Shechonge, A., Ngatunga, B. P., Tamatamah, R., Bradbeer, S. J., Sweke, E., Smith, A., Turner, G. F., & Genner, M. J. (2019). Population genetic evidence for a unique resource of Nile tilapia in Lake Tanganyika, East Africa. *Environmental Biology of Fishes*, *102*(8), 1107-1117. https://doi.org/10.1007/s10641-019-00895-2

Population genetic evidence for a unique resource of Nile tilapia in Lake Tanganyika, East Africa

Asilatu Shechonge[1,2] , Benjamin P. Ngatunga[2] , Rashid Tamatamah[1,2], Stephanie J. Bradbeer[3], Emmanuel Sweke[2] , Alan Smith[4] , George F. Turner[5] and Martin J. Genner[3]

[1] Department of Aquatic Sciences and Fisheries, University of Dar es Salaam, P.O. Box 35064, Dar es Salaam, Tanzania.

[2] Tanzania Fisheries Research Institute (TAFIRI) PO. Box 9750. Dar es Salaam. Tanzania.

[3] School of Biological Sciences, Life Sciences Building, 24 Tyndall Avenue, University of Bristol. Bristol. BS8 1TQ. United Kingdom.

[4] Evolutionary and Environmental Genomics Group, School of Environmental Sciences, University of Hull, Hull HU5 7RX, United Kingdom.

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

Author to whom correspondence should be addressed: Tel: +44 117 39 41182

email: m.genner@bristol.ac.uk

Nile tilapia (*Oreochromis niloticus*) is one of the most important species in Tanzania for inland fisheries and aquaculture. Although indigenous to the country, it is only naturally distributed within the margins of Lake Tanganyika and peripheral water bodies. The widespread distribution across other parts of the country is a consequence of introductions that started in the 1950s. We investigated the population genetic structure of Nile tilapia across Tanzania using nuclear microsatellite markers, and compared the head and body morphology of populations using geometric morphometric analyses. We found the Lake Tanganyika population to be genetically distinct from the introduced populations. However, there were no clear morphological differences in head and body shape that distinguished the Lake Tanganyika population from the others. We conclude that the Lake Tanganyika population of Nile tilapia represents a unique genetic resource within the country. We suggest that Nile tilapia aquaculture within the Lake Tanganyika catchment should be restricted to the indigenous strain.

Keywords: invasive species, hybridization, conservation genetics, stock structure.

Introduction

The ability of species to adapt to changing environmental conditions is dependent on the availability of standing genetic variation on which selection can act (Hoban et al. 2103). Both capture fisheries and aquaculture practices can deplete genetic diversity through the effects of size-selective harvesting (Frost et al. 2006; Pinsky & Palumbi 2013). Moreover, since many capture fisheries and aquaculture enterprises globally are based on species that have been introduced from other regions of the world, then such populations may particularly prone to founder events and episodes of strong selection associated with adaptation to new environments (Willoughby et al. 2018). Thus, the identification and conservation of natural genetic resources of species widely used in both aquaculture and capture fisheries could in the long-term help to mitigate against losses of genetic diversity and sustain fisheries production (Lind et al. 2012a).

Global production of Nile tilapia *Oreochromis niloticus* (L. 1758) within aquaculture and capture fisheries has been growing at an exponential rate since the 1990s (FAO 2018), and it is now one of the most widely cultured and fished species across tropical and subtropical freshwaters, including those of China, southeast Asia, north Africa, the Levant and central America (Deines et al. 2016). Moreover, since Nile tilapia is becoming a major aquaculture species in sub-Saharan Africa, the production of this species is likely be substantially increased as demand for farmed fish increases over the coming decades in line with human population growth.

Nile tilapia has a primary natural distribution in lakes and slow flowing rivers across the Nile and Niger basins of northern Africa (Trewavas 1983). Across its natural range it is extensively exploited in capture fisheries, and it has also been successfully introduced to natural water bodies and impoundments throughout much of tropical Africa. One of the earliest and most notable introductions of Nile tilapia was into Lake Victoria in the 1950s, initially as an

accidental ‘contaminant’ of stocks of *Coptodon zillii* (Gervais 1848), before deliberate introductions to boost fisheries production (Trewavas 1983). The species subsequently underwent a major population increase in Lake Victoria (Goudswaard et al. 2002), and now supports an important fishery with estimated landings of ~70,000 tonnes in 2010 (Kolding et al. 2014).

Several spatially separated distinct subspecies of Nile tilapia have been recognised in Africa based on morphological differences (Trewavas 1983), consistent with strong natural population genetic substructure within the range of this species (Agnèse et al.1997; Bezault et al. 2011). This natural spatial diversity has the potential to be compromised by interbreeding with introduced populations following escapes from aquaculture facilities, or following deliberate introductions aimed at improving capture fisheries. Already, some genetically and phenotypically distinct native populations of Nile tilapia are considered threatened because of hybridization with invading species, for example the blue spotted tilapia (*Oreochromis leucostictus*) (Ndiwa et al. 2014).

Nile tilapia from Lake Tanganyika is the most southerly population within the natural range of the species. The evidence that Nile tilapia is native to Lake Tanganyika comes from capture records that date as far back as 1906 (Trewavas 1983; Van Steenberge et al. 2011), before the first continuous aquaculture and fisheries improvement research activities in East Africa that took place during the mid-20th century (EAFFRO 1967). Lake Tanganyika is within Congo drainage, and thus is presently disconnected from other parts of the natural range of the species. Precisely how Nile tilapia arrived in Lake Tanganyika is unclear, but it is possible that it arrived naturally from Lake Kivu within the last 9,500-14,000 years, after volcanic activity blocked the northern connection of Lake Kivu to the Nile system, forming the Ruzizi river which flows into the northern Lake Tanganyika (Snoeks et al. 1997; Danley et al. 2012). In support of this scenario is evidence that Nile tilapia is native to Lake Kivu (Snoeks et al. 1997),

which has a history of faunal connectivity with Lake Tanganyika, for example through shared distributions of the migratory cyprinids *Raiamas moori* (Boulenger 1900) and *Labeobarbus altianalis* (Boulenger 1900) (Snoeks et al. 1997).

Although several studies have tested for genetic evidence of hybridization between invasive Nile tilapia and indigenous *Oreochromis* within East Africa (Nyingi et al. 2007; Ndiwa et al. 2014; Shechonge et al. 2018; Bradbeer et al. 2019), there have been few studies of population-genetic differentiation among Nile tilapia populations of the region (Agnèse et al. 1997; Fuerst et al. 2000; Nyingi et al. 2009; Bezault et al. 2011), and none have considered variation among populations in Tanzania. Thus, here we test for population-level genetic differences among populations of Nile tilapia in Tanzania, focussing on comparisons between the indigenous Lake Tanganyika Nile tilapia and populations known to be introduced elsewhere in the country for aquaculture and fisheries improvement. We also test for morphological differences between the Lake Tanganyika population and the introduced populations.

Methods

Sampling

We collected samples of Nile tilapia from eight locations during 2015 and 2016, within the catchments of the Pangani River and Lakes Victoria, Eyasi and Tanganyika (Table 1; Fig. 1). Samples were collected from artisanal fishers or from experimental fishing using a seine net or gill net. Samples from fishers were already dead at the time of collection, while live fish collected from the nets were subjected to an overdose of clove oil (eugenol) anaesthetic on landing. Individual fish were pinned out with the head facing left, photographed from a standard orientation, and individually labelled. From each fish, we collected a tissue sample (fin clip)

preserved in absolute ethanol. Whole fish were then preserved in absolute ethanol, before transfer to 70% IMS for long term storage.

DNA extraction and microsatellite genotyping.

A piece of fin tissue approximately 3 x 3 mm was air dried, and the DNA was extracted using the Promega Wizard DNA extraction kit. Individual samples were then analysed to quantify variation at 17 microsatellite loci (Supplementary Information Table 1), sourced from Saju et al. (2010) and Liu et al. (2013). PCR was performed in a volume of 10µl, consisting of 1µl DNA (~5ng), 5µl Mastermix and 4µl primer mix (10mM). Each primer was labelled with one dye from the ABI DS-33 set (either 6-FAM, VIC, PET, NED). PCR amplifications were conducted within one of two multiplex PCR amplifications. PCR conditions for each multiplex consisting of one denaturation step of 15 minutes at 95^{o}C, followed by 35 cycles of 30 seconds denaturation at 94^{o}C, 90 seconds annealing at 57^{o}C and 60 seconds extension at 72^{o}C, followed by a final extension step of 30 minutes at 60^{o}C. Samples were run on an ABI 3500 automated sequencer against a LIZ 500 size standard, and allele sizes for each locus were identified using GeneMapper 4.1 (Applied Biosystems, MA).

Molecular data analysis

Individual loci were checked for significant deviation from Hardy-Weinberg equilibrium using Arlequin 3.5 (Excoffier and Lischer 2010). In the 126 tests of deviation from Hardy Weinberg Equilibrium across the 17 loci, 23 were significant at $P < 0.05$, and in 21 of those cases observed heterozygosity was lower than expected heterozygosity. However, only one locus (OM-01) showed a consistent deviation from Hardy-Weinberg equilibrium across populations (a

heterozygote deficit), and data for this locus were excluded from further analysis. To compare genetic diversity among populations, we calculated a standardised allelic richness for each locus within in each population using rarefaction within HP-Rare, selecting the option for a sample of 10 "genes" (Kalinowski 2005). We tested for significant differences in rarefied allelic richness among populations we used a general linear model in R 3.6.0 (R Core Team 2019), followed by estimation of least square means and implementation of Tukey's *post-hoc* tests using the R package lsmeans (Lenth 2016).

To quantify population genetic subdivision, we used F_{ST} calculated in Genepop 4.2. (Rousset 2008), alongside Exact tests based on 10,000 dememorisation steps, and 100 batches of 10,000 iterations. To ordinate genetic differences among individuals we used Principal Component Analysis (PCA) implemented in adegenet 2.1.1 (Jombart & Ahmed 2011) in R 3.6.0. To estimate the probability of individual membership to *K* populations we used Structure 2.3.4 (Pritchard et al. 2000), with the admixture model, no location priors, and 10 runs each with 100,000 burn-in steps and 100,000 recorded steps. The Structure output was then entered into Clumpak (Kopelman et al. 2015) to estimate the optimal number of populations present in the dataset using the Evanno method (Evanno et al. 2005). The probability of membership of individuals to those clusters was then graphically illustrated.

Morphological analyses

The left side of each specimen was photographed in a standard orientation, alongside a scale bar. Images were loaded into tpsDIG 2.26 (Rohlf 2015), using a file generated in tpsUtil 1.74 (Rohlf 2015) and a total of 24 landmarks were placed on a calibrated image of each individual (Fig. 2). The resultant landmark coordinates were then aligned using a Procrustes analysis in MorphoJ 1.06 (Klingenberg 2011), and the generated Procrustes coordinates were used in a

pooled between-groups regression against centroid size, generating size-standardised residual Procrustes coordinates. These size-standardised Procrustes coordinates were then used within a Principal Components Analysis (PCA) to ordinate observed shape differences among individuals in MorphoJ 1.07a (Klingenberg et al. 2011) We tested the significance of shape differences between populations along the two primary axis of shape variation (PC1 and PC2) using a general linear model in R 3.6.0, followed by Tukey's *post-hoc* tests of pairwise differences between populations.

Results

Population genetic structure and genetic diversity

Overall, there were highly significant genetic differences among the eight populations (Global F_{ST} = 0.249; Exact test $P < 0.001$). Between the population pairs, F_{ST} ranged from 0.016 to 0.431 (Table 2), and all populations were significantly different (Exact tests, $P < 0.001$). Principal Component Analysis (PCA) separated three clusters of individuals along PCA axes 1 and 2. One cluster comprised the population from Lake Tanganyika, a second cluster comprised the population from Mwamapuli, and the third cluster comprised individuals sampled from other locations (Fig. 3). Within this third cluster, populations from the eastern Pangani system (Kerenge, Kumba, Pangani Falls) were tightly clustered, while the populations from the western Pangani system (Nyumba ya Mungu, Kivulini) were closely clustered with those from Lake Victoria (Fig. 3).

The optimum number of genetic clusters in the dataset, according the Structure analysis applying the Evanno method was $K = 7$ (Fig. 4). The analysis indicated that the populations from Lake Tanganyika, Mwamapuli, Lake Kumba and Nyumba ya Mungu were largely distinct

from one another, and the other populations. Meanwhile, the populations from Lake Victoria and Kivulini were similar in allelic composition. The populations from Kerenge and Pangani Falls were similar, albeit heterogeneous, with some individuals sharing considerable allelic similarity with the Lake Victoria population (Fig. 4).

Rarefied allelic richness differed significantly among loci ($F_{16,105} = 9.213$, $P < 0.001$), and among populations ($F_{7,105} = 7.561$, $P < 0.001$; Table 1). In *post-hoc* comparisons, the Pangani Falls population had elevated diversity relative to those from Kivulini, Lake Kumba, Mwamipuli and Lake Tanganyika, while the Lake Kumba population had lower genetic diversity than Kerenge, Mwamipuli and Lake Tanganyika (Table 1; Supplementary Information Table 2).

Morphological differences among populations

Principal Component axis 1 captured variation in head and eye size, with individuals with positive PC1 scores possessing relative elongate snouts and larger eyes than individuals with negative scores. Principal Component axis 2 captured variation in body depth, with individuals with positive PC2 scores possessing shallower body depth than individuals with negative scores (Figure 5). Overall there was a highly significant differences among populations along these two axes of morphological variation ($F_{7,126} = 9.599$, $P < 0.001$). In *post-hoc* tests we found significant morphological differences in 11 of the 28 pairwise comparisons (Table 3). However, we found no clear evidence of morphological separation of the Lake Tanganyika population from the introduced populations sampled elsewhere in Tanzania. Instead, the Lake Tanganyika population overlapped in morphospace with most populations.

Discussion

Our results demonstrate that Nile tilapia collected around the margins of north-eastern Lake Tanganyika are genetically distinct from those sampled elsewhere in Tanzania, despite the lack of any clear diagnostic morphological differences. The apparent genetic uniqueness of this Tanganyika population is consistent with a long-period of separation from other populations sampled in Tanzania.

It seems unlikely that the samples we obtained are exclusively a result of recent colonisation of the sampled region by an invasive strain, but it is not unusual for fish to escape aquaculture facilities and introgress with wild stocks (Faust et al. 2018; Wringe et al. 2018), and this can have consequences for ecologically-important phenotypes of the wild populations (Bolstad et al. 2017). We cannot rule out the possibility that the Nile tilapia samples we collected from the Lake Tanganyika catchment are contaminated with recent escapes from aquaculture systems within the basin. For example, the Chitralada strain of Nile tilapia from Thailand has been reported in aquaculture within Burundi (https://bit.ly/2JvI0N3; https://bit.ly/2EfAGB9), and thus is potentially inside the Lake Tanganyika catchment. Contamination from genetically similar non-native stocks could explain the apparently high allelic similarity between two individuals from the Lake Tanganyika and those from Nyumba-ya-Mungu dam (Fig. 4). However, further sampling of Nile tilapia across its native and introduced range across Africa is required to test for introgression between indigenous and introduced strains.

233 Genetic structuring of introduced populations

234

235 It is commonplace to find population genetic structuring among naturally occurring populations
236 of Nile tilapia (Table 4). Nevertheless, our finding of the substantial genetic structure among
237 the non-native populations of Nile tilapia in Tanzania (average F_{ST} = 0.191, standard deviation
238 0.092) is perhaps surprising given the relatively recent introductions of the species into the
239 country. The most plausible explanation is that the high levels of genetic differentiation are
240 driven by demographic processes that influence genetic diversity, including founder events
241 and/or selection, perhaps associated with fisheries activity. In experimental conditions, Eguia
242 et al. (2005) showed strong genetic divergence (F_{ST} = 0.130) between a control and size-
243 selected populations of Nile tilapia over as few as four generations. Spatial connectivity may
244 also have affected genetic similarity of the populations from Lake Kumba, Kerenge and Pangani
245 falls which are near one another and connected by flowing waterways. Finally, the timescale of
246 divergence may have been influenced the extent of genetic divergence observed. For example,
247 the populations from the Kivulini fishponds and Lake Victoria are genetically similar, which
248 was expected given that Lake Victoria was cited as the original source of the fish we sampled
249 from the newly constructed ponds by the owner at the time of sampling.

250

251 Another explanation for the presence of population genetic structure among our studied
252 introduced populations is that they were seeded from multiple geographically distinct sources.
253 Different Nile tilapia strains commonly used in aquaculture in Asia, for example, have clear
254 genetic differences when studied using microsatellite loci (Sukmanomon et al. 2012; Table 4).
255 Certainly, not all Nile tilapia in the country are from the same source, as shown by the recent
256 arrival of the Chitralada strain at ponds in Dar es Salaam (Shechonge et al. 2019). A further
257 explanation is that genetic differentiation is partially linked to hybridization with other

Oreochromis species. Relatively rare hybridization events between *O. niloticus* and native species are known from multiple locations relevant to our sampling, including satellite lakes of Lake Victoria [*O. esculentus* (Graham 1928); Angienda et al. 2011], the Pangani falls dam [*O. korogwe* (Lowe 1955); Bradbeer et al. 2019] and Nyumba ya Mungu [*O. jipe* (Lowe 1955); Bradbeer et al. 2019].

Aquaculture potential and the conservation of an indigenous genetic resource

Increased aquaculture production is required to meet demands for fish protein from the growing human population (FAO, 2018). At present, the aquaculture production potential of the Lake Tanganyika Nile tilapia population is unknown. We are unaware of any aquaculture facilities using this strain, and typically aquaculture in the Tanzanian sector of the Lake Tanganyika catchment focusses primarily on the other large-bodied indigenous species *Oreochromis tanganicae* (Günther 1894) and *Oreochromis malagarasi* Trewavas 1983. Controlled growth trials of these two species, alongside indigenous Nile tilapia, would inform us of their collective aquaculture potential as the industry expands to support the growing human population of the region.

An expanding aquaculture industry requires strains of farmed fish that are resistant to emerging diseases and are able to thrive given the specific environmental conditions. The increasing importance of Nile tilapia in global aquaculture implies that genetic resources will be required to facilitate the selective breeding of improved varieties (Eknath and Hulata 2009; Lind et al. 2012b). Our results indicating unique status of the Lake Tanganyika population imply that it should be valued for its potential to contribute to future selective breeding programmes. The introduction of Nile tilapia from other sources into the catchment could potentially lead to

intraspecific hybridization and the dilution or loss of this unique genetic resource. Already at least one potentially invasive populations of Nile tilapia of uncertain provenance is present in the upper Malagarasi river connected to Lake Tanganyika (Shechonge et al. 2019). Given the uncertainty regarding the outcome of direct contact between non-native and native strains of Nile tilapia, we suggest that further development of Nile tilapia aquaculture and fisheries in the region should be based on the indigenous population to reduce the likelihood of erosion of the Lake Tanganyika Nile tilapia genetic resource.

Concluding remarks

Key questions remaining from this study relate to the processes that have driven the patterns of spatial genetic variation in Tanzania, and to answer these requires more extensive sampling of both Nile tilapia and native *Oreochromis* populations in Tanzania. It also requires sampling of wild stocks in neighbouring countries, as well as the high-performance commercially farmed strains from which introduced broodstock could have been sourced. With the recent availability of high-quality reference genomes of Nile tilapia (Brawand et al. 2014; Conte et al. 2017), it is now possible to accurately conduct genome-wide analyses to quantify intraspecific gene flow, introgression and reconstruct population demography, and to map traits beneficial for fisheries production on the genome. Such information will further clarify the value of the Lake Tanganyika Nile tilapia population as a genetic resource, while potentially verifying and explaining the patterns of population genetic structuring we have recovered in this study. Knowledge of the genomic composition of populations in a comparative framework would also inform future investigations of phenotypic traits that could be useful for aquaculture and capture fisheries development, and potentially inform the development of future strains of this globally important species.

Acknowledgements. The work was supported by Royal Society-Leverhulme Trust Africa Awards AA100023 and AA130107 to MJG, BPN and GFT, and a BBSRC award BB/M026736/1 to GFT and MJG. The Tanzania Commission for Science and Technology (COSTECH) provided fieldwork permits. We thank Carlos Gracida Juarez and staff from the Tanzania Fisheries Research Institute for contributions to fieldwork, and Jack Harrington for laboratory support.

References

Agnèse JF, Adépo-Gourène B, Abban EK, Fermon Y (1997) Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). Heredity 79: 88-96.

Angienda PO, Lee HJ, Elmer KR, Abila R, Waindi EN, Meyer A (2011) Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. Conservation Genetics 12: 243-255.

Bezault E, Balaresque P, Toguyeni A, Fermon Y, Araki H, Baroiller JF, Rognon X (2011) Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. BMC Genetics 12: 102.

Bolstad GH, Hindar K, Robertsen G, Jonsson B, Sægrov H, Diserud OH, Fiske P, Jensen AJ, Urdal K, Næsje TF, Barlaup BT (2017) Gene flow from domesticated escapes alters the life history of wild Atlantic salmon. Nature Ecology & Evolution 1: 0124.

Bradbeer SJ, Harrington J, Watson H, Warraich A, Shechonge A, Smith A, Tamatamah R, Ngatunga BP, Turner GF, Genner MJ (2019) Limited hybridization between introduced

and Critically Endangered indigenous tilapia fishes in northern Tanzania. Hydrobiologia 832 : 257-268.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, ... Turner-Maier J (2014) The genomic substrate for adaptive radiation in African cichlid fish. Nature 513: 375-381.

Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD (2017) A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. BMC Genomics 18: 341.

Danley PD, Husemann M, Ding B, DiPietro LM, Beverly EJ, Peppe DJ (2012) The impact of the geologic history and paleoclimate on the diversification of East African cichlids. International Journal of Evolutionary Biology 2012: 574851.

Deines AM, Wittmann ME, Deines JM, Lodge DM (2016) Tradeoffs among ecosystem services associated with global tilapia introductions. Reviews in Fisheries Science and Aquaculture 24: 178-191.

EAFFRO (1967) The history and research results of the East African Freshwater Fisheries Research Organization from 1946 - 1966. East African Freshwater Fisheries Research Organization, Jinja, Uganda.

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

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

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567.

355 Faust E, Halvorsen KT, Andersen P, Knutsen H, André C (2018) Cleaner fish escape salmon
356 farms and hybridize with local wrasse populations. Royal Society Open Science 5:
357 171752.
358 FAO (2018) The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable
359 development goals. FAO, Rome.
360 Frost LA, Evans BS, Jerry DR (2006) Loss of genetic diversity due to hatchery culture practices
361 in barramundi (*Lates calcarifer*). Aquaculture 261: 1056-1064.
362 Fuerst PA, Mwanja WW, Kaufman L (2000) The genetic history of the introduced Nile tilapia
363 of Lake Victoria (Uganda – E. Africa): the population structure of *Oreochromis niloticus*
364 (Pisces: Cichlidae) revealed by DNA microsatellite markers. In Tilapia Aquaculture in
365 the 21st Century. Proceedings from the Fifth International Symposium on Tilapia in
366 Aquaculture (eds. K. Fitzsimmons and J. Carvalho) Ministry of Agriculture, Brazil and
367 Aquaculture CRSP. Rio de Janeiro.
368 Goudswaard PC, Witte F, Katunzi EFB (2002) The tilapiine fish stock of Lake Victoria before
369 and after the Nile perch upsurge. Journal of Fish Biology 60: 838-856.
370 Gu DE, Mu XD, Xu M, Luo D, Wei H, Li YY, Zhu YJ, Luo JR, Hu YC (2016) Identification
371 of wild tilapia species in the main rivers of south China using mitochondrial control
372 region sequence and morphology. Biochemical Systematics and Ecology 65: 100-107.
373 Hassanien HA, Gilbey J (2005) Genetic diversity and differentiation of Nile tilapia
374 (*Oreochromis niloticus*) revealed by DNA microsatellites. Aquaculture Research 36:
375 1450-1457.
376 Hoban SM, Hauffe HC, Pérez-Espona S, Arntzen JW, Bertorelle G, Bryja J, Frith K, Gaggiotti
377 OE, Galbusera P, Godoy JA, Hoelzel AR (2013) Bringing genetic diversity to the
378 forefront of conservation policy and management. Conservation Genetics Resources 5:
379 593-598.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27: 3070-3071.

Kalinowski ST (2005) HP-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5: 187-189.

Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. Molecular Ecology Resources 11: 353-357.

Kolding J, Medard M, Mkumbo O, van Zwieten PAM (2014) Status, trends and management of the Lake Victoria Fisheries. In Welcomme, R.L, J. Valbo-Jørgensen and A.S. Halls (eds.). Inland fisheries evolution and management – case studies from four continents. FAO Fisheries and Aquaculture Technical Paper 579. FAO, Rome.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15: 1179-1191.

Lenth R (2016) Least-Squares Means: The R Package lsmeans. Journal of Statistical Software 69: 1-33.

Lind CE, Ponzoni RW, Nguyen NH, Khaw HL (2012a) Selective breeding in fish and conservation of genetic resources for aquaculture. Reproduction in Domestic Animals 47: 255-263.

Lind CE, Brummett RE, Ponzoni RW (2012b) Exploitation and conservation of fish genetic resources in Africa: issues and priorities for aquaculture development and research. Reviews in Aquaculture 4: 125-141.

Liu F, Sun F, Li J, Xia JH, Lin G, Tu RJ, Yue GH (2013) A microsatellite-based linkage map of salt tolerant tilapia (*Oreochromis mossambicus* x *Oreochromis* spp.) and mapping of sex-determining loci. BMC Genomics 14: 58.

Ndiwa TC, Nyingi DW, Agnèse J-F (2014) An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Loboi drainage, Kenya. PloS One, 9: e106972.

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

Pinsky ML, Palumbi SR (2014) Meta-analysis reveals lower genetic diversity in overfished populations. Molecular Ecology 23: 29-39.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rohlf FJ (2015) The tps series of software. Hystrix, the Italian Journal of Mammalogy 26: 1-4.

Romana-Eguia MRR, Ikeda M, Basiao ZU, Taniguchi N (2005) Genetic changes during mass selection for growth in Nile tilapia, *Oreochromis niloticus* (L.), assessed by microsatellites. Aquaculture Research 36L 69-78.

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

Rutten MJM, Komen H, Deerenberg RM, Siwek M, Bovenhuis H (2004) Genetic characterization of four strains of Nile tilapia (*Oreochromis niloticus* L.) using microsatellite markers. Animal Genetics 35, 93-97.

Saju JM., Lee WJ, Orban L (2010) Characterization of nine novel microsatellites isolated from Mozambique tilapia, *Oreochromis mossambicus*. Conservation Genetics Resources 2: 385-387.

Shechonge A, Ngatunga BP, Tamatamah R, Bradbeer SJ, Harrington J, Ford AG, Turner GF, Genner MJ (2018) Losing cichlid fish biodiversity: genetic and morphological homogenization of tilapia following colonization by introduced species. Conservation Genetics 19: 1199-1209.

Shechonge A, Ngatunga BP, Bradbeer SJ, Day JJ, Freer JJ, Ford AG, Kihedu J, Richmond T, Mzighani S, Smith AM, Sweke EA, Tamatamah R, Tyers A, Turner GF, Genner MJ (2019) Widespread colonization of Tanzanian catchments by introduced *Oreochromis* tilapia fishes: the legacy from decades of deliberate introduction. Hydrobiologia, 832: 235-253.

Snoeks J, De Vos L, Van den Audenaerde DT (1997) The ichthyogeography of Lake Kivu. South African Journal of Science 93: 579-584.

Sukmanomon S, Kamonrat W, Poompuang S, Nguyen TT, Bartley DM, May B, Na-Nakorn U (2012) Genetic changes, intra-and inter-specific introgression in farmed Nile tilapia (*Oreochromis niloticus*) in Thailand. Aquaculture 324: 44-54.

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

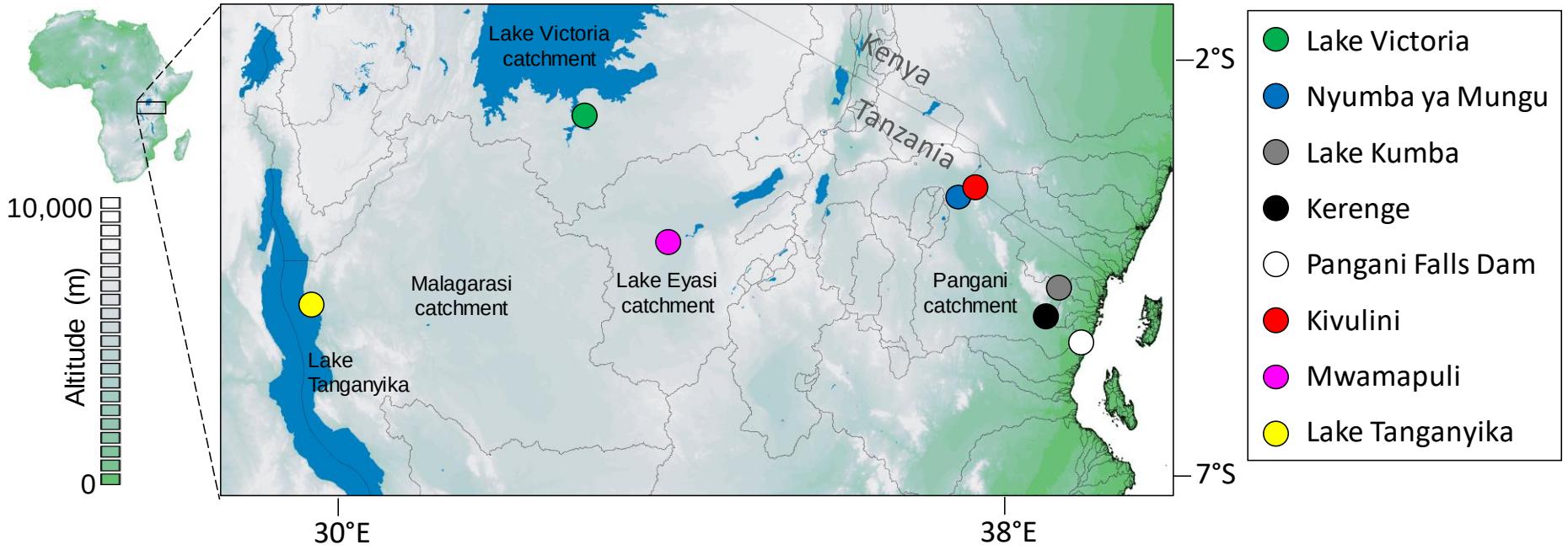
Van Steenberge M, Vanhove M, Risasi DM, N'sibula TM, Bukinga FM, Pariselle A, Gillardin C, Vreven E, Raeymaekers J, Huyse T, Volckaert F (2011) A recent inventory of the fishes of the north-western and central western coast of Lake Tanganyika (Democratic Republic Congo). Acta Ichthyologica et Piscatoria 41: 201-214.

Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR (2018) Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. Molecular Ecology 27: 4041-4051.

Wringe BF, Jeffery NW, Stanley RR, Hamilton LC, Anderson EC, Fleming IA, Grant C, Dempson JB, Veinott G, Duffy SJ, Bradbury IR (2018) Extensive hybridization following

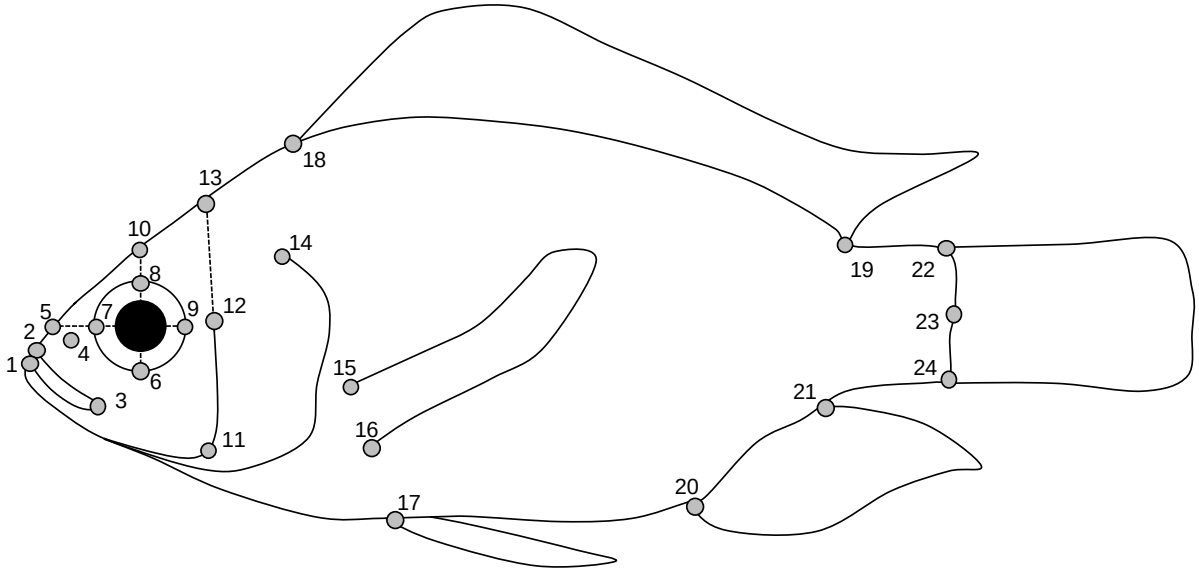
453 a large escape of domesticated Atlantic salmon in the Northwest Atlantic.
454 Communications Biology 1: 108
455 Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nature Reviews
456 Genetics 2: 983-989.

457



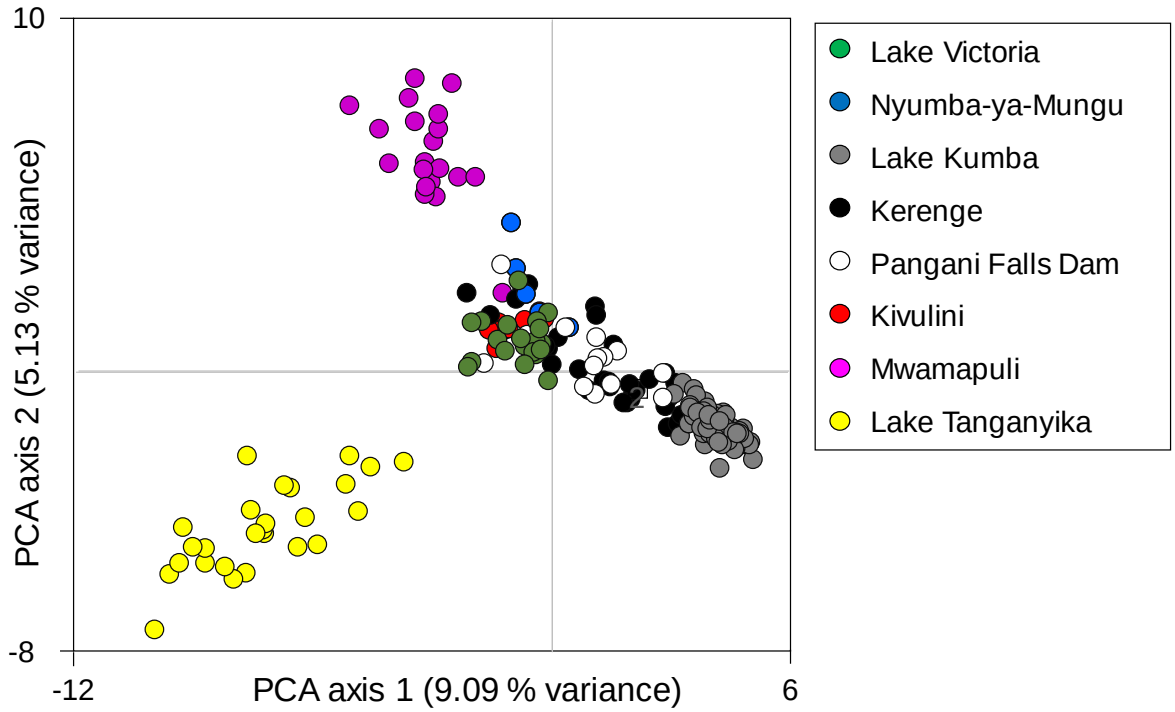
458

459 **Fig. 1** Locations of eight sampling sites in northern Tanzania



460

461 **Fig. 2** Landmarks used in geometric morphometric analysis of Nile tilapia.



462

463

464 **Fig. 3** Principal Component ordination plot illustrating genetic differences among
465 individuals.

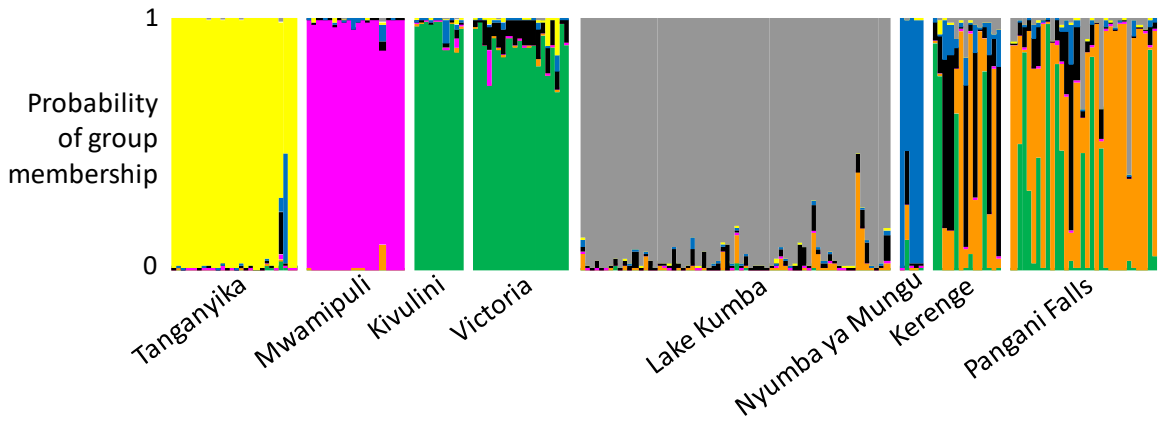
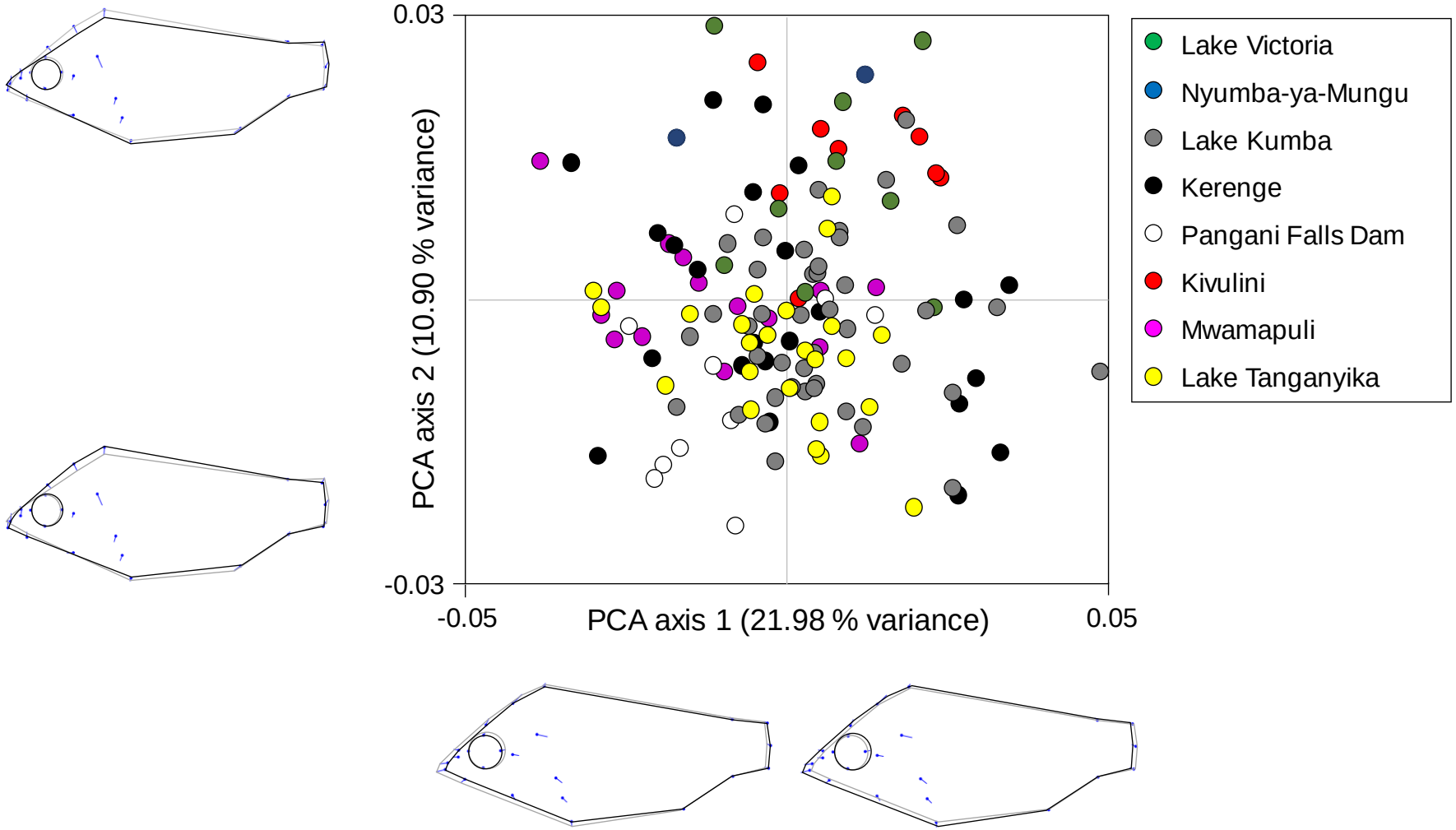


Fig. 4 Structure plot of probability of that individuals belong to genetic groupings. The optimal number of groups shown is *K*=7, following the Evanno method. Each colour represents one genetic grouping.



470

471 **Fig. 5** Principal Component ordination plot illustrating shape variation among *O. niloticus* populations. Shape variation is illustrated using

472 outlined lollipop plots, with darker lines indicative of phenotypes at the extremes of each axis.

Table 1. Sampling localities and sample sizes for molecular and morphological analyses. RAR = Rarefied allelic richness (measured across 10 "genes") in HPRare (Kalinowski 2005).

Site name	Coordinates	Sampling dates	Sampling method	N genetics	N morphology	RAR (± 95% CI)
Lake Tanganyika*	4.859 °S, 29.621°E 4.907°S, 29.665°E 5 211°S, 29.842°E	27-29 / 07 / 2016	Artisanal fishers	26	24	3.27 (0.39)
Mwamapuli	4.356°S, 33.876°E	02 / 08 / 2016	Seine net	20	15	3.22 (0.39)
Kivulini	3.479°S, 37.589°E	14 / 08 / 2015	Seine net	10	9	2.98 (0.37)
Kerenge	5.032°S, 38.548°E	12 / 08 / 2015	Seine net	30	23	3.65 (0.37)
Lake Kumba	4.806°S, 38.621°E	12 / 08 / 2015	Artisanal fishers	64	42	2.32 (0.37)
Nyumba ya Mungu	3.612°S, 37.459°E	14 / 08 / 2015	Artisanal fishers	5	2	3.42 (0.39)
Pangani falls	5.347°S, 38.645°E	19 / 08 / 2015	Gill net	14	10	4.14 (0.37)
Lake Victoria**	2.627°S, 32.899°E 2.588°S, 32.855°E	04-06 / 08 / 2016	Artisanal fishers	20	9	3.14 (0.43)

*samples from 3 sites, n for genetics: Ujiji n = 4; Malagarasi n=2; Kigoma n=20. Samples were pooled for analyses was no evidence of significant genetic structuring among them (Global F_{ST} = -0.011; Exact test P = 0.575).

** samples from 2 sites, n for genetics: Lake Malimbe n = 14, Mwanza Gulf n = 6. Samples were pooled for analyses as there was no evidence of significant genetic structuring among them (Global F_{ST} = -0.008; Exact test P = 0.215).

Table 2. Genetic differences among populations (F_{ST}). All comparisons were highly significantly different ($P < 0.001$) in pairwise exact tests.

Population	Lake Tanganyika	Mwamapuli	Kivulini	Kerenge	Lake Kumba	Nyumba ya Mungu	Pangani Falls
Mwamapuli	0.254						
Kivulini	0.284	0.228					
Kerenge	0.267	0.182	0.123				
Lake Kumba	0.431	0.325	0.344	0.149			
Nyumba ya Mungu	0.320	0.235	0.239	0.145	0.347		
Pangani Falls	0.253	0.171	0.122	0.016	0.158	0.088	
Lake Victoria	0.269	0.212	0.087	0.130	0.340	0.237	0.133

Table 3. Results of *post-hoc* Tukey's tests (p-values) indicating significance of morphological differences among populations, as captured along the first two axes of morphological variation (PC1 and PC2; Figure 5).

Population	Lake Tanganyika	Mwamapuli	Kivulini	Kerenge	Lake Kumba	Nyumba ya Mungu	Pangani Falls
Mwamapuli	0.957						
Kivulini	< 0.001	< 0.001					
Kerenge	0.834	0.272	0.002				
Lake Kumba	0.249	0.032	0.005	0.995			
Nyumba ya Mungu	0.426	0.196	0.999	0.798	0.906		
Pangani Falls	0.465	0.975	< 0.001	0.043	0.004	0.067	
Lake Victoria	0.002	0.000	0.991	0.055	0.122	1.000	< 0.001

Table 4. F_{ST} values reported from published population genetic studies of Nile tilapia using microsatellite markers.

Studied populations	Number of microsatellite markers	Location	Number of comparisons (strains or populations)	Mean F_{ST}	Standard deviation of F_{ST}	Maximum F_{ST}	Minimum F_{ST}	Reference
Native range	9	Africa	10	0.340	0.177	0.723	0.054	Bezault et al. (2011)
Native range	6	Egypt	5	0.035	-	-	-	Hassanien & Gilbey (2005)
Native range	6	Kenya	4	0.216	0.050	0.290	0.127	Nyingi et al. (2009)
Native range	16	Kenya	6	0.164	0.099	0.352	0.018	Ndiwa et al. (2014)
Introduced range (feral)	8	Kenya	4	0.042	0.018	0.069	0.020	Angienda et al. (2010)
Introduced range (feral)	10	China	5	0.207	0.150	0.376	0.030	Gu et al. (2014)
Within culture (non feral)	14	Thailand	7	0.087	0.087	0.194	0.012	Sukmanomon et al. (2012)
Within culture (non feral)	14	Global	4	0.176	0.093	0.333	0.084	Rutten et al. (2004)

Supplementary Information Table 1 Genetic diversity of 17 microsatellite loci at the sampling locations. N = sample size, HO = Observed heterozygosity, HE = Expected heterozygosity, P = probability of Hardy-Weinberg equilibrium.

Population	Locus	OM-01	OM-03	OM-04	OM-09	OMO043	OMO100	OMO248	OMO093	OMO114	OMO129	OMO161	OMO219	OMO229	OMO337	OMO391	OMO392	OMO397
Lake Tanganyika	N	23	26	-	26	26	26	24	19	24	25	17	19	25	25	22	16	25
	N alleles	13	6	-	8	4	6	3	3	6	4	5	4	4	3	4	4	7
	HO	0.478	0.615	-	0.615	0.462	0.692	0.500	0.368	0.792	0.600	0.529	0.316	0.640	0.120	0.273	0.750	0.640
	HE	0.848	0.673	-	0.809	0.482	0.732	0.401	0.681	0.650	0.541	0.683	0.286	0.644	0.256	0.253	0.554	0.788
	P	< 0.001	0.073	-	0.202	0.873	0.743	0.725	0.004	0.451	0.180	0.009	1.000	0.587	0.015	1.000	0.345	0.069
Mwamapuli	N	20	20	-	20	20	20	20	18	20	20	19	19	20	-	19	19	20
	N alleles	8	5	-	8	5	4	5	3	4	3	2	6	5	-	4	3	4
	HO	0.550	0.500	-	0.650	0.550	0.700	0.650	0.500	0.700	0.500	0.263	0.842	0.650	-	0.632	0.316	0.300
	HE	0.788	0.729	-	0.794	0.553	0.694	0.685	0.624	0.724	0.627	0.422	0.770	0.719	-	0.636	0.522	0.350
	P	0.034	0.076	-	0.112	0.478	0.607	0.706	0.182	0.397	0.103	0.125	0.836	0.095	-	0.926	0.098	0.575
Kivulini	N	9	9	9	9	9	9	9	10	10	10	10	10	10	-	10	10	10
	N alleles	5	5	2	4	3	6	4	2	4	4	3	2	4	-	3	4	6
	HO	0.333	0.778	0.111	0.778	0.444	1.000	0.556	0.800	0.900	0.800	0.600	0.300	0.600	-	0.200	0.500	0.900
	HE	0.778	0.680	0.111	0.575	0.451	0.837	0.471	0.505	0.684	0.595	0.584	0.395	0.489	-	0.195	0.489	0.832
	P	0.005	0.117	1.000	0.762	0.250	0.628	1.000	0.173	0.545	0.581	0.449	0.480	1.000	-	1.000	0.446	0.321
Kerenge	N	28	30	28	30	30	30	30	30	30	30	30	30	30	30	30	30	30
	N alleles	12	10	4	8	5	7	3	3	6	6	3	3	7	2	6	4	7
	HO	0.643	0.700	0.179	0.767	0.367	0.767	0.667	0.633	0.867	0.533	0.633	0.567	0.667	0.267	0.800	0.500	0.567
	HE	0.904	0.764	0.424	0.602	0.328	0.762	0.621	0.671	0.781	0.686	0.660	0.635	0.676	0.325	0.773	0.580	0.802
	P	0.001	0.772	< 0.001	0.967	1.000	0.055	0.177	1.000	0.369	0.010	0.103	0.534	0.897	0.305	0.040	0.346	0.028

Supplementary Information Table 1 continued

Population	Locus	OM-01	OM-03	OM-04	OM-09	OMO043	OMO100	OMO248	OMO093	OMO114	OMO129	OMO161	OMO219	OMO229	OMO337	OMO391	OMO392	OMO397
Lake Kumba	N	62	63	64	63	63	63	63	60	64	64	62	57	64	64	64	61	64
	N alleles	5	4	3	5	2	5	4	3	4	3	2	4	4	2	4	3	5
	HO	0.500	0.714	0.047	0.921	0.016	0.206	0.397	0.583	0.609	0.297	0.532	0.544	0.281	0.281	0.531	0.541	0.563
	HE	0.544	0.682	0.046	0.592	0.016	0.230	0.427	0.618	0.561	0.374	0.504	0.499	0.315	0.496	0.589	0.505	0.622
	P	0.020	0.848	1.000	< 0.001	1.000	0.162	0.763	0.116	0.586	0.136	0.799	0.708	0.002	0.001	0.712	0.852	0.427
Nyumba-ya-	N	3	5	4	5	5	5	5	-	5	5	5	5	5	-	5	5	5
Mungu	N alleles	4	4	2	4	4	3	5	-	3	3	3	3	5	-	4	3	6
	HO	0.333	0.600	0.500	1.000	0.400	0.600	0.800	-	1.000	0.600	0.800	0.000	1.000	-	1.000	0.400	0.800
	HE	0.867	0.778	0.429	0.711	0.778	0.644	0.756	-	0.644	0.511	0.644	0.622	0.844	-	0.733	0.733	0.778
	P	0.067	0.693	1.000	0.428	0.048	1.000	0.487	-	0.173	1.000	0.619	0.016	0.846	-	0.387	0.544	0.872
Pangani Falls	N	14	13	9	14	14	14	14	14	14	13	14	14	14	14	14	14	14
	N alleles	10	6	6	6	5	4	4	3	6	7	4	4	8	2	5	5	7
	HO	0.429	0.692	0.444	0.929	0.429	0.500	0.429	0.500	0.571	0.769	0.571	0.500	0.786	0.214	0.643	0.643	0.714
	HE	0.902	0.806	0.719	0.688	0.479	0.730	0.611	0.574	0.828	0.855	0.696	0.696	0.849	0.389	0.746	0.675	0.870
	P	< 0.001	0.207	0.023	0.651	0.221	0.030	0.276	0.533	0.048	0.052	0.830	0.421	0.187	0.142	0.350	0.810	0.230
Lake Victoria	N	18	19	10	18	6	-	20	-	19	18	-	20	20	-	20	19	20
	N alleles	11	8	2	8	3	-	4	-	4	3	-	4	5	-	2	3	6
	HO	0.556	0.789	0.000	0.500	0.500	-	0.250	-	0.684	0.444	-	0.500	0.750	-	0.200	0.632	0.650
	HE	0.867	0.815	0.189	0.784	0.621	-	0.233	-	0.698	0.532	-	0.596	0.626	-	0.185	0.496	0.673
	P	0.001	0.321	0.053	0.003	0.655	-	1.000	-	0.419	0.786	-	0.568	0.968	-	1.000	0.421	0.330

Supplementary Information Table 2 *Post-hoc* tests of differences in genetic diversity (Rarefied Allelic Richness) among populations.

Population pair	P
Kerenge - Kivulini	0.2222
Kerenge – Kumba	0.0001
Kerenge – Mwamipuli	0.7768
Kerenge – Nyumba ya Mungu	0.9916
Kerenge – Pangani falls	0.5909
Kerenge – Lake Tanganyika	0.8648
Kerenge – Lake Victoria	0.6729
Kivulini – Kumba	0.2296
Kivulini – Mwamipuli	0.9882
Kivulini - Nyumba ya Mungu	0.7478
Kivulini – Pangani falls	0.0009
Kivulini – Lake Tanganyika	0.9667
Kivulini – Lake Victoria	0.9993
Kumba – Mwamipuli	0.0301
Kumba - Nyumba ya Mungu	0.0028
Kumba – Pangani falls	< 0.0001
Kumba – Lake Tanganyika	0.01790
Kumba – Lake Victoria	0.1039
Mwamipuli - Nyumba ya Mungu	0.9962
Mwamipuli – Pangani falls	0.0229
Mwamipuli – Laek Tanganyika	1.0000
Mwamipuli – Lake Victoria	1.0000
Nyumba ya Mungu – Pangani falls	0.1560
Nyumba ya Mungu – Lake Tanganyika	0.9993
Nyumba ya Mungu – Lake Victoria	0.9816
Pangani falls – Lake Tanganyika	0.0380
Pangani falls– Lake Victoria	0.0192
Lake Tanganyika – Lake Victoria	0.9999