

Molecular genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East African natural and stocked populations

Papius Dias Tibihika (✉ papiust@yahoo.com)

Universität für Bodenkultur Wien <https://orcid.org/0000-0001-6918-9903>

Manuel Curto

Universität für Bodenkultur Wien Department für Integrative Biologie und Biodiversitätsforschung

Esayas Alemayehu

Universität für Bodenkultur Wien Department Wasser-Atmosphäre-Umwelt

Herwig Waidbacher

Universität für Bodenkultur Wien Department Wasser-Atmosphäre-Umwelt

Charles Maseembe

Makerere University

Peter Akoll

Makerere University

Harald Meimberg

Universität für Bodenkultur Wien Department für Integrative Biologie und Biodiversitätsforschung

Research article

Keywords: cichlids, fish translocations, genetic structure, gene flow, bottleneck

Posted Date: January 3rd, 2020

DOI: <https://doi.org/10.21203/rs.2.11849/v4>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 30th, 2020. See the published version at <https://doi.org/10.1186/s12862-020-1583-0>.

Abstract

Background The need for enhancing the productivity of fisheries in Africa triggered the introduction of non-native fish, causing dramatic changes to local species. In East Africa, the extensive translocation of Nile tilapia (*Oreochromis niloticus*) is one of the major factors in this respect. Using 40 microsatellite loci with SSR-GBS techniques, we amplified a total of 664 individuals to investigate the genetic structure of *O. niloticus* from East Africa in comparison to Ethiopian and Burkina Faso populations. **Results** All three African regions were characterized by independent gene-pools, however, the Ethiopian population from lake Tana showed to be more divergent than expected suggesting that it might be a different species. In East Africa, the genetic structure was congruent with both geographical location and anthropogenic activities. *O. niloticus* from Lake Turkana (Kenya) was isolated, while in Uganda, despite populations being rather similar to each other, two main natural catchments were able to be defined. We show that these two groups contributed to the gene-pool of different non-native populations. Moreover, admixture and possible hybridization with other tilapiine species may have contributed to the genetic divergence found in some populations such as Lake Victoria. We detected other factors that might be affecting Nile tilapia genetic variation. For example, most of the populations have gone through a reduction of genetic diversity, which can be a consequence of bottleneck caused by overfishing, genetic erosion due to fragmentation or founder effect resulting from stocking activities. **Conclusions** The anthropogenic activities particularly in the East African *O. niloticus* translocations, promoted admixture and contact with the native congenics which may contribute to outbreeding depression and hence compromising the sustainability of the species in the region.

Background

Nile tilapia, *Oreochromis niloticus*, is native to the Levant and African freshwater systems e.g., in the Western part of the continent (e.g., Senegal, Gambia, Niger, Benue, Chad) as well as to many of the East African rivers (e.g. R. Nile) and Rift Valley Lakes like, Albert, Turkana, George, Edward, Tanganyika, Kivu, etc. (1, 2). Although *O. niloticus* is native to Africa, the cichlid is naturally absent in the world's largest tropical freshwater body, Lake Victoria and the neighboring Lakes Kyoga and Nabugabo as well as many of the East African satellite lakes (2–5). These lakes were naturally inhabited by two tilapiine species; *O. variabilis* (Nyasalapia) and *O. esculentus* (Ngege) (2–4). For more than nine decades, *O. niloticus* has been intentionally dispersed worldwide, in particular for aquaculture and restocking programs (2, 6). In East Africa, various fish introductions are reported, starting in the 1920s. For example, *O. niloticus*, and other tilapiines e.g. Athi River Tilapia (*Tilapia spilurus nigra*, Günther 1894) as well as Black bass (*Micropterus salmoides*), being initially translocated for enhancing fisheries productivity in water bodies naturally considered as unproductive like the southwestern Uganda high-altitude lakes (7, 8). A case in point is Lake Bunyonyi which was stocked in the 1920s with individuals of *O. niloticus* from Lake Edward (7). Similarly, in the 1950s, several tilapiine species were stocked into Lakes Victoria, Nabugabo and Kyoga to counteract the decline of native fish species (*O. variabilis* and *O. esculentus*) (3–5, 9). The introduced species; *O. niloticus*, *O. leucosticus* (Blue-spotted tilapia), *Coptodon zillii* (Red-belly tilapia) and *O. melanopleura*, were all suspected to originate from Lake Albert (3, 4, 9, 10). However, some introductions might have also

originated from Lake Edward and Lake Turkana into Lake Victoria basin (2, 4, 8). Following these introductions, the indigenous fish species in Lakes Victoria, Kyoga, and Nabugabo, significantly declined in the 1980s, coinciding with the dramatic increase in the stock size of the non-native *O. niloticus* (3, 4, 11). The potential reasons for the declined native fish species (*O. variabilis* and *O. esculentus*) were suspected to a combination of factors including; competition, over fishing, as well as predation pressures from another introduced species, the Nile perch (*Lates niloticus*) (3, 5). But one additional factor might have been hybridization between the native cichlids (*O. variabilis* and *O. esculentus*) and the introduced tilapiines, particularly *O. niloticus* (4, 9, 11). Based on these events, the expanded distribution of *O. niloticus* in East Africa complicates the differentiation and identification of genetic units for management and conservation. For example, the population considered as non-native *O. niloticus* in Lakes Victoria and Kyoga might have genetically diverged via admixture and hybridization with the indigenous species (4, 9, 12). The loss of indigenous *O. mossambicus* due to hybridization with the introduced *O. niloticus* has been reported in South Africa (13). The situation in East Africa may have worsened with the recent boom of fish hatcheries and aquaculture production systems (14). In this context, feral populations resulting from escapees might be an additional and serious threat to natural systems.

In nearly the last two decades, the East African countries have been developing measures for fisheries sustainable exploitation through the implementation of co-management strategies (15). Nonetheless, conservation and management of the already admixed species might not be achieved if the genetic structure of the species in question is not well understood, as the stocks are difficult to define (16). Therefore, with respect to the East African *O. niloticus*, as the species were potentially affected by various anthropogenic activities, a thorough characterization of the populations at the molecular level might be needed.

Based on the earliest studies, East African *O. niloticus* diversity has been studied using both traditional morphometric methods and molecular markers, which led to contradictory patterns in the species description. For example, using biometrics and counts, seven *O. niloticus* subspecies from African different regions or lakes were described (2). Nevertheless, this classification was contradicted by subsequent studies using morphometric analyses accompanied by allozyme markers, which indicated that the *O. niloticus* strain from Lake Edward is closely related to that of the lower Nile (Egypt) (17). Also, other investigations using restriction endonuclease mitochondrial DNA found that *O. niloticus* from Lake Tana is distinct, contrary to the earlier traditional morphometric and meristic analyses (18). Other earlier molecular genetic studies employing allozymes and restriction fragment length polymorphism (RFLP) of mitochondrial DNA (mtDNA), as well as randomly amplified polymorphic DNA (RAPD) for investigating the demography of *O. niloticus* populations in East Africa, shed some additional light to these incongruences (1, 19, 20). Some of these studies reported that *O. niloticus* populations from Albert Nile (the Egyptian stretch of River Nile) are distinct from the West African populations, also contrary to earlier morphometric studies (1). Furthermore, these past investigations based on traditional markers indicated conflicting results amongst. For example, findings from a combination of allozymes and restriction fragment length polymorphism (RFLP) of mtDNA indicated that *O. niloticus* from Lake Tana is clustered with Lake Edward and the Kenyan Lake Turkana system, which differs from the findings based on restriction endonuclease

analysis of mtDNA (1, 18). These results are inconsistent probably because of the different methodological approaches used that comprise different information content (21–26). Additionally, the markers used so far have low resolving power to characterize variation within and between populations, and the genetic fingerprinting markers like RAPD cannot discern between homozygotes and heterozygotes (22). The lack of methodologies with high discriminating power in the past studies, therefore, suggests that the genetic structure patterns of the East African *O. niloticus* are insufficiently documented.

In the present study, we utilize nuclear microsatellite markers, simple sequence repeats (SSRs), to typify the *O. niloticus* in East Africa using next-generation sequencing. SSR loci have been proven robust when investigating the genetic structure of *O. niloticus*, particularly, using SSR genotyping by sequencing (SSR-GBS) (27). SSR-GBS approaches are useful because they reduce size homoplasy, which is one of the constraints of traditional SSR fragment length analysis (28, 29). However, SSR-GBS is not without drawbacks (30). For example, the presence of stutter complicates allele calling for di-nucleotides, null alleles due to mutation on primer binding sites, and it does not recover genomic information hence overestimating events that had a small impact on the gene-pool. Although generally, the use of SSR fragment length analysis can yield information for delineating populations, the recent studies in East Africa that have used this approach on *O. niloticus* were limited to few water bodies in Kenya, with the broader scope of the African Great Lakes missing (31–33). It is important to conduct a comparative study of various water bodies where *O. niloticus* is present (native and non-native with possible admixture). Such research would provide information on the genetic structure and diversity which would establish a firm base for management and conservation of these cichlids (34).

Here, we explicitly investigate the genetic structure of *O. niloticus*, in East Africa including some populations from Ethiopia and West Africa (Burkina Faso), representing the Sub-Saharan African Great Lakes. We compare natural/native with introduced/non-native *O. niloticus* populations, including other populations from aquaculture systems. With this approach, we investigate the impact of anthropogenic activities, particularly the translocations, on the *O. niloticus*' gene pool. This is especially important to evaluate the genetic integrity of native stocks. We hypothesized that anthropogenic activities have affected the genetic divergence of *O. niloticus* populations, particularly in environments where the species was introduced. We also predict that the geographical context exhibited by aquatic interconnectivity may influence the genetic homogeneity of cichlid in such environments. We test these hypotheses by answering the following research questions: 1) Does the genetic structure of the East African *O. niloticus* populations differ from those outside the region? 2) To what extent does the genetic structure of the East African *O. niloticus* populations reflect the geography and anthropogenic activities associated with the pathways of the translocation?

Results

Variability of SSR loci

In total, 13,530,228 paired reads were produced for genotyping, from which 9,579,578 passed the quality control steps, which were later used for allele calling. Genetic variation results for the 40 SSR loci are presented in the Additional file, Table S2. The number of alleles per locus had a mean value of 33.8 ± 20.5 , ranging from seven to 84, with a total of 1,352 alleles generated across all loci. Overall, 80% of the loci exhibited expected heterozygosity (H_e) values greater than 0.5. Polymorphic Information Content (PIC) was generally congruent with H_e , with 78% of loci indicating values of greater than 0.5 (additional file, Table S2).

Genetic structure

The UPGMA dendrogram showed that all East African populations were more similar to each other than to the other regions (Figure 2). In this case, the three Ethiopian populations (Hashenge, Ziway, and Chamo) formed the most distant group followed by Burkina Faso and the other Ethiopian water body, Lake Tana. Among the East African natives, the largest separation was between the Kenyan, Lake Turkana, and the Ugandan water bodies. In Uganda, with exception of Lake Victoria, the non-native lakes and fish farms grouped with a native population: the southern Ugandan high-altitude Lakes (Kayumbu and Mulehe) with a group comprised by Lakes George, Edward and Kazinga Chanel; Lake Kyoga populations- and Sindi Farm with River Nile; and Bagena and Rwitabingi farms with Albert. Four subpopulations of Lake Victoria (Gaba, Masese, Kakyanga, Kamuwunga) formed a sister group to the River Nile one. The Lake Victoria subpopulation Sango Bay showed the highest degree of divergence in Uganda.

Neighbor network results showed a similar pattern to the UPGMA dendrogram both at regional and local levels (Figure 3). In this case, however, Burkina Faso was observed to be closer to the Ugandan populations. In general, network results reflected two Ugandan catchment groups: the George, Kazinga Channel, and Edward group together with the non-native Ugandan highland lakes, and on the other end, Albert and River Nile systems together with the non-native Lake Kyoga and all fish farms. Interestingly Lake Victoria exhibited an intermediate position between both groups with the subpopulation from Sango Bay showing a long branch, suggesting high genetic differentiation. Overall, most of the non-native populations (including farms) showed longer branches than the natives (Figure 3).

Genetic distance between individuals which was visualized through principal coordinates analysis (PCoA), analysis showed a separation of population groups based on geographic regions (Figure 4a). Samples formed four groups when analyzed at the regional/country level (Figure 4a): two groups with individuals from Ethiopia, one with individuals from East Africa, and another intermediate group with samples from both regions. The composition of these groups was clearer when the distance between the native individuals was plotted (Figure 4b). At this level, Lake Turkana clustered with Burkina Faso, and a division between the three Ethiopian Lakes (Hashenge, Chamo, and Ziway) and Lake Tana was clearly observed. Amongst East African populations, the separation between Lake Turkana and the remaining native populations was evident (Figure 4b). Individuals found in the Ugandan native populations were divided into two main groups (Figure 5a). One group was composed of Lake Albert and River Nile individuals while the other by Lake Edward, Kazinga Channel, and Lake George. This division was less evident when

individuals from non-native and fish farm populations were included in the analysis (Figure 5b). Here, some individuals from Sango Bay formed a separate group from the remaining Ugandan individuals. A further group composed of Lake Hashenge individuals was found when only Ethiopian individuals were plotted (Figure 5c). Substructure within the same lake was only evident for Lakes Victoria and Kyoga (Figure 6).

The Bayesian analysis with STRUCTURE was portrayed based on the optimal K values. For all populations, the best K was 10, all native populations, K = 7, East African native populations, K = 2, Ugandan native populations, K = 2, and all Ugandan populations including farms, K = 4 (additional file, Figure S2). *O. niloticus* populations from each African region were assigned to different groups (Figure 7a). Within each region, the same assignments were observed with Lakes Tana and Turkana isolated from the rest of Ethiopians and East African populations, respectively (Figure 7a). Among the Ugandan native populations, clustering was also congruent with the two water systems, as indicated earlier by both network and PCoA analyses, see Figures 7b and 7c. However, there were cases where the non-native populations showed independent clusters from the native. For example, in all analyses, Lake Victoria clusters differed from other populations even when only Ugandan *O. niloticus* were included in the analysis (Figure 7c). Apparently, admixture was more evident amongst the East African populations but mostly detected when only non-native populations were considered (Figure 7c).

Gene flow between population

Results from recent migration rates estimated with BayesAss indicated that Lakes Kyoga and George were the main sources of migration (Figure 8), with values for other populations generally falling below (<2%). Noticeable gene flow was from Lakes Kyoga to Victoria and George to Edward (27%), Kyoga to Albert (25%), Kyoga to Bagena farm (23%), Kyoga to Sindi farm, River Nile and Rwitabingi farm (22%), George to Kazinga Channel (21%) and finally George to Mulehe (20.4 %) (Figure 8). Migration rates estimated through Genalex were congruent with BayesAss, but with the difference that the *O. niloticus* population from Lake Victoria was also a source of migrants (Additional file, Table S3).

Genetic differentiation, diversity, and isolation by distance

Genetic differentiation of *O. niloticus* was consistent with the STRUCTURE results. For instance, the F_{st} values clearly demonstrated that the East African *O. niloticus* populations are genetically distant from the Ethiopian and West African populations (Figure 9a). Despite *O. niloticus* populations from River Nile and Lake Kyoga showing relatively high F_{st} values, results from the East African populations generally showed low genetic differentiation. Also, the East African *O. niloticus* populations were genetically more diverse when compared to either Ethiopian or Burkina Faso (Figures 9b-d). Based on all statistics, the non-native Lake Victoria and native Lake Turkana *O. niloticus* populations were the most genetically diverse. On the other hand, Lake Kyoga and River Nile *O. niloticus* populations were consistently the least diverse even when investigated at the subpopulation level (additional file, Figure S4).

Results from the Garza-Williamson index (G-W), generally indicated that nearly all of the studied populations went through a bottleneck, apart from the Ethiopian Lake Tana (Figure 10a). In the analysis, only Lake Tana exhibited G-W values >0.5 (0.56 ± 0.44). Regarding population genetic diversity, however, Lakes Victoria and Turkana showed the highest number of private alleles (Figure 10b).

When we partitioned Lake Victoria to assess the genetic diversity patterns within the water body, generally one sub-population was distinguished from the others (Figure 11). Sango Bay, in particular, was isolated based on F_{st} values, and consistently exhibited higher genetic diversity indices (N_a , H_e and A_r) (Figure 11).

Mantel tests for isolation by distance (IBD) across all samples showed a positive correlation between geographical and genetic distance ($R^2 = 0.30$) (Figure 12a). However, the strong correlation ($R^2 = 0.67$) between the populations was only found when Burkina Faso was excluded from the analysis (Figure 12b). The genetic differentiation between the East African and the Ethiopian populations appears to inflate this correlation. Similarly, a strong IBD was also found amongst East African populations (Figure 12c), which was not the case when only Ugandan populations (excluding Turkana) were considered (Figure 12d).

Discussion

Fisheries and fishery products are vital in the developing world but heavily threatened through various anthropogenic activities which may compromise the continuity of the resources (35). One aspect of the anthropogenic threats is the change or alteration of the natural genetic structure of fish stocks through admixture (36, 37). Understanding the admixture of stocks is only possible if the source populations can be differentiated using genetic markers. We show the importance of SSR-GBS for a deeper understanding of population dynamics, in particular, the East African *O. niloticus*, towards the alignment of management and conservation strategies. In this study, we investigated the phylogeographical patterns and we found large differences between lakes (e. g. Lake Tana) and also differences between natural water catchments that allow populations to be identified. Here, we discuss the current state of *O. niloticus* in reference to phylogeographical patterns and anthropogenic activities.

Phylogeography of East African *O. niloticus*

In all analyses, we found a clear differentiation among all three African regions included in this study (East Africa, Burkina Faso, and Ethiopia), indicating a low degree of connectivity amidst them and highlighting the high level of differentiation between regions. Lake Tana was completely distinct from the remaining populations. This applies not only to the Ethiopian populations but also to the East African ones. So, the genetic distance in Ethiopia is higher than between the East African and West Africa populations, indicating a divergence higher than we would expect within a species. These results are consistent with previous genetic reports (18), but not the findings of the subspecies treatment based on the traditional morphometric and meristics (2). This high level of differentiation argues for a revision of the species delimitation for these populations.

Lake Tana lies in the Ethiopian mountains and is isolated from the Lakes in the Rift valley (38). This might explain the high degree of differentiation of this lake because of the lack of connectivity and divergent ecological conditions. Contrary, Lake Hashenge which is also in the Ethiopian mountains is related to the Rift Valley lakes. Lake Hashenge is reported to have been stocked with *O. niloticus* following mass mortalities of the native species (39). The native status of this lake is unclear since it could have been restocked with *O. niloticus* that originated from the Rift Valley Lakes. Besides that, we see a slight differentiation in PCoA between Lake Hashenge and the Rift Valley Lakes in Ethiopia, which may reflect an unsampled source of stocking or differentiation accumulated because of the high degree of isolation of the lake.

In East Africa, genetic structure reflected different catchments. The population from Lake Turkana was genetically distinct from the Ugandan populations which is expected given its high geographical isolation (40). Our findings concur with the previous works that treated the Turkana population as a different subspecies (*O. vulcani*) (2). The high diversity and number of private alleles found in Lake Turkana can be a consequence of this isolation. The East African arid, Lake Turkana, naturally is also characterized by a remarkable genetic diversity. One factor might be introgression perhaps from anthropogenic activities or influx of gene flow from River Omo (Ethiopia). However, this is not clear and a better sampling from the region needs to be included to evaluate the extent of the observed current genetic structure of the population.

In Uganda, despite the high degree of connectivity and proximity between the water bodies, *O. niloticus* populations were clearly structured. These reflected three main groups: 1) (Lakes George and Edward, as well as Kazinga Channel, 2) Lake Albert, River Nile, and Kyoga and 3) Lake Victoria system. The 2nd and 3rd groups are discussed in more detail under anthropogenic activities subsection. The 1st group, Lakes George and Edward are connected via the Kazinga channel which also explains the high natural migration rates between these populations. The different genetic structure between the western Rift Valley Lakes (Edward-George-Kazinga Channel and Albert) was conserved despite being connected through River Semliki that flows from Lake Edward and Albert (41). The strong rapids and falls present in this river (41, 42), might constitute a strong barrier to gene flow, which maintains these systems apart. These findings are congruent with recent work on *O. niloticus* geometric morphometrics (43) but do not concur with past studies (2, 20). This incongruity might be associated with different methodological approaches utilized between the earliest studies and the current one. For example, using morphometric and meristic methods, *O. niloticus* from the Edward-George system and Albert was treated as one subspecies; *O. niloticus eduardianus* (2). However, inference from traditional morphometrics are weak due to the lack of informative characters (18). Similarly, while we used SSR-GBS techniques, (20) employed random amplified polymorphic DNA (RAPD) markers, which due to their dominance genotypic nature, provide only part of the information content (22).

Anthropogenic activities-fish translocations

In East Africa, we know that *O. niloticus* was introduced into several water bodies through stocking activities. We were able to genetically track these translocation events to both non-native water bodies and fish farms. All genetic structure analyses and migration rates showed that the two Ugandan groups (the George-Edward complex and Lake Albert) contributed to the stocking of different water bodies. *O. niloticus* from the southwestern Ugandan high-altitude Lakes; Mulehe and Kayumbu, originated from the Western Rift Valley Lakes—Edward and George. For the 2nd group, Lake Kyoga and River Nile (Victoria Nile) are genetically similar to Lake Albert, suggesting that, the latter population might have contributed genes to the gene-pool of the former systems. Although Lake Kyoga is connected to Lake Albert via River Nile, their genetic similarity is unlikely related to the consequence of natural migration via water flow. The main reason here is the natural occurrence of Murchison Falls on the River Nile that acts as a barrier between the systems (3, 41). For this matter, the genetic similarity between River Nile, Lakes Kyoga, and Albert populations may have resulted in stocking regimes using the latter as source (3).

Fish farms seem to have sourced fish seed from multiple populations, resulting in admixed stocks. Our results show that Lakes Albert, and Kyoga, as well as River Nile, contributed to the gene pool of the farmed populations (Figures 3, 8 & 7c). Based on genetic distance, Lake Albert was the main contributor to Rwitabingi and Bagena farms while Kyoga to Sindi farm. However, we also observed a high amount of gene flow from Kyoga to Rwitabingi and all these farms appeared to be admixed with other populations including Lake Victoria. Apart from farms, evidence of admixture was probable in the East African natural populations, which seems to have been promoted by anthropogenic activities (3, 9). This is supported by the fact that when non-native populations were unconsidered in the STRUCTURE and PCoA analyses, signals of admixture were minimal, and clear genetic structure assignments could be observed. In East African, admixture in *O. niloticus* populations may stem from three main processes: 1) translocation from multiple sources into the non-native water bodies, 2) back translocation from non-native to native populations, and 3) hybridization of *O. niloticus* with congeneric species promoted by translocations.

The first and third processes may explain partly the genetic variation found in the 3rd group; Lake Victoria (see above the three Ugandan groups). Although *O. niloticus* in Lake Victoria is generally isolated, based on the distance neighbor Network tree (Figure 3), the population occupied an intermediate position between the above described; 1st and 2nd, Ugandan groups. Thus, it is clearly possible that multiple stockings might have contributed to the gene-pool indicated by the Lake Victoria population. For example, (2) suggests that introductions into Lake Victoria may have originated from Lake Edward, with other authors suggesting multiple sources (4, 5, 12, 44), which support our results. The highly diverse and differentiated gene-pool in Lake Victoria could have originated from the admixture of several lineages due to multiple sources.

On the other hand, possible hybridization of the introduced *O. niloticus* with the indigenous relative species (*O. variabilis* and *O. esculentus*) in Lake Victoria may explain some of the genetic variation patterns found in this lake. First, this lake together with Turkana showed values of private alleles up to four times higher than the remain populations. This genetic variation could have originated from introgression by species that have not been included in the analysis. Similarly, the probable hybridization may explain the high

genetic diversity and divergent gene-pool detected in the system. Within Lake Victoria, the Sango Bay subpopulation appears to be an extreme case from this by showing the highest degree of genetic divergence. Remarkable genetic differentiation in Sango Bay was noticed only when compared with the remaining subpopulations within the lake, but also with the other East African populations. In this case, during the boom of the *O. niloticus* population in Lake Victoria (3–5, 45), a larger portion of the native species' genetic materials may have been introduced into *O. niloticus* gene-pool. This is just a hypothesis since, in this study, we cannot directly test for hybridization because we did not include samples of *O. niloticus* congenics. However, hybridization involving *O. niloticus* and other tilapiines has been reported to be relatively frequent and it needs to be considered (9, 33, 46, 47).

If admixture/hybridization shaped the gene-pool of Lake Victoria, it may have adaptive consequences and compromise the sustainability of *O. niloticus*. Although hybridization may lead to heterosis/hybrid vigor (48, 49), admixture is usually reported to have negative consequences (37, 50). Introgression can contribute to outbreeding depression either by the introduction of maladaptive alleles or through the dilution of alleles important for local adaptation (51). In more drastic scenarios, hybridization can result in genomic incompatibilities contributing to a fast reduction of population fitness (51). Alternatively, the hybrids may potentially exhibit more fitness and subsequently extirpate the parental lines (46). The observed genetic structure of *O. niloticus* populations in Lake Victoria was unexpected and has not been reported before, which calls for further investigations for taxonomic recognition.

Evidence for the second process of admixture was only found in Lake Albert. In the structure analysis, this population showed admixture with Lake Kyoga. We also found significant migrations from Lake Kyoga to Lake Albert. These results indicated that admixture with respect to translocations not only contributes to non-native populations but also to native ones. The sequence of gene flow from Lake Kyoga to Albert is not clear as none of the previous reports have indicated this. However, it is likely that aquaculture activities might be contributing to the observed gene flow between Lakes Kyoga and Albert.

Anthropogenic activities-Consequences of overfishing

Some water bodies, especially Lake Kyoga and River Nile showed low genetic variability and evidence of bottleneck with respect to G-W estimations. Given the recent stocking of these water bodies, this pattern may be explained by the founder effects. Nevertheless other anthropogenic activities need to be considered as well. High loss of genetic diversity among populations, particularly, in fishes has been attributed to over-exploitation (52). This might be the case for the L. Kyoga population. For example, although *O. niloticus* boosted the capture fisheries in the Lake Victoria basin (Lakes Victoria and Kyoga) following introductions, the species was subsequently overexploited between the 1970s and 80s (53, 54). This was reflected in the dramatic decline of the stock sizes and increased fecundity, which are clear indicators of overfishing (54). The low diversity in River Nile could be linked to low gene-flow connectivity with other water bodies due to hydro-electric power dams that have been constructed along the river (the upper Nile of the Ugandan side), which increases the effect of genetic drift. However, this needs to be

assessed in further analyses, especially when additional samples are collected in sections of the lower Nile (below Murchison falls), where apparently there are no dams.

Implications for management and outlook

Overall, we found evidence that anthropogenic activities affected the gene-pool of the East African *O. niloticus*. The main consequence might have been admixture and potentially hybridization between different stocks and species respectively. In the long term, this may have negative effects on population fitness due to outbreeding depression and genetic swamping. Thus, management measures should inhibit any form of unauthorized spread of fish in the aquatic ecosystems. The Western or Albertine Rift Valley lakes (Edward-George) may be ideal broodstock sources for subsequent breeding programs and aquaculture, as these systems seem not to be admixed. To avoid an influx of feral populations, a proper environmental impact assessment should be prioritized before implementation. Genetic diversity might also have been affected by overfishing and the construction of hydropower dams, which should also be taken into consideration in future management options.

Conclusion

Our results were congruent with the hypothesis that anthropogenic activities affected the genetic structure of *O. niloticus* populations in East Africa. The genetic variation of some populations, especially from Lake Victoria, corresponded with possible hybridization of *O. niloticus* with native congeneric species, which may have been mediated by anthropogenic activities. This study also contributed to the knowledge of *O. niloticus* phylogeography in East Africa. In this case, we found several new genetic groups such as the populations from Lake Tana, Victoria and the two natural catchments in Uganda. Some of these may require further taxonomic exploration. Additionally, we show that gene-flow among the East African *O. niloticus* populations was not entirely from native to non-native environments, but also from non-native to native environments likely through aquaculture and restocking programs. Moreover, this study shows the importance of molecular markers, in particular, the use of SSR-GBS in cataloging populations. Further studies should include *O. niloticus* samples from other regions such as the lower Nile (below Murchison Falls), Lake Kivu (Rwanda), Tanganyika and Baringo as well as the congenics for a more comprehensive picture.

Methods

Sampling/study areas

We collected *O. niloticus* specimens from three water body types: a) those where *O. niloticus* is native, b) where introduced, and c) from fish farms (Figure 1), following our earlier sampling design (43). Most samples were collected by local fishermen using gill nets set overnight. At Lake Turkana, a seine net was utilized. From Ethiopia and Burkina Faso, four and one native populations were sampled, respectively. Considering the large extent of Lake Victoria and multiple *O. niloticus* introductions into the world's largest

tropical freshwater body, we sampled five locations to assess possible genetic heterogeneity within the system (Figure 1). Similarly, in other relatively large lakes like Lake Edward, Kyoga, and Albert, we sampled two locations each for subsequent subpopulation analyses (Table 1). A total of 664 samples were collected from 18 water bodies during several field excursions in 2016. From every single fish, a muscle tissue sample (approx. 30 mg) was extracted from the dorsal region, preserved in absolute ethanol contained in 2 ml Eppendorf tubes and later stored in a freezer until genotyping at the Institute for Integrative Nature Conservation Research-University of Natural Resources and Applied Life Sciences Vienna (BOKU), Austria. Sampling was conducted in collaboration with respective authorities per region and therefore no special permission was required. In all cases, the fish were already dead when obtained from the fishermen, therefore no special treatment for the animals was administered in the process. As contamination of the specimens was not likely during sampling with gill nets, great care and attention were provided for during seining on Lake Turkana. The non-native and farm populations were only sampled in Uganda. Here, we refer to the non-native populations like those found in the high-altitude satellite lakes of south-western Uganda (Lakes Mulehe and Kayumbu) as well as in lower altitude lakes (Lake Victoria and Kyoga) (43). The three sampled fish farms include; Rwitabingi (located near River Nile and Lake Kyoga), Bagena and Sindi from South-western Uganda. The rest of the populations are regarded as native (Figure 1; Table 1).

Table 1 Details of the sampling sites and the total number of individuals collected per water body and location/site. The indigenous *O. niloticus* populations, are also herein referred to as natives and introduced, non-natives and farms are the pond culture systems

Lakes/River	Sample Site	Local name	No.	Country	Pop. nature	Coordinates		Elev. (m)
Albert	Ntoroko	Ngege	21	Uganda	Indigenous	N.01.05206 ⁰	E030.53464 ⁰	618
Albert	Kyehooro	Ngege	16	Uganda	Indigenous	N.01.50990 ⁰	E030.93610 ⁰	615
Edward	Rwenshama	Ngege	27	Uganda	Indigenous	S.00.40459 ⁰	E029.77283 ⁰	908
Edward	Kazinga	Ngege	22	Uganda	Indigenous	S.00.20783 ⁰	E029.89252 ⁰	914
George	Hamukungu	Ngege	35	Uganda	Indigenous	S.00.01739 ⁰	E030.08698 ⁰	916
Kazinga Ch	Katungulu	Ngege	21	Uganda	Indigenous	S.00.12541 ⁰	E030.04744 ⁰	915
R. Nile (VN)	Kibuye	Ngege	24	Uganda	Indigenous*	N.01.18734 ⁰	E032.96865 ⁰	1062
Kyoga	Kibuye	Ngege	44	Uganda	Introduced	N.01.40028 ⁰	E032.57949 ⁰	1034
Kyoga	Bukungu	Ngege	22	Uganda	Introduced	N.01.43873 ⁰	E032.86809 ⁰	1045
Victoria	Kakyanga	Ngege	30	Uganda	Introduced	N.00.18079 ⁰	E032.29332 ⁰	1136
Victoria	Gaba	Ngege	26	Uganda	Introduced	N.00.25819 ⁰	E032.63727 ⁰	1146
Victoria	Masese	Ngege	23	Uganda	Introduced	N.00.43650 ⁰	E033.24081 ⁰	1136
Victoria	Sango Bay	Ngege	25	Uganda	Introduced	N.00.86772 ⁰	E031.71332 ⁰	1129
Victoria	Kamuwunga	Ngege	25	Uganda	Introduced	S.00.12747 ⁰	E031.93999 ⁰	1139
Mulehe	kisoro	Ngege	25	Uganda	Introduced	S.01.21345 ⁰	E029.72668 ⁰	1801
Kayumbu	Kisoro	Ngege	30	Uganda	Introduced	S.01.34679 ⁰	E029.78446 ⁰	1901
Rwitabingi	Kamuli	Ngege	29	Uganda	Fish farm	N.00.97116 ⁰	E033.13924 ⁰	1069
Bagen	Kisoro	Ngege	31	Uganda	Fish farm	S.01.25617 ⁰	E029.73622 ⁰	1857
Sindi	Kabale	Ngege	25	Uganda	Fish farm	S.01.17578 ⁰	E030.06198 ⁰	1733
Turkana	Longech	Ngege/Sato	35	Kenya	Indigenous	N.03.55617 ⁰	E035.91599 ⁰	364
Ziway	Ziway	Koroso	27	Ethiopia	Indigenous	N8.00730866	E38.8413922	1636
Tana	Tana	Koroso	32	Ethiopia	Indigenous	N12.0266003	E37.3036142	1831
Hashenge	Hashenge	Koroso	26	Ethiopia	Indigenous*	N12.5746028	E39.4966667	2443
Chamo	Chamo	Koroso	25	Ethiopia	Indigenous	N5.82128333	E37.5747222	1110
Loumbila	Loumbila	Tegr-pere	18	Burkina Fs	Indigenous*	12.5142528"N	01.3972222"w	276

Kazinga Ch=Kazinga Channel, VN=Victoria Nile, Burkina Fs=Burkina Faso, and Population nature with the asterisk symbol (*) implies that the population might not be indigenous, No.=Number of samples, Pop. =Population, and Elev.=Elevation

Genotyping

Genomic DNA extraction was conducted using magnetic beads based on the MagSi-DNA Vegetal kit (MagnaMedics, Geleen, Netherlands) and a magnetic separator, SL-MagSep96 (Steinbrenner, Germany) (27, 30). We used microsatellite markers (27), to which we added 15 extra primers (Table 2, see also Table S3). The SSR primers were designed and tested following our earlier work (27), using the same shotgun sequencing data present in the sequence read archive database (SRA) under the reference number SRX3398501. Screened primers were then grouped into three multiplexes and used to prepare amplicon SSR-GBS libraries using the same approach and specifications of (27). The PCR products were then pooled and sent for paired-end 300bp sequencing in Illumina MiSeq, at the Genomics Service Unit in

Ludwig Maximilian Universität, München, Germany. The raw sequence data were deposited in the GenBank, sequence read archive database (SRA) under the project PRJNA550300 with the accession numbers, SRR9587388 to SRR9587270. Sequences generated by Illumina, were subsequently quality checked and controlled, which were later used for alleles calling as described in (27, 30) using the scripts from the SSR-GBS pipeline (<https://github.com/mcurto/SSR-GBS-pipeline>). The resulting codominant matrix and information for which sequences correspond to each allele can be found in the additional file (see the file named "Second_additional fileAllelesList & matrix_"). For subsequent analyses, all loci and samples with missing genotypes $\geq 50\%$ were excluded, leaving a total number of 40 markers (additional file, Tables S1, S3). Other studies have indicated that many SSR loci are not necessary in order to detect population structure (55, 56), so we did not see the need of developing additional markers to the 40 already in use.

Table 2 15 new primer pairs developed in the present study. The other 26 tested primers developed by (27) can be found in the additional file section, Table S1.

Locus	F: Primer sequence (5'-3')	R: Primer sequence (5'-3')	Repeat motif	Asr
Ti39	TACCTGCCAGTCATGTGCTG	TGCTCAGACTGGTCCCTTCT	(ATGG)8	368-420
Ti41	TCGCAGCTGCTCCTGTTTAA	TTGTGCACGTGGACATGTTG	(AAAC)11	381-471
Ti43	ATTGCCATCACCAGGAACCA	TGCTAGCCCAGAGCATTGGA	(GAATA)6	425-478
Ti44	TGCTCCTGACTCAGCATCAC	GCAGCACTCTGACATGAAGC	(GAAAA)6	419-469
Ti49	TCGAAGTAGCGTGAAAACCT	ACAACAACAACAGGTCGGGA	(TGT)8	395-403
Ti50	CCTGTGACAGACTGGTGACC	ACACTGATGCGGTTTACGGT	(ATGG)7	442-517
Ti51	TGCTAAACGCCAGCTGATGA	TTACCACACGATGTCGCAGG	(TGT)8	401-428
Ti52	GAGAAACGTCCAGTGGCAGA	TTTCGATCTGCTGCCCTTT	(TAT)8	373-429
Ti54	TTTCTTGCCAGCAAAAACAGT	CAGATTCTTCCAGTGCTTGTGC	(GGAT)7	390-480
Ti55	GAGCCCAGACAGCAGACAAT	AGGACCTTCTATGGCCCTGT	(TCTA)7	417-491
Ti56	TGCAGTGAATTTGGCACCTG	AGCCTGAGATACCTGTGCCT	(TGTT)6	310-462
Ti57	CAGTGGGAGGAAGCTCCAAA	GCTGCATGGATCCAATAGGC	(TCCA)7	400-444
Ti59	ATGGACTTAAGCTGCACCCC	TGAGCATTTGACCCCAGCAT	(AGGA)6	429-461
Ti60	GAGCCGCCATAGTGTCACCT	CCTGCTCTCACTCAAAGAGGG	(ATCC)7	473-516
Ti61	GCTACACAGGAAAGCAGAGC	ACTCAATGCTGGACGTGACC	(TGGA)6	474-501

F=Forward and R=Reverse, Asr=Allelic size range

Genetic Structure

Genetic structure was first assessed by calculating the genetic distance between individuals and then visualized through Principal Coordinate Analysis (PCoA), all conducted in GenAlex Version 6.5 (57). Genetic similarity between populations was evaluated by plotting a Neighbor-Net tree based on Nei's genetic distance (58), using the program, SplitsTree4 version, 4.14.8 (59). We also constructed UPGMA

dendrograms for making inferences on the hierarchical clustering using Nei's genetic distance as implemented in Populations-1.2.32 (60). Support values were estimated with 1000 bootstrap replicates based on loci resampling. Neighbor-Net tree and the UPGMA dendrogram were conducted with the inclusion of subpopulations, when applicable to evaluate possible substructure within the populations. Genetic structure was further investigated using the program, STRUCTURE Version 2.3.4 (61). STRUCTURE clusters individuals into hypothetical populations through optimization of Hardy-Weinberg equilibrium (62). STRUCTURE was run from $K = 1 - 35$ for 10,000 Markov chain Monte Carlo (MCMC) generations after a burn-in length of 10,000 generations (63), whereby each run was iterated 20 times. The program's default settings for the admixture model and allele frequencies correlated were implemented. Detection of optimal K was done with STRUCTURE HARVESTER (64) using the delta K (ΔK) statistic, which is the second-order rate of change ($\ln P(D)$) across successive K values (63, 65). In this context, STRUCTURE HARVESTER uses ΔK to identify the highest value and henceforth the best K . Results from multiple replicates were summarized using the online pipeline Clumpak program (66) available at <http://clumpak.tau.ac.il/>. Similar analyses were performed for Lake Victoria within populations.

Migration rates and number of migrants per generation (N_m)

Recent migratory rates and the number of migrants per generation were determined as proxy estimates of gene flow among the *O. niloticus* populations. However, recent migratory rates were only estimated for the East African populations, since the corresponding water bodies are the most affected by anthropogenic activities such as fish translocations. Pairwise recent migration rates were estimated using BayesAss Version 3.0 (67). Here, the program was run for 200,000,000 iterations, discarding the first 100,000,000 generations and sampling every 1000th generation (68). Only results with a 95% confidence interval of a fraction of migrants per population above 0.01 were considered significant. Recent migration rates were used because most of the fish translocations in the region, seemingly were recent. Additionally, we estimated the number of migrants (N_m) per generation between population pairs, to validate the recent migration rates using GenAlex program. Consequently, we present both, the percentage of migrants estimated in BayesAss and the number of migrants between population pairs against the fixation index (F_{st}) values.

Genetic diversity, differentiation, and isolation by distance (IBD)

Genetic diversity and differentiation indices between *O. niloticus* populations throughout East Africa and beyond were examined using the following indices: expected heterozygosity (H_e), observed heterozygosity (H_o), number of alleles (N_a), allelic richness (A_r), fixation index (F_{st}), private alleles, and Garza-Williamson index (G-W). N_a , F_{st} , G-W and H_e per population were analyzed using the program Arlequin Version 3.5 (69). H_o , H_e , N_a and PIC per locus were determined through Cervus version 3.0.7 (70). A_r was analyzed using the rarefaction algorithm implemented in the Hp-rare program (71). G-W was used to explore the

possibility of bottlenecks amongst the populations. If G-W values are closer to zero, it implies that the populations went through a bottleneck, but when the values are close to one, the populations are in a stable phase (72). To test whether the genetic diversity and differentiation of *O. niloticus* populations conform to isolation by distance (IBD), we plotted genetic distance (F_{st}) against the geographical distance (GGD in kilometers) and conducted correlation analyses using Mantel test (999 permutations) implemented in GenAlex Version 6.5 (57).

Declarations

Ethics approval and consent to participate

Field excursions were conducted together with respective authorities per region and therefore no special permission was required. Throughout the sampling regimes, fish were always dead when obtained from the local fishermen, thus no special treatment for the animals was administered in the process.

Consent for publication

Not applicable

Availability of data and material

Raw sequence data were submitted to the sequence read archive (SRA) database and can be accessed under the reference number PRJNA550300

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the Austrian Partnership Program in Higher Education and Research for Development (APPEAR) through the STRECAFISH project, a program of the Austrian Development Cooperation (ADC) and implemented by the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH).

Authors' contributions

PDT, MC, and HM developed the research concept and performed experimental design, laboratory work, bioinformatics, and data analysis with contributions of EA, CM, and HW. PDT, EA, HW, PA, and CM conducted fieldwork. HM provided laboratory space/analytical tools. PDT and MC led the writing process with substantial contributions of the other authors. All co-authors read and approved the final version of the manuscript.

Acknowledgments

We are indebted to Kachwekano Zonal Agricultural Research and Development Institute, National Agricultural Research Organization, Uganda for their great support during field activities. We are grateful to the local governments, in particular, the respective field Fisheries Officers from East Africa and the Ethiopian Fishery and Aquatic Life Research Centre, for their assistance and advice during field excursions. Special thanks to Ms. Howot Teshome and Dr. Mano Komandan who transported our Ethiopian and Burkina Faso samples respectively to Vienna, Austria.

References

1. Agnès J-F, Adépo-Gourène B, Abban EK, Fermon Y. Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity*. 1997;79(1):88–96.
2. Trewavas E. Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*. *London: British Museum (Natural History)*; 1983.
3. Ogutu-Ohwayo R. The reduction in fish species diversity in Lakes Victoria and Kyoga (East Africa) following human exploitation and the introduction of non-native fishes. *Journal of Fish Biology*. 1990;37:207–8.
4. Balirwa JS. The evolution of the fishery of *Oreochromis niloticus* (Pisces: Cichlidae) in Lake Victoria. *Hydrobiologia*. 1992 232(1):85–9.
5. Kaufman L. Catastrophic change in species-rich freshwater ecosystems. *BioScience*. 1992;42(11):846–58.
6. Canonico GC, Arthington A, McCrary JK, Thieme ML. The effects of introduced tilapias on native biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems*. 2005;15(5):463–83.
7. Worthington EB. A report on the fisheries of Uganda investigated by the Cambridge Expedition to the East African Lakes, 1930–31. 1932.
8. Agnès J-F, Adépo-Gourène B, Pouyaud L. Natural hybridization in tilapias. *Genetics and Aquaculture in Africa*. 1998:95–105.
9. Welcomme R. Recent changes in the stocks of *Tilapia* in Lake Victoria. *Nature*. 1966;212(5057):52.
10. Mboya D, Manyala J, Ngugi C. Fish introductions and their impact on the biodiversity and the fisheries of Lake Victoria. Fisheries Department, Kisumu, Kenya, 11 pp. 2001.
11. Mwanja WW. Genetic biodiversity and evolution of two principal fisheries species groups, the Labeine and Tilapiine, of Lake Victoria Region, East Africa: The Ohio State University; 2000.
12. Pullin RS. *Tilapia genetic resources for aquaculture*. Proceedings of the Workshop on Tilapia Genetic Resources for Aquaculture, 23–24 March 1987. The WorldFish Center Working Papers. 1988.

- 13.Zengeya TA, Booth AJ, Chimimba CT. Broad niche overlap between invasive Nile tilapia *Oreochromis niloticus* and indigenous congeners in southern Africa: Should we be concerned? *Entropy*. 2015;17(7):4959–73.
- 14.Rothuis A, Turenhout M, van Duijn A, Roem A, Rurangwa E, Katunzi E, et al. Aquaculture in East Africa: a regional approach. IMARES/LEI; 2014.
- 15.Nunan F, Hara M, Onyango P. Institutions and co-management in east African inland and Malawi fisheries: A critical perspective. *World Development*. 2015;70:203–14.
- 16.Allendorf FW, Luikart G. Conservation and the genetics of populations. *Mammalia*. 2007;2007:189–97.
- 17.Vreven EJ, Adépo-Gourène B, Agnèsè J-F, Teugels G. Morphometric and allozyme variation in natural populations and cultured strains of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Belgian Journal of Zoology*. 1998;128(1):23–34.
- 18.Seyoum S, Kornfield I. Identification of the subspecies of *Oreochromis niloticus* (Pisces: Cichlidae) using restriction endonuclease analysis of mitochondrial DNA. *Aquaculture*. 1992;102(1–2):29–42.
- 19.Mwanja WW, Booton G, Kaufman L, Chandler M, Fuerst P, Donaldson E, et al., editors. Population and stock characterization of Lake Victoria. Tilapine fishes based on RAPD markers. *Aquaculture Biotechnology Symposium Proceedings of the International Congress on the Biology of Fishes* (ed, by EM Donaldson & DD Mackinlay); 1996.
- 20.Mwanja WW, Booton GC, Kaufman L, Fuerst PA. A profile of the introduced *Oreochromis niloticus* (Pisces: Teleostei) populations in Lake Victoria Region in relation to its putative origin of Lakes Edward and Albert (Uganda-E. Africa) based on random amplified polymorphic DNA analysis. *African Journal of Biotechnology*. 2008;7(11).
- 21.Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam KN, et al. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International Journal of Molecular Sciences*. 2013;14(11):22499–528.
- 22.Wirgin II, Waldman JR. What DNA can do for you. *Fisheries*. 1994;19(7):16–27.
- 23.Okumuş İ, Çiftçi Y. Fish population genetics and molecular markers: II-molecular markers and their applications in fisheries and aquaculture. *Turkish Journal of Fisheries and Aquatic Sciences*. 2003;3(1):51–79.
- 24.Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, et al. Current trends in microsatellite genotyping. *Molecular ecology resources*. 2011;11(4):591–611.
- 25.Hedrick PW. Conservation genetics: where are we now? *Trends in Ecology and Evolution*. 2001;16(11):629–36.

26. Liu Z, Cordes J. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*. 2004;238(1):1–37.
27. Tibihika PD, Curto M, Dornstauder-Schrammel E, Winter S, Alemayehu E, Waidbacher H, et al. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*. 2019:1–16.
28. Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*. 2013;22(11):2841–7.
29. Vartia S, Villanueva-Cañas JL, Finarelli J, Farrell ED, Collins PC, Hughes GM, et al. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society Open Science*. 2016;3(1):150565.
30. Curto M, Winter S, Seiter A, Schmid L, Scheicher K, Barthel LM, et al. Application of a SSR-GBS marker system on investigation of European Hedgehog species and their hybrid zone dynamics. *Ecology and Evolution*. 2019.
31. Ndiwa TC, Nyingi DW, Agnese J-F. An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Lobo drainage, Kenya. *PloS One*. 2014;9(9):1–15.
32. Nyingi D, De Vos L, Aman R, Agnès J-F. Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758)(Cichlidae; Teleostei) in the Lobo Swamp (Kenya). *Aquaculture*. 2009;297(1–4):57–63.
33. Angienda PO, Lee HJ, Elmer KR, Abila R, Waindi EN, Meyer A. 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*. 2011;12(1):243–55.
34. Frankham R, Briscoe DA, Ballou JD. *Introduction to conservation genetics*: Cambridge university press; 2002.
35. Welcomme R, Cowx IG, Coates D, Béné C, Funge-Smith S, Halls A, et al. Inland capture fisheries. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010;365(1554):2881–96.
36. Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, et al. Hybridization and extinction. *Evolutionary Applications*. 2016;9(7):892–908.
37. Rhymer JM, Simberloff D. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*. 1996;27(1):83–109.
38. Vijverberg J, Dejen E, Getahun A, Nagelkerke LA. The composition of fish communities of nine Ethiopian lakes along a north-south gradient: threats and possible solutions. *Animal Biology*. 2012;62(3):315–35.

39. Tsegay T. Characteristics and Status of the Highland Lake Hashenge Fishery, Tigray, Northern Ethiopia. *International Journal of Aquaculture and Fishery Sciences* 2017;3(4):71–6.
40. Odada EO, Olago DO, Bugenyi F, Kulindwa K, Karimumuryango J, West K, et al. Environmental assessment of the East African rift valley lakes. *Aquatic Sciences*. 2003;65(3):254–71.
41. Dumont HJ. *The Nile: origin, environments, limnology and human use*: Springer Science & Business Media; 2009.
42. Ogutu-Ohwayo R. 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*. 1990;27(2):81–96.
43. Tibihika PD, Waidbacher H, Masembe C, Curto M, Sabatino S, Alemayehu E, et al. Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa. *Environmental Biology of Fishes*. 2018;101(3):363–81.
44. Fuerst P, Mwanja W, Kaufman L, editors. *The genetic history of the introduced Nile tilapia of Lake Victoria (Uganda–E. Africa): the population structure of *Oreochromis niloticus* (Pisces: Cichlidae) revealed by DNA microsatellite markers*. *Tilapia Aquaculture in the 21st Century Proceedings from the Fifth International Symposium on Tilapia in Aquaculture*; 2000.
45. Balirwa JS, Chapman CA, Chapman LJ, Cowx IG, Geheb K, Kaufman L, et al. Biodiversity and fishery sustainability in the Lake Victoria basin: an unexpected marriage? *BioScience*. 2003;53(8):703–15.
46. Deines A, Bbole I, Katongo C, Feder J, Lodge D. Hybridisation between native *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River, Zambia. *African Journal of Aquatic Science*. 2014;39(1):23–34.
47. Shechonge A, Ngatunga BP, Tamatamah R, Bradbeer SJ, Harrington J, Ford AG, et al. Losing cichlid fish biodiversity: genetic and morphological homogenization of tilapia following colonization by introduced species. *Conservation Genetics*. 2018;19(5):1199–209.
48. Benvenuto C, Tabone E, Vercken E, Sorbier N, Colombel E, Warot S, et al. Intraspecific variability in the parasitoid wasp *Trichogramma chilonis*: can we predict the outcome of hybridization? *Evolutionary Applications*. 2012;5(5):498–510.
49. Dodson JJ, Bourret A, Barrette MF, Turgeon J, Daigle G, Legault M, et al. Intraspecific genetic admixture and the morphological diversification of an estuarine fish population complex. *PloS one*. 2015;10(4):e0123172.
50. Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*. 2001;16(11):613–22.

51. Roberts DG, Gray CA, West RJ, Ayre DJ. Marine genetic swamping: hybrids replace an obligately estuarine fish. *Molecular Ecology*. 2010;19(3):508–20.
52. Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N. Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*. 2008;23(6):327–37.
53. Ogutu-Ohwayo R, Odongkara K, Okello W, Mbabazi D, Wandera S, Ndawula L, et al. Variations and changes in habitat, productivity, composition of aquatic biota and fisheries of the Kyoga lake system: lessons for management. *African Journal of Aquatic Science*. 2013;38(sup1):1–14.
54. Njiru M, Getabu A, Jembe T, Ngugi C, Owili M, Van der Knaap M. Management of the Nile tilapia (*Oreochromis niloticus* (L.)) fishery in the Kenyan portion of Lake Victoria, in light of changes in its life history and ecology. *Lakes & Reservoirs: Research & Management*. 2008;13(2):117–24.
55. Neophytou C, Torutaeva E, Winter S, Meimberg H, Hasenauer H, Curto M. Analysis of microsatellite loci in tree of heaven (*Ailanthus altissima* (Mill.) Swingle) using SSR-GBS. *Tree Genetics & Genomes*. 2018;14(6):82.
56. Arthofer W, Heussler C, Krapf P, Schlick-Steiner BC, Steiner FM. Identifying the minimum number of microsatellite loci needed to assess population genetic structure: A case study in fly culturing. Taylor & Francis. 2018;12(1):13–22.
57. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources*. 2006;6(1):288–95.
58. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978;89(3):583–90.
59. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*. 2005;23(2):254–67.
60. Langella O. Populations 1.2.32: population genetic software (individuals or population distances, phylogenetic trees). http://www.bioinformatics.org/project/?group_id = 84. 2011.
61. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*. 2009;9(5):1322–32.
62. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*. 2013;4.
63. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 2005;14(8):2611–20.
64. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 2012;4(2):359–61.

- 65.Chaluvadi SR, Khanam S, Aly MA, Bennetzen JL. Genetic diversity and population structure of native and introduced date palm (*Phoenix dactylifera*) germplasm in the United Arab Emirates. *Tropical plant biology*. 2014;7(1):30–41.
- 66.Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*. 2015;15(5):1179–91.
- 67.Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*. 2003;163(3):1177–91.
- 68.Curto M, Puppo P, Kratschmer S, Meimberg H. Genetic diversity and differentiation patterns in *Micromeria* from the Canary Islands are congruent with multiple colonization dynamics and the establishment of species syngameons. *BMC Evolutionary Biology*. 2017;17(198):1–16.
- 69.Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 2010;10(3):564–7.
- 70.Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*. 2007;16(5):1099–106.
- 71.Kalinowski ST. Hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Resources*. 2005;5(1):187–9.
- 72.Excoffier L, Estoup A, Cornuet J-M. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*. 2005;169(3):1727–38.

Figures

Figure 2

Genetic structure of *O. niloticus* populations based on UPGMA dendrogram. Node values correspond to bootstrap values.

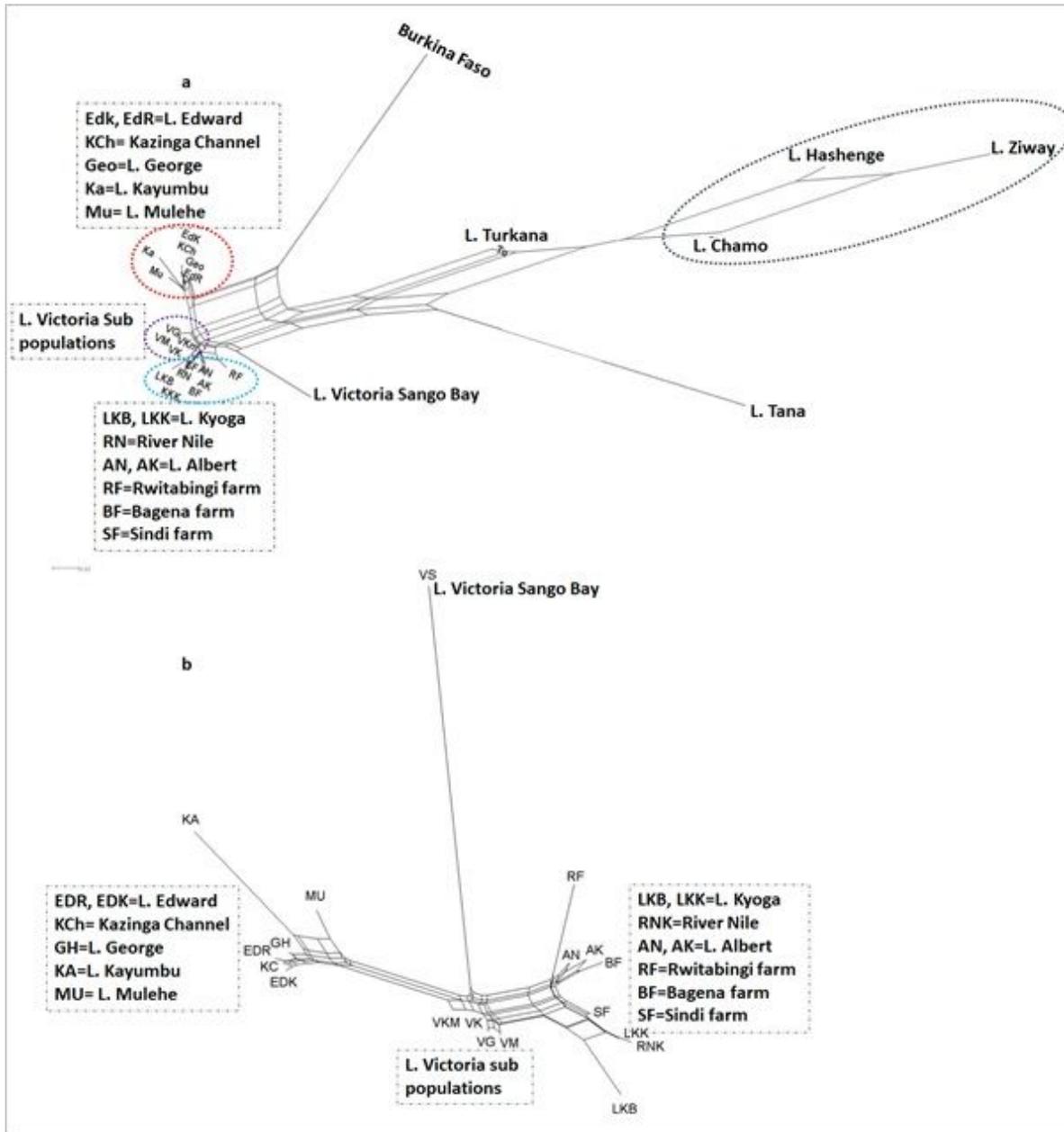


Figure 3

Genetic structure based on unrooted network tree illustrating population relationships based on genetic distance. a represents a network tree for all the populations and b for only the Ugandan populations. Dotted oval and rectangular shapes depict closely related genetic groups.

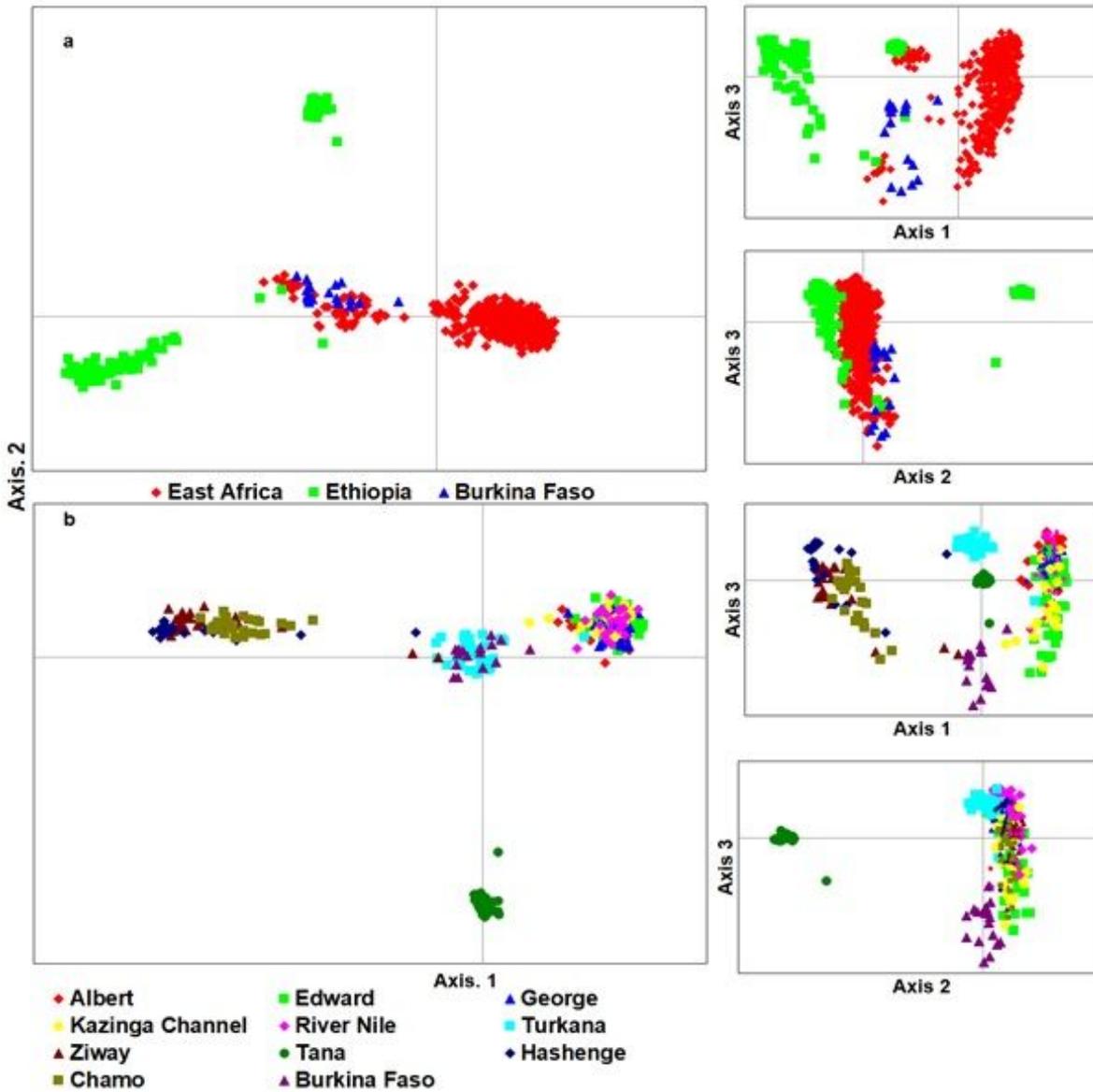


Figure 4

Genetic scatter plots of *O. niloticus* exhibited by Principal Coordinate Analysis (PCoA). (a) populations per region, (b) all indigenous populations. PCoA was constructed with respect to unbiased Nei's genetic distance among individuals

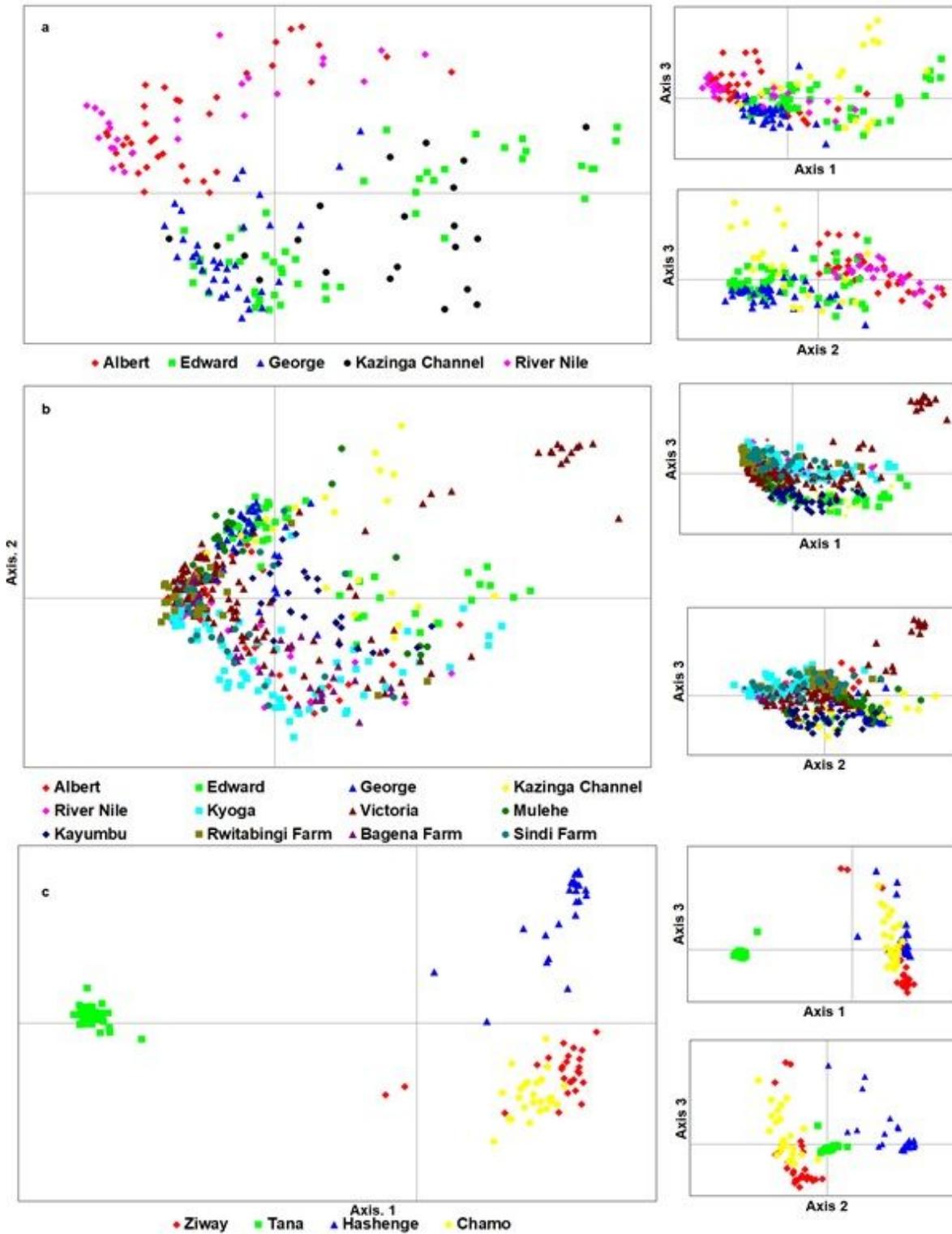


Figure 5

Genetic scatter plots of *O. niloticus* based on Principal Coordinate Analysis (PCoA). (a) Ugandan native populations, (b) all Ugandan populations including non-natives, natives, and farms, and (c) all Ethiopian populations

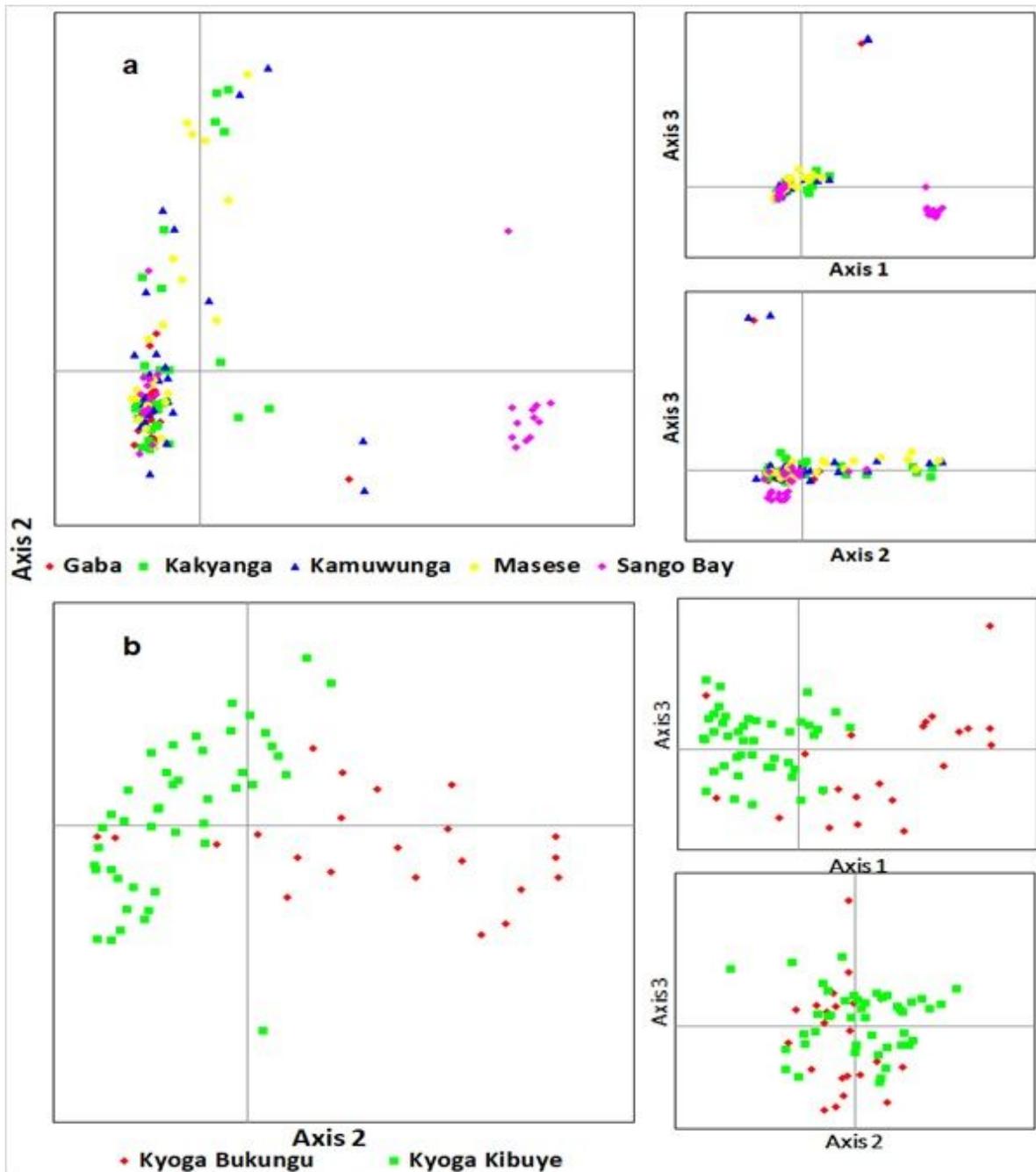


Figure 6

Genetic scatter plots of *O. niloticus* exhibited by PCoA within Lakes Victoria (a) and Kyoga (b) populations

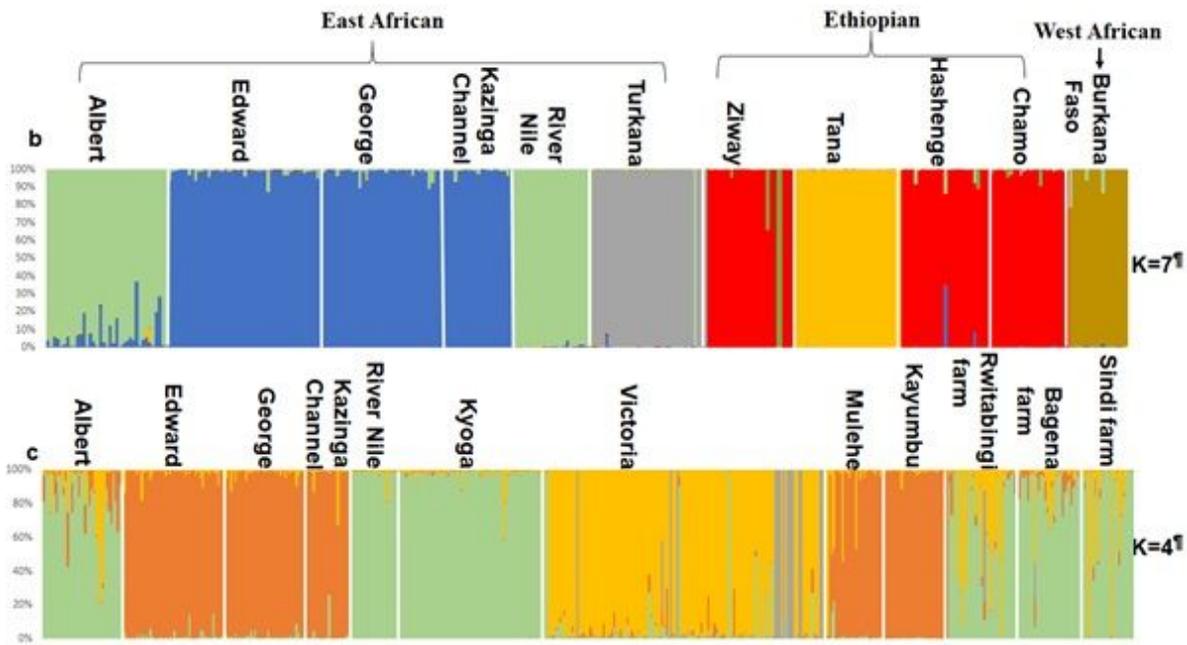


Figure 7

Bayesian clustering for genetic assignments of *O. niloticus* populations. (a) represents all populations, (b) all indigenous populations, and (c) all Ugandan populations including indigenous, non-indigenous and farms. K_s with a superscript symbol (¶) indicates the optimal K values based on STRUCTURE HARVESTER analyses.

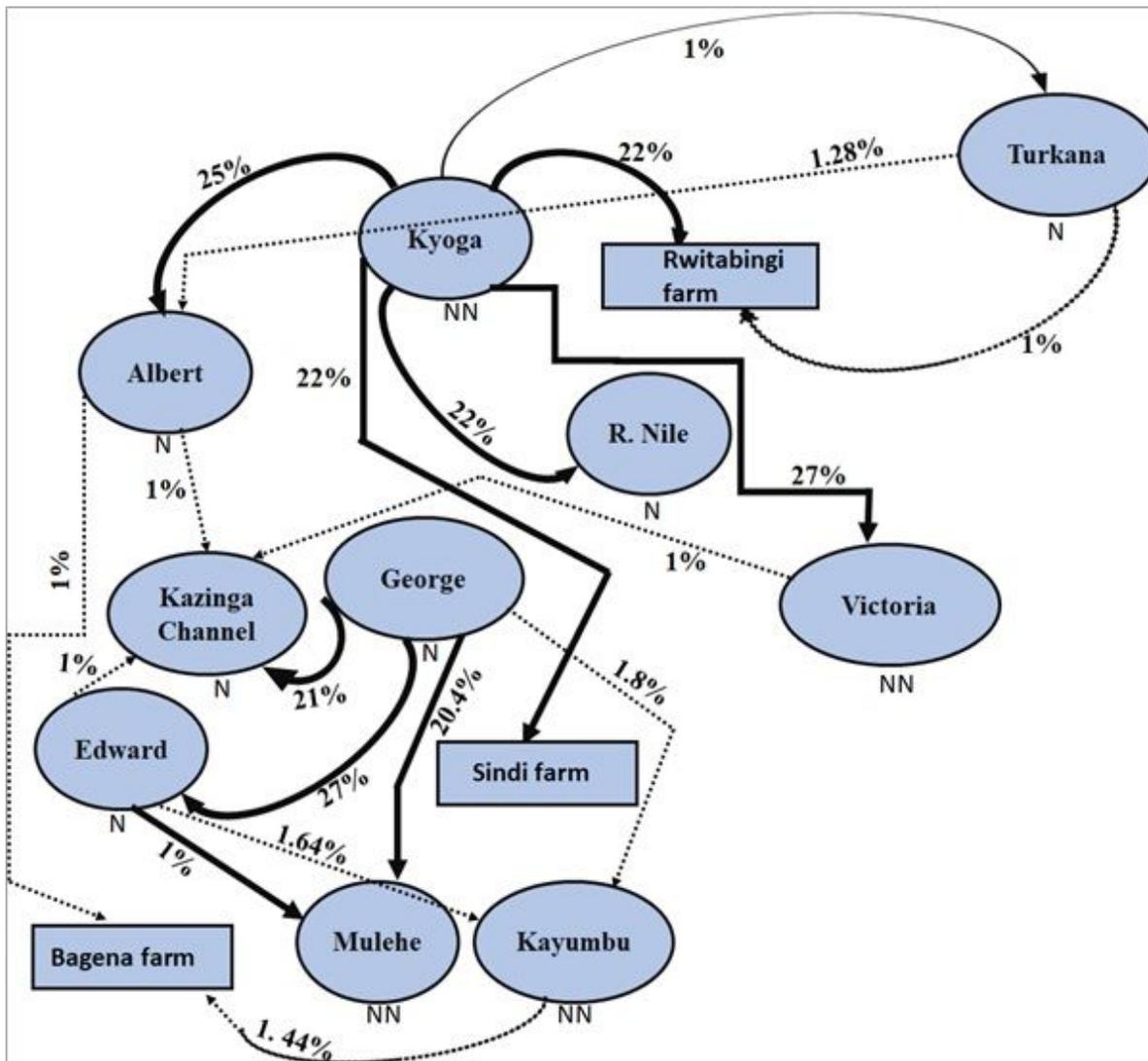


Figure 8

Bayesian inference of recent migratory rates for the 13 East African *O. niloticus* populations. Oval light blue and rectangular light-blue shapes indicate natural and farm populations, respectively. The arrows contain percentage values showing the direction and magnitude of gene flow. Darker and thick arrows represent stronger gene flow, while thin, dotted arrows indicate weaker gene flow. Native and non-native populations are indicated by the letters, "N" and "NN", respectively. This analysis is based on BayesAss program and for GenAlex program, see the additional file, Table S3

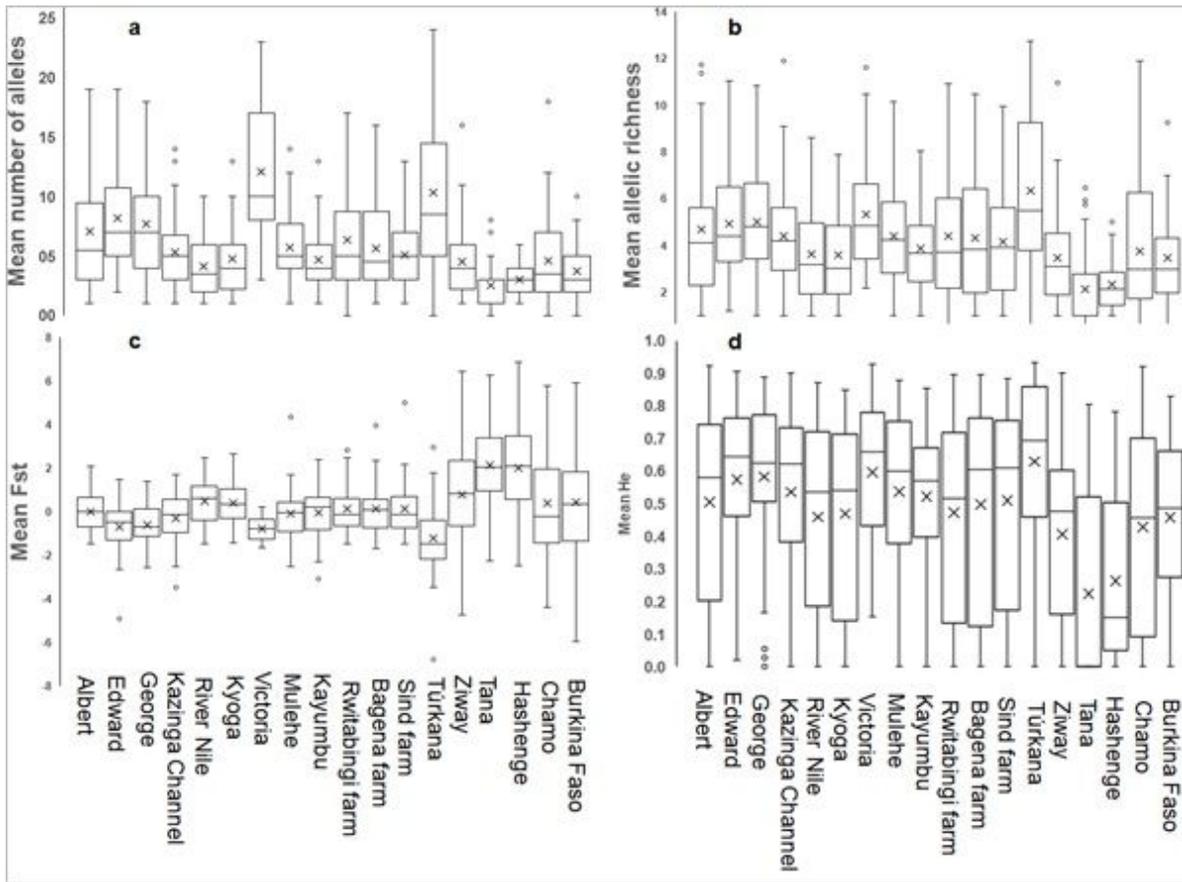


Figure 9

Genetic diversity and differentiation indices. a number of alleles, b allelic richness, c fixation index (Fst) and d expected heterozygosity

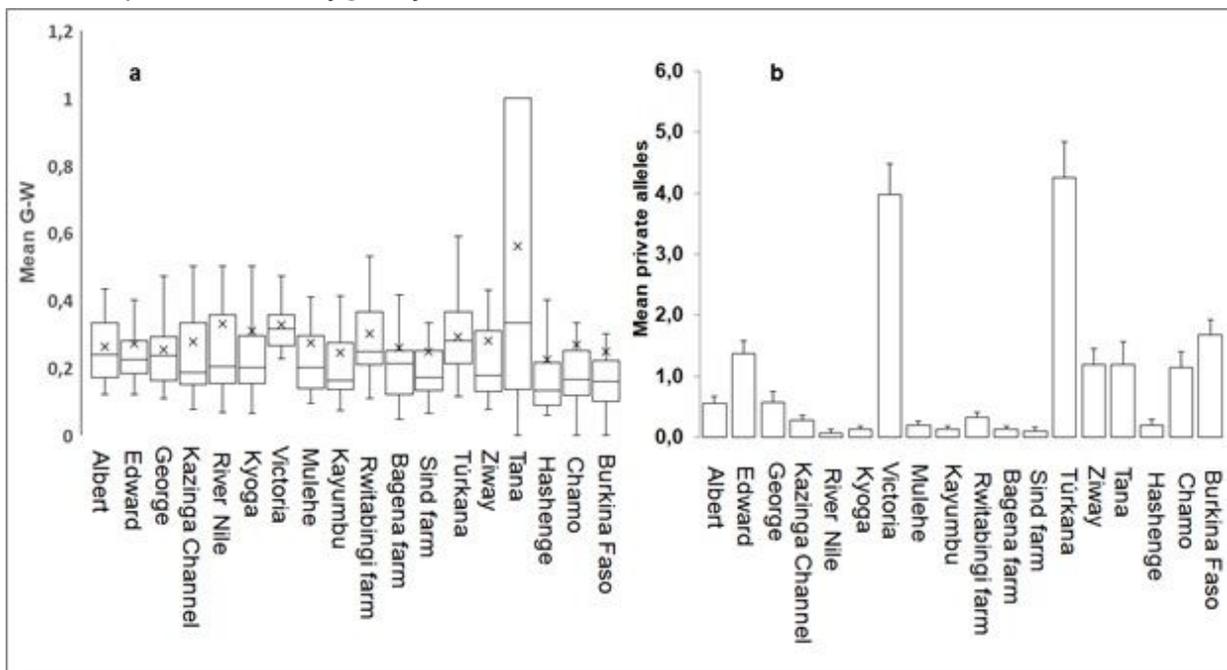


Figure 10

Estimations of population bottleneck derived from Garza-Williamson Index (G-W) (a) and measure of genetic diversity based on private alleles (b)

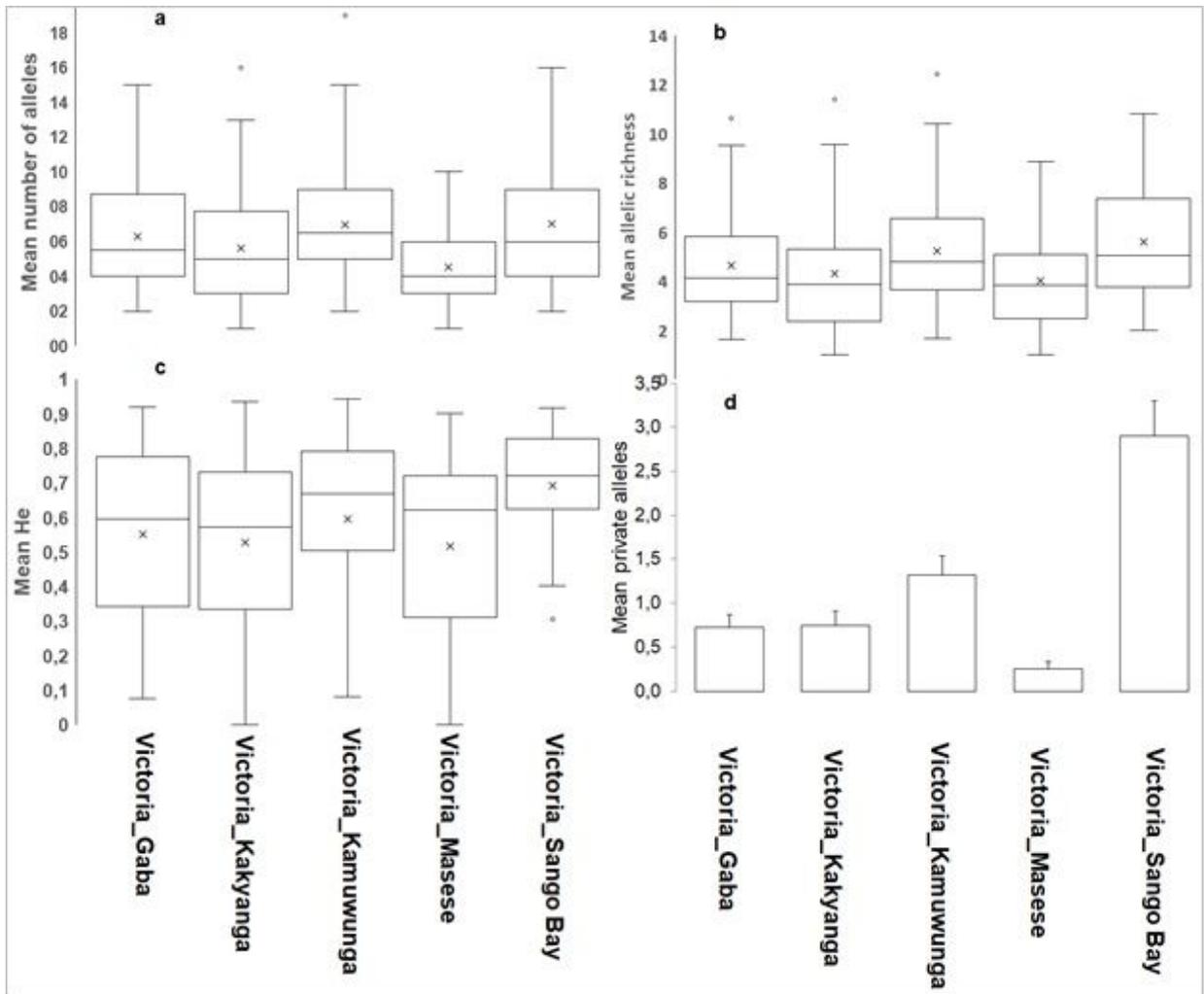


Figure 11

Genetic diversity of Lake Victoria within the population. a number of alleles, b allelic richness c expected heterozygosity and d private alleles

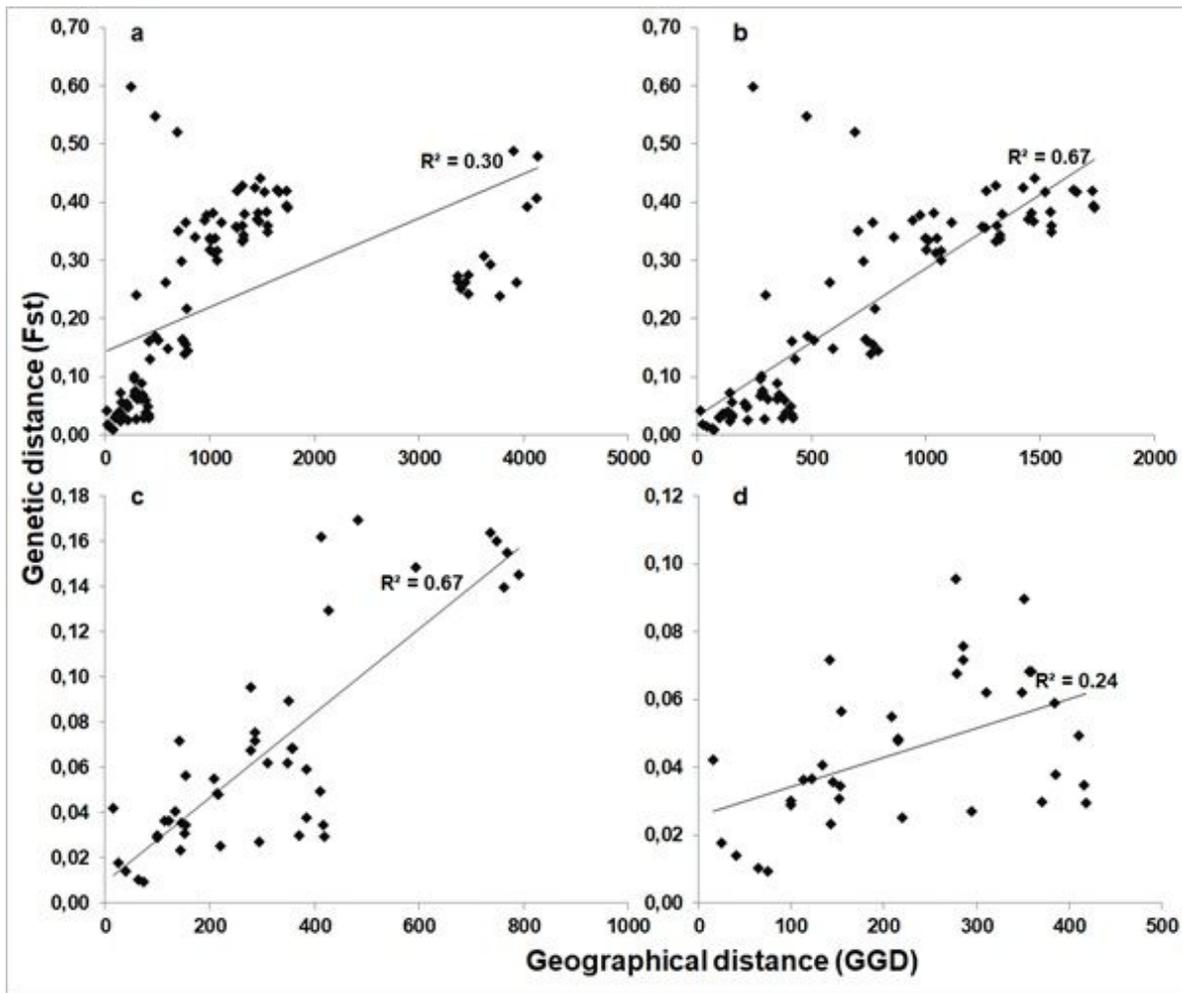


Figure 12

Mantel tests for correlations between genetic distance (Fst) and Euclidean geographical distance (GGD in Km) for *O. niloticus* populations. a represents isolation by distance (IBD) between all populations, b all populations without Burkina Faso, c East African, and d only Ugandan populations. References 1. Tibihika PD, Waidbacher H, Masembe C, Curto M, Sabatino S, Alemayehu E, et al. Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa. *Environmental Biology of Fishes*. 2018;101(3):363-81.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile.docx](#)
- [SecondadditionalfileAllelesListmatrix.txt](#)