Molecular Phylogenetics and Evolution 136 (2019) 215–226

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Molecular phylogeny of *Oreochromis* (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple colonisation of adverse aquatic environments

Antonia G.P. Ford[a,*], Thomas R. Bullen[a], Longson Pang[a], Martin J. Genner[b], Roger Bills[c], Tomáš Flouri[a], Benjamin P. Ngatunga[d], Lukas Rüber[e,f], Ulrich K. Schliewen[g], Ole Seehausen[f,h], Asilatu Shechonge[d,i], Melanie L.J. Stiassny[j], George F. Turner[k], Julia J. Day[a,*]

[a] Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK
[b] School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK
[c] South African Institute for Aquatic Biodiversity, Private Bag 1015, 6140 Grahamstown, South Africa
[d] Tanzania Fisheries Research Institute (TAFIRI), P.O. Box 9750, Dar es Salaam, Tanzania
[e] Naturhistorisches Museum Bern, Bernastrasse 15, 3005 Bern, Switzerland
[f] Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland
[g] Bavarian State Collection of Zoology, Department of Ichthyology, Münchhausenstr. 21, 81247 München, Germany
[h] Department of Fish Ecology and Evolution, Center for Ecology, Evolution and Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland
[i] Department of Aquatic Sciences and Fisheries, University of Dar es Salaam, P.O. Box 35064, Dar es Salaam, Tanzania
[j] Department of Ichthyology, American Museum of Natural History, New York, NY, USA
[k] School of Biological Sciences, Bangor University, Bangor LL57 2UW, UK

ARTICLE INFO

Keywords:
Tilapia
Alcolapia
Ancestral state reconstruction
Mito-nuclear discordance
Introgression
Incomplete lineage sorting
Taxonomic revision

ABSTRACT

Although the majority of cichlid diversity occurs in the African Great Lakes, these fish have also diversified across the African continent. Such continental radiations, occurring in both rivers and lakes have received far less attention than lacustrine radiations despite some members, such as the oreochromine cichlids (commonly referred to as 'tilapia'), having significant scientific and socio-economic importance both within and beyond their native range. Unique among cichlids, several species of the genus *Oreochromis* exhibit adaptation to soda conditions (including tolerance to elevated temperatures and salinity), which are of interest from evolutionary biology research and aquaculture perspectives. Questions remain regarding the factors facilitating the diversification of this group, which to date have not been addressed within a phylogenetic framework. Here we present the first comprehensive (32/37 described species) multi-marker molecular phylogeny of *Oreochromis* and closely related *Alcolapia*, based on mitochondrial (1583 bp) and nuclear (3092 bp) sequence data. We show widespread discordance between nuclear DNA and mitochondrial DNA trees. This could be the result of incomplete lineage sorting and/or introgression in mitochondrial loci, although we did not find a strong signal for the latter. Based on our nuclear phylogeny we demonstrate that adaptation to adverse conditions (elevated salinity, temperature, or alkalinity) has occurred multiple times within *Oreochromis*, but that adaptation to extreme (soda) conditions (high salinity, temperature, and alkalinity) has likely arisen once in the lineage leading to *O. amphimelas* and *Alcolapia*. We also show *Alcolapia* is nested within *Oreochromis*, which is in agreement with previous studies, and here revise the taxonomy to synonymise the genus in *Oreochromis*, retaining the designation as subgenus *Oreochromis* (*Alcolapia*).

1. Introduction

The propensity for African cichlid fishes to form adaptive radiations within lacustrine environments has received considerable research attention (Turner, 2007; Seehausen, 2015), whereas the processes promoting diversification within largely riverine lineages are less well

* Corresponding authors at: Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK (J.J. Day). Department of Life Sciences, Whitelands College, University of Roehampton, Holybourne Avenue, London SW15 4JD, UK (A.G.P. Ford).
E-mail addresses: antonia.ford@roehampton.ac.uk (A.G.P. Ford), ",0,0,2
>j.day@ucl.ac.uk (J.J. Day).

https://doi.org/10.1016/j.ympev.2019.04.008
Received 7 August 2018; Received in revised form 26 March 2019; Accepted 6 April 2019
Available online 08 April 2019
1055-7903/

Table 1
Alcolapia and *Oreochromis* adaptation to extreme environmental conditions.

Species	Natural distribution	Temp°C	Salinity ‰	pH	Ref.
A. alcalica (Hilgendorf 1905)	Natron, TZA	20–42	> 40	> 10	2,3
A. grahami (Boulenger 1912)	Magadi, Nakuru, KEN	20–42	> 40	> 10	1,2,3
A. latilabris (Seegers & Tichy 1999)	Natron, TZA	20–42	> 40	> 10	3
A. ndalalani (Seegers & Tichy 1999)	Natron, TZA	20–42	> 40	> 10	3
O. amphimelas (Hilgendorf 1905)	Soda lakes, TZA	20–30	58	> 9	1,5
O. andersonii (Castelnau 1861)	South-central Africa	18–33	20	nd	1
O. angolensis (Trewavas 1973)	Southern Africa	nd	nd	nd	2
O. aureus (Steindachner 1864)	Eurasia, Africa,	12–32	45	nd	1,2
O. chungruruensis (Ahl 1924)	Chungruru, TZA	nd	freshwater	nd	2
O. esculentus (Graham 1928)	Nile, East African Lakes	23–29	freshwater	nd	1,2
O. hunteri (Günther 1889)	Chala, TZA	nd	nd	nd	2
O. ismailiaensis (Mekkaway 1995)	EGY	nd	nd	nd	
O. jipe (Lowe 1955)	Jipe, Pangani, TZA	nd	nd	nd	2
O. karomo (Poll 1948)	Tanganyika, E. Africa	nd	nd	nd	2
O. karongae (Trewavas 1941)	Malawi, E. Africa	nd	nd	nd	2
O. korogwe (Lowe 1955)	Eastern Africa	nd	freshwater	nd	2
O. lepidurus (Boulenger 1899)	Central Africa	nd	nd	nd	2
O. leucostictus (Trewavas 1933)	Edward, George, UGA	15–38	freshwater	7–9	2
O. lidole (Trewavas 1941)	Malawi, Chungruru, TZA	nd	nd	nd	2
O. macrochir (Boulenger 1912)	S. Africa	18–32	20	nd	1,2
O. malagarasi (Trewavas 1983)	Eastern Africa	nd	nd	nd	2
O. mortimeri (Trewavas 1966)	Southern Africa	19–32	26	nd	1,2
O. mossambicus (Peters 1852)	SE Africa, widely introduced	17–35	> 100	nd	1
O. mweruensis (Trewavas 1983)	Congo River system	nd	nd	nd	2
O. niloticus sp 'Bogoria'	Lake Bogoria, KEN	36	nd	7	4
O. niloticus baringoensis (Trewavas 1983)	Baringo, KEN	nd	nd	nd	2
O. niloticus cancellatus (Nichols 1923)	Awash Basin, ETH	17–26	nd	nd	2
O. niloticus eduardianus (Boulenger 1912)	Edward, UGA	nd	nd	nd	2
O. niloticus filoa (Trewavas 1983)	Hot springs, Awash, ETH	32–39	nd	nd	2
O. niloticus niloticus (Linnaeus 1758)	NE Africa	14–32	30	nd	1
O. niloticus sugutae (Daget 1991)	Karpeddo soda springs, Suguta system, KEN	33–38	nd	nd	2
O. niloticus tana (Seyoum 1992)	Lake Tana, ETH	nd	nd	nd	2
O. niloticus vulcani (Trewavas 1933)	Crater lake, Turkana, KEN	nd	nd	nd	2
O. placidus placidus (Trewavas 1941)	Southeastern Africa	nd	freshwater	nd	2
O. placidus ruvumae (Trewavas 1966)	Ruvuma, SE Africa	nd	nd	nd	2
O. rukwaensis (Hilgendorf 1903)	Lake Rukwa, TZA	nd	nd	nd	2
O. saka (Lowe 1953)	Lake Malawi, East Africa	nd	nd	nd	2
O. salinicola (Poll 1948)	Central Africa	nd	25–35	nd	2
O. schwebischi (Sauvage 1884)	West-Central Africa	nd	nd	nd	2
O. shiranus chilwae (Trewavas1966)	Lake Chilwa, MWI	21–37	30	nd	1,2
O. shiranus shiranus (Boulenger 1896)	Lake Malawi and drainage	nd	nd	nd	2
O. spilurus niger (Günther 1894)	Kibwezi River, KEN	19–32	nd	nd	1,2
O. spilurus percivali (Boulenger 1912)	Hot springs, KEN	20–38	'alkaline'	nd	1, 2
O. spilurus spilurus (Günther 1894)	KEN	20–31	nd	nd	2
O. squamipinnis (Günther 1864)	Lake Malawi	nd	nd	nd	2
O. tanganicae (Günther 1893)	Lake Tanganyika	nd	nd	nd	2
O. upembae (Thys van den Audenaerde 1964)	Congo river basin	nd	nd	nd	2
O. urolepis urolepis (Norman 1922)	TNZ	25–38	> 35	8.4	2
O. variabilis (Boulenger 1906)	Lake Victoria and drainage	23–28	nd	nd	1,2

Taxa in bold indicate samples included in the present study (see Table S1). Temperatures and salinity/alkalinity conditions are the maximum at which the species naturally occur. Several studies have shown species are able to tolerate/survive higher levels in laboratory conditions (though fewer have explored successful reproduction at extreme conditions). For the present study, temperature tolerance > 35 °C, salinity tolerance > 30‰, and pH tolerance > pH 9 were considered to represent elevated environmental tolerance; and tolerance to all three elevated parameters to indicate soda adaptation. See Table S3 for details of how species were coded for ancestral reconstruction analyses. References: 1. Philippart and Ruwet 1982; 2. Trewavas 1983; 3. Seegers et al., 1999. 4. Ndiwa et al., 2014. 5. Present study.
nd: minimal data indicating elevated tolerance, presumed not to occur in soda conditions.

known. One of the most species rich and widely distributed lineages of African cichlids is the oreochromine group. Oreochromine cichlids have significant scientific and socio-economic importance both within and beyond their native range – providing a major food source for fisheries and aquaculture, a biocontrol agent for aquatic plants and, together with other cichlid groups, they are a 'model' system for evolutionary research (Kobayashi et al., 2015; Yue et al., 2016; Brawand et al., 2014). However, the phylogenetic relationships of the species from the most diverse group – *Oreochromis* Günther 1889 – are poorly understood.

As defined by Dunz and Schliewen (2013), the tribe Oreochromini is comprised of the mouthbrooding lineages formerly assigned to the tilapiine tribe: *Alcolapia, Danakilia, Iranocichla, Oreochromis, Sarotherodon, Tristramella,* along with four *Sarotherodon*-derived genera endemic to the Cameroonian crater Lake Barombi Mbo (*Konia, Myaka, Pungu, Stomatepia*). The most species rich genus, *Oreochromis*, is found throughout Sub-Saharan Africa, as well as the Nile basin and Middle East (Trewavas, 1983). The Oreochromini largely occur in rivers, floodplains and shallow lakes (75% of species occur in rivers, or rivers and lakes; Trewavas, 1983) and while members occur in every larger lake of Africa (and in many small ones), they have only formed low diversity radiations in very few lakes (Lowe-McConnell, 1959; Trewavas, 1982; Klett and Meyer, 2002; Seehausen, 2007), and these usually are lakes that lack haplochromine cichlids (Seehausen, 2007).

Oreochromine cichlids are attractive targets for aquaculture, with at least eight species actively farmed globally. The Nile tilapia

(*Oreochromis niloticus*) is one of the most widely farmed aquaculture species globally, but other species of *Oreochromis* are farmed in many local regions on a smaller scale, and are also important capture fisheries species (FAO, 2016). However, such aquaculture and capture fishery improvement initiatives can have significant environmental impact where fish escape and establish local populations. Many species of *Oreochromis* have a substantial invasive potential, exhibiting trophic plasticity that enables broad resource use, and a tendency to hybridise (Genner et al., 2013; Shechonge et al., 2019), and now four species of *Oreochromis* are listed on the Global Invasive Species Index (IUCN, 2018). Despite certain species exhibiting strong invasive potential, some species within the genus are severely range restricted and thus are threatened, with 19 species (50%) categorised as "Near Threatened to Critically Endangered", of which seven species are listed as Critically Endangered (IUCN Red List, v. 2017/3).

Although the higher-level taxonomic categories of cichlids formerly referred to as "Tilapia" have received recent attention (e.g., Dunz and Schliewen, 2013), questions remain regarding the factors facilitating the diversification within *Oreochromis*. However, previous phylogenetic studies included limited taxonomic coverage of *Oreochromis*, and delivered conflicting results. These have been largely based on mtDNA markers, with the exception of Schwarzer et al. (2009), Dunz and Schliewen (2013), and Matschiner et al. (2017) who included nuclear loci, but only a maximum of four species of *Oreochromis* regarding the earlier studies, and seven species for the latter study. A recent study investigating actinopterygian relationships (Rabosky et al., 2018), included 17 species of *Oreochromis* and two species of *Alcolapia* for both mitochondrial and nuclear data (expanding slightly on Rabosky et al., 2013), and while all taxa in this study included ND2, inclusion of nuclear data was very limited, with only *O. niloticus* and *O. tanganicae* having reasonable coverage.

The soda lake cichlid species in the genus *Alcolapia* were originally described as a member of *Tilapia* and subsequently included by Trewavas (1983) as subspecies within a subgenus of *Oreochromis,* as *O. (Alcolapia) alcalicus alcalicus* and *O. (Alcolapia) alcalicus grahami*. Seegers et al. (1999) elevated the subgenus *Alcolapia* to genus based on mtDNA sequence data, and later revised the species name *alcalicus* to *alcalica* to agree with the feminine genus (Seegers, 2008a, 2008b). However, while molecular analyses have consistently resolved a monophyletic *Alcolapia*, that is shown to nest within *Oreochromis*, the specific relationship to *Oreochromis* is unresolved, with various taxa having been hypothesised to be the sister species to the *Alcolapia* clade: *O. amphimelas* (Seegers et al., 1999; Nagl et al., 2001), *O. malagarasi* (Seegers et al., 1999), *O. tanganicae* (Schwarzer et al., 2009; Dunz and Schliewen, 2013) and *O. variabilis* (Kavembe et al., 2013; Matschiner et al., 2017; Rabosky et al., 2018). Other than *O. amphimelas,* the other three species previously suggested as alternative sister taxa to *Alcolapia* (i.e. *O. malagarasi*, *O. tanganicae, O. variabilis*) are not found in soda lake conditions. Based on the lack of a densely sampled phylogeny for *Oreochromis* and *Alcolapia*, we generated a comprehensive phylogeny for the group using multiple markers from the mitochondrial and nuclear genomes with near-complete coverage of sampled loci.

1.1. Adaptation to adverse environments

Of the 44 currently recognised species and subspecies of *Oreochromis*, excluding *Alcolapia*, nine are known to tolerate or occur in environments with elevated salinity or temperature, but only *O. amphimelas* and *Alcolapia* have adapted to the high alkalinity and low dissolved oxygen levels found in soda lake conditions (see Table 1). Tolerance to temperature and salinity are intrinsically linked, with lower temperature ranges tolerated in saline conditions than in freshwater in some species of tilapia (reviewed in Philippart and Ruwet, 1982), and these two parameters are considered to be the determining factors in the distribution of several species. Several species of *Oreochromis* are euryhaline and acclimatise to a range of levels of salinity (from freshwater through to seawater), including *O. urolepis* and *O. mossambicus* that naturally occur in estuarine conditions and have been widely used in aquaculture due to their salinity tolerance (e.g., Riedel and Costa-Pierce, 2005; Sardella and Brauner, 2008; Ulotu et al., 2016). The typically freshwater *O. niloticus* (composed of seven subspecies) includes members that have successfully colonised thermal hot springs (Bezault et al., 2007; Ndiwa et al., 2014), while other *Oreochromis* species thrive in the high-pH volcanic springs feeding the soda lakes of East Africa (Trewavas, 1983). In particular, *O. amphimelas* occurs in the springs and main water bodies of the seasonal soda-like Lakes Manyara, Eyasi, Sulungali, Kitangiri, and Singida, Tanzania. This species co-occurs with introduced populations of *O. niloticus* in the Sulungali, Kitangiri and Singida lakes and alongside introduced populations of *O. esculentus* in Kitangiri and Singida. The genus *Alcolapia* has diversified in the even more extreme environment of the nearby soda lakes: Lakes Natron in Tanzania and Magadi in Kenya (Seegers et al., 1999; Ford et al., 2015; 2016) and occurs in alkaline hydrothermal springs feeding the lakes. These hot, hyper saline and highly alkaline soda springs represent some of the most extreme aquatic environments known supporting fish life. It is unclear whether adaptation to soda waters occurred early in the evolution of this lineage or arose more recently; and whether it arose only once or multiple times. Resolving the species relationships of the genera *Oreochromis* and *Alcolapia* will enable these hypotheses to be tested.

Based on our comprehensive multi-marker phylogeny, we examine the taxonomic status of the genus *Alcolapia* to test whether *Alcolapia* and *Oreochromis* are reciprocally monophyletic. We also test for multiple origins of a) tolerance to increased salinity and temperature typical of soda conditions, and b) two male secondary sexual characteristics: extended jaws and the genital tassel (long bifid filaments that develop in the breeding season on the genital papillae), that have historically been used to delimit subgenera of *Oreochromis*.

2. Materials and methods

2.1. Species sampling

A total of 28 species from 33 described species of *Oreochromis,* as well as all four species of *Alcolapia* (using taxonomic data compiled by Eschmeyer et al. (2018)), are included in this study. Multiple wild samples of each species, and where possible subspecies, from across Africa were included totalling 105 samples (Appendix A, Table S1). Specimens were collected mainly using seine nets, with specimens euthanised with an overdose of clove oil and subsequently preserved in 70–80% ethanol. Fin-clips or muscle tissue were taken for molecular analysis and were preserved in 96–100% ethanol. While this study endeavoured to produce a comprehensive phylogeny of all species of *Oreochromis*, five species could not be obtained, including *O. aureus* (although common in some countries as a food fish, we only included wild samples with known localities), *O. ismailiaensis*, *O. lidole, O. saka,* and *O. spilurus*. However, of these species we believe *O. lidole* to be functionally extinct based on our personal extensive field observations (GFT, MJG), and it is likely that *O. saka* is a geographic variant (and junior synonym) of *O. karongae* (Turner and Robinson, 1991). *Oreochromis ismailiaensis* has also not been seen since the original description and the type locality is reported to have been converted to a concrete channel devoid of fish life (A. Dunz pers. comm.), so this too may be extinct. As such (excluding *O. lidole, O. ismailiaensis* and *O. saka* as extinct or invalid species), our study represents ~93% of all *Oreochromis* species, and all *Alcolapia* species (94% when *Oreochromis* and *Alcolapia* are combined). We included two *Sarotherodon* species (*S. galilaeus* and *S. mvogoi*) based on their close association to *Oreochromis* and *Alcolapia* (Dunz and Schliewen, 2013), and the more distantly related *Coptodon rendalli* as an outgroup taxa. For the mtDNA dataset only, a further 16 samples were included based on ND2 GenBank sequences (Table S1). These included samples referred to as *Oreochromis*

Table 2
Best partitioning schemes and evolutionary models using PartitionFinder.

Partitions	Best fitting model
nuDNA exon vs. intron	
nuDNA exons	HKY + I + G
nuDNA introns	GTR + I + G
nuDNA loci	
bmp4	HKY + I + G
S7, ccng1	HKY + G
gapdhs, tyr, b2m	HKY + I + G
mtDNA	
ND2 pos1, 12S	HKY + I + G
ND2 pos2	TRN + I
ND2 pos3	TRN + G
mtDNA	GTR + I + G

aureus (three samples), as well as coverage of other Oreochromine genera, including five of the ten genera (Dunz and Schliewen, 2013), *Iranocichla*, *Konia*, *Sarotherodon* (eight further species), *Stomatepia,* and *Tristramella*, allowing us to investigate the monophyly of *Oreochromis*.

2.2. Molecular markers

As cichlids are known to exhibit mito-nuclear discordance (e.g., Rognon and Guyomard, 2003; Seehausen et al., 2003; Schliewen and Klee, 2004; Egger et al., 2007; Alter et al., 2017), we selected six nuclear loci to target a presumed rapidly evolving clade based on the age of the oldest member of this tribe (Dunz and Schliewen, 2013). These included the recently developed exon-primed intron crossing (EPIC) markers bmp4, ccng1, gapdhs, tyr, b2m (Meyer and Salzburger, 2012) and the nuclear intron S7 intron 1 (Chow and Hazama, 1998), which has previously been used in cichlid studies (Schelly et al., 2006; Schwarzer et al., 2009). We also selected two mitochondrial (mtDNA) genes: NADH dehydrogenase 2 (ND2), frequently used in cichlid phylogenetics (e.g. Klett and Meyer, 2002; Day et al., 2007; Schwarzer et al., 2009; Dunz and Schliewen, 2013), and 16S rRNA (e.g. Farias et al., 1999; Schwarzer et al., 2009) to determine if there was discordance between nuclear and mitochondrial datasets. We sequenced all samples where possible (see Appendix A, Table 1).

2.3. DNA Extraction, amplification and sequencing

Genomic DNA was extracted from fin clip samples or muscle tissue stored in 95% ethanol using the Qiagen DNeasy Blood and Tissue kit. The molecular markers were amplified using the Polymerase Chain Reaction. 1 µl of the DNA extraction was added to 12.5 µl of MyTaq™ Mix (Bioline, UK), 9.5 µl of water, and 1 µl each of the 10 M forward and reverse primers (Sigma-Aldrich, UK), to give a 25 µl total reaction volume. The primers and reaction conditions for each gene are shown in Appendix A, Table S2. Cleaned PCR products were sequenced on a 3730xl DNA Analyser (Applied Biosystems).

2.4. Alignment and partitioning scheme

Sequence data was edited using Sequencher 5.4 (Gene Codes Corporation, Ann Arbor, MI USA) and GENEIOUS v. 6.0.6 (Biomatters) (Kearse et al., 2012) in which contigs were assembled from the forward and reverse sequences. Sequences were subsequently aligned using MUSCLE in GENEIOUS (Kearse et al., 2012) using default parameters, with alignments checked for stop codons and reading frame shifts. The concatenated nuclear dataset (101 samples) included a total of 3092 bp: bmp4 (482 bp), ccng1 (650 bp), gapdhs (436 bp), tyr (561 bp), b2m (482 bp) and S7 (481 bp). We, however, removed a hyper-variable region (42 bp) from exon 2 of the b2m loci (from 144 and 185 bp) as it was shown to contain two dominant haplotypes and a high density (14) of variable sites. As the possible causes of patterns like this include strong selection, introgression from more divergent species, or paralogous sequences (although our sequences of b2m aligned against a previous dataset for this region [Meyer and Salzburger, 2012]), they would likely erroneously influence phylogenetic reconstruction. The final dataset for the nuclear data was therefore 3050 bp. The mitochondrial dataset (116 samples) included a total of 1582 bp: ND2 (1047 bp) and 16S (535 bp), with each dataset analysed separately.

The optimal partitioning scheme and model choices were assessed with PartitionFinder v.2.1.1 (Lanfear et al., 2017) using the greedy algorithm and assessed using the Bayesian Information Criteria (BIC). For the nuclear concatenated analyses we defined two subsets for the nuclear (nuDNA) dataset (combining introns vs. exons) and four subsets for the mitochondrial (mtDNA) dataset (each codon position for ND2, plus 16S). For the species tree analyses it was more appropriate to treat the six nuclear loci as separate genes (as opposed to the exon vs. intron partitions in the concatenated analysis), as the most important factor in species tree analysis is variation in the gene tree, as genes will (more or less) have different histories. We also checked to see if implementing the partitions by loci made any difference to the outcome of the nuclear concatenated analysis. To obtain evolutionary models for these partitions we re-ran PartitionFinder. Resulting models (see Table 2) were implemented in subsequent phylogenetic analyses.

2.5. Phylogenetic inference

Phylogenetic analyses were run on 1) the concatenated nuDNA dataset, and 2) the concatenated mtDNA data. Each dataset was analysed using Bayesian Phylogenetic Inference (BI) and Maximum Likelihood (ML), and the nuDNA dataset was also analysed using a Bayesian multispecies coalescent approach. Bayesian Phylogenetic Inference was implemented using MrBayes v.3.2.6 (Ronquist et al., 2012) in which analyses were run for 50,000,000 generations using four Markov chains (three heated, one cold, heating parameter 0.4) with default priors, implementing the models as defined by PartitionFinder. Maximum Likelihood analyses were run using GARLI v.2.01 (Zwickl, 2006), again implementing the models defined by PartitionFinder, and running 100 bootstrap (BS) replicates. All analyses were run on CIPRES Science Gateway server (Miller et al., 2010). The convergence of MCMC runs and burn-in were assessed in Tracer v.1.7 (Rambaut et al., 2018) and FigTree v.1.4 (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualise trees.

2.6. Non-parametric likelihood-based tests

The alternative phylogenetic topologies generated from the mt- and nuDNA datasets were evaluated using the approximately unbiased (AU) test of Shimodaira (2002), who considers this test more accurate than the Shimodaira-Hasegawa (SH) test. ML trees for each dataset were generated in Garli v.2.01 (Zwickl, 2006) using matrices containing the same one sample of each ingroup and outgroup species. These trees were imported into PAUP v.4.0b 10 (Swofford, 2002) and site likelihood scores generated. The resulting site likelihood scores were imported and run in the program CONSEL v0.20 (Shimodaira and Hasegawa, 2001).

2.7. Calibration selection and species trees analyses

Although our study was not focused on investigating divergence dates, generation of a species tree to investigate species relationships and trait evolution required the selection of calibrations. The fossil record of the Oreochromini is represented by several fossils. The oldest of these, †*Sarotherodon martyni* (Van Couvering, 1982) from the Ngorora Formation, Lake Turkana, Kenya (late Miocene: 12.0–9.3 Ma), is considered to belong to *Oreochromis* (Murray and Stewart, 1999),

although this affinity has been debated (Carnevale et al., 2003). A recent discovery of eight well preserved fossil skeletons from the nearby site, Kabchore, which is middle Miocene (12.5 Ma) are attributed to *Oreochromis* based on unique character combinations and meristic traits (Penk et al., 2018). Based on these fossils (and the occurrence of †*Sarotherodon martyni*), the older age is used to constrain *Oreochromis* and is preferred here to †*Oreochromis lorenzoi* (Carnevale et al., 2003) from the upper Miocene Messinian (7.246 Ma and 5.333 Ma), which has previously been applied as a constraint (Schwarzer et al., 2009). As the fossil record represents a minimum age, we also used a secondary calibration from Rabosky et al. (2018), which is based on a large teleost wide dataset and calibrated with a larger number of fossils, to provide a maximum age (14.71 Ma) for this node.

The Bayesian multispecies coalescent (MSC) method was applied to our nuclear data because it accounts for the coalescent process and therefore accommodates ILS (Flouri et al., 2018). Since our study is focused on closely related young species we removed the outgroup species *Coptodon* spp. to reduce rate heterogeneity. We initially used the program *BEAST v2.4.8 (Bouckaert et al., 2014). Clock models were unlinked across loci, in which a log-normal relaxed clock was selected for the S7 and gapdhs loci, while a strict clock was selected for all other loci based on the ucldStdev values (which were < 1) from an initial run. Site models were implemented based on the results of PartitionFinder in which the nuDNA data was partitioned according to loci (see Section 2.4) for these analyses. Analyses were run using the Birth-Death (BD) Model. We applied the maximum and minimum calibrations described above using a uniform prior, selecting 'use originate' = true. Population sizes in the multi-species coalescent were modelled using the 'piecewise linear and constant root' setting.

A further *BEAST analysis of all loci (nuclear and mitochondrial DNA) was also performed as this is required for running JML software (see Section 2.8). All settings were the same as described for the nuclear tree species analysis. The mtDNA data was treated as a single locus (see Table 2) in which a strict clock was preferred as indicated by the ucldStdev value which was < 1. For each dataset (i.e. nuDNA, and nuDNA + mtDNA) three analyses were run with chain length 500,000,000, and convergence was checked in Tracer v.1.7, as for the MrBayes analysis. Most ESS values for the nuDNA *BEAST analyses were > 200, with treePriors and clock rates > 100, but some values (posterior, species coalescent and popMean) were between 96 and 93. All ESS values for the nuDNA + mtDNA species tree were > 200. Runs were combined using TreeAnnotator, and resampled, discarding 25% burnin, and visualised using FigTree v.1.4.3.

We also used the program Bayesian Phylogenetics and Phylogeography (BPP) v 4.0 (Flouri et al., 2018) (method A01) and performed runs of 30,000,000 MCMC iterations testing alternative priors. The calibrations previously described were also applied to this analysis. Although we did not obtain convergence across the runs on the maximum a posteriori tree (MAP) we obtained convergence on the majority rule consensus (MRC) tree, which was identical across all runs. We tried two priors on the root age (tau prior) 1) Inv-Gamma (3,0.002), and 2) Inv-Gamma (2,0.05), and both priors yielded the same (MRC) trees.

2.8. Testing for hybridisation

Due to discordance between mitochondrial and nuclear trees (see Results 3.1) we attempted to investigate if we could distinguish between introgression and incomplete lineage sorting (ILS) using the method described by Joly et al. (2009), implemented in JML v.1.3.1 (Joly, 2012). This method uses posterior predictive checking by comparing the minimum sequence distance between two species to test if the minimum distance is smaller than expected under a regime not accounting for hybridisation. However, we acknowledge that our dataset does not include many samples per species, and that loci from our nuclear dataset are short, and therefore lack power of longer sequences (see Joly et al., 2009). As this method uses the posterior distributions of species trees, population sizes and divergence times, we generated a further coalescent tree from all loci (see Section 2.7), ensuring that all loci were unlinked to obtain mutation rates needed for JML. JModelTest v.2.1.10 (Guindon and Gascuel, 2003; Darriba et al., 2012) was implemented to obtain models for individual loci and also to obtain settings for the .jml.ctl file of JML (state frequencies, transition-transversion ratio, proportion of invariable sites, gamma rate heterogeneity). We ran 1000 simulations for each of the three selected loci, mtDNA (locus rate = 0.0038), bmp4 (locus rate = 0.00036) and tyr (locus rate = 0.000656) pairwise distance comparisons. A Benjamini-Hochberg correction was applied to the results using R v.3.5 (R Core Team, 2018).

2.9. Trait analyses

We tested for correlated evolution of traits using BayesTraits v3.0.1 (Pagel, 2004; Pagel and Meade, 2006), which was also used for ancestral state reconstruction. We focused on trait data for tolerance to soda conditions (salinity, temperature, and pH) with species states derived from the literature (see Table 1). As tolerance to elevated salinity and temperature was most common, we tested for correlated evolution between these traits, and separately reconstructed ancestral states for soda adaptation, defined as comprising all three tolerance traits found in a species: salinity, temperature, and pH. We also examined the correlated evolution of phenotypic secondary sexual male characteristics that have been used in previous taxonomic analysis: the genital tassel, and extended male jaws. Although the two traits have not previously been implicated to have correlated evolution, the two traits are used as diagnostic characters to separate clades and subgenera of *Oreochromis*, and the traits do not co-occur, so we tested whether their presence/absence was correlated. We ran the ancestral state reconstruction analyses on the nuclear datasets only, using both the MrBayes (non-ultrametric) and *BEAST (ultrametric) phylogenies. We did not ultrametricise the MrBayes tree, as branch lengths can have a substantial effect on ancestral state reconstructions (McCann et al., 2016), and following recommendations from Cusimano and Renner (2014) to run reconstructions on more than one type of branch length depiction we therefore include both the phylogram and ultrametric trees. The MrBayes nuclear concatenated tree was pruned to include only one sample per species (or subspecies, where relevant). One sample each was included for subspecies of three species (specifically: *O. niloticus niloticus*; *O. niloticus cancellatus*; *O. niloticus filoa*; and *O. placidus placidus*; *O. placidus ruvumae;* and *O. shiranus shiranus, O. shiranus chilwae*), as plausibly the subspecies may exhibit different adaptations/morphology from each other and so were coded separately. The resulting tree contained 40 tips (including subspecies of *Oreochromis*, species of *Alcolapia*, and the three outgroup species). Details of the specimens included in the pruned tree are given in Appendix A, Table S3. No pruning was required for the *BEAST species tree as there was only one tip per species (total of 36 tips), which also did not include all the respective subspecies of *O. shiranus* and *O. niloticus* as the availability of only single samples meant that they could not be included in the multispecies coalescent approach. The BayesTraits analysis tested the tolerance (salinity and temperature) and morphological characters separately, using the Discrete analysis mode (testing two binary characters) and comparing the independent model (no correlation of shifts between the two traits) with dependent (correlated shifts) models (see Table S3 for how traits were coded per species). Each analysis (independent and dependent) was run for 5 independent runs, with 10 million MCMC iterations, with the first million discarded as burnin, and the stepping stone sampler set to use 100 stones and run each stone for 10^4 iterations. Tree branch lengths were scaled to a mean of 0.1 using the ScaleTrees command, following the recommendations of the BayesTrait manual. Priors for the rate parameters (the rate coefficient for the gain/loss of each trait respectively) were set based on initial

Maximum Likelihood runs in BayesTrait. Based on the ML rates estimates, the MCMC analysis on the MrBayes phylogeny used a hyperprior for all rate coefficients specifying an exponential prior seeded from a uniform distribution on the interval 0–10, and those on the *BEAST trees used an exponential prior seeded from a uniform distribution on the interval 0–1. Mixing of the chains was checked in the output files to ensure acceptance rates were in the range 20–40%; convergence and ESS were checked using Tracer v1.6. The results for the independent and dependent models based on the marginal likelihood from the stepping stone sampler (expressed on a natural log scale) were compared using Log Bayes Factors calculated as:

$$\text{Log Bayes Factors} = 2(\text{log marginal likelihood complex model} - \text{log marginal likelihood simple model})$$

Interpretation of the values was based on Gilks (1996) as described in the BayesTraits manual, specifically: a log BF factor of < 2 is interpreted as weak evidence, > 2 as positive evidence, 5–10 as strong evidence, and > 10 as very strong evidence.

The tests of correlated evolution suggested that tolerance to salinity and temperature were correlated traits (Results Section 3.3). However, as salinity and thermal tolerance were predominantly concentrated in one clade (*Alcolapia* and *O. amphimelas*), which could potentially generate a spurious pattern of correlation (Uyeda et al., 2018), we ran two separate ancestral reconstructions coding the two traits either as independent or dependent (correlated) traits. The mean values of the proportional likelihoods for each node was calculated with Excel. We also separately reconstructed a 'soda' adaptation trait for which tip species were coded as present if the species exhibited elevated tolerance to all three parameters (temperature, salinity, and pH). We reconstructed ancestral traits for the phenotypic traits (genital tassel and male jaw morphology) independently, as they did not exhibit evidence of correlated trait evolution (see Results). Input files were prepared for BayesTraits using TreeGraph2 (Stöver and Müller, 2010), and results of the ancestral state reconstruction analysis were visualised using ggtree 1.12.14 (Yu et al., 2017) in R v3.5 (R Core Team, 2018). For the analyses using the *BEAST trees, ancestral state reconstructions were conducted on a reduced set of trees (resampled to 25,001 trees in TreeAnnotator), and visualised by plotting on the Maximum Clade Credibility tree.

3. Results

3.1. Phylogenetic relationships

The nuclear species tree (Fig. 1), BI (Supplementary Fig. S1) and ML concatenated (data not shown, but BS values included on Fig. S1) trees were reasonably similar in topology, particularly regarding subclades, although there were several instances of taxa occurring in alternative positions between the concatenated and species trees, most notably the sister group of *Alcolapia* (discussed below). The mtDNA concatenated trees generated using BI and ML were largely congruent (see Fig. 1b, Supplementary Figs. S2). However, comparison of nu- and mtDNA trees showed high levels of mito-nuclear discordance from the strikingly different placement of certain clades and taxa (Fig. 1), and there are instances where different groups appear to be well resolved and monophyletic in trees built from the different sets of loci (Fig. 1, Supplementary Figs. S1, S2). The AU test based on the ML concatenated trees resulted in the mtDNA tree ($-\ln$ L 7024.31) being significantly worse fit to the data ($P < 0.001$) than the nuclear tree ($-\ln$ L 6516.82), but we discuss both topologies below. As mitochondrial data is prone to introgression in cichlids, we used the nuclear species trees for downstream analyses. A multispecies coalescent approach is preferred as it accounts for gene tree- species tree incongruence that arise due to population level processes, and is particularly suitable for more recently diverged groups (e.g. Ogilvie et al., 2016, 2017 and refs therein), although we acknowledge that the nuclear genome may also be introgressed in this group of fishes.

3.1.1. The sister group of Alcolapia

Both the *BEAST (Fig. 1a) and BPP (Supplementary Fig. S3) versions of the nuDNA species tree resolved *O. esculentus* (a freshwater species) as the sister group to *Alcolapia* (extreme soda-lakes), with *O. amphimelas* (seasonal soda-lakes) as sister to these lineages. This is however not well supported (0.55/0.69 respectively), and a preliminary *BEAST analysis including the 42 bp hyper-variable data (removed from subsequent analyses due to high variability) conversely resolved *O. amphimelas* as the sister group (data not shown). The nuDNA concatenated phylogenies (Fig. 1b, Appendix A Fig. S1) (with or without the 42 bp region) resolved the sister group to the *Alcolapia* clade as *O. amphimelas* (BPP 1/0.90 [concatenated/species tree]; BS 91), with *O. esculentus* resolved (BPP 1/0.90; BS 89) as sister to the *Alcolapia* + *O. amphimelas* group. All these species are closely distributed geographically occurring in NW Tanzania and SW Kenya. The nuDNA concatenated tree (Appendix A Fig. S1) shows that there is also some resolution within the *Alcolapia* clade itself, with the geographically isolated *A. grahami* (Lake Magadi, Kenya) resolved (BPP 0.99) and most closely related to the 'northern' *A. alcalica* from Lake Natron, although there is poor resolution regarding the relationships among the three sympatric Lake Natron species ('southern' *A. alcalica*, *A. latilabris*, *A. ndalalani*). In contrast, the mtDNA concatenated phylogeny (Fig. 1c, Appendix A Fig. S2) placed the *Alcolapia* clade as the sister group to (BPP 0.86/BS 56) a mixed assemblage of six species (including *O. esculentus*), which are not themselves phylogenetically resolved, with *O. amphimelas* resolved as the sister taxon to this clade. There is also no resolution of the relationships among the constituent species of *Alcolapia*. When the ND2 data alone is analysed, *Alcolapia* + *O. amphimelas* are sister taxa, albeit with low support, and the 'mixed assemblage' is the sister group to this clade (data not shown) indicating differing signals from the mtDNA loci.

3.1.2. Relationships of African Great lake taxa

There are also disparities in the placement of some of the African Great Lake taxa. For example, the Lake Tanganyika basin species *O. tanganicae* and *O. malagarasi* are sister taxa in all the nuDNA phylogenies, but are not close relatives in the mtDNA tree. In the mtDNA tree a sister relationship of *O. malagarasi* and *O. upembae* (occurs in the Congo Basin, and East Africa, specifically within the Lake Tanganyika catchment) is resolved, which is in line with Trewavas's (1983) reporting of close relationships between these two species based on phenotypic characters. Other placements that differ between the nuDNA and mtDNA phylogenies included riverine taxa, *O. niloticus* (a wide-ranging taxon whose native range is across the Nilo-Sudan ichthyo-province) which groups with *O. lepidurus* (occurs in the Lower Congo River) and *O. upembae* in the nuDNA tree. Conversely *O. niloticus* groups with *O. angolensis* (Quanza ichthyo-province) and *O. schwebischi* (Lower Guinea Forest ichthyo-province) in the mtDNA tree. In the mtDNA trees subspecies *O. niloticus cancellatus* and *O. niloticus filoa* (Appendix A Fig. S2) were consistently resolved as being more closely related to each other than to *O. niloticus niloticus,* but there was poorer resolution of *O. niloticus* subspecies in the nuDNA concatenated tree (Appendix A Fig. S1). Notably the Lower Congo River species *O. lepidurus* is closely related to a taxon that occurs in Lake Tanganyika (or its catchment) in either tree: it is the sister taxon to *O. tanganicae* in the mtDNA tree, and sister to the clade comprising *O. upembae* and *O. niloticus* in the nuDNA tree. A connection between these water bodies has been demonstrated in other lineages such as lamprologine cichlids (e.g. Day et al., 2007) and mastacembelid spiny eels (Day et al., 2017).

However, there are areas of some congruence between the trees built from mtDNA versus nuDNA. In particular, species from the Lake Malawi catchment and the formerly connected Ruvuma (also known as the Rovuma) catchment (*O. shiranus, O. squamipinnis, O. karongae, O.*

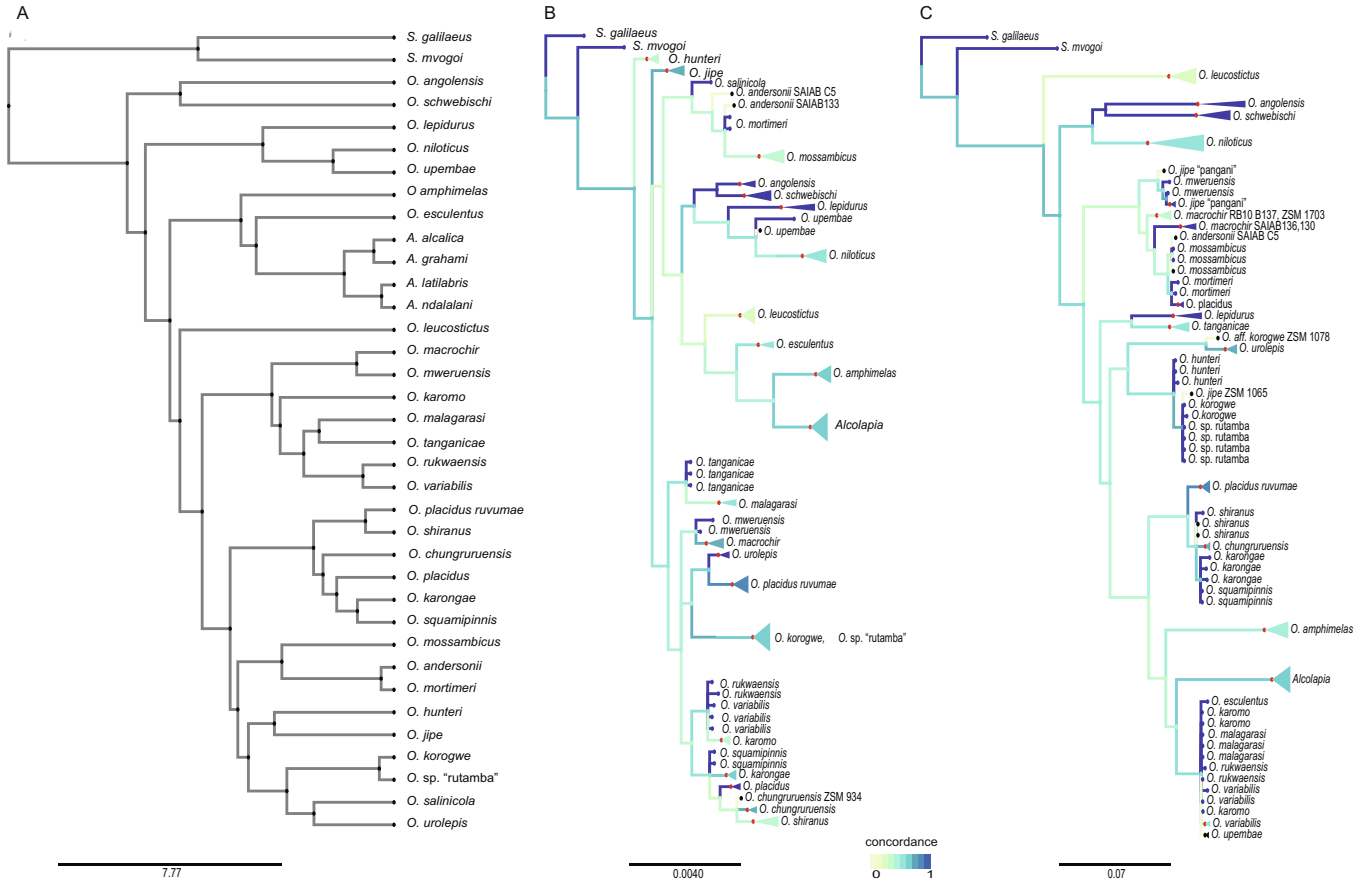

Fig. 1. Comparison of the (A) nuclear DNA species tree (generated using *BEAST); (B) nuclear DNA Bayesian concatenated tree and (C) mitochondrial DNA Bayesian concatenated tree (generated using MrBayes). Figures B + C are compared using Phylo.io software (Robinson et al., 2016). Species that are monophyletic are collapsed. Differences and similarities of both trees are highlighted on the branches, in which the lighter to darker colour indicates the similarity of best matching subtrees between the two trees. Black circles below nodes show > 95% BPP support values (see supplementary Figs. S1 and S2 for complete trees and support values for BPP and BS). *Coptodon* spp. are removed for clarity.

chungruruensis, O. placidus ruvumae) form a well supported clade in both the mtDNA and nuDNA species trees. The concatenated nuclear tree supports a variation of this grouping, with the exception of *O. placidus ruvumae,* which conversely groups within a larger clade that is sister to Lake Malawi catchment species. In both nuclear trees (species and concatenated), *O. placidus* (Zambezi river) is also a member of this group, whereas this taxon groups with other largely southern African species in the mtDNA tree.

3.1.3. Pangani relationships

In both the nuDNA concatenated, and mtDNA trees, *O. jipe* ("pangani") and the type species of the genus, *O. hunteri* (both endemic to the upper Pangani system in the Tanzania/Kenya border region of the East African ichthyo-province) are not monophyletic. This result is also suggested from the mtDNA tree. In this tree *O. jipe* and *O. mweruensis* (from the Zambezi ichthyo-province) form a clade in which constituent species are non-monophyletic, although the *O. jipe* sample ZSM 1065 (not included in the nuDNA analysis) forms a clade with *O. hunteri* and *O. korogwe*, highlighting that mtDNA loci may not always correctly resolve species boundaries. However, when our nuclear data was analysed using species tree methods *O. hunteri* and *O. jipe* are resolved as monophyletic, although with weak support. A recent study focused on *O. hunteri* (Moser et al., 2018) based on mtDNA control region (830 bp) also supported a close relationship between *O. jipe* and *O. hunteri*, but this study only included a small subset of *Oreochromis* species.

3.1.4. Oreochromis monophyly

The inclusion of samples from GenBank within the mtDNA analysis (see Supplementary Fig. S2) allowed us to test the monophyly of *Oreochromis*. With the exception of sequences uploaded as *O. aureus* (for which we did not have access to voucher specimens), *Oreochromis* was resolved as monophyletic, or rather as paraphyletic with *Alcolapia* nested within it. The position of *O. aureus* (a Nilo-Sudanic species) grouped within the Lake Barombi Mbo radiation, and outside of *Oreochromis* likely implies that the sampled taxa are hybrids or were originally mis-identified. However, although *Oreochromis aureus* has not been included in recent molecular studies of tilapiine relationships (e.g. Schwarzer et al., 2009; Dunz and Schliewen, 2013), an allozyme study previously resolved *O. aureus* at the base of the *Oreochromis* clade (Pouyaud and Agnèse, 1995). Of the four GenBank *O. aureus* samples used to sequence NADH, only one is in a published paper (DQ465029.1; Cnaani et al., 2008), and was collected from stocks of *O. aureus* in Israel that were originally sourced from Lakes Hula (Israel) and Manzala (Egypt), both of which are also inhabited by *S. galilaeus*. Trewavas (1983) notes several characters that distinguish *S. galilaeus* from co-occurring *O. aureus* and *O. niloticus*, including the depth of the pre-orbital bone. Intergeneric hybrids between *Oreochromis* and *Sarotherodon* are viable and have been produced in aquaculture strain development (using *in vitro* fertilisation; Bezault et al., 2012), but we are not aware of any reports of intergeneric crosses in the wild, especially in natural sympatry zones. Further samples of *O. aureus* are required to investigate this placement fully and ensure that *S. galilaeus* samples have not been mis-identified as *O. aureus*.

3.2. Assessment of hybridisation

Results from the JML analysis revealed no support for introgression, with no significant signal of hybridisation after applying the Benjamini-Hochberg correction in any of the loci tested (mtDNA, tyr, bmp4). These results suggest that incomplete lineage sorting could explain the incongruence in mtDNA and nuclear datasets (rather than a signal of introgression). However, the large number of pairwise comparisons (630) mean that testing across the entire phylogeny is unlikely to uncover signals of hybridisation, and future analyses may focus on specific pairwise comparisons of interest. We suggest that additional data would need to be included to help refine this analysis, but that ultimately examination of species of *Oreochromis* using genome-wide data would provide a powerful approach to test hybridisation hypotheses.

3.3. Diversification and ancestral state reconstruction

The BayesTraits analysis for environmental tolerance traits (salinity and thermal tolerance) using the Discrete model for the nuclear *BEAST species tree gave a marginal log likelihood for the independent model of −26.50 (with a standard deviation of 0.04 across 5 runs), while the dependent model had a marginal log likelihood of −24.72 (s.d. 0.02). These resulted in a log Bayes Factor of 3.56, indicating moderate support for the dependent model, suggesting that the shifts in adaptation to salinity and thermal tolerance may be correlated. Results when using the nuclear concatenated MrBayes tree were more conclusive, with the 5 runs resulting in a log Bayes Factor of 6.78, indicating strong support for correlation of the traits. For the morphological traits (genital tassel and extended male jaw morphology) using the nuclear *BEAST species tree the independent model marginal log likelihood was −39.82 (s.d. 0.06) while that for the dependent model was −39.18 (s.d. 0.04). The log BF was 1.30, indicating that there was no support for the complex model (dependence of trait shifts), and suggesting that the two phenotypic traits do not exhibit correlated rate shifts. Results when using the nuclear concatenated MrBayes tree were similar, with the 5 runs resulting in a log Bayes Factor of −0.22, indicating no support for correlation of the phenotypic traits.

The ancestral state reconstruction indicated that adaptation to increased salinity and/or temperature has occurred multiple times within the genus *Oreochromis*, but that adaptation to soda lake conditions has likely occurred once, in the lineage leading to *Alcolapia* + *O. esculentus* + *O. amphimelas* (Fig. 2A). The results were similar (and more conclusive) in the analysis using the concatenated nuclear phylogeny (Fig. S4A), where soda adaptation is likely to have occurred only once in the lineage leading to *Alcolapia* + *O. amphimelas*. Reconstructing thermal and salinity tolerance independently also showed that adaptation is likely to have occurred multiple times throughout the phylogeny (Fig. S5). The reconstruction of the phenotypic traits (Fig. 2B) indicated that the secondary sexual characteristics or extended jaw and genital tassel are not correlated, and that both are likely to have evolved multiple times, suggesting that they are not useful taxonomic diagnostic characteristics for this genus. The reconstruction of phenotypic characters on the MrBayes phylogeny gave similar results to that using the *BEAST phylogeny, and suggested that the morphological traits evolved independently in multiple clades (Fig. S4B).

4. Discussion

Here, we present the first comprehensive phylogenetic analysis of the cichlid genus *Oreochromis* using multi-marker molecular datasets comprising nuclear and mitochondrial loci, and reveal high levels of incongruence between them. This incongruence could either represent incomplete lineage sorting and/or introgression, and while we did not find a signal for the latter mechanism it cannot be ruled out. Incongruence as a result of these mechanisms is commonly reported in other cichlid studies (e.g., Genner and Turner, 2012; Willis et al., 2014; Meier et al., 2017). We suggest that the nuDNA phylogeny is likely to be a better estimate of the species tree, although we acknowledge that additional data would help to resolve conflicting relationships. Based on the nuDNA trees, we show that adaptation to adverse environmental conditions (i.e. increased salinity and temperature) has occurred multiple times, but that adaptation to extreme (soda) conditions is likely to have occurred once. Irrespective of the dataset (nuDNA or mtDNA), we demonstrate that the *Alcolapia* clade is consistently resolved within the genus *Oreochromis*.

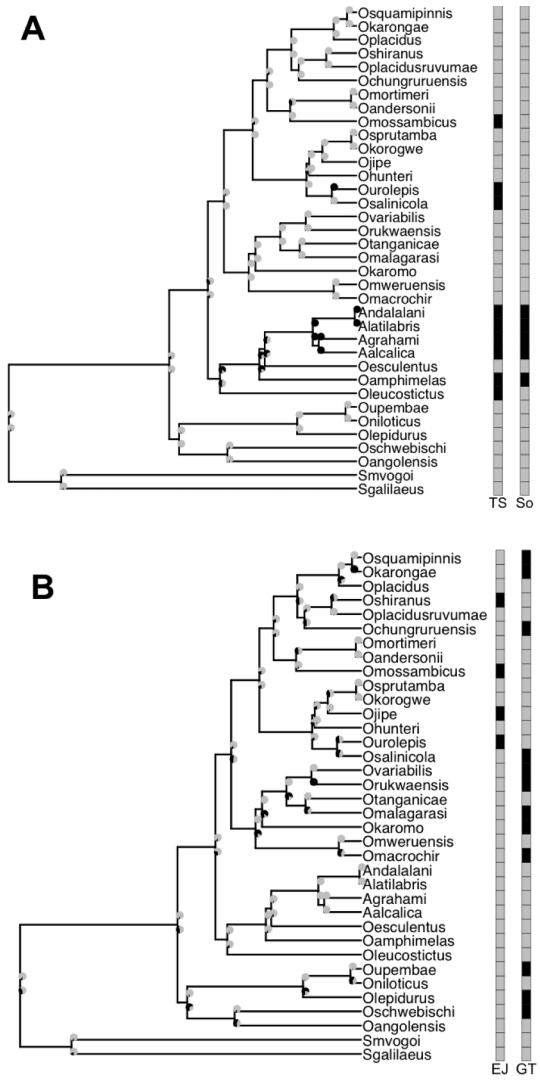

Fig. 2. Ancestral state reconstruction from BayesTraits analysis based on the species tree nuclear phylogeny (generated using *BEAST). (A) Ancestral state reconstruction of thermal/salinity tolerance (TS) and soda adaptation (So). Colours at tips represent adaptation reported in extant species (listed in Table 1) (black = present; grey = absent). Pie charts at internal nodes indicate probability of presence/absence of ancestor exhibiting adaptation to soda conditions, from BayesTraits analysis. (pie above node: temperature/salinity tolerance ancestral state reconstruction; pie below node: soda adaptation ancestral state reconstruction). (B) Ancestral state reconstruction of phenotypic male secondary sexual characteristics: genital tassel (GT) and extended jaw morphology (EJ). Colours at tips represent phenotypic characters in extant species (black = present; grey = absent). Pie charts at internal nodes indicate probability of presence/absence of ancestor exhibiting trait, from BayesTraits analysis (pie above node: extended jaw morphology ancestral state reconstruction; pie below node: genital tassel ancestral state reconstruction).

4.1. Taxonomy of the genus Oreochromis: effects of mito-nuclear discordance

We observed substantial mito-nuclear discordance in the phylogenetic analysis (Fig. 1). We note that several of the relationships in the nuclear analysis are concordant with previous phenotypic classifications based on phenotypic features, but that certain relationships within the mtDNA analysis are also plausible based on phenotypic and geographical range data. The results suggest reticulate evolution could have played a role in the existing genetic relationships, and could be a result of the fact that many tilapiine species are known to have been widely translocated for stocking and aquaculture purposes, although we do not find significant evidence of introgression using JML analysis. Alternative explanations include incomplete lineage sorting that is widely reported in cichlids (e.g., Genner and Turner, 2012; Meier et al., 2017; Meyer et al., 2017). Previous analysis of mito-nuclear discordance in the tribe Oreochromini suggested that ancient hybridisation was the more probable explanation of this discordance (Dunz and Schliewen, 2013).

The translocation and establishment of non-native species (*O. leucostictus, O. niloticus*) within Tanzania has been recently documented (Shechonge et al., 2019). We note two additional non-native species occurrences in Tanzania from our sampling: *O. leucostictus* (AF042-04) sampled from fish ponds in the Lake Eyasi basin (reportedly stocked from streams surrounding Lake Eyasi), and *O. urolepis* (AF014-01) sampled from irrigation canals draining into Lake Manyara. Male specimens collected from the latter site also exhibited the distinctive extended jaw morphology of sexually mature male *O. urolepis*. To our knowledge, this is the first report of *O. urolepis* sampled in a natural water body outside of its native catchments, that includes the Wami, Ruvu and Rufiji basins. However, we are aware of aquaculture centres rearing *O. urolepis* in Tanga (Mmochi, 2017).

We also find taxonomic discrepancy in the resolution of *O. placidus* samples from different geographic areas. Trewavas and Teugels (1991) synonymised *O. placidus placidus* and *O. placidus ruvumae*, but other researchers have reported substantial morphological differences between *O. placidus* specimens from the type locality (Buzi River) and those found in the Ruvuma River (Bills, 2004). Our data support this latter finding as in both the nu- and mtDNA phylogenies, *O. placidus ruvumae* samples form a distinct clade from *O. placidus placidus*. However, we find that within the mtDNA tree and nuDNA species tree *O. placidus ruvumae* is either sister or groups within the Lake Malawi catchment clade, whereas in the concatenated nuDNA tree is it more distantly related. The close relationship supported in several of these trees is consistent with recent suggestions that the Ruvuma River was formerly the outflow of Lake Malawi (Ivory et al., 2016), supported by the shared presence of *O. shiranus* in the Lake Malawi catchment, in Lake Chiuta (Ruvuma catchment) and Lake Chilwa (endorheic but formerly connected to Ruvuma) (Trewavas, 1983). However, the mitochondrial and nuclear trees disagree with morphological hypotheses (Trewavas, 1983), although these are not based on cladistic analyses, which placed *O. squamipinnis, O. karongae* and *O. chungruruensis* with the other tasseled species in the subgenus *Nyasalapia,* but grouped *O. shiranus* and *O. placidus* with other species from East coast rivers showing enlarged male jaws in a division of subgenus *Oreochromis*.

The molecular phylogenies and ancestral state reconstruction (Fig. 2B) suggest that the male secondary sexual phenotypic characteristics previously used to group species do not represent phylogenetically conserved characters. Specifically, the presence of a genital tassel in males, a defining character of the *Nyasalapia* subgenus rank erected by Thys van den Audenaerde (1968), does not reliably distinguish clades resolved in our molecular phylogeny, and is likely to have evolved multiple times. Trewavas (1983) reported that other than genital tassel, there were no defining phenotypic characteristics distinguishing *Nyasalapia* from the rest of *Oreochromis* and suggested that the subgeneric value of the tassel was open to question.

Alcolapia is consistently resolved within *Oreochromis* irrespective of dataset and forms a strongly supported clade with *O. amphimelas* and *O. esculentus* irrespective of nuclear analyses performed. All previous phylogenetic studies including the two genera have also resolved *Alcolapia* within *Oreochromis*, albeit with less comprehensive sampling of either genus (Seegers et al., 1999; Nagl et al., 2001; Schwarzer et al., 2009; Dunz and Schliewen, 2013; Kavembe et al., 2013; Matschiner et al., 2017; Rabosky et al., 2018). However, the sister group of *Alcolapia* is contentious, as *O. amphimelas* is strongly supported as its sister group based on concatenated nuclear analyses, whereas, species tree analyses placed *O. esculentus* as sister, albeit with weak support. A species tree analysis of the nuclear data without removing the hypervariable 42 bp (data not shown) did support the *Alcolapia* + *O. amphimelas* relationship, and it is likely that the uncertainty in the species tree may be attributed to a lack of synapomorphies.

4.2. Systematics account: revised classification of Alcolapia

Based on the results of our phylogenetic analyses, we propose a revised classification of *Alcolapia*. Given the position of *Alcolapia* within the comprehensively sampled molecular phylogenies presented here, and concordant with previous molecular work, we propose that *Alcolapia* is synonymised with the genus *Oreochromis*, retaining the subgeneric allocation of *Alcolapia*. We recognise *Alcolapia* Thys van den Audenaerde, 1969 as a subgenus within *Oreochromis* Günther, 1889. The synonymy is necessary to reflect a monophyletic taxonomy; *Oreochromis* is paraphyletic unless *Alcolapia* is subsumed within it. While both Trewavas (1983) and Seegers and Tichy (1999) noted several morphological characters for the diagnosis of *Alcolapia* as a subgenus, most of the characters were shared with *O. amphimelas* and several overlapped with other *Oreochromis* species. The elevation of *Alcolapia* to genus rank was ultimately based on molecular (mtDNA) data (Seegers et al., 1999). However, this is not supported by subsequent molecular analyses, including the present study. As such, the revised classification we propose here follows the previous taxonomy of Seegers and Tichy (1999), namely: *Oreochromis (Alcolapia) alcalicus*, *Oreochromis (Alcolapia) grahami*, *Oreochromis (Alcolapia) latilabris*, and *Oreochromis (Alcolapia) ndalalani*.

An alternative solution to the synonymisation would be to split *Oreochromis* into several genera representing the constituent reciprocally monophyletic groups; however, given the differences in groups between datasets, the current data would not support this taxonomic treatment. In addition to the molecular evidence for synonymising these genera, we note the phenotypic similarities between the species of *Oreochromis* (*Alcolapia)* and *O. amphimelas* reported by Trewavas (1983).

4.3. Timing and colonisation of extreme conditions

Our study shows that there has been repeated adaptation to elevated salinity and temperature throughout the evolutionary history of *Oreochromis* irrespective of phylogenetic hypothesis, but that adaptation to soda conditions (high temperature, salinity, alkalinity, and low dissolved oxygen) has likely occurred once. The placement of *O. esculentus* in a clade with *Alcolapia* and *O. amphimelas* in the species tree analyses (Fig. 1A, Supplementary S3), raises the possibility that the adaptation to soda conditions was also gained in *O. esculentus* (Fig. 2A), although it does not currently inhabit soda conditions (Table 1). While *O. esculentus* is native to the freshwater Lake Victoria (Njiru et al, 2005; Hallerman and Hilsdorf, 2014), our samples are from Lake Rukwa, a relatively shallow saline lake, where they have been introduced from Lake Victoria (Seegers, 1996). Lake Rukwa became saline around 5,500 years ago (Barker et al, 2002), and although conditions (pH 9.19–9.26; temperature 27–32 °C; salinity: 6000 mg/L Haberyan, 1987; Bathymetric survey report, 2014) are not as extreme as the springs inhabited by *Oreochromis (Alcolapia)* (e.g. pH 8–12, temperature

27–42 °C, salinity 34,000 mg/L) or *O. amphimelas*, they are much higher than Lake Victoria, for example (pH: 8.2–9, temperature: 23–28 [surface water temperature], salinity: 97 mg/L) (vanden Bossche and Bernacsek, 1990). Of course, Lake Victoria has experienced phases of desiccation (Johnson et al., 1996) and so it is likely that species such as *O. esculentus* will have encountered periods of higher salinity in the past. This is probably true of many other *Oreochromis* populations and thus it is perhaps unsurprising given that there has been multiple adaptations to these conditions throughout the evolution of this group, indicating a likely shared genetic mechanism allowing fish from this genus to adapt to such conditions.

Our age estimates of the diversification of the *Oreochromis (Alcolapia)* adaptive radiation at 1.73 Ma (95% HPD: 3.20–0.57) coincide with estimates of the date of basin formation for Lakes Natron and Magadi at 1.7 Ma and the formation of the single palaeolake (Orolonga) preceding Natron and Magadi, around *c.*700 Ka (Eugster, 1986). The Lake Magadi species *O. (A.) grahami* is estimated to have diverged from the Lake Natron species *O.* (*A.*) *alcalicus* at 0.70 Ma (95% HPD: 1.55–0.007 Ma), suggesting these taxa may have already diverged prior to the separation of the lakes. However, although these dates are generally well before the time that these lakes are suggested to have separated, during an aridity event dated at *c.*11 ka (Williamson et al., 1993) the upper bound is after this timeframe. It is likely that the nuclear loci selected for this study are too slowly evolving to provide a precise estimate of divergence dates for a potentially recent radiation (Ford et al., 2015). More nuclear loci and broader outgroup sampling including additional fossil calibrations should be considered in future studies to test the accuracy of these dates.

It has been suggested that *O. amphimelas* (occurs in Lakes Manyara, Eyasi, Sulungali, Kitangiri and Singida) might be a close relative of *Oreochromis (Alcolapia)* based on adaptation to soda conditions – although *O. amphimelas* experiences less extreme conditions (Trewavas, 1983) – supporting the findings presented here. There is also a hydrological connection between the two basins, with groundwater flowing northwards from Manyara to Natron and the lowest border of Manyara (~80 m above the current lake level) forming an overspill to Natron (Hillaire-Marcel and Casanova, 1987; Bachofer et al., 2014). However, the suggestion that the two basins were joined in a palaeolake as recently as 10 ka (Holdship, 1976) has not been supported by geological evidence (Casanova and Hillaire-Marcel, 1992). As for Magadi, fossils indicate that Lake Manyara also was inhabited by a larger freshwater cichlid in a previous palaeolake (Schluter et al., 1992). However, unlike Natron-Magadi, Manyara now contains only one native species (*O. amphimelas*).

5. Conclusions

The comprehensive molecular phylogenies of *Oreochromis* presented here represent the first attempt to clarify species relationships for this relatively young and widespread African cichlid genus. Trees reconstructed from nuclear sequence data likely give the best hypothesis of relationships, although we recognise that nuclear genomes may also show introgression (e.g. Nevado et al., 2011). The multispecies coalescent approach is preferred over concatenation methods, as it accounts for gene tree- species tree incongruence that arise due to population level processes and has been shown to be especially suited to the estimation of shallower evolutionary relationships (e.g. Ogilvie et al., 2016, 2017; Flouri et al., 2018), but we acknowledge that we include a limited number of loci for these types of analyses. To clarify relationships and better understand the substantial mito-nuclear discordance that we report, further work should focus on genomic data e.g. ultraconserved element or reduced representation genome-wide sequencing to resolve potential causes of this incongruence and better clarify the areas of conflict (currently in progress). Genomic data will also enable the likely shared genetic mechanism allowing these fish to adapt to adverse aquatic conditions to be investigated. Our results highlight the importance of establishing species relationships within this genus, which contains several commercially important, and many endangered, species.

Funding

We acknowledge the Systematics Research Fund, the Genetics Society Heredity Fieldwork grant, and UCL Graduate scholarship (AGPF); BBSRC/NERC SYNTAX award (JJD); Royal Society Leverhulme Trust Africa Awards (MJG, GFT, BPN).

Data accessibility

Sequencing reads are deposited in NCBI. Nexus files of aligned sequenced data used in this study are available from the Supplementary material and corresponding author(s) on request.

Acknowledgements

We thank Simon Joly and Andrew Meade for advice in the use of the JML and BayesTraits software programs, respectively. We also thank Wendy Hart for sequencing support, and Etienne Bezault, Emmanuel Vreven, and Julia Schwarzer for providing additional samples. The associate editor Guillermo Orti and reviewers, Michael Matschiner and Prosanta Chakrabarty, provided helpful comments on an earlier version of this manuscript. We thank Wanja Nyingi and staff at the National Museums of Kenya for assistance with fieldwork and sample collection.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.04.008.

References

Alter, A.E., Munshi-South, J., Stiassny, M.L.J., 2017. Genome wide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River. Mol. Ecol. 26, 1401–1419.

Bachofer, F., Quénéhervé, G., Märker, M., 2014. The delineation of paleo-shorelines in the Lake Manyara basin using TerraSAR-X data. Remote Sensing 6, 2195–2212.

Barker, P., Telford, R., Gasse, F., Thevenon, F., 2002. Late Pleistocene and Holocene palaeohydrology of Lake Rukwa, Tanzania, inferred from diatom analysis. Palaeogeogr. Palaeoclimatol. Palaeoecol. 187, 295–305.

Bezault, E., Feder, C., Derivaz, M., et al., 2007. Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. Aquaculture 272, S3–S16.

Bezault, E., Rognon, X., Clota, F., Gharbi, K., Baroiller, J.-F., Chevassus, B., 2012. Analysis of the meiotic segregation in intergeneric hybrids of tilapias. Int. J. Evolut. Biol. 2012, 817562.

Bills, I.R., 2004. A survey of the fishes and fisheries in the Niassa Reserve, Niassa and Cabo Delgado Provinces, Mozambique (10-30/8/2003). SAIAB Investigational Report 69.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M.A., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C.P., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O., Di Palma, F., 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513, 375–381.

Carnevale, G., Sorbini, C., Landini, W., 2003. *Oreochromis lorenzoi*, a new species of tilapiine cichlid from the Late Miocene of Central Italy. J. Vertebr. Paleontol. 23, 508–516.

Casanova, J., Hillaire-Marcel, C., 1992. Chronology and paleohydrology of Late Quaternary high lake levels in the Manyara basin (Tanzania) from isotopic data (^{18}O, ^{13}C, ^{14}C, Th/U) on fossil stromatolites. Quat. Res. 38, 205–226.

Chow, S., Hazama, K., 1998. Universal PCR primers for S7 ribosomal protein gene introns

in fish. Mol. Ecol. 7, 1255–1256.
Cnaani, A., Lee, B.Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., et al., 2008. Genetics of sex determination in tilapiine species. Sex Dev. 2, 43–54.
Cusimano, N., Renner, S.S., 2014. Ultrametric trees or phylograms for ancestral state reconstruction – Does it matter? Taxon 36, 721–726.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 (8), 772.
Day, J.J., Santini, S., Garcia-Moreno, J., 2007. Phylogenetic relationships of the Lake Tanganyika cichlid tribe Lamprologini: the story from mitochondrial DNA. Mol. Phylogenet. Evol. 45, 629–642.
Day, J.J., Fages, A., Brown, K.J., Vreven, E.J., Stiassny, M.L.J., Bills, R., Friel, J.P., Rüber, L., 2017. Multiple independent colonizations into the Congo Basin during the continental radiation of African *Mastacembelus* spiny-eels. J. Biogeogr. 44, 2308–2318.
Dunz, A.R., Schliewen, U.K., 2013. Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as "Tilapia". Mol. Phylogenet. Evol. 68, 64–80.
Egger, B., Koblmüller, S., Sturmbauer, C., Sefc, K.M., 2007. Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. BMC Evol. Biol. 7, 137.
Eschmeyer, W.N., 2018. Catalog of fishes electronic version. Updated 2018. California Academy of Sciences.
Eugster, H.P., 1986. Lake Magadi, Kenya: A model for rift valley hydrochemistry and sedimentation? Sedimentation in African 2833 rifts. In: In: Frostick, L.E. (Ed.), Geol. Soc. Spec. Pap., vol. 25. pp. 177–189.
FAO., 2016. The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all. Rome.
Farias, I.P., Ortí, G., Sampaio, I., Schneider, H., Meyer, A., 1999. Mitochondrial DNA phylogeny of the family Cichlidae: monophyly and fast molecular evolution of the Neotropical assemblage. J. Mol. Evol. 48, 703–711.
Ford, A.G.P., Dasmahapatra, K.K., Rüber, L., Gharbi, K., Day, J.J., 2015. High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. Mol. Ecol. 24, 3421–3440.
Ford, A.G.P., Rüber, L., Newton, J., Dasmahapatra, K.K., Balarin, J.D., Bruun, K., Day, J.J., 2016. Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. Evolution 70, 2718–2735.
Flouri, T., Jiao, X., Rannala, B., Yang, Z., 2018. Species tree inference with BPP using genomic sequences and the multispecies coalescent. Mol. Biol. Evol. 35, 2585–2593.
Guindon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst. Biol. 52, 696–704.
Genner, M.J., Turner, G.F., 2012. Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. Mol. Biol. Evol. 29, 195–206.
Genner, M.J., Connell, E., Shechonge, A., Smith, A., Swanstrom, J., Mzighani, S., Mwijage, A., Ngatunga, B.P., Turner, G.F., 2013. Nile tilapia invades the Lake Malawi catchment. Afr. J. Aquat. Sci. 38 (Supplement 1), 85–90.
Gilks, W.R., 1996. Introducing Markov chain Monte Carlo. Markov chain Monte Carlo in practice. Springer.
Hallerman, E., Hilsdorf, A.W.S., 2014. Conservation genetics of tilapias: seeking to define appropriate units for management. Israeli J. Aquac. 2014, 1–18.
Hillaire-Marcel, C., Casanova, J., 1987. Isotopic hydrology and paleohydrology of the Magadi (Kenya) – Natron (Tanzania) basin during the Late Quaternary. Palaeogeogr. Palaeoclimatol. Palaeoecol. 58, 155–181.
Holdship, S.A., 1976. The Palaeolimnology of Lake Manyara, Tanzania: a Diatom Analysis of a 56 meter Sediment Core (Ph.D. Dissertation). Duke University.
IUCN, 2018. Global invasive species database. Available at: < http://www.iucngisd.org/gisd/ > . (Accessed April 2018).
The IUCN Red List of Threatened Species. Version 2017-3. < www.iucnredlist.org > . (Accessed April 2018).
Ivory, S.J., Blome, M.W., King, J.W., McGlue, M.M., Cole, J.E., Cohen, A.S., 2016. Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. PNAS 113, 11895–11900.
Johnson, T.C., Scholz, C.A., Talbot, M.R., Kelts, K., Ricketts, R.D., Ngobi, G., Beuning, K., Ssemmanda, I., McGill, J.W., 1996. Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. Science 273, 1091–1093.
Joly, S., 2012. jml: testing hybridization from species trees. Mol. Ecol. Resour. 12, 179–184.
Joly, S., McLenachan, P.A., Lockhart, P.J., 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174, e54–e70.
Kavembe, G.D., Machado-Schiaffino, G., Meyer, A., 2013. Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. Hydrobiologia 739, 55–71.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Klett, V., Meyer, A., 2002. What, if anything, is a Tilapia? – mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. Mol. Biol. Evol. 19, 865–883.
Kobayashi, M., Msangi, S., Batka, M., Vannuccini, S., Dey, M.M., Anderson, J.L., 2015. Fish to 2030: the role and opportunity for aquaculture. Aquac. Econ. Manage. 19, 282–300.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.
Lowe-McConnell, R.H., 1959. Breeding behaviour patterns and ecological differences between Tilapia species and their significance for evolution within the genus Tilapia (Pisces: Cichlidae). In: Proceedings of the Zoological Society of London, pp. 1–30.
Matschiner, M., Musilová, Z., Barth, J.M.I., Starostová, Z., Salzburger, W., Steel, M., Bouckaert, R., 2017. Bayesian phylogenetic estimation of clade ages supports Trans-Atlantic dispersal of cichlid fishes. Syst. Biol. 66, 3–22.
McCann, J., Schneeweiss, G.M., Stuessy, T.F., Villaseñor, J.L., Weiss-Schneeweiss, H., 2016. The Impact of reconstruction methods, phylogenetic uncertainty and branch lengths on inference of chromosome number evolution in American Daisies (*Melampodium*, Asteraceae). PLoS ONE 11, e0162299.
Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L., Seehausen, O., 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8, 14363.
Meyer, B.S., Salzburger, W., 2012. A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. Mol. Ecol. Resour. 12, 1097–1104.
Meyer, B.S., Matschiner, M., Salzburger, W., 2017. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Syst. Biol. 66, 531–550.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop, New Orleans, LA, pp. 1–8.
Mmochi, A.J., 2017. Growth Rates Of Selected *Oreochromis* Species Cultured at Different Salinities. World Aquaculture meeting, Abstract.
Moser, F.N., Rijssel, J.V., Ngatunga, B., Mwaiko, S., Seehausen, O., 2018. The origin and future of an endangered crater lake endemic; phylogeography and ecology of *Oreochromis hunteri* and its invasive relatives. Hydrobiologica. https://doi.org/10.1007/s10750-018-3780-z.
Murray, A.M., Stewart, K.M., 1999. A new species of tilapiine cichlid from the Pliocene, Middle Awash, Ethiopia. J. Vertebr. Paleontol. 19, 293–301.
Nagl, S., Tichy, H., Mayer, W.E., et al., 2001. Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. 20, 361–374.
Ndiwa, T.C., Nyingi, D.W., Agnese, J.F., 2014. An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Loboi Drainage, Kenya. PLoS ONE 9 (9), e106972.
Nevado, B., Fazalova, V., Backeljau, T., Hanssens, M., Verheyen, E., 2011. Repeated unidirectional introgression of nuclear and mitochondrial DNA between four congeneric Tanganyikan cichlids. Mol. Biol. Evol. 28, 2253–2267.
Njiru, M., Waithaka, E., Muchiri, M., van Knaap, M., Cowx, I.G., 2005. Exotic introductions to the fishery of Lake Victoria: What are the management options? Lakes Reservoirs Res. Manage. 10, 147.
Ogilvie, H.A., Heled, J., Xie, D., Drummond, A.J., 2016. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. Syst. Biol. 65, 381–396.
Ogilvie, H.A., Bouckaert, R.R., Drummond, A., 2017. StarBEAST2 brings faster species tree inference and accuarate estimation of substitution rates. Mol. Biol. Evol. 34, 2101–2114.
Pagel, M., 2004. Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53, 673–684.
Pagel, M., Meade, A., 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am. Nat. 167, 808–825.
Penk, S., Schliewen, U., Altner, M., Reichenbacher, B., 201.8 The earliest record of the Oreochromini (Cichlidae: Pisces) from a Miocene lacustrine deposit in the East African Rift Valley. 5th International Paleontological Congress. Abstract.
Philippart, J.-C., Ruwet, J.-C., 1982. Ecology and distribution of tilapias, p. 15-59. In: Pullin, R.S.V., Lowe-McConnell, R.H. (Eds.), The biology and culture of tilapias. ICLARM Conference Proceedings 7. International Center for Living Aquatic Resources Management, Manila, Philippines, pp. 432.
Pouyaud, L., Agnèse, J.F., 1995. Phylogenetic relationships between 21 species of three tilapiine genera Tilapia, Sarotherodon and Oreochromis using allozyme data. J. Fish Biol. 47, 26–38.
Core Team, R., 2018. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J., Alfaro, M.E., 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 4, 1958.
Rabosky, D.L., Chang, J., Pascal, O., Cowman, P.F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T.J., Coll, M., Alfaro, M.E., 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature 559, 392–395.
Rambaut, R., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904.
Riedel, R., Costa-Pierce, B.A., 2005. Feeding ecology of Salton Sea Tilapia (*Oreochromis* spp). Bull. Southern California Acad. Sci. 104, 26–36.
Robinson, O., Dylus, D., Dessimoz, C., 2016. Phylo.io: Interactive viewing and comparison of large phylogenetic trees on the web. Mol. Biol. Evol. 33, 2163–2166.
Rognon, X., Guyomard, R., 2003. Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. Mol. Ecol. 12, 435–445.
Ronquist, F., Teslenko, M., van der Mark, P., et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Sardella, B.A., Brauner, C.J., 2008. The effect of elevated salinity on 'California' Mozambique tilapia (Oreochromis mossambicus x O. urolepis hornorum) metabolism. Comp. Biochem. Physiol. C Toxicol. Pharmacol. 148, 430–436.
Schelly, R., Salzburger, W., Koblmüller, S., Duftner, N., Sturmbauer, C., 2006. Phylogenetic relationships of the lamprologine cichlid genus Lepidiolamprologus (Teleostei: Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. Mol. Phylogenet. Evol. 38, 426–438.
Schliewen, U.R., Klee, B., 2004. Reticulate sympatric speciation in Cameroonian crater lake cichlids. Front. Zool. 1, 5.

Schluter, T., Kohring, R., Mehl, J., 1992. Hyperostotic fish bones ("Tilly bones") from presumably Pliocene phosphorites of the Lake Manyara area, northern Tanzania. Paläontologische Zeitschrift 66, 129–136.
Schwarzer, J., Misof, B., Tautz, D., Schliewen, U., 2009. The root of the East African cichlid radiations. BMC Evol. Biol. 9, 186.
Seegers, L., 2008a. The fishes collected by G A Fischer in East Africa in 1883 and 1885/86. Zoosyst. Evol. 84, 149–195.
Seegers, L., 2008b. The fishes collected by G A Fischer in East Africa in 1883 and 1885/86. Zoosyst. Evol. 84, 49–95.
Seegers, L., Tichy, H., 1999. The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with description of two new species. Ichthyolo. Explor. Freshwaters 10, 97–146.
Seegers, L., Sonnenberg, R., Yamamoto, R., 1999. Molecular analysis of the *Alcolapia* flock from lakes Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their systematics and evolution. Ichthyolo. Explor. Freshwaters 10, 175–199.
Seegers, L., 1996. The Fishes of the Lake Rukwa Drainage. Africa Museum, Tervuren, Belgium.
Seehausen, O., Koetsier, E., Schneider, M.V., Chapman, L.J., Chapman, C.A., Knight, M.E., Turner, G.F., van Alphen, J.M., Bills, R., 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. Proc. R. Soc. Lond. B 270, 129–137.
Seehausen, O., 2007. Chance, historical contingency and ecological determinism jointly determine the rate of adaptive radiation. Heredity 99, 361–363.
Seehausen, O., 2015. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. New Phytol. 207, 304–312.
Shechonge, A., Ngatunga, B.P., Bradbeer, S.J., Day, J.J., Ford, A.G.P., Kihedu, J., Richmond, T., Mzighani, S., Smith, A.M., Sweke, E.A., Tamatamah, R., Tyers, A.R., Turner, G.F., Genner, M.J., 2019. Widespread colonisation of Tanzanian catchments by introduced *Oreochromis* tilapia fishes: the legacy from decades of deliberate introduction. Hydrobiologia 832, 235.
Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.
Stöver, B.C., Müller, K.F., 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC Bioinf. 11, 7.
Swofford, D.L., 2002. PAUP*: Phylogenetic analysis using parsimony (* and other methods), version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
Thys van den Audenaerde, D.F.E., 1968. An annotated bibliography of tilapia (Pisces, Cichlidae). Documentation zoologique No. 14, Musée royal d'Afrique centrale, Tervuren.
Trewavas, E., 1982a. Tilapias: taxonomy and speciation. In: Pullin, R., Lowe-McConnel, R.H. (Eds.), The Biology and Culture of Tilapias. ICLARM Conference Proceedings, Manila, Philippines, pp. 3–13.
Trewavas, E., Teugels, G.G., 1991. Oreochromis. In: In: Daget, J., Gosse, J.-P., Teugels, G.G., Thys van den Audenaerde, D.F.E. (Eds.), Checklist of the Freshwater Fishes of Africa (CLOFFA), vol. 4. pp. 307–346 ISNB, Brussels; MRAC, Tervuren; and ORSTOM, Paris.
Trewavas, E., 1983. Tilapiine Fishes of the genera Sarotherodon, Oreochromis and Danakilia. British Museum (Natural History), London.
Turner, G.F., Robinson, R.L., 1991. Ecology, morphology and taxonomy of the Lake Malawi *Oreochromis (Nyasalapia)* species flock. Ann. Mus. Roy. Afr. Centr. Sc. Zool. 262, 23–28.
Turner, G.F., 2007. Adaptive radiation of cichlid fish. Curr. Biol. 17, R827–R831.
Ulotu, E.E., Mmochi, A.J., Lamtane, H.A., 2016. Effect of salinity on the survival and growth of Rufiji Tilapia (*Oreochromis urolepis urolepis*) fry. WIO J. Mar. Sci. 15, 31–37.
Uyeda, J.C., Zenil-Ferguson, R., Pennell, M.W., 2018. Rethinking phylogenetic comparative methods. Syst. Biol. 67, 1091–1109.
Van Couvering, J., 1982. Fossil cichlid fish of Africa. pp. 29: 1–103. Special Papers in Palaeontology. The Palaeontological Association, London.
Vanden Bossche, J.-P., Bernacsek, G.M., 1990. Source Book for the Inland Fishery Resources of Africa Vol. 1 FAO, Rome.
Williamson, D., Taieb, M., Damnati, B., Icole, M., Thouveny, N., 1993. Equatorial extension of the younger Dryas event rock magnetic evidence from Lake Magadi (Kenya). Global Planet. Change 7, 235–242.
Willis, S.C., Farias, I.P., Ortí, G., 2014. Testing mitochondrial capture and deep coalescence in amazonian cichlid fishes (Cichlidae: *Cichla*). Evolution 68, 256–268.
Yu, G., Smith, D., Zhu, H., Guan, Y., Lam, T.T., 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36.
Yue, G.H., Lin, H.R., Li, J.L., 2016. Tilapia is the fish for next-generation aquaculture. Int. J. Marine Sci. Ocean Technol. 3, 11–13.
Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion. Ph.D. dissertation. The University of Texas at Austin.