and Save rivers southwards) was delimited due to apparent faunal separation from the more northerly Zambezi system (including the Buzi and Pungwe rivers). The East Africa area contains mainly rivers flowing east into the Indian Ocean. The Lower Guinea comprised multiple rivers flowing west into the Atlantic Ocean. The Nile system (including Lake Turkana) was treated as distinct from the West Africa region, which included Niger, Upper Guinea rivers and Lake Chad systems.
For the Bayes-Lagrange S-DEC analysis we used 1000 trees randomly sampled from the posterior sample of trees obtained from the BI analysis in BEAST. We assumed an ancestral range to include no more than four areas, given the maximum observed range of species is three areas. We gave no prior range constraints, but we assumed dispersal constraints that allowed only movement between adjacent catchments.

3. Results

3.1 Phylogenetic relationships

The ML phylogeny of the concatenated dataset was topologically similar to the BI consensus tree of the same data (Fig. 1). African chedrins were resolved as a single monophyletic group with an origin 19.97 Ma (95% CI range 14.46 to 26.21 Ma). Our focal Opsaridium genus was resolved as paraphyletic with respect to Raiamas salmolucius and Raiamas moori. This broader “Opsaridium” clade containing R. salmolucius and R. moori was dated to 14.85 Ma (95% CI range 10.35 to 19.26 Ma).

Within the “Opsaridium” clade, all species with multiple specimens were resolved as monophyletic. The Lake Malawi species did not comprise a monophyletic group. Opsaridium microcephalum, an endemic of the Lake Malawi catchment, was resolved as a sister taxon to O. peringueyi found in the eastward flowing rivers of southern Africa south of the Zambezi, with an estimated divergence time of 2.87 Ma (95% CI range 1.46 to 4.63 Ma). Opsaridium microlepis, another endemic of the Lake Malawi catchment, was resolved as a sister taxon of O. loveridgii from the Rufiji-Ruaha basin, with an estimated divergence time of 10.24 Ma (95% CI 5.91 to 14.97 Ma). O. tweddleorum, present only in Lake Malawi and the neighbouring Lower Shire tributary of the Zambezi, was resolved as a sister taxon to O. leleupi from the SE Congo, with an estimated divergence time of 8.65 Ma (95% CI range 4.56 to 13.82 Ma). The phylogenetic hypotheses of monophyly of the Lake Malawi Opsaridium, and monophyly of the genus Opsaridium, were both rejected (SH tests, both $P < 0.001$).

3.2 Ancestral area reconstructions
The results provided a strong indication that three *Opsaridium* present in the Lake Malawi catchment (*O. microcephalum*, *O. microlepis* and *O. tweddleorum*) have colonised the Lake Malawi biogeographic area independently. The most likely scenarios (66.0% probability; Fig 3) were that the common ancestor of *O. microcephalum*, *O. microlepis* and *O. tweddleorum* had a Congo distribution (66.0% probability; Fig 3), or a Malawi-Congo distribution (27.1% probability; Fig 3).

4. Discussion

4.1 Phylogenetic relationships of *Opsaridium* and *Raiamas*

Chedrins are a distinct component of the diverse cyprinid Danioninae subfamily. Within this group, previous work has strongly indicated the monophyly of the African clade, but paraphyly of the genus *Raiamas* with respect other African chedrins (Tang et al., 2010; Liao et al., 2012), and our results were consistent with this analysis. Previous analyses had placed *Opsaridium* as monophyletic [maximum likelihood and Bayesian inference in Liao et al. (2012)], or non-monophyletic with respect to *Chelaethiops bibie* (Tang et al., 2010), and *Raiamas salmolucius* [maximum parsimony in Liao et al. (2012)]. Our analyses have demonstrated clearer evidence of paraphyly of *Opsaridium* with respect to *R. salmolucius* and *R. moorii*.

Howes (1980) proposed twelve morphological characters to diagnose *Opsaridium*. Of these only four were confirmed by Stiassny et al. (2006), including an extended anal fin in mature males, well-developed axial lobes, large and granular tubercles, and a distinctively marked dorsal fin. A fifth character, a large dorsal fin with 13-15 branched rays was not mentioned by Stiassny et al. (2006). Further studies of more species of *Opsaridium* and *Raiamas* by Liao et al. (2012) demonstrated that none of these characters are diagnostic for *Opsaridium*. Our results, demonstrating paraphyly of *Opsaridium* with the inclusion of two *Raiamas* species, thus add weight to conclusions of Liao et al. (2012) that a comprehensive review of the African chedrin taxonomy is required.

4.2. Origins of the Lake Malawi chedrins

The Lake Malawi mpasa (*O. microlepis*) was found to share a most recent common ancestor with *O. loveridgii* from the neighbouring Ruaha-Rufiji system in Tanzania. It was estimated
that the divergence between these species took place on average 10.24 million years ago. The
two catchments are currently separated by the Kipengere/Livingstone mountain range that
originated as part of the western rift during the Late Miocene (15-30 million years ago; Burke
and Gunnell, 2008), with rifting continuing through the Pliocene. Evidence of a shared fish
fauna between the systems comes from the fossils of the Chiwondo beds in northern Malawi,
dated to between 2 and 3.75 Ma (Stewart and Murray, 2013). There, remains of claroteid
catfishes and tigerfish (*Hydrocynus*) are present, which are currently absent from the extant
Lake Malawi catchment fauna, and only occur together in one neighbouring drainage, the
Ruaha (Stewart and Murray, 2013). Molecular evidence from cichlids is also supportive of a
connection between the Malawi and Ruaha catchments between 2 to 7 million years ago
(Genner et al., 2015). It is notable that the area occupied by the ancestral species of *O. microlepis* and *O. loveridgii* was ambiguous, with reconstructed areas including the Congo, Malawi basin and East Coast all possible. On the basis of the combined evidence, it is possible
that the common ancestor of *O. microlepis* and *O. loveridgii* occupied a broader area across
central and eastern Africa.

The Lake Malawi catchment sanjika (*O. microcephalum*) was found to share a most recent
common ancestor with *O. peringueyi*, a species restricted to river systems south of the Zambezi.
The divergence was estimated to have taken place on average 2.87 million years ago. The
reconstructions of the range of common ancestor were ambiguous, with suggested ranges
encompassing the Congo, Malawi, Zambezi and South East coast areas. The apparent absence
of representatives of this clade from the geographically intermediate Zambezi system is
intriguing, particularly given the presence of *O. zambezense* in this system, suggesting suitable
habitat for *Opsaridium* is currently present. It is possible that competition with *O. zambezense*
contributed to the extirpation of the *O. microcephalum - O. peringueyi* lineage from the
Zambezi. Alternatively, or additionally, it is possible that historic events, such as the East
African megadroughts (Cohen et al., 2007; Moore and Eckardt, 2012), have extirpated
populations of the species. In this case it is possible that Lake Malawi may have acted as a
refuge for the *O. microcephalum* lineage during drought periods, particularly given evidence
the species is capable of spawning in fully lacustrine conditions (Tweddle & Turner 2014).

Dwarf sanjika (*O. tweddleorum*) is present in both Lake Malawi catchment and tributaries of
the Lower Shire River that is part of the Zambezi catchment. The species was found to share
common ancestry with *Opsaridium leleupi*, known only from the Upper Lualaba river (Fig 2),
with an estimated divergence time of 8.65 million years ago. Again, the ancestral area
reconstructions of this species were ambiguous, with the Congo, Zambezi and Lake Malawi
areas variously supported. It is likely the ancestral species had a broader distribution, but also
the apparent absence of representatives of the clade in geographically intermediate Zambezi
system is notable. This is again suggestive of historic extirpation of common ancestors from
Zambezi system, and that Lake Malawi may have acted as refuge during drought periods, in
turn allowing allopatric diversification.

It seems likely that the relationships within the “Opsaridium” clade will be further clarified
through sampling species not studied here, all of which have strongly restricted distributions
within tributaries of the Congo system. *Opsaridium engrauloides* is a species known only from
one type specimen of the Ubangui River at Bangui, Central African Republic. *Opsaridium
maculicauda* is known only from type specimens collected at Tshikapa in the Upper Kasai
system, Democratic Republic of the Congo. *Opsaridium splendens* is distributed in the
Malagarasi system of Tanzania and Burundi (Fig. 2). These species could be particularly
important given evidence that other fish species may have colonised the Malawi catchment
from the Congo system, possibly via the Luangwa river (Tweddle and Skelton, 2008; Egger et
al., 2015).

4.3. Origins of the “Opsaridium” clade.

Our analyses of ancestral areas revealed the Congo Basin as the most likely origin of much of
the “Opsaridium” clade (Fig. 3). The Congo is the most ancient drainage basin on the continent
with an age of at least 65 million years, although the present geographic structure of the basin
dates to ~5 million years ago (Stankiewicz and de Wit, 2006). The substantial catchment (ca.
3.8 mil km) will have ensured flow in downstream habitats even during regional droughts (e.g.
Dalibard et al., 2014). A recent study estimated that the basin harbours more than 1,250 species
of freshwater fish (Snoeks et al., 2011), so this stability may have helped to supporting high
phylogenetic diversity and species richness of riverine species present. However, whether the
basin has acted as a source of diversity for other regions of Africa is more equivocal, as
although several studies have considered historical phylogeography of African freshwater
fishes across broad spatial scales (e.g. Koblmüller et al., 2008b; Goodier et al., 2011; Brown et
al., 2010), few have reconstructed ancestral areas. Among the exceptions are studies of
*Synodontis* catfishes that have attributed basal nodes of the genus primarily to the West Africa,
the Nilo-Sudan or Congo regions, whereas the East African and Southern African taxa appear to be more derived (Day et al., 2013; Pinton et al., 2013). Notably, however, there was also evidence of Congo drainage species being derived from East and Southern African clades (Day et al., 2013; Pinton et al., 2013).

4.4. Assembly of the Lake Malawi catchment fauna

Over large spatial scales species richness within river systems is related to key environmental variables, including the catchment area, the energy input into the system, and the historical stability of the catchment (Oberdorff et al., 2011), including the frequency of river capture events. These factors in turn will affect the likelihood of the different processes that affect community assembly, most importantly immigration, speciation and extinction. Our study supports the hypothesis that river assemblages within the Lake Malawi catchment have formed through a process of immigration and within-catchment adaptation, rather than within-catchment speciation. This reinforces the importance of natural barriers within and across river networks in preventing gene flow, and allowing allopatric adaptive divergence to take place. The data are also supportive of periodic between-catchment dispersal events across African watersheds, which are often characterised by flat wetlands that periodically flood and allow fish movement between drainage systems (Beadle, 1974). The underlying reasons why most of the riverine fish lineages in the Lake Malawi catchment have failed to undergo speciation and adaptive radiation within these river systems remains unclear. One possibility is that inherent environmental instability in shallow river systems that surround large lakes favours generalist strategies and high dispersal that prevent the local adaptation and assortative mating that often accompany speciation in fishes.

5. Summary

A combined analysis of mitochondrial and nuclear sequence data revealed that the genus *Opsaridium* is not a monophyletic clade, but instead includes representatives of the genus *Raiamas* from the broader Congo catchment, including Lake Rukwa and Lake Tanganyika. The results also suggest that the three *Opsaridium* species in Lake Malawi catchment arrived from multiple independent colonisation events, as opposed to radiating within the catchment. Much of the diversity within the *Opsaridium* group is found within the Congo region that appears to have been an important source for diversity in the Zambezi system, including Lake
Malawi. Further resolution of the phylogeny of the African chedrin clade will help to improve our understanding of the role of Congo system in acting as a source of diversity within East and Southern Africa.

Acknowledgements

The work was supported by a Commonwealth Scholarship MWCS-2012-249 to HS, and Royal Society-Leverhulme Trust Africa Awards AA100023 and AA130107 to MG and BPN. We thank A. Smith, G. Turner, J. Swanstrom, C. Katongo, S. Mzi ghani and TAFIRI researchers for assistance with sample collection. We thank G. Orti and two reviewers for helpful comments.

References


Figure Legends

Fig. 1. Time calibrated phylogeny of chedrins inferred from Bayesian analysis of mitochondrial and nuclear DNA sequences. Blue node bars represent the 95% highest posterior density (HPD) intervals of inferred divergence time estimates. Values at nodes indicate Bayesian posterior probabilities and maximum likelihood bootstrap percentages, respectively. Underlined taxa are from the Lake Malawi catchment.

Fig. 2. Distribution of species within the “Opsaridium” clade across Africa. Data sourced from GBIF (http://www.gbif.org/), downloaded 31 July 2015. Records conflicting with distributional information of O. zambezense in Skelton (1996) and Skelton (2001) were removed.

Fig. 3. Ancestral area reconstruction using the Bayes Lagrange Statistical Dispersal-Extinction-Cladogenesis (S-DEC) analyses as implemented in RASP (Yu et al. 2015). Ancestral distribution range with a low percentage likelihood (<15%) are merged for each node and coloured black. Coloured circles indicate the percentage probability of area occupancy. Multiple colours in a segment indicate a distribution shared between areas.
Fig. 2
Fig. 3

Divergence time (Ma)
Table 1: Collection information (coordinates in degrees and decimal minutes) and Genbank accessions for specimens newly sequenced for this study. n/a indicates no sequence was available.

<table>
<thead>
<tr>
<th>Species</th>
<th>Location</th>
<th>Catchment</th>
<th>Latitude (S)</th>
<th>Longitude (E)</th>
<th>Date</th>
<th>COI Accession</th>
<th>RAG1 Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opsaridium leleupi</td>
<td>Kisanfu River (a) Congo</td>
<td></td>
<td>10°45.840'</td>
<td>25°57.840'</td>
<td>09-09-2008*</td>
<td>KX810223</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Kisanfu River (b) Kicrossing</td>
<td>Congo</td>
<td>10°47.970'</td>
<td>25°58.859'</td>
<td>09-12-2008*</td>
<td>KX810224</td>
<td>n/a</td>
</tr>
<tr>
<td>Opsaridium loveridgii</td>
<td>Mngeta (a) Rufiji (Kilombero)</td>
<td></td>
<td>08°20.457'</td>
<td>36°05.227'</td>
<td>07-03-2014</td>
<td>KX810212</td>
<td>KX810226</td>
</tr>
<tr>
<td></td>
<td>Mngeta (b) Rufiji (Kilombero)</td>
<td></td>
<td>08°20.457'</td>
<td>36°05.227'</td>
<td>07-03-2014</td>
<td>KX810213</td>
<td>n/a</td>
</tr>
<tr>
<td>Opsaridium microcephalum</td>
<td>Malombe Lake Malawi</td>
<td></td>
<td>12°30.807'</td>
<td>34°11.755'</td>
<td>25-01-2014</td>
<td>KX810216</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Bua Lake Malawi</td>
<td></td>
<td>12°47.194'</td>
<td>34°11.755'</td>
<td>25-01-2014</td>
<td>KU941767</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Dwangwa Lake Malawi</td>
<td></td>
<td>12°30.807'</td>
<td>34°11.755'</td>
<td>25-01-2014</td>
<td>KU941691</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Linthipe Lake Malawi</td>
<td></td>
<td>13°47.333'</td>
<td>34°26.088'</td>
<td>22-01-2014</td>
<td>KU941590</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Kamuzu Dam Lake Malawi</td>
<td></td>
<td>14°10.465'</td>
<td>33°38.566'</td>
<td>23-08-2013</td>
<td>KX810217</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>North Rukuru Lake Malawi</td>
<td></td>
<td>09°55.032'</td>
<td>33°55.676'</td>
<td>28-01-2014</td>
<td>KU941673</td>
<td>KX810228</td>
</tr>
<tr>
<td>Opsaridium microlepis</td>
<td>Malombe Lake Malawi</td>
<td></td>
<td>12°30.807'</td>
<td>34°06.971'</td>
<td>25-01-2014</td>
<td>KU941767</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Dwangwa Lake Malawi</td>
<td></td>
<td>12°47.194'</td>
<td>34°11.755'</td>
<td>25-01-2014</td>
<td>KU941712</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Bua Lake Malawi</td>
<td></td>
<td>12°47.194'</td>
<td>34°11.755'</td>
<td>25-01-2014</td>
<td>KU941767</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>North Rukuru Lake Malawi</td>
<td></td>
<td>09°55.032'</td>
<td>33°55.676'</td>
<td>28-01-2014</td>
<td>KU941758</td>
<td>KX810230</td>
</tr>
<tr>
<td></td>
<td>Linthipe Lake Malawi</td>
<td></td>
<td>13°47.333'</td>
<td>34°26.088'</td>
<td>22-01-2014</td>
<td>KU941691</td>
<td>KX810229</td>
</tr>
<tr>
<td>Opsaridium tweddleorum</td>
<td>Mtakataka Lake Malawi</td>
<td></td>
<td>13°47.333'</td>
<td>34°26.088'</td>
<td>22-01-2014</td>
<td>KU941691</td>
<td>KX810229</td>
</tr>
<tr>
<td></td>
<td>Khuyu Lake Malawi</td>
<td></td>
<td>12°16.735'</td>
<td>33°59.939'</td>
<td>26-01-2014</td>
<td>KU941806</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Ruhuhu Lake Malawi</td>
<td></td>
<td>10°37.603'</td>
<td>35°39.192'</td>
<td>09-09-2012</td>
<td>KU941831</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>North Rumpfi Lake Malawi</td>
<td></td>
<td>10°41.109'</td>
<td>34°10.883'</td>
<td>27-01-2014</td>
<td>KU941793</td>
<td>KX810236</td>
</tr>
<tr>
<td></td>
<td>Lumbadzi Lake Malawi</td>
<td></td>
<td>13°47.640'</td>
<td>33°59.300'</td>
<td>23-08-2013</td>
<td>n/a</td>
<td>KX810235</td>
</tr>
<tr>
<td></td>
<td>Lingadzi Lake Malawi</td>
<td></td>
<td>13°32.061'</td>
<td>34°17.064'</td>
<td>23-01-2014</td>
<td>n/a</td>
<td>KX810234</td>
</tr>
<tr>
<td>Opsaridium zambezense</td>
<td>Lutembe (a) Zambezi (Luangwa)</td>
<td></td>
<td>13°42.383'</td>
<td>32°29.366'</td>
<td>21-06-2010</td>
<td>KX810214</td>
<td>KX810227</td>
</tr>
<tr>
<td></td>
<td>Lutembe (b) Zambezi (Luangwa)</td>
<td></td>
<td>13°42.383'</td>
<td>32°29.366'</td>
<td>21-06-2010</td>
<td>KX810215</td>
<td>n/a</td>
</tr>
<tr>
<td>Raiamas moorii</td>
<td>Kavimira River (a) Lake Tangan’ika</td>
<td></td>
<td>03°21.156'</td>
<td>29°09.054'</td>
<td>10-09-2013</td>
<td>KX810221</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Kavimira River (b) Lake Tangan’ika</td>
<td></td>
<td>03°21.156'</td>
<td>29°09.054'</td>
<td>10-09-2013</td>
<td>KX810220</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>Kalimabenge River (a) Lake Tangan’ika</td>
<td></td>
<td>03°24.434'</td>
<td>29°07.585'</td>
<td>10-09-2013</td>
<td>KX810222</td>
<td>KX810233</td>
</tr>
<tr>
<td></td>
<td>Kalimabenge River (b) Lake Tangan’ika</td>
<td></td>
<td>03°24.434'</td>
<td>29°07.585'</td>
<td>10-09-2013</td>
<td>KX810219</td>
<td>KX810231</td>
</tr>
<tr>
<td></td>
<td>Lake Rukwa Lake Rukwa</td>
<td></td>
<td>08°23.850</td>
<td>32°54.110'</td>
<td>01-09-2012</td>
<td>n/a</td>
<td>KX810232</td>
</tr>
<tr>
<td>Engraulicypris sardella</td>
<td>Lake Malawi Lake Malawi</td>
<td></td>
<td>14°04.172'</td>
<td>34°55.034'</td>
<td>10-01-2015</td>
<td>KX810225</td>
<td>TBC</td>
</tr>
</tbody>
</table>

*collectors: K. Magellan, J. Harvey, G. Winnaar
Table 2: Taxa, voucher specimens, and accession numbers for the gene sequences from GenBank used in this study. n/a indicates no sequence was available.

<table>
<thead>
<tr>
<th>Taxon group</th>
<th>Species</th>
<th>COI</th>
<th>cytb</th>
<th>RAG1</th>
<th>RH1</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chelaethiops cf. congicus</td>
<td>HM418170</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>International Barcode of Life submission</td>
</tr>
<tr>
<td></td>
<td>Chelaethiops congicus</td>
<td>HM880211</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>International Barcode of Life submission</td>
</tr>
<tr>
<td></td>
<td>Chelaethiops elongatus</td>
<td>JX196996</td>
<td>JX197006</td>
<td>JX197014</td>
<td>JX197021</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Chelaethiops elongatus</td>
<td>JX196996</td>
<td>JX197021</td>
<td>JX197014</td>
<td>JX197006</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Engraulicypris sardella</td>
<td>JX196997</td>
<td>JX197022</td>
<td>JX197015</td>
<td>JX197007</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Leptocypris lujae</td>
<td>KT193306</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>Decru et al. unpublished</td>
</tr>
<tr>
<td></td>
<td>Leptocypris sp.</td>
<td>n/a</td>
<td>n/a</td>
<td>HM224052</td>
<td>HM223933</td>
<td>Tang et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Leptocypris weeksi</td>
<td>HM418175</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>International Barcode of Life submission</td>
</tr>
<tr>
<td></td>
<td>Leptocypris weynsi</td>
<td>HM418188</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>International Barcode of Life submission</td>
</tr>
<tr>
<td></td>
<td>Mesobola brevianalis</td>
<td>HM224176</td>
<td>HM224295</td>
<td>HM224054</td>
<td>HM223935</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Neobola bottegoi</td>
<td>HM224178</td>
<td>HM224296</td>
<td>HM224056</td>
<td>HM223936</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Opsaridium boweni</td>
<td>JX197000</td>
<td>JX197009</td>
<td>JX197016</td>
<td>JX197023</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Opsaridium peringueyi</td>
<td>HM224192</td>
<td>HM224311</td>
<td>HM224072</td>
<td>HM223954</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Opsaridium zambezense</td>
<td>HM224194</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Rafflesia batesii</td>
<td>JX197002</td>
<td>JX197010</td>
<td>JX197017</td>
<td>JX197024</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Rafflesia buchholzi</td>
<td>HM224213</td>
<td>HM224330</td>
<td>HM224090</td>
<td>HM223975</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Rafflesia cf. christyi</td>
<td>HM418152</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>International Barcode of Life submission</td>
</tr>
<tr>
<td></td>
<td>Rafflesia cf. christyi</td>
<td>KT193013</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>Decru et al. unpublished</td>
</tr>
<tr>
<td></td>
<td>Rafflesia christyi</td>
<td>JX197003</td>
<td>JX197011</td>
<td>JX197018</td>
<td>JX197025</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Rafflesia khebei</td>
<td>HM418197</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>International Barcode of Life submission</td>
</tr>
<tr>
<td></td>
<td>Rafflesia salmolucia</td>
<td>JX197004</td>
<td>JX197012</td>
<td>JX197019</td>
<td>JX197026</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Rafflesia senegalensis</td>
<td>HM224215</td>
<td>HM224332</td>
<td>HM224093</td>
<td>HM223978</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Rastrineobola argentea</td>
<td>JX197005</td>
<td>n/a</td>
<td>JX197020</td>
<td>JX197027</td>
<td>Liao et al. 2012</td>
</tr>
<tr>
<td>Asiatic chedrins</td>
<td>Barilius barila</td>
<td>HM224138</td>
<td>HM224257</td>
<td>HM224020</td>
<td>HM223900</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Barilius vagra</td>
<td>HM224140</td>
<td>HM224259</td>
<td>HM224022</td>
<td>HM223901</td>
<td>Tang et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Cabdio morar</td>
<td>AP011335</td>
<td>AP011335</td>
<td>FJ531343</td>
<td>EU711105</td>
<td>Tang et al. 2010, Chen &amp; Mayden 2009</td>
</tr>
<tr>
<td>Species</td>
<td>Accession Numbers</td>
<td>References</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------</td>
<td>------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Opsarius bakeri</em></td>
<td>HM224197, HM224315, HM224076, HM223959</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Opsarius barnoides</em></td>
<td>HM224199, HM224317, HM224078, HM223961</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Opsarius canarensis</em></td>
<td>HM224200, HM224318, HM224079, HM223962</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Raiamas bola</em></td>
<td>HM224212, HM224329, HM224089, HM223974</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Raiamas guttatus</em></td>
<td>HM224214, HM224331, HM224091, HM223976</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Raiamas sp. “Myanmar”</em></td>
<td>HM224217, HM224334, HM224095, HM223980</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Salmostoma bacaila</em></td>
<td>AP011223, AP011223, HM224128, HM224010</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Salmostoma phulo</em></td>
<td>HM224248, HM224379, HM224133, HM224013</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Securicula gora</em></td>
<td>HM224250, HM224381, HM224135, HM224015</td>
<td>Tang et al. 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outgroups</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Danio rerio</em></td>
<td>AC024175, AC024175, U71093, AF105152</td>
<td>Milam et al. unpublished, Willett et al. 2007, Robinson et al. 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hypentelium nigricans</em></td>
<td>KF29987, n/a, EU711134, n/a</td>
<td>Bentley and Wiley unpublished, Mayden et al. 2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ictiobus bubalus</em></td>
<td>KF29996, n/a, EF056353, n/a</td>
<td>Bentley and Wiley unpublished, Slechtova et al. 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>