

Newly discovered cichlid fish biodiversity threatened by hybridization with non-native species

Blackwell, Tabitha; Ford, Antonia G. P.; Ciezarek, Adam G.; Bradbeer, Stephanie J.; Gracida Juarez, Carlos A.; Smith, Alan M.; Ngatunga, Benjamin P.; Shechonge, Asilatu; Tamatamah, Rashid; Etherington, Graham; Haerty, Wilfried; Di Palma, Federica; Turner, George F.; Genner, Martin J.

Molecular Ecology

DOI:
[10.1111/mec.15638](https://doi.org/10.1111/mec.15638)

Published: 01/02/2021

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
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. (2021). Newly discovered cichlid fish biodiversity threatened by hybridization with non-native species. *Molecular Ecology*, 30(4), 895-911.
<https://doi.org/10.1111/mec.15638>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **Newly discovered cichlid fish biodiversity threatened by**
2 **hybridization with non-native species**

3
4 Tabitha Blackwell^{1*}, Antonia G.P. Ford^{2*}, Adam G. Ciezarek^{3*}, Stephanie J. Bradbeer¹,
5 Carlos A. Gracida Juarez¹, Alan M. Smith⁴, Benjamin P. Ngatunga⁵, Asilatu Shechonge⁵,
6 Rashid Tamatamah⁵, Graham Etherington³, Wilfried Haerty³, Federica Di Palma^{3,6},
7 George F. Turner⁷ and Martin J. Genner¹

8
9 ¹ School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall
10 Avenue, Bristol, BS81TQ, United Kingdom

11 ² Department of Life Sciences, Whitelands College, Holybourne Avenue, University of
12 Roehampton, London, SW154JD, United Kingdom

13 ³ Earlham Institute, Norwich Research Park Innovation Centre, Colney Ln, Norwich, NR47UZ,
14 United Kingdom

15 ⁴Department of Biological Sciences University of Hull, Hull, HU6 7RX

16 ⁵Tanzania Fisheries Research Institute (TAFIRI) P.O. Box 9750. Dar es Salaam. Tanzania.

17 ⁶ Department of Biological and Medical Sciences, University of East Anglia, Norwich,
18 NR47TJ, UK

19 ⁷ School of Biological Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, United
20 Kingdom

21 *Authors contributed equally

22
23 Correspondence: Martin J. Genner, School of Biological Sciences, University of Bristol, Life
24 Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, United Kingdom. Email
25 m.genner@bristol.ac.uk

26 **Abstract**

27

28 Invasive freshwater fish systems are known to readily hybridize with indigenous congeneric
29 species, driving loss of unique and irreplaceable genetic resources. Here we reveal that newly
30 discovered (2013-2016) evolutionarily significant populations of Korogwe tilapia (*Oreochromis*
31 *korogwe*) from southern Tanzania are threatened by hybridization with the larger invasive Nile
32 tilapia (*Oreochromis niloticus*). We use a combination of morphology, microsatellite allele
33 frequencies and whole genome sequences to show that *O. korogwe* from southern lakes
34 (Nambawala, Rutamba and Mitupa) are distinct from geographically-disjunct populations in
35 northern Tanzania (Zigi River and Mlingano Dam). We also provide genetic evidence of *O.*
36 *korogwe* x *niloticus* hybrids in three southern lakes and demonstrate heterogeneity in the
37 extent of admixture across the genome. Finally, using the least admixed genomic regions we
38 estimate that the northern and southern *O. korogwe* populations most plausibly diverged
39 approximately 140,000 years ago, suggesting that the geographical separation of the northern
40 and southern groups is not a result of a recent translocation, and instead these populations
41 represent independent evolutionarily significant units. We conclude that these newly-
42 discovered and phenotypically unique cichlid populations are already threatened by
43 hybridization with an invasive species, and propose that these irreplaceable genetic resources
44 would benefit from conservation interventions.

45

46 **Keywords:** Introgression, admixture, biodiversity conservation, cichlid fishes, population
47 genomics, geometric morphometrics.

48 Freshwater ecosystems are undergoing rapid changes in biodiversity due to the interacting
49 effects of habitat degradation, over-exploitation, water pollution, flow modification and species
50 invasion (Sala *et al.* 2000; Dudgeon *et al.* 2006; Millennium Ecosystem Assessment, 2016).
51 As human population sizes continue to rise, and climate change becomes an ever-increasing
52 threat, these impacts are predicted to grow (Martinuzzi *et al.* 2014; Arroita *et al.* 2017; Kalacska
53 *et al.* 2017). A specific issue is hybridization between introduced species and native fish
54 species. This has been reported in closely-related species from multiple fish families, including
55 the salmonids (Muhlfield *et al.* 2014; Mandeville *et al.* 2019), cichlids (Firmat *et al.* 2013;
56 Shechonge *et al.* 2018) and cyprinids (Almodóvar *et al.* 2012; Hata *et al.* 2019), and is likely
57 to become increasingly common due to the spread of freshwater species for aquaculture and
58 inland fisheries enhancement (Deines *et al.* 2014). However, the full evolutionary and
59 ecological consequences of hybridization between invasive and native species are typically
60 unclear, and further studies of the impact of hybridization events on native biodiversity are
61 required.

62
63 African inland fisheries depend heavily on “Tilapias” (Brummett & Williams, 2000), a group of
64 cichlids that includes the commercially important genera *Oreochromis*, *Sarotherodon* and
65 *Coptodon*. Among the most favoured of these species is the Nile tilapia, *Oreochromis niloticus*,
66 which has broad physiological tolerances of environmental conditions, potential for rapid
67 growth, and thus has been widely translocated across the continent (Josupeit, 2010; Dienes
68 *et al.* 2014). However, because of these traits the species is also highly invasive within its
69 introduced range (Ogutu-Ohwayo, 1990; Canonico *et al.* 2005; Deines *et al.* 2017). Moreover,
70 *O. niloticus* is also known to hybridize with native *Oreochromis* species at the locations where
71 it has been introduced, for example with *Oreochromis mossambicus* in Southern Africa
72 (D’Amato, 2007), *Oreochromis esculentus* in Lake Victoria (Angienda *et al.* 2011) and
73 *Oreochromis urolepis* and *Oreochromis jipe* in Tanzania (Shechonge *et al.* 2018; Bradbeer *et al.*
74 *et al.* 2019). However, despite the growing concern surrounding the impacts of hybridization on
75 native *Oreochromis* populations, the potential loss of unique native genetic diversity due to
76 hybridization with *O. niloticus* remains poorly studied. This is an important area to study
77 because shifts in cichlid fish biodiversity and community composition can lead to fundamental
78 changes in ecosystem functioning (Lévêque 1995), and loss of potential valuable genomic
79 resources for future *Oreochromis* aquaculture strain development (Eknath & Hulata 2009; Lind
80 *et al.* 2012).

81
82 Tanzania has a rich diversity of *Oreochromis* species, and preservation of these natural
83 species and its genetic diversity has been recognized as an important conservation goal, given
84 threats of changing environment and hybridization with invasive *Oreochromis* species

85 (Shechonge *et al.* 2018). Recently (between 2013 and 2016) populations of *Oreochromis*
86 *korogwe* were discovered in three lakes in southern Tanzania near Lindi (Lakes Rutamba,
87 Nambawala and Mitupa; hereafter referred to as ‘southern populations’; Fig. 1). Previously this
88 species was only known from the Pangani and Zigi river catchments in northern Tanzania
89 (hereafter referred to as ‘northern populations’; Fig. 1), some 500 km north of Lindi (Trewavas,
90 1983; Bradbeer *et al.* 2018; Shechonge *et al.* 2019); the holotype is a specimen from Korogwe
91 in the Pangani catchment (Lowe, 1955). The close evolutionary relationship between
92 representatives of the northern and southern populations has been confirmed in a recent
93 genus-level phylogeny, based on ~3000 bp of nuclear DNA across six loci and ~1500bp of
94 mtDNA (Ford *et al.* 2019, where they were referred to as *O. korogwe* and *O. sp.* Rutamba,
95 respectively). Importantly, the rivers between Lindi and the Pangani are populated naturally
96 only by *O. urolepis*. Such a large geographic discontinuity in the apparent natural distribution
97 of *Oreochromis* is not known in any other species (Trewavas 1983, Shechonge *et al.* 2019),
98 and is rare in other African freshwater fishes (e.g. Skelton 2001). Importantly, in all three of the
99 southern lakes studied, the invasive *O. niloticus* was also found, and the presence of
100 phenotypically intermediate individuals suggested the presence of hybrids.

101
102 In this study we aimed to characterize the diversity and origins of the newly discovered
103 southern populations of *O. korogwe*. We first quantified the extent of hybridization between
104 these populations and invasive Nile tilapia. We then evaluated the possibility that the southern
105 population could be a newly recognized evolutionarily significant unit (*sensu* Fraser &
106 Bernatchez 2001), by comparing genetic and morphological differences with northern *O.*
107 *korogwe*. We also investigate varying levels of admixture across the genome from *O. niloticus*
108 into southern *O. korogwe*. These results demonstrate that an evolutionarily significant unit is
109 threatened by hybridization with an invasive species, and add to a growing body of evidence
110 for the heterogenous nature of admixture across genomes during hybridization events.

111 112 **Materials and Methods**

113
114 *Study sites and sample collection.*

115 *Oreochromis korogwe*, *O. niloticus* and their potential hybrids were collected from southern
116 Tanzania (Lake Rutamba, Lake Nambawala, and Lake Mitupa) on the 14 August 2013, 2-4
117 May 2015 and 21-27 October 2016 (Fig. 1; Table 1). Samples of *O. korogwe* were collected
118 from northern Tanzania (Zigi River and Mlingano Dam) on the 18 August 2015 (Fig. 1; Table
119 1). Samples were collected either using multi-mesh gill nets, a seine net, or from purchasing
120 from local fishermen. Multi-mesh nets measured 30m in length with a stretched depth of 1.5m
121 height, and 12 panels each 2.5 meters long. Mesh sizes for panels were in the following order

122 43mm, 19.5mm, 6.25mm, 10mm, 55mm, Need 8mm, 12.5mm, 24mm, 15.5mm, 5mm, 35mm
123 and 29mm. The seine net measured 30 m in length, 1.5 m in height with 25.4 mm mesh and
124 fine mesh cod end.

125
126 Other samples used for this study were *O. placidus rovumae* from Lake Chidya in the Ruvuma
127 catchment sampled on 18 August 2013, *O. placidus rovumae* from the Ruvuma river sampled
128 on 16 August 2013, *O. placidus rovumae* from the Muhuwesi river (Ruvuma drainage) sampled
129 on 17 August 2013, *O. urolepis* from Lake Lugongwe near Utete on the Rufiji river sampled on
130 11 March 2015, *O. urolepis* from Mbuyuni pool on the Wami river sampled on 22 August 2015,
131 and *O. niloticus* from within its native (rather than introduced) distribution in Lake Albert,
132 Uganda, sampled on 29 October 2015 (Tables S1, S2). Field collected samples were
133 preserved either in 96-100% ethanol or DMSO salt buffer.

134 135 *Population genetics – microsatellite genotyping*

136 DNA was extracted from fin clips using the Wizard kit from Promega (Madison, WI). Samples
137 were genotyped at 13 microsatellite loci (Table S3), sourced from Saju *et al.* (2010) and Liu *et*
138 *al.* (2013), within two multiplex reactions for each sample. The first contained 6 loci and the
139 second 7 loci. Polymerase Chain Reaction (PCR) was performed using solutions comprising:
140 1 μ l DNA, 0.2 μ l of each 10 μ M forward primer, 0.2 μ l of each 10 μ M reverse primer, 5 μ l 2x Qiagen
141 Multiplex PCR Master Mix, and made up to 10 μ l using RNase-free water. PCR was conducted
142 on a 3PRIME X/02 thermocycler (Techne), with the following settings: an initial denaturation
143 at 95°C for 60 seconds, followed by 35 cycles of 94°C for 30 seconds, 57°C for 90 seconds,
144 and 72°C for 60 seconds. The final extension stage was 60°C for 30 minutes. Products were
145 genotyped on an Applied Biosystems 3500 Genetic Analyser alongside a LIZ500 size
146 standard. Peaks were identified automatically using the software Genemapper v4.1 (Applied
147 Biosystems; CA) and checked manually for accuracy. Arlequin v3.5 (Excoffier and Lischer,
148 2010) was used to summarize genetic diversity of populations and test for deviations from
149 Hardy Weinberg Equilibrium.

150 151 *Population genetics – microsatellite evidence of hybridization in the southern lakes*

152 Potential hybrid individuals between *O. korogwe* and *O. niloticus* were identified from
153 microsatellite data using a two-step process. 1) For all three lakes simultaneously, the
154 find.clusters function in the R package adegenet v2.1.1 (Jombart and Ahmed 2011) was
155 applied, selecting max.n.clust = 40, and the maximum number of principal components, to
156 make a preliminary assignment of individuals to two genetic clusters ($K = 2$), representing *O.*
157 *korogwe* and *O. niloticus*. 2) Structure v2.3.4 (Pritchard *et al.* 2000) was used to quantify
158 probability of assignment of individuals to the two species. Structure runs used $K = 2$ with the

159 adegenet find.clusters assignments as a prior. The admixture model was used, with each run
160 including 100,000 steps as burn-in, followed by 100,000 sampled steps. Runs were repeated
161 a total of 10 times, and Structure results were summarized across the runs using Clumpak
162 (Kopelman *et al.* 2015), with putatively purebred individuals identified as those possessing >
163 0.9 probability of belonging to either *O. korogwe* or *O. niloticus*, and the remainder considered
164 to be putative *O. niloticus* x *korogwe* hybrids. To ordinate the genetic structure present within
165 the southern lakes, a Factorial Correspondence Analysis in Genetix v4.05 was used (Belkhir
166 *et al.* 1999).

167

168 *Population genetics – microsatellite differences between northern and southern O. korogwe.*

169 The genetic structure of putative purebreds from the southern *O. korogwe* populations (Lake
170 Nambawala and Lake Rutamba) to the northern *O. korogwe* populations (Zigi River and
171 Mlingano Dam) was compared, as well as *O. placidus* (Lake Chidya) and *O. urolepis* (Rufiji
172 river at Utete) (Table S4). *Oreochromis korogwe* individuals from Lake Mitupa were not
173 included in the analysis due to the small sample size of purebred individuals ($n = 6$). Structure
174 v2.3.4 (Pritchard *et al.* 2000) was used to assess population genetic structure, using sampling
175 location as a prior. The admixture model was selected, with each run including 100,000 steps
176 as burn-in, followed by 100,000 sampled steps. Runs for each potential number of clusters K
177 (between 2 and 6), were repeated a total of 10 times, and the results were summarized using
178 Clumpak (Kopelman *et al.* 2015). Within Clumpak the Evanno method (Evanno *et al.* 2005)
179 was used to identify the optimal number of clusters present in the data. A Factorial
180 Correspondence Analysis in Genetix 4.05 was used to ordinate the genetic structure (Belkhir
181 *et al.* 1999). Genetic structure among the populations was estimated in Genepop v4.2
182 (Rousset, 2008) using F_{ST} and the significance of differences among populations was
183 estimated using Exact tests with default settings.

184

185 *Whole genome resequencing - library preparation and data analysis*

186 Twelve samples were processed for whole genome resequencing, comprising two *O. niloticus*
187 specimens, two *O. urolepis* specimens, two *O. placidus* specimens, three specimens from a
188 northern *O. korogwe* population (Mlingano Dam) and three specimens from a southern *O.*
189 *korogwe* population (Lake Nambawala) (Tables S1 and S2). The selection of these specimens
190 was based on phenotypic characters, and they were all assumed to be purebred at the time of
191 selection for WGS analysis. DNA was extracted from fin clips using a PureLink Genomic DNA
192 extraction kit (ThermoFisher, MA, USA). Genomic libraries were prepared using the Illumina
193 TruSeq HT paired-end read protocol, by Earlham Institute Pipelines department. Samples
194 were sequenced using an Illumina HiSeq 2500 with version 4 chemistry (10 samples per lane;
195 target 5X coverage per sample) and a 125bp paired end read metric. Initial data handling and

196 quality analysis included demultiplexing and conversion to FASTQ files, followed by use of
197 FASTQC (Andrews, 2010) for quality analysis of FASTQ files.

198

199 *Whole genome resequencing - Read mapping and SNP calling*

200 Reads were mapped against the “GCF_001858045.2” reference *Oreochromis niloticus*
201 assembly (Conte *et al.* 2019) from NCBI, using the default settings of BWA-MEM v0.7.17 (Li
202 2013), with the output bam files subjected to samtools v1.9 (Li *et al.* 2009) fixmate prior to
203 being sorted by co-ordinate. Duplicate reads were then marked using picardtools (v1.140;
204 <http://broadinstitute.github.io/picard>). SNPs were then called using gatk (v4.1.6.0) (McKenna
205 *et al.* 2010). First, HaplotypeCaller was used on each sample, using min-pruning 1, min-
206 dangling-branch-length 1 and heterozygosity 0.01. All samples were collated using
207 GenomicsDBImport, before joint-genotyping with GenotypeGVCFs. SNPs within 5 base pairs
208 of an indel were removed using BCFtools v1.10.2, and then SNPs with total depth exceeding
209 180 (average exceeding 15x coverage per sample), quality-by-depth less than 2, FS greater
210 than 10, MQ less than 30, MQRankSum less than -2, ReadPosRankSum less than -2 or SOR
211 greater than 3 were filtered using GATK VariantFiltration (Table S5). Individual genotypes with
212 depth less than 3 were replaced with a no-call. BCFtools v1.10.2 was then used to remove
213 sites which overlapped with indels in some samples, and remove SNPs which fell in scaffolds
214 other than the inferred linkage groups.

215

216 *Whole genome resequencing – PCA, ADMIXTURE and phylogenetic analysis*

217 For PCA and ADMIXTURE analysis, biallelic SNPs within the linkage groups, with a minor-
218 allele count of at least 3 and less than 25% missing taxa per site were extracted. These were
219 filtered for linkage-disequilibrium using PLINK v2.0.0 (Purcell *et al.* 2007), removing SNPs with
220 $r^2 > 0.2$ in sliding windows of 50 SNPs, with 10 SNP overlap. PCA analysis on the resulting
221 160,883 SNPs was then carried out in PLINK, with the top 20 principal components reported.
222 To investigate population membership, we used Bayesian clustering in ADMIXTURE v1.3.0
223 (Alexander *et al.* 2009) on the same SNP dataset. which uses a similar algorithm to the
224 Structure program used for the microsatellite analysis, but runs more quickly on large datasets.
225 ADMIXTURE analysis was run using the main algorithm, from $K = 1$ to $K = 6$, with default
226 values for cross-validation error estimation.

227

228 For the nuclear phylogeny, SNPs with at least one homozygous reference and one
229 homozygous alternate site were extracted. A phylogenetic tree was inferred using RAxML
230 v8.0.20 (Stamatakis 2014) and the GTRGAMMA model of evolution, with the Lewis
231 ascertainment bias correction and 200 rapid bootstraps. To examine the mitochondrial
232 phylogeny, *de novo* assemblies were produced from the raw reads for each individual, using

233 mtArchitect (Lobon *et al.* 2016), which accounts for nuclear mitochondrial DNA segments.
234 These assemblies were aligned using MAFFT v7.271 (Kato and Standley 2013). A
235 phylogenetic tree was then inferred using RAxML, the GTRGAMMA model of evolution and
236 200 rapid bootstraps.

237

238 *Whole genome resequencing - differentiation across the genome*

239 Relative genetic differentiation between populations (Weir and Cockerham F_{ST}) as well as
240 absolute sequence divergence within (π) and between (Dxy) populations were calculated in
241 non-overlapping 50kb windows using popgenWindows.py
242 (https://github.com/simonhmartin/genomics_general). For this analysis, SNPs were filtered to
243 include only sites with at least two individuals per population. Both π and Dxy require counts
244 of all sites in a window, including SNPs and monomorphic sites. To get the number of callable
245 sites across the genome, we used the CallableLoci function within GATK v3.7.0 (McKenna *et al.*
246 2010) and a custom script to get counts in each 50kb window. Inferred values of Dxy and
247 π from popgenWindows.py were then corrected to account for monomorphic sites, which were
248 not in the input vcf, by multiplying them by the number of SNPs in the windows, and then
249 dividing by the total number of callable sites. The *O. placidus* samples were not included as
250 one specimen was evidently a hybrid (see Results).

251

252 We also used Twisst (Martin and Van Belleghem 2017) to explore phylogenetic relationships
253 across the genome. Although we did not perform phasing and imputation for the main whole
254 genome dataset analysis due to the small sample size, it is useful for phylogenetic analysis
255 and likely to be accurate over the short (50-SNP) regions considered in the Twisst analysis
256 (discussed further in Martin & Belleghem 2017). We therefore first performed phasing and
257 imputation of biallelic SNPs with minor-allele count of at least three and less than 3 missing
258 taxa using Beagle v4.1 (Browning and Browning 2007) with a window size of 10,000 and
259 overlap of 1000 SNPs. Phylogenetic trees were inferred over sliding 50-SNP windows
260 (requiring at least 40 SNPs per individual), with a 10 SNP overlap using IQtree v1.6.12
261 (Nguyen *et al.* 2015) using the best fit model for each, with ascertainment bias correction,
262 using scripts modified from genomics_general
263 (https://github.com/simonhmartin/genomics_general). We then ran Twisst to calculate
264 topology weightings for each window using the method 'complete'. A smoothing parameter
265 was applied with a loess span of 500,000 base pairs, with a 25,000 spacing.

266

267 *Divergence times*

268 We used estimates of Dxy to estimate divergence times between *O. korogwe* from Mlingano,
269 and Nambawala. To convert estimates of Dxy to divergence times, we used the genome-wide

270 mutation (μ) estimate of 3.5×10^{-9} (95% confidence interval: 1.6×10^{-9} to 4.6×10^{-9}) per bp per
271 generation as recently estimated for haplochromine cichlids in Malinsky *et al.* (2018) and
272 assumed a generation time of one year. This was chosen because studies of wild populations
273 of *Oreochromis* species suggest that generation time varies from 3-36 months and is
274 dependent on habitat and population density, with populations in shallow-water and inshore
275 habitats maturing at 12 months or less (Lowe-McConnell 1982). Given the small adult body
276 size of *O. korogwe* and its occurrence in shallow eutrophic water bodies, we used a generation
277 time at the lower end of this range of 1 year.

278
279 Estimates of Dxy between the Mlingano and Nambawala *korogwe* will be increased in genomic
280 regions involved with introgression or incomplete lineage sorting. Using the Twisst output, we
281 identified windows where the weighting of the species tree was 1, i.e. there is no evidence for
282 discordance. Using bedtools (v2.28.0) (Quinlan and Hall 2010), we found the 50kb windows
283 overlapping these regions, and used Dxy from these regions to get a measure of divergence
284 in windows supporting the species tree.

285
286

287 *D3 statistics*

288 The genotypes used for sliding window F_{ST} , Dxy and pi analysis were using to calculate
289 pairwise-distances between each individual, in 50kb non-overlapping windows across the
290 genome, using *distMat.py* (https://github.com/simonhmartin/genomics_general). This
291 pairwise-distance was corrected using the number of callable sites per window (see that
292 section of the methods). *D3* statistics can be used to test for introgression between either P3
293 and P2 or P3 and P1 in a three-taxon phylogeny (P3,(P2,P1));, without the presence of an
294 outgroup, using genetic distances. Introgression would be expected to result in reduced
295 genetic distance between the two taxon in question. Using the equation $D3 = (dP1P3 - dP2P3) / (dP1P3 + dP2P3)$
296 / (dP1P3 + dP2P3); where $dP1P3$ is the distance between P1 and P3 and $dP2P3$ is the
297 distance between P2 and P3, a result where *D3* is significantly less than 0 indicates
298 introgression between P1 and P3, whereas a result where *D3* is significantly greater than 0
299 indicates introgression between P2 and P3 (Hahn and Hibbins 2019). Significance was
300 assessed by 1000 block bootstrap replicates, with the standard deviation used to calculate p
301 values using the overall mean *D3*. The test was carried out between all trios of species where
302 P1 was an individual from *O. korogwe* Nambawala, P2 was an individual from *O. korogwe*
303 Mlingano and P3 was an individual from either *Oreochromis niloticus* or *Oreochromis urolepis*.

304
305
306

307 *Geometric morphometrics – analyses of individuals from the southern lakes*

308 Ethanol preserved specimens were photographed on their left side in standard orientation with
309 a scale. The image was calibrated for size and 24 landmarks (Fig. S1) were placed onto the
310 image of each specimen using tpsDIG 1.40 (Rohlf, 2004). All microsatellite-genotyped fish
311 (See below) were included in geometric morphometrics, except for specimens of *O. korogwe*
312 where pelvic fins were naturally absent. Landmark data were subjected to a Procrustes
313 analysis in MorphoJ 1.06 (Klingenberg, 2011). Individuals were assigned to one of three
314 groups based on Structure results (purebred *O. niloticus*, purebred *O. korogwe*, hybrid *O.*
315 *niloticus* x *korogwe*). The Procrustes coordinates were then regressed against centroid size in
316 MorphoJ 1.06, and the size standardized residuals from this regression analysis were then
317 used in a stepwise Discriminant Analysis in SPSS 24 (IBM, London), with purebred *O. niloticus*
318 and purebred *O. korogwe* placed in *a-priori* known categories, and hybrid individuals
319 uncategorized.

320

321 Linear morphometric measurements were taken from each genotyped specimen collected in
322 2016 using digital calipers, following methods outlined in Barel *et al.* (1977) and Snoeks (2004).
323 The following measures were made: standard length, body depth, head length, caudal
324 peduncle length, caudal peduncle depth, dorsal fin base length, anal fin base length, pectoral
325 fin base length, pelvic fin length, caudal fin length, head width, snout length, eye length,
326 interorbital width and lower jaw length. Measurements were \log_{10} transformed and size-
327 standardized residuals generated from a linear regression against standard length. Individuals
328 were assigned to the three different groups based on Structure results (purebred *O. niloticus*,
329 purebred *O. korogwe*, hybrid *O. niloticus* x *korogwe*). The size-standardized residuals were
330 used in a Discriminant Analysis in SPSS 24 (IBM, London), with purebred *O. niloticus* and
331 purebred *O. korogwe* placed in *a-priori* known categories, and hybrid individuals remaining
332 uncategorized.

333

334 *Morphological comparisons between northern and southern O. korogwe*

335 The morphology of genetically purebred *O. korogwe* from Lakes Rutamba and Nambawala
336 (identified from microsatellite data) was compared to individuals from the Mlingano Dam and
337 Zigi River in northern Tanzania. Geometric morphometric landmarks and linear morphometric
338 measurement data were collected using the methods described above. The geometric
339 morphometric landmarks were subjected to a Procrustes standardization and the resultant
340 Procrustes coordinates were subjected to a pooled within-group regression against centroid
341 size, generating size standardized residuals. These residuals were used in a Canonical
342 Variates Analysis in MorphoJ 1.06, and a Discriminant Analysis in SPSS 24. Linear
343 morphometric measurements were \log_{10} transformed. A small number (9 of 2000) of

344 measurements were interpolated using Bayesian PCA in the R package *pcaMethods*
345 (*Stacklies et al.* 2007), allowing individuals with absent pelvic fins or damaged fins to be
346 included in analyses. We then pooled within-group regressions of each variable against
347 standard length, treating each of the four populations as a group. The size-standardized
348 residuals generated from these regressions were then used in a Discriminant Analysis in SPSS
349 24.

350

351 **Results**

352

353 *Population genetics - microsatellite analysis of purebred and hybrid Oreochromis in southern* 354 *lakes*

355 Using Structure, the majority of individuals were assigned to one of the two parent species with
356 a probability of >90%. Individuals that were not able to be assigned to a single species with a
357 probability of >90% were considered hybrids. In total these hybrids comprised 29% of
358 individuals sampled from Lake Mitupa (2 of 7), 27% of individuals from Lake Nambawala (6 of
359 22), and 6% of individuals from Lake Rutamba (2 of 32) (Fig. 2a,b).

360

361 *Morphological comparisons of purebred and hybrid Oreochromis in southern lakes*

362 Discriminant Analysis of geometric morphometric data demonstrated that *O. niloticus* and *O.*
363 *korogwe* individuals could be reliably separated (Wilk's $\lambda = 0.272$, $\chi^2 = 37.054$, $P < 0.001$) with
364 30 of 32 purebred individuals correctly classified (Table S6). Equally, Discriminant Analysis
365 using linear morphometric measurements showed that that *O. niloticus* and *O. korogwe*
366 individuals could be reliably separated (Wilk's $\lambda = 0.314$, $\chi^2 = 32.401$, $P < 0.001$), with 29 of 32
367 purebred individuals correctly classified (Table S6). Typically, *O. niloticus* were characterized
368 as possessing a longer and broader head (Table S7). Hybrid morphospace overlapped with
369 that of purebred species in both datasets (Fig. 2c).

370

371 *Population genetics – microsatellite genetic structure among Oreochromis populations*

372 Structure analyses indicated the optimum number of genetically distinct populations across
373 the six sampled populations was $K = 5$, with the southern populations from neighbouring lakes
374 Rutamba and Nambawala resolved as genetically homogeneous group (Fig 3a). All *O.*
375 *korogwe* were genetically distinct from reference populations of *O. urolepis* from the Rufiji river
376 and *O. placidus* from Lake Chidya in ordination plots (Fig. 3b). Analysis including only *O.*
377 *korogwe* revealed the Zigi river and Mlingano dam populations to be distinct from one another,
378 and to both populations from the south (Fig. 3c). In pairwise comparisons, all populations
379 showed highly significant genetic differences, with exception of *O. korogwe* from Lakes

380 Rutamba and Nambawala (Table 2). No populations showed clear patterns of significant
381 deviation from Hardy-Weinberg Equilibrium in microsatellite loci (Table S4).

382

383 *Morphological comparisons of northern and southern O. korogwe*

384 Discriminant Function Analysis of both the geometric morphometric data and the traditional
385 morphometric data demonstrated substantial differences between the northern and southern
386 *O. korogwe* groups (Fig. 3d,e), with the majority of individuals being able to be classified by
387 sampling site using either linear traditional measurement data (74 of 80 individuals), or
388 geometric morphometric data (84 of 88 individuals; Table S8). Discriminant Function Axis 1
389 separated northern and southern populations in both morphological datasets. In the linear
390 measurements this axis indicated *O. korogwe* from the northern populations to have shallower
391 body depth, a less deep caudal peduncle, a narrower interorbital width and shorter pectoral
392 fins, relative to southern populations (Table S9). Wireframe diagrams indicated northern *O.*
393 *korogwe* populations had smaller eyes and shallower body dimensions than southern
394 populations (Fig. 3f).

395

396 *Whole genome resequencing: phylogenomic analyses*

397 Illumina sequencing resulted in an average of 22 million reads per sample (range: 20.53 to
398 24.40 million), and mapping rates to the *O. niloticus* reference genome of 97.39 to 99.18%
399 (Table S1). Mean sequencing coverage across the dataset was 5.29X, with approximately half
400 the genome covered with a sequencing depth of at least 5X (Table S1). The filtered datasets
401 and number of SNPs used for downstream analysis are given in Table S10. ADMIXTURE
402 analysis of all 12 samples suggested cross-validation minima at $K = 2$ and $K = 5$, indicating
403 the most likely number of clusters in the dataset (Fig. S2). At $K = 5$, there was a clear separation
404 of the northern and southern *O. korogwe* populations alongside the other species, supported
405 by groupings in PCA (Fig. 4a). The ADMIXTURE analysis also indicated that one *O. placidus*
406 sample was likely an early-generation *O. placidus* x *niloticus* hybrid or backcross, with
407 approximately 40% *O. niloticus* cluster membership (Fig. 4b).

408

409 Maximum likelihood phylogenetic analysis indicated that the *O. placidus* hybrid was likely the
410 result of a female *O. niloticus* x male *O. placidus* cross, as the (maternally inherited) mtDNA
411 of the sample clustered with *O. niloticus* (Fig. 4d). Otherwise, there was a clear separation of
412 *O. urolepis*, *O. niloticus*, *O. placidus* and the two *O. korogwe* populations in both the nuclear
413 and mtDNA phylogenies (Fig. 4c-d).

414

415

416

417 *Whole genome resequencing: differentiation across the genome and timescale of divergence*
418 Differentiation (F_{ST}) was highest among interspecific comparisons (Fig. 5a-f). Between the
419 northern (Mlingano Dam) and southern (Nambawala) *O. korogwe* populations, most 50kb
420 windows had low differentiation, but there were prominent regions of the genome showing very
421 high F_{ST} differentiation (Fig. 5e). Notably, there were regions of relatively low genetic
422 differentiation between the *O. niloticus* and *O. korogwe* sampled from Nambawala where the
423 two species are sympatric (Fig. 5f), but these were not apparent in the comparison between
424 the fully allopatric *O. niloticus* and *O. korogwe* from Mlingano Dam (Fig. 5d). Sections of low
425 F_{ST} were also present in the comparison of *O. korogwe* from Nambawala and *O. urolepis*. In
426 general, regions of low F_{ST} showed no clear pattern of being associated with areas of elevated
427 or depleted genomic diversity (π) in the focal species (Fig. S3). However, it was notable that
428 in all species LG3 had substantially higher variability in genetic diversity relative to other
429 linkage groups, and possessed higher absolute sequence divergence in both intraspecific and
430 interspecific comparisons (Fig. S4).

431
432 Phylogenetic relationships across the genome, generated using Twisst, provided evidence of
433 admixture that was heterogeneous across the genome (Fig. 5g). The two *O. korogwe*
434 populations were resolved as sister taxa across most of the genome. However, for substantive
435 sections of the genome, a phylogeny supported *O. niloticus* and the southern *O. korogwe*
436 (Nambawala) as sister taxa, and *O. urolepis* and northern *O. korogwe* (Mlingano Dam) as
437 sisters. Notably, these tracts of the genome consistent with interspecific hybridization
438 corresponded with both the low F_{ST} regions *O. niloticus* and the southern *O. korogwe*
439 (Nambawala) (Fig 5f), and low F_{ST} region between *O. urolepis* and the northern *O. korogwe*
440 (Mlingano) (Fig 5b). $D3$ statistics consistently provided strong statistical support for scenarios
441 of both decreased genetic distance between *O. niloticus* and southern *O. korogwe* in
442 Nambawala compared to between *O. niloticus* and northern *O. korogwe*, and between *O.*
443 *urolepis* and the northern *O. korogwe* at the Mlingano Dam compared to between *O. urolepis*
444 and southern *O. korogwe* (Table S11).

445
446 Overall absolute sequence divergence (D_{xy}) between the northern (Mlingano Dam) and
447 southern (Nambawala) *O. korogwe* populations was 0.0009 (Fig. S4). Applying the genome-
448 wide mutation (μ) rate estimate of 3.5×10^{-9} (95% confidence interval: 1.6×10^{-9} to 4.6×10^{-9})
449 from Malinsky *et al.* (2018), with a generation time of one year, gave a genome-wide
450 divergence time estimate of 271 KYA (95% CI: 206-594 KYA). Using only those regions of the
451 genome consistent with the hypothesis of the northern and southern *O. korogwe* being sister
452 taxa, the overall absolute sequence divergence (D_{xy}) was 0.0005, providing a divergence time
453 estimate of 144 KYA (95% CI: 109-315 KYA).

454 **Discussion**

455

456 *Population structure of southern and northern O. korogwe.*

457

458 This study confirmed the distinctness of all sampled *O. korogwe* populations from two other
459 species of *Oreochromis* naturally present in coastal rivers of Tanzania, namely *O. placidus*
460 and *O. urolepis*. The results also demonstrated a close evolutionary relationship between *O.*
461 *korogwe* individuals in northern and southern Tanzania. Nevertheless, there has been
462 extensive morphological divergence between the northern and southern *O. korogwe*, and
463 based on least admixed sections of the genome, this divergence took place approximately
464 140,000 years ago. Therefore, the data are consistent with these taxa representing
465 independent evolutionarily significant units. The presence of a 500 km gap between the
466 sampled northern and southern populations of *O. korogwe* in Tanzania, is intriguing. In tilapiine
467 cichlids the presence of such gaps is typically due to human intervention. For example,
468 stocking has resulted in *O. niloticus* having a broad discontinuous distribution across Africa,
469 and further afield (Deines *et al.* 2014). However, our results are consistent with the current
470 distribution of *O. korogwe* being natural. The distribution may have arisen from a natural long-
471 distance colonization event, or perhaps that the species once had a wider distribution that has
472 been disrupted through either extirpation or introgression with *O. urolepis*, a species that neatly
473 fits the gap between northern and southern *O. korogwe* (Ford *et al.* 2019; Shechonge *et al.*
474 2019).

475

476 *Morphological variation among O. korogwe populations*

477 Our results showed that the northern and southern *O. korogwe* populations are largely distinct
478 in characters such as body depth, fin length and eye size morphology. The populations are
479 sufficiently divergent in morphology to warrant consideration of these as distinct species under
480 morphological species concepts. The anatomical divergence may be accompanied by
481 ecological differences, as variation in craniofacial morphology and body shape are often
482 related to resource use patterns in cichlids. For example, variation in eye size is related to
483 visual environment (Hahn *et al.* 2017), and fin morphology is related to patterns of habitat use
484 (Colombo *et al.* 2016). Little is known about the feeding habits of *O. korogwe* and detailed
485 analysis of diets and foraging environments within the sampled locations are required to
486 explore functions of the morphological variation observed. Given the allopatric nature of the
487 populations, further ecologically and developmentally-focussed work would also help to reveal
488 if the observed divergence can be attributed to fixed genetic differences, or alternatively
489 variation between environments during development (Parsons *et al.* 2011; Schneider and
490 Meyer, 2017).

491 Our microsatellite-based results also confirmed the presence of hybrids between *O. korogwe*
492 and invasive *O. niloticus* in all three of the southern lakes, with a frequency of between 6 and
493 29% of sampled individuals. This level of hybridization is likely to be an underestimate if
494 purebreds are present (Boecklen & Howard, 1997), which our genome-wide analyses also
495 support. Such hybridization between native and non-native species commonly occurs when
496 invader is closely-related to the native species, and the species pair are still reproductively
497 compatible due to an absence of strong reproductive barriers that typically isolate naturally
498 sympatric taxa (Horreo *et al.* 2011, Gainsford, 2014). It is not fully understood what factors
499 influence the extent of reproductive isolation among *Oreochromis* species. However, it is
500 notable that like many African mouthbrooding cichlids, *Oreochromis* exhibit traits indicative of
501 sexual selection based on male colours or the characteristics of breeding territory (Trewavas
502 1983). It is possible that in this case hybridization between *O. korogwe* or *O. niloticus* takes
503 place due to both species possessing dark male breeding colours (Genner *et al.* 2018). Female
504 mating decisions also biased towards larger individuals in *Oreochromis* species, most likely
505 due to the influence of male-male competition on breeding territory acquisition (Nelson 1995;
506 Fessehaye *et al.* 2006). Hence, is also conceivable that larger *O. niloticus* males have
507 effectively excluded smaller *O. korogwe* males from suitable breeding habitats; but detailed
508 survey and experimental work is required to test this hypothesis, including tests of sex-biases
509 in the direction of hybridization (e.g. Hayden *et al.* 2010; Rognon & Guyomard, 2003).

510

511 *Heterogeneity of admixture across the genome*

512

513 We conducted genome-wide scans of F_{ST} and Dxy between *O. niloticus*, *O. urolepis* and *O.*
514 *korogwe* populations. F_{ST} between the northern (Mlingano Dam) and southern (Nambawala)
515 *O. korogwe* populations was typically low across all linkage groups, with peaks of high F_{ST} that
516 may reflect genomic regions under directional selection. These peaks of the F_{ST} were not
517 clustered, and these regions associated loci associated with the divergent phenotypes of these
518 populations. These patterns are characteristic of early stage speciation under geographical
519 isolation (Seehausen *et al.* 2014).

520

521 Between *O. urolepis* and *O. niloticus* a consistent pattern of high F_{ST} was present, reflecting
522 the long divergence. On linkage group 3, F_{ST} was lower, and but it is notable that this shows
523 an unusually high level of nucleotide (π) diversity in all our studied *Oreochromis* populations
524 (Fig. S3), as well as a high level of absolute sequence divergence (Dxy) between all
525 populations (Fig. S4). On account of this linkage group being 2-3 times larger than any other
526 in the *Oreochromis* genome (Fig. 5; Conte *et al.* 2019), LG3 has been referred to as a
527 megachromosome, and is likely to consist of a fusion with an ancestral B-chromosome (Conte

528 *et al.* 2020). It is rich in long-coding RNA, genes related to immune response and regulation,
529 and repetitive elements. It has also been reported as containing a sex-determination locus in
530 *Oreochromis*, albeit not in *O. niloticus* itself (Conte *et al.* 2020). Collectively, the high genetic
531 diversity of this linkage group explains the relatively low F_{ST} observed between *O. urolepis* and
532 *O. niloticus*, and between other species pairs.

533
534 In comparisons between *O. niloticus* and southern *O. korogwe* from Lake Nambawala, there
535 was considerable heterogeneity in F_{ST} across the genome. There were notable long-tracts of
536 relatively low F_{ST} , most conspicuously on linkage groups 1, 7, 9 10, 17, 20 and 23. Many of
537 these were paralleled by low F_{ST} between *O. urolepis* and *O. korogwe* from Lake Nambawala.
538 However, the regions of low differentiation were not present in comparisons between *O.*
539 *niloticus* and northern *O. korogwe* from the Mlingano Dam, or between *O. urolepis* and *O.*
540 *korogwe* from the Mlingano Dam. This is suggestive of the observed patterns of substantive
541 genomic heterogeneity being reflective of admixture events in the south of Tanzania, after the
542 split from northern *O. korogwe* approximately 140,000 years ago.

543
544 Given our microsatellite evidence of individuals of *O. korogwe x niloticus* hybrid ancestry within
545 Lake Nambawala, tracts of low F_{ST} between *O. korogwe x O. niloticus* plausibly reflect
546 hybridization between in the southern region. The analysis of phylogenetic relationships of the
547 focal populations in this study using Twisst show that although the species tree relationship is
548 most common across the genome, there is a substantial difference in the frequency of the two
549 discordant relationships, which under incomplete lineage sorting alone would be expected to
550 have the same frequency, The observed excess of the discordant topology grouping *O.*
551 *niloticus* with *O. korogwe* Nambawala and *O. urolepis* with *O. korogwe* Mlingano (green in
552 Figure 5g) therefore suggests introgression between *O. niloticus* and *O. korogwe* Nambawala
553 or between *O. urolepis* and *O. korogwe* Mlingano. Supporting this, all *D3* analysis suggest
554 significantly lower genetic distances between *O. niloticus* and *O. korogwe* Nambawala and
555 between *O. urolepis* and *O. korogwe* Mlingano, than otherwise expected under a model of no-
556 hybridization. However, this three-taxon analysis can be confounded by introgression events
557 involving taxa that have not been included in the analysis. Introgression between *O. niloticus*
558 and *O. korogwe* Nambawala, for example, would increase average the genetic distance
559 between *O. korogwe* Nambawala and *O. urolepis*, as the genetic distance between *O. urolepis*
560 and *O. niloticus* is greater than between *O. urolepis* and *O. korogwe* Nambawala. A single
561 introgression event, between *O. niloticus* and *O. korogwe* Nambawala, could therefore explain
562 both positive results.

563

564 The genomic regions of this introgression highlighted by the Twisst analysis overlap with the
565 low F_{ST} regions between *O. niloticus* and *O. korogwe* Nambawala, but such low F_{ST} regions
566 are not observed between *O. urolepis* and *O. korogwe* Mlingano. The most congruent
567 interpretation of these F_{ST} results is introgression between *O. niloticus* and *O. korogwe*
568 Nambawala. The parallel regions of low F_{ST} present between *O. korogwe* from Lake
569 Nambawala and *O. urolepis* are unusual however, given that *O. urolepis* has never been
570 recorded inside Lake Nambawala, or elsewhere in the known range of *O. korogwe* (Shechonge
571 *et al.* 2019). One possible explanation for this pattern is that the introduced *O. niloticus*
572 population in Lake Nambawala could itself comprise *O. urolepis* x *niloticus* hybrids, as these
573 species are known to hybridise elsewhere in Tanzania (Shechonge *et al.* 2018), and it is
574 plausible that Nambawala was stocked from a hybrid population. Alternatively, these low F_{ST}
575 tracts may reflect recent admixture of ancestral variation shared by both *O. urolepis* and *O.*
576 *niloticus*. We have not sequenced the *O. niloticus* from Lake Nambawala to test for the
577 presence of recent introgression with *O. urolepis*, but this may be enlightening. We must also
578 note that the low sample sizes (n=2 to 3 individuals) will have limited the accuracy and reliability
579 of F_{ST} , Dxy and pi statistics. Further studies with more comprehensive phylogenetic and
580 population sampling with greater sample sizes may be able to untangle the nature of
581 introgression events with more precision.

582
583 Extensive heterogeneity in the extent of admixture across genomes has been reported in
584 multiple studies of closely related species, including trees (Wang *et al.* 2020), insects (Martin
585 *et al.* 2019, Valencia-Motoya *et al.* 2020) and cichlid fish (Gante *et al.* 2016, Svarel *et al.*
586 2020). Tracts of the southern *O. korogwe* genome with extensive evidence for hybridization
587 (e.g. LG7, LG9 and LG17), may have resulted from introgressed alleles in those regions being
588 favoured by selection. In North America hybridization between introduced rainbow trout
589 (*Oncorhynchus mykiss*) and native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*), has
590 led to multiple genomic variants being shared between the species, with selection repeatedly
591 favouring some introduced alleles within the native species (Bay *et al.* 2019). Adaptive
592 introgression has similarly been suggested to have led to multiple beneficial traits arising from
593 close-relatives in many species groups, including Darwin's finches (Lamichhaney *et al.* 2015),
594 snowshoe hares (Jones *et al.* 2018) and multiple plant taxa (Suarez-Gonzalez *et al.* 2018).

595
596 In comparisons of *O. korogwe* from Lake Nambawala and *O. niloticus*, regions of the genome
597 with low levels of introgression (e.g. LG6, LG16 and LG19). This may be due to the presence
598 of "barrier" loci that reduce gene flow and maintain species boundaries (Elmer *et al.* 2019). It
599 is been shown that hybridization can suppress recombination rates in some genomic regions
600 of hybrid trout (Ostberg *et al.* 2013). It has also been proposed that recombination is

601 particularly strongly suppressed near genes associated with reproductive isolation among
602 parent species, due to hybrids have a low relative fitness (Hvala *et al.* 2018). In particular,
603 hybridization could lead to the breakup of coadapted “supergene” clusters, leading to low
604 fitness hybrids, and so these large genomic regions would in principle be among most resistant
605 to introgression. Positive associations between recombination rate of genome and admixture
606 have been described in humans and swordtail fishes (Schumer *et al.* 2018), as well as
607 sympatric pairs of *Heliconius* butterflies (Martin *et al.* 2019). However, accurate estimations of
608 recombination rate require genotype data from more extensive population sampling than has
609 been undertaken for our study, so this remains an untested yet plausible explanation for at
610 least some of the heterogeneity observed.

611

612 *Conservation implications*

613 Our results support the concept that the northern and southern *O. korogwe* populations are
614 long-diverged and phenotypically-divergent evolutionarily significant units. These may require
615 consideration as discrete species, which will have implications for the biodiversity of tilapias of
616 East Africa. However, the results also illustrate that genetic structure within the newly
617 discovered populations of *O. korogwe* has already been impacted by the invasive species *O.*
618 *niloticus*. Similarly, the results also show *O. niloticus* has hybridized with *O. placidus* in the
619 neighbouring Ruvuma drainage. Species introductions can have non-reversible impacts on
620 genetic diversity (Dudgeon *et al.* 2006), and therefore the presence of this highly invasive
621 species in these lakes is of considerable concern for the long-term conservation status for
622 these populations. Hybridization could have larger impacts on the genetic diversity of this
623 population over time, especially given evidence from other lakes where *O. niloticus* have been
624 introduced (e.g. Deines *et al.* 2014) and given the lack of understanding of the long-term fitness
625 consequences of these interaction. Although there is some evidence that hybridization could
626 introduce advantageous alleles into the population, our findings suggest that these southern
627 *O. korogwe* populations are likely to be locally adapted to the southern lakes. Therefore,
628 introgression may have negative outcomes for the genetic uniqueness of the *O. korogwe*
629 populations at least.

630

631 Our results clearly demonstrate an ongoing threat to unique southern *O. korogwe* populations,
632 and long-term monitoring of the genetic and phenotypic diversity within the studied lakes will
633 yield insights into changes of their status. We suggest that clear conservation actions could be
634 implemented. Given the removal of *O. niloticus* from the southern lakes would be impractical,
635 conservation of the unique genetic resources within the southern lakes would be best done
636 through the identification of potential ark sites. For this research we sampled three of the water
637 bodies in close proximity to the towns of Lindi and Rutamba, and it is possible that *O. korogwe*

638 populations unaffected by *O. niloticus* are present in four additional proximate water bodies
639 that we have not yet been surveyed. Each of these potential ark lakes will need to be
640 intensively investigated to determine the species of fish present, and the potential for *O.*
641 *niloticus* colonisation via natural waterways. In the absence of the suitable ark sites, the *ex-*
642 *situ* conservation could be implemented. In both conservation strategies, genome-wide
643 sequencing would be useful to confirm the genetic purity of the stocks, as this study has shown
644 a clear signal of introgression in individuals of *O. korogwe* from Lake Nambawala that were
645 assumed to purebred on the basis of the phenotypes. Therefore, this study underlines the
646 value of using genome-wide sequencing for assessing the conservation status of taxa under
647 threat from hybridization with introduced species.

648

649 **Acknowledgements**

650 The work was funded by Royal Society-Leverhulme Trust Africa Awards AA100023 and
651 AA130107 to MJG, BPN and GFT, BBSRC award BB/M026736/1 to GFT, MJG and FdP, and
652 and BBSRC award BB/P016774/1 to WH and FdP We thank the Tanzania Commission for
653 Research and Technology (COSTECH) for fieldwork approval and permits, and staff of the
654 Tanzania Fisheries Research Institute for contributions to fieldwork. We thank Nasser Kazosi
655 for help with sample collection in Uganda, made under permit number IMP/GEN/2014/06. This
656 research was supported in part by the NBI Computing infrastructure for Science (CiS)
657 group through use of the CiS high-performance computing cluster for the analysis of the whole
658 genome resequence data.

659

660 **Author Contributions**

661 GFT, MJG and FDP conceived the study. MJG, GFT and BPN designed fieldwork and
662 sampling. TB, SJB, AGPF, CAGJ, BPN, AS, GFT, RT and MJG conducted or supervised
663 fieldwork, or collected data. TB, AGPF, AGC, MJG, GE and WH designed and performed the
664 analysis. TB, AGPF and MJG wrote the first draft of the manuscript. All authors commented
665 on and edited the final manuscript.

666

667 **Data Accessibility Statement**

668 - Microsatellite genotype data - <https://doi.org/10.5061/dryad.ht76hdcv>

669 - Morphological data - <https://doi.org/10.5061/dryad.ht76hdcv>

670 - DNA resequencing data (raw reads) - European Nucleotide Archive; Project number:
671 PRJEB36772; Accessions: ERS4308617- ERS4308628

672 - DNA resequencing analysis files - <https://doi.org/10.5061/dryad.ht76hdcv>

673

674

675 **References**

676

677 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry
678 in unrelated individuals. *Genome Research*, 19, 1655–1664.

679 Almodóvar, A., Nicola, G.G., Leal, S., Torralva, M. and Elvira, B. (2012). Natural hybridization
680 with invasive bleak *Alburnus alburnus* threatens the survival of Iberian endemic
681 calandino *Squalius alburnoides* complex and Southern Iberian chub *Squalius*
682 *pyrenaicus*. *Biological Invasions*, 14, 2237-2242.

683 Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
684 Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

685 Angienda, P.O., Lee, H.J., Elmer, K.R., Abila, R., Waindi, E.N., & Meyer, A. (2011). Genetic
686 structure and gene flow in an endangered native fish (*Oreochromis esculentus*)
687 compared to invasive Nile Tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa.
688 *Conservation Genetics*, 12, 243-255.

689 Arroita, M., Flores, L., Larrañaga, A., Martínez, A., Martínez-Santos, M., Pereda, O., Ruiz-
690 Romera, E., Solagaistua, L., & Elozegi, A. (2017). Water abstraction impacts stream
691 ecosystem functioning via a wetted-channel contraction. *Freshwater Biology*, 62, 243-
692 257.

693 Barel, C.D.N., van Oijen, M.J.P., Witte, F., & Witte-Maas, E.L.M. (1977). An introduction to the
694 taxonomy and morphology of the Haplochromine Cichlidae from Lake Victoria.
695 *Netherlands Journal of Zoology*, 27, 381-389.

696 Bay, R.A., Taylor, E.B. & Schluter, D. (2019). Parallel introgression and selection on introduced
697 alleles in a native species. *Molecular Ecology*, 28, 2802-2813.

698 Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (1999). GENETIX 4.05, logiciel
699 sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations,
700 Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

701 Boecklen, W.J. & Howard, D.J. (1997). Genetic analysis of hybrid zones: numbers of markers
702 and power of resolution. *Ecology*, 78, 2611-2616.

703 Bradbeer, S.J., Harrington, J., Watson, H., Warraich, A., Shechonge, A., Smith, A. ...Genner,
704 M.J. (2019). Limited hybridization between introduced and critically endangered
705 indigenous tilapia fishes in northern Tanzania. *Hydrobiologia*, 832, 257-268.

706 Brawand, D., Wagner, C., Li, Y., Malinsky, M., Keller, I., Fan, S. ... Di Palma, F. (2014). The
707 genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513, 375–381.

708 Browning, S. R. & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-
709 data inference for whole-genome association studies by use of localized haplotype
710 clustering. *American Journal of Human Genetics*, 81, 1084-1097.

711 Brummett, R.E., & Williams, M.J. (2000). The evolution of aquaculture in African rural and
712 economic development. *Ecological Economics*, 33, 193-203.

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

716 Colombo, M., Indermaur, A., Meyer, B.S., & Salzburger, W. (2016). Habitat use and its
717 implication to functional morphology: niche partitioning and the evolution of locomotory
718 morphology in Lake Tanganyikan cichlids (Perciformes: Cichlidae). *Biological Journal of
719 the Linnean Society*, 117, 536-550.

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

723 Conte, M.A., Clark, F.B., Roberts, R.B., Xu, L., Tao, W., Zhou, Q., Wang, D. and Kocher, T.D.
724 (2020). Evolution of a sex megachromosome. bioRxiv.
725 <https://doi.org/10.1101/2020.07.02.182808>

726 D'Amato, M.E., Esterhuysen, M.M., Van Der Waal, B.C., Brink, D. & Volckaert, F.A. (2007).
727 Hybridization and phylogeography of the Mozambique tilapia *Oreochromis
728 mossambicus* in southern Africa evidenced by mitochondrial and microsatellite DNA
729 genotyping. *Conservation Genetics*, 8, 475-488.

730 Deines, A.M., Bhole, I., Katongo, C., Feder, J.L. & Lodge, D.M. (2014). Hybridization between
731 native *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River,
732 Zambia. *African Journal of Aquatic Science*, 39, 23-34.

733 Deines, A.M., Bunnell, D.B., Rogers, M.W., Bennion, D., Woelmer, W., Sayers, M.J., ... Beard
734 Jr., T.D. (2017). The contribution of lakes to global inland fisheries harvest. *Frontiers in
735 Ecology and the Environment*, 15, 293-298.

736 Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z., Knowler, D.J., Leveque, C., ...
737 Sullican, C.A. (2006). Freshwater biodiversity: importance, threats, status and
738 conservation challenges. *Biological Reviews*, 81, 163-182.

739 Eknath, A.E. and Hulata, G., 2009. Use and exchange of genetic resources of Nile tilapia
740 (*Oreochromis niloticus*). *Reviews in Aquaculture*, 1, 197-213.

741 Elmer, K.R., 2017. Barrier loci and progress towards evolutionary generalities. *Journal of
742 Evolutionary Biology*, 30, 1491-1493.

743 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
744 using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.

745 Excoffier, L. & Lischer, H.E.L. (2010). Arlequin suite version 3.5: A new series of programs to
746 perform population genetics analyses under Linux and Windows. *Molecular Ecology
747 Resources*, 10, 564-567.

748 Fessehaye, Y., El-bialy, Z., Rezk, M.A., Crooijmans, R., Bovenhuis, H. & Komen, H. (2006).
749 Mating systems and male reproductive success in Nile tilapia (*Oreochromis niloticus*) in
750 breeding hapas: a microsatellite analysis. *Aquaculture*, 256, 148-158.

751 Firmat, C., Alibert, P., Losseau, M., Baroiller, J.F. & Schliewen, U.K. (2013). Successive
752 invasion-mediated interspecific hybridizations and population structure in the
753 endangered cichlid *Oreochromis mossambicus*. *PLoS ONE*, 8, e63880.

754 Fraser, D.J. & Bernatchez, L. (2001). Adaptive evolutionary conservation: towards a unified
755 concept for defining conservation units. *Molecular Ecology*, 10, 2741-2752.

756 Ford, A.G.P., Bullen, T.R., Pang, L., Genner, M.J., Bills, R., Flouri, T., ... Day, J.J. (2019).
757 Molecular phylogeny of *Oreochromis* (Cichlidae: Oreochromini) reveals mito-nuclear
758 discordance and multiple colonisation of adverse aquatic environments. *Molecular*
759 *Phylogenetics and Evolution*. **136**, 215-226.

760 Gainsford, A. (2014). Hierarchical behavior, habitat use, and species size differences shape
761 evolutionary outcomes of hybridization in a coral reef fish. *Journal of Evolutionary*
762 *Biology*, 28, 205-222.

763 Gante, H.F., Matschiner, M., Malmstrøm, M., Jakobsen, K.S., Jentoft, S. & Salzburger, W.,
764 (2016). Genomics of speciation and introgression in Princess cichlid fishes from Lake
765 Tanganyika. *Molecular Ecology*, 25, 6143-6161.

766 Genner, M.J., Ngatunga, B.P. Turner (2018) A Guide to the Tilapia fishes of Tanzania.
767 Available at: <http://bit.ly/TZTilapiaGuide>. Accessed 27 July 2020.

768 Hahn, C., Genner, M.J., Turner, G.F. & Joyce, D.A. (2017). The genomic basis of adaptation
769 to the deep water 'twilight zone' in Lake Malawi cichlid fishes. *Evolution Letters*, 1-4,
770 184–198.

771 Hahn, M.W. & Hibbins, M.S. (2019). A three-sample test for introgression. *Molecular Biology*
772 *and Evolution*, 36, 2878-2882.

773 Hata, H., Uemura, Y., Ouchi, K. and Matsuba, H., 2019. Hybridization between an endangered
774 freshwater fish and an introduced congeneric species and consequent genetic
775 introgression. *PloS ONE*, 14, e0212452.

776 Hayden, B., Pulcini, D., Kelly-Quinn, M., O'Grady, M., Caffrey, J., McGrath, A., & Mariani, S.
777 (2010). Hybridization between two cyprinid fishes in a novel habitat: genetics,
778 morphology and life-history traits. *BMC Evolutionary Biology*, 10, 1-11.

779 Horreo, J.L., Ayllon, F., Perez, J., Beall, E., and Garcia-Vazquez, E. (2011). Interspecific
780 hybridization, a matter of pioneering? Insights from Atlantic salmon and Brown trout.
781 *Journal of Heredity*, 102, 237-242.

782 Hvala, J.A., Frayer, M.E. and Payseur, B.A. (2018). Signatures of hybridization and speciation
783 in genomic patterns of ancestry. *Evolution*, 72, 1540-1552.

784 Jombart, T., & Ahmed, I. (2011). Adegnet 1.3-1: new tools for the analysis of genome-wide
785 SNP Data. *Bioinformatics*, 21, 3070-1.

786 Jones, M.R., Mills, L.S., Alves, P.C., Callahan, C.M., Alves, J.M., Lafferty, D.J., ... Good, J.M.
787 (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe
788 hares. *Science*, 360, 1355-1358.

789 Josupeit, H. (2010). *World supply and demand of tilapia 2010*. Paper presented at the Third
790 International Technical and Trade Conference and Exposition on Tilapia, 27–29 October
791 2010. Kuala Lumpur: Malaysia.

792 Kalacska, M., Arroyo-Mora, J.P., Lucanus, O., & Kische-Machumu, M.A. (2017). Land cover,
793 land use, and climate change impacts on endemic cichlid habitats in northern Tanzania.
794 *Remote Sensing*, 9, 623.

795 Katoh, K. & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7:
796 improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772-
797 780.

798 Klingenberg, C.P. (2011). MorphoJ: an integrated software package for geometric
799 morphometrics. *Molecular Ecology Resources*, 11, 353-357.

800 Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. (2015). Clumpak:
801 a program for identifying clustering modes and packaging population structure
802 inferences across K. *Molecular Ecology Resources*, 15, 1179-1191.

803 Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A.,
804 ... Grant, B.R. (2015). Evolution of Darwin's finches and their beaks revealed by genome
805 sequencing. *Nature*, 518, 371-375.

806 Lévêque, C., 1995. Role and consequences of fish diversity in the functioning of African
807 freshwater ecosystems: a review. *Aquatic Living Resources*, 8, 59-78.

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

811 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-
812 MEM. *ArXiv Preprint*. doi:arXiv:1303.3997.

813 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The
814 Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-2079.

815 Liu, F., Sun, F., Li, J., Xia, J.H., Lin, G., Tu, R.J., & Yue, G.H. (2013). A microsatellite-based
816 linkage map of salt tolerant tilapia (*Oreochromis mossambicus* x *Oreochromis spp.*) and
817 mapping of sex-determining loci. *BMC Genomics*, 14, 1-14.

818 Lowe, R.H. (1955). New species of *Tilapia* (Cichlidæ) from Lake Jipe and the Pangani River,
819 East Africa, with notes on the biology of these and Lake Victoria species grown in ponds.
820 *Bulletin of the British Museum of Natural History*, 2, 350–368.

821 Lowe-McConnell, R.H. (1982). Tilapias in fish communities. In: *The biology and culture of*
822 *tilapia*. Eds: Pullin, R.S.V. and Lowe-McConnell, R.H. International Centre for Living
823 Aquatic Resources Management, Manila.

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

827 Mandeville, E.G., Walters, A.W., Nordberg, B.J., Higgins, K.H., Burckhardt, J.C. & Wagner,
828 C.E. (2019). Variable hybridization outcomes in trout are predicted by historical fish
829 stocking and environmental context. *Molecular Ecology*, 28, 3738-3755.

830 Martin, S.H., & Jiggins, C.D. (2017). Interpreting the genomic landscape of introgression.
831 *Current Opinion in Genetics & Development*, 47, 69-74.

832 Martin, S.H. & Van Belleghem, S.M. (2017). Exploring evolutionary relationships across the
833 genome using topology weighting. *Genetics*, 206, 429-438.

834 Martin, S.H., Davey, J.W., Salazar, C., & Jiggins, C.D. (2019). Recombination rate variation
835 shapes barriers to introgression across butterfly genomes. *PLoS Biology*, 17, e2006288.

836 Martinuzzi, S., Januchowski-Hartley, S.R., Pracheil, B., McIntyre, P.B., Plantinga, A.J., Lewis,
837 D.J., & Radeloff, V.C. (2014). Threats and opportunities for freshwater conservation
838 under future land use change scenarios in the United States. *Global Change Biology*,
839 20, 113-124.

840 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo,
841 M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-
842 generation DNA sequencing data. *Genome Research*, 20, 1297–1303.

843 Millennium Ecosystem Assessment (2016). Ecosystems and human well-being: Synthesis.
844 Washington, DC: Island Press.

845 Muhlfeld, C.C., Kovach, R.P., Jones, L.A., Al-Chokhachy, R., Boyer, M.C., Leary, R.F., Lowe,
846 W.H., Luikart, G. & Allendorf, F.W. (2014). Invasive hybridization in a threatened species
847 is accelerated by climate change. *Nature Climate Change*, 4, 620-624.

848 Nelson, C.M. (1995). Male size, spawning pit size and female mate choice in a lekking cichlid
849 fish. *Animal Behaviour* 50, 1587–1599.

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

853 Ogutu-Ohwayo, R. (1990). The decline of the native fishes of lakes Victoria and Kyoga (East
854 Africa) and the impact of introduced species, especially the Nile perch, *Lates niloticus*,
855 and the Nile tilapia, *Oreochromis niloticus*. *Environmental Biology of Fishes*, 27, 81-96.

856 Ostberg, C.O., Hauser, L., Pritchard, V.L., Garza, J.C. and Naish, K.A. (2013). Chromosome
857 rearrangements, recombination suppression, and limited segregation distortion in

858 hybrids between Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and rainbow
859 trout (*O. mykiss*). *BMC Genomics*, 14, 570.

860 Parsons, K.J., Son, Y.H., & Albertson, R.C. (2011). Hybridization promotes evolvability in
861 African cichlids: Connections between transgressive segregation and phenotypic
862 integration. *Evolutionary Biology*, 38, 306-315.

863 Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
864 multilocus genotype data. *Genetics*, 155, 945-959.

865 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., ... Sham,
866 P.C. (2007). PLINK: a toolset for whole genome association and population-based
867 linkage analysis. *American Journal of Human Genetics*, 81, 559–575.

868 Quinlan, A.R. & Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
869 features. *Bioinformatics*, 26, 841-842.

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

872 Rohlf, F.J. (2004). TpsDig Version 1.4. Department of Ecology and Evolution. State University
873 of New York at Stony Brook, New York.

874 Rousset F. (2008). Genepop'007: a complete re-implementation of the genepop software for
875 Windows and Linux. *Molecular Ecology Resources*, 8, 103-106.

876 Saju, J.M., Lee, W-J., & Orban, L. (2010). Characterization of nine novel microsatellites
877 isolated from Mozambique tilapia, *Oreochromis mossambicus*. *Conservation Genetics*
878 *Resources*, 2, 287–385.

879 Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, R., Bloomfield, J., Dirzo, R., ... Wall, D.H.
880 (2000). Global biodiversity scenarios for the year 2100. *Science*, 287, 1770-1774.

881 Schneider, R.F., & Meyer, A. (2017). How plasticity, genetic assimilation and cryptic genetic
882 variation may contribute to adaptive radiations. *Molecular Ecology*, 26, 330-350.

883 Schumer, M., Xu, C., Powell, D.L., Durvasula, A., Skov, L., Holland, C., ... Przeworski, M.
884 (2018). Natural selection interacts with recombination to shape the evolution of hybrid
885 genomes. *Science*, 360, 656-660.

886 Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A.,
887 Peichel, C.L., Saetre, G.P., Bank, C., Brännström, Å. & Brelsford, A. (2014). Genomics
888 and the origin of species. *Nature Reviews Genetics*, 15, 176-192.

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

893 Shechonge, A., Ngatunga, B.P., Bradbeer, S.J., Day, J.J., Freer, J.J., Ford, A.G.P., ... Genner,
894 M.J. (2019). Widespread colonization of Tanzanian catchments by introduced

895 *Oreochromis tilapia* fishes: the legacy from decades of deliberate introduction.
896 *Hydrobiologia*, 832, 235-253.

897 Skelton, P.H., (2001). *A complete guide to the freshwater fishes of southern Africa*. Capetown:
898 Struik.

899 Snoeks, J., Hanssens, M., Ngatunga, B., Chisambo, J., & Nyasulu, T. (2004). *Identification,*
900 *distribution and taxonomy of Malawian cichlids: a look to the future*. In: M.J. Snoeks, ed.
901 2004. The cichlid diversity of Lake Malawi/Nyasa.Niassa: identification, distribution and
902 taxonomy. 354-356.

903 Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007) pcaMethods - a
904 Bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, 223,
905 1164-1167.

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

910 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
911 large phylogenies. *Bioinformatics*, 30, 1312-1313.

912 Suarez-Gonzalez, A., Lexer, C., & Cronk, Q.C. (2018). Adaptive introgression: a plant
913 perspective. *Biology Letters*, 14, 20170688.

914 Trewavas, E. (1983). *Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia*.
915 London, UK: British Museum (Natural History).

916 Valencia-Montoya, W.A., Elfekih, S., North, H.L., Meier, J.I., Warren, I.A., Tay, W.T., Gordon,
917 K.H., Specht, A., Paula-Moraes, S.V., Rane, R. & Walsh, T.K. (2020). Adaptive
918 introgression across semipermeable species boundaries between local *Helicoverpa zea*
919 and invasive *Helicoverpa armigera* moths. *Molecular Biology and Evolution*, msaa108,
920 <https://doi.org/10.1093/molbev/msaa108>.

921 Wang, J., Street, N.R., Park, E.J., Liu, J. & Ingvarsson, P.K. (2020). Evidence for widespread
922 selection in shaping the genomic landscape during speciation of *Populus*. *Molecular*
923 *Ecology*, 29, 1120-1136.

924

925 **Table 1.** Sample sizes for southern comparison analysis of *O. korogwe*, *O. niloticus*, and
 926 individuals of hybrid origin (comparison 1) and comparisons of southern and northern *O.*
 927 *korogwe* populations and reference *O. urolepis* and *O. placidus* (comparison 2).
 928

Site	Species	Microsatellite	Linear (conventional) measures	Geometric morphometric
<u>Comparison 1: southern comparison analysis of <i>O. korogwe</i>, <i>O. niloticus</i>, and hybrids</u>				
Lake Mitupa	<i>O. korogwe</i> (M-OK)	2	-	-
	<i>O. niloticus</i> (M-ON)	3	3	3
	Hybrid (M-OK x M-ON)	2	1	1
Lake Rutamba	<i>O. korogwe</i> (R-OK)	17	9	9
	<i>O. niloticus</i> (R-ON)	13	6	6
	Hybrid (R-OK x R-ON)	2	2	2
Lake Nambawala	<i>O. korogwe</i> (N-OK)	10	9	9
	<i>O. niloticus</i> (N-ON)	6	4	4
	Hybrid (OK x ON)	6	5	5
<u>Comparison 2: southern and northern <i>O. korogwe</i> populations, and reference species</u>				
Mlingano dam	<i>O. korogwe</i> (MI-OK)	40	34	40
Zigi River	<i>O. korogwe</i> (Z-OK)	16	23	29
Lake Chidya	<i>O. placidus</i> (C-OP)	10	-	-
Rufiji River	<i>O. urolepis</i> (RR-OU)	26	-	-
Lake Nambawala	<i>O. korogwe</i> (N-OK)	10	9	10
Lake Rutamba	<i>O. korogwe</i> (R-OK)	17	14	9

929

930 **Table 2.** Matrix of F_{ST} pairwise comparisons (below left) and corresponding P values from
 931 Exact tests (above right).

	<i>O. placidus</i> Chidya	<i>O. korogwe</i> Zigi	<i>O. korogwe</i> Mlingano	<i>O. urolepis</i> Rufiji	<i>O. korogwe</i> Rutamba	<i>O. korogwe</i> Nambawala
<i>O. placidus</i> Lake Chidya		<0.001	<0.001	<0.001	<0.001	<0.001
<i>O. korogwe</i> Zigi river	0.547		<0.001	<0.001	<0.001	<0.001
<i>O. korogwe</i> Mlingano dam	0.761	0.341		<0.001	<0.001	<0.001
<i>O. urolepis</i> Rufiji river	0.229	0.455	0.612		<0.001	<0.001
<i>O. korogwe</i> Lake Rutamba	0.659	0.358	0.378	0.511		0.473
<i>O. korogwe</i> Lake Nambawala	0.618	0.415	0.470	0.461	0.011	

932

933 **Figure Legends**

934

935 **Figure 1. Sampling sites and example specimens of focal populations.** a) northern *O.*
936 *korogwe* male, b) northern *O. korogwe* female, c) southern *O. korogwe* male, d) southern *O.*
937 *korogwe* female. Pink and purple filled circles indicate northern *O. korogwe* populations
938 sampled, darker blue filled circles locations of the southern *O. korogwe* populations sampled.
939 Grey and black filled circles indicate the sampling locations of *O. urolepis* (Wami and Rufiji
940 and rivers, respectively). The orange filled circles indicate the sampling location of *O.*
941 *placidus* (Lake Chidya).

942

943 **Figure 2.** Genetic and morphological contrasts between *O. korogwe*, *O. niloticus* and *O.*
944 *korogwe* x *niloticus* hybrids. a) Structure assignment of individuals to populations ($K = 2$) using
945 microsatellite data from *Oreochromis* from the southern lakes. Filled black symbols indicate
946 individuals of putative hybrid origin. b) images of *O. korogwe* (top), *O. korogwe* x *niloticus*
947 (middle) and *O. niloticus* (bottom). c) Discriminant function axes illustrate distinctive
948 morphology of purebred *O. korogwe* (blue)s and *O. niloticus* (red) *O. korogwe* x *niloticus* hybrid
949 individuals which overlap in morphospace with parent taxa.

950

951 **Figure 3.** Genetic and morphological analysis of focal populations of *O. korogwe*, and
952 reference populations of *O. urolepis* (Utete), and *O. placidus* (Lake Chidya). a) Structure
953 analysis of the six populations, using $K = 5$. b) Factorial correspondence analysis (FCA) of all
954 populations from all six sites, c) FCA of the four *O. korogwe* populations, d-e) Discriminant
955 Function analysis (DFA) of the four *O. korogwe* populations using linear and geometric
956 measures respectively, and f) Wireframe analysis from Canonical Variates Analysis (CVA)
957 showing geometric morphometric divergence between northern (light blue lines) and
958 southern (dark blue lines) populations.

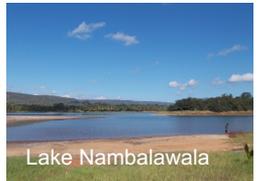
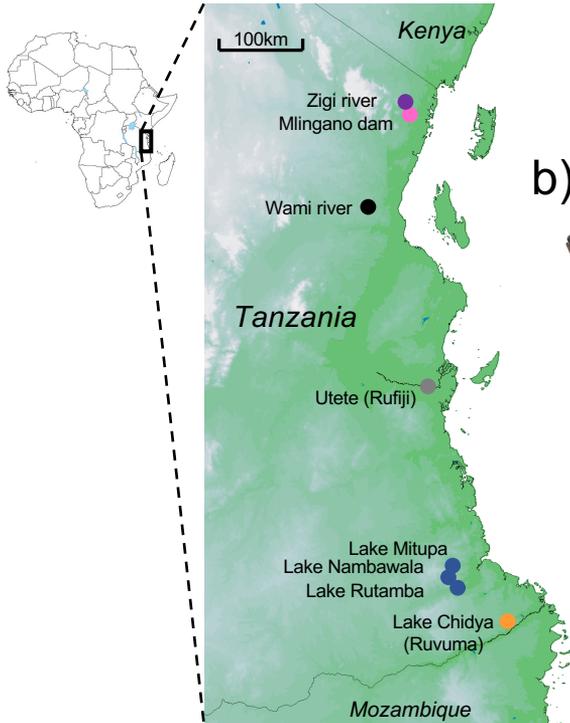
959

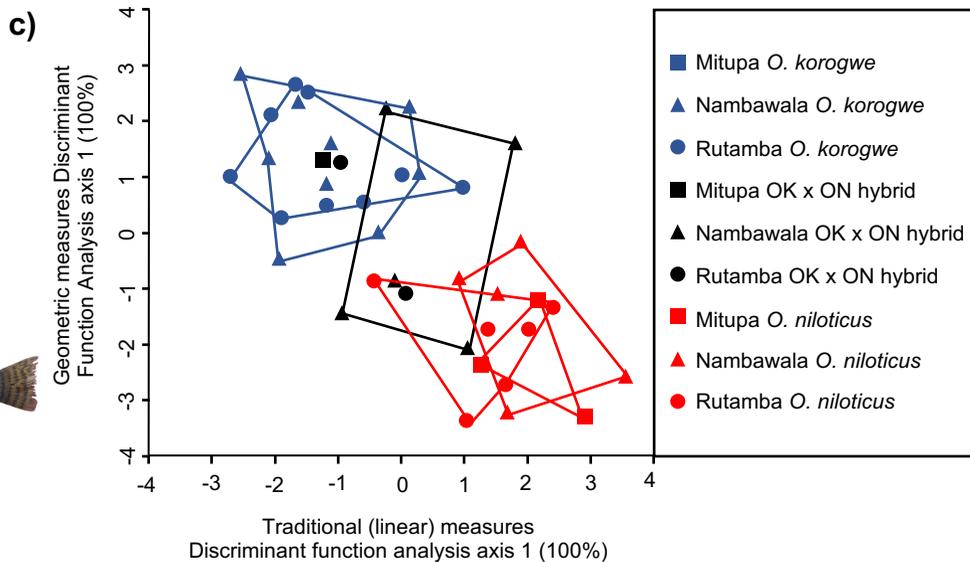
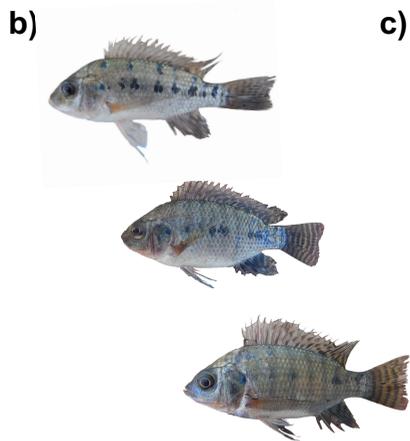
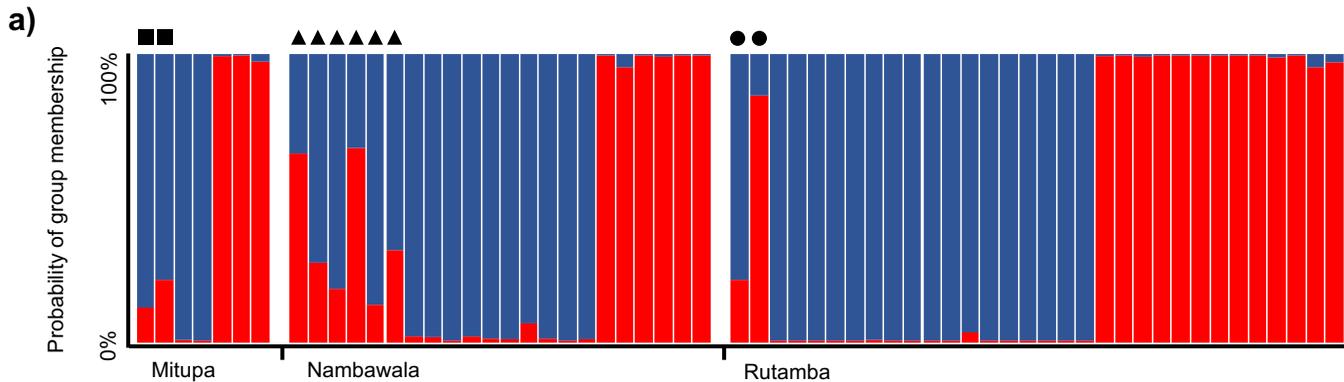
960 **Figure 4.** Analyses of genome-wide data. a) Principal Component Analysis (PCA) of all
961 variants, b) Admixture analysis of all variances, c) phylogeny based on nuclear genome
962 variants, using RAxML GTR+ Γ model. d) phylogeny based on mitochondrial genome variants,
963 using RAxML GTR+ Γ model. Scale bars in changes per bp. Values on nodes indicate
964 bootstrap support values for 1000 bootstraps, those >70% shown.

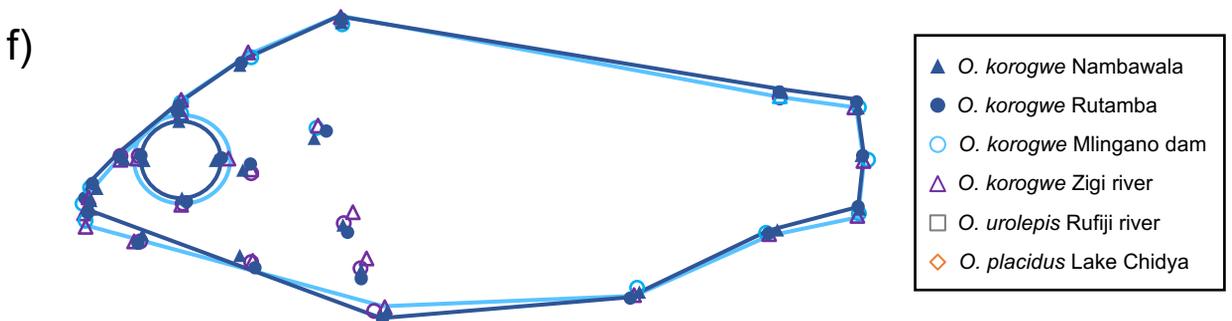
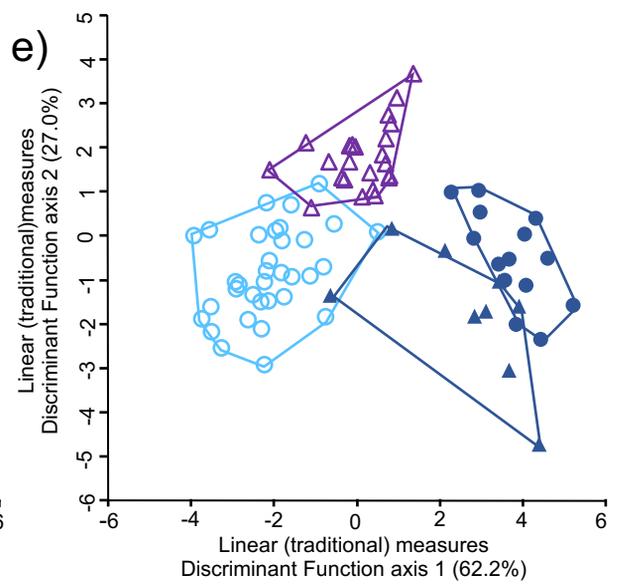
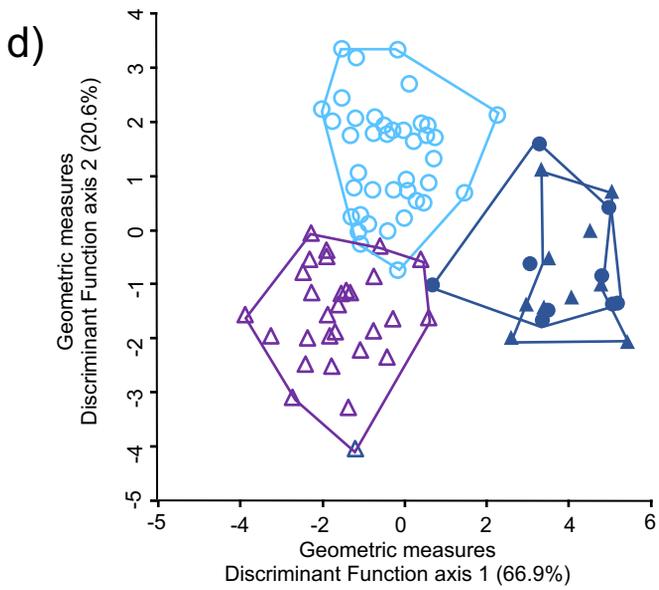
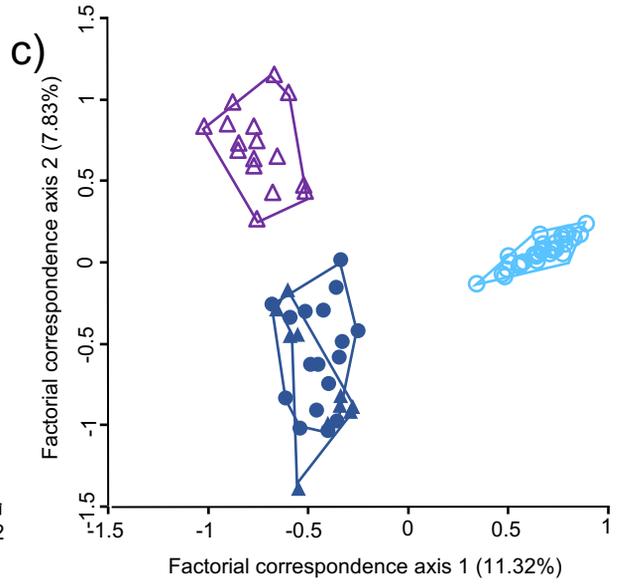
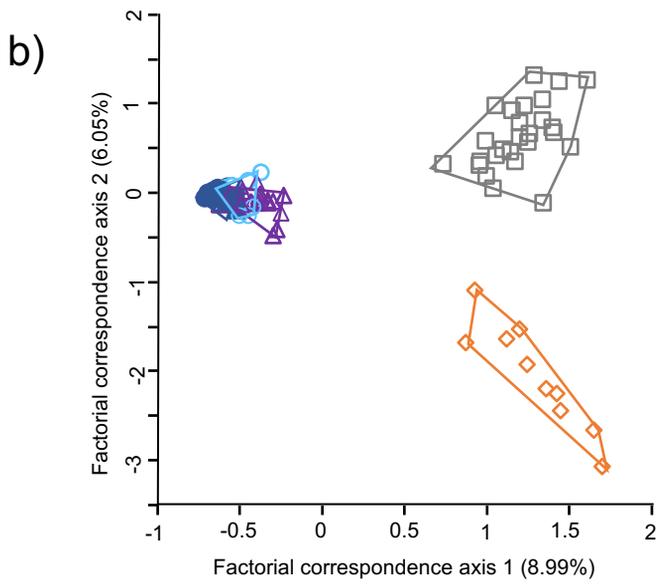
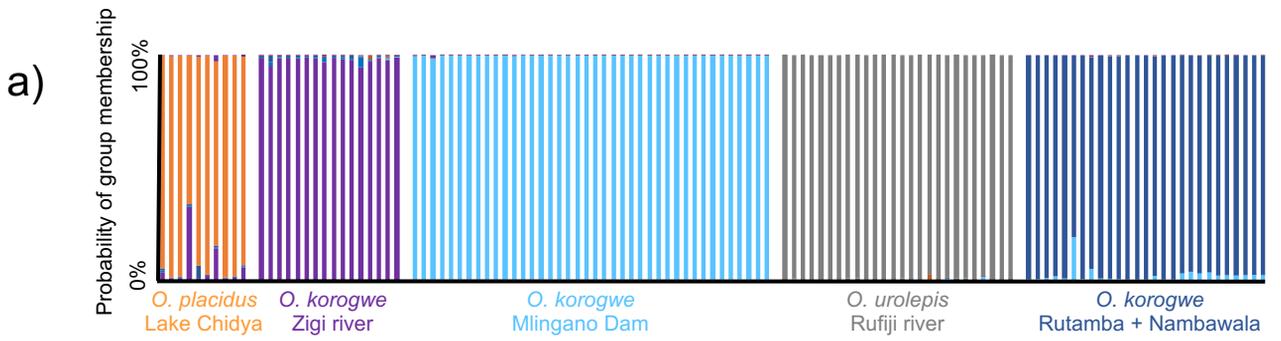
965

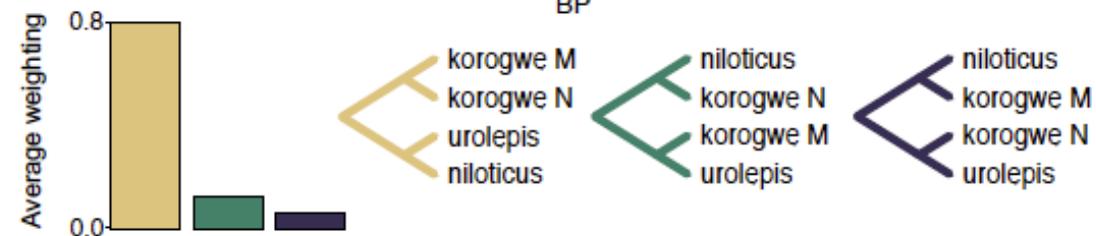
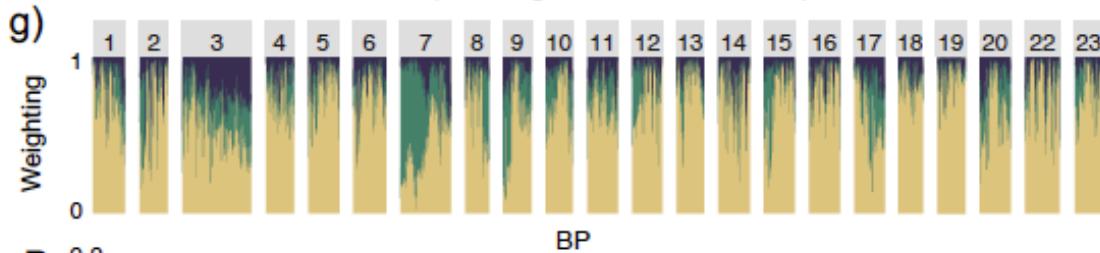
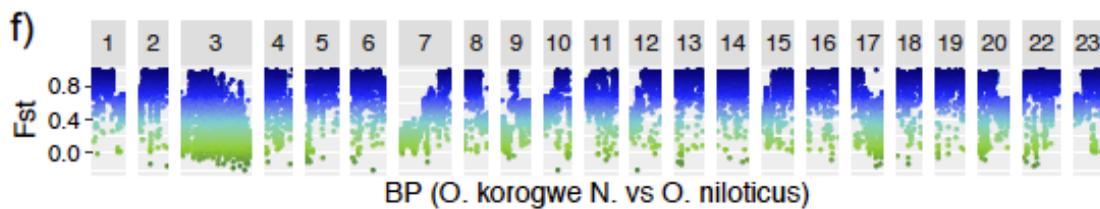
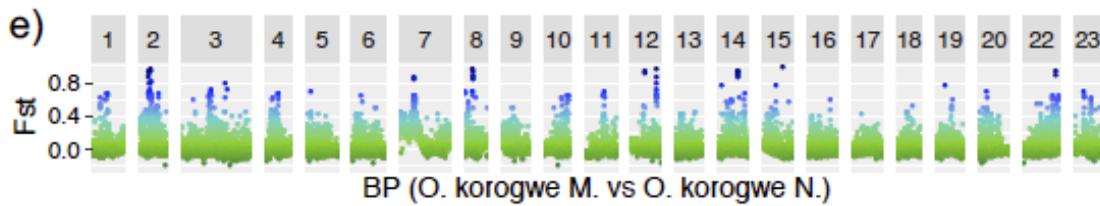
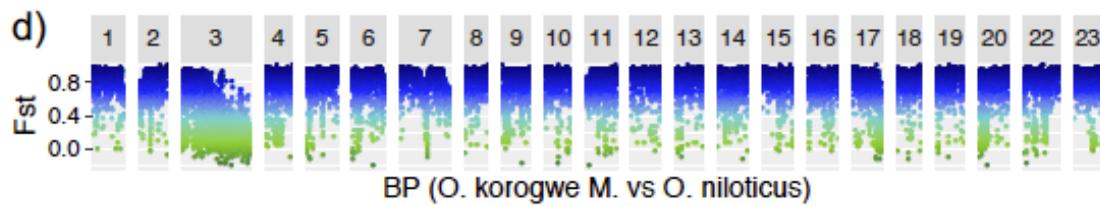
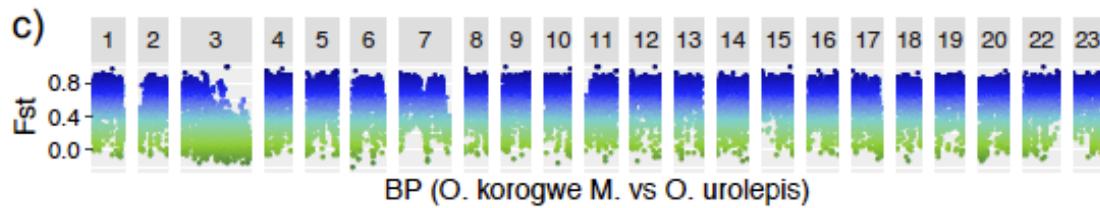
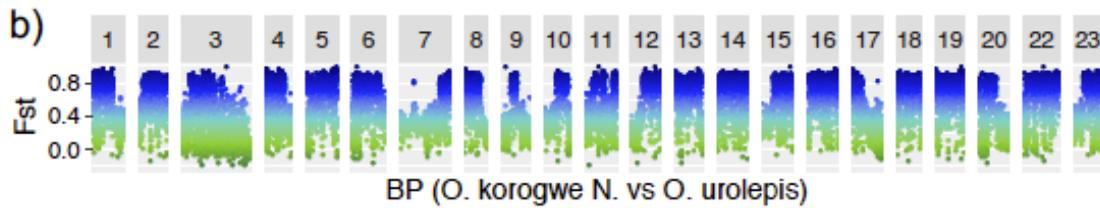
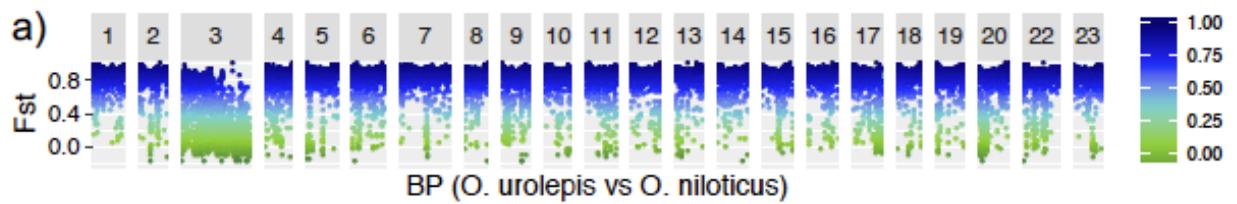
966

967 **Figure 5.** a-f) Pairwise sliding window F_{ST} between populations across genome linkage
968 groups, in 50-kb windows, between combinations of *O. niloticus*, *O. urolepis*, southern *O.*
969 *korogwe* N (Lake Nambawala), northern *O. korogwe* M (Mlingano Dam). g) Phylogenetic
970 representation across genomes of four populations, as estimated by Twisst. Three possible
971 phylogenies for the four taxa are illustrated below, and their colours correspond to relative
972 weightings in plot above. The linkage groups are labelled according to the numbering of the
973 linkage groups in the reference genome.









for:

Newly discovered cichlid fish biodiversity threatened by hybridization with non-native species

Tabitha Blackwell, Antonia G. P. Ford, Adam G. Ciezarek, Stephanie J. Bradbeer,
Carlos A. Gracida Juarez, Alan M. Smith, Benjamin P. Ngatunga, Asilatu Shechonge, Rashid Tamatamah,
Graham Etherington, Wilfried Haerty, Federica Di Palma, George F. Turner and Martin J. Genner

Table of Contents

Supporting Text. Commands for whole genome resequence (WGR) data analysis	2
Table S1: Whole genome resequencing sample details	4
Table S2: Samples used in each of the analyses	5
Table S3: Microsatellite loci primer sequences	11
Table S4. Genetic (microsatellite) diversity of the focal populations	12
Table S5. WGR quality filtering thresholds applied to the SNP and indel datasets.	13
Table S6. Classification results from Discriminant Function Analysis I	14
Table S7. Correlation of traits with Discriminant Function Axis I (Analysis I)	14
Table S8. Classification results from Discriminant Function Analysis II	15
Table S9. Correlation of traits with Discriminant Function Axes (Analysis II)	15
Table S10. Data files number of SNPs by analysis.	16
Table S11. Results of the D3 statistics	17
Figure S1. Landmarks used in the geometric morphometric analysis.	18
Figure S2. WGR population genetic and phylogenetic analysis.	19
Figure S3. Within population nucleotide diversity (π) across linkage groups	20
Figure S4. Absolute genetic divergence (D_{xy}) between population pairs	21

Supporting Text. Commands for whole genome resequence (WGR) data analysis

Genotyping (Per sample HaplotypeCaller):

```
gatk --java-options "-Xmx30g" HaplotypeCaller -R $Reference0 -I $inbam -O
1_HaploCaller/$outname -ERC GVCF --min-pruning 1 --min-dangling-branch-length 1 -
-heterozygosity 0.01 -G StandardAnnotation -G AS_StandardAnnotation --native-
pair-hmm-threads 4
```

Joint Genotyping:

```
gatk --java-options "-Xmx100g" GenomicsDBImport -R $Reference --genomicsdb-
workspace-path GenDB --intervals intervals.list --max-num-intervals-to-import-in-
parallel 30 --overwrite-existing-genomicsdb-workspace --tmp-dir gatktmp $(ls
1_HaploCaller/*.vcf.gz | sed 's/^/-V /g' | tr '\n' ' ')
gatk --java-options "-Xmx100g" GenotypeGVCFs -R $Reference -V gendb://GenDB -O
oreo_genotype.g.vcf.gz
```

Filtering:

```
bcftools filter -G 5 -e 'TYPE != "snp" || ALT="*"' -Oz -o
oreo_genotype_snp_G5.g.vcf.gz oreo_genotype.g.vcf.gz
bcftools index oreo_genotype_snp_G5.g.vcf.gz
tabix oreo_genotype_snp_G5.g.vcf.gz
gatk VariantFiltration \
  -R $Reference0 \
  -V oreo_genotype_snp_G5.g.vcf.gz \
  -O oreo_nucrawfiltersnps.vcf.gz \
  --filter-expression "QD < 2.0 || FS > 10.0 || MQ < 30.0 || MQRankSum < -2.0 ||
ReadPosRankSum < -2.0 || SOR > 3.0 || DP > 180.0" \
  --filter-name "filter" \
  --genotype-filter-expression "DP < 3.0" \
  --genotype-filter-name "lowCov" \
  --set-filtered-genotype-to-no-call true
bcftools view -e 'FILTER!="PASS"' -Oz -o oreo_nucfiltersnps.vcf.gz
oreo_nucrawfiltersnps.vcf.gz.gz
```

Phylogeny reconstruction using RAXML (using full sequence mtDNA data):

```
raxmlHPC-PTHREADS-AVX2 -s Mitodenovo.aln -x $RANDOM -p $RANDOM -f a -# 200 -n
Mito -m GTRGAMMA -T 4
```

Phylogeny reconstruction with RAXML using ascertainment bias correction for SNP data (note that heterozygotes were removed using BCFtools tools first):

```
raxmlHPC-PTHREADS-AVX2 -s raxml.min4.phy -x $RANDOM -p $RANDOM -f a -# 200 -n
test -m ASC_GTRGAMMA --asc-corr=lewis -T 16.support nucSNPs.raxml.support_backup
```

Pruning for linkage disequilibrium in PLINK:

```
plink2 --vcf oreo_nucfiltersnps_biallelic_lg.vcf.gz --indep-pairwise 50 10 0.2 --  
out oreo_ldfilter --allow-extra-chr --set-all-var-ids @:#  
plink2 --vcf oreo_nucfiltersnps_biallelic_lg.vcf.gz --extract  
oreo_ldfilter.prune.in --make-bed --out oreo_pruned --allow-extra-chr --set-all-  
var-ids @:#
```

Run ADMIXTURE for K=1-6 and PCA:

```
plink2 --pca 20 --out pruned_pca --bed oreo_pruned.bed --bim oreo_pruned.bim --  
fam oreo_pruned.fam --allow extra-chr
```

```
cut -f1 oreo_pruned.bim | sort -u | while read line; do counter=$((counter + 1));  
echo $line $counter; sed -i "s/${line}/${counter}/g" oreo_pruned.bim ; done  
for k in {1..6}; do echo $k; admixture --cv oreo_pruned.bed $k > ${k}.o 2>  
${k}.e; done
```

Table S1. Whole genome resequencing sample details and sequencing statistics.

Sample sequencing name	Sample name	ENA sample accession	Species	Collection Location	Sequencing statistics: reads aligned to <i>O. niloticus</i>				
					Paired reads (n)	Mapped (%)	Paired (%)	Mean coverage (X)	>5x coverage (% of genome)
1657_LIB19618_LDI16937_GGCTAC_L005	T2J5	ERS4308617	<i>O.urolepis</i>	Ligongwe Utete (Rufiji)	24,395,019	98.92	88.76	5.83	59.40
1657_LIB19643_LDI16962_CGATGT_L008	T6A2	ERS4308618	<i>O.urolepis</i>	Lower Wami	22,518,296	98.99	88.67	5.41	54.67
1689_LIB19659_LDI16978_ATGTCA_L002	U1A1	ERS4308619	<i>O.niloticus</i>	Uganda	20,528,899	99.21	93.14	5.02	52.13
1689_LIB19660_LDI16979_CCGTCC_L002	U3A3	ERS4308620	<i>O.niloticus</i>	Uganda	23,467,486	99.18	93.33	5.74	60.83
1720_LIB20174_LDI17702_TGACCA_L001	83-2013	ERS4308621	<i>O.placidus</i>	Rovuma	20,921,460	98.75	88.09	4.97	47.81
1720_LIB20175_LDI17703_CAGATC_L001	120-2013	ERS4308622	<i>O.placidus</i>	Rovuma	22,595,615	98.94	90.46	5.45	55.80
1720_LIB20179_LDI17707_CACGAT_L001	T3J2	ERS4308623	<i>O.korogwe-N</i>	Nambawala, Lindi	21,822,242	98.89	88.65	5.21	51.51
1720_LIB20180_LDI17708_CAGGCG_L001	T3J4	ERS4308624	<i>O.korogwe-N</i>	Nambawala, Lindi	22,517,046	98.93	88.26	5.38	53.48
1720_LIB20181_LDI17709_CATGGC_L001	T3J6	ERS4308625	<i>O.korogwe-N</i>	Nambawala, Lindi	22,700,204	98.99	88.88	5.43	53.90
1720_LIB20191_LDI17719_CATGGC_L002	P4A10	ERS4308626	<i>O.korogwe-M</i>	Mlingano Dam	21,088,993	98.27	87.63	5.01	48.55
1720_LIB20192_LDI17720_CGGAAT_L002	P4B1	ERS4308627	<i>O.korogwe-M</i>	Mlingano Dam	20,979,050	97.39	86.48	4.93	47.64
1720_LIB20193_LDI17721_TCGGCA_L002	P4B2	ERS4308628	<i>O.korogwe-M</i>	Mlingano Dam	21,566,100	98.11	87.25	5.10	49.34
MEAN:					<u>22,091,701</u>	<u>98.71</u>	<u>89.13</u>	<u>5.29</u>	<u>52.92</u>

Table S2. Samples used in each of the analyses

- A** Morphological analysis of Southern populations
- B** Microsatellite analysis of Southern populations
- C** Traditional morphological analysis North and South *O. korogwe*
- D** Geometric analysis North and South *O. korogwe*
- E** Microsatellite analysis 6 populations
- F** Whole Genome Resequencing

Species	Sample Code	Collection Site	Collection Date	Latitude (decimals)	Longitude (decimals)	Collector(s)	A	B	C	D	E	F
<i>O. korogwe</i>	korogwe_Mitupa_M177	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga		Y				
<i>O. korogwe</i>	korogwe_Mitupa_M178	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga		Y				
<i>O. korogwe</i>	korogwe_Mitupa_M179	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga		Y				
<i>O. korogwe</i>	korogwe_Mlingano_P4A10	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	Y
<i>O. korogwe</i>	korogwe_Mlingano_P4A6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4A7	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4A8	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4A9	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B1	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	Y
<i>O. korogwe</i>	korogwe_Mlingano_P4B10	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B2	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	Y
<i>O. korogwe</i>	korogwe_Mlingano_P4B3	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B4	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B5	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B7	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B8	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4B9	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4C1	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4C2	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4C3	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4C4	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4C5	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P4C6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P5A10	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P5B1	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P5B10	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	
<i>O. korogwe</i>	korogwe_Mlingano_P5B2	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	

<i>O. niloticus</i>	niloticus_Rutamba_R225	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	
<i>O. niloticus</i>	niloticus_Rutamba_R29	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	
<i>O. niloticus</i>	niloticus_Uganda_U1A1	Uganda	29_10_2015	-	-	N. Kazosi			Y
<i>O. niloticus</i>	niloticus_Uganda_U3A3	Uganda	29_10_2015	-	-	N. Kazosi			Y
<i>O. placidus</i>	placidus_Chidya_142A	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_142B	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_142D	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_143A	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_143C	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_143D	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_144A	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_144C	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_144D	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Chidya_145C	Lake Chidya	18_08_2013	-10.597	40.155	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Rovuma_120-2013	Muguwesi, Rovuma	17_08_2013	-10.847	37.474	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. placidus</i>	placidus_Rovuma_83-2013	Rovuma River	16_08_2013	-11.414	38.492	A. Shechonge; B.P. Ngatunga; M. Genner			Y
<i>O. urolepis</i>	urolepis_Utete_T2J05	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A01	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A02	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A03	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A04	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A05	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A06	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A07	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A08	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A09	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3A10	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B01	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B02	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B03	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B04	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B05	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B06	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B07	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B08	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B09	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3B10	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3C01	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3C02	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y

<i>O. urolepis</i>	urolepis_Utete_T3C03	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3C04	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3C05	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Utete_T3C06	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga			Y
<i>O. urolepis</i>	urolepis_Wami_T6A02	Mbuyuni, Wami	22_07_2015	-6.25149	38.6875	G. Turner			Y
OKxON hybrid	hybrid_Mitupa_M182	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga	Y	Y	
OKxON hybrid	hybrid_Nambawala_N174	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y	
OKxON hybrid	hybrid_Nambawala_N65	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y	
OKxON hybrid	hybrid_Nambawala_N70	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y	
OKxON hybrid	hybrid_Nambawala_T4A5 (=TXA5)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge	Y	Y	
OKxON hybrid	hybrid_Nambawala_T4C5 (=TXC5)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge	Y	Y	
OKxON hybrid	hybrid_Rutamba_R144	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	
OKxON hybrid	hybrid_Rutamba_R245	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	

Table S3: Microsatellite loci primer sequences and sources.

Marker name	Genbank Accession	Primer sequence (forward)	Primer sequence (reverse)	Motif
OMO043	JX204857	GGGGTCATTTCGGTTTATTGGTTAT	AGGGCAGGTCACGGGTTTCG	(TTTG)8
OMO093	JX204891	AAGCCCCACATAGACGACCAGAGA	CAGAAACGGTGCCTGTTCCAGAA	(CAT)8
OMO100	JX204895	CCTTCCCCACCACTACCCTCATAA	CCCGCCCCACACCTGACGA	(ATT)18
OMO114	JX204905	ACGCCTTAATGCTGCCTTCAAGA	TGATGCTCACCCCGTTCCTCA	(GTT)11
OMO129	JX204914	TTGGCAGGCTAAGTACTATTTTCAT	GAGCGAATGGTTGTCTGTCTCT	(CCAT)9
OMO161	JX204924	ACTTTGACAAAAGAAGTGAACAA	AGGGGAGGAGAAAATAAACTGTAT	(TAA)10
OMO219	JX204964	ATCCCCTTCTTTCCATCCCTGTC	AAGGCCTCTGTGAGCTGATTGATT	(TTTTG)10
OMO229	JX204973	GCGACTTTTTCTTTGCACATTTTT	AACTGAACCGCCATCATAATCATC	(GTT)9
OMO248	JX204987	AAAGACACAAAGAGAACTAATCA	GGATGAATATTTAAAATCAGTCAG	(TCA)9
OMO337	JX205052	TAGGAGAGGCATAGGTTGTCAAAT	CAAGAGTCTAGGAGGGAATCAAAA	(GTTT)7
OMO361	JX205069	TGACAGCGAGCCAGAATGGAAGTA	AAAAGTGAAAGGGGCACAGTGAGG	(CTT)17
OMO391	GR699257	AGACATCTGTACGCTCTTACGAA	AGTGCTAGAGGGAAGGGGCTGTA	(GAT)9
OMO392	GR698887	CTGGCTTAACTTCTCTACTGGACA	TCTACTCAAACCTGGCAACAAAAC	(GAATA)7
OMO397	GR693794	ACGCGTGTTTGAGATATTTAGATT	GAACAAACAAGGGGAGTGG	(GATT)7
OM-01	GU391020	TTTAAAGTTACACAGCAGTACAAAG	TTGTAGCATTTC AACACAGTCTC	(GT)20
OM-03	GU391022	CTTTTTAATGAGCAACTTTTAAGTC	TGTGAATTTGACAACTTCCTTTC	(GATA)47
OM-04	GU391022	AGCTCAAAACCTCATACAAAGG	GCAGAGATGTCAGATGTTGTTC	(GACA)6 (GATA)16
OM-09	GU391028	GGCTACAACACCTGGATGG	TTGGGCTTACTGAAGCTGAC	(GT)26

Table S4. Genetic (microsatellite) diversity of the focal populations of *O. korogwe*, *O. urolepis* and *O. placidus*. N - number of individuals, NA - number alleles, Ho - Observed heterozygosity, He - Expected heterozygosity, P - probability of Hardy Weinberg Equilibrium.

Site	Species		OMO219	OMO229	OMO337	OMO391	OMO392	OMO397	OMO09	OMO043	OMO129	OMO03	OMO04	OMO01	OMO114	
Mlingano	<i>O. korogwe</i>	N	33	-	-	-	40	40	-	-	-	-	35	35	40	
		NA	3	-	-	-	3	3	-	-	-	-	8	3	2	
		Ho	0.21	-	-	-	0.45	0.58	-	-	-	-	0.83	0.71	0.33	
		He	0.25	-	-	-	0.38	0.63	-	-	-	-	0.82	0.57	0.28	
		P	0.03	-	-	-	0.6	0.45	-	-	-	-	0.99	0.21	0.56	
Zigi River	<i>O. korogwe</i>	N	12	16	-	15	9	16	16	16	-	5	-	14	14	
		NA	2	3	-	4	3	3	2	3	-	4	-	3	2	
		Ho	0.17	0.19	-	0.8	0.33	0.81	0.13	0.13	-	0	-	0	0.07	0
		He	0.16	0.28	-	0.65	0.54	0.66	0.31	0.23	-	0.8	-	0.47	0.14	
		P	1	0.05	-	0.82	0.17	0.5	0.05	0.19	-	<0.001	-	<0.001	0.04	
Lake Chidya	<i>O. placidus</i>	N	8	10	10	10	-	9	5	-	10	10	-	-	-	
		NA	2	5	3	2	-	4	5	=	5	11	5	-	-	
		Ho	0.13	0.9	0.3	0.1	-	0.56	0.2	-	0.7	0.3	0.7	-	-	
		He	0.13	0.73	0.43	0.1	-	0.66	0.87	-	0.62	0.94	0.62	-	-	
		P	1	0.73	0.09	1	-	0.21	<0.001	-	0.77	<0.001	0.77	-	-	
Rufiji	<i>O. urolepis</i>	N	26	25	26	26	26	26	22	25	26	21	19	22	25	
		NA	7	8	5	4	3	8	18	4	4	4	15	18	18	7
		Ho	0.77	0.84	0.27	0.54	0.35	0.77	0.86	0.52	0.31	0.52	0.74	0.73	0.84	
		He	0.79	0.79	0.67	0.69	0.3	0.81	0.88	0.7	0.28	0.92	0.96	0.9	0.77	
		P	0.42	0.47	<0.001	0.22	1	0.46	0.44	0.03	1	<0.001	<0.01	0.05	0.6	
Rutamba	<i>O. korogwe</i>	N	16	-	17	17	11	17	16	-	17	8	17	13	13	
		NA	2	-	2	2	3	3	2	-	2	5	2	5	2	
		Ho	0.19	-	0	0.12	0.45	0.29	0.25	-	0	0.13	0.06	0.15	0.62	
		He	0.5	-	0.11	0.11	0.65	0.27	0.31	-	0.51	0.81	0.06	0.63	0.52	
		P	0.03	-	0.03	1	0.04	1	0.43	-	<0.001	<0.001	1	<0.001	0.6	
Rutamba	<i>O. niloticus</i>	N	12	13	-	-	12	12	12	-	-	11	13	6	8	
		NA	4	2	2	-	4	2	3	-	-	5	2	3	3	
		Ho	0.08	0.54	0.08	-	0.5	0.33	0.17	-	-	0.82	0.08	0.17	0.38	
		He	0.72	0.51	0.08	-	0.64	0.39	0.65	-	-	0.77	0.08	0.32	0.64	
		P	<0.001	1	1	-	0.08	1	<0.001	-	-	0.36	1	0.09	0.14	
Nambawala	<i>O. korogwe</i>	N	7	10	10	-	4	10	10	-	9	1	10	4	4	
		NA	2	2	2	-	3	2	2	-	2	2	2	3	2	
		Ho	0.29	0.3	0	-	0.5	0.2	0.2	-	0	1	0.2	0.25	0.25	
		He	0.26	0.27	0.19	-	0.61	0.19	0.34	-	0.47	1	0.19	0.61	0.25	
		P	1	1	0.05	-	0.43	1	0.31	-	<0.001	1	1	0.14	1	
Nambawala	<i>O. niloticus</i>	N	6	6	-	-	6	6	6	6	6	6	6	6	6	
		NA	2	2	-	-	3	3	2	2	2	2	2	6	4	
		Ho	0.33	0.5	-	-	1	0.17	0.33	0.17	0.17	0.17	0.17	0.5	0.5	
		He	0.48	0.53	-	-	0.71	0.62	0.3	0.17	0.17	0.17	0.17	0.8	0.79	
		P	1	1	-	-	0.58	0.03	1	1	1	1	1	0.06	0.32	

Table S5. WGR quality filtering thresholds applied to the SNP and indel datasets.

	Nuclear SNP
QD	< 2.0
FS	> 10.0
SOR	> 3.0
MQ	< 30.0
MQRankSum	< -2.0
ReadPosRankSum	< -2.0
InbreedingCoeff	-
DP	> 180.0
Per-Sample DP	< 3.0

Filtering parameters (definitions per the GATK website: <https://gatkforums.broadinstitute.org/>)

QD: QualbyDepth - variant confidence divided by the unfiltered depth of non-reference samples.

FS: FisherStrand - Phred-scaled p-value using Fisher's Exact Test to detect strand bias

SOR: StrandOddsRatio - aims to evaluate whether there is strand bias in the data

MQ: RMSMappingQuality - Root Mean Square of the mapping quality of the reads across all samples

MQRankSum: MappingQualityRankSumTest - The u-based z-approximation from the Mann-Whitney Rank Sum Test for mapping qualities

ReadPosRankSum: ReadPosRankSumTest - the u-based z-approximation from the Mann-Whitney Rank Sum Test for the distance from the end of the read for reads with the alternate allele

DP: Depth (mean coverage across all samples) - aims to eliminate sites with excessive coverage caused by alignment artifacts

Per-sample DP: Depth per individual sample (minimum coverage).

Table S6. Classification results from Discriminant Function Analysis I. Original and predicted group membership results from Discriminant function analysis of *O. korogwe*, *O. niloticus* and identified hybrids in the southern lakes, using traditional methods and geometric morphometric analysis.

Measurements	Original group	Classified group		Total
		<i>O. korogwe</i>	<i>O. niloticus</i>	
Linear (traditional)	<i>O. korogwe</i>	16	2	18
	<i>O. niloticus</i>	1	13	14
	Hybrids (OK x ON)	6	2	8
Geometric morphometric	<i>O. korogwe</i>	17	1	18
	<i>O. niloticus</i>	1	13	14
	Hybrids (OK x ON)	4	4	8

Table S7. Correlation of traits with Discriminant Function axes I. Correlation of traits with Discriminant Function Axis 1 separating *O. niloticus* from *O. korogwe* in the southern lakes, using linear (traditional) measurements.

Trait	Correlation with Axis 1
Head Width	0.533
Head Length	0.392
Anal fin base length	-0.370
Eye length	0.367
Body depth	0.205
Inter orbital width	0.192
Pelvic fin length	0.165
Caudal fin length	-0.120
Caudal peduncle length	-0.115
Pectoral fin base length	-0.105
Snout length	-0.080
Dorsal fin base length	-0.058
Caudal peduncle depth	-0.050
Lower Jaw length	-0.013

Table S8. Classification results from Discriminant Function Analysis II. Classification results from Discriminant Function Analysis of four populations of *O. korogwe* from a) traditional measures of morphology and b) geometric morphometric measures.

Measurements	Original group	Classified group				Total
		K-M	K-Z	K-N	K-R	
Linear (traditional)	O. korogwe Mlingano (K-M)	31	3	0	0	34
	O. korogwe Zigi (K-Z)	0	23	0	0	23
	O. korogwe Nambawala (K-N)	0	0	13	1	14
	O. korogwe Rutamba (K-R)	1	1	0	7	9
Geometric morphometric	O. korogwe Mlingano (K-M)	10	0	0	0	10
	O. korogwe Zigi (K-Z)	0	8	1	0	9
	O. korogwe Nambawala (K-N)	0	0	28	1	29
	O. korogwe Rutamba (K-R)	0	0	2	38	40

Table S9. Correlation of traits with Discriminant Function axes II. Correlation of traits with Discriminant Function axes separating *O. korogwe* populations, using linear (traditional) measurements.

Trait	Correlation with DF Axis 1	Correlation with DF Axis 2	Correlation with DF Axis 3
Anal fin base length	0.148	-0.036	-0.045
Body depth	0.314	-0.061	0.172
Caudal fin length	-0.090	0.291	0.828
Caudal peduncle depth	0.446	-0.049	0.157
Caudal peduncle length	0.089	0.133	-0.105
Dorsal fin base length	0.169	-0.050	-0.276
Eye length	-0.197	0.430	0.246
Head length	0.030	0.174	0.500
Head width	-0.130	-0.019	0.476
Inter-orbital width	0.226	0.458	0.337
Lower jaw length	0.141	-0.086	0.511
Pectoral fin length	0.365	0.031	0.149
Pelvic fin length	-0.090	0.470	0.197
Snout length	0.016	0.338	0.318

Table S10. Data files and number of SNPs by analysis. Data files correspond to code in Supporting Text.

Description	File	Analysis	SNPs (n)
Quality-filtered biallelic nuclear SNPs excluding unplaced scaffolds	oreo_nucfiltersnps_biallelic_lg_2perpop.vcf.gz	FST, Dxy, pi	4,072,183
Quality-filtered biallelic nuclear SNPs excluding <i>O. placidus</i> hybrid, SNPs excluding heterozygote only sites	oreo_nucfiltersnps_biallelic_lg_noldfilter.phy	RAML ASC	5,992,590
Quality-filtered biallelic nuclear SNPs LD pruned	oreo_ld50_10_0.2_pruned.bim	Admixture	160,883

Table S11. Results of the D3 statistics, testing for statistically significant deviations from expected genetic distance between pairs of individuals.

P1	P2	P3	Mean D3	SD D3	p value
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4A10	Oniloticus__LakeAlbert__U1A1	-0.043266683	0.002428077	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B1	Oniloticus__LakeAlbert__U1A1	-0.031307302	0.002423252	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B2	Oniloticus__LakeAlbert__U1A1	-0.043255295	0.002480584	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4A10	Oniloticus__LakeAlbert__U3A3	-0.043062768	0.002406678	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B1	Oniloticus__LakeAlbert__U3A3	-0.031222912	0.002447765	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B2	Oniloticus__LakeAlbert__U3A3	-0.043104307	0.0023835	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4A10	Oniloticus__LakeAlbert__U1A1	-0.01744169	0.001630196	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B1	Oniloticus__LakeAlbert__U1A1	-0.005470223	0.001669989	0.001054365
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B2	Oniloticus__LakeAlbert__U1A1	-0.017430138	0.001641556	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4A10	Oniloticus__LakeAlbert__U3A3	-0.017495855	0.001679343	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B1	Oniloticus__LakeAlbert__U3A3	-0.005641	0.001544814	0.00026064
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B2	Oniloticus__LakeAlbert__U3A3	-0.017538009	0.001620035	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4A10	Oniloticus__LakeAlbert__U1A1	-0.026655445	0.002170778	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B1	Oniloticus__LakeAlbert__U1A1	-0.014685469	0.002199143	2.43E-11
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B2	Oniloticus__LakeAlbert__U1A1	-0.026642954	0.002192584	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4A10	Oniloticus__LakeAlbert__U3A3	-0.026472375	0.002240804	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B1	Oniloticus__LakeAlbert__U3A3	-0.014620779	0.002181374	2.05E-11
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B2	Oniloticus__LakeAlbert__U3A3	-0.026513275	0.002200194	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4A10	Ourolepis__LakeLugongwe__T2J5	0.049471943	0.00242394	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B1	Ourolepis__LakeLugongwe__T2J5	0.037784552	0.002365374	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B2	Ourolepis__LakeLugongwe__T2J5	0.050077397	0.00241608	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4A10	Ourolepis__Mbuyunipool__T6A2	0.049724793	0.00254084	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B1	Ourolepis__Mbuyunipool__T6A2	0.037858938	0.002454694	0
Okorogwe__Nambawala__T3J2	Okorogwe__MlinganoDam__P4B2	Ourolepis__Mbuyunipool__T6A2	0.050036855	0.002508508	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4A10	Ourolepis__LakeLugongwe__T2J5	0.025566503	0.001844267	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B1	Ourolepis__LakeLugongwe__T2J5	0.013864769	0.001747232	2.00E-15
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B2	Ourolepis__LakeLugongwe__T2J5	0.02617948	0.001897977	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4A10	Ourolepis__Mbuyunipool__T6A2	0.025979047	0.001885167	0
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B1	Ourolepis__Mbuyunipool__T6A2	0.014100537	0.001754005	8.88E-16
Okorogwe__Nambawala__T3J4	Okorogwe__MlinganoDam__P4B2	Ourolepis__Mbuyunipool__T6A2	0.026296578	0.00188004	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4A10	Ourolepis__LakeLugongwe__T2J5	0.036884685	0.002462671	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B1	Ourolepis__LakeLugongwe__T2J5	0.025187948	0.00234499	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B2	Ourolepis__LakeLugongwe__T2J5	0.037491545	0.002448547	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4A10	Ourolepis__Mbuyunipool__T6A2	0.03715063	0.002506768	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B1	Ourolepis__Mbuyunipool__T6A2	0.025275681	0.002416399	0
Okorogwe__Nambawala__T3J6	Okorogwe__MlinganoDam__P4B2	Ourolepis__Mbuyunipool__T6A2	0.037463557	0.002480429	0

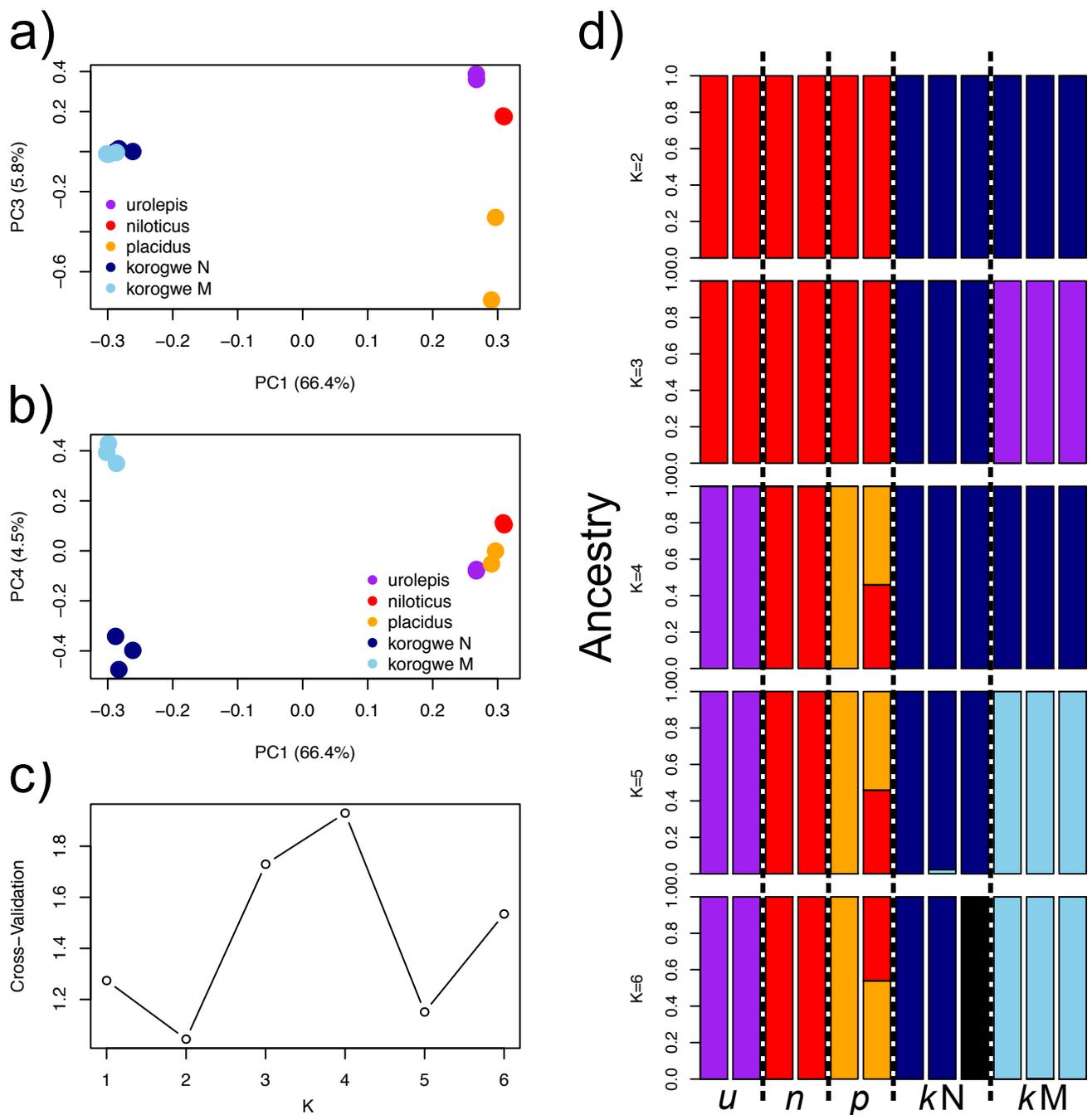


Figure S2. WGR population genetic and phylogenetic analysis. a-b) PCA analysis of LD-pruned nuclear (116,901) SNPs, c) Cross-validation error for admixture analysis of K=1-6, based on 400,680 LD-pruned nuclear SNPs, d) Admixture cluster membership for K=2-5. Species codes used here are u = *O. urolepis*, n = *O. niloticus*, p = *O. placidus rovumae*, kN = *O. korogwe* Nambalwala (southern), kM = *O. korogwe* Mlingano northern). Note colours correspond with genetic clusters, and individual colours are selected to best correspond with populations in K=5.

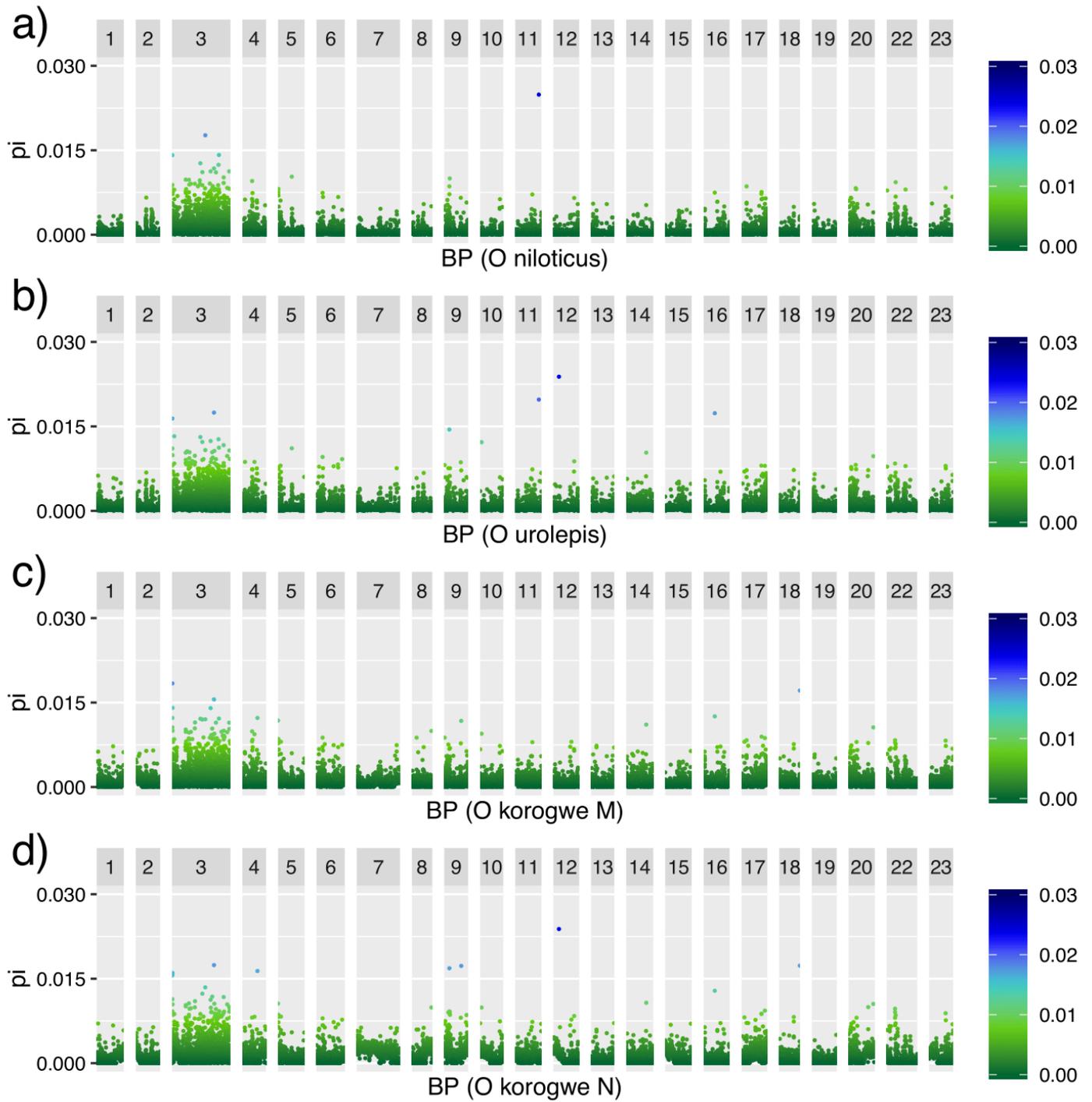


Figure S3. Within population nucleotide diversity (π) across linkage groups, estimated with whole genome data, in non-overlapping 50kb windows.

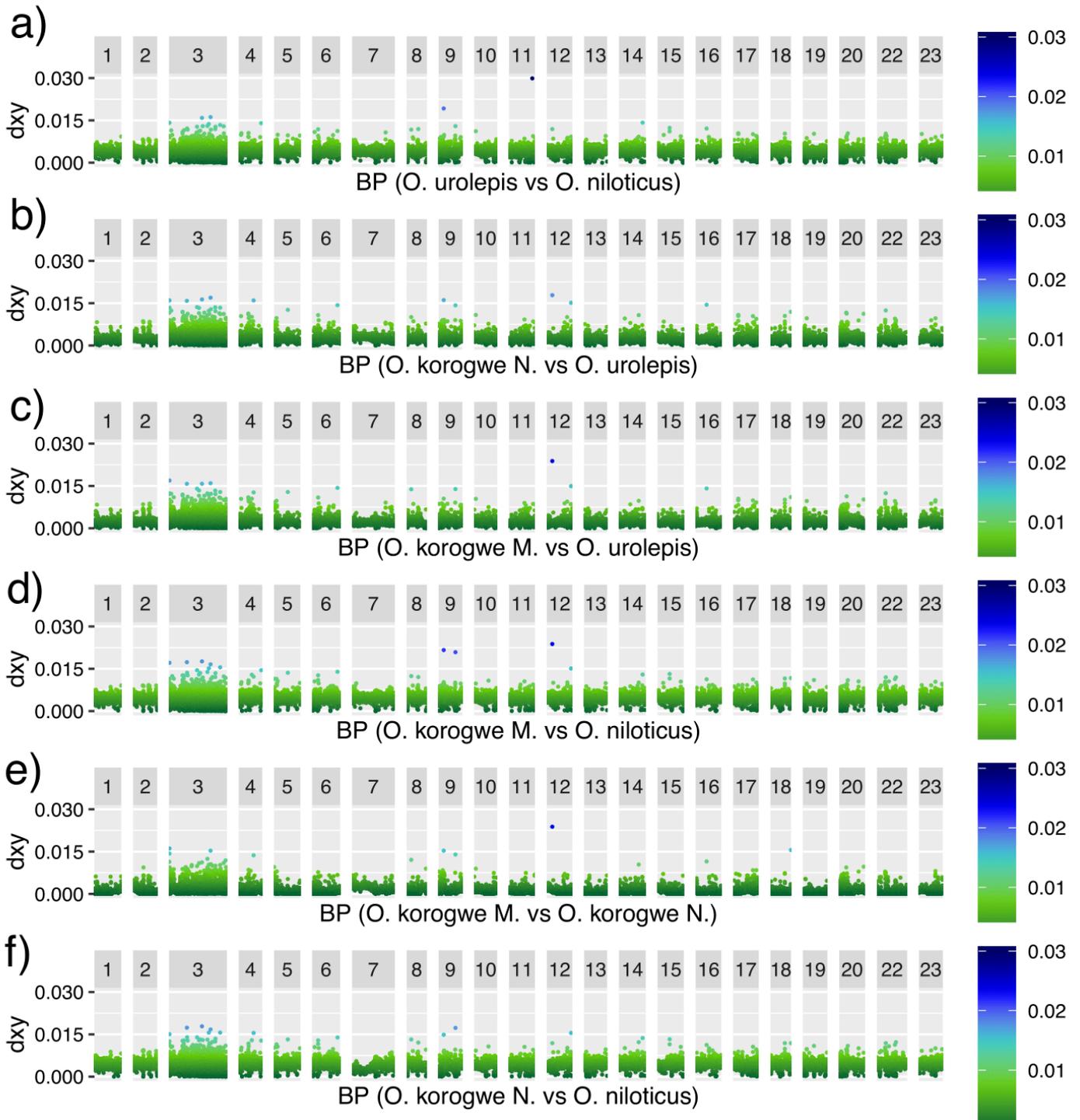


Figure S4. Absolute genetic divergence (D_{xy}) between population pairs across linkage groups, as estimated using whole genome data, in non-overlapping 50kb windows.