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

Papius Dias Tibihika (papiust@ahoo.com)
Universitat fur Bodenkultur Wien https://orcid.org/0000-0001-6918-9903
Manuel Curto
Universitat fur Bodenkultur Wien Department fur Integrative Biologie und Biodiversitatsforschung
Esayas Alemayehu
Universitat fur Bodenkultur Wien Department Wasser-Atmosphare-Umwelt
Herwig Waidbacher
Universitat fur Bodenkultur Wien Department Wasser-Atmosphare-Umwelt
Charles Masembe
Makerere University
Peter Akoll
Makerere University
Harald Meimberg
Universitat fur Bodenkultur Wien Department fur Integrative Biologie und Biodiversitatsforschung

Research article

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

Posted Date: July 23rd, 2019

DOI: https://doi.org/10.21203/rs.2.11849/v1

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 stoking activities. Conclusions The anthropogenic activities particularly in the East African O. niloticus translocations, promoted admixture and contact with the native congenerics 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 freshwater systems in Western Africa (e.g., Senegal, Gambia, Niger, Benue, Chad) as well as to rivers in East Africa and many of the Rift Valley Lakes (1, 2). For more than nine decades, this species has been intentionally dispersed worldwide, in particular for aquaculture and restocking programs (2, 3). In East Africa, various fish introductions are reported, starting in the 1920s. *O. niloticus,* and other species like Tilapia nigra (*Tilapia spilurus nigra)*and Black bass (*Micropterus salmoides)* had been used for stocking initially to build up fisheries in lakes considered as unproductive like the southwestern Uganda high-altitude lakes, and Naivasha in Kenya (4, 5). For example, Lake Bunyonyi was stocked with individuals from Lake Edward in 1920 (4).

In the 1950s, several tilapiine species were stocked into Lake Victoria, Nabugabo and Kyoga to counteract the decline of native fish species (6–9). The introduced species (*O. niloticus, O. leucosticus, Tilapia zillii* and *Tilapia melanopleura)* were all suspected to originate from Lake Albert (Balirwa, 1992; Ogutu-Ohwayo, 1990; Welcomme, 1966). However, some introductions might have also originated from Lake Edward and Lake Turkana (2, 5, 7). Following these introductions, the indigenous fish species, such as *O. variabilis* and Ngege*O. esculentus),*, significantly declined (6, 7, 10). The reason for the declined native fish species was suspected to a combination of factors including; competition, over fishing, as well as predation pressures from another introduced species, the Nile perch (*Lates niloticus)* (6, 8). But one additional factor might have been hybridization with the introduced *O. niloticus (*7, 9, 10). The expanded distribution of *O. niloticus* in East Africacomplicates differentiation and identification of genetic units for management and conservation. For example, the population considered as nonnative in Lake Victoria and Kyoga might have genetically diverged from the native populations (7, 9, 11). The loss of indigenous *O. mossambicus* due to hybridization with the introduced *O. niloticus* has been reported in South Africa (12). The situation in East Africa may have worsened with the recent boom of fish hatcheries and aquaculture production systems (13). Feral populations resulting from escape 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 (14). 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 (15). Therefore, in particular for the East African *O. niloticus,* as the species were particularly affected by various anthropogenic activities, a thorough characterization of the populations might be needed.

The earliest studies in this respect used traditional morphometric and meristic analyses. Using biometrics and counts, Trewavas (2) described seven subspecies of *O. niloticus* in Africa and associated them to different regions or lakes. Nevertheless, this classification was contradicted as misleading (16, 17). For example, Vreven, Teugels (16) used morphometric analyses accompanied with allozyme markers and found that, contrary to findings from Trewavas (2), the strain from Lake Edward was closely related to the lower Nile (Egypt). Moreover, Seyoum and Kornfield (1992) using restriction endonuclease mitochondrial DNA, found *O. niloticus* in Lake Tana distinct, contrary to findings from Trewavas (2). Other earlier genetic studies concerning *O. niloticus* in East Africaare reflected by Agnèse, Adépo-Gourène (1), Mwanja, Booton (18), Mwanja, Booton (19). Agnèse, Adépo-Gourène (1) employed allozymes and restriction fragment length polymorphism (RFLP) of mitochondrial DNA (mtDNA) and Mwanja, Booton (18), Mwanja, Booton (19) randomly amplified polymorphic DNA (RAPD). Interestingly, Agnèse, Adépo-Gourène (1) reported that *O. niloticus* populations from Albert Nile (the Egyptian stretch of River Nile) are distinct from the West African populations, contrary to earlier reports by Trewavas (2). Furthermore, Agnèse, Adépo-Gourène (1) indicated that *O. niloticus* inLake Tana is clustered together with Lake Edward and the Kenyan Lake Turkana system, which is contrary to the findings from Seyoum and Kornfield (17). These findings are conflicting probably because of the different methodological approaches used comprise different information content (20–25). 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 (21). The lack of methodologies with high discriminating power, 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 (SSR) to typify the *O. niloticus* in East Africa. Using next-generation sequencing, SSR loci have been proven to be robust when determining the genetic structure of *O. niloticus,* particularly, using microsatellite genotyping by sequencing (SSR-GBS) (26). Albeit recent studies using SSR have also focused on *O. niloticus* (27–29), their work was limited to few waterbodies in Kenya, with the broader scope of the African Great Lakes missing. It is important, however, to conduct a comparative study of various waterbodies where *O. niloticus* is present (native and non-native with possible admixture). Such research would not only be informative in the context of the cichlid's genetic structure and diversity but also would establish a firm base for management and for conservation options (30). SSR-GBS approaches are useful in getting rid size homoplasy, which is one of the constraints of traditional fragment length analysis (31, 32).

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 populations with populations that had been stocked and artificial ponds of aquaculture. With this approach, we investigate the impact of anthropogenic activities on the O. *niloticus'* gene pool. We hypothesized that anthropogenic activities have affected the genetic divergence of some *O. niloticus* populations,particularly in environments where the species was introduced. In a similar trend, we 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 is related to geography and anthropogenic activities including the pathways of the translocation processes?

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 and were used for allele calling. Genetic variation results for the 40 SSR loci are presented in additional file, Table S2. The number of alleles per locus ranged from seven to 84, with a total of 1,352 alleles generated across all loci. Overall, 80% of the loci

exhibited expected heterozygosity (He) values greater than 0.5. Polymorphic Information Content (PIC) was generally congruent to He, 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 2.1). In this case, the three Ethiopian populations (Hasshenge, Ziway, and Chamo) formed the most distant group followed by Burkina Faso and the other Ethiopian waterbody, 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 and one Lake Kyoga sub-population, the nonnative lakes and fish farms grouped with a native population: the southern Ugandan high-altitude Lakes (Kayumbu and Mulehe) with Lake George; Kyoga-Kibuye and Sindi Farm with River Nile; and Bagena and Rwitabingi farms with Albert. The three sub-populations of Lake Victoria (Gaba, Masese, Kakyanga) formed an independent group while the other (Sango Bay) showed the highest degree of divergence in East Africa.

Neighbor network results showed a similar pattern to the UPGMA dendrogram both at regional and local levels (Figure 2.2). 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 Chanel, and Edward group together with the nonnative Ugandan highland lakes, and on the other end, Albert and River Nile systems together with the nonnative 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 nonnative populations (including farms) showed longer branches than the natives (Figure 2.2).

Genetic distance between individuals which was visualized through a PCoA analysis showed a separation of population groups based on geographic regions (Figure 3.1a). Samples formed four groups when analyzed at regional/country level (Figure 3.1a): 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 3.1b). 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, the separation between Lake Turkana and the remaining native populations was evident (Figure 3.1b). Individuals found in the Ugandan native waterbodies divided into two main groups (Figure 3.2a). One group was composed by Lake Albert and River Nile individuals while the other by Lake Edward, Kazinga Chanel, and George. This division was less evident when individuals from nonnative waterbodies and fish farms were included in the analysis (Figure 3.2b). 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 3.2c). Substructure within the same lake was only evident for Lakes Victoria and Kyoga (Figure 3.3).

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). Comparatively, we also present other K values representing the total number of populations per analysis. *O. niloticus* populations from each African region were assigned to different groups (Figure 4a). Within each region, the same assignments were observed with Lakes Tana and Turkana isolated from the rest of Ethiopians and East African populations, respectively. Among the Ugandan native populations, clustering was also congruent with the two water systems, as indicated earlier by both network and PCoA analyses, see Figure 4b and 4c. However, there were cases where the nonnative 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 4c). Admixture was more evident amongst the East African populations but mostly detected when nonnative populations were considered than was for natives (Figure 4c).

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 5), 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 and to Mulehe (21% and 20.4 %) respectively (Figure 5). Migration rates estimated through Genalex were congruent with BayesAss, but with the difference that Lake Victoria was also a source of migrants (additional file, Table S3).

Genetic differentiation, diversity, and isolation by distance

Genetic differentiation was consistent with the STRUCTURE results. For instance, the Fst values clearly demonstrated that the East African populations are genetically distant from the Ethiopian and West African populations (Figure 6.1a). Despite River Nile and Lake Kyoga showing relatively high Fst values, results from the East African populations generally showed low genetic differentiation. Also, the East African populations were genetically more diverse when compared to either Ethiopian or Burkina Faso (Fig. 6.1b-d). Based on all statistics, the non-native Lake Victoria and native Lake Turkana populations were the most genetically diverse. Lake Kyoga and River Nile were consistently the least diverse even when investigated at 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 6.2a). In the analysis, only Lake Tana exhibited G-W values >0.5. Regarding population genetic diversity, however, Lakes Victoria and Turkana showed the highest number of private alleles (Figure 6.2b).

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 6.3). Sango Bay, in particular, was isolated based on Fst values, and consistently exhibited higher genetic diversity indices (Na, He and Ar) (Figure 6.3).

Mantel tests for isolation by distance (IBD) across all samples showed a positive correlation between geographical and genetic distance (R^2 = 0.30) (Figure 7¶). However, the strong correlation (R^2 = 0.67) between the populations was only found when Burkina Faso was excluded from the analysis (Figure 7†). 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 7§), which was not the case when only Ugandan populations were considered (Figure 7‡).

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 (33). One aspect of the anthropogenic threats is the change or alteration of the natural genetic structure of fish stocks through admixture (34, 35). To understand the admixture of stocks, it 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, *O. niloticus,* towards management and conservation genetics. 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 allows identifying populations. 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. Tana was completely distinct from the remaining populations. This applies not to only to 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. Seyoum and Kornfield (17) with allozymes found this population to be different and thus contradicting the subspecies treatment from Trewavas (2). This high level of differentiation was confirmed in concurrence to these past studies and so the revision of the species delimitation for these populations is suggested.

Lake Tana lies in the Ethiopian mountains and is isolated from the Lakes in the Rift valley (36). 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 (37). 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 stoking 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 are expected given its high geographical isolation (38). Trewavas (2) treated the population from this lake as a different subspecies (*O. vulcani)*.. The high diversity and number of private alleles found in Lake Turkana can be a consequence of this isolation. The East African arid Lake 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 waterbodies, 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 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 (39). The strong rapids and falls present in this river (6, 39), 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 (40) but do not concur with past studies (2, 19). This incongruity might be associated with different methodological approaches utilized between the earlier studies and the current. For example, Trewavas (2) using morphometric and meristics treated *O. niloticus* from the Edward-George system and Albert as one subspecies (*O. niloticus eduardianus)*.. However, the use of traditional morphometrics has been strongly labeled as weak in identifying species due to the lack of informative characters (17). Similarly, while we used SSR-GBS techniques, Mwanja, Booton (19) employed random amplified polymorphic DNA (RAPD) markers, which due to their dominance genotypic nature provide only part of the information (21).

Anthropogenic activities-fish translocations

In East Africa, we know that *O. niloticus* was introduced into several water bodies through stocking activities and here we show their genetic impacts. Referring to both genetic structure and migration rate analyses, the two Ugandan groups (the George-Edward complex and Lake Albert) contributed to the stocking of different water bodies. Apparently, *O. niloticus* fromthe 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 waterbody 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 is the natural occurrence of Murchison Falls on the River Nile that acts as a barrier between the systems (39, 41). For this matter, the genetic similarity between River Nile, Lakes Kyoga, and Albert may result in stocking regimes using the latter as source as reported by Ogutu-Ohwayo (41).

Fish farms seem to have sourced 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. 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 showed to be admixed with other populations including Lake Victoria. Apart from farms, evidence of admixture was noticed in the East African natural populations, which seems to have been promoted by anthropogenic activities. This is supported by the fact that when non-native populations were not considered in the Structure and PCoA analyses, signals of admixture were reduced, 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, in the distance Network, 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 Lake Victoria. Trewavas (2) suggests that introductions into Lake Victoria may have originated from Lake Edward, with other authors suggesting multiple sources (7, 8, 11, 42), 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 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 from other 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, 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 not only noticed when compared with the remaining subpopulations within the Lake, but also with the other East African populations. In this case, during the boom of *O. niloticus* population in Lake Victoria (7, 8, 41, 43), a larger portion of the native species genetic variation may have been introduced in its gene-pool. However, in this study, we cannot directly test for hybridization since we did not include samples of *O. niloticus* congenerics. However, hybridization involving *O. niloticus* and other tilapiines has been reported to be relatively frequent and it needs to be considered (9, 29, 44, 45).

The reported admixture/hybridization especially in Lake Victoria, may have adaptive consequences and compromise the sustainability of *O. niloticus.* Although hybridization may lead to heterosis/hybrid vigor (46, 47), the admixture is usually reported to have negative consequences (35, 48). Introgression can contribute to outbreeding depression either by the introduction of maladaptive alleles or through the dilution of alleles important for local adaptation (Roberts, Gray, West, & Ayre, 2010). In more drastic scenarios, hybridization can result in genomic incompatibilities contributing to a fast reduction of population fitness (49). Alternatively, the hybrids my potentially exhibit more fitness and subsequently extirpate the parental lines (44). 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 waterbody showed to be admixed with Lake Kyoga. Besides we found significant migrations from Lake Kyoga to Lake Albert. These results

indicated that admixture with respect to translocations does not only contribute 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 signals of genetic erosion. These were indicated by low genetic variability and evidence of bottleneck with respect to G-W estimations. In a previous study, Mwanja, Booton (19) had already reported this to be the case for Lake Kyoga. High loss of genetic diversity among populations, particularly, in fishes has been attributed to over-exploitation (50). This might be the case for L. Kyoga population, but other factors may also play a role. The low diversity in River Nile could be linked to the hydro-electric power dams that have been constructed along the river (Ugandan side, upper Nile), which might be restricting gene flow and increasing the effect of 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 no dams. Nevertheless, Nile tilapia was introduced in these waterbodies recently and the low genetic diversity would be a consequence of founder effect.

Implications for management and outlook

Overall, we found evidence that anthropogenic activities affected the gene-pool of East African *O. niloticus.* The main consequence was admixture and possible hybridization between different stocks and species respectively. In the long term, this may have negative effects on populations fitness due to outbreeding depression and genetic swamping. Thus, management measures should inhibit any form of artificial 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 un admixed from external sources. To avoid an influx of feral populations and other infrastructure developments, a proper environmental impact assessment should be prioritized before implementation. We also found evidence of genetic erosion that in the long term might lead to inbreeding depression and loss of adaptive potential from these populations. Some of the potential causes were overfishing and the construction of hydropower dams, which should also be taken into consideration in future managing practices.

Conclusions

Our results were congruent with the hypothesis that anthropogenic activities affected the genetic structure of *O. niloticus* in East Africa. The genetic variation of some populations, especially from Lake Victoria, corresponded with possible hybridization of *O. niloticus* with native congeneric species, 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. 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 lower Nile (below Murchison Falls), Lake Kivu (Rwanda), Tanganyika and Baringo as well as the congenerics for a more comprehensive picture.

Methods

Sampling/study areas

Following the earlier sampling design from Tibihika, Waidbacher (40), we collected *O. niloticus* specimen from three waterbody types: a) those where *O. niloticus* is native, b) where it was introduced, and c) from fish farms (Figure 1). Most

samples were collected by local fishermen using gill nets set overnight, at the location of 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 waterbodies during several field excursions in the year, 2016. From every single fish, a muscle tissue sample (approx. 3 g) from the dorsal region was extracted, preserved in absolute ethanol 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 together 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. The non-native and farm locations 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) (40). 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

Genotyping

Genomic DNA extraction followed the modified MagSi-DNA Vegetal kit protocol, using magnetic beads (MagSi-DNA beadsMagnaMedics) and a magnetic separator, SL-MagSep96 (Steinbrenner, Germany). For a detailed description, see Tibihika, Curto (26). We used microsatellite markers (26), to which we added 15 extra primers (Table 2). The SSR primers were designed and tested analogous to Tibihika, Curto (26), using the same short gun 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 (26). The PCR products were then pooled and sent for paired-end 300bp sequencing in Illumina MiSeq, at the Genomics Service Unit in Ludwig Maximillian Universität, München, Germany. Allele calling was performed based on amplicon sequence information using the script from the SSR-GBS pipeline (https://github.com/mcurto/SSR-GBS-pipeline). For subsequent analyses, all loci and samples with missing genotypes ≥50% were excluded, leaving a total number of 40 markers (additional file, Table S1).

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

Genetic Structure

Genetic structure was first assessed by calculating the absolute number of different alleles and unbiased Nei's genetic distance between individuals. These statistics were obtained and visualized through Principal Coordinate Analysis (PCoA), conducted in GenAlex Version 6.5 (51). Genetic similarity between populations was evaluated by plotting a Neighbor-Net tree based on Nei's genetic distance using the program, SplitsTree4 version, 4.14.8 (52). We also constructed UPGMA dendrograms for making inferences on the hierarchical clustering between populations using PAUP* version 4.0a (53). Neighbor-Net tree and the UPGMA dendrogram were conducted with the inclusion of subpopulations, when applicable to evaluate possible substructure within the waterbodies. Genetic structure was further investigated using the program, STRUCTURE Version 2.3.4 (54). STRUCTURE clusters individuals into hypothetical populations through optimization of Hardy-Weinberg equilibrium (55). 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 (56), 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 (57) using the delta K (ΔK) statistic, which is the second order rate of change (lnP(D)) across successive K values (56, 58). 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 (59) available at http://clumpak.tau.ac.il/. Similar analyses were performed for Lake Victoria within populations.

Migration rates and number of migrants per generation (Nm)

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 (60). Here, the program was run for 200, 000,000 iterations, discarding the first 100,000,000 generations and sampling every 1000^{th} generation (61). Only results with a 95% confident 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 number of migrants (Nm) 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 (Fst) 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 (He), observed heterozygosity (Ho), number of alleles (Na), allelic richness (Ar), fixation index (Fst), private alleles, and Garza-Williamson index (G-W). Na, Fst, G-W and He per population were analyzed using the program Arlequin Version 3.5 (62). Ho, He, Na and PIC per locus were determined through Cervus version 3.0.7 (63). Ar was analyzed using the rare-faction algorithm implemented in Hp-rare program (64). 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 (65). To test whether the genetic diversity and differentiation of *O. niloticus* populations conform to isolation by distance (IBD), we plotted genetic distance (Fst) against the geographical distance (GGD in kilometers) and conducted correlation analyses using Mantel test (999 permutations) implemented in GenAlex Version 6.5 (51).

Declarations

List of abbreviations

Not applicable

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, therefore 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 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.Agnèse 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.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.

4.Worthington EB. A report on the fisheries of Uganda investigated by the Cambridge Expedition to the East African Lakes, 1930–31. 1932.

5.Agnèse J-F, Adépo-Gourène B, Pouyaud L. Natural hybridization in tilapias. Genetics and Aquaculture in Africa. 1998:95–105.

6.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.

7.Balirwa JS. The evolution of the fishery of Oreochromis niloticus (Pisces: Cichlidae) in Lake Victoria. Hydrobiologia. 1992;232(1):85–9.

8.Kaufman L. Catastrophic change in species-rich freshwater ecosystems. BioScience. 1992;42(11):846–58.

9.Welcomme R. Recent changes in the stocks of Tilapia in Lake Victoria. Nature. 1966;212(5057):52.

10.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.

11.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.

12.Zengeya TA, Booth AJ, Chimimba CT. Broad niche overlap between invasive Nile tilapia *Oreochromis niloticus* and indigenous congenerics in southern Africa: Should we be concerned? Entropy. 2015;17(7):4959–73.

13.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.

14.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.

15.Allendorf FW, Luikart G. Conservation and the genetics of populations. Mammalia. 2007;2007:189–97.

16.Vreven E, Teugels G, Adepo-Gourene B, Agnese J. Morphometric and allozyme variation in natural populations and cultured strains of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). Belgian Journal of Zoology (Belgium). 1998;128(1):23.

17.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.

18.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.

19.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).

20.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.

21.Wirgin II, Waldman JR. What DNA can do for you. Fisheries. 1994;19(7):16–27.

22.Okumuş İ, Çiftci 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.

23.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.

24.Hedrick PW. Conservation genetics: where are we now? Trends in Ecology and Evolution. 2001;16(11):629–36.

25.Liu Z, Cordes J. DNA marker technologies and their applications in aquaculture genetics. Aquaculture. 2004;238(1):1–37.

26.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 the genetic diversity of the East African Oreochromis niloticus. Conservation Genetics. 2018:1–16.

27.Ndiwa TC, Nyingi DW, Agnese J-F. An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Loboi drainage, Kenya. PloS One. 2014;9(9):1–15.

28.Nyingi D, De Vos L, Aman R, Agnèse J-F. 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. 2009;297(1–4):57–63.

29.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.

30.Frankham R, Briscoe DA, Ballou JD. Introduction to conservation genetics: Cambridge university press; 2002.

31.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.

32.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.

33.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.

34.Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, et al. Hybridization and extinction. Evolutionary Applications. 2016;9(7):892–908.

35.Rhymer JM, Simberloff D. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics. 1996;27(1):83–109.

36.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.

37.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.

38.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.

39.Dumont HJ. The Nile: origin, environments, limnology and human use: Springer Science & Business Media; 2009.

40.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.

41.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.

42.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.

43.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.

44.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.

45.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.

46.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.

47.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.

48.Allendorf FW, Leary RF, Spruell P, Wenburg JK. The problems with hybrids: setting conservation guidelines. Trends in Ecology & Evolution. 2001;16(11):613–22.

49.Roberts DG, Gray CA, West RJ, Ayre DJ. Marine genetic swamping: hybrids replace an obligately estuarine fish. Molecular Ecology. 2010;19(3):508–20.

50.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.

51.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.

52.Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution. 2005;23(2):254–67.

53.Swofford DL. Phylogenetic analysis using parsimony (* and other methods) PAUP* Version 4.0a. Sinauer Associates, Sunderland, Massachusetts; 2002.

54.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.

55.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.

56.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.

57.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.

58.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.

59.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.

60.Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. Genetics. 2003;163(3):1177–91.

61.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.

62.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.

63.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.

64.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.

65.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.

Tables

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.01.052060	E030.534640	618
Albert	Kyehooro	Ngege	16	Uganda	Indigenous	N.01.509900	E030.936100	615
Edward	Rwenshama	Ngege	27	Uganda	Indigenous	S.00.404590	E029.772830	908
Edward	Kazinga	Ngege	22	Uganda	Indigenous	S.00.207830	E029.892520	914
George	Hamukungu	Ngege	35	Uganda	Indigenous	S.00.017390	E030.086980	916
Kazinga Ch	Katungulu	Ngege	21	Uganda	Indigenous	S.00.125410	E030.047440	915
R. Nile (VN)	Kibuye	Ngege	24	Uganda	Indigenous*	N.01.187340	E032.968650	1062
Kyoga	Kibuye	Ngege	44	Uganda	Introduced	N.01.400280	E032.579490	1034
Kyoga	Bukungu	Ngege	22	Uganda	Introduced	N.01.438730	E032.868090	1045
Victoria	Kakyanga	Ngege	30	Uganda	Introduced	N.00.180790	E032.293320	1136
Victoria	Gaba	Ngege	26	Uganda	Introduced	N.00.258190	E032.637270	1146
Victoria	Masese	Ngege	23	Uganda	Introduced	N.00.436500	E033.240810	1136
Victoria	Sango Bay	Ngege	25	Uganda	Introduced	N.00.867720	E031.713320	1129
Victoria	Kamuwunga	Ngege	25	Uganda	Introduced	S.00.127470	E031.939990	1139
Mulehe	kisoro	Ngege	25	Uganda	Introduced	S.01.213450	E029.726680	1801
Kayumbu	Kisoro	Ngege	30	Uganda	Introduced	S.01.346790	E029.784460	1901
Rwitabingi	Kamuli	Ngege	29	Uganda	Fish farm	N.00.971160	E033.139240	1069
Bagena	Kisoro	Ngege	31	Uganda	Fish farm	S.01.256170	E029.736220	1857
Sindi	Kabale	Ngege	25	Uganda	Fish farm	S.01.175780	E030.061980	1733
Turkana	Longech	Ngege/Sato	35	Kenya	Indigenous	N.03.556170	E035.915990	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

Table 2 15 new primer pairs developed in the present study. Other 26 tested primers developed by (26) 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	TGCTAGCCCAGAGCATTTGA	(GAATA)6	425-478
Ti44	TGCTCCTGACTCAGCATCAC	GCAGCACTCTGACATGAAGC	(GAAAA)6	419-469
Ti49	TCGAAGTAGCGTGGAAAACCT	ACAACAACAACAGGTCGGGA	(TGT)8	395-403
Ti50	CCTGTGACAGACTGGTGACC	ACACTGATGCGGTTTACGGT	(ATGG)7	442-517
Ti51	TGCTAAACGCCAGCTGATGA	TTACCACACGATGTCGCAGG	(TGT)8	401-428
Ti52	GAGAAACGTCCAGTGGCAGA	TTTCGATCTGCTGCCCCTTT	(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	GAGCCGCCATAGTGTCACTT	CCTGCTCTCACTCAAAGAGGG	(ATCC)7	473-516
Ti61	GCTACACAGGAAAGCAGAGC	ACTCAATGCTGGACGTGACC	(TGGA)6	474-501

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

Figures

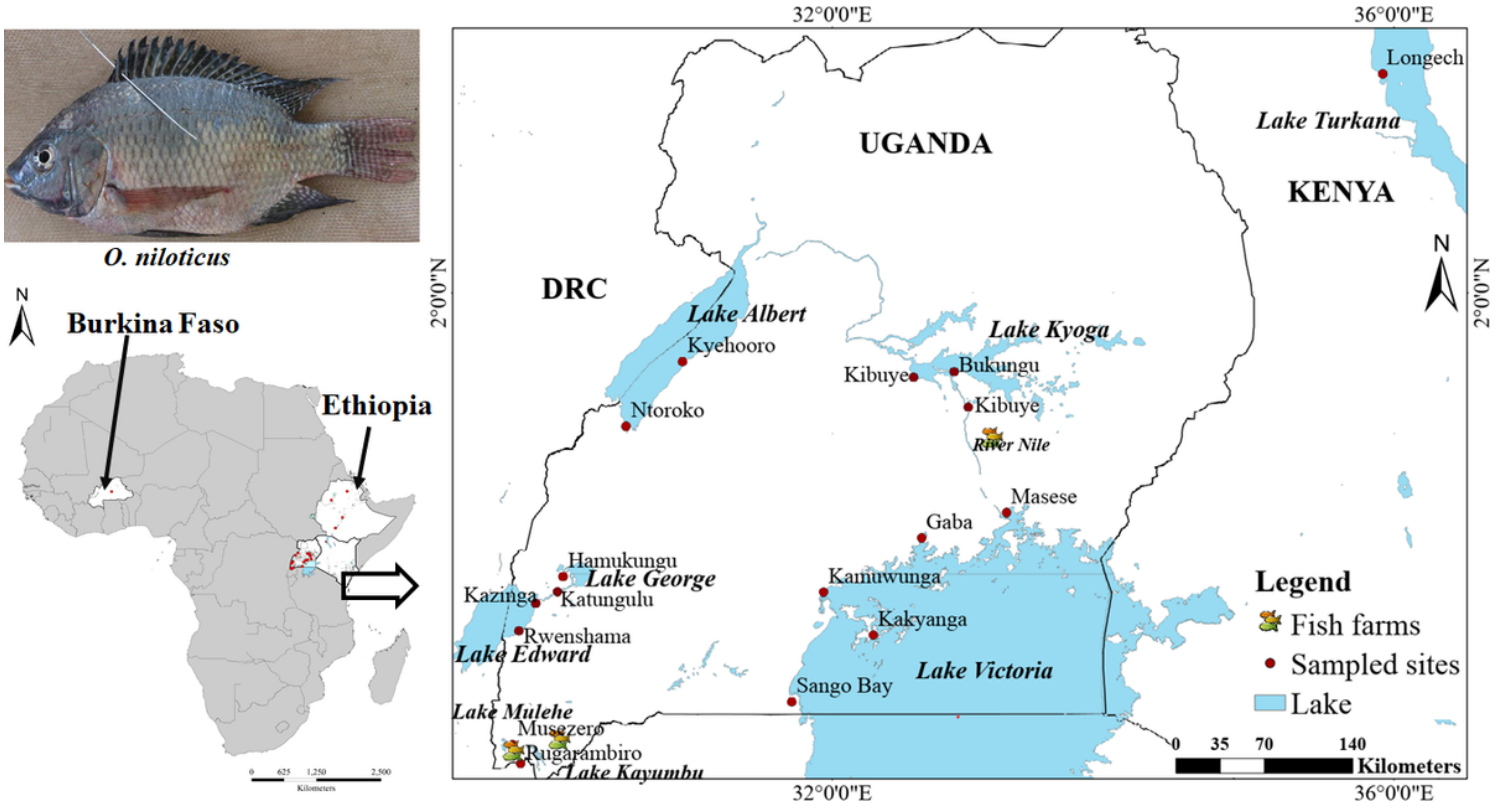

Figure 1

Sample sources and sizes. The top panel shows a map for the sampling locations with a focus on East Africa, Ethiopia, and Burkina Faso. The bottom tables show the details of the sampling sites and the total number of individuals collected (Left) and Lake Victoria sub-population (right). In the tables, Natives are indigenous O. niloticus populations, non-natives are introduced, and farms are the pond culture systems.

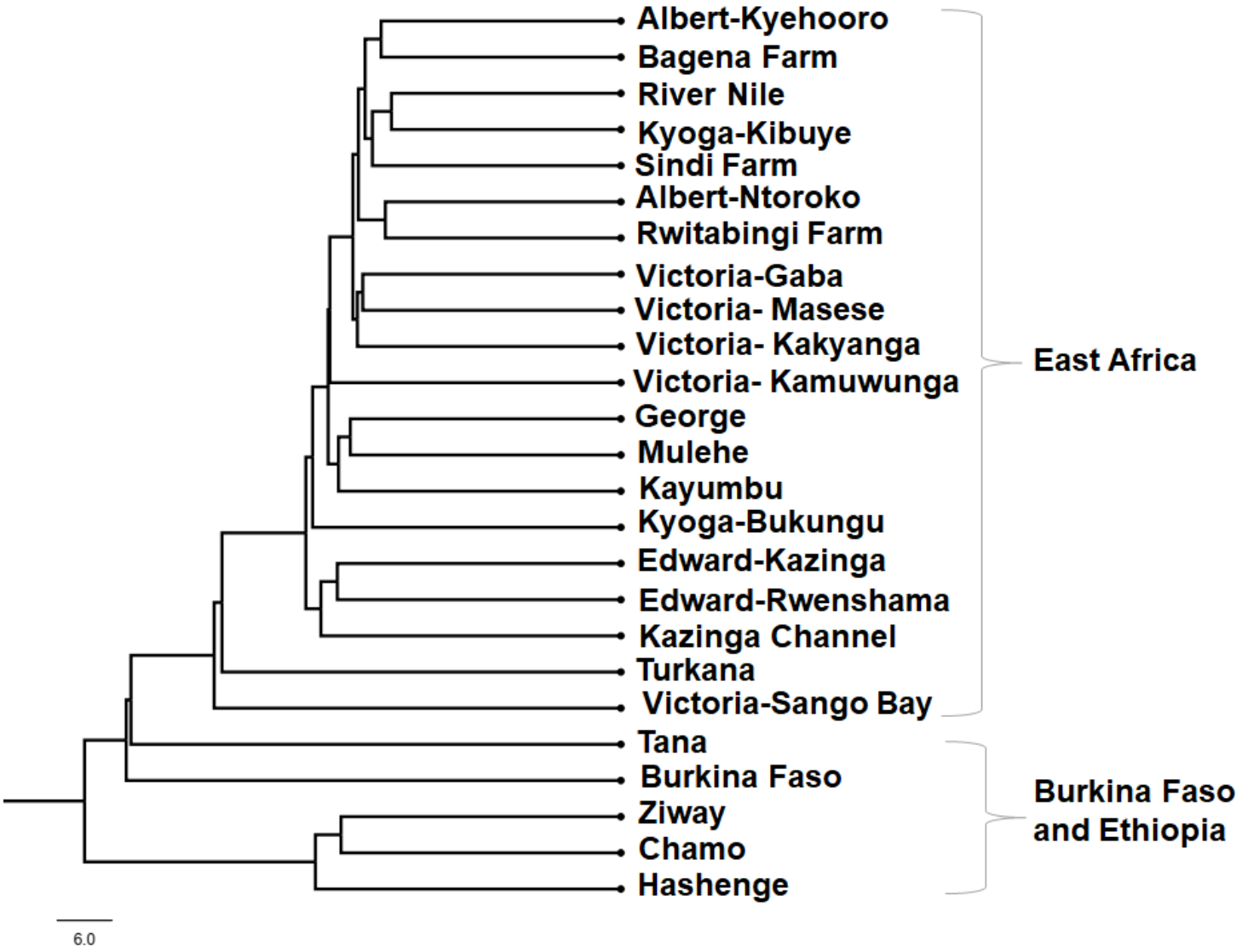

Figure 2

(Fig 2.1) Genetic structure of O. niloticus populations based on UPGMA dendrogram

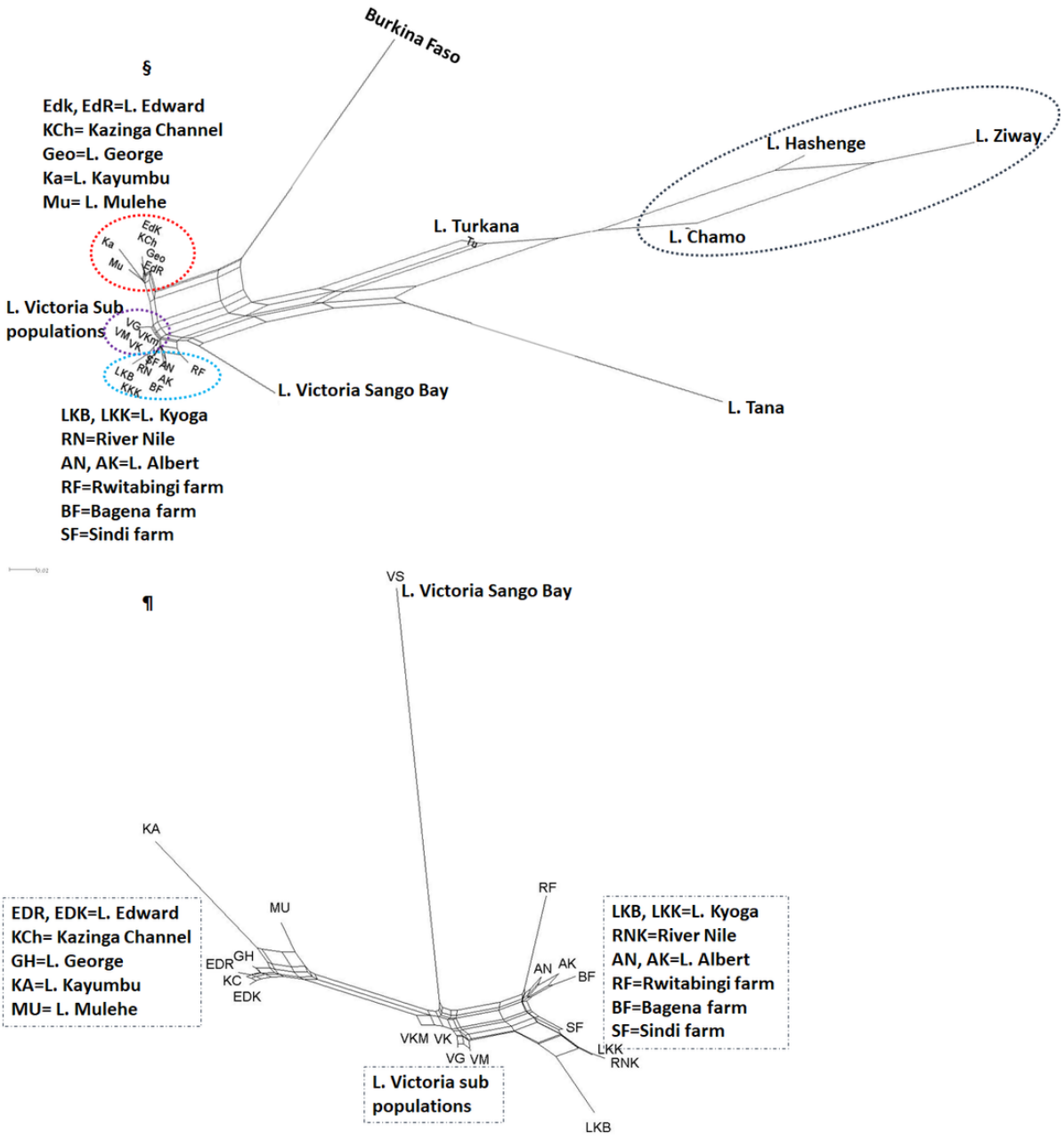

Figure 3

(Fig 2.2) Genetic structure based on unrooted network tree illustrating population relationships based on distance. § represents a tree for all the populations and ¶ for only the Ugandan populations. Dotted oval and rectangular shapes depict closely related genetic groups.

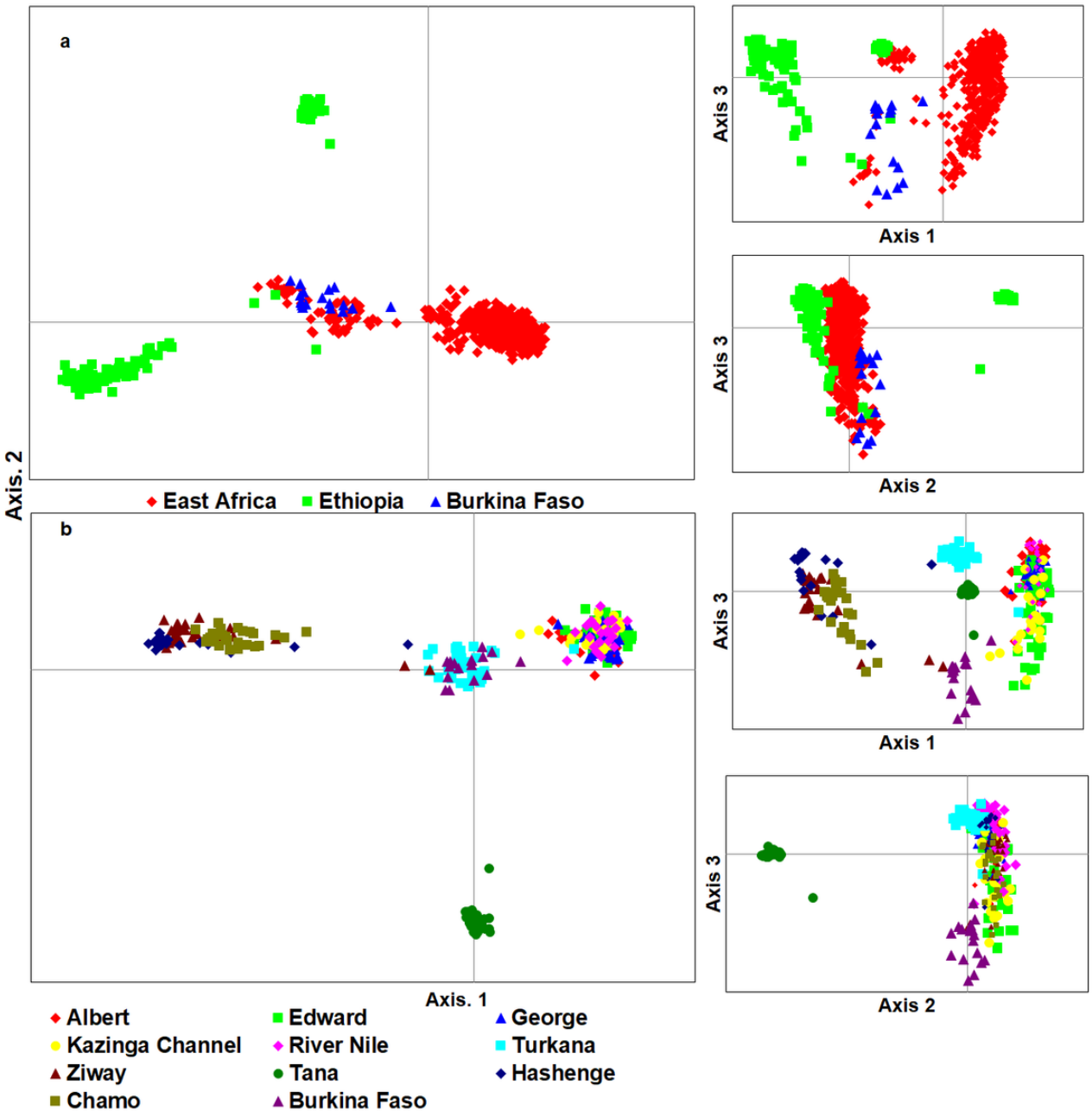

Figure 4

(Fig 3.1) 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

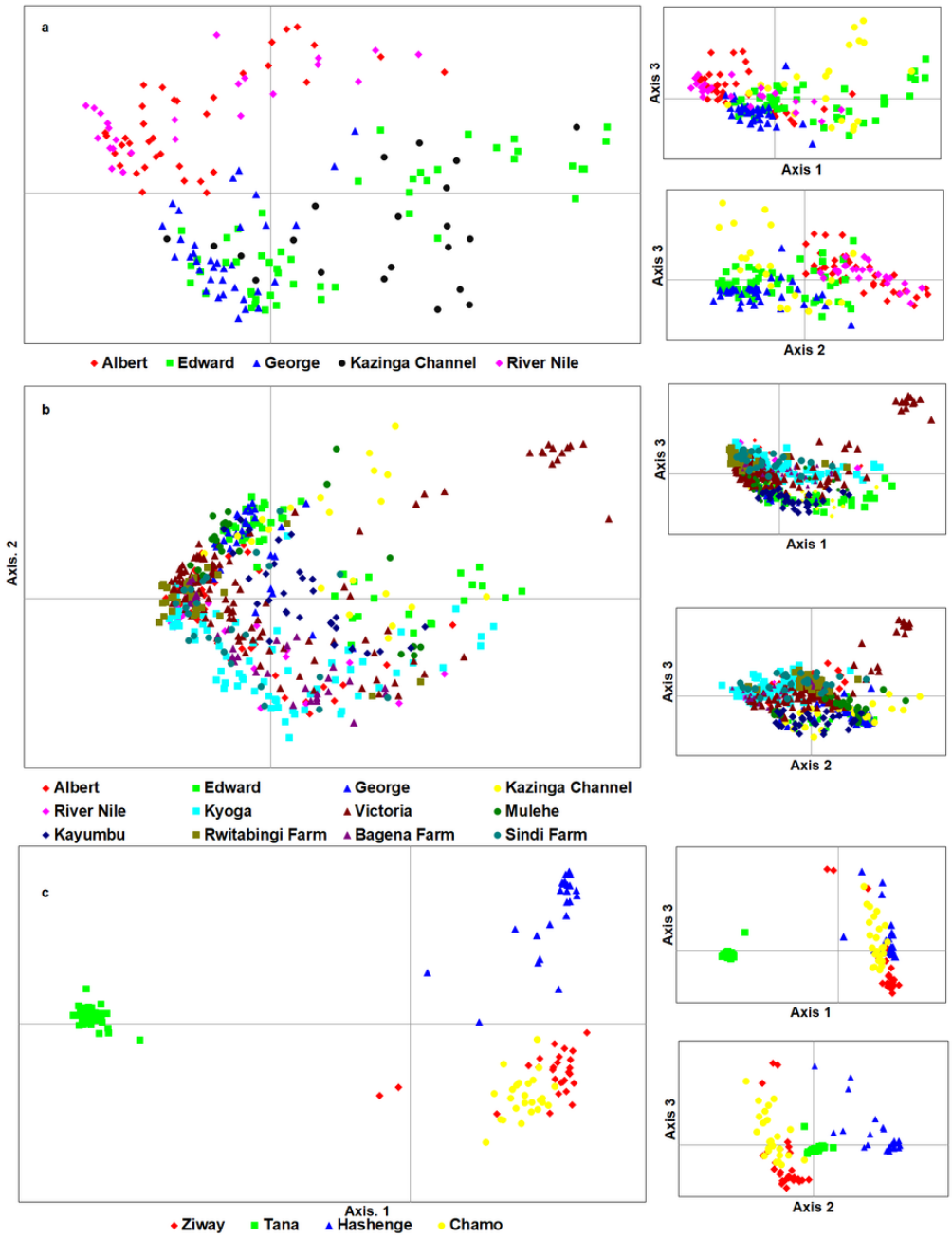

Figure 5

(Fig 3.2) 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

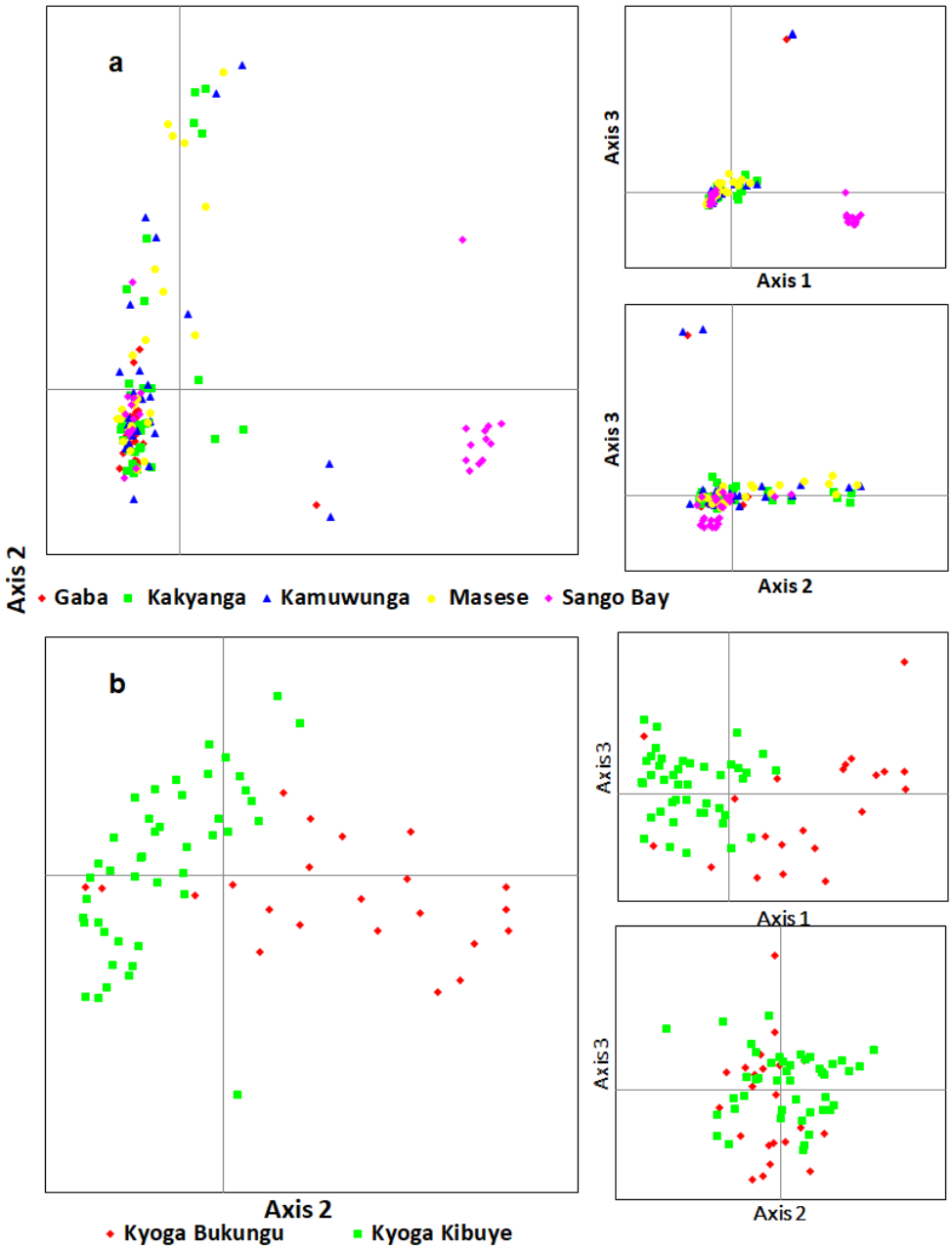

Figure 6

(Fig 3.3) Genetic scatter plots of O. niloticus exhibited by PCoA within Lakes Victoria (a) and Kyoga (b) populations

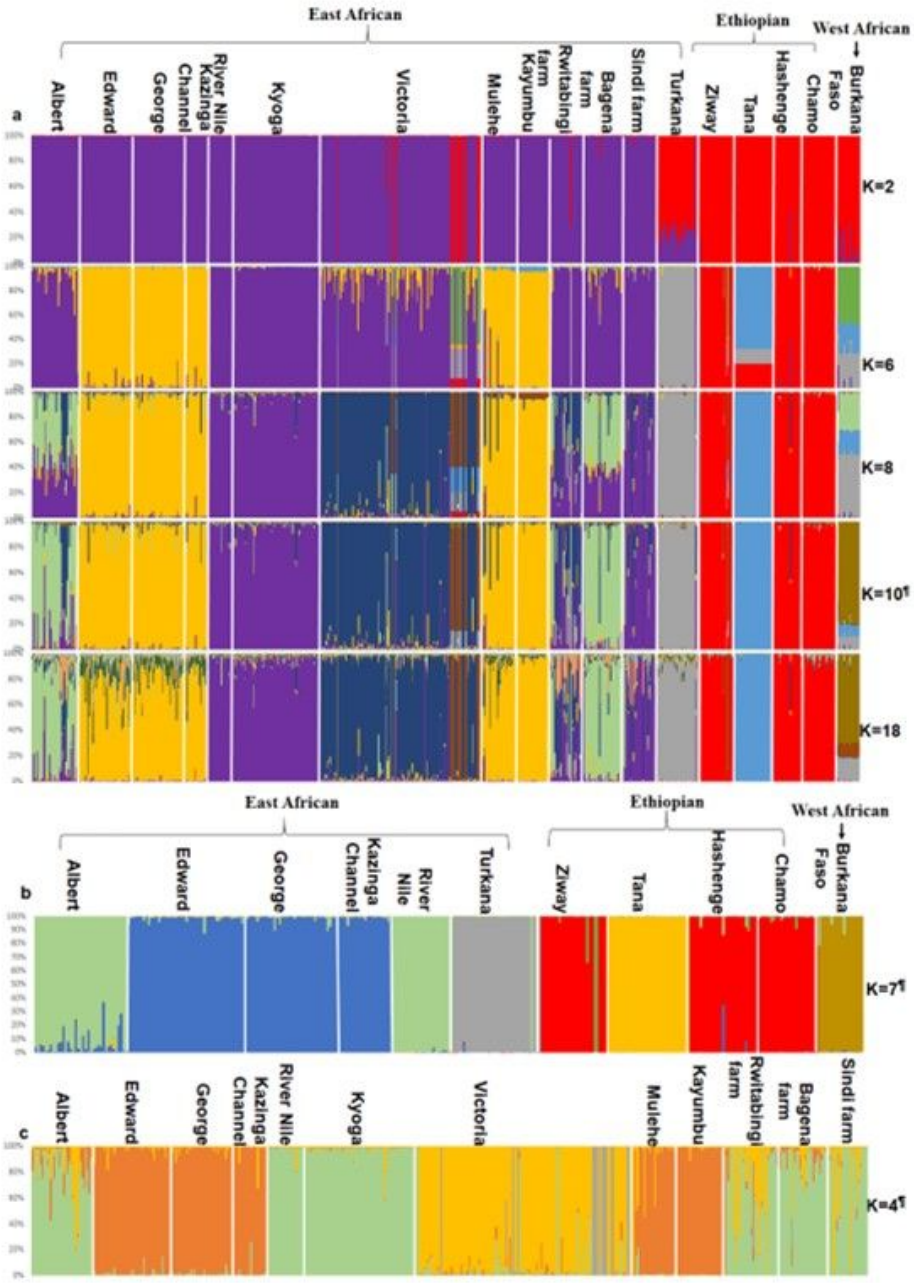

Figure 7

(Fig 4) 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. Ks with a superscript symbol (¶) indicates the optimal K values based on STRUCTURE HARVESTER analyses.

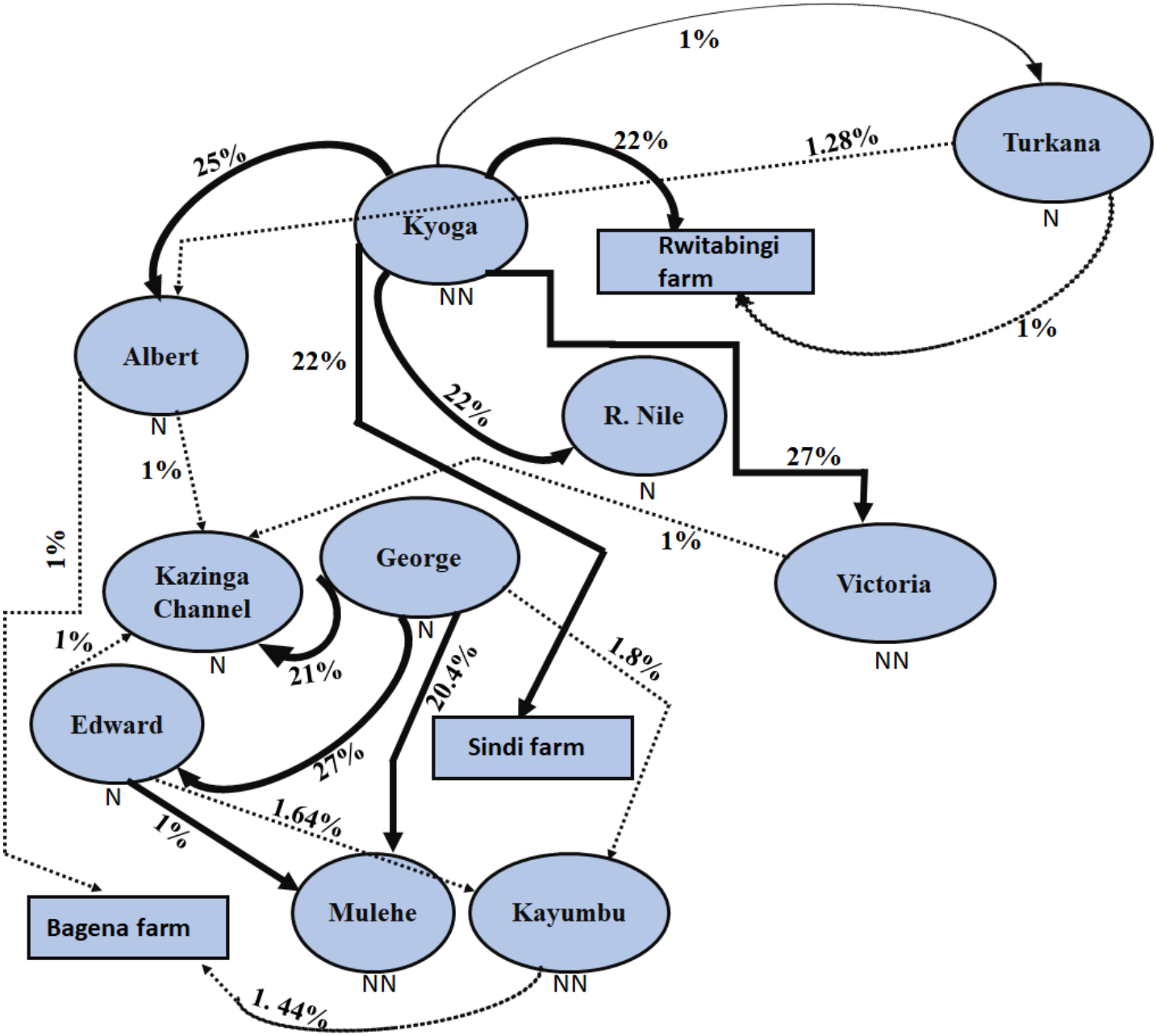

Figure 8

(Fig 5) Bayesian inference of recent migratory rates for the 13 East African O. niloticus populations. Oval light blue and rectangular 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 supplementary materials, Table S3

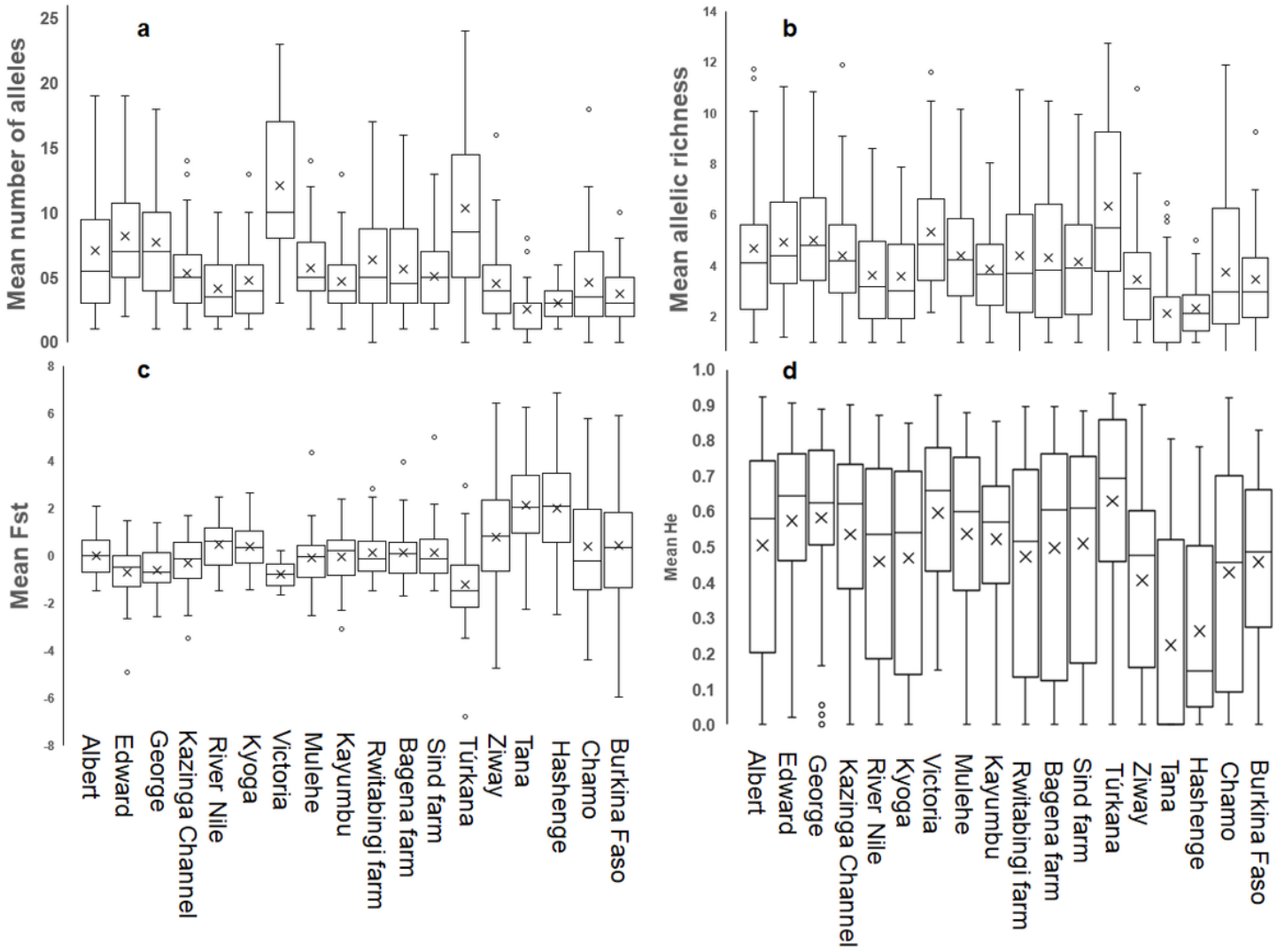

Figure 9

(Fig 6.1) Genetic diversity and differentiation indices. a number of alleles, b allelic richness, c fixation index (Fst) and d expected heterozygosity

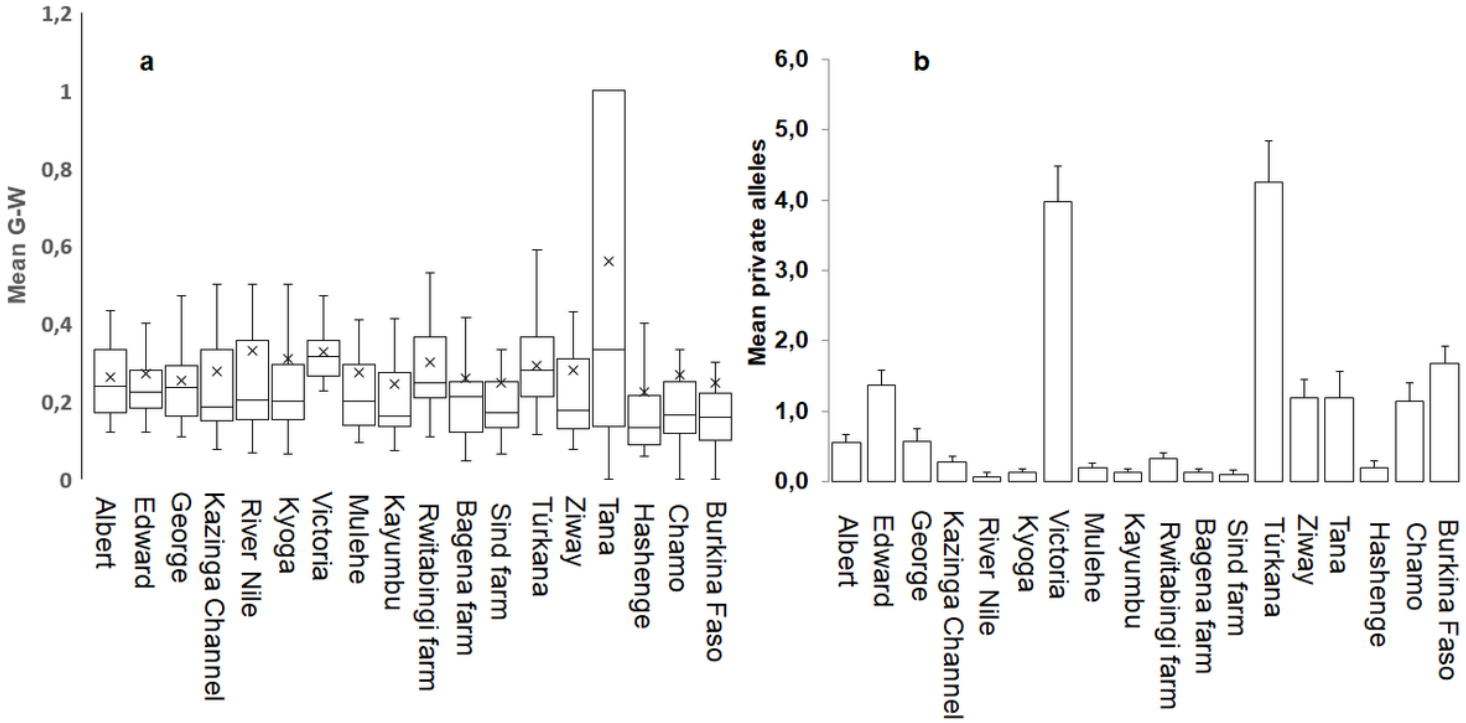

Figure 10

(Fig 6.2) Estimations of population bottleneck derived from Garza-Williamson Index (G-W) (a) and measure of genetic diversity based on private alleles (b)

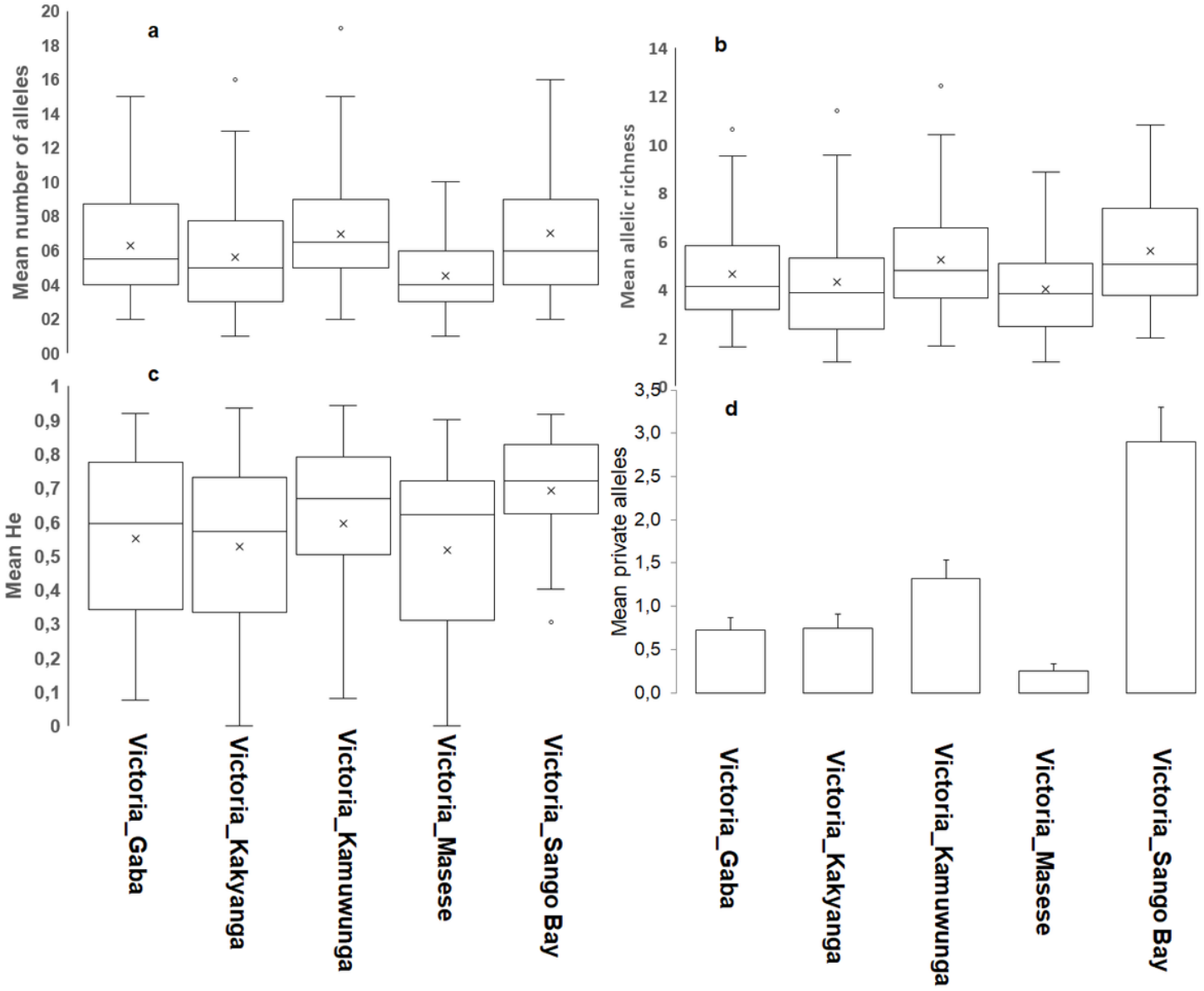

Figure 11

(Fig 6.3) Genetic diversity of Lake Victoria within the population. a number of alleles, b allelic richness c expected heterozygosity and d private alleles

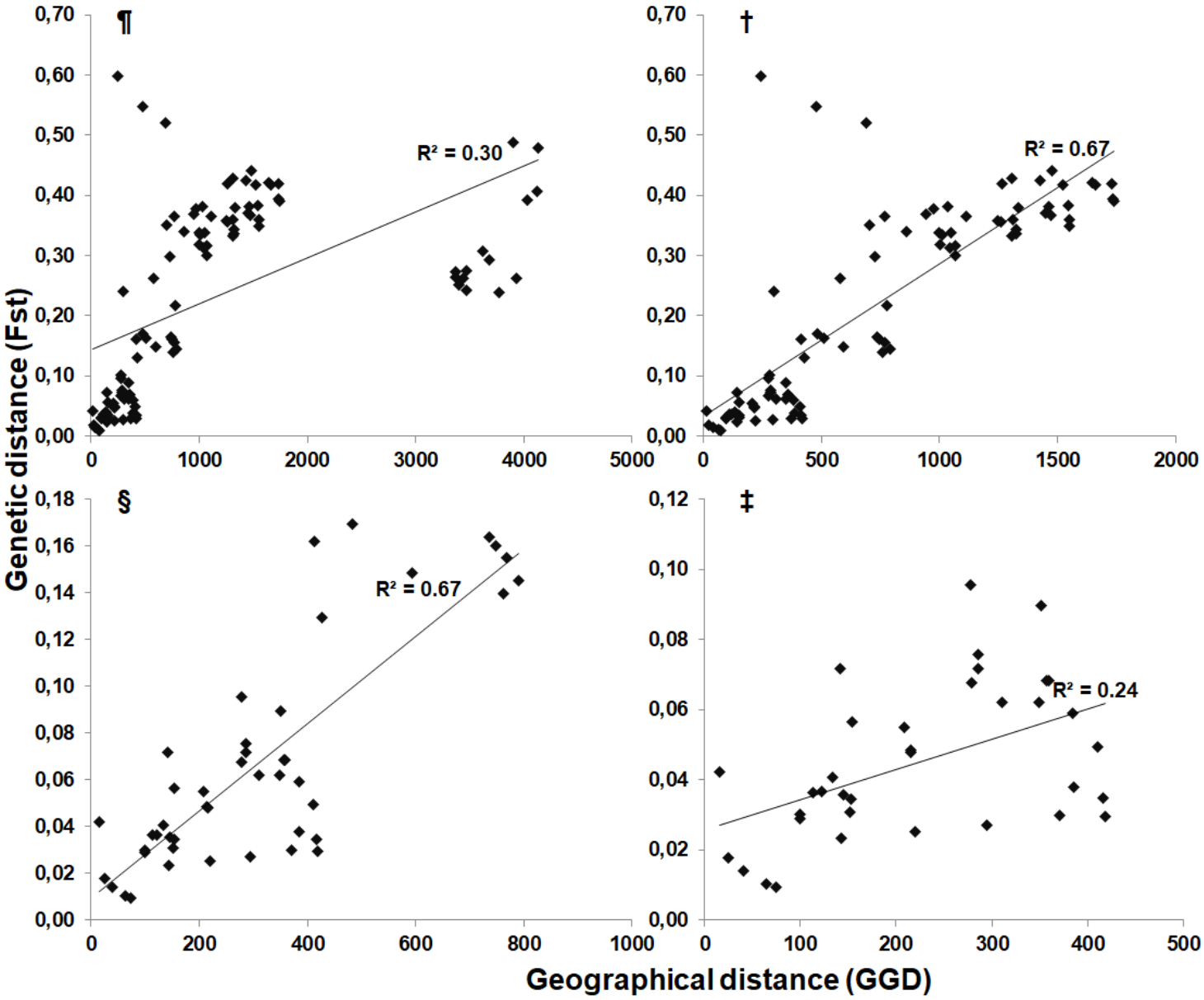

Figure 12

(Fig 7) Mantel tests for correlations between genetic distance (Fst) and geographical distance (GGD in km) for O. niloticus populations. ¶ represents isolation by distance (IBD) between all populations, † all populations without Burkina Faso, § East African, and ‡ only Ugandan populations

Supplementary Files

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

- supplement1.docx