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

Published: 01/02/2021

Peer reviewed version

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

Tabitha Blackwell[1*], Antonia G.P. Ford[2*], Adam G. Ciezarek[3*], Stephanie J. Bradbeer[1], Carlos A. Gracida Juarez[1], Alan M. Smith[4], Benjamin P. Ngatunga[5], Asilatu Shechonge[5], Rashid Tamatamah[5], Graham Etherington[3], Wilfried Haerty[3], Federica Di Palma[3,6], George F. Turner[7] and Martin J. Genner[1]

[1] School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, BS81TQ, United Kingdom

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

[3] Earlham Institute, Norwich Research Park Innovation Centre, Colney Ln, Norwich, NR47UZ, United Kingdom

[4]Department of Biological Sciences University of Hull, Hull, HU6 7RX

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

[6] Department of Biological and Medical Sciences, University of East Anglia, Norwich, NR47TJ, UK

[7] School of Biological Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, United Kingdom

*Authors contributed equally

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

Abstract

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

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

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

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

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

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

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

Materials and Methods

Study sites and sample collection.

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

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

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

Population genetics – microsatellite genotyping

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

Population genetics – microsatellite evidence of hybridization in the southern lakes

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

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

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

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

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

Whole genome resequencing - Read mapping and SNP calling

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

Whole genome resequencing – PCA, ADMIXTURE and phylogenetic analysis

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

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

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

Whole genome resequencing - differentiation across the genome

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

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

Divergence times

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

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

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

D3 statistics

The genotypes used for sliding window F_{ST}, Dxy and pi analysis were using to calculate pairwise-distances between each individual, in 50kb non-overlapping windows across the genome, using *distMat.py* (https://github.com/simonhmartin/genomics_general). This pairwise-distance was corrected using the number of callable sites per window (see that section of the methods). *D3* statistics can be used to test for introgression between either P3 and P2 or P3 and P1 in a three-taxon phylogeny (P3,(P2,P1));, without the presence of an outgroup, using genetic distances. Introgression would be expected to result in reduced genetic distance between the two taxon in question. Using the equation $D3 = (d\text{P1P3} - d\text{P2P3}) / (d\text{P1P3} + d\text{P2P3})$; where *d*P1P3 is the distance between P1 and P3 and *d*P2P3 is the distance between P2 and P3, a result where *D3* is significantly less than 0 indicates introgression between P1 and P3, whereas a result where *D3* is significantly greater than 0 indicates introgression between P2 and P3 (Hahn and Hibbins 2019). Significance was assessed by 1000 block bootstrap replicates, with the standard deviation used to calculate p values using the overall mean *D3*. The test was carried out between all trios of species where P1 was an individual from *O. korogwe* Nambawala, P2 was an individual from *O. korogwe* Mlingano and P3 was an individual from either *Oreochromis niloticus* or *Oreochromis urolepis*.

Geometric morphometrics – analyses of individuals from the southern lakes

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

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

Morphological comparisons between northern and southern O. korogwe

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

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

Results

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

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

Morphological comparisons of purebred and hybrid Oreochromis in southern lakes

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

Population genetics – microsatellite genetic structure among Oreochromis populations

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

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

Morphological comparisons of northern and southern O. korogwe

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

Whole genome resequencing: phylogenomic analyses

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

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

Whole genome resequencing: differentiation across the genome and timescale of divergence

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

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

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

Discussion

Population structure of southern and northern O. korogwe.

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

Morphological variation among O. korogwe populations

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

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

Heterogeneity of admixture across the genome

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

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

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

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

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

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

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

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

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

Conservation implications

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

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

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

Acknowledgements

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

Author Contributions

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

Data Accessibility Statement

- Microsatellite genotype data - https://doi.org/10.5061/dryad.ht76hdrcv
- Morphological data - https://doi.org/10.5061/dryad.ht76hdrcv
- DNA resequencing data (raw reads) - European Nucleotide Archive; Project number: PRJEB36772; Accessions: ERS4308617- ERS4308628
- DNA resequencing analysis files - https://doi.org/10.5061/dryad.ht76hdrcv

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., & Elosegi, 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., Esterhuyse, 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., Bbole, 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). Adegenet 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., & Kishe-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.

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

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

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

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

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

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

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

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

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

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

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

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

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

Ostberg, C.O., Hauser, L., Pritchard, V.L., Garza, J.C. and Naish, K.A. (2013). Chromosome 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

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

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

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

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

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

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

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

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

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

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

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
Comparison 1: southern comparison analysis of *O. korogwe*, *O. niloticus*, and hybrids				
Lake Mitupa	*O. korogwe* (M-OK)	2	-	-
	O. niloticus (M-ON)	3	3	3
	Hybrid (M-OK **x** M-ON)	2	1	1
Lake Rutamba	*O. korogwe* (R-OK)	17	9	9
	O. niloticus (R–ON)	13	6	6
	Hybrid (R-OK x R-ON)	2	2	2
Lake Nambawala	*O. korogwe* (N-OK)	10	9	9
	O. niloticus (N-ON)	6	4	4
	Hybrid (OK x ON)	6	5	5
Comparison 2: southern and northern *O. korogwe* populations, and reference species				
Mlingano dam	*O. korogwe* (Ml-OK)	40	34	40
Zigi River	*O. korogwe* (Z-OK)	16	23	29
Lake Chidya	*O. placidus* (C-OP)	10	-	-
Rufiji River	*O. urolepis* (RR-OU)	26	-	-
Lake Nambawala	*O. korogwe* (N-OK)	10	9	10
Lake Rutamba	*O. korogwe* (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).

	O. placidus Chidya	*O. korogwe* Zigi	*O. korogwe* Mlingano	*O. urolepis* Rufiji	*O. korogwe* Rutamba	*O. korogwe* Nambawala
O. placidus Lake Chidya		<0.001	<0.001	<0.001	<0.001	<0.001
O. korogwe Zigi river	0.547		<0.001	<0.001	<0.001	<0.001
O. korogwe Mlingano dam	0.761	0.341		<0.001	<0.001	<0.001
O. urolepis Rufiji river	0.229	0.455	0.612		<0.001	<0.001
O. korogwe Lake Rutamba	0.659	0.358	0.378	0.511		0.473
O. korogwe 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.

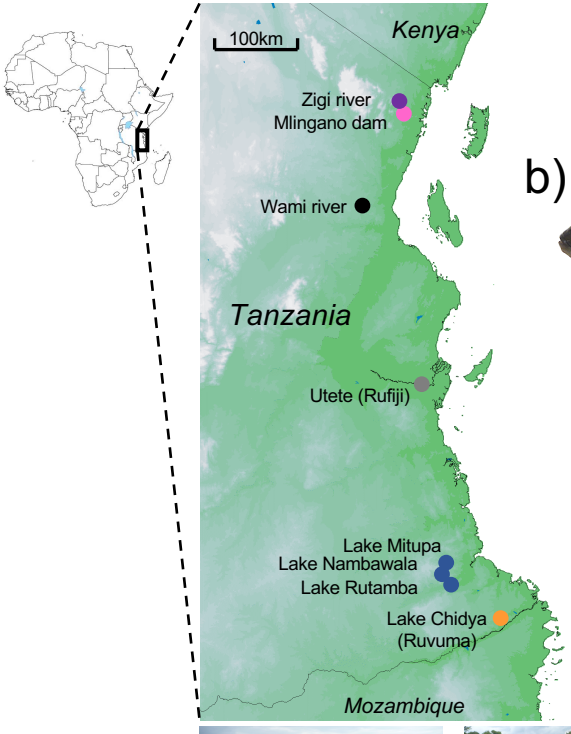
959

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

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


100km
Kenya
Zigi river
Mlingano dam
Wami river
Tanzania
Utete (Rufiji)
Lake Mitupa
Lake Nambawala
Lake Rutamba
Lake Chidya
(Ruvuma)
Mozambique

a)

b)

c)

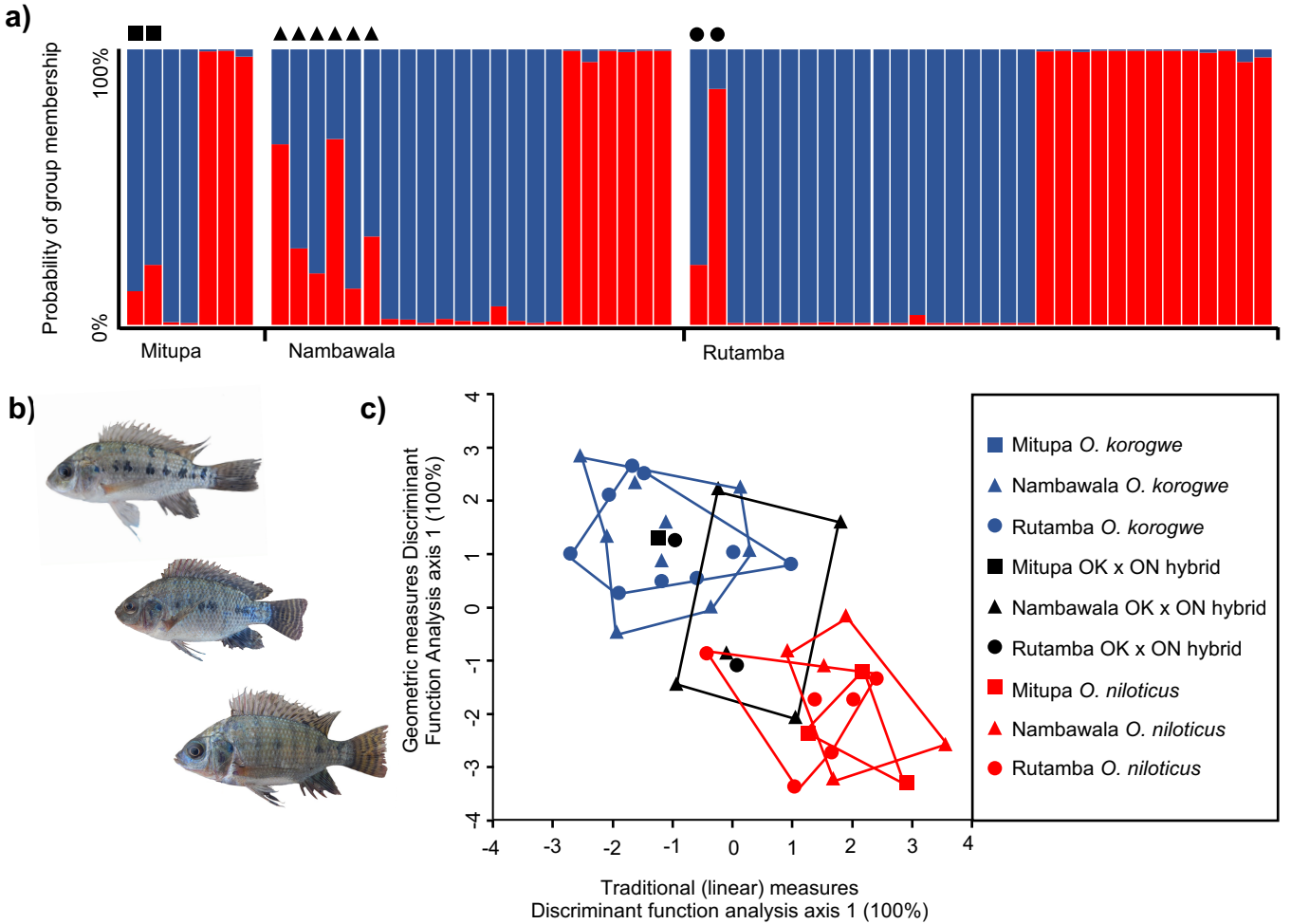
d)

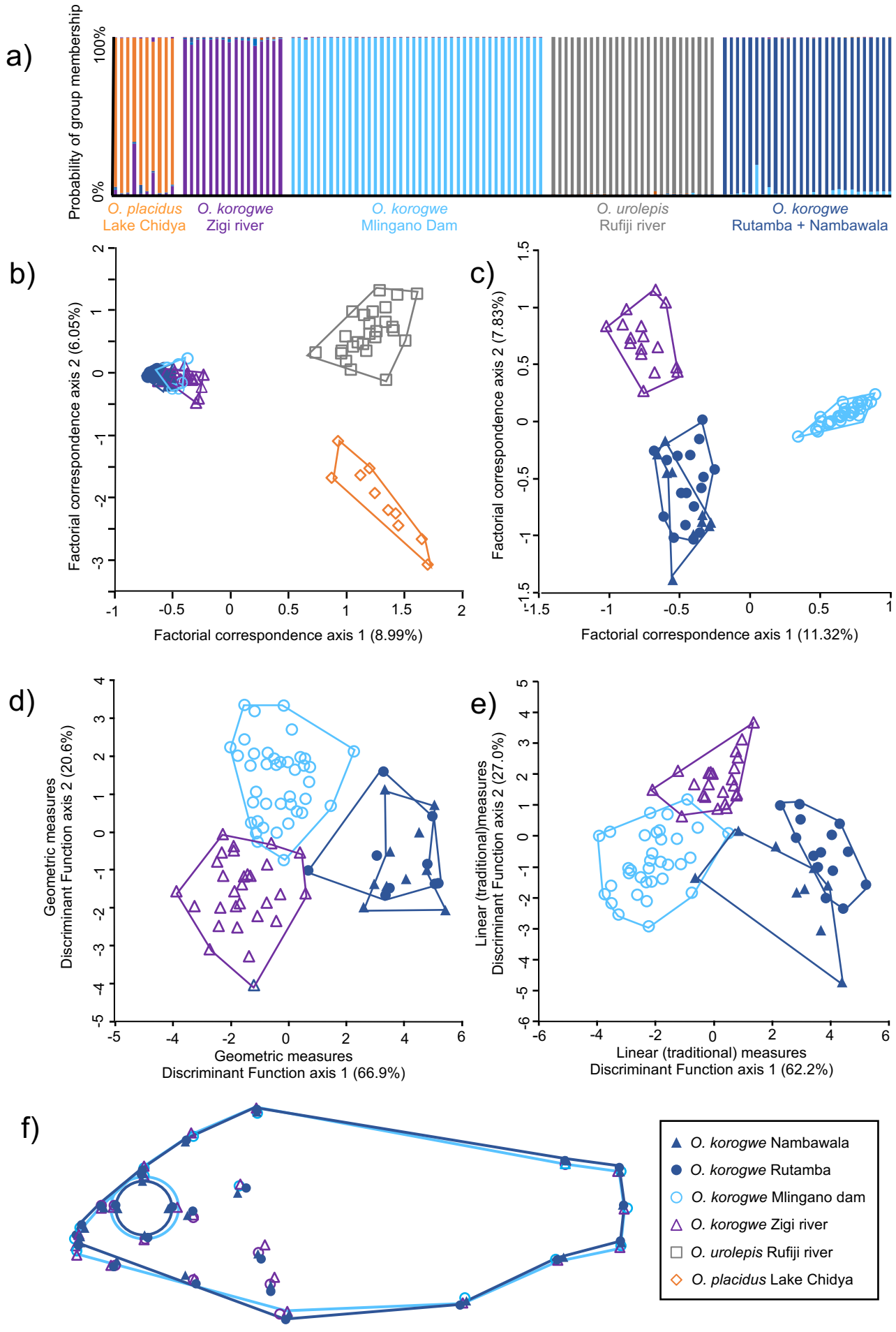

Zigi river

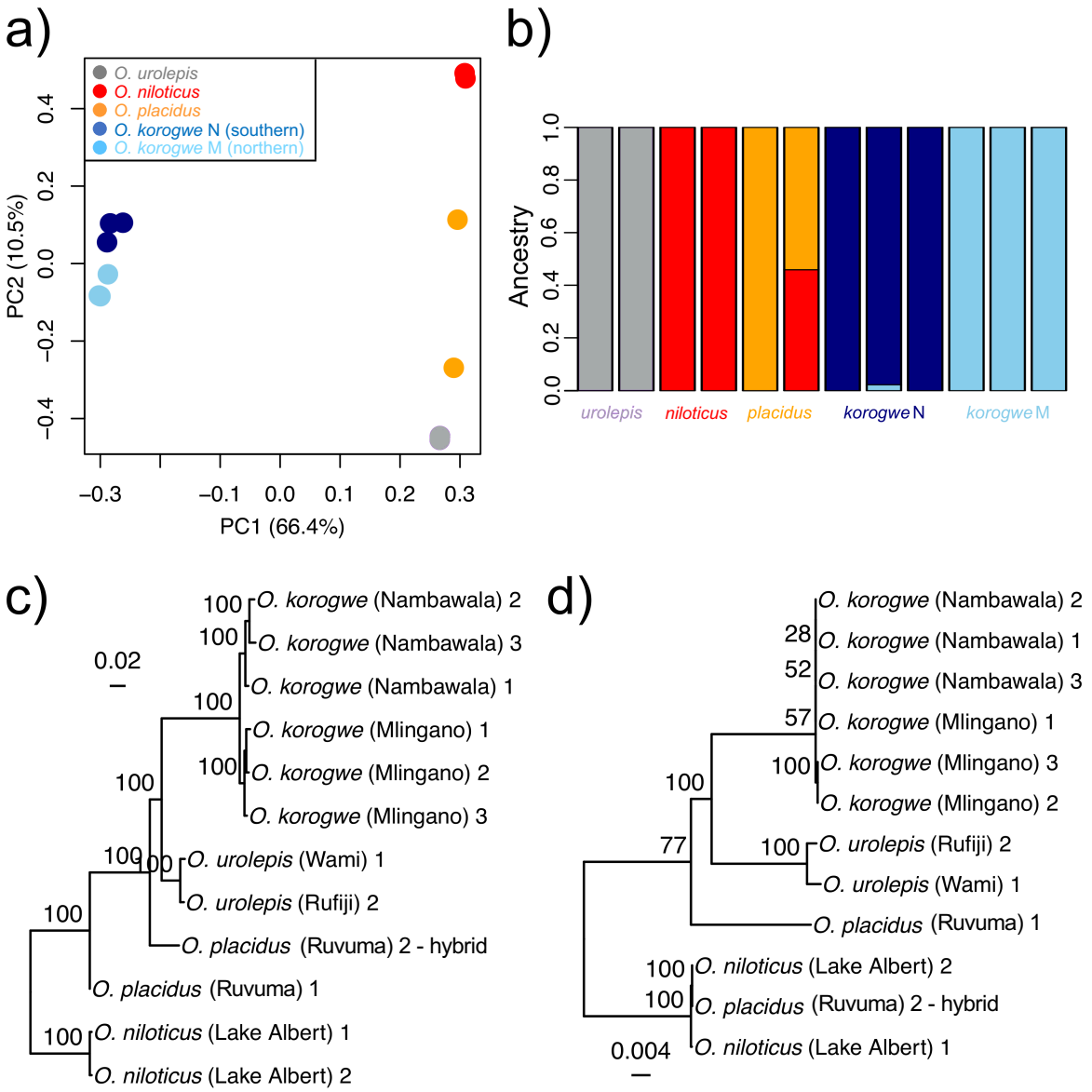

Mlingano dam

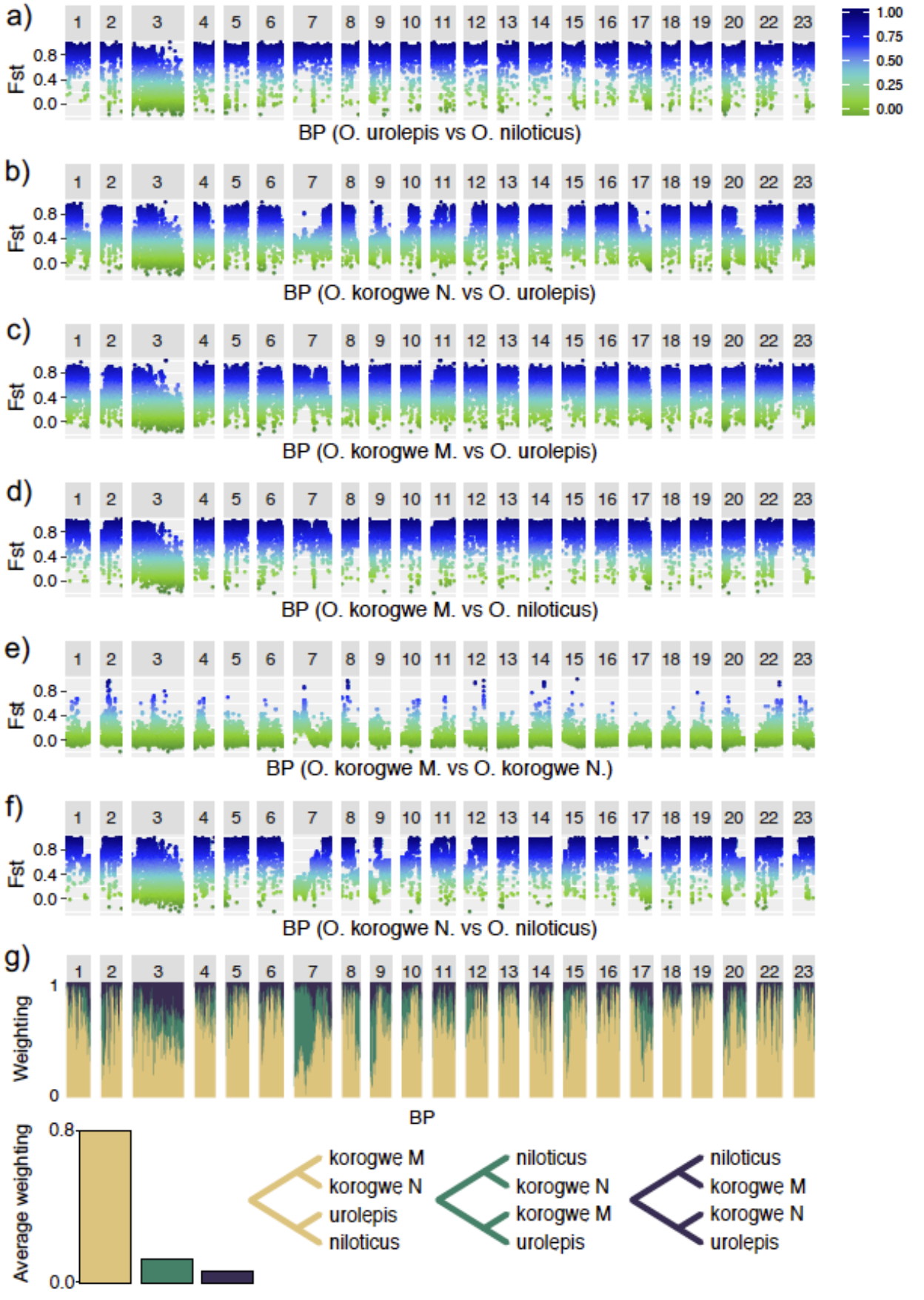

Lake Rutamba


Lake Nambalawala


a)
Probability of group membership
100%
0%
Mitupa
Nambawala
Rutamba
b)
c)
Geometric measures Discriminant Function Analysis axis 1 (100%)
Traditional (linear) measures Discriminant function analysis axis 1 (100%)
-4
-3
-2
-1
0
1
2
3
4
Mitupa O. korogwe
Nambawala O. korogwe
Rutamba O. korogwe
Mitupa OK x ON hybrid
Nambawala OK x ON hybrid
Rutamba OK x ON hybrid
Mitupa O. niloticus
Nambawala O. niloticus
Rutamba O. niloticus


a)
Probability of group membership
100%
0%
O. placidus
Lake Chidya
O. korogwe
Zigi river
O. korogwe
Mlingano Dam
O. urolepis
Rufiji river
O. korogwe
Rutamba + Nambawala
b)
Factorial correspondence axis 2 (6.05%)
Factorial correspondence axis 1 (8.99%)
c)
Factorial correspondence axis 2 (7.83%)
Factorial correspondence axis 1 (11.32%)
d)
Geometric measures
Discriminant Function axis 2 (20.6%)
Geometric measures
Discriminant Function axis 1 (66.9%)
e)
Linear (traditional)measures
Discriminant Function axis 2 (27.0%)
Linear (traditional) measures
Discriminant Function axis 1 (62.2%)
f)
O. korogwe Nambawala
O. korogwe Rutamba
O. korogwe Mlingano dam
O. korogwe Zigi river
O. urolepis Rufiji river
O. placidus Lake Chidya


a)
O. urolepis
O. niloticus
O. placidus
O. korogwe N (southern)
O. korogwe M (northern)
PC2 (10.5%)
PC1 (66.4%)
b)
Ancestry
urolepis
niloticus
placidus
korogwe N
korogwe M
c)
O. korogwe (Nambawala) 2
O. korogwe (Nambawala) 3
O. korogwe (Nambawala) 1
O. korogwe (Mlingano) 1
O. korogwe (Mlingano) 2
O. korogwe (Mlingano) 3
O. urolepis (Wami) 1
O. urolepis (Rufiji) 2
O. placidus (Ruvuma) 2 - hybrid
O. placidus (Ruvuma) 1
O. niloticus (Lake Albert) 1
O. niloticus (Lake Albert) 2
0.02
d)
O. korogwe (Nambawala) 2
O. korogwe (Nambawala) 1
O. korogwe (Nambawala) 3
O. korogwe (Mlingano) 1
O. korogwe (Mlingano) 3
O. korogwe (Mlingano) 2
O. urolepis (Rufiji) 2
O. urolepis (Wami) 1
O. placidus (Ruvuma) 1
O. niloticus (Lake Albert) 2
O. placidus (Ruvuma) 2 - hybrid
O. niloticus (Lake Albert) 1
0.004


a)
b)
c)
d)
e)
f)
g)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 22 23
Fst
0.8
0.4
0.0
1.00
0.75
0.50
0.25
0.00
BP (O. urolepis vs O. niloticus)
BP (O. korogwe N. vs O. urolepis)
BP (O. korogwe M. vs O. urolepis)
BP (O. korogwe M. vs O. niloticus)
BP (O. korogwe M. vs O. korogwe N.)
BP (O. korogwe N. vs O. niloticus)
Weighting
1
0
BP
Average weighting
0.8
0.0
korogwe M
korogwe N
urolepis
niloticus
niloticus
korogwe N
korogwe M
urolepis
niloticus
korogwe M
korogwe N
urolepis

MOLECULAR ECOLOGY

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

Genotyping (Per sample HaplotypeCaller):

```
gatk --java-options "-Xmx30g" HaplotypeCaller -R $ReferenceO -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 $ReferenceO \
  -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.

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

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
O. korogwe	korogwe_Mitupa_M177	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga		Y				
O. korogwe	korogwe_Mitupa_M178	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga		Y				
O. korogwe	korogwe_Mitupa_M179	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga		Y				
O. korogwe	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
O. korogwe	korogwe_Mlingano_P4A6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4A7	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4A8	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4A9	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	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
O. korogwe	korogwe_Mlingano_P4B10	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4B2	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	Y
O. korogwe	korogwe_Mlingano_P4B3	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4B4	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	
O. korogwe	korogwe_Mlingano_P4B5	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4B6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4B7	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4B8	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4B9	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4C1	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4C2	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4C3	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4C4	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4C5	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P4C6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	
O. korogwe	korogwe_Mlingano_P5A10	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P5B1	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y	
O. korogwe	korogwe_Mlingano_P5B10	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	
O. korogwe	korogwe_Mlingano_P5B2	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y	

O. korogwe	korogwe_Mlingano_P5B3	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5B4	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5B5	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5B6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5B7	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5B8	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5B9	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5C1	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5C2	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	Y		
O. korogwe	korogwe_Mlingano_P5C3	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5C4	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5C5	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5C6	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5C7	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Mlingano_P5C8	Mlingano Dam	18_08_2015	-5.122	38.857	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	Y		
O. korogwe	korogwe_Nambawala_N63	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y		
O. korogwe	korogwe_Nambawala_N66	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y		
O. korogwe	korogwe_Nambawala_N71	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y		
O. korogwe	korogwe_Nambawala_N75	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y		
O. korogwe	korogwe_Nambawala_T3J2	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge						Y	
O. korogwe	korogwe_Nambawala_T3J4	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge						Y	
O. korogwe	korogwe_Nambawala_T3J6	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge						Y	
O. korogwe	korogwe_Nambawala_T4A6 (=TXA6)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge		Y		Y			
O. korogwe	korogwe_Nambawala_T4A7 (=TXA7)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge	Y	Y	Y	Y			
O. korogwe	korogwe_Nambawala_T4A9 (=TXA9)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge	Y	Y	Y	Y			
O. korogwe	korogwe_Nambawala_T4C1 (=TXC1)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge	Y	Y	Y	Y			
O. korogwe	korogwe_Nambawala_T4C2 (=TXC2)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge	Y		Y	Y			
O. korogwe	korogwe_Nambawala_T4C4 (=TXC4)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge	Y	Y	Y	Y			
O. korogwe	korogwe_Rutamba_33A	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y					
O. korogwe	korogwe_Rutamba_33B	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y					
O. korogwe	korogwe_Rutamba_33C	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y					
O. korogwe	korogwe_Rutamba_34B	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y					
O. korogwe	korogwe_Rutamba_34F	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y					
O. korogwe	korogwe_Rutamba_42C	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y			Y		
O. korogwe	korogwe_Rutamba_43B	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y			Y		
O. korogwe	korogwe_Rutamba_43C	Rutamba, Lindi	14_08_2013	-10.032	39.46	A. Shechonge; B.P. Ngatunga; M. Genner		Y			Y		
O. korogwe	korogwe_Rutamba_R17	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y		
O. korogwe	korogwe_Rutamba_R207	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y		
O. korogwe	korogwe_Rutamba_R210	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y		

O. korogwe	korogwe_Rutamba_R215	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y
O. korogwe	korogwe_Rutamba_R22	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y			
O. korogwe	korogwe_Rutamba_R222	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y
O. korogwe	korogwe_Rutamba_R23	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y
O. korogwe	korogwe_Rutamba_R240	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y
O. korogwe	korogwe_Rutamba_R240	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y
O. korogwe	korogwe_Rutamba_R244	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y	Y		Y
O. korogwe	korogwe_Rutamba_R244	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y	Y		Y
O. korogwe	korogwe_Rutamba_R246	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y
O. korogwe	korogwe_Rutamba_R247	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y	Y	Y	Y
O. korogwe	korogwe_Rutamba_R82	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y	Y		Y
O. korogwe	korogwe_Rutamba_R83	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y	Y		Y
O. korogwe	korogwe_Rutamba_R86	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y	Y		Y
O. korogwe	korogwe_Rutamba_R88	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y	Y		Y
O. korogwe	korogwe_Zigi_P4C10	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_P4C7	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_P4C8	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_P4C9	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_P4D1	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_P4D8	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y		
O. korogwe	korogwe_Zigi_S20K01	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K010	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K011	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K012	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K013	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K014	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_S20K015	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_S20K016	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_S20K017	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K018	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K019	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K02	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K020	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K021	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K022	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_S20K023	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K024	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K03	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K04	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	

O. korogwe	korogwe_Zigi_S20K05	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_S20K06	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K07	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K08	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner			Y	Y	
O. korogwe	korogwe_Zigi_S20K09	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner				Y	
O. korogwe	korogwe_Zigi_Z10	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z12	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z17	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z19	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z2	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z20	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z21	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z22	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z24	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z3	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z4	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z5	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z6	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z7	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z8	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. korogwe	korogwe_Zigi_Z9	Zigi River	18_08_2015	-5.042	38.898	A. Shechonge; S. Bradbeer; B.P. Ngatunga; M. Genner					Y
O. niloticus	niloticus_Mitupa_M183	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga	Y	Y			
O. niloticus	niloticus_Mitupa_M184	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga	Y	Y			
O. niloticus	niloticus_Mitupa_M185	Mitupa, Lindi	21-27_10_2016	-10.011	39.451	T. Blackwell; B.P. Ngatunga	Y	Y			
O. niloticus	niloticus_Nambawala_N165	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga		Y			
O. niloticus	niloticus_Nambawala_N166	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga		Y			
O. niloticus	niloticus_Nambawala_N167	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga		Y			
O. niloticus	niloticus_Nambawala_N170	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga		Y			
O. niloticus	niloticus_Nambawala_N171	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga		Y			
O. niloticus	niloticus_Nambawala_N172	Nambawala, Lindi	21-27_10_2016	-10.044	39.453	T. Blackwell; B.P. Ngatunga	Y	Y			
O. niloticus	niloticus_Nambawala_T4C3 (=TXC3)	Nambawala, Lindi	03_06_2015	-10.044	39.453	A.Shechonge		Y			
O. niloticus	niloticus_Rutamba_R13	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y			
O. niloticus	niloticus_Rutamba_R14	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y			
O. niloticus	niloticus_Rutamba_R152	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y			
O. niloticus	niloticus_Rutamba_R165	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y				
O. niloticus	niloticus_Rutamba_R167	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y				
O. niloticus	niloticus_Rutamba_R171	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y				
O. niloticus	niloticus_Rutamba_R18	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga		Y			
O. niloticus	niloticus_Rutamba_R223	Rutamba, Lindi	21-27_10_2016	-10.032	39.46	T. Blackwell; B.P. Ngatunga	Y	Y			

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

O. urolepis	urolepis_Utete_T3C03	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga				Y	
O. urolepis	urolepis_Utete_T3C04	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga				Y	
O. urolepis	urolepis_Utete_T3C05	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga				Y	
O. urolepis	urolepis_Utete_T3C06	Lugongwe Utete	11_03_2015	-8	38.759	A. Shechonge; B.P. Ngatunga				Y	
O. urolepis	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	GGGGTCATTCGGTTTATTGGTTAT	AGGGCAGGTCACGGGTTCG	(TTTG)8
OMO093	JX204891	AAGCCCCACATAGACGACCAGAGA	CAGAAACGGTGCCTGTTCCAGAA	(CAT)8
OMO100	JX204895	CCTTCCCCACCACTACCCTCATAA	CCCGCCCACACCTGACGA	(ATT)18
OMO114	JX204905	ACGCCTTAATGCTGCCTTCAAGA	TGATGCTCACCCCGTTCCTCA	(GTT)11
OMO129	JX204914	TTGGCAGGCTAAGTACTATTTCAT	GAGCGAATGGTTGTCTGTCTCT	(CCAT)9
OMO161	JX204924	ACTTTGACAAAAGAAGTGTAACAA	AGGGGAGGAGAAAATAAACTGTAT	(TAA)10
OMO219	JX204964	ATCCCCTTCTTTCCATCCCTGTC	AAGGCCTCTGTGAGCTGATTGATT	(TTTTG)10
OMO229	JX204973	GCGACTTTTTCTTTGCACATTTTT	AACTGAACCGCCATCATAATCATC	(GTT)9
OMO248	JX204987	AAAGACACAAAGAGAAACTAATCA	GGATGAATATTTAAAATCAGTCAG	(TCA)9
OMO337	JX205052	TAGGAGAGGCATAGGTTGTCAAAT	CAAGAGTCTAGGAGGGAATCAAAA	(GTTT)7
OMO361	JX205069	TGACAGCGAGCCAGAATGGAAGTA	AAAAGTGAAAGGGGCACAGTGAGG	(CTT)17
OMO391	GR699257	AGACATCTGTACGCTCTTTACGAA	AGTGCTAGAGGGAAGGGGCTGTA	(GAT)9
OMO392	GR698887	CTGGCTTAACTTCTCTACTGGACA	TCTACTCAAAACTGGCAACAAAAC	(GAATA)7
OMO397	GR693794	ACGCGTGTTTGAGATATTTAGATT	GAACAAACAAGGGGAGTGG	(GATT)7
OM-01	GU391020	TTTAAAGTTACACAGCAGTACAAAG	TTGTAGCATTTCAACACAGTCTC	(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	*O. korogwe*	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	*O. korogwe*	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	-	0.07	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	*O. placidus*	N	8	10	10	10	-	9	5	-	10	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	*O. urolepis*	N	26	25	26	26	26	26	22	25	26	21	19	22	25
		NA	7	8	5	4	3	8	18	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	*O. korogwe*	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	*O. niloticus*	N	12	13	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	*O. korogwe*	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	*O. niloticus*	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
		O. korogwe	*O. niloticus*	
Linear (traditional)	*O. korogwe*	16	2	18
	O. niloticus	1	13	14
	Hybrids (OK x ON)	6	2	8
Geometric morphometric	*O. korogwe*	17	1	18
	O. niloticus	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.

		Classified group				
Measurements	Original group	K-M	K-Z	K-N	K-R	Total
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	O. korogwe Mlingano (K-M)	10	0	0	0	10
morphometric	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 *O. placidus* hybrid, SNPs excluding heterozygote only sites	oreo_nucfiltersnps_biallelic_lg_noldfilter.phy	RAxML 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

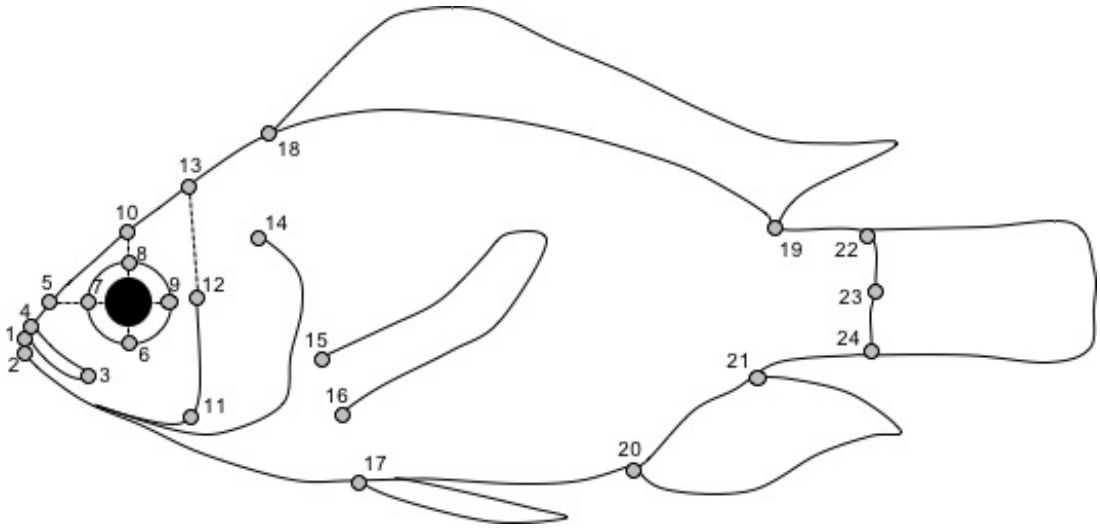


Figure S1. Landmarks used in the geometric morphometric analysis.

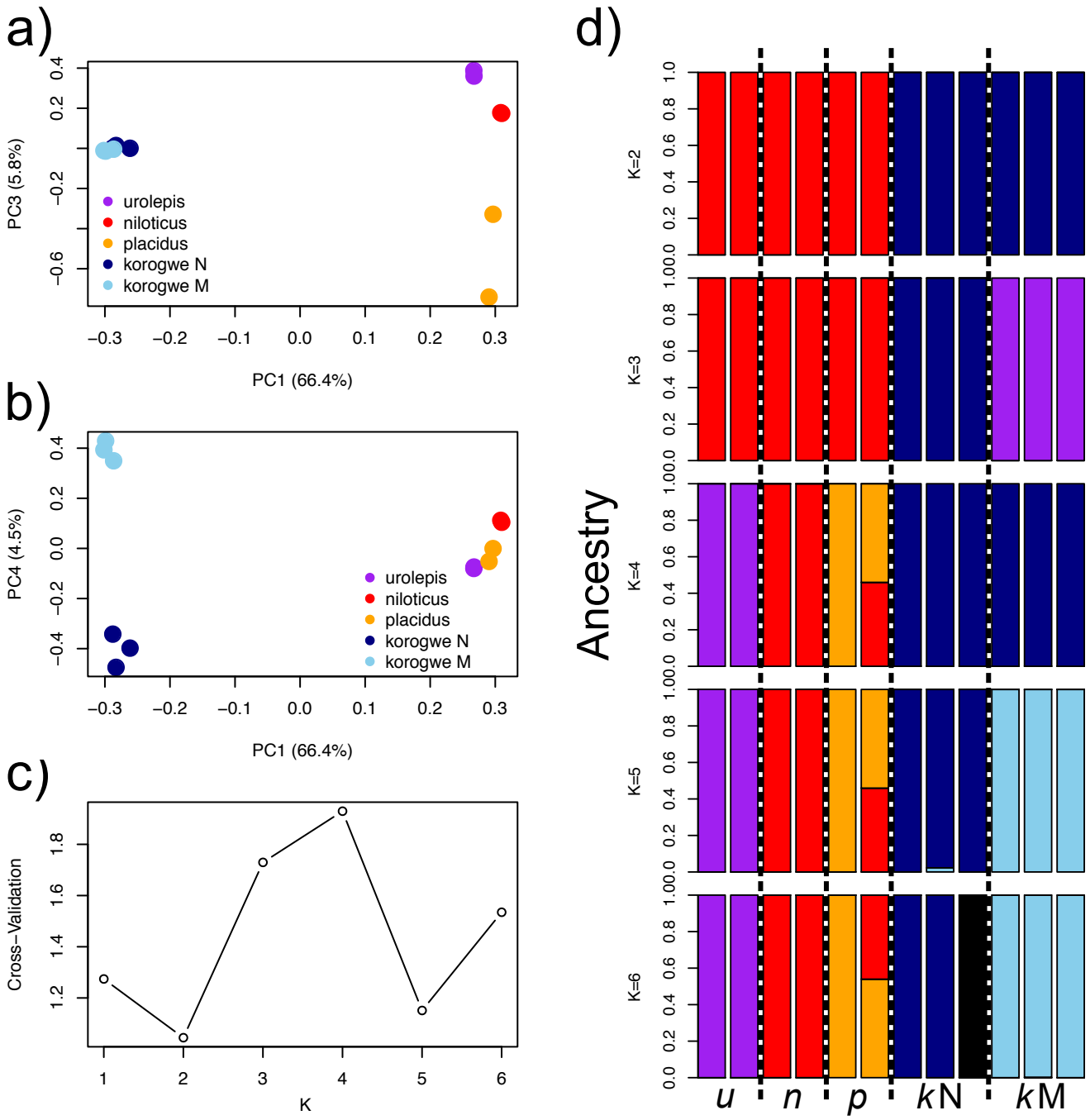


Figure S2. WGR population genetic and phylogenetic analysis. a-b) PCA analysis of LD-pruned nuclear (116,901) SNPs, c) Cross-validation error for admixture analysis of K=1-6, based on 400,680 LD-pruned nuclear SNPs, d) Admixture cluster membership for K=2-5. Species codes used here are u = *O. urolepis*, 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.

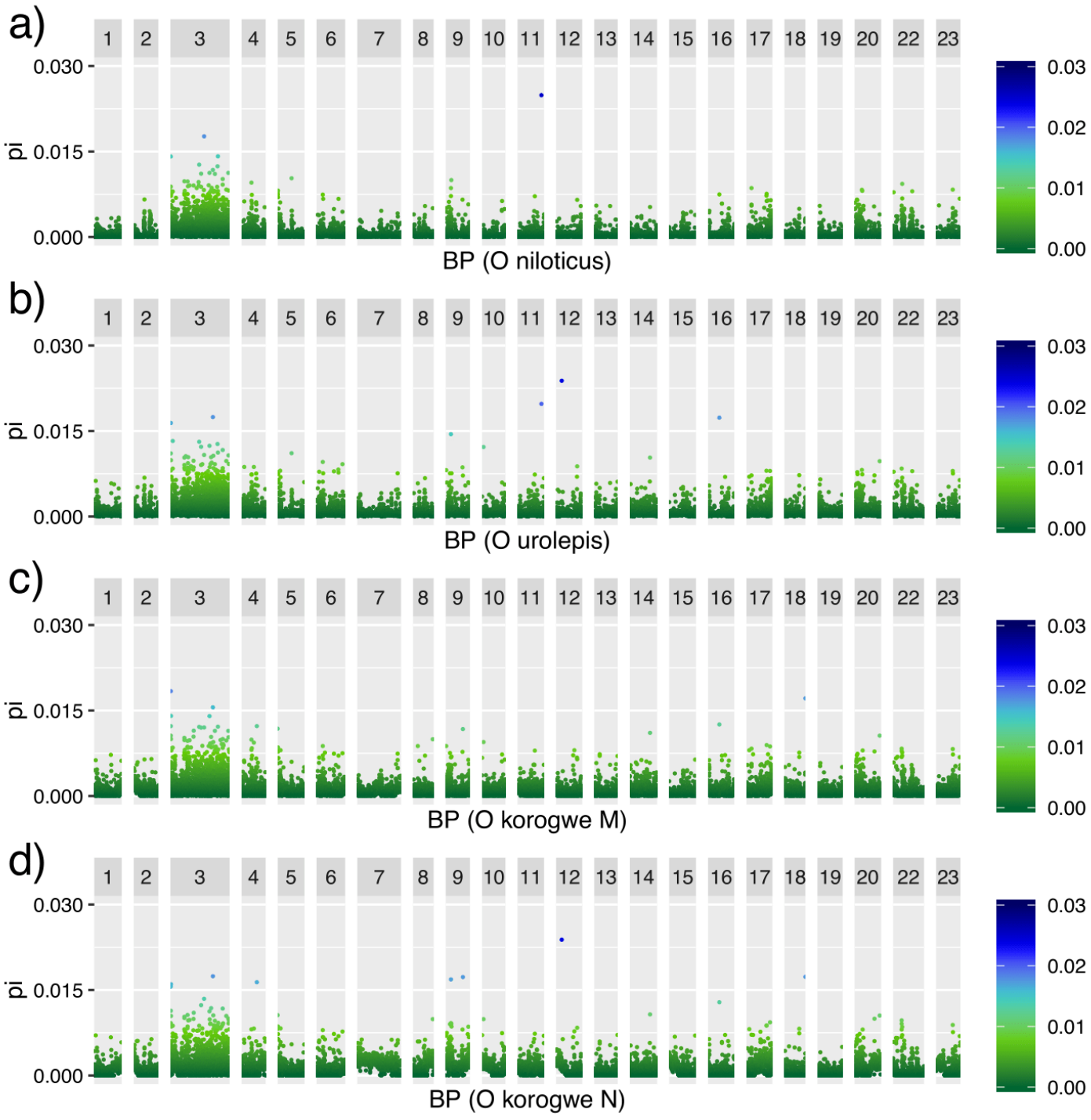


Figure S3. Within population nucleotide diversity (pi) across linkage groups, estimated with whole genome data, in non-overlapping 50kb windows.

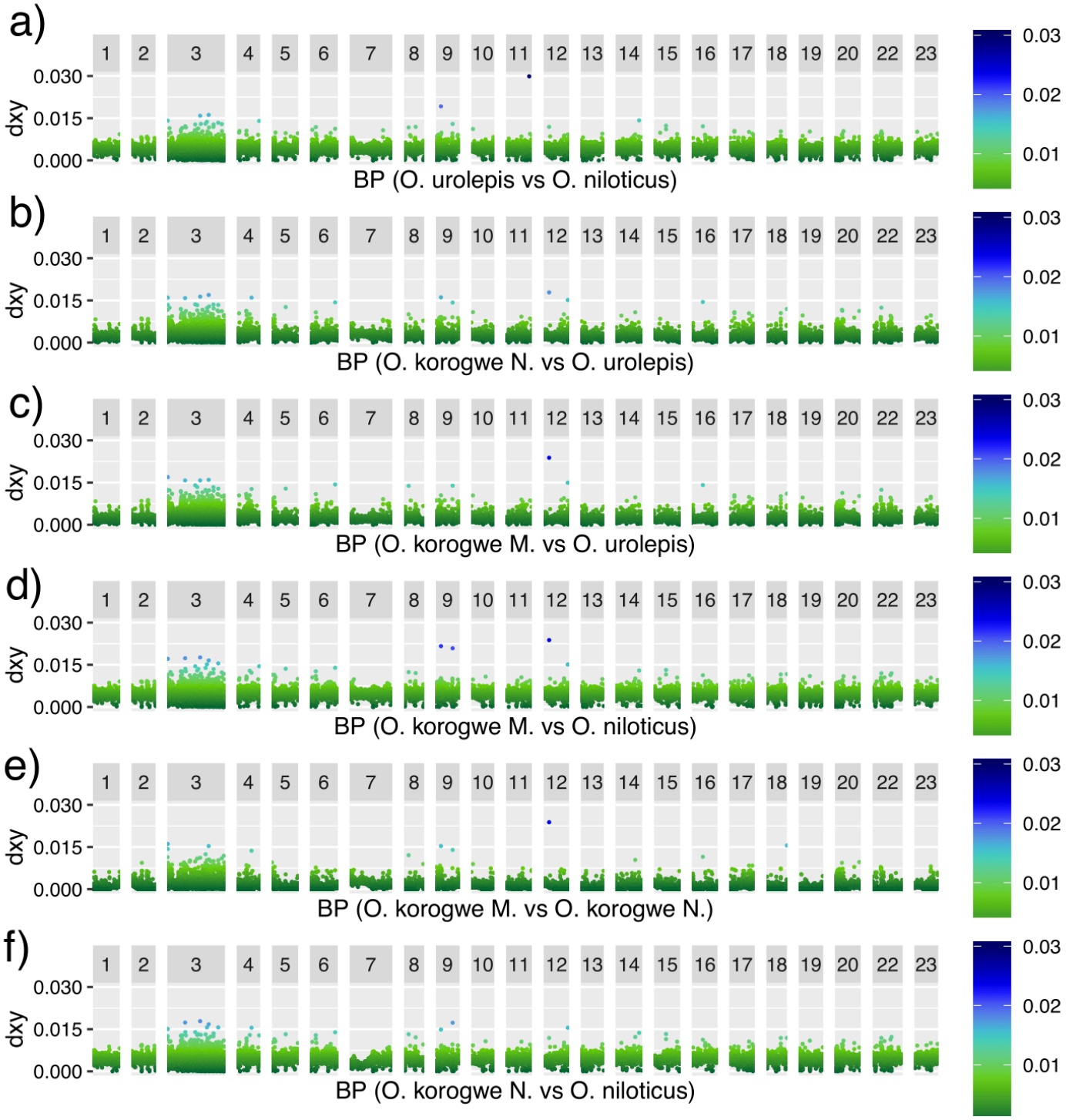


Figure S4. Absolute genetic divergence (Dxy) between population pairs across linkage groups, as estimated using whole genome data, in non-overlapping 50kb windows.