Ancient and ongoing hybridization in the *Oreochromis* cichlid fishes

Adam G Ciezarek[1]*, Tarang K Mehta[1], Angela Man[1], Antonia GP Ford[2], Geraldine Dorcas Kavembe[3], Nasser Kasozi[4], Benjamin P Ngatunga[5], Asilatu H Shechonge[5], Rashid Tamatamah[5], Dorothy Wanja Nyingi[6], Avner Cnaani[7], Federica Di Palma[8], George F Turner[9], Martin J Genner[10], Wilfried Haerty[1]

*- corresponding author

1 - Earlham Institute, Norwich Research Park, Colney Ln, Norwich NR4 7UZ, UK

2 - School of Life and Health Sciences, Whitelands College, University of Roehampton, London, SW15 4NA, UK

3 - Department of Life Sciences, South Eastern Kenya University, P.O. Box 170 Kitui, 90200, Kenya

4 - National Agricultural Research Organisation, Aquaculture Research and Development Centre, P.O. Box 530, Kampala, Uganda

5 - Tanzania Fisheries Research Institute, Dar es Salaam, P.O. Box 9750, Tanzania

6 - Ichthyology Section, National Museums of Kenya, P.O. Box 40658, Nairobi 00100, Kenya

7 - Institute of Animal Science, Agricultural Research Organization, Rishon LeZion 7528809, Israel

8 - School of Biological Sciences, University of East Anglia, Norwich NR4 7TU, UK

9 - School of Natural Sciences, Bangor University, Bangor, LL57 2UW, United Kingdom,

10 - School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK

Abstract

Hybridization may enable adaptive diversification by generating unique genetic combinations when hybrid lineages are faced with ecological opportunity. Conversely, hybridization with exotic species may be detrimental to native biodiversity, by leading to homogenisation and the loss of important genetic material associated with local adaptation. Here we studied both ancient and contemporary hybridization in cichlid fishes of the genus *Oreochromis* (tilapia), which are among the most important fish for global aquaculture. We use whole genome resequencing of 575 individuals from 23 species, focussing on Tanzania, a natural hotspot of tilapia diversity, and a country where hybridization between exotic and native species in the natural environment has been previously reported. We reconstruct the first genome-scale phylogeny of the genus and reveal prevalent ancient gene flow across the *Oreochromis* phylogeny. This introgression has not led to large-scale adaptive radiation as seen in other cichlid lineages. We identify multiple cases of contemporary hybridization between native and introduced species in the wild, linked to the use of non-native species in aquaculture improvement and stocking for capture fisheries. Our study shows how ancient hybridization contributed to modern tilapia diversity, but is now a threat to both the genetic integrity of wild populations and the long-term prospects of the global tilapia aquaculture industry.

Introduction

Global aquaculture production has increased dramatically in recent decades, and is projected to take a key role in meeting the food demands of a growing human population, whilst alleviating nutritional poverty (Ahmed et al., 2019; Hall et al., 2017; Naylor et al., 2021). Tilapia of the genus *Oreochromis* (Cichlidae: Oreochromini) are a group of cichlid fish native to Africa and the Middle East, and are now the second largest group (by tonnage produced) of any aquaculture fish globally, following carp (Cyprinidae). *Oreochromis* production is dominated by a single species, the Nile tilapia, *Oreochromis niloticus*, which accounted for 4.5 out of the total 5.5 million tonnes of tilapia produced in 2020, the third most of any fish species globally. The remaining ~1 million tonnes are only accurately recorded to the genus level (FAO, 2022). The success of Nile tilapia has been attributed to high growth rates as well as tolerance of a wide range of environmental conditions, such as temperature, salinity, pH and low dissolved oxygen enable (El-Sayed & Fitzsimmons, 2023). In principle, the use of wild relatives of the Nile tilapia in breeding programmes could be used to further enhance production (Lind et al., 2012). However, despite this importance for food security, many of the 37 described *Oreochromis* species (Ford et al., 2019) are poorly known.

Although *Oreochromis* tilapia are widespread across Africa, subject to major capture fisheries, and key aquaculture species, considerably more attention has been given to the evolution of haplochromine cichlid systems (Joyce et al., 2011; Salzburger, 2018; Seehausen, 2006; Sturmbauer, 1998). This focus has been due to the propensity of haplochromines to form expansive lacustrine adaptive radiations, characterised by rapid speciation and extensive ecomorphological diversification (Malinsky et al., 2018; McGee et al., 2020). This occurred despite low levels of genetic diversity, a high proportion of which is shared between species due to incomplete lineage sorting and widespread interspecific introgression (Svardal et al., 2021). It has been proposed that extensive lacustrine radiations of haplochromine cichlids have been facilitated by the introgression arising from hybridization of ancestral lineages (Meier et al., 2017; Svardal et al., 2020). This may fuel

68 differentiation and diversification by supplying novel combinations of genetic variants
69 (Singhal et al., 2021), and sharing of ancient alleles that have already been filtered by
70 selection (Marques et al., 2019). By contrast, ongoing or recurring hybridization may
71 decrease differentiation and homogenise gene pools, therefore reducing the propensity of a
72 lineage to radiate.

73 In contrast to the haplochromines, extensive radiation has not taken place in the
74 *Oreochromis* (Seehausen, 2006). Most *Oreochromis* species occupy rivers and are allopatric
75 to apparent sister taxa (Ford et al., 2019; Trewavas, 1983)*;* the only known *Oreochromis*
76 lacustrine radiations are small and limited to Lake Malawi and Lake Natron (3 species each).
77 To date, comprehensive phylogenomic examinations of ancestral introgression occurring
78 during the evolution of *Oreochromis* are lacking, and therefore it is currently not clear what
79 role ancestral introgression has played in *Oreochromis* diversification*.*

80 Given the high productivity of some *Oreochromis* species in aquaculture and capture
81 fisheries, they have been widely introduced across tropical and subtropical freshwaters
82 globally (Canonico et al., 2005; Lévêque, 1995). In Africa, introductions to the natural
83 environment have resulted in widely documented contemporary hybridization (Blackwell et
84 al., 2020; Bradbeer et al., 2019; Champneys et al., 2021; Ciezarek et al., 2022; Shechonge
85 et al., 2018). The risk of hybridisation between native and exotic *Oreochromis* species is
86 likely to increase with the projected growth of aquaculture across Africa (Prabu et al., 2019).
87 Detailed analysis of the extent of hybridization between co-occurring tilapia species will
88 therefore be valuable for identifying species and populations of conservation concern.

89 This study focuses on the genomic implications on ancestral and contemporary
90 introgression in *Oreochromis*, focusing on wild populations in Tanzania. The country is a
91 hotspot for diversity of the genus with at least 21 species present, including a minimum of
92 three species that have been widely introduced outside their native range, namely *O.*
93 *niloticus*, *O. leucostictus* and *O. esculentus* (Shechonge et al., 2019). To reconstruct the
94 phylogenetic relationships between species, and to quantify ancestral and modern
95 hybridization, we present a genome-wide resequencing dataset of the genus *Oreochromis*.

We provide a first phylogenomic analysis of the group and show that *Oreochromis* is typified by a high degree of ancestral introgression. We also identify a species of hybrid origin, the Lake Kiungululu tilapia, *O. chungruruensis*. We further find that anthropogenic activities have resulted in widespread recent hybridization, highlighting a potential risk to the conservation of species and their genetic diversity.

Materials and methods

Sampling and sequencing

We sampled and generated sequencing data for 433 individuals, and combined this with previously published data from 142 individuals (Table S1) from across southern and eastern Africa (Figure 1). Samples were from 65 locations, and 21 drainage basins across Tanzania, as well as South Africa, Kenya, Angola, Malawi, Mozambique and Uganda. Individuals were morphologically identified to species using information outlined in (Genner et al., 2018). Individuals were classified as hybrids if they had characteristics intermediate between different species. Our final dataset comprised of 25 species, including the outgroups *Maylandia zebra* and *Sarotherodon galilaeus* and 23 out of the 37 species of *Oreochromis* currently recognised, including *Oreochromis (Alcolapia) grahami* (Fricke et al., 2023). Fish specimens were mostly obtained from local fishers or through experimental fishing surveys. Live fish were euthanised with an overdose of anaesthetic (MS-222). Samples for subsequent DNA extraction were collected by clipping off the right pectoral fin, which was placed in a labelled vial of ethanol. Whole voucher specimens were retained, with each individual pinned on a styrofoam board, photographed and preserved in formalin or ethanol.

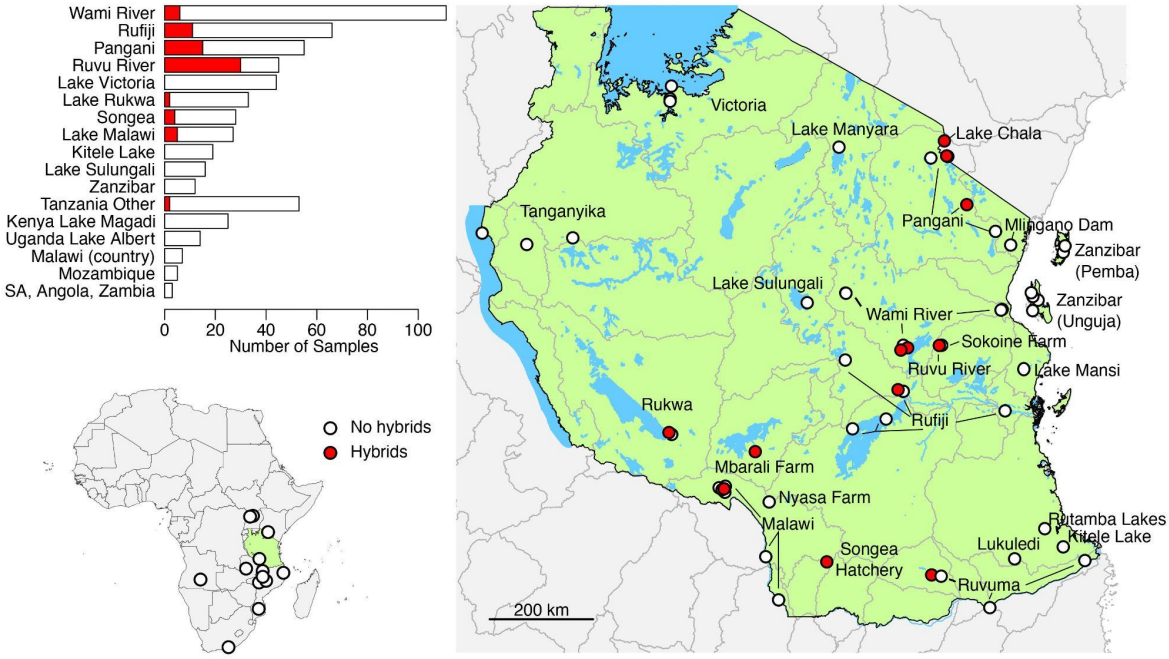


Figure 1. Locations of the *Oreochromis* sampled across Africa (bottom left) and, within Tanzania (right). Sample sites where contemporary hybrids were identified from genomic data are highlighted in red. Key drainage basins are labelled. The bar chart (top left) shows the total number of specimens (contemporary hybrids in red, non-hybrids in white) located per drainage basin within Tanzania, for all basins with more than ten samples, or country outside of Tanzania. All basins within Tanzania with fewer than ten specimens sampled are grouped into “Tanzania Other”. Abbreviation: SA - South Africa. Geographic coordinates of sampling locations are shown in Table S1.

DNA was extracted from fin clips using a PureLink Genomic DNA extraction kit (Life Technologies). Sequencing was conducted using either PCR-free library preparation with Illumina HiSeq 2500 (250bp paired-end), automated KAPA DNA library preparation with Illumina HiSeq 2500 (125bp paired-end), LITE library preparation with Illumina HiSeq 4000 (150bp paired-end) or LITE library preparation with NovaSeq S4 (150bp paired-end).

SNP calling

136 SNPs were called against previously published near chromosome-level assemblies of *M.*
137 *zebra* (GCF_000238955.4) and *O. niloticus* (GCF_001858045.2) (Conte et al., 2019). Raw
138 reads were checked for adapter sequence contamination using BBMerge, within BBMap
139 (v38.06) (Bushnell, 2014). Where necessary, reads were trimmed for adapters using fastp
140 (v0.20.0) (Chen et al., 2018), with quality-trimming disabled. Reads were then mapped
141 separately against both assemblies using the mem function in bwa (v0.30.0) (Li & Durbin,
142 2009), with raw mappings sorted by coordinate and having mate coordinates and insert size
143 fields added using samtools (v1.9) (Li et al., 2009). Joint genotyping was then carried out on
144 all 575 samples using bcftools (v1.10.2) (Li et al., 2009). First, bcftools mpileup was used
145 with a minimum base and mapping qualities of 30. Multi-allelic variant calling was then
146 carried out using bcftools call. Variants were filtered if they were within three base pairs of
147 any other variant, if variant quality score was less than 30, if depth at a given site (across all
148 samples) was less than 500 or greater than 9000, or if minor allele count was less than
149 three.
150
151 *Backbone phylogeny reconstruction*
152 To construct a species-level phylogeny, we identified a panel of 91 reference individuals that
153 were unlikely to be recent hybrids (Table S1). These individuals were identified using
154 morphology, evidence of monophyly in a full genome neighbour-joining tree of individuals not
155 classified morphologically as hybrid, and evidence of monophyly in a mitochondrial genome
156 neighbour-joining tree of those individuals (Supplementary Data 1). Using this reference
157 panel, separate phylogenetic trees were inferred for each of the *M. zebra* and *O. niloticus*
158 mapped datasets. Trees were constructed using both a maximum-likelihood SNP-based
159 concatenation tree, inferred using IQ-TREE (v1.6.12) (Nguyen et al., 2015), as well as a
160 multi-species coalescent summary-based tree, inferred using ASTRAL (v5.4.12) (Rabiee et
161 al., 2019; Zhang et al., 2018), based on trees inferred from recombination-free 10kb
162 windows across the genome with IQ-TREE (Supplementary Methods). Robinson-Foulds
163 (RF) distances (Robinson & Foulds, 1981) were calculated between each 10kb-window tree

164 and the inferred ASTRAL species tree using the ETE 3 python package (Huerta-Cepas et
165 al., 2016), in order to further quantify how widespread phylogenetic discordance was. For
166 species where the inferred phylogenetic trees raised taxonomic questions, such as a lack of
167 monophyly of all individuals assigned to a species, exploratory Principal Component
168 Analysis (PCA) was carried out on all individuals of the relevant species, but excluding those
169 inferred to be a contemporary hybrid (see below), using plink (v2.0.0) (Purcell et al., 2007).
170 The PCAs on each subset of individuals used all biallelic SNPs found with a minor allele
171 count of at least 3 in this subset of individuals, and SNPs were pruned for linkage
172 disequilibrium (r^2>0.6) over 20kb windows using bcftools.
173
174 *Assessment of ancestral introgression*
175 Ancestral introgression between species was assessed using the panel of 91 reference
176 individuals, and both the *M. zebra* and *O. niloticus* mapping datasets. We used both SNP-
177 based D statistics (Green et al., 2010), as well as *f*-branch statistics (Malinsky et al., 2018),
178 calculated using Dsuite (v0.4) (Malinsky et al., 2021), and genome-wide 10kb window
179 phylogenetic-tree-based analyses. These tree-based analyses included the branch-length
180 tests (BLT) and discordant count tests (DCT), using the scripts provided by Suvorov et al.,
181 (2022), and chi-squared tests of discordant topology frequencies inferred using IQ-TREE
182 (Minh et al., 2020; Suvorov et al., 2022) (Supplementary Methods). To test whether
183 introgression was more likely between species sharing a drainage basin or between closely
184 or more distantly related species, for each species we identified the species counterpart with
185 the highest *f*-branch (i.e. the highest degree of gene flow). We then inferred whether these
186 species pairs were more likely to have shared modern drainage basins than expected by
187 chance using a randomised permutation test. Similarly, we identified for each species
188 whether their high *f*-branch was more or less phylogenetically related than the average of all
189 the species compared (z <= -2 for less phylogenetically distant, z >= 2 for more
190 phylogenetically distant). We then tested whether we observed more species with

191 significantly phylogenetically similar or dissimilar counterparts than expected by chance
192 using a permutation test.
193
194 *Analysis of putative introgression events*
195 Putative introgression events, identified by the *f*-branch analysis, were further investigated to
196 identify introgressed genomic regions. This was carried out using a novel statistic (Dwt;
197 Dweighted_topo), derived from topology weightings calculated from 'topology weighting by
198 iterative sampling of subtrees' (Twisst) (Martin & Van Belleghem, 2017), computed along the
199 22 linkage groups of the *M. zebra* reference assembly (Supplementary Methods). Dwt is
200 designed to highlight putatively introgressed regions. Dwt = 0 indicates regions where the
201 species tree (sptree) has the highest weighting. Dwt > 0 indicates where disc1 (the putative
202 introgression topology) is most heavily weighted, with 1 indicating it is fully supported (no
203 weighting for either of the other topology). Similarly, Dwt < 0 indicates where disc2 (the other
204 alternate topology) is most heavily weighted. $Dwt = \frac{disc1 - disc2}{disc1 + disc2} *$
205 $\frac{max(sptree, disc1, disc2) - sptree}{max(sptree, disc1, disc2) + sptree}$.
206 Putative introgression events where the Twisst analysis showed no clear excess of
207 either the species-tree topology or hybridization topology were further investigated to identify
208 signals of hybrid speciation. Following Olave et al. (2022), we used ADMIXTURE analyses
209 of the relevant species at both K=2 and K=3, as well as assessment of species-diagnostic
210 SNPs. Putative introgression dates were also estimated based on estimates of Dxy and π in
211 phylogenetically identified introgressed regions of the genome, calculated using
212 genomics_general (https://github.com/simonhmartin/genomics_general) (see Supplementary
213 Methods).
214
215 *Contemporary hybridization*
216 The 486 individuals that were not among the reference panel were tested for evidence of
217 recent hybridization. We carried out separate analyses in each sampled water drainage

218 basin within Tanzania. For each drainage basin, biallelic SNPs were extracted for each
219 individual sampled, alongside reference individuals for all the taxa recorded in the water
220 body by Shechonge et al. (2019). SNPs with a minor allele count of at least 3 were retained,
221 and SNPs were pruned for linkage disequilibrium (r^2>0.6) over 20kb windows using bcftools.
222 For each drainage basin, a supervised ADMIXTURE (Alexander et al., 2009) analysis was
223 carried out only involving the relevant species or populations, with a K value equal to the
224 number of locally recorded taxa for the ADMIXTURE analysis, and each reference species
225 assigned to their known group. Test individuals were assigned a hybrid status if they had
226 multiple ancestry components where the lower end of the standard error was at least 0.1.
227 This cut-off was used to identify up to second generation backcrosses. We further confirmed
228 that all of the reference individuals of the same species were tightly clustered using PCAs
229 performed in plink, and comprised a monophyletic group with no long branch lengths, in
230 phylogenetic trees inferred with IQ-TREE (v2.0) with automated model detection, 5
231 independent runs and 1,000 rapid bootstraps. These analyses were carried out with the
232 same SNP set as the ADMIXTURE runs, except it was further filtered for sites where at least
233 one individual had each of the homozygous reference and alternate alleles for the IQ-TREE
234 analysis. We further characterised contemporary hybridization using panels of species-
235 diagnostic SNPs (Supplementary Methods).
236
237 **Results**
238 Mapping, phylogenomic and ancestral introgression analyses were carried out against both
239 the *O. niloticus* and *M. zebra* reference assemblies separately. Results were similar and
240 conclusions identical (Supplementary Materials), therefore only the *M. zebra* reference
241 analyses are reported below.
242
243 *Sampling, sequencing, read mapping and SNP calling of 23 species.*
244 An average of 41.3 million paired-end reads were sequenced for the 433 individuals (range
245 10.4 - 106 million) (Figure 1). Mapped reads from all 575 individuals (including the previously

246 published data for 142 individuals) had average depth of 6.3 (range 1.7-18.6), and average
247 properly paired mapping of 75% (range 45-89%). In total 55,949,298 filtered SNPs were
248 called. Full mitochondrial genomes of at least 10,000bp were successfully assembled for
249 437 out of the 450 individuals not morphologically identified as hybrid, with an average
250 length of 16,541bp (range 11,048-17,126bp).
251
252 *Phylogenetic inference*
253 A total of 1,692,136 non-coding SNPs were used for the maximum-likelihood phylogenetic
254 inference with 15,310 recombination-free 10kb windows used for ASTRAL. Most of the
255 taxonomic species produced monophyletic groups, and therefore collapsed to a single node.
256 Those that were not monophyletic were instead collapsed to the population level (Figure 2a).
257 There were some inconsistencies between the ASTRAL and maximum-likelihood trees,
258 particularly with regard to the phylogenetic placement of *O. tanganicae, O. placidus placidus*
259 and *O. korogwe* (Figure S2-5). The ASTRAL results also suggest *O. hunteri* to be a
260 monophyletic species, whereas the ML trees do not (Figure S2-5). All following analyses are
261 reported using the ASTRAL species tree, given that unlike the maximum-likelihood tree it
262 accounts for the likely widespread incomplete lineage sorting in the dataset. This is indicated
263 by the high levels of phylogenetic discordance shown by high Robinson-Foulds (RF)
264 distances between each 10kb window tree and the species tree. No 10kb window tree had a
265 RF distance from the species tree of less than 0.08 (Figure S1).

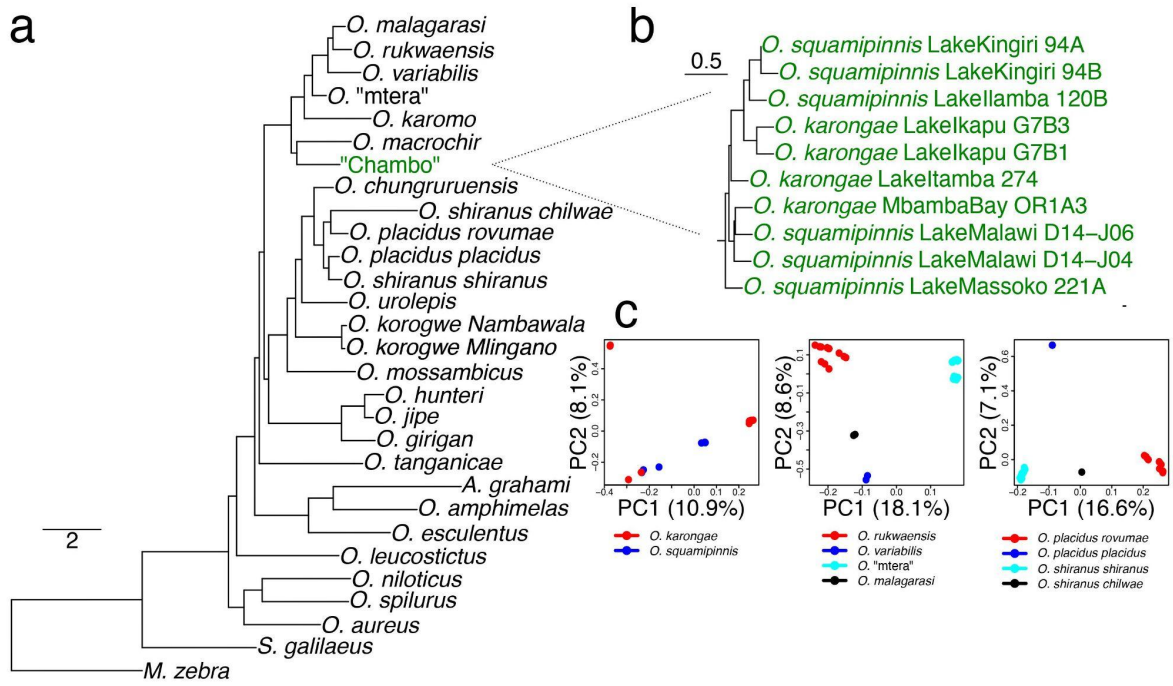


Figure 2. a) Phylogenetic tree inferred using ASTRAL from the *M. zebra* mapping dataset, with conspecific individuals collapsed. Branch lengths are in coalescent units. All nodes had posterior probability of 1.0. b) Phylogenetic tree inferred using ASTRAL from the *M. zebra* mapping dataset, without species collapsed, highlighting only the individuals from the Lake Malawi "Chambo" assemblage of *O. squamipinnis* and *O. karongae*. c) PCA analyses of the major taxonomic questions raised by this phylogenetic tree, showing the "Chambo" group (left), *O. rukwaensis* being distinct from *O.* "mtera" (middle), and subspecies of *O. placidus* and *O. shiranus* not supporting current classification (right).

There were instances of taxonomic species where individuals from different populations did not form monophyletic groups (Figure 2a, Figure S2), and PCA analyses further confirmed a lack of clustering of taxonomically identified species (Figure 2c). *O. rukwaensis* individuals from the Rukwa drainage basin were resolved as a clade but were phylogenetically distinct from individuals collected from the Mtera Reservoir, which had provisionally been identified as *O. rukwaensis* based on morphology (Shechonge et al., 2019). Here, we refer to this population as *O. "*mtera*"*. We found no evidence of a taxonomic

split between *O. shiranus* and *O. placidus*, and instead *O. shiranus shiranus* was resolved as sister to *O. placidus placidus*, whereas *O. shiranus chilwae* is sister to *O. placidus rovumae*. Similarly, the analysed "Chambo" species, *O. squamipinnis* and *O. karongae*, are not reciprocally monophyletic in our anlayses (Figure 2b,c).

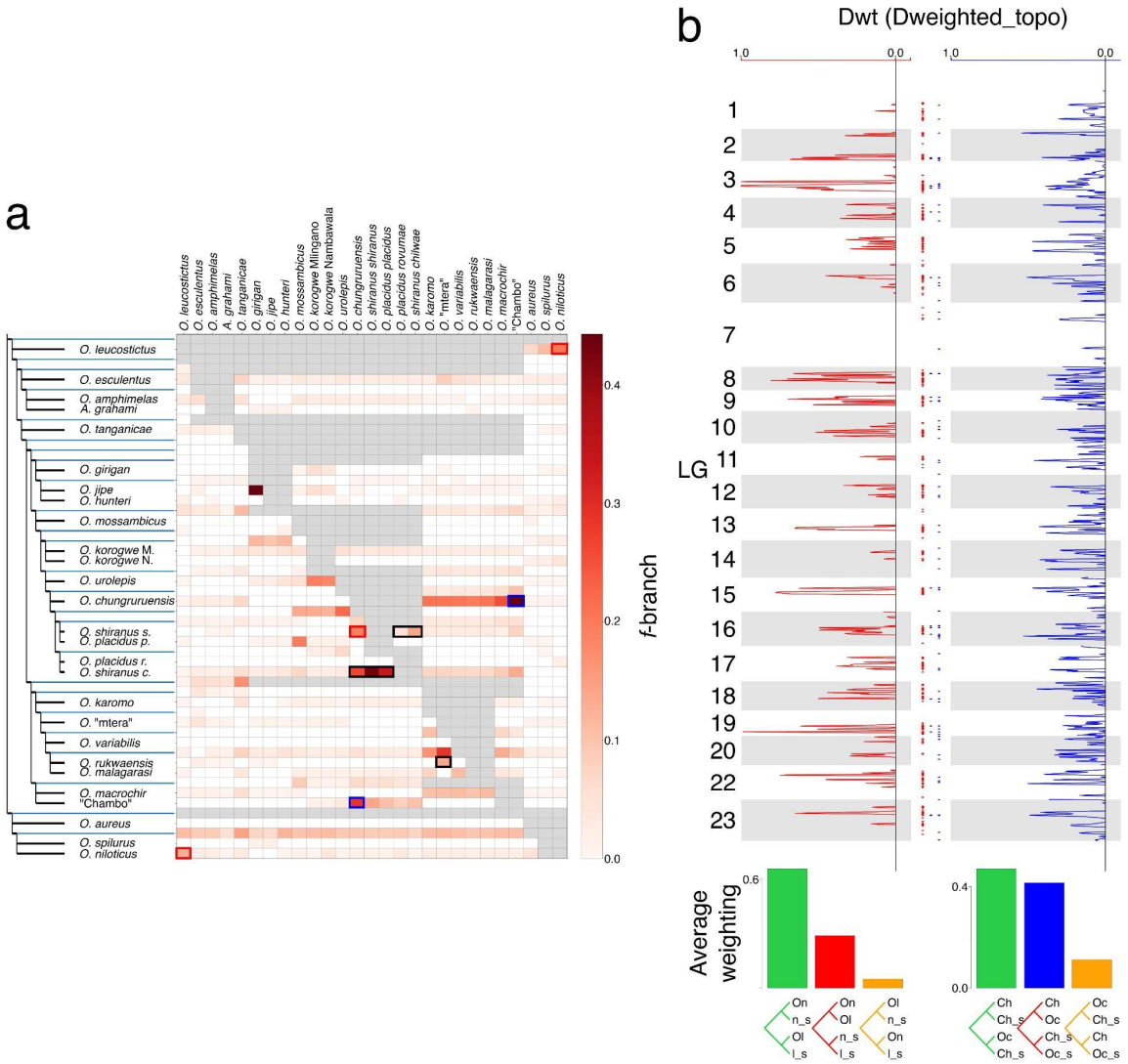
Widespread ancestral introgression

There was consistently a high degree of ancestral introgression across the *Oreochromis* phylogeny evident in our analyses. Using the genome-wide phylogenetic trees constructed from 10kb windows, 44,030 out of the 113,564 trios tested with the blt and dct tests showed significant deviations from expected values under incomplete lineage sorting alone, consistent with introgression. Similarly, the frequencies of 10kb window-tree discordant topologies indicated introgression at 42/89 nodes in the ASTRAL tree. For the SNP-based D statistics, 2,658 out of the 2,925 tested trios showed significantly higher levels of allele sharing between non-sister taxa than expected, with Dp values indicating a range of introgression proportions ranging from 0-0.45 (Table S2; for this analysis, individuals were pooled into populations or species, hence the lower number of tested trios than the blt and dct analyses).

F-branch statistics were calculated between each branch in the tree and each tip, revealing a few branches which appeared to exhibit relatively large amounts of gene flow (Figure 3a). Notably high *f*-branch values were recorded associated with two of the taxonomic uncertainties highlighted above; namely the *O. shiranus*- *O. placidus* group (especially involving *O. shiranus chilwae*) and between *O. rukwaensis* and *O. "mtera"* (Figure 3a).

We identified the counterpart for each species which had the highest *f*-branch value. These species pairs occupied the same drainage basin more often than expected by chance (p=0.0002). This was the case for seven species (compared to simulated range of 0-8 species): *O. leucostictus* / *O. niloticus*, *O. girigan* / *O. korogwe* Mlingano, *O. karomo* / *O. tanganicae*, "Chambo" / *O. chungrurensis*, *O. jipe* / *O. girigan*, *O. chungruruensis* /

“Chambo”, and *O. shiranus shiranus* / *O. chungruruensis*. Three out of these seven species pairs were related to a single putative hybrid speciation event of *O. chungruruensis* (see results below). After removing these species from the analysis, shared drainage basin less strongly predicted the *f*-branch values (p=0.03; 4 species, simulated range 0-7). Similarly, the species pairs with evidence of introgression were more likely to be more closely related than the mean phylogenetic distance between each species (Z-score <= -2) than expected by chance (p=0.0004). This was the case for four species pairs (compared to simulated range 0-4 species): *O. jipe* with *O. girigan*, *O. amphimelas* with *O. esculentus*, *O. rukwaensis* with *O.* “mtera” and *O. malagarasi* with *O. variabilis*.



320

Figure 3. a) Heatmap of the *f*-branch values calculated between each node (y axis) and tip (x axis) of the *M. zebra* mapping ASTRAL tree. Boxes indicating the introgression from "Chambo" into *O. chungruruensis* are highlighted in blue, and between *O. leucostictus* and *O. niloticus* are highlighted in red. Boxes indicating introgression amongst the clades associated with taxonomic uncertainties are highlighted in black. b) Normalised Dwt statistics across the linkage groups showing regions with a stronger weighting for the introgression tree (left of the solid line), where *O. leucostictus* and *O. niloticus* are sister (left; red), and where "Chambo" and *O. chungruruensis* are sister (right; blue). Dots represent stretches of the genome >=50kb where the four Twisst analyses consistently have Dwt = 1.0 showing introgression between *O. leucostictus* and *O. niloticus* (red), *O. squamipinnis* and *O. chungruruensis* (blue), and overlap between both comparisons (black). Abbreviations: On - *O. niloticus*, Ol - *O. leucostictus*, On_s - *O. spilurus* + *O. aureus*; Ol_s - *O. variabilis* + *O. esculentus* + *A. grahami,* Ch - "Chambo", Oc - *O. chungrurensis*, Ch_s - *O. rukwaensis* + *O. variabilis* + *O. macrochir*, Oc_s - *O. placidus rovumae* + *O. korogwe* Mlingano + *O. mossambicus*).

Based on the results of *f*-branch analyses and Dp values, we focused on ancestral introgression between *O. niloticus* and *O. leucostictus* and between *O. chungruruensis* and the "Chambo" group. These were the species pairs indicating the highest proportions of introgression according to Dp (Table S2), as well as highest *f*-branch values between non-sister or closely related branches (Figure S6). Four analyses were carried out to assess gene flow in each case (Supplementary Materials). All analysis from the Dweighted_topo statistics show generally similar putatively introgressed regions between the different analyses, although some species combinations suggested more than others (Figure S7). In both cases, very little of the genome had a higher weighting for the other alternate topology (non-species or introgression tree), as would have been indicated by Dwt < 0 (Figure 3b). In total, 52Mb of the genome was consistently found to be introgressed between *O. leucostictus* and *O. niloticus,* and 4Mb introgressed between *O. chungruensis* and "Chambo" (Figure 3b), from 761Mb in the *M. zebra* reference genome linkage groups. There was

349 significantly more of the genome introgressed in both comparisons (816Kb, with regions
350 across all 22 linkage groups; Figure 3b) than would be expected randomly (permutation test
351 *p*=0.0002; Figure S8).

352 The nature of *O. chungruruensis* introgression with “Chambo” and *O. shiranus*
353 *shiranus* was investigated by ADMIXTURE analysis assuming two populations (K=2; cross-
354 validation score = 0.48), with individuals of “Chambo”, *O. chungruruensis* and *O. shiranus*
355 *shiranus* from the Lake Malawi basin (*O. placidus* and *O. shiranus chilwae* were not
356 recorded here so was not used). This suggested that 47-48% of the *O. chungruruensis*
357 genome had affinity with “Chambo”, and 52-53% with *O. shiranus shianus*. When assuming
358 three populations (K=3; cross-validation score =0.58), *O. chungruruensis* formed its own
359 group (Figure 4). Diagnostic SNPs were also identified for “Chambo”, the *O. shiranus* - *O.*
360 *placidus* group and *O. chungruruensis* (Supplementary Results). Note that diagnostic SNPs
361 could not be detected for *O. shiranus shiranus* alone, and so the monophyletic *O. shiranus*-
362 *O. placidus* group as a whole was used. These showed similar frequencies of SNPs fixed for
363 either the reference or the homozygous alternate alleles of both “Chambo” and *O. shiranus* -
364 *O. placidus*, with around half of diagnostic SNPs heterozygous (Figure 5b). Across the 22
365 linkage groups, 4,300 windows (each of length 200 SNPs) were found where all haplotypes
366 *O. chungruruensis* and *O. shiranus*- *O. placidus* were sister taxa, compared to 2,247
367 windows where *O. chungruruensis* and “Chambo” were sister. Based on values of π and
368 Dxy within these windows, we inferred introgression times of 21,520 (CI 16,374-47,075)
369 years ago between *O. shiranus/ O. placidus* and *O. chungruruensis* and 13,565 (CI 10,321-
370 29,673) years ago between “Chambo” and *O. chungruruensis*.

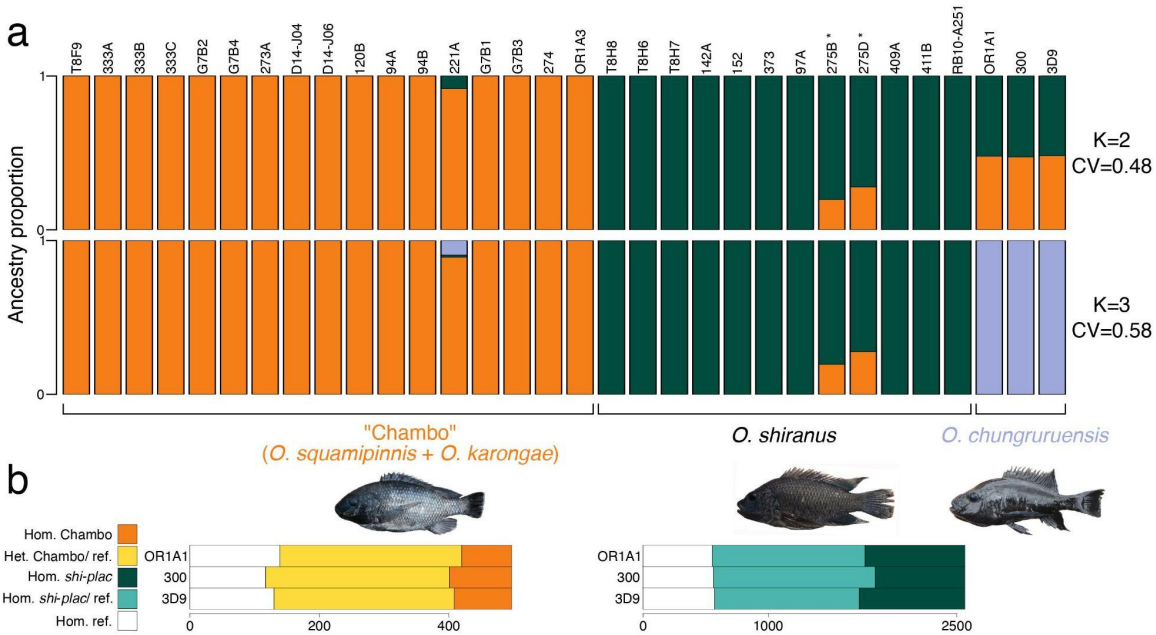


Figure 4. a) ADMIXTURE analysis of all *O. shiranus*- *O. placidus*, “Chambo” and *O. chungruruensis* at *K*=2 and *K*=3. b) The proportion of species diagnostic SNPs which are either homozygous for the alternate allele (Hom.), heterozygous (Het.), or homozygous for the reference allele (ref.) in the three *O. chungruruensis* individuals (abbreviation *shi-plac* - *O. shiranus* + *O. placidus*). * indicates genetically identified *O. shiranus shiranus* x “Chambo” hybrids in Lake Itamba.

Contemporary hybridization ongoing across Tanzania

Eleven different drainage basins and three aquaculture establishments across Tanzania were tested for the presence of hybrids (Figure 1). A total of 76 hybrids (Table 1) were identified. These were identified from across seven drainage basins: Ruvu River (30/46 tested individuals identified as hybrid), Pangani River (15/51), Rufiji River (11/56), Wami River (6/109), Lake Malawi (2/12), Lake Rukwa (2/29) and Ruvuma River (1/3) and Lake Chala (1/1). Hybrids were also identified at two aquaculture establishments: the Songea hatchery (7/50) and Mbarali Farm (1/24). It should be stressed that these numbers should not be taken as representative, as sampling was non-random and biased towards individuals thought to be hybrids on the basis of morphology.

389

390 **Table 1.** Recent hybrids identified from the supervised ADMIXTURE analysis. See Table S1

391 for exact locations and ancestry components of each hybrid. * indicates the individual with

392 the "Bandia" phenotype with Lake Chala.

393

Species pair	Number of hybrids	Drainage basins recorded (number; locations)
O. niloticus x *O. urolepis*	27	Ruvu River (10; Mindu Reservoir), Rufiji (9; Kidatu), Pangani (7; Lake Jipe), Wami River (1; Kilosa)
O. leucostictus x *O. urolepis*	20	Ruvu River (17; Mindu), Wami River (3; Kilosa (2), Lake Nala (1))
O. niloticus x *O. leucostictus* x *O. urolepis*	5	Ruvu River (3; Mindu), Wami River (2; Kilosa)
O. niloticus x *O. shiranus*	5	Songea hatchery (5)
O. niloticus x *O.* "mtera"	3	Rufiji (2; Kidatu), Mbarali Farm (1)
O. shiranus x "Chambo"	2	Lake Malawi (2; Lake Itamba)
O. leucostictus x *O. shiranus*	2	Songea hatchery (2)
O. girigan x *O. jipe*	2	Pangani (2; Lake Jipe)
O. girigan x *O. niloticus*	2	Pangani (2; Lake Jipe)
O. girigan x *O. jipe* x *O. niloticus*	2	Pangani (2; Lake Jipe)
O. girigan x *O. esculentus*	2	Pangani (2; Lake Jipe)
O. niloticus x *O. placidus rovumae*	1	Ruvuma (1; Rovuma)
O. leucostictus x *O. rukwaensis*	1	Lake Rukwa (1; Lake Rukwa)
O. esculentus x *O. leucostictus*	1	Lake Rukwa (1; Lake Rukwa)
**O. rukwaensis* x *O. urolepis*	1	Lake Chala (1; Lake Chala)

394

395 Most of the hybrids in our samples (68%) were between *O. urolepis, O. leucostictus* and *O.*

396 *niloticus*, with none directly between *O. leucostictus* and *O. niloticus*. Species diagnostic

397 SNPs, fixed in the confidently non-hybrid individuals of each of these three species, and with

398 a low frequency (<0.1) in other non-hybrid individuals, were identified for each, with numbers
399 ranging from 571 in *O. niloticus*, to 3,147 in *O. leucostictus* and 4,177 in *O. urolepis*. The
400 hybrids identified by ADMIXTURE between the three species were investigated for these
401 diagnostic SNPs, with results generally concordant with the ADMIXTURE ancestry
402 components (Figure 5). The individuals with ADMIXTURE ancestry components close to
403 50% had high levels of heterozygosity for both species, potentially indicating F1 hybrids.
404 When ancestry components rose above 50% for a species, it was generally associated with
405 more homozygous species diagnostic SNPs, suggesting backcrossing.
406

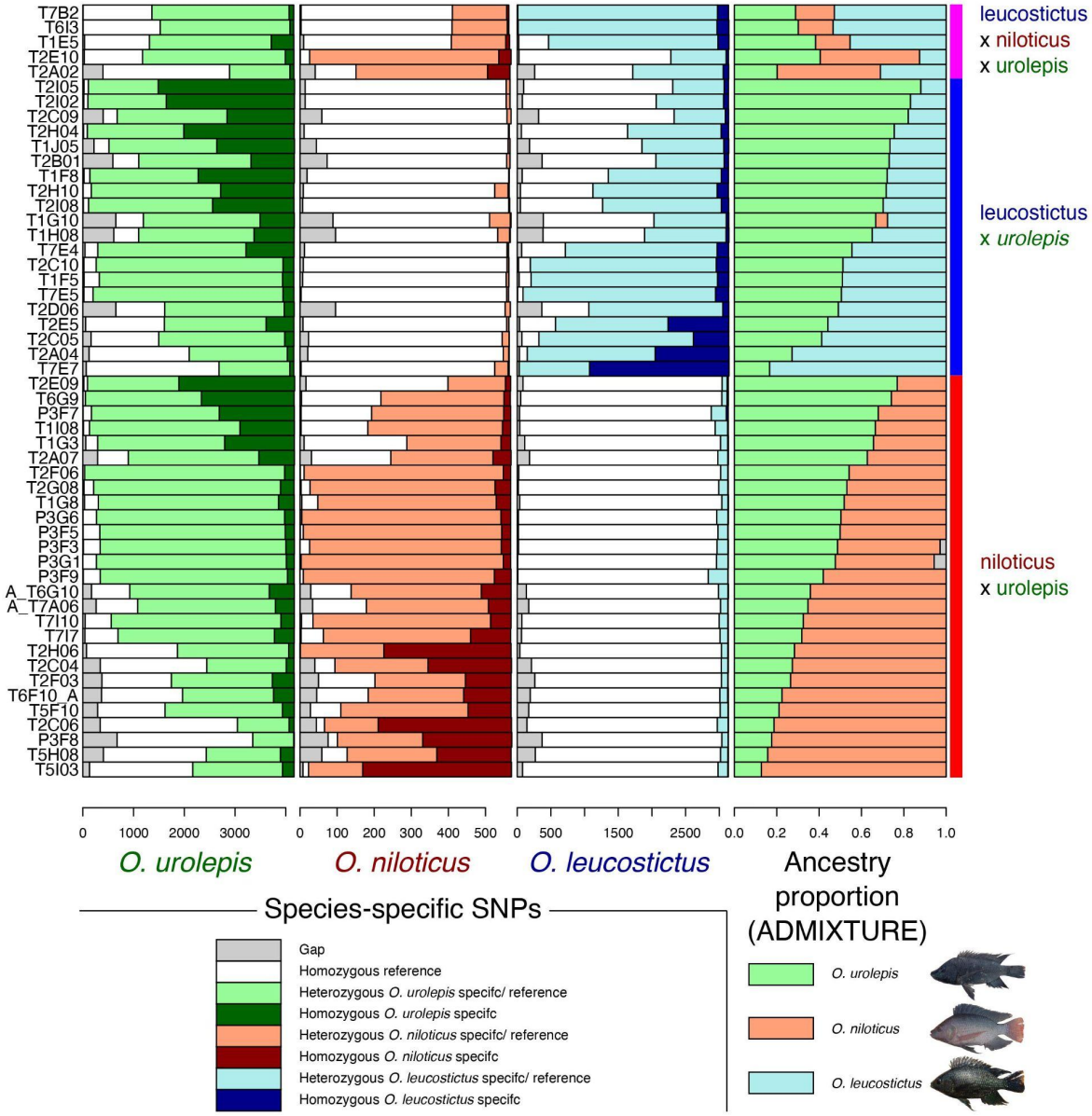


Figure 5. Species diagnostic SNP frequency for each individual inferred to be a hybrid between the potential parent species of *O. leucostictus*, *O. niloticus* and *O. urolepis* from ADMIXTURE analyses (ancestry components on far-right) (one per row). Diagnostic SNPs shown for *O. urolepis* (left), *O. niloticus* (mid-left), *O. leucostictus* (mid-right).

Only 51/124 individuals morphologically suggested to be possible hybrids were genetically confirmed, ranging across the Rufiji, Ruvu River, Wami River, Zanzibar, as well as the Songea hatchery. 24 out of the remaining 321 test individuals were genetically, but not morphologically identified as hybrids. These included individuals morphologically

assigned to *O. urolepis* from Mbarali Farm, the Ruvu River and the Wami River, *O. shiranus shiranus* from Lake Itamba and Songea hatchey, *O. niloticus* from Lake Jipe*, O. girigan* from Lake Jipe, *O. esculentus* from Lake Rukwa, *O. rukwaensis* from Lake Rukwa, *O. pangani* from Lake Kalimau and *O. placidus rovumae* from the Muhuwesi River (a tributary of the Ruvuma River).

The 'Bandia' individual, from a potentially hybrid population found in Lake Chala of unknown origin, was identified as a probable *O. rukwaensis* x *O. urolepis* hybrid, according to a panel of species diagnostic SNPs, as well as ADMIXTURE and Twisst analysis (see Supplementary Results).

Discussion

Here, we present evidence that the genus *Oreochromis*, a group of cichlid fish important for global aquaculture, has undergone introgressive hybridization multiple times in the course of its evolution, and is now experiencing large-scale contemporary hybridization resulting from anthropogenic translocations. Almost all evidence of contemporary hybrids involved populations introduced to boost aquaculture or fisheries production, threatening the conservation of wild populations in Africa.

Phylogenomic resolution of Oreochromis and taxonomic questions

We inferred a first phylogenomic backbone tree for the clade, building on previous work that was based on a handful of nuclear and mitochondrial markers (Ford et al., 2019). In accordance with Ford et al., (2019), we found that the *Alcolapia* group, comprised of species adapted to extreme soda lake environments (Trewavas, 1983), is nested within the *Oreochromis*, forming a clade with *O. esculentus* and *O. amphimelas*. This supports the conclusions of Ford et al. (2019) that *Alcolapia* is best considered as a subgenus within *Oreochromis*. We consistently found *O. amphimelas,* to be sister to *A. grahami.* This would suggest a single origin of tolerance of soda-lake conditions, which is present in *O. amphimelas*, albeit less extreme than *Alcolapia* (Trewavas, 1983). However, unlike Ford et

al., (2019), our analyses suggests that the species previously placed in the subgenus *Nyasalapia* ("Chambo", *O. rukwaensis*, *O. chungruruensis*, *O. variabilis*, *O. malagarasi*, *O. macrochir* and *O. karomo*) (Trewavas, 1983) do form a monophyletic clade, with the exception of *O. chungruruensis*, which has a hybrid origin between the "Chambo" and the *O. shiranus- O. placidus* group.

Our study raises several taxonomic questions related to species classifications within the *Oreochromis*. Two species of the subgenus *Nyaslapia*, *O. squamipinnis* and *O. karongae*, which are morphologically similar and endemic to Lake Malawi, comprise a single clade, but were not resolved as reciprocally monophyletic. Within the main lake, they have been differentiated largely on the basis of male breeding dress: male *O. karongae* are uniformly black (apart from white fin margins and pinkish genital papilla), while *O. squampinnis* males have a contrasting blaze of colour, generally blue, on the top of the head. In Lake Malawi itself, these traits are associated with other differences, including breeding seasonality, microhabitat and trophic morphology (Trewavas, 1983), but populations in crater lakes have not been thoroughly studied and have just been provisionally identified on the basis of male colour. Our results do not support a phylogenetic division based on male colour, particularly when the crater lake populations are considered. Ongoing admixture between species or parallel evolution may explain the evolution of these traits. The low number of specimens from Lake Malawi, as opposed to elsewhere in the catchment, prevents us from drawing clear conclusions about its populations. We were unable to sample the third *Nyasalapia* species from Lake Malawi, *O. lidole,* which has not been seen alive since 2007 and may be extinct (Konings, 2018). Genomic data on this species, potentially from museum specimens, would help with a taxonomic reassessment of the group.

Our results suggest that the current taxonomic status of *O. shiranus* and *O. placidus* needs revision. Each of these species is presently split into two geographically separated subspecies (Trewavas, 1983). Specifically, *O. shiranus shiranus* is found in Lake Malawi and the Upper-Middle Shire River, while *O. shiranus chilwae* is found in Lake Chilwa and Chiuta.

Meanwhile, *O. placidus placidus* is found in the Lower Shire and Lower Zambezi, while *O. placidus ruvumae* is found in the Ruvuma system. Our results suggest that the current groupings are incorrect and should be replaced by a Lake Malawi/ Shire/ Zambezi taxon and a Chilwa/ Chiuta/ Ruvuma taxon, which more closely reflects historic and present river system connectivity. There are few, if any, morphological features supporting the current classification.

We found that the *Oreochromis* population from the Mtera Reservoir (on the Ruaha section of the Rufiji system) did not cluster with *O. rukwaensis* sampled from the neighbouring Lake Rukwa catchment. Instead, it was sister to a clade consisting of *O. rukwaensis, O. variabilis* and *O. malagarasi.* This first genomic assessment of the population is consistent with the Mtera population representing a candidate new species, which we refer to as *O.* sp "mtera". There was some apparent introgression between *O. rukwaensis* and *O.* "mtera", perhaps explaining the morphological similarity that led to previous assumptions of their conspecific status (Genner et al., 2018).

Widespread ancestral introgression

We found that there was widespread, although uneven, introgression across the *Oreochromis* phylogeny. The degree of introgression was only weakly predicted by phylogenetic distance or whether species occupy the same drainage basin in the modern day. This is, however difficult to test given the highly correlated nature and lack of independence of introgression statistics, which are calculated on different subsets of four taxon across a phylogenetic tree. We may therefore have lacked power to identify these influences. We highlight two notable instances of introgression between relatively phylogenetically divergent species, primarily between the "Chambo" and *O. shiranus* (likely resulting in *O. chungruruensis*), and between *O. niloticus* and *O. leucostictus.* These latter two species coexist in their indigenous habitat of the Albertine Rift lakes Edward, George and Albert, and have been co-translocated to the same water bodies (e.g. Lake Victoria), yet are rarely found to hybridize in the modern day (Ciezarek et al., 2022; Geletu & Zhao, 2022).

Our results may be reflective of gene flow between the two species at early stages of the speciation continuum (Stankowski & Ravinet, 2021).

We found that *O. chungruruensis* could have arisen as a result of hybrid speciation between “Chambo” and either *O. shiranus shiranus* or the wider *O. shiranus/ O. placidus* group. Interestingly, *O. chungruruensis* is currently classified as a member of the subgenus *Nyasalapia* alongside *O. squamipinnis* and *O. karongae* (Trewavas, 1983), which was supported by previous studies based on a handful of nuclear and mitochondrial markers (Ford et al., 2019). *O. chungruruensis* is endemic to Lake Kiungululu, an isolated crater lake within the northern sector of the Lake Malawi basin. Neither “Chambo”, or *O. shiranus* have been recorded in this lake, although “Chambo” and *O. shiranus shiranus* are found throughout the Malawi basin and co-occur in several crater lakes in the northern sector (Kingiri, Ilamba, Itamba and Ikapu). Indeed, our analysis finds two contemporary hybrids between “Chambo” and *O. shiranus shiranus* in Lake Itamba. The age of Lake Kiungulu is not known, although the nearby crater Lake Masoko has been estimated to be around 50,000 years old (Garcin et al., 2006), relatively soon before our estimated introgression ≤30,000 years ago. This may suggest a scenario where *O. chungruruensis* originated from hybridization between ancestral populations of both within the lake.

Contemporary hybridization among Oreochromis species

Within Tanzania, there has been a long history of translocations of *Oreochromis* species for capture fisheries, dating back to at least the 1950s, in addition to ongoing aquaculture translocations, leading to the widespread colonisation by three species: *O. niloticus*, *O. leucostictus* and *O. esculentus* (Shechonge et al., 2019). Interspecific hybridization has long been recorded between *Oreochromis* species based on SNP panels, mitochondrial data and microsatellites (Geletu & Zhao, 2022), although studies based on genome-wide data are lacking. Our study has particularly demonstrated that hybridization between the invasive *O. niloticus, O. leucostictus* and the native *O. urolepis*, previously recorded in Tanzania

528 (Ciezarek et al., 2022) is widespread. Although we found no evidence of hybrids between
529 the invasive *O. leucosticus* and *O. niloticus*, we did find five individuals which had ancestry
530 components of over 0.1 (10%) for both species as well as *O. urolepis*. This suggests that
531 there may be barriers preventing *O. leucostictus* and *O. niloticus* from directly hybridizing in
532 the modern day, despite the ancestral introgression we inferred between them. Hybridization
533 from either species with *O. urolepis* may however mediate gene flow between them.
534 Previous studies, based on mitochondrial data have recorded introgression directly between
535 the two species (Ndiwa et al., 2014), although Nyingi et al. (2009) found this introgression
536 signal was not present in the nuclear genome. Furthermore, it is not clear whether there was
537 thorough species sampling in these studies, raising the possibility of 'ghost' introgression
538 explaining these findings. Further work is necessary to determine whether there are genomic
539 or behavioural barriers reducing the likelihood of *O. leucostictus* and *O. niloticus* hybridizing
540 directly.

541 Interestingly, the Lake Chala Bandia individual appears to be a case of hybridization
542 between two species not native to the lake (*O. urolepis* and *O. rukwaensis*), plausibly in farm
543 ponds prior to its introduction to the lake. Previous studies on Bandia have suggested some
544 clustering with *O. urolepis*, but with a high degree of diversity in both mitochondrial DNA and
545 morphology (Dieleman et al., 2019; Moser et al., 2019). This is consistent with our finding of
546 hybrid origin. Our study is however, based on only a single Bandia individual, and so more
547 detailed population genomic analysis would be welcome.

548

549 *Concluding remarks*

550 Several caveats need to be considered when assessing our findings. Critically, our
551 study did not include every existing species or population of *Oreochromis* found across the
552 globe. These missing or 'ghost' taxa, in addition to extinct populations of *Oreochromis* may
553 have significantly misled analyses of ancestral introgression (Tricou et al., 2022a, 2022b).
554 Extensive sampling of tilapia at both the species and population levels across Africa and the
555 Middle East will be necessary to assess this, in addition to the development of bioinformatic

556 tools which can accurately infer phylogenetic networks with large datasets, including
557 genome-scale data and many individuals. Ghost taxa may have influenced the identification
558 of recent hybrids, particularly in water bodies where *Oreochromis* was less thoroughly
559 sampled. There is always the possibility for unsampled species to be responsible for the
560 observed signals.

561 Our results highlight the complex relationship between farmed animals and their wild
562 relatives. The diversity that exists within the wild may be used to enhance that of farmed
563 populations by selective breeding (Bentsen et al., 2017). However, this potentially important
564 diversity may also be lost due to unintentional hybridization with farm escapees. This
565 accidental hybridization may also in turn have a harmful effect on farm populations, by
566 potentially introducing alleles less beneficial for farming. This dynamic is particularly
567 pronounced in *Oreochromis*, and requires ongoing genomic monitoring and assessment, to
568 meet both food security and conservation outcomes.

569

570 **Data availability**

571 All raw-read data generated will be made available in the European Nucleotide Archive upon
572 publication. All SNP datasets and inferred phylogenetic trees will be made available in dryad
573 upon publication. All custom R and python scripts will be made available on github upon
574 publication.

575

576 **Acknowledgments**

577 The authors declare no competing interests. Sequencing and library construction was
578 delivered via the BBSRC National Capability in Genomics and Single Cell (BB/CCG1720/1)
579 at Earlham Institute by members of the Genomics Pipelines Group. The computing
580 infrastructure and data centre delivered via the NBI Research Computing are supported by
581 the Biotechnology and Biological Sciences Research Council (BBSRC), part of UK Research
582 and Innovation, Core Capability Grant BB/CCG1720/1.

583

584 This work was funded by BBSRC/NERC award BB/M026736/1 (GFT, MJG, FdP), and Royal
585 Society Leverhulme Trust Africa Awards AA100023 and AA130107 (MJG, BPN, GFT, RT).
586 WH and FDP acknowledge the support of the Biotechnology and Biological Sciences
587 Research Council (BBSRC), part of UK Research and Innovation; this research was funded
588 by the BBSRC Core Strategic Programme Grants BB/CSP1720/1 and its constituent work
589 packages - BBS/E/T/000PR9818. AC, WH, FDP were supported through BBSRC GCRF
590 funding (BB/P028098/1). Samples from Tanzania were collected across 2013-2016 under
591 permit numbers 2012-393-ER-2011-103, 2014-374-ER-2011-103, 2015-83-NA-2011-103,
592 and 2016-293-NA-2011-103. Reference samples from Uganda (Lake Albert) were collected
593 in 2015 under permit number FS06/15/04. This study was approved by the Agricultural
594 Research Organization Committee for Ethics in Using Experimental Animals and was carried
595 out in compliance with the current laws governing biological research in Israel (Approval
596 number: IL-568/15). We are grateful for the technical and logistical support from staff of the
597 Tanzania Fisheries Research Institute. We thank Alan Smith, Steph Bradbeer, Carlos
598 Gracida Juarez and María Lorena Romero-Martinez for assistance with sample collection
599 and DNA extraction.
600

601 **References**

602 Ahmed, N., Thompson, S., & Glaser, M. (2019). Global Aquaculture Productivity,
603 Environmental Sustainability, and Climate Change Adaptability. *Environmental*
604 *Management*, *63*(2), 159–172.
605 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of
606 ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.
607 Bentsen, H. B., Gjerde, B., Eknath, A. E., de Vera, M. S. P., Velasco, R. R., Danting, J. C.,
608 Dionisio, E. E., Longalong, F. M., Reyes, R. A., Abella, T. A., Tayamen, M. M., &
609 Ponzoni, R. W. (2017). Genetic improvement of farmed tilapias: Response to five
610 generations of selection for increased body weight at harvest in Oreochromis niloticus

and the further impact of the project. *Aquaculture* , *468*, 206–217.

Blackwell, T., Ford, A. G. P., Ciezarek, A. G., Bradbeer, S. J., Gracida Juarez, C. A., Smith, A. M., Ngatunga, B. P., Shechonge, A., Tamatamah, R., Etherington, G., Haerty, W., Di Palma, F., Turner, G. F., & Genner, M. J. (2020). Newly discovered cichlid fish biodiversity threatened by hybridization with non-native species. *Molecular Ecology*, *3o*(4), 895–911.

Bradbeer, S. J., Harrington, J., Watson, H., Warraich, A., Shechonge, A., Smith, A., Tamatamah, R., Ngatunga, B. P., Turner, G. F., & Genner, M. J. (2019). Limited hybridization between introduced and Critically Endangered indigenous tilapia fishes in northern Tanzania. *Hydrobiologia*, *832*(1), 257–268.

Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner* (No. LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Canonico, G. C., Arthington, A., McCrary, J. K., & Thieme, M. L. (2005). The effects of introduced tilapias on native biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *15*(5), 463–483.

Champneys, T., Genner, M. J., & Ioannou, C. C. (2021). Invasive Nile tilapia dominates a threatened indigenous tilapia in competition over shelter. *Hydrobiologia*, *848*(16), 3747–3762.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , *34*(17), i884–i890.

Ciezarek, A., Ford, A. G. P., Etherington, G. J., Kasozi, N., Malinsky, M., Mehta, T. K., Penso-Dolfin, L., Ngatunga, B. P., Shechonge, A., Tamatamah, R., Haerty, W., Di Palma, F., Genner, M. J., & Turner, G. F. (2022). Whole genome resequencing data enables a targeted SNP panel for conservation and aquaculture of Oreochromis cichlid fishes. *Aquaculture* , *548*, 737637.

Conte, M. A., Joshi, R., Moore, E. C., Nandamuri, S. P., Gammerdinger, W. J., Roberts, R. B., Carleton, K. L., Lien, S., & Kocher, T. D. (2019). Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *GigaScience*, *8*(4), giz030.

Dieleman, J., Muschick, M., Nyingi, W. D., & Verschuren, D. (2019). Species integrity and origin of Oreochromis hunteri (Pisces: Cichlidae), endemic to crater Lake Chala (Kenya–Tanzania). *Hydrobiologia*, *832*(1), 269–282.

El-Sayed, A.-F. M., & Fitzsimmons, K. (2023). From Africa to the world—The journey of Nile tilapia. *Reviews in Aquaculture*, *15*(S1), 6–21.

FAO. (2022). *The State of World Fisheries and Aquaculture 2022*.

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., Stiassny, 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.

Fricke, R., Eschmeyer, W. N., & van der Laan, R. (2023). Eschmeyer's Catalog of Fishes.

Garcin, Y., Vincens, A., Williamson, D., Guiot, J., & Buchet, G. (2006). Wet phases in tropical southern Africa during the last glacial period. *Geophysical Research Letters*, *33*(7).

Geletu, T. T., & Zhao, J. (2022). Genetic resources of Nile tilapia (Oreochromis niloticus Linnaeus, 1758) in its native range and aquaculture. *Hydrobiologia*, 1–21.

Genner, M. J., Turner, G. F., & Ngatunga, B. P. (2018). A guide to the tilapia fishes of Tanzania.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722.

Hall, C., Dawson, T. P., Macdiarmid, J. I., Matthews, R. B., & Smith, P. (2017). The impact of population growth and climate change on food security in Africa: looking ahead to 2050. *International Journal of Agricultural Sustainability*, *15*(2), 124–135.

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, analysis, and visualization of phylogenomic Data. *Molecular Biology and Evolution*, *33*(6), 1635–1638.

Joyce, D. A., Lunt, D. H., Genner, M. J., Turner, G. F., Bills, R., & Seehausen, O. (2011). Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology: CB*, *21*(3), R108–R109.

Konings, A. (2018). *Oreochromis lidole. The IUCN Red List Of Threatened Species 2018: E. T61276A47243265*.

Lévêque, C. (1995). Role and consequences of fish diversity in the functioning of African freshwater ecosystems: a review. *Aquatic Living Resources*, *8*(1), 59–78.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Lind, C. E., Brummett, R. E., & Ponzoni, R. W. (2012). Exploitation and conservation of fish genetic resources in Africa: issues and priorities for aquaculture development and research. *Reviews in Aquaculture*, *4*(3), 125–141.

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, *21*(2), 584–595.

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, *2*(12), 1940–1955.

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544.

Martin, S. H., & Van Belleghem, S. M. (2017). Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics*, *206*(1), 429–438.

McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., Kishe, M. A., O'Meara, B., Bruggmann, R., Excoffier, L., & Seehausen, O. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, *586*(7827), 75–79.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O.

(2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*, 14363.

Minh, B. Q., Hahn, M. W., & Lanfear, R. (2020). New Methods to Calculate Concordance Factors for Phylogenomic Datasets. *Molecular Biology and Evolution*, *37*(9), 2727–2733.

Moser, F. N., van Rijssel, J. C., Ngatunga, B., Mwaiko, S., & Seehausen, O. (2019). The origin and future of an endangered crater lake endemic; phylogeography and ecology of Oreochromis hunteri and its invasive relatives. *Hydrobiologia*, *832*(1), 283–296.

Naylor, R. L., Hardy, R. W., Buschmann, A. H., Bush, S. R., Cao, L., Klinger, D. H., Little, D. C., Lubchenco, J., Shumway, S. E., & Troell, M. (2021). A 20-year retrospective review of global aquaculture. *Nature*, *591*(7851), 551–563.

Ndiwa, T. C., Nyingi, D. W., & Agnese, J.-F. (2014). An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Loboi drainage, Kenya. *PloS One*, *9*(9), e106972.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Nyingi, D., De Vos, L., Aman, R., & Agnèse, J.-F. (2009). Genetic characterization of an unknown and endangered native population of the Nile tilapia Oreochromis niloticus (Linnaeus, 1758) (Cichlidae; Teleostei) in the Loboi Swamp (Kenya). *Aquaculture* , *297*(1), 57–63.

Olave, M., Nater, A., Kautt, A. F., & Meyer, A. (2022). Early stages of sympatric homoploid hybrid speciation in crater lake cichlid fishes. *Nature Communications*, *13*(1), 5893.

Prabu, E., Rajagopalsamy, C. B. T., Ahilan, B., Jeevagan, I. J. M., & Renuhadevi, M. (2019). Tilapia – An excellent candidate species for world aquaculture: a review. *Annual Research & Review in Biology*, 1–14.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for

whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575.

Rabiee, M., Sayyari, E., & Mirarab, S. (2019). Multi-allele species reconstruction using ASTRAL. *Molecular Phylogenetics and Evolution*, *130*, 286–296.

Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, *53*(1), 131–147.

Salzburger, W. (2018). Understanding explosive diversification through cichlid fish genomics. *Nature Reviews. Genetics*, *19*(11), 705–717.

Seehausen, O. (2006). African cichlid fish: a model system in adaptive radiation research. *Proceedings. Biological Sciences / The Royal Society*, *273*(1597), 1987–1998.

Shechonge, A., Ngatunga, B. P., Bradbeer, S. J., Day, J. J., Freer, J. J., Ford, A. G. P., Kihedu, J., Richmond, T., Mzighani, S., Smith, A. M., Sweke, E. A., Tamatamah, R., Tyers, A. M., Turner, G. F., & Genner, M. J. (2019). Widespread colonisation of Tanzanian catchments by introduced *Oreochromis* tilapia fishes: the legacy from decades of deliberate introduction. *Hydrobiologia*, *832*(1), 235–253.

Shechonge, A., Ngatunga, B. P., Tamatamah, R., Bradbeer, S. J., Harrington, J., Ford, A. G. P., Turner, G. F., & Genner, M. J. (2018). Losing cichlid fish biodiversity: genetic and morphological homogenization of tilapia following colonization by introduced species. *Conservation Genetics* , *19*(5), 1199–1209.

Singhal, S., Derryberry, G. E., Bravo, G. A., Derryberry, E. P., Brumfield, R. T., & Harvey, M. G. (2021). The dynamics of introgression across an avian radiation. *Evolution Letters*, *5*(6), 568–581.

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution; International Journal of Organic Evolution*, *75*(6), 1256–1273.

Sturmbauer, C. (1998). Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. *Journal of Fish Biology*, *53*(sA), 18–36.

Suvorov, A., Kim, B. Y., Wang, J., Armstrong, E. E., Peede, D., D'Agostino, E. R. R., Price, D. K., Waddell, P., Lang, M., Courtier-Orgogozo, V., David, J. R., Petrov, D., Matute, D.

R., Schrider, D. R., & Comeault, A. A. (2022). Widespread introgression across a phylogeny of 155 Drosophila genomes. *Current Biology: CB*, *32*(1), 111–123.

Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., Genner, M. J., Turner, G. F., & Durbin, R. (2020). Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Molecular Biology and Evolution*, *37*(4), 1100–1113.

Svardal, H., Salzburger, W., & Malinsky, M. (2021). Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes. *Annual Review of Animal Biosciences*, *9*, 55–79.

Trewavas, E. (1983). Tilapia Fishes of the Genera Sarotherodon, Oreochromis and Danakilia. *British Museum (Natural History), London*, *152*.

Tricou, T., Tannier, E., & de Vienne, D. M. (2022a). Ghost lineages can invalidate or even reverse several results on gene flow. *PLoS Biology*, *20*(9), e3001776.

Tricou, T., Tannier, E., & de Vienne, D. M. (2022b). Ghost lineages highly influence the interpretation of introgression tests. *Systematic Biology*, *71*(5), 1147-1158.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(Suppl 6), 153.