



Ford, A. G. P., Bullen, T. R., Pang, L., Genner, M. J., Bills, R., Flouri, T., Ngatunga, B. P., Rüber, L., Schliewen, U. K., Seehausen, O., Shechonge, A., Stiasny, M. L. J., Turner, G. F., & Day, J. J. (2019). Molecular phylogeny of *Oreochromis* (Cichlidae, Oreochromini) reveals mito-nuclear discordance and multiple colonisation of adverse aquatic environments. *Molecular Phylogenetics and Evolution*, 136, 215-226. <https://doi.org/10.1016/j.ympev.2019.04.008>

Peer reviewed version

License (if available):  
CC BY-NC-ND

Link to published version (if available):  
[10.1016/j.ympev.2019.04.008](https://doi.org/10.1016/j.ympev.2019.04.008)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://www.sciencedirect.com/science/article/pii/S1055790318305062>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/>

1 **Running title: Molecular phylogeny of *Oreochromis* cichlid fishes**

2

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

6

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

10

11 <sup>a</sup>Department of Genetics, Evolution and Environment, University College London, Gower Street, London,  
12 WC1E 6BT, U.K.

13 <sup>b</sup> Current address: Department of Life Sciences, Whitelands College, University of Roehampton,  
14 Holybourne Avenue, London, SW15 4JD, U.K.

15 <sup>c</sup> School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol.  
16 BS8 1TQ, U.K.

17 <sup>d</sup> South African Institute for Aquatic Biodiversity, Private Bag 1015, 6140 Grahamstown, South Africa

18 <sup>e</sup> Tanzania Fisheries Research Institute (TAFIRI) PO. Box 9750. Dar es Salaam. Tanzania

19 <sup>f</sup> Naturhistorisches Museum Bern, Bernastrasse 15, 3005 Bern, Switzerland

20 <sup>g</sup> Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse  
21 6, 3012 Bern, Switzerland

22 <sup>h</sup> Bavarian State Collection of Zoology, Department of Ichthyology, Münchhausenstr. 21, 81247  
23 München, Germany

24 <sup>i</sup> Department of Fish Ecology and Evolution, Center for Ecology, Evolution and Biogeochemistry, EAWAG  
25 Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland

26

27 <sup>j</sup> Department of Aquatic Sciences and Fisheries, University of Dar es Salaam, P.O. Box 35064, Dar es  
28 Salaam, Tanzania

29 <sup>k</sup> Department of Ichthyology, American Museum of Natural History, New York, NY, USA

30 <sup>l</sup> School of Biological Sciences, Bangor University, Bangor, LL57 2UW, UK.

31

32 \* E-mail address: [j.day@ucl.ac.uk](mailto:j.day@ucl.ac.uk), [antonia.ford@roehampton.ac.uk](mailto:antonia.ford@roehampton.ac.uk)

33

34 Abstract

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

55

56 Keywords: *Tilapia*; *Alcolapia*; ancestral state reconstruction; mito-nuclear discordance; introgression;  
57 incomplete lineage sorting; taxonomic revision.

58

## 59 Introduction

60 The propensity for African cichlids to form adaptive radiations within lacustrine environments has  
61 received considerable research attention (Turner 2007, Seehausen 2015), but the processes promoting  
62 diversification within largely riverine lineages are less well known. One of the most species rich and  
63 widely distributed lineages of African cichlids is the oreochromine group. Oreochromine cichlids have  
64 significant scientific and socio-economic importance both within and beyond their native range –  
65 providing a major food source for fisheries and aquaculture, a biocontrol agent for aquatic plants and,  
66 together with other cichlid groups, they are a ‘model’ system for evolutionary research (Kobayashi et al.  
67 2015; Yue et al. 2016; Brawand et al. 2014). However, the phylogenetic relationships of the species of  
68 the most diverse group - *Oreochromis* Günther 1889 - are poorly understood.

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

79 Oreochromine cichlids are attractive targets for aquaculture, with at least eight species actively  
80 farmed globally. The Nile tilapia (*Oreochromis niloticus*) is one of the most widely farmed aquaculture  
81 species globally, but other species of *Oreochromis* are farmed in many local regions on a smaller scale,

82 and are also important capture fisheries species (FAO 2016). However, such aquaculture and capture  
83 fishery improvement initiatives can have significant environmental impact where fish escape and  
84 establish local populations. Many species of *Oreochromis* have a substantial invasive potential,  
85 exhibiting trophic plasticity that enables broad resource use, and a tendency to hybridise (Genner et al.  
86 2013; Shechonge et al. 2019), and now four species of *Oreochromis* are listed on the Global Invasive  
87 Species Index (IUCN 2018). Despite certain species exhibiting strong invasive potential, some species  
88 within the the genus are severely range restricted and thus are threatened, with 19 species (50%)  
89 categorised as "Near Threatened to Critically Endangered", of which seven species are listed as  
90 Critically Endangered (IUCN Red List, v. 2017/3).

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

102         The soda lake cichlid species in the genus *Alcolapia* were originally described as a member of  
103 *Tilapia* and subsequently included by Trewavas (1983) as subspecies within a subgenus of *Oreochromis*,  
104 as *O. (Alcolapia) alcalicus alcalicus* and *O. (Alcolapia) alcalicus grahami*. Seegers et al. (1999) elevated  
105 the subgenus *Alcolapia* to genus based on mtDNA sequence data, and later revised the species name  
106 *alcalicus* to *alcalica* to agree with the feminine genus (Seegers 2008). However, while molecular  
107 analyses have consistently resolved a monophyletic *Alcolapia*, that is shown to nest within *Oreochromis*,  
108 the specific relationship to *Oreochromis* is unresolved, with various taxa having been hypothesised to  
109 be the sister species to the *Alcolapia* clade: *O. amphimelas* (Seegers *et al.* 1999; Nagl *et al.* 2001), *O.*

110 *malagarasi* (Seegers *et al.* 1999), *O. tanganyicae* (Schwarzer *et al.* 2009, Dunz & Schliewen 2013) and *O.*  
111 *variabilis* (Kavembe *et al.* 2013, Matschiner *et al.* 2017, Rabosky *et al.* 2018). Other than *O. amphimelas*,  
112 the other three species previously suggested as alternative sister taxa to *Alcolapia* (i.e. *O. malagarasi*, *O.*  
113 *tanganyicae*, *O. variabilis*) are not found in soda lake conditions. Based on the lack of a densely sampled  
114 phylogeny for *Oreochromis* and *Alcolapia*, we generated a comprehensive phylogeny for the group using  
115 multiple markers from the mitochondrial and nuclear genomes with near-complete coverage of sampled  
116 loci.

117

#### 118 Adaptation to adverse environments

119 Of the 44 currently recognised species and subspecies of *Oreochromis*, excluding *Alcolapia*, nine are  
120 known to tolerate or occur in environments with elevated salinity or temperature, but only *O.*  
121 *amphimelas* and *Alcolapia* have adapted to the high alkalinity and low dissolved oxygen levels found in  
122 soda lake conditions (see Table 1). Tolerance to temperature and salinity are intrinsically linked, with  
123 lower temperature ranges tolerated in saline conditions than in freshwater in some species of tilapia  
124 (reviewed in Philippart & Ruwet 1982), and these two parameters are considered to be the determining  
125 factors in the distribution of several species. Several species of *Oreochromis* are euryhaline and  
126 acclimatise to a range of levels of salinity (from freshwater through to seawater), including *O. urolepis*  
127 and *O. mossambicus* that naturally occur in estuarine conditions and have been widely used in  
128 aquaculture due to their salinity tolerance (e.g., Riedel & Costa-Pierce 2005; Sardella & Brauner 2008;  
129 Ulotu *et al.* 2016). The typically freshwater *O. niloticus* (composed of seven subspecies) includes  
130 members that have successfully colonised thermal hot springs (Bezault *et al.* 2007, Ndiwa *et al.* 2014),  
131 while other *Oreochromis* species thrive in the high-pH volcanic springs feeding the soda lakes of East  
132 Africa (Trewavas 1983). In particular, *O. amphimelas* occurs in the springs and main water bodies of the  
133 seasonal soda-like Lakes Manyara, Eyasi, Sulungali, Kitangiri, and Singida, Tanzania. This species co-  
134 occurs with introduced populations of *O. niloticus* in the Sulungali, Kitangiri and Singida lakes and  
135 alongside introduced populations of *O. esculentus* in Kitangiri and Singida. The genus *Alcolapia* has  
136 diversified in the even more extreme environment of the nearby soda lakes: Lakes Natron in Tanzania  
137 and Magadi in Kenya (Seegers *et al.* 1999, Ford *et al.* 2015, 2016) and occurs in alkaline hydrothermal

138 springs feeding the lakes. These hot, hyper saline and highly alkaline soda springs represent some of the  
139 most extreme aquatic environments known supporting fish life. It is unclear whether adaptation to soda  
140 waters occurred early in the evolution of this lineage or arose more recently; and whether it arose only  
141 once or multiple times. Resolving the species relationships of the genera *Oreochromis* and *Alcolapia* will  
142 enable these hypotheses to be tested.

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

149

## 150 2. Materials and methods

### 151 2.1 Species sampling

152 A total of 28 species from 33 described species of *Oreochromis*, as well as all four species of *Alcolapia*  
153 (using taxonomic data compiled by Eschmeyer et al. 2018), are included in this study. Multiple wild  
154 samples of each species, and where possible subspecies, from across Africa were included totalling 105  
155 samples (Appendix A, Table S1). Specimens were collected mainly using seine nets, with specimens  
156 euthanised with an overdose of clove oil and subsequently preserved in 70-80% ethanol. Fin-clips or  
157 muscle tissue were taken for molecular analysis and were preserved in 96-100% ethanol. While this  
158 study endeavoured to produce a comprehensive phylogeny of all species of *Oreochromis*, five species  
159 could not be obtained, including *O. aureus* (although common in some countries as a food fish, we only  
160 included wild samples with known localities), *O. ismailiaensis*, *O. lidole*, *O. saka*, and *O. spilurus*. However,  
161 of these species we believe *O. lidole* to be functionally extinct based on our personal extensive field  
162 observations (GFT, MJG), and it is likely that *O. saka* is a geographic variant (and junior synonym) of *O.*  
163 *karongae* (Turner & Robinson 1991). *Oreochromis ismailiaensis* has also not been seen since the original  
164 description and the type locality is reported to have been converted to a concrete channel devoid of fish  
165 life (A. Dunz pers. comm.), so this too may be extinct. As such (excluding *O. lidole*, *O. ismailiaensis* and *O.*

166 *saka* as extinct or invalid species), our study represents ~93% of all *Oreochromis* species, and all  
167 *Alcolapia* species (94% when *Oreochromis* and *Alcolapia* are combined). We included two *Sarotherodon*  
168 species (*S. galilaeus* and *S. mvogoi*) based on their close association to *Oreochromis* and *Alcolapia* (Dunz  
169 and Schliewen 2013), and the more distantly related *Coptodon rendalli* as an outgroup taxa. For the  
170 mtDNA dataset only, a further 16 samples were included based on ND2 GenBank sequences (Table S1).  
171 These included samples referred to as *Oreochromis aureus* (three samples), as well as coverage of other  
172 Oreochromine genera, including five of the ten genera (Dunz and Schliewen 2013), *Iranocichla*, *Konia*,  
173 *Sarotherodon* (eight further species), *Stomatepia*, and *Tristramella*, allowing us to investigate the  
174 monophyly of *Oreochromis*.

175

## 176 2.2 Molecular markers

177 As cichlids are known to exhibit mito-nuclear discordance (e.g., Rognon & Guyomard 2003; Seehausen  
178 2003; Schliewen & Klee 2004; Egger et al. 2007, Alter et al. 2017), we selected six nuclear loci to target  
179 a presumed rapidly evolving clade based on the age of the oldest member of this tribe (Dunz & Schliewen  
180 2013). These included the recently developed exon-primed intron crossing (EPIC) markers BMP4,  
181 CCNG1, GAPDHS, TYR, b2m (Meyer & Salzburger, 2012) and the nuclear intron S7 intron 1 (Chow &  
182 Hazama, 1998), which has previously been used in cichlid studies (Schelly et al., 2006; Schwarzer et al.,  
183 2009). We also selected two mitochondrial (mtDNA) genes: NADH dehydrogenase 2 (ND2), frequently  
184 used in cichlid phylogenetics (e.g. Klett & Meyer, 2002; Day et al., 2007; Schwarzer et al., 2009; Dunz &  
185 Schliewen, 2013), and 16S rRNA (e.g. Farias et al., 1999; Schwarzer et al., 2009) to determine if there  
186 was discordance between nuclear and mitochondrial datasets. We sequenced all samples where  
187 possible (see Appendix A, Table 1).

188

## 189 2.3 DNA Extraction, amplification and sequencing

190 Genomic DNA was extracted from fin clip samples or muscle tissue stored in 95% ethanol using the  
191 Qiagen DNeasy Blood and Tissue kit. The molecular markers were amplified using the Polymerase Chain  
192 Reaction. 1µl of the DNA extraction was added to 12.5µl of MyTaq™ Mix (Bioline, UK), 9.5µl of water,  
193 and 1µl each of the 10M forward and reverse primers (Sigma-Aldrich, UK), to give a 25µl total reaction



194 volume. The primers and reaction conditions for each gene are shown in Appendix A, Table S2. Cleaned  
195 PCR products were sequenced on a 3730xl DNA Analyser (Applied Biosystems).

196

#### 197 2.4 Alignment and partitioning scheme

198 Sequence data was edited using Sequencher 5.4 (Gene Codes Corporation, Ann Arbor, MI USA) and  
199 GENEIOUS v. 6.0.6 (Biomatters) (Kearse et al. 2012) in which contigs were assembled from the forward  
200 and reverse sequences. Sequences were subsequently aligned using MUSCLE in GENEIOUS (Kearse et  
201 al. 2012) using default parameters, with alignments checked for stop codons and reading frame shifts.  
202 The concatenated nuclear dataset (101 samples) included a total of 3092 bp: BMP4 (482 bp), CCNG1  
203 (650 bp), GAPDHS (436 bp), TYR (561 bp), b2m (482 bp) and S7 (481 bp). We, however, removed a  
204 hyper-variable region (42 bp) from exon 2 of the b2m loci (from 144 and 185 bp) as it was shown to  
205 contain two dominant haplotypes and a high density (14) of variable sites. As the possible causes of  
206 patterns like this include strong selection, introgression from more divergent species, or paralogous  
207 sequences (although our sequences of b2m aligned against a previous dataset for this region [Meyer &  
208 Salzburger, 2012]), they would likely erroneously influence phylogenetic reconstruction. The final  
209 dataset for the nuclear data was therefore 3050 bp. The mitochondrial dataset (116 samples) included  
210 a total of 1582 bp: ND2 (1047 bp) and 16S (535 bp), with each dataset analysed separately.

211 The optimal partitioning scheme and model choices were assessed with PartitionFinder v.2.1.1 (Lanfear  
212 et al., 2017) using the greedy algorithm (Huelsenbeck and Ronquist, 2001) and assessed using the  
213 Bayesian Information Criteria (BIC). For the nuclear concatenated analyses we defined two subsets for  
214 the nuclear (nuDNA) dataset (combining introns vs. exons) and four subsets for the mitochondrial  
215 (mtDNA) dataset (each codon position for ND2, plus 16S). For the species tree analyses it was more  
216 appropriate to treat the six nuclear loci as separate genes (as opposed to the exon vs. intron partitions  
217 in the concatenated analysis), as the most important factor in species tree analysis is variation in the  
218 gene tree, as genes will (more or less) have different histories. We also checked to see if implementing  
219 the partitions by loci made any difference to the outcome of the nuclear concatenated analysis. To obtain  
220 evolutionary models for these partitions we re-ran PartitionFinder. Resulting models (see Table 2) were  
221 implemented in subsequent phylogenetic analyses.

## 222 2.5 Phylogenetic inference

223 Phylogenetic analyses were run on 1) the concatenated nuDNA dataset, and 2) the concatenated mtDNA  
224 data. Each dataset was analysed using Bayesian Phylogenetic Inference (BI) and Maximum Likelihood  
225 (ML), and the nuDNA dataset was also analysed using a Bayesian multispecies coalescent approach.  
226 Bayesian Phylogenetic Inference was implemented using MrBayes v.3.2.6 (Ronquist et al. 2012) in  
227 which analyses were run for 50,000,000 generations using four Markov chains (three heated, one cold,  
228 heating parameter 0.4) with default priors, implementing the models as defined by PartitionFinder.  
229 Maximum Likelihood analyses were run using GARLI v.2.01 (Zwickl 2006), again implementing the  
230 models defined by PartitionFinder, and running 100 bootstrap (BS) replicates. All analyses were run on  
231 CIPRES Science Gateway server (Miller et al., 2010). The convergence of MCMC runs and burn-in were  
232 assessed in Tracer v.1.7 (Rambaut et al., 2018) and FigTree v.1.4 (Rambaut, 2009) was used to visualise  
233 trees.

234

## 235 2.6 Non-parametric likelihood-based tests

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

243

## 244 2.7 Calibration selection and species trees analyses

245 Although our study was not focused on investigating divergence dates, generation of a species tree to  
246 investigate species relationships and trait evolution required the selection of calibrations. The fossil

247 record of the Oreochromini is represented by several fossils. The oldest of these, †*Sarotherodon martyni*  
248 (Van Couvering, 1982) from the Ngorora Formation, Lake Turkana, Kenya (late Miocene: 12.0–9.3 Ma),  
249 is considered to belong to *Oreochromis* (Murray and Stewart 1999), although this affinity has been  
250 debated (Carnevale et al. 2003). A recent discovery of eight well preserved fossil skeletons from the  
251 nearby site, Kabchore, which is middle Miocene (12.5 Ma) are attributed to *Oreochromis* based on  
252 unique character combinations and meristic traits (Penk et al. 2018). Based on these fossils (and the  
253 occurrence of †*Sarotherodon martyni*), the older age is used to constrain *Oreochromis* and is preferred  
254 here to †*Oreochromis lorenzoi* (Carnevale et al. 2003) from the upper Miocene Messinian (7.246 Ma and  
255 5.333 Ma), which has previously been applied as a constraint (Schwarzer et al. 2009). As the fossil record  
256 represents a minimum age, we also used a secondary calibration from Rabosky et al. (2018), which is  
257 based on a large teleost wide dataset and calibrated with a larger number of fossils, to provide a  
258 maximum age (14.71 Ma) for this node.

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

270 A further \*BEAST analysis of all loci (nuclear and mitochondrial DNA) was also performed as this is  
271 required for running JML software (see section 2.8). All settings were the same as described for the  
272 nuclear tree species analysis. The mtDNA data was treated as a single locus (see Table 2) in which a  
273 strict clock was preferred as indicated by the uclStdev value which was <1. For each dataset (i.e.

274 nuDNA, and nuDNA + mtDNA) three analyses were run with chain length 500,000,000, and convergence  
275 was checked in Tracer v.1.7, as for the MrBayes analysis. Most ESS values for the nuDNA \*BEAST  
276 analyses were >200, with treePriors and clock rates >100, but some values (posterior, species  
277 coalescent and popMean) were between 96-93. All ESS values for the nuDNA + mtDNA species tree were  
278 >200. Runs were combined using TreeAnnotator, and resampled, discarding 25% burnin, and visualised  
279 using FigTree v.1.4.3.

280 We also used the program Bayesian Phylogenetics and Phylogeography (BPP) v 4.0 (Flouri et al.  
281 2018) (method A01) and performed runs of 30,000,000 MCMC iterations testing alternative priors. The  
282 calibrations previously described were also applied to this analysis. Although we did not obtain  
283 convergence across the runs on the maximum a posteriori tree (MAP) we obtained convergence on the  
284 majority rule consensus (MRC) tree, which was identical across all runs. We tried two priors on the root  
285 age (tau prior) 1) Inv-Gamma (3,0.002), and 2) Inv-Gamma (2,0.05), and both priors yielded the same  
286 (MRC) trees.

287

## 288 2.8 Testing for hybridisation

289 Due to discordance between mitochondrial and nuclear trees (see Results 3.1) we attempted to  
290 investigate if we could distinguish between introgression and incomplete lineage sorting (ILS) using the  
291 method described by Joly et al. (2009), implemented in JML v.1.3.1 (Joly 2012). This method uses  
292 posterior predictive checking by comparing the minimum sequence distance between two species to  
293 test if the minimum distance is smaller than expected under a regime not accounting for hybridisation.  
294 However, we acknowledge that our dataset does not include many samples per species, and that loci  
295 from our nuclear dataset are short, and therefore lack power of longer sequences (see Joly et al. 2009).  
296 As this method uses the posterior distributions of species trees, population sizes and divergence times,  
297 we generated a further coalescent tree from all loci (see section 2.7), ensuring that all loci were unlinked  
298 to obtain mutation rates needed for JML. JModelTest v.2.1.10 (Guindon and Gascuel, 2003; Darriba et al.  
299 2012) was implemented to obtain models for individual loci and also to obtain settings for the .jml.ctl

300 file of JML (state frequencies, transition-transversion ratio, proportion of invariable sites, gamma rate  
301 heterogeneity). We ran 1000 simulations for each of the three selected loci, mtDNA (locus rate = 0.0038),  
302 BMP4 (locus rate = 0.00036) and TYR (locus rate = 0.000656) pairwise distance comparisons. A  
303 Benjamini-Hochberg correction was applied to the results using R v.3.5 (R Core Team, 2018).

304

## 305 2.9 Trait analyses

306 We tested for correlated evolution of traits using BayesTraits v3.0.1 (Pagel et al. 2004; Pagel & Meade  
307 2006), which was also used for ancestral state reconstruction. We focused on trait data for tolerance to  
308 soda conditions (salinity, temperature, and pH) with species states derived from the literature (see  
309 Table 1). As tolerance to elevated salinity and temperature was most common, we tested for correlated  
310 evolution between these traits, and separately reconstructed ancestral states for soda adaptation,  
311 defined as comprising all three tolerance traits found in a species: salinity, temperature, and pH. We also  
312 examined the correlated evolution of phenotypic secondary sexual male characteristics that have been  
313 used in previous taxonomic analysis: the genital tassel, and extended male jaws. Although the two traits  
314 have not previously been implicated to have correlated evolution, the two traits are used as diagnostic  
315 characters to separate clades and subgenera of *Oreochromis*, and the traits do not co-occur, so we tested  
316 whether their presence/absence was correlated. We ran the ancestral state reconstruction analyses on  
317 the nuclear datasets only, using both the MrBayes (non-ultrametric) and \*BEAST (ultrametric)  
318 phylogenies. We did not ultrametricise the MrBayes tree, as branch lengths can have a substantial effect  
319 on ancestral state reconstructions (McCann et al. 2016), and following recommendations from  
320 Cusimano & Renner (2014) to run reconstructions on more than one type of branch length depiction we  
321 therefore include both the phylogram and ultrametric trees. The MrBayes nuclear concatenated tree  
322 was pruned to include only one sample per species (or subspecies, where relevant). One sample each  
323 was included for subspecies of three species (specifically: *O. niloticus niloticus*; *O. niloticus cancellatus*;  
324 *O. niloticus filoa*; and *O. placidus placidus*; *O. placidus ruvumae*; and *O. shiranus shiranus*, *O. shiranus*  
325 *chilwae*), as plausibly the subspecies may exhibit different adaptations / morphology from each other  
326 and so were coded separately. The resulting tree contained 40 tips (including subspecies of *Oreochromis*,

327 species of *Alcolapia*, and the three outgroup species). Details of the specimens included in the pruned  
328 tree are given in Appendix A, Table S3. No pruning was required for the \*BEAST species tree as there  
329 was only one tip per species (total of 36 tips), which also did not include all the respective subspecies of  
330 *O. shiranus* and *O. niloticus* as the availability of only single samples meant that they could not be  
331 included in the multispecies coalescent approach. The BayesTraits analysis tested the tolerance (salinity  
332 and temperature) and morphological characters separately, using the Discrete analysis mode (testing  
333 two binary characters) and comparing the independent model (no correlation of shifts between the two  
334 traits) with dependent (correlated shifts) models (see Table S3 for how traits were coded per species).  
335 Each analysis (independent and dependent) was run for 5 independent runs, with 10 million MCMC  
336 iterations, with the first million discarded as burnin, and the stepping stone sampler set to use 100  
337 stones and run each stone for  $10^4$  iterations. Tree branch lengths were scaled to a mean of 0.1 using the  
338 ScaleTrees command, following the recommendations of the BayesTrait manual. Priors for the rate  
339 parameters (the rate coefficient for the gain/loss of each trait respectively) were set based on initial  
340 Maximum Likelihood runs in BayesTrait. Based on the ML rates estimates, the MCMC analysis on the  
341 MrBayes phylogeny used a hyperprior for all rate coefficients specifying an exponential prior seeded  
342 from a uniform distribution on the interval 0 to 10, and those on the \*BEAST trees used an exponential  
343 prior seeded from a uniform distribution on the interval 0 to 1. Mixing of the chains was checked in the  
344 output files to ensure acceptance rates were in the range 20-40%; convergence and ESS were checked  
345 using Tracer v1.6. The results for the independent and dependent models based on the marginal  
346 likelihood from the stepping stone sampler (expressed on a natural log scale) were compared using Log  
347 Bayes Factors calculated as:

$$348 \text{ Log Bayes Factors} = 2(\log \text{ marginal likelihood complex model} - \log \text{ marginal likelihood simple model})$$

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

352 The tests of correlated evolution suggested that tolerance to salinity and temperature were  
353 correlated traits (Results Section 3.3). However, as salinity and thermal tolerance were predominantly

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

366

### 367 3. Results

#### 368 3.1 Phylogenetic relationships

369 The nuclear species tree (Figure 1), BI (Supplementary Figure S1) and ML concatenated (data not  
370 shown, but BS values included on Figure S1) trees were reasonably similar in topology, particularly  
371 regarding subclades, although there were several instances of taxa occurring in alternative positions  
372 between the concatenated and species trees, most notably the sister group of *Alcolapia* (discussed  
373 below). The mtDNA concatenated trees generated using BI and ML were largely congruent (see Figures  
374 1b, Supplementary Figures S2). However, comparison of nu- and mtDNA trees showed high levels of  
375 mito-nuclear discordance from the strikingly different placement of certain clades and taxa (Figure 1),  
376 and there are instances where different groups appear to be well resolved and monophyletic in trees  
377 built from the different sets of loci (Figure 1, Supplementary Figures S1, S2). The AU test based on the  
378 ML concatenated trees resulted in the mtDNA tree (-ln L 7024.31) being significantly worse fit to the  
379 data ( $P < 0.001$ ) than the nuclear tree (-ln L 6516.82), but we discuss both topologies below. As  
380 mitochondrial data is prone to introgression in cichlids, we used the nuclear species trees for

381 downstream analyses. A multispecies coalescent approach is preferred as it accounts for gene tree-  
382 species tree incongruence that arise due to population level processes, and is particularly suitable for  
383 more recently diverged groups (e.g. Ogilvie et al. 2016, 2017 and refs therein), although we acknowledge  
384 that the nuclear genome may also be introgressed in this group of fishes.

385  
386 *The sister group of Alcolapia*

387 Both the \*BEAST (Figure 1a) and BPP (Supplementary Figure S3) versions of the nuDNA species tree  
388 resolved *O. esculentus* (a freshwater species) as the sister group to *Alcolapia* (extreme soda-lakes), with  
389 *O. amphimelas* (seasonal soda-lakes) as sister to these lineages. This is however not well supported (0.55  
390 / 0.69 respectively), and a preliminary \*BEAST analysis including the 42 bp hyper-variable data  
391 (removed from subsequent analyses due to high variability) conversely resolved *O. amphimelas* as the  
392 sister group (data not shown). The nuDNA concatenated phylogenies (Figure 1b, Appendix A Figure  
393 S1) (with or without the 42 bp region) resolved the sister group to the *Alcolapia* clade as *O. amphimelas*  
394 (BPP 1/0.90 [concatenated/species tree]; BS 91), with *O. esculentus* resolved (BPP 1/0.90; BS 89) as  
395 sister to the *Alcolapia* + *O. amphimelas* group. All these species are closely distributed geographically  
396 occurring in NW Tanzania and SW Kenya. The nuDNA concatenated tree (Appendix A Figure S1) shows  
397 that there is also some resolution within the *Alcolapia* clade itself, with the geographically isolated *A.*  
398 *grahami* (Lake Magadi, Kenya) resolved (BPP 0.99) and most closely related to the 'northern' *A. alcalica*  
399 from Lake Natron, although there is poor resolution regarding the relationships among the three  
400 sympatric Lake Natron species ('southern' *A. alcalica*, *A. latilabris*, *A. ndalalani*). In contrast, the mtDNA  
401 concatenated phylogeny (Figure 1c, Appendix A Figure S2) placed the *Alcolapia* clade as the sister  
402 group to (BPP 0.86/BS 56) a mixed assemblage of six species (including *O. esculentus*), which are not  
403 themselves phylogenetically resolved, with *O. amphimelas* resolved as the sister taxon to this clade.  
404 There is also no resolution of the relationships among the constituent species of *Alcolapia*. When the  
405 ND2 data alone is analysed, *Alcolapia* + *O. amphimelas* are sister taxa, albeit with low support, and the  
406 'mixed assemblage' is the sister group to this clade (data not shown) indicating differing signals from  
407 the mtDNA loci.



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

427 However, there are areas of some congruence between the trees built from mtDNA versus  
428 nuDNA. In particular, species from the Lake Malawi catchment and the formerly connected Ruvuma  
429 (also known as the Rovuma) catchment (*O. shiranus*, *O. squamipinnis*, *O. karongae*, *O. chunguruensis*, *O.*  
430 *placidus ruvumae*) form a well supported clade in both the mtDNA and nuDNA species trees. The  
431 concatenated nuclear tree supports a variation of this grouping, with the exception of *O. placidus*  
432 *ruvumae*, which conversely groups within a larger clade that is sister to Lake Malawi catchment species.  
433 In both nuclear trees (species and concatenated), *O. placidus* (Zambezi river) is also a member of this  
434 group, whereas this taxon groups with other largely southern African species in the mtDNA tree.

435 *Pangani relationships*

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

447

448 *Oreochromis monophyly*

449 The inclusion of samples from GenBank within the mtDNA analysis (see Supplementary Figures S2)  
450 allowed us to test the monophyly of *Oreochromis*. With the exception of sequences uploaded as *O. aureus*  
451 (for which we did not have access to voucher specimens), *Oreochromis* was resolved as monophyletic,  
452 or rather as paraphyletic with *Alcolapia* nested within it. The position of *O. aureus* (a Nilo-Sudanic  
453 species) grouped within the Lake Barombi Mbo radiation, and outside of *Oreochromis* likely implies that  
454 the sampled taxa are hybrids or were originally mis-identified. However, although *Oreochromis aureus*  
455 has not been included in recent molecular studies of tilapiine relationships (e.g. Schwarzer et al 2009;  
456 Dunz and Schliewen 2013), an allozyme study previously resolved *O. aureus* at the base of the  
457 *Oreochromis* clade (Pouyaud and Agnèsè 1995). Of the four GenBank *O. aureus* samples used to sequence  
458 NADH, only one is in a published paper (DQ465029.1; Cnaani et al. 2008), and was collected from stocks  
459 of *O. aureus* in Israel that were originally sourced from Lakes Hula (Israel) and Manzala (Egypt), both of  
460 which are also inhabited by *S. galilaeus*. Trewavas (1983) notes several characters that distinguish *S.*  
461 *galilaeus* from co-occurring *O. aureus* and *O. niloticus*, including the depth of the pre-orbital bone.

462 Intergeneric hybrids between *Oreochromis* and *Sarotherodon* are viable and have been produced in  
463 aquaculture strain development (using *in vitro* fertilisation; Bezault et al. 2012), but we are not aware  
464 of any reports of intergeneric crosses in the wild, especially in natural sympatry zones. Further samples  
465 of *O. aureus* are required to investigate this placement fully and ensure that *S. galilaeus* samples have  
466 not been mis-identified as *O. aureus*.

467

468

### 469 3.2 Assessment of hybridisation

470 Results from the JML analysis revealed no support for introgression, with no significant signal of  
471 hybridisation after applying the Benjamini-Hochberg correction in any of the loci tested (mtDNA, TYR,  
472 BMP4). These results suggest that incomplete lineage sorting could explain the incongruence in mtDNA  
473 and nuclear datasets (rather than a signal introgression). However, the large number of pairwise  
474 comparisons (630) mean that testing across the entire phylogeny is unlikely to uncover signals of  
475 hybridisation, and future analyses may focus on specific pairwise comparisons of interest. We suggest  
476 that additional data would need to be included to help refine this analysis, but that ultimately  
477 examination of species of *Oreochromis* using genome-wide data would provide a powerful approach to  
478 test hybridisation hypotheses.

479

### 480 3.3 Diversification and ancestral state reconstruction

481 The BayesTraits analysis for environmental tolerance traits (salinity and thermal tolerance) using the  
482 Discrete model for the nuclear \*BEAST species tree gave a marginal log likelihood for the independent  
483 model of -26.50 (with a standard deviation of 0.04 across 5 runs), while the dependent model had a  
484 marginal log likelihood of -24.72 (s.d. 0.02). These resulted in a log Bayes Factor of 3.56, indicating  
485 moderate support for the dependent model, suggesting that the shifts in adaptation to salinity and  
486 thermal tolerance may be correlated. Results when using the nuclear concatenated MrBayes tree were  
487 more conclusive, with the 5 runs resulting in a log Bayes Factor of 6.78, indicating strong support for  
488 correlation of the traits. For the morphological traits (genital tassel and extended male jaw morphology)  
489 using the nuclear \*BEAST species tree the independent model marginal log likelihood was -39.82 (s.d.

490 0.06) while that for the dependent model was -39.18 (s.d. 0.04). The log BF was 1.30, indicating that  
491 there was no support for the complex model (dependence of trait shifts), and suggesting that the two  
492 phenotypic traits do not exhibit correlated rate shifts. Results when using the nuclear concatenated  
493 MrBayes tree were similar, with the 5 runs resulting in a log Bayes Factor of -0.22, indicating no support  
494 for correlation of the phenotypic traits.

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

508

#### 509 4. Discussion

510 Here, we present the first comprehensive phylogenetic analysis of the cichlid genus *Oreochromis* using  
511 multi-marker molecular datasets comprising nuclear and mitochondrial loci, and reveal high levels of  
512 incongruence between them. This incongruence could either represent incomplete lineage sorting  
513 and/or introgression, and while we did not find a signal for the latter mechanism it cannot be ruled out.  
514 Incongruence as a result of these mechanisms is commonly reported in other cichlid studies (e.g.,  
515 Genner & Turner 2012; Willis et al. 2013; Meier et al. 2017). We suggest that the nuDNA phylogeny is  
516 likely to be a better estimate of the species tree, although we acknowledge that additional data would  
517 help to resolve conflicting relationships. Based on the nuDNA trees, we show that adaptation to adverse

518 environmental conditions (i.e. increased salinity and temperature) has occurred multiple times, but that  
519 adaptation to extreme (soda) conditions is likely to have occurred once. Irrespective of the dataset  
520 (nuDNA or mtDNA), we demonstrate that the *Alcolapia* clade is consistently resolved within the genus  
521 *Oreochromis*.

522

#### 523 4.1 Taxonomy of the genus *Oreochromis*: effects of mito-nuclear discordance

524 We observed substantial mito-nuclear discordance in the phylogenetic analysis (Figure 1). We note that  
525 several of the relationships in the nuclear analysis are concordant with previous phenotypic  
526 classifications based on phenotypic features, but that certain relationships within the mtDNA analysis  
527 are also plausible based on phenotypic and geographical range data. The results suggest reticulate  
528 evolution could have played a role in the existing genetic relationships, and could be a result of the fact  
529 that many tilapiine species are known to have been widely translocated for stocking and aquaculture  
530 purposes, although we do not find significant evidence of introgression using JML analysis. Alternative  
531 explanations include incomplete lineage sorting that is widely reported in cichlids (e.g., Genner & Turner  
532 2012; Meier et al. 2017; Meyer et al. 2017). Previous analysis of mito-nuclear discordance in the tribe  
533 Oreochromini suggested that ancient hybridisation was the more probable explanation of this  
534 discordance (Dunz & Schliewen 2013).

535 The translocation and establishment of non-native species (*O. leucostictus*, *O. niloticus*) within  
536 Tanzania has been recently documented (Shechonge et al. 2018). We note two additional non-native  
537 species occurrences in Tanzania from our sampling: *O. leucostictus* (AF042-04) sampled from fish ponds  
538 in the Lake Eyasi basin (reportedly stocked from streams surrounding Lake Eyasi), and *O. urolepis*  
539 (AF014-01) sampled from irrigation canals draining into Lake Manyara. Male specimens collected from  
540 the latter site also exhibited the distinctive extended jaw morphology of sexually mature male *O.*  
541 *urolepis*. To our knowledge, this is the first report of *O. urolepis* sampled in a natural water body outside  
542 of its native catchments, that includes the Wami, Ruvu and Rufiji basins. However, we are aware of  
543 aquaculture centres rearing *O. urolepis* in Tanga (Mmochi 2017).

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

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

568 *Alcolapia* is consistently resolved within *Oreochromis* irrespective of dataset and forms a strongly  
569 supported clade with *O. amphimelas* and *O. esculentus* irrespective of nuclear analyses performed. All  
570 previous phylogenetic studies including the two genera have also resolved *Alcolapia* within  
571 *Oreochromis*, albeit with less comprehensive sampling of either genus (Seegers *et al.* 1999; Nagl *et al.*

572 2001, Schwarzer *et al.* 2009, Dunz & Schlieven 2013, Kavembe *et al.* 2013, Matschiner *et al.* 2017,  
573 Rabosky *et al.* 2018). However, the sister group of *Alcolapia* is contentious, as *O. amphimelas* is strongly  
574 supported as its sister group based on concatenated nuclear analyses, whereas, species tree analyses  
575 placed *O. esculentus* as sister, albeit with weak support. A species tree analysis of the nuclear data  
576 without removing the hyper-variable 42 bp (data not shown) did support the *Alcolapia* + *O. amphimelas*  
577 relationship, and it is likely that the uncertainty in the species tree may be attributed to a lack of  
578 synapomorphies.

579

#### 580 4.2 Systematics Account: Revised classification of *Alcolapia*

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

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

600

#### 601 4.3 Timing and colonisation of extreme conditions

602 Our study shows that there has been repeated adaptation to elevated salinity and temperature  
603 throughout the evolutionary history of *Oreochromis* irrespective of phylogenetic hypothesis, but that  
604 adaptation to soda conditions (high temperature, salinity, alkalinity, and low dissolved oxygen) has  
605 likely occurred once. The placement of *O. esculentus* in a clade with *Alcolapia* and *O. amphimelas* in the  
606 species tree analyses (Figure 1A, Supplementary S3), raises the possibility that the adaptation to soda  
607 conditions was also gained in *O. esculentus* (Figure 2A), although it does not currently inhabit soda  
608 conditions (Table 1). While *O. esculentus* is native to the freshwater Lake Victoria (Njiru et al, 2005;  
609 Hallerman and Hilsdorf, 2014), our samples are from Lake Rukwa, a relatively shallow saline lake, where  
610 they have been introduced from Lake Victoria (Seegers, 1996). Lake Rukwa became saline around 5,500  
611 years ago (Barker et al, 2002), and although conditions (pH 9.19-9.26; temperature 27-32°C; salinity:  
612 6,000 mg/L, Haberyan, 1987; Bathymetric survey report, 2014) are not as extreme as the springs  
613 inhabited by *Oreochromis (Alcolapia)* (e.g. pH 8-12, temperature 27-42°C, salinity 34,000 mg/L) or *O.*  
614 *amphimelas*, they are much higher than Lake Victoria, for example (pH: 8.2-9, temperature: 23-28  
615 [surface water temperature], salinity: 97 mg/L) (vanden Bossche & Bernacsek 1990). Of course, Lake  
616 Victoria has experienced phases of desiccation (Johnson et al. 1996) and so it is likely that species such  
617 as *O. esculentus* will have encountered periods of higher salinity in the past. This is probably true of  
618 many other *Oreochromis* populations and thus it is perhaps unsurprising given that there has been  
619 multiple adaptations to these conditions throughout the evolution of this group, indicating a likely  
620 shared genetic mechanism allowing fish from this genus to adapt to such conditions.

621 Our age estimates of the diversification of the *Oreochromis (Alcolapia)* adaptive radiation at 1.73  
622 Ma (95% HPD: 3.20-0.57) coincide with estimates of the date of basin formation for Lakes Natron and  
623 Magadi at 1.7 Ma and the formation of the single palaeolake (Orolonga) preceding Natron and Magadi,  
624 around c.700 Ka (Eugster 1986). The Lake Magadi species *O. (A.) grahami* is estimated to have diverged  
625 from the Lake Natron species *O. (A.) alcalicus* at 0.70 Ma (95% HPD: 1.55-0.007 Ma), suggesting these  
626 taxa may have already diverged prior to the separation of the lakes. However, although these dates are



627 generally well before the time that these lakes are suggested to have separated, during an aridity event  
628 dated at c.11 ka (Williamson et al. 1993) the upper bound is after this timeframe. It is likely that the  
629 nuclear loci selected for this study are too slowly evolving to provide a precise estimate of divergence  
630 dates for a potentially recent radiation (Ford et al. 2015). More nuclear loci and broader outgroup  
631 sampling including additional fossil calibrations should be considered in future studies to test the  
632 accuracy of these dates.

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

644

## 645 5 Conclusions

646 The comprehensive molecular phylogenies of *Oreochromis* presented here represent the first attempt  
647 to clarify species relationships for this relatively young and widespread African cichlid genus. Trees  
648 reconstructed from nuclear sequence data likely give the best hypothesis of relationships, although we  
649 recognise that nuclear genomes may also show introgression (e.g. Nevado et al. 2011). The multispecies  
650 coalescent approach is preferred over concatenation methods, as it accounts for gene tree- species tree  
651 incongruence that arise due to population level processes and has been shown to be especially suited to  
652 the estimation of shallower evolutionary relationships (e.g. Ogilvie et al. 2016, 2017; Flouri et al. 2018),  
653 but we acknowledge that we include a limited number of loci for these types of analyses. To clarify  
654 relationships and better understand the substantial mito-nuclear discordance that we report, further

655 work should focus on genomic data e.g. ultraconserved element or reduced representation genome-  
656 wide sequencing to resolve potential causes of this incongruence and better clarify the areas of conflict  
657 (currently in progress). Genomic data will also enable the likely shared genetic mechanism allowing  
658 these fish to adapt to adverse aquatic conditions to be investigated. Our results highlight the importance  
659 of establishing species relationships within this genus, which contains several commercially important,  
660 and many endangered, species.

661

## 662 Funding

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

666

## 667 Acknowledgements

668 We thank Simon Joly and Andrew Meade for advice in the use of the JML and BayesTraits software  
669 programs, respectively. We also thank Wendy Hart for sequencing support, and Etienne Bezault,  
670 Emmanuel Vreven, and Julia Schwarzer for providing additional samples. The associate editor Guillermo  
671 Orti and reviewers, Michael Matschiner and Prosanta Chakrabarty, provided helpful comments on an  
672 earlier version of this manuscript.

673

## 674 Appendix A. Supplementary material

675 Table S1. Specimen list and GPS coordinates of samples included in the present study, including  
676 GenBank accession numbers for molecular sequences used in phylogenetic analysis.

677 Table S2. Primers and reaction conditions for each gene.

678 Table S3. Trait coding (adaptation to high temperature, high salinity and high pH, and phenotypic traits)  
679 for each species, as implemented in the BayesTraits analysis.

680 Figure S1. Bayesian phylogeny of *Oreochromis* based on concatenated nuclear data (3092 bp) including  
681 101 samples (*Coptodon* sp. samples [outgroup] are removed from the figure). Support values shown

682 below nodes are Bayesian Posterior Probabilities (BPP). Those above the nodes are bootstrap (BS)  
683 values generated using Maximum Likelihood.

684 Figure S2. Bayesian phylogeny of *Oreochromis* based on concatenated mitochondrial data (1582 bp)  
685 including 116 samples, including additional sampling of other Oreochromini taxa (*Coptodon* spp.  
686 samples [outgroup] are removed from the figure). Support values shown below nodes are Bayesian  
687 Posterior Probabilities (BPP). Those above the nodes are bootstrap (BS) values generated using  
688 Maximum Likelihood.

689 Figure S3. Species tree based on nuclear data and generated using BPP. Support values are given below  
690 branches.

691 Figure S4. Ancestral state reconstruction from BayesTraits analysis based on the MrBayes nuclear  
692 concatenated phylogeny. Details, abbreviations and colour coding are as per Figure 2 in the main paper.  
693 A) Ancestral state reconstruction of thermal/salinity tolerance (TS) and soda adaptation (So). Pie charts  
694 at internal nodes indicate probability of presence/absence of ancestor exhibiting adaptation to soda  
695 conditions, from BayesTraits analysis. (pie above node: temperature/salinity tolerance ancestral state  
696 reconstruction; pie below node: soda adaptation ancestral state reconstruction). B) Ancestral state  
697 reconstruction of phenotypic male secondary sexual characteristics: genital tassel (GT) and extended  
698 jaw morphology (EJ). (pie above node: extended jaw morphology ancestral state reconstruction; pie  
699 below node: genital tassel ancestral state reconstruction).

700 Figure S5. Ancestral state reconstruction of thermal and salinity tolerance as independent traits. A)  
701 Reconstruction using the \*BEAST species trees; B) reconstruction using the MrBayes concatenated  
702 nuclear phylogeny. (pie above node: thermal tolerance ancestral state reconstruction; pie below node:  
703 salinity tolerance ancestral state reconstruction).

704

## 705 **References**

706 Alter AE, Munshi-South J, Stiasny MLJ (2017) Genome wide SNP data reveal cryptic phylogeographic  
707 structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River.  
708 *Molecular Ecology*, 26, 1401-1419.

709 Bacher F, Quénéhervé G, Märker M (2014) The delineation of paleo-shorelines in the Lake Manyara  
710 basin using TerraSAR-X data. *Remote Sensing*, 6, 2195-2212.

711 Barker P, Telford R, Gasse F, Thevenon F (2002) Late Pleistocene and Holocene palaeohydrology of Lake  
712 Rukwa, Tanzania, inferred from diatom analysis. *Palaeogeography, Palaeoclimatology,*  
713 *Palaeoecology*, 187, 295–305.

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

717 Bezault E, Rognon X, Clota F, Gharbi K, Baroiller J-F, Chevassus B (2012) Analysis of the meiotic  
718 segregation in intergeneric hybrids of tilapias. *International Journal of Evolutionary Biology*, 2012,  
719 817562.

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

722 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ  
723 (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational*  
724 *Biology*, 10, e1003537.

725 Carnevale G, Sorbini C, Landini W (2003) *Oreochromis lorenzoi*, a new species of tilapiine cichlid from  
726 the Late Miocene of Central Italy. *Journal of Vertebrate Paleontology*, 23, 508–516.

727 Casanova J, Hillaire-Marcel C (1992) Chronology and paleohydrology of Late Quaternary high lake levels  
728 in the Manyara basin (Tanzania) from isotopic data (<sup>18</sup>O, <sup>13</sup>C, <sup>14</sup>C, Th/U) on fossil stromatolites.  
729 *Quaternary Research*, 38, 205–226.

730 Chow S and Hazama K (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish.  
731 *Molecular Ecology*, 7, 1255-1256.

732 Cnaani A, Lee BY, Zilberman N, Ozouf-Costaz C, Hulata G, et al (2008) Genetics of sex determination in  
733 tilapiine species. *Sex Dev*, 2, 43–54.

734 Cusimano N and Renner SS (2014) Ultrametric trees or phylograms for ancestral state reconstruction -  
735 Does it matter? *Taxon*, 36: 721–726.

736 Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and  
737 parallel computing. *Nature Methods*, 9(8), 772.

738 Day JJ, Santini S, Garcia-Moreno J (2007) Phylogenetic relationships of the Lake Tanganyika cichlid tribe  
739 Lamprologini: The story from mitochondrial DNA. *Molecular Phylogenetics and Evolution*, 45, 629-  
740 642.

741 Day JJ, Fages A, Brown KJ, Vreven EJ, Stiassny MLJ, Bills R, Friel JP, Rüber L (2017) Multiple independent  
742 colonizations into the Congo Basin during the continental radiation of African *Mastacembelus* spiny-  
743 eels. *Journal of Biogeography*, 44, 2308-2318.

744 Dunz AR, Schlieven UK (2013) Molecular phylogeny and revised classification of the haplotilapiine  
745 cichlid fishes formerly referred to as “Tilapia.” *Molecular Phylogenetics and Evolution*, 68, 64–80.

746 Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space  
747 complexity. *BMC Bioinformatics*, 19, 1–19.

748 Egger B, Koblmüller S, Sturmbauer C, Sefc KM (2007) Nuclear and mitochondrial data reveal different  
749 evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. *BMC Evolutionary Biology*,  
750 7, 137.

751 Eschmeyer WN (2018) *Catalog of fishes electronic version. Updated 2018*. California Academy of Sciences.

752 FAO. 2016. The State of World Fisheries and Aquaculture 2016. Contributing to food security and  
753 nutrition for all. Rome.

754 Farias IP, Ortí G, Sampaio I, Schneider H, Meyer A (1999) Mitochondrial DNA phylogeny of the family  
755 Cichlidae: monophyly and fast molecular evolution of the Neotropical assemblage. *Journal of*  
756 *Molecular Evolution*, 48, 703–711.

757 Ford AGP, Dasmahapatra KK, Rüber L, Gharbi K, Day JJ (2015) High levels of interspecific gene flow in  
758 an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*,  
759 24, 3421–3440.

760 Ford AGP, Rüber L, Newton J, Dasmahapatra KK, Balarin JD, Bruun K, Day JJ (2016) Niche divergence  
761 facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution*,  
762 70, 2718–2735.

763 Flouri T, Jiao X, Rannala B, Yang Z (2018) Species tree inference with BPP using genomic sequences and  
764 the multispecies coalescent. *Molecular Biology and Evolution*, 35, 2585–2593.

765 Guindon S and Gascuel O (2003). A simple, fast and accurate method to estimate large phylogenies by  
766 maximum-likelihood. *Systematic Biology*, 52, 696-704.

767 Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF (2007) Age of cichlids:  
768 new dates for ancient lake fish radiations. *Molecular Biology and Evolution*, 24, 1269-1282.

769 Genner MJ, Turner GF (2012) Ancient hybridization and phenotypic novelty within Lake Malawi's  
770 cichlid fish radiation. *Molecular Biology and Evolution*, 29, 195–206.

771 Genner MJ, Connell E, Shechonge A, Smith A, Swanstrom J, Mzighani S, Mwijage A, Ngatunga BP, Turner  
772 GF (2013) Nile tilapia invades the Lake Malawi catchment. *African Journal of Aquatic Science*, 38  
773 (Supplement 1), 85–90.

774 Gilks WR (1996). *Introducing Markov chain Monte Carlo. Markov chain Monte Carlo in practice*. Springer.

775 Hall BG (2011) *Phylogenetic Trees Made Easy: A How-To Manual*. Sinauer Associates, Sunderland, MA.

776 Hallerman E, Hilsdorf AWS (2014) Conservation genetics of tilapias: Seeking to define appropriate units  
777 for management. *The Israeli Journal of Aquaculture*, 2014, 1–18.

778 Hillaire-Marcel C, Casanova J (1987) Isotopic hydrology and paleohydrology of the Magadi (Kenya) -  
779 Natron (Tanzania) basin during the Late Quaternary. *Palaeogeography, Palaeoclimatology,*  
780 *Palaeoecology*, 58, 155-181.

781 IUCN (2018) Global invasive species database. Available at: <http://www.iucngisd.org/gisd/>. Accessed  
782 April 2018.

783 The IUCN Red List of Threatened Species. Version 2017-3. <[www.iucnredlist.org](http://www.iucnredlist.org)>. Accessed April 2018.

784 Ivory SJ, Blome MW, King JW, McGlue MM, Cole JE & Cohen AS (2016) Environmental change explains  
785 cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. *Proceedings of the National*  
786 *Academy of Sciences of the United States of America*, 113, 11895-11900.

787 Johnson TC, Scholz CA, Talbot MR, Kelts K, Ricketts RD, Ngobi G, Beuning K, Ssemmanda I, McGill JW  
788 (1996) Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science*,  
789 273, 1091-1093.

790 Joly S (2012) jml: testing hybridization from species trees. *Molecular Ecology Resources*, 12, 179–184.

791 Joly S, McLenachan PA, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and  
792 incomplete lineage sorting. *The American Naturalist*, 174, e54-e70.

793 Kavembe GD, Machado-Schiaffino G, Meyer A (2013) Pronounced genetic differentiation of small,  
794 isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya.  
795 *Hydrobiologia*, 739, 55–71.

796 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S,  
797 Duran C, Thierer T, Ashton B, Mentjies P, Drummond A (2012) Geneious Basic: an integrated and  
798 extendable desktop software platform for the organization and analysis of sequence data.  
799 *Bioinformatics*, 28, 1647–1649.

800 Klett V, Meyer A (2002) What, if anything, is a Tilapia? - Mitochondrial ND2 phylogeny of tilapiines and  
801 the evolution of parental care systems in the African cichlid fishes. *Molecular Biology and Evolution*,  
802 19, 865–883.

803 Kobayashi M, Msangi S, Batka M, Vannuccini S, Dey MM, Anderson JL (2015) Fish to 2030: The Role and  
804 Opportunity for Aquaculture. *Aquaculture Economics and Management*, 19, 282–300.

805 Koblmüller S, Schliewen UK, Duftner N *et al.* (2008) Age and spread of the haplochromine cichlid fishes  
806 in Africa. *Molecular Phylogenetics and Evolution*, 49, 153–169.

807 Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., Calcott, B. (2017) PartitionFinder 2: new methods  
808 for selecting partitioned models of evolution for molecular and morphological phylogenetic  
809 analyses. *Molecular Biology and Evolution*, 34, 772-773.

810 Lowe-McConnell RH (1959) Breeding behaviour patterns and ecological differences between Tilapia  
811 species and their significance for evolution within the genus Tilapia (Pisces: Cichlidae). *Proceedings*  
812 *of the Zoological Society of London*, 132, 1–30.

813 Matschiner M, Musilová Z, Barth JMI, Starostová Z, Salzburger W, Steel M, Bouckaert R (2017)  
814 Bayesian phylogenetic estimation of clade ages supports Trans-Atlantic dispersal of cichlid fishes.  
815 *Systematic Biology*, 66, 3-22.

816 Mazzuchelli J, Kocher TD, Yang F, Martins C (2012) Integrating cytogenetics and genomics in  
817 comparative evolutionary studies of cichlid fish. *BMC Genomics*, 13, 463.

818 McCann J, Schneeweiss GM, Stuessy TF, Villaseñor JL, Weiss-Schneeweiss H (2016) The Impact of  
819 reconstruction methods, phylogenetic uncertainty and branch lengths on inference of chromosome  
820 number evolution in American Daisies (*Melampodium*, Asteraceae). *PLoS ONE*, 11, e0162299.

821 Meade A. (2011) BayesTrees 1.1. Available at: <http://www.evolution.rdg.ac.uk/BayesTrees.html>

822 Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O (2017) Ancient hybridization fuels  
823 rapid cichlid fish adaptive radiations. *Nature Communications*, 8, 14363.

824 Meyer BS, Salzburger W (2012) A novel primer set for multilocus phylogenetic inference in East African  
825 cichlid fishes. *Molecular Ecology Resources*, 12, 1097–1104.

826 Meyer BS, Matschiner M, Salzburger W (2017) Disentangling incomplete lineage sorting and  
827 introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Systematic Biology*,  
828 66, 531–550.

829 Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES science gateway for inference of large  
830 phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop, New Orleans,  
831 LA, 14 Nov 2010, pp 1–8.

832 Mmochi AJ (2017) Growth rates of selected *Oreochromis* species cultured at different salinities. World  
833 Aquaculture meeting. Abstract.

834 Moser FN, Rijssel JV, Ngatunga B, Mwaiko S, Seehausen O (2018) The origin and future of an endangered  
835 crater lake endemic; phylogeography and ecology of *Oreochromis hunteri* and its invasive relatives.  
836 *Hydrobiologica*. doi.org/10.1007/s10750-018-3780-z

837 Murray AM, Stewart KM (1999) A new species of tilapiine cichlid from the Pliocene, Middle Awash,  
838 Ethiopia. *Journal of Vertebrate Paleontology*, 19, 293–301.

839 Nagl S, Tichy H, Mayer WE *et al.* (2001) Classification and phylogenetic relationships of African tilapiine  
840 fishes inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 20, 361–  
841 374.

842 Ndiwa TC, Nyingi DW, Agnese JF (2014) An important natural genetic resource of *Oreochromis niloticus*  
843 (Linnaeus, 1758) threatened by aquaculture activities in Lobo Drainage, Kenya. *PLoS ONE* 9(9),  
844 e106972.



- 845 Nevado B, Fazalova V, Backeljau T, Hanssens M, Verheyen E (2011) Repeated unidirectional  
846 introgression of nuclear and mitochondrial DNA between four congeneric Tanganyikan cichlids.  
847 *Molecular Biology and Evolution*, 28, 2253–2267.
- 848 Njiru M, Waithaka E, Muchiri M, van Knaap M, Cowx IG (2005) Exotic introductions to the fishery of Lake  
849 Victoria: What are the management options? *Lakes & Reservoirs: Research and Management*, 10,  
850 147.
- 851 Ogilvie HA, Heled J, Xie D, Drummond AJ (2016) Computational performance and statistical accuracy of  
852 \*BEAST and comparisons with other methods. *Systematic Biology*, 65, 381-396.
- 853 Ogilvie HA, Bouckaert RR, Drummond A (2017) StarBEAST2 brings faster species tree inference  
854 and accurate estimation of substitution rates. *Molecular Biology and Evolution*, 34, 2101-2114.
- 855 Pagel M. (2004) Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*,  
856 53, 673–684.
- 857 Pagel M, Meade A (2006) Bayesian analysis of correlated evolution of discrete characters by reversible-  
858 jump Markov chain Monte Carlo. *The American Naturalist*, 167, 808–825.
- 859 Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language.  
860 *Bioinformatics*, 20, 289-290.
- 861 Penk S, Schliewen U, Altner M, Reichenbacher, B (2018) The earliest record of the Oreochromini  
862 (Cichlidae: Pisces) from a Miocene lacustrine deposit in the East African Rift Valley. 5th International  
863 Paleontological Congress. Abstract.
- 864 Philippart J-C, Ruwet J-C (1982) Ecology and distribution of tilapias, p. 15-59. In RSV Pullin and RH  
865 Lowe-McConnell (eds.) The biology and culture of tilapias. ICLARM Conference Proceedings 7, 432 p.  
866 International Center for Living Aquatic Resources Management, Manila, Philippines.
- 867 Poletto AB, Ferreira IA, Cabral-de-Mello D *et al.* (2010) Chromosome differentiation patterns during  
868 cichlid fish evolution. *BMC Genetics*, 11, 50.
- 869 Pouyaud L, Agnès JF (1995) Phylogenetic relationships between 21 species of three tilapiine genera  
870 *Tilapia*, *Sarotherodon* and *Oreochromis* using allozyme data. *Journal of Fish Biology*, 47, 26–38.

871 R Core Team (2018) R: A language and Environment for Statistical Computing. R Foundation for  
872 Statistical Computing, Vienna, Austria.

873 Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME (2013) Rates of speciation  
874 and morphological evolution are correlated across the largest vertebrate radiation. *Nature*  
875 *Communications*, 4: 1958.

876 Rabosky DL, Chang J, Pascal O, Cowman, PF, Sallan L, Friedman M, Kaschner K, Garilao C, Near TJ, Coll M,  
877 Alfaro ME (2018) An inverse latitudinal gradient in speciation rate for marine fishes *Nature*, 559,  
878 392–395.

879 Rambaut R, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian  
880 phylogenetics using Tracer 1.7. *Systematic Biology*, 67, 901–904.

881 Regan C (1920) The classification of the fishes of the family Cichlidae. *Annals and Magazine of Natural*  
882 *History*, 5, 1–71.

883 Riedel R, Costa-Pierce BA (2005) Feeding ecology of Salton Sea Tilapia (*Oreochromis* spp). *Bulletin of the*  
884 *Southern California Academy of Sciences*, 104, 26–36.

885 Robinson O, Dylus D, Dessimoz C (2016) Phylo.io: Interactive viewing and comparison of large  
886 phylogenetic trees on the web. *Molecular Biology and Evolution*, 33, 2163-2166.

887 Rognon X, Guyomard R (2003) Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to  
888 *O. niloticus* in West Africa. *Molecular Ecology*, 12, 435–445.

889 Ronquist F, Teslenko M, van der Mark P, *et al.* (2012) MrBayes 3.2: Efficient Bayesian phylogenetic  
890 inference and model choice across a large model space. *Systematic Biology*, 61, 539–542.

891 Schliwen UR, Klee B (2004) Reticulate sympatric speciation in Cameroonian crater lake cichlids.  
892 *Frontiers in Zoology*, 1, 5.

893 Schluter T, Kohring R, Mehl J (1992) Hyperostotic fish bones (“Tilly bones”) from presumably Pliocene  
894 phosphorites of the Lake Manyara area, northern Tanzania. *Paläontologische Zeitschrift*, 66, 129–  
895 136.

896 Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.

897 Schwarzer J, Misof B, Tautz D, Schliwen U (2009) The root of the East African cichlid radiations. *BMC*  
898 *Evolutionary Biology*, 9, 186.

899 Seegers L (2008) The fishes collected by G A Fischer in East Africa in 1883 and 1885/86. *Zoosystematics*  
900 *and Evolution*, 84, 149–195.

901 Seegers L, Tichy H (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and  
902 Magadi, Tanzania and Kenya, with description of two new species. *Ichthyological Explorations of*  
903 *Freshwaters*, 10, 97–146.

904 Seegers L, Sonnenberg R, Yamamoto R (1999) Molecular analysis of the *Alcolapia* flock from lakes  
905 Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their  
906 systematics and evolution. *Ichthyological Explorations of Freshwaters*, 10, 175–199.

907 Seegers L (2008) The fishes collected by G A Fischer in East Africa in 1883 and 1885/86. *Zoosystematics*  
908 *and Evolution*, 84, 49–95.

909 Seehausen O, Koetsier E, Schneider MV, Chapman LJ, Chapman CA, Knight ME, Turner GF, van Alphen  
910 JM, Bills R (2003) Nuclear markers reveal unexpected genetic variation and a Congolese–Nilotic  
911 origin of the Lake Victoria cichlid species flock. *Proc. R. Soc. Lond. B*, 270, 129–137.

912 Seehausen O (2007) Chance, historical contingency and ecological determinism jointly determine the  
913 rate of adaptive radiation. *Heredity*, 99, 361–363.

914 Seehausen O (2015) Process and pattern in cichlid radiations - inferences for understanding unusually  
915 high rates of evolutionary diversification. *New Phytologist*, 207, 304–312.

916 Shechonge A, Ngatunga BP, Bradbeer SJ, Day JJ, Ford AGP, Kihedu J, Richmond T, Mzighani S, Smith AM,  
917 Sweke EA, Tamatamah R, Tyers AR, Turner GF, Genner MJ (2019) Widespread colonisation of  
918 Tanzanian catchments by introduced *Oreochromis* tilapia fishes: the legacy from decades of  
919 deliberate introduction. *Hydrobiologia*, 832, 235.

920 Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection.  
921 *Bioinformatics* 17, 1246–1247.

922 Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51, 492–  
923 508.

924 Stöver B C, Müller K F (2010) TreeGraph 2: Combining and visualizing evidence from different  
925 phylogenetic analyses. *BMC Bioinformatics*, 11, 7.

- 926 Swofford DL (2000) PAUP\*: Phylogenetic analysis using parsimony (\* and other methods), version  
927 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
- 928 Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using  
929 maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology  
930 and Evolution*, 28, 2731–2739.
- 931 Thys van den Audenaerde DFE (1968) An annotated bibliography of tilapia (Pisces, Cichlidae).  
932 Documentation zoologique No. 14, Musée royal d'Afrique centrale, Tervuren.
- 933 Trewavas E (1982b) Tilapias: taxonomy and speciation. In: *The Biology and Culture of Tilapias* (eds  
934 Pullin R, Lowe-McConnel RH), pp. 3–13. ICLARM Conference Proceedings, Manila, Philippines.
- 935 Trewavas E (1982) Tilapias: taxonomy and speciation. In: *The Biology and Culture of Tilapias* (eds Pullin  
936 R, Lowe-McConnel RH), pp. 3–13. ICLARM Conference Proceedings, Manila, Philippines.
- 937 Trewavas E (1983) *Tilapiine Fishes of the genera Sarotherodon, Oreochromis and Danakilia*. British  
938 Museum (Natural History), London.
- 939 Turner GF, Robinson RL (1991) Ecology, morphology and taxonomy of the Lake Malawi *Oreochromis*  
940 (*Nyasalapia*) species flock. *Ann. Mus. Roy. Afr. Centr. Sc. Zool.*, 262, 23–28.
- 941 Turner GF (2007) Adaptive radiation of cichlid fish. *Current Biology*, 17, R827–R831.
- 942 Uyeda JC, Zenil-Ferguson R, Pennell MW (2018) Rethinking phylogenetic comparative methods.  
943 *Systematic Biology*, 67, 1091–1109.
- 944 Van Couvering J (1982) Fossil cichlid fish of Africa. In: *Special Papers in Palaeontology*, pp. 29: 1–103.  
945 The Palaeontological Association, London.
- 946 Vanden Bossche J-P, Bernacsek GM (1990) Source Book for the Inland Fishery Resources of Africa, Vol.  
947 1, FAO, Rome.
- 948 Wagner CE, Harmon LJ, Seehausen O (2014) Cichlid species-area relationships are shaped by adaptive  
949 radiations that scale with area. *Ecology Letters*, 17, 583–592.
- 950 Willis SC, Farias IP, Ortí G. (2014) Testing mitochondrial capture and deep coalescence in amazonian  
951 cichlid fishes (Cichlidae: *Cichla*). *Evolution*, 68, 256–268.

952 Yu G, Smith D, Zhu H, Guan Y, Lam TT (2017) ggtree: an R package for visualization and annotation of  
953 phylogenetic trees with their covariates and other associated data. *Methods in Ecology and*  
954 *Evolution*, 8, 28–36.

955 Yue GH, Lin HR, Li JL (2016) Tilapia is the fish for next-generation aquaculture. *Int J Marine Sci Ocean*  
956 *Technol*, 3, 11–13.

957 Zwickl, D. J (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological  
958 sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of  
959 Texas at Austin.

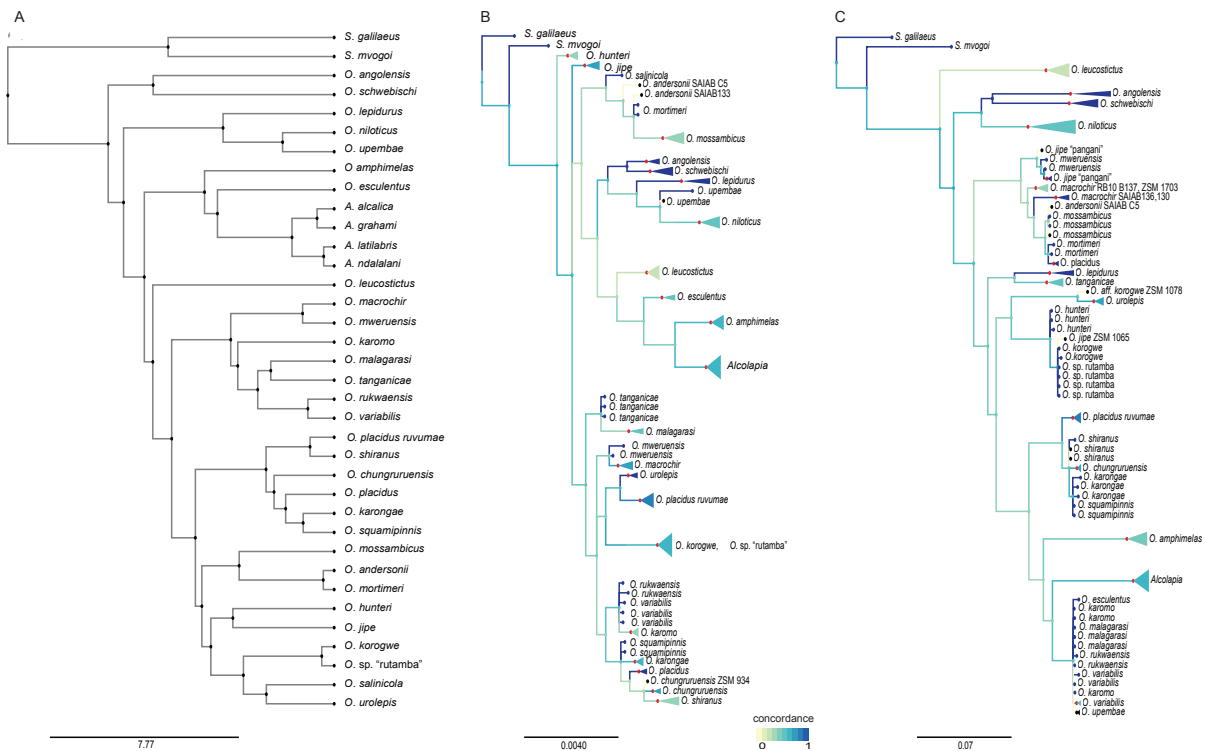
960

961

962 **Figures**

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

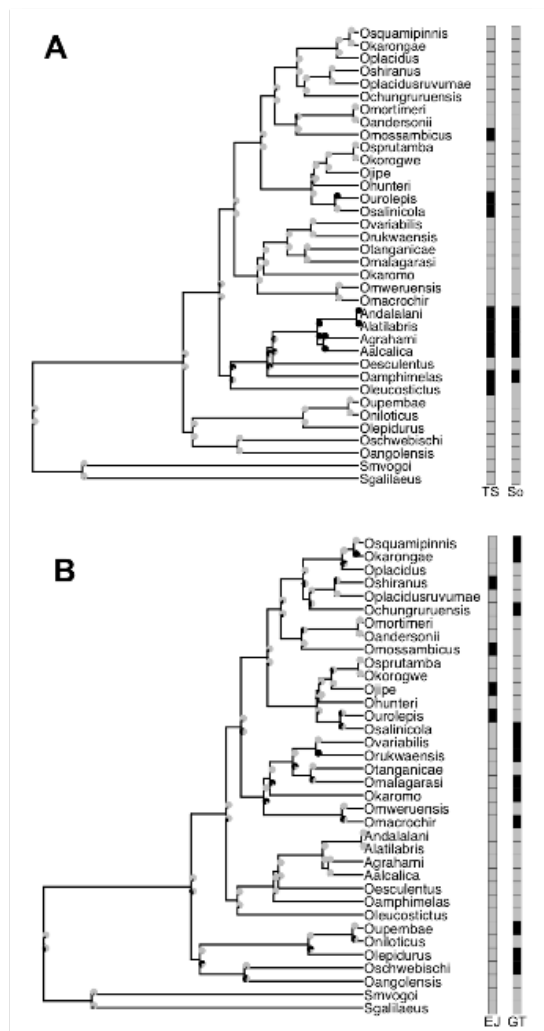
970



971

972

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



985

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

## 986 Tables

987 Table 1. *Oreochromis* adaptations to soda conditions.

988 Taxa in bold indicate samples included in the present study (see Table S1). Temperatures and  
989 salinity/alkalinity conditions are the maximum at which the species naturally occur. Several studies  
990 have shown species are able to tolerate/survive higher levels in laboratory conditions (though fewer  
991 have explored successful reproduction at extreme conditions). For the present study, temperature  
992 tolerance >35°C, salinity tolerance >30‰, and pH tolerance >pH 9 were considered to represent



993 elevated environmental tolerance; and tolerance to all three elevated parameters to indicate soda  
 994 adaptation. Species names in grey were coded as temperature/saline tolerant for ancestral state  
 995 reconstruction analyses, and those in dark grey were also coded as soda tolerant (see Table S3).  
 996 References: 1. Philippiart & Ruwet 1982; 2. Trewavas 1983; 3. Seegers et al. 1999. 4. Ndiwa et al.  
 997 2014. 5. Present study.  
 998 nd: minimal data indicating elevated tolerance, presumed not to occur in soda conditions.  
 999

1000

1001 Table 2. Substitution models by partition used for the Bayesian and ML analyses.

<b>Partitions</b>	<b>Best Fitting Model</b>
<b>nuDNA exon vs. intron</b>	
nuDNA exons	HKY+I+G 1003
nuDNA introns	GTR+I+G
<b>nuDNA loci</b> 1004	
BMP4	HKY+I+G
S7, CCNG1	HKY+G
GAPDHS, TYR, b2m	HKY+I+G
<b>mtDNA</b>	
ND2 pos1, 12S	HKY+I+G
ND2 pos2	TRN+I
ND2 pos3	TRN+G
mtDNA	GTR+I+G