



Universität für Bodenkultur Wien
Department für Integrative Biologie

Multi-level analysis on local adaptation of the East African Nile tilapia (*Oreochromis niloticus*, Linn, 1758) from native and non-native environments

Papius Dias Tibihika

Dissertation submitted for obtaining a doctorate degree at the University of Natural
Resources and Applied Life Sciences Vienna

Department of Integrative Biology and Biodiversity Research (DIB), Institute for Integrative
Nature Conservation Research

Thesis supervisor: Univ. Prof. Dipl.-Biol. Dr. rer. nat. Harald Meimberg

Co-supervisors: Ao. Prof. Dr. phil. Herwig Waidbacher, Dr. Manuel Curto, and Assoc. Prof. Dr.
Charles Masembe

Vienna, May 2019

Addresses

Univ. Prof. Dipl.-Biol. Dr. rer. nat. Harald Meimberg and Dr. Manuel Curto

Department of Integrative Biology and Biodiversity Research (DIB), Institute for Integrative Nature Conservation Research (INF)-University of Natural Resources and Applied Life Sciences Vienna

Ao. Univ. Prof. Dr. phil. Herwig Waidbacher,

Institute for Hydrobiology and Aquatic Ecosystems Management-University of Natural Resources and Applied Life Sciences Vienna (BOKU)

Assoc. Prof. Dr. Charles Masembe

Department of Zoology, Entomology and Fisheries Sciences-Makerere University Kampala, P. O. Box 7062, Kampala, Uganda



Nile tilapia (*Oreochromis niloticus*, Linnaeus 1758), from Lake Albert, East Africa (Author picture)

Acknowledgements

This Ph. D journey has been one of its kind that I would have neither walked alone nor obtained a grant if all was up to me solely. This was ideally a fulfilled God's promise and therefore, I would like to say, THANK YOU LORD.

First, I thank vehemently my supervisor; Prof. Harald Meimberg (Harry) who accepted me as his student, welcomed me to his Laboratory and further provided me a good office next to his. Harry, thank you also for coming up with new and interesting ideas from time to time while at the Institute and even when I was away in Uganda for field work. You streamlined my Ph. D research focus and it is what I am today because of you. Thank you for encouraging and guiding me when I could not see results especially when genotyping outputs could not materialise, despite spending funds. Thank you for dedicating much of your time during progress meetings, discussing my papers and the desire to see that my Ph. D journey was consistently pursued and goals achieved. Harry, thank you for your love, care and enormous support you gave me when my daughter was very ill. I remember when I told you this bad news, you quickly arranged for a flight and even gave me a lot of money for travel. I have no other form of thanking you other than saying "may the good Lord reward you for me" I have learnt a lot from you and my prayer is that, I keep working under your leadership for I still need your advanced scientific expertise. Apart from academic, thank you again for parenting me especially when we went out for drinks and eats, quite often you footed all my bills.

I would like to thank in a special way all the staff members of our Institute, more especially, Dr. Manuel Curto and Eva Dornstauder-Schrammel. Manuel and Eva, I can never forget my first-time encounter with you in 2016 when my supervisor handed me over to you and you quickly took me all over what you do in Meimberg lab. Since, 2016 until now, I have worked with you, but never did you complain that I am over occupying you, nor even hesitated to take care and work on my project. I owe you, Manuel and Eva, as you are a team of great minds. Manuel, thank for your bioinformatics skills and again for accepting to travel to Uganda, a period when I was in my last field work. Your presence in my country and your scientific mentorship, reminds me a lot that you are indeed more than a true friend to me. Manuel, you are a selfless, kind and generous man, thank you my friend for always helping me financially when I was in crisis. To my office-mates; Dr. Silvia Winter, Dr. Bernhard Splechtna, and Seiberl Margit thank you for being awesomely true friends. To all our institute staff members, Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn Kriechbaum Monika, Priv.-Doz. Dr. Kropf Matthias, Prochazka Barbara, Horner David Christopher, Kratschmer Sophie, Dr.nat.techn

Pachinger Baerbel, Krenn Philipina Yoko, Plenk Kristina and Abdalla Elzein, thank you for making the institute my home.

I am indebted to the staff of the Institute for Hydrobiology and Aquatic Ecosystems Management (BOKU), particularly Prof. Herwig Waidbacher, Gerold Winkler, Meulenbroek Paul, Haslinger Nina, Seebacher Martin and Drexler Silke-Silvia, as well as Dr. Peter Akoll of Makerere University for interviewing me in Uganda and henceforth recommending my Ph.D. application for a grant under the STRECAFISH project. I was further grateful for your recommendations towards my project funding, despite being above the 35-year age bracket required under APPEAR-OeAD funding. THANK YOU. Gerold, I would like to thank you in a special way for your generosity, selfless and whole heartedly contributions you gave me when my daughter was very ill in Uganda at a critical moment when I was preparing my thesis. I don't know how I can thank you, but the fact that we all believe in the Almighty Father, I say "God bless your efforts". Thank you again for your orientation during my first challenging days in Vienna, "I call you parents".

To my colleagues under the same STRECAFISH project, Esayas Alemayehu, Kevin Obiero and Zipporah Gichana, thank you for being good friends. Whenever I would be with you, you always made my life easy as you often reminded me that God is in control.

I would also like to acknowledge my colleagues in NARO (National Agricultural Research Organization) specifically National Fisheries Resources Research Institute (NaFiRRI), Dr. Richard Ogutu-Ohwayo, Dr. John Stephen Baliwa and Dr. Wilson Waiswa Mwanja, whose literature generally provided me with an enormous insight into the historical background of the East African Freshwater bodies, particularly the Nile tilapia fish species of my interest. Thank you. In the same respect, I am gratified by my mother Institution; National Agricultural Research Organization-Kachwekano Zonal Agricultural Research and Development Institute, for approving and releasing me for a three-year study leave. I would like also to thank the staff at Makerere University-Kampala, particularly Assoc. Prof. Charles Masembe for accepting to store my samples during field work excursions.

It is quite very difficult and disturbing to stay away from a family specifically when it is a young one, as in my case. I know I have indirectly tortured you greatly my dear wife, Ms. Doreen Kakama and my kids; Valerie Ayebare, Emmy Amutuheire, Leila Mercy and Kennedy Aheisibwe, for a long time being away from you. Great thanks for your patience and enduring the pain! To my late Grand Mother, Valeria Bakehemura, you brought me up in this world, but unfortunately you never lived

longer to enjoy or see the fruits from my hands, RIP Mukaaka. In this regard therefore, I would like to dedicate this thesis to all my family members.

This thesis was funded by the Austrian Partnership Programme in Higher Education and Research for Development – APPEAR through STRECAFISH project, a programme of the Austrian Development Cooperation (ADC) and implemented by the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH). Genotyping and bioinformatics were supported by the Department of Integrative Biology and Biodiversity Research-University of Natural Resources and Life Sciences Vienna (BOKU). Thesis printing was supported with funds from the Institute for Integrative Nature Conservation Research (INF)-University of Natural Resources and Applied Life Sciences Vienna (BOKU).

*Ask and it will be given to you; seek and you will find; knock and the door will be opened to you;
Matthew 7:7*

Abstract

The Sub-Saharan Great Lakes Region has a rich ichthyofauna most notably, the cichlids. Among the cichlids, the most important and commercially exploited are the tilapine species of which *O. niloticus* Linn, 1758 (Nile tilapia) is the most common. This vital resource has been subjected to immense anthropogenic threats primarily through fish translocations, overfishing, habitat destruction, climate change, alien invasive aquatic weeds etc. Hereby, the broader aim of this thesis was to investigate the impact of anthropogenic activities on the evolutionary trajectory of the East African *O. niloticus* particularly in nonnative habitats following the aforementioned activities in the region. This was investigated using the hypothesis that anthropogenic activities might have altered the natural morphology and genetic structure of *O. niloticus* through admixture and introgression with the indigenous congeneric species. Geometric morphometrics and molecular genetic markers were used to delineate the cichlid in the region from different East African localities, covering low and high-altitude environments. Samples were collected from three locality types; farms, lakes with natural but non-native (introduced) fish and lakes with native (indigenous) fish. The farms and non-native localities were only limited to Uganda. Additionally, samples at the continental level were collected, including populations from the West African country of Burkina Faso and Ethiopia to test and compare the genetic structure with the East African ones. For morphometric analysis only the East African populations were included. During geometric morphometrics analysis, two thin plate spline (Tps) programs; Tps Utility (TpsUtil) and Tps digitizer (TpsDig) were utilized. These programs were used to acquire landmarks on fish digital photographs and henceforth generation of x, y coordinates which were then imported into MorphoJ program for subsequent multivariate analyses (chapter 2). At molecular level, 64 microsatellites (SSR) loci were developed and 26 were employed to test the four East African/Uganda populations. Here, the latest genotyping by sequencing approach using SSR loci (SSR-GBS) was compared with the traditional fragmented length analysis by assessing the information content recovered with the two approaches (Chapter 3). SSR-GBS was then applied in Chapter four to investigate all the sub-Saharan Nile tilapia populations. A total of 40 SSR loci were used for amplifying all the present studied populations. Geometric morphometric analysis generally revealed that all the East African populations were somewhat homogeneous in shape, implying close relation. However, when principle component scores (PC1 and PC2) were plotted against the axes, it was clear that the non-natives; Victoria and Kyoga, were morphologically divergent from for instance some native populations (Kazinga Channel and Edward). But the non-native Victoria and Kyoga appeared undifferentiated from some natives like Albert, George, River Nile and Turkana. The shape variation among the East African Nile tilapia populations was suspected

to admixture and hybridization, partly induced by anthropogenic activities. However, other environmental cues like water turbulences (for example in River Nile), predation threat effects (e.g. Nile perch in native lakes; Albert and Turkana), severe and adverse environmental conditions (like declined water levels aggravated by high temperatures in L. Turkana), were taken into consideration. Apparently, use of SSR-GBS genotyping approach yielded more information content than the traditional based amplicon length procedures. For instance, SSR-GBS method did better to reveal the genetic diversity of the East African *O. niloticus* than the conventional amplicon length (AL). These results were consistent with a detailed genetic structure generated by SSR-GBS which was undetectable with AL. Additionally, based on polymorphic information content (PIC), it was evident that the SSR-GBS showed more polymorphic loci compared to AL. More importantly the markers developed in this thesis and the genotyping techniques were demonstrated how they may contribute useful information for further investigations regarding the East African Nile tilapia for sustainable exploitation and conservation. By evaluating the genetic diversity and variability patterns among Nile tilapia populations at sub-Saharan level, the Ethiopian populations were indicated more divergent than the East African ones. This was particularly prominent for lake Tana which indicated a high level of distinctiveness. In East Africa, genetic structure analyses revealed four *O. niloticus* groups which were consistently supported by Fst results. Moreover, the East African populations genetically differed from their Ethiopian and West African counterparts. Also, *O. niloticus* from Lake Turkana (Kenya) was isolated, while the Ugandan populations were grouped according geographical locations/water catchments and anthropogenic activities i. Signals of genetic erosion were detected in Ethiopian and some East African populations being likely a consequence of either bottlenecks associated with fragmentation or founder effects. *O. niloticus* in Lakes Tana, Victoria and the cluster of Lakes Edward, George as well as Kazinga Channel, may require further exploration for taxonomic recognition. Apart from Lake Tana, the level of isolation of the cichlid in Lake Victoria and the genetic group in the western Rift Valley Lakes, have not been reported before. This thesis indicates that anthropogenic activities have contributed to a significant change in the genotypic composition of *O. niloticus* which may compromise the species' sustainability. Over all, the molecular analyses were congruent to the geometric morphometrics, which shows the robustness and importance of the current approach in understanding the state the of the East African *O. niloticus*. In the ongoing work, out of scope of this thesis, these populations are followed at mitochondrial genome level to precisely investigate the extent of introgression and phyletic patterns.

Keywords: Nile tilapia. Geometric morphometrics. Genotyping by sequencing. Admixture. Geneflow

Zusammenfassung

Die subsaharische Region der Großen Seen verfügt über eine reiche Fischfauna; insbesondere sind hier die Cichliden zu erwähnen. Innerhalb der Cichliden sind die wichtigsten und kommerziell am meisten genutzten die Tilapia Arten, von denen *O. niloticus* Linn, 1758 (Nile tilapia) der häufigste ist. Diese wichtige Ressource ist vor allem durch Translokation und Überfischung sowie weitere anthropogene Faktoren wie Lebensraumzerstörung, Klimawandel, invasive Wasserunkräuter usw. einer enormen Bedrohung ausgesetzt. Deswegen ist das übergeordnete Ziel dieser Arbeit, den Einfluss menschlicher Aktivitäten auf die evolutionäre Entwicklung des ostafrikanischen *O. niloticus* speziell in nicht ursprünglichen, durch die vorher genannten regionalen Aktivitäten veränderten Lebensräumen zu untersuchen. Dies wurde anhand der Hypothese untersucht, dass anthropogene Aktivitäten die natürliche Morphologie und genetische Struktur von *O. niloticus* durch Besatz und Vermischung mit heimischen gleichartigen Arten durch Introgression verändert haben könnten. In dieser Arbeit setzte ich geometrische Morphometrie und molekulargenetische Marker ein, um den Buntbarsch der Region von anderen ostafrikanischen Lokaltäten, welche in höher oder tiefer liegenden Regionen vorkommen, zu unterscheiden. Stichproben wurden von drei unterschiedlichen Typen von Lokaltäten genommen: Fischfarmen, Seen mit natürlichen aber nicht heimischen (also eingeführten) Fischen und Seen mit heimischen (indigenen) Fischen. Fischfarmen und die nicht heimischen Lokaltäten beschränkten sich auf Uganda. Zusätzlich bezog ich auf kontinentaler Ebene weitere Stichproben von Populationen aus Ländern West-Afrikas wie Burkina Faso und Äthiopien, um die genetische Struktur mit den ostafrikanischen Individuen zu vergleichen. Für die morphometrischen Analysen wurden nur die Ost-Afrikanischen Populationen verwendet. Für die geometrisch-morphometrischen Analysen verwendete ich zwei Thin-Plate-Spline (Tps) Programme; Tps Utility (TpsUtil) and Tps digitizer (TpsDig). Diese Programme wurden verwendet, um Orientierungspunkte auf digitalen Fotos von Fischen zu bestimmen und dafür schließlich x, y Koordinaten zu erhalten, die dann in das Programm MorphoJ importiert wurden für die weitere multivariate Analyse (Kapitel 2). Auf molekularer Ebene entwickelte ich zuerst 64 Mikrosatelliten (SSR) Loci, 26 davon wurden verwendet um die vier ostafrikanischen/Uganda Populationen zu testen. Dafür testete ich das neueste Verfahren der Genotypisierung durch Sequenzierung mit SSR loci (SSR-GBS) und verglich die Ergebnisse mit der traditionellen Genotypisierung anhand der Länge von Fragmenten. (Kapitel 3). SSR-GBS wurde dann in Kapitel vier verwendet, um alle Sub-Saharische Nil Tilapia Populationen zu untersuchen. Insgesamt wurden 40 SSR-Loci zur Amplifikation aller vorliegenden untersuchten Populationen verwendet. Die geometrisch-morphometrische Analyse zeigte für alle ostafrikanischen Populationen eine im allgemeinen relativ einheitliche Form. Eine grafische

Darstellung der Hauptkomponentenanalyse zeigte aber, dass sich die nicht-heimischen Populationen von Lake Vicotria und Lake Kyoga von einigen heimischen Populationen (von Kazinga Kanal und Lake Edward) morphologisch unterscheiden, von anderen heimischen Populationen (Lake Albert, Lake George, dem Nil und Turkana) aber nicht. Diese Unterschiede in der Form innerhalb der ostafrikanischen Nile tilapia Populationen wurden auf Vermischung und Hybridisierung teilweise durch menschliche Aktivitäten zurückgeführt. Aber darüber hinaus wurden auch andere Umweltfaktoren wie Turbulenzen (z.B. im Nilfluss), Effekte der Bedrohung durch Prädation (z.B. Nilbarsch in den heimischen Seen Albert und Turkana), raue und ungünstige Umweltbedingungen (wie durch hohe Temperaturen hervorgerufene niedrige Wasserstände in L. Turkana) in Betracht gezogen. Offensichtlich hat der SSR-GBS Genotypisierungsansatz genauere Informationen geliefert als die traditionelle Methode basierend auf Amplicon Länge (AL). Zum Beispiel erkannte die SSR-GBS Methode die genetische Diversität innerhalb der ostafrikanischen *O. niloticus* besser als die herkömmliche Methode. Das zeigte sich auch dadurch, dass es mit SSR-GBS möglich war eine detaillierte genetische Struktur festzustellen, die von der AL Methode nicht erkannt wurde. Darüber hinaus zeigte die SSR-GBS mehr polymorphe Loci als AL, basierend auf dem polymorphen Informationsgehalt (polymorphic information content (PIC)). Noch wichtiger erscheint aber die Tatsache, dass demonstriert werden konnte, wie die in dieser Arbeit entwickelten Marker und die Genotypisierungs Techniken wichtige Information zu weiteren Untersuchungen des ostafrikanischen Nil Tilapia die nachhaltige Nutzung und den Schutz betreffend beitragen könnten. Wenn die genetische Diversität und Variabilitätsmuster der Nile tilapia Populationen auf subsaharischem Maßstab betrachtet wurden, zeigten die äthiopischen Populationen größere Unterschiede als die ostafrikanischen. Das war besonders deutlich für die Population von Lake Tana, welche ein hohes Maß an Eigenständigkeit erkennen ließ. Auf Basis der genetischen Struktur können in Ostafrika vier Gruppen unterschieden werden, in Übereinstimmung mit Fst Ergebnissen. Darüber hinaus unterschieden sich die ostafrikanischen Populationen von ihren äthiopischen und westafrikanischen Pendants. *O. niloticus* von Lake Turkana (Kenya) erscheint in der Analyse isoliert aufgrund seiner Lage. Im Gegensatz dazu zeigten die Populationen in Uganda eine enge Beziehung, obgleich geographischer Ort, Wasser Einzugsgebiete und anthropogene Aktivitäten einige genetische Cluster beeinflusst haben. Anzeichen genetischer Erosion wurde in Äthiopien und einigen Ostafrikanischen Populationen gefunden, wahrscheinlich als Resultat eines Flaschenhalseffektes aufgrund von Fragmentierung oder eines Gründereffektes. Zur taxonomischen Bestimmung von *O. niloticus* in den Lakes Tana, Victoria und den Clustern von Lake Edward, Lake George und Kazinga Kanal erscheinen weitere Untersuchungen notwendig. Außer für Lake Tana wurde zuvor noch nie über eine ähnlich große

Isolation berichtet, wie in dieser Arbeit für die Buntbarsche in Lake Victoria und die genetische Gruppe in den Seen des westlichen Grabenbruchs. Diese Arbeit zeigt, dass anthropogene Aktivitäten zu einer signifikanten Veränderung der genotypischen Zusammensetzung von *O. niloticus* beigetragen haben, was die Nachhaltigkeit der Art gefährden könnte. Im Großen und Ganzen stimmten die Ergebnisse der molekularen Analysen und der geometrisch-morphometrischen überein, was die Robustheit und die Wichtigkeit dieses Ansatzes unterstreicht zu einem besseren Verständnis des Status des ostafrikanischen *O. niloticus* beizutragen. Die weiteren Arbeiten, außerhalb dieser Dissertation, werden diese Populationen auf Basis mitochondrialer Genome weiter verfolgt um das Ausmaß der Introgression und die phyletischen Muster genau feststellen zu können.

Included papers, also herein referred to as subsequent chapters

Tibihika PD, Waidbacher H, Masembe C, Curto M, Sabatino S, Alemayehu E, Meulenbroek P, Akoll P, Meimberg H (2018) Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa. *Environmental Biology of Fishes* 101 (3):363–381. <https://doi.org/10.1007/s10641-017-0704-0>

Tibihika PD, Curto M, Dornstauder-Schrammel E, Winter S, Alemayehu E, Waidbacher H, Meimberg H (2018) Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*. 1-16. <https://doi.org/10.1007/s10592-018-1136-x>

Tibihika PD, Curto M, Alemayehu E, Waidbacher H, Masembe C, Meimberg H. Genetic structure evidence of anthropogenic threats in the East African fisheries: the paradigm of Nile tilapia (*Oreochromis niloticus*, L. 1758) (under revision)

Table of Contents

Addresses.....	ii
Acknowledgements.....	iii
Abstract.....	vii
Zusammenfassung.....	ix
Included papers, also herein referred to as subsequent chapters	xii
1. Chapter 1	3
1.1. General introduction.....	3
1.2. Overview of the East African Nile tilapia and congeners	6
1.3. Traditional attempted studies on Nile tilapia	8
1.3.1. Advances in morphometric and genetic markers	10
1.4. A brief overview of the waterbodies included in this thesis	12
1.4.1. Lakes Kayumbu and Mulehe system	12
1.4.2. Lakes Edward and George system	15
1.4.3. Lake Albert system.....	15
1.4.4. Lakes Victoria, Kyoga and River Nile system	19
1.4.5. Lake Turkana system	20
1.4.6. The Ethiopian Lakes; Tana, Chamo, Hashenge and Ziway, and Burkina Faso- Loumbila reservoir	22
1.5. Significance and general objective of the thesis	24
1.5.1. Organization of chapters and general research questions	26
1.6. Thesis general structure and overview of the research approaches	27
1.7. References	32
2. Chapter 2	39
1 st Paper. Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (<i>Oreochromis niloticus</i> , L. 1758) in East Africa	39
3. Chapter 3	59
2 nd paper. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African <i>Oreochromis niloticus</i>	59
4. Chapter 4	76
3 rd paper. Genetic structure evidence of anthropogenic threats in the East African fisheries: the paradigm of Nile tilapia (<i>Oreochromis niloticus</i>)	76
5. Chapter 5	114
5.1 General discussion.....	114

5.1.1	Morphological divergence and local adaptation of the East African Nile tilapia (<i>Oreochromis niloticus</i> , L. 1758).....	115
5.1.2	Genetic diversity and differentiation of Nile tilapia among regions.....	118
5.1.3	Nile tilapia gene flow patterns delineated by migration rates/anthropogenic fish translocations.....	120
5.1.4	Nile tilapia genetic structure in the context of ecogeographical location	124
5.1.5	Use of microsatellite loci (SSR) and the strengths of genotyping by sequencing (GBS) under next generation sequencing (NGS)	127
6.	Chapter 6.....	129
6.1.	Conclusions and outlook	129
6.2.	References	131
	Publications	135
	About the author.....	136
	Appendix.....	137
	Supplementary materials for chapter three	137
	Supplementary materials for chapter four.....	152

1. Chapter 1

1.1. General introduction

Several fish genera and subgenera exist in the cichlidae family which are collectively classified as tilapias (George, 1995). Tilapias exhibit a high degree of parental care which translates into good survival rates of the hatchlings and perhaps explains their ability to adapt to various habitats. On the basis of the reproductive behaviors, tilapiine fishes are classified into three main genera: *Tilapia*, *Sarotherodon* and *Oreochromis*, but also a rare fourth genera exists – *Danakilia* (George, 1995; Trewavas, 1983). Generally, the spawning strategies of these genera differ from each other. For example, the reproductive strategy of the genus *Tilapia* involves both, male and female brood parents in preparing spawning nests in shallow waters where the female deposits the eggs and the male subsequently fertilizes them. Representatives of *Tilapia* generally remain monogamous during the breeding period. After spawning, both parents guard the nests and ventilate the fertilized eggs, and, after hatching, they continue to protect their offspring until they are able to live independently (George, 1995). This spawning behavior is typical for *Tilapia* (e.g., *Tilapia zillii*) which are classified as substrate spawners (Trewavas, 1983). *Oreochromis* and *Sarotherodon*, in contrast, belong to mouthbrooders. However, also these two genera, *Oreochromis* and *Sarotherodon*, exhibit differences. For instance, *Oreochromis* males prepare circular and bowl-like nests in shallow waters where they remain on guard until visited by gravid females (George, 1995). After fertilization, the females pick the eggs in their buccal cavity where the eggs are incubated until hatching. Once the females take up the eggs, they separate from the males, where the latter remain on the nests, waiting to mate with additional females – a practice of polygamy (George, 1995). Following hatching of eggs, the females continue to protect the fry by providing refugia into their buccal cavity, following any threats, until they are able to swim freely and independently. *Oreochromis niloticus* is an example of this spectacular breeding strategy and their closest relatives include *O. esculentus* (Trewavas, 1983). Other closely-related mouthbrooders include *O. mossambicus*, *O. aureus* (blue tilapia) and *O. macrochir* (tassled tilapia) (George, 1995). The genus *Sarotherodon* differs from *Oreochromis* in breeding strategy, whereby they are either biparental or one of the parents take care of the fertilized eggs and so to the fry. For instance, both parents in *Sarotherodon galilaeus* carry the eggs in their buccal cavity after fertilization, whereas males in *Sarotherodon melanotheron* are solely responsible for this task (George, 1995; Trewavas, 1983).

Generally, tilapiines are native to Africa where they are essentially a food source. Due to their importance for food, they have been dubbed the “aquatic chicken” or “miracle fish” (Eknath et al.,

1998; George, 1995). Despite being native to Africa, tilapias have been widely distributed and currently constitute one of the most important farmed fish species worldwide (Diana, 2009; Vicente & Fonseca-Alves, 2013). Nile tilapia, in particular, are more widely translocated than other tilapia species due to their vital attributes, such as fast growth rates, resistance to diseases and pests, high fecundity, or broad food spectrum. These traits make Nile tilapia suitable for aquaculture production systems (Baliwa, 1998; Njiru et al., 2006; Vicente & Fonseca-Alves, 2013). Although carp (*Cyprinus carpio*) is still the most-farmed fish worldwide, the above-mentioned unique traits of Nile tilapia might allow the cichlid to surpass the cyprinid in aquaculture production (Vicente & Fonseca-Alves, 2013; Zengeya, Booth, & Chimimba, 2015). Nonetheless, presently Nile tilapia is one of the most important species in East Africa for both, capture fisheries and aquaculture (Njiru et al., 2006).

Traditionally, East Africa comprised of three countries; Uganda, Kenya and Tanzania (Otieno & Awange, 2006), despite three more countries – South Sudan, Rwanda and Burundi –which have been recently included. The region is well-known for its enormous inland freshwater resources, and is, therefore, called the Great Lakes Region (Ogutu-Ohwayo, Hecky, Cohen, & Kaufman, 1997; Salzburger, Van Bocxlaer, & Cohen, 2014). Moreover, East Africa is a hotspot for endemic ichthyofauna and if sustainably managed, the East African Great Lakes would significantly contribute to global fish productivity and supply (Lowe-McConnell, 2009).

The East African freshwater fisheries have for a long time been a vital and cheap source of aquatic proteins for the rapidly growing riparian human populations (Odongkara, Abila, & Onyango, 2005; Scullion, 2007). In addition, the region's freshwater ecosystems significantly contribute to overall employment. In Uganda, for example, the fisheries industry directly employs more than 700,000 people and more than 1.2 million people rely either directly or indirectly on fishery related products (Scullion, 2007). In Tanzania, more than 4 million people are either directly or indirectly employed in the fisheries sector (Odongkara et al., 2005). The East African inland fisheries predominantly rely on the major freshwater bodies, of which Lake Victoria, the world's second largest freshwater body, is the most famous (Baliwa et al., 2003; Getabu, Tumwebaze, & MacLennan, 2003). The contribution to foreign exchange from Lake Victoria capture fisheries alone ranged from 50–112 million US\$ annually for Uganda, Kenya and Tanzania (Odongkara et al., 2005). However, the regional aquatic ecosystems have undergone drastic and dramatic changes, in particular to Lakes Victoria, Kyoga and Nabugabo basins, following various anthropogenic activities (Ogutu-Ohwayo et al., 1997).

Lake Victoria is famous for its large native fish biodiversity, including hundreds of cichlid species, more than 40 cyprinids, catfish and other species (Witte, Wanink, & Kishe-Machumu, 2007). However, influx of multiple anthropogenic activities, such as overfishing, habitat destruction, non-native and invasive water hyacinth (*Eichhornia crassipes*) and eutrophication, have altered the biotic community in this ecosystem (Baliirwa, 1998; Baliirwa et al., 2003; Ogutu-Ohwayo, 1990; Ogutu-Ohwayo et al., 1997). Throughout the 20th century, overfishing increased and was especially enormous in the 1950s and 1960s. During this period, already significant evidence of species mass overexploitation was reported (Baliirwa et al., 2003; Ogutu-Ohwayo, 1990). By 1970s, many of the indigenous and eminent fish species which comprised of *O. esculentus*, *O. variabilis*, *Bagrus docmak*, *Labeo victorianus* and *Mormyrus kannume*, had greatly declined (Baliirwa et al., 2003). Despite overfishing, one additional major factor in the East African freshwater fisheries was the fish introductions/translocations. For instance, between 1950 and 1960s, many species, including Nile perch (*Lates niloticus*), Nile tilapia (*O. niloticus*), *O. leucosticus* and *Tilapia zilli*, were transferred from Lake Albert into Lake Victoria, Kyoga, Nabugabo, as well as several other systems (Baliirwa, 1992; Ogutu-Ohwayo, 1990). Generally, the motive for these introductions was to elevate fish productivity, but with separate roles per species. For example; 1) Nile perch was introduced to predate and convert the less preferred haplochromine cichlids into better table size fish, 2) *T. zilli* was introduced to utilize the aquatic vegetation, perhaps which resulted from eutrophication and 3) Nile tilapia was introduced to supplement the dwindled/overfished native stocks (Ogutu-Ohwayo, 1990). Nevertheless, other reports indicate that Nile tilapia was introduced accidentally or mistakenly for *T. zillii* (Trewavas, 1983).

Albeit many of the indigenous East African fish stocks had already declined due to overfishing, the introduction of alien species lead to a dramatic disappearance of native tilapiines such as *O. variabilis* and *O. esculentus* (Ogutu-Ohwayo, 1990; Ogutu-Ohwayo et al., 1997). Lake Victoria, which once was home to hundreds of indigenous fishes, currently is inhabited by only three species: the native pelagic minnow/small cyprinid *Rastrineobola argentea* – locally called Omena (Kenya), Mukene (Uganda) and Dagaa (Tanzania) – and the two introduced species, Nile tilapia and Nile perch (Baliirwa et al., 2003; Ogutu-Ohwayo, 1990). Although some stakeholders condemned the introduction of Nile perch and Nile tilapia, others welcomed the alien species, which they considered as saviors based on the improved fishery productivity in terms of catch size (Kaufman, 1992). Overall, the collapse of Nile tilapia congeners was attributed to a combination of multiple factors such as overfishing, eutrophication, competition, admixture and predation (Baliirwa, 1992; Mwanja, 2000; Ogutu-Ohwayo, 1990; Ogutu-Ohwayo et al., 1997). It remains an open question, however, why Nile tilapia successfully endured the aforementioned pressures

which possibly led to the collapse of its relatives. One would expect, for example, if predation pressure from Nile perch indeed contributed to the disappearance of the indigenous species, Nile tilapia to be similarly affected. However, this was not the case. Possible reasons for the perseverance of Nile tilapia include; the cichlid has originally coexisted with Nile perch in native environments such as Lakes Albert and Turkana, and this may have stimulated Nile tilapia to develop predator-evading strategies (Baliirwa, 1994; Ogutu-Ohwayo, 1990). Additionally, Nile tilapia grows faster consistent with a large size, more fecund, omnivorous or habitat/trophic diversification which provides a higher reproductive success when compared to the indigenous species (Baliirwa, 1992; Ogutu-Ohwayo, 1990).

Morphologically, Nile tilapia resembles most of the other cichlids, but the species can easily be identified by the interrupted lateral line. Also, Nile tilapia has a long dorsal fin which is sub-divided into portions; the forward is heavily spiny compared to the ray-finned rear part. But one critical feature that distinguishes Nile tilapia from other tilapiines is the presence of the vertical bars (sometimes called bands) on the caudal fin (Popma & Masser, 1999).

Albeit Nile perch and Nile tilapia currently dominate the East African commercial fisheries, the levels of overexploitation intensify annually mainly due to human population explosion consistent with high market fish demand (Njiru et al., 2007). This has been further exacerbated by the high and reliable foreign markets, for the Nile perch's white fillet in many European countries (Kirema Mukasa, 2012). Consequently, Nile tilapia remains a vital fish for catering all the local domestic needs and inclusive in all aspects of aquaculture. Thus, the East African Nile tilapia is a fish of great economic importance to the local populations, and studies focusing on the evolutionary responses to the anthropogenic threats may be vital for contributing to the sustainable exploitation and management of the species

1.2. Overview of the East African Nile tilapia and congenics

In Uganda, Nile tilapia is native to Lakes Albert, Edward, George, and the lower River, in Kenya, the species naturally inhabits Lakes Turkana and Baringo, in Tanzania Lake Tanganyika, and in Rwanda Lake Kivu (Baliirwa, 1992; Ogutu-Ohwayo, 1990; Trewavas, 1983). Nevertheless, the distribution of the species in the region seems complex and confounding based on various non-native habitats occupied by the cichlid (Baliirwa, 1992; Ogutu-Ohwayo, 1990; Ogutu-Ohwayo et al., 1997). Originally, the species never existed in Lakes Victoria, Kyoga and numerous regional satellites lakes, which were inhabited by native *O. variabilis* and *O. esculentus* (Ngege) tilapiines. These two native tilapias were important for the local commercial fishery (Baliirwa, 1992; Kaufman, 1992; Lowe-McConnell, 1975; Ogutu-Ohwayo, 1990; Pullin, 1988). Compared to Nile tilapia, *O.*

variabilis is rather a small sized fish but for consumption purposes, *O. esculentus* is the most preferred and considered the most delicious tilapiine (Balirwa et al., 2003; Kaufman, 1992). In terms of aquaculture, the growth rates of *O. esculentus* is generally slower than Nile tilapia. *O. esculentus* has also been reported to forage mostly on phytoplankton and can effectively provide an important ecological control for the algal blooms in the Lake Victoria catchment (Balirwa, 1992; Kaufman, 1992). Despite the ecological importance and food resource from *O. esculentus*, it is probable that, the anthropogenic threats on the species in L. Victoria and perhaps in Kyoga were evidenced way-back in the 1920s through overfishing (Anderson, 1961; Dumont, 2009). Overexploitation of the native cichlids was further exacerbated in the 1950s when controversial regulatory policies recommended the use of smaller gill net sizes (< 12.7 cm) (Anderson, 1961). Apart from the anthropogenic impacts, the native littoral tilapiines (e.g. *O. variabilis*) were further hit by heavy rainfall events in the 1960s which led to the destruction of the spawning and nursery grounds (Kaufman, 1992). This natural calamity had negative effects on fish populations and perhaps contributed to the decline of the native species. The massive decline in fish catches of the important Ngege fishery must have triggered new ideas such as the introduction of alien fish species, such as Nile tilapia and Nile perch, into the Lake Victoria basin (Balirwa, 1994; Beverton, 1959; Kaufman, 1992; Ogutu-Ohwayo, 1990).

The dynamic and multifaceted distribution of Nile tilapia in East Africa complicates the determination and identification of genetic resources for management and conservation. In fact, it is alleged that what is observed as Nile tilapia especially in Lakes Victoria and Kyoga might be something different (Pullin, 1988). The introduction of Nile tilapia into these waterbodies may have forced the contact with native species and hence promoted introgressive hybridization or introgressive segregation in the species. Through the recent expansion of fish hatcheries and aquaculture production systems in East African (Rothuis et al., 2014), more feral populations might be prevalent and intensifying this effect. The outcomes of these interactions might be detrimental for the future of the species (Harrison & Larson, 2014; Mooney & Cleland, 2001). In general, hybridization may be useful for taxa adaptation especially in new environments because of the establishment of heterosis traits allowing niche expansion (Epifanio & Philipp, 2000). However, hybrid characters may have a lower fitness than the parental populations and if introgression proceeds, extinction of the descendants may follow, through outbreeding depression (Agnèse, Adépo-Gourène, & Pouyaud, 1998; Epifanio & Philipp, 2000). Agnèse et al. (1998) further elucidates that, hybrids may inherit lower adaptive values than the pure ancestral lines and may eventually extinct following prior disappearance of the parents. Invasion of Nile tilapia into the non-native environments and the subsequent disappearance of native species through the combined

effects of hybridization and competition has been reported in South Africa (Zengeya, Robertson, Booth, & Chimimba, 2013). It is widely known that, because Nile tilapia is considered omnivorous, its influx into habitats previously occupied by other species may result into competition and even displacement of the native biota. Nile tilapia is tolerant to different environmental conditions and is a prolific breeder, which might also explain why this cichlid is invasive (Vicente & Fonseca-Alves, 2013). It should be noted that, the extent of Nile tilapia admixture with the congeneric species or outbreeding, aggravated by anthropogenic activities in the Lake Victoria basin and other satellites, is less documented. Although Baliwa (1992) reported some form of hybridization within the genus *Oreochromis* in L. Victorian waters, this was based on visual observations without empirical investigations.

Apart from stocking of Nile tilapia in Lakes Victoria and Kyoga, the species has been transplanted into numerous naturally unproductive minor lakes. These programs did not succeed regarding fish performance as fish sizes are below marketable standards (personal observation). The unexpected unproductiveness of the stocked satellite waters might be a result of maladaptation due to founder effects which might be a pressing concern. Furthermore, overfishing of East African Nile tilapia is immense and could have also impacted demographic bottlenecks on the populations in respect to the declined stock sizes (Nei, Maruyama, & Chakraborty, 1975; Spencer, Nelgel, & Leberg, 2000; Tanya & Kumar, 2010). In this context, the diminishing populations may lead to increased homozygosity or decreased heterozygosity (inbreeding), hence reducing overall genetic diversity and leading to unfit or less adapted generations (Halleran, 2003; Keller & Waller, 2002). Although it can be suspected that, in capture fisheries, Nile tilapia might undergo bottleneck effects, analogous consequences might also be prevalent among this species in aquaculture facilities.

Aquaculture, and especially the pond production systems, has been in the region for more than five decades (Dadzie, 1992). Despite interventions like water quality control measures, available complete feeds and improved culture production systems, the enterprise is still underperforming. The underperformance of farmed Nile tilapia might, however, be a consequence of unfit genotypes, which arise from culture systems especially when species are in captivity for many generations. This is because, lengthy captivity of Nile tilapia, leads to build-up of inbreeding as the populations are always small and isolated (Biswas, Morita, Yoshizaki, Maita, & Takeuchi, 2005; Langen, Schwarzer, Kullmann, Bakker, & Thünken, 2011).

1.3. Traditional attempted studies on Nile tilapia

The earliest investigations on Nile tilapia in Africa generally involved phenotypic methodologies based on biometrics, with relatively few molecular genetic studies. For example, the African

tilapiine, morphologically were comprehensively studied by Trewavas (1983), as well as by Vreven, Teugels, Adepo-Gourene, and Agnese (1998), who employed traditional methods. Traditional morphometrics and meristics involve linear measurements on organisms like head length, orbit size, body depth, total length, standard length, and counts like number of teeth, scales, vertebral column etc. respectively (Cadrin, 2000). Trewavas (1983) biometrically categorized Nile tilapia into seven sub species. On the other hand, Vreven et al. (1998) generally studied one population (Lake Edward) from Uganda. These authors reported that, Nile tilapia populations from the lower Egyptian River Nile were related to the populations from L. Edward, herein referred to as *O. niloticus eduardianus*. Although, traditional morphometric approaches have been widely employed to delineate fish and other organisms, these methods are considered less informative and relatively biased (Maderbacher et al., 2008; Petrtyl, Kalous, & Memiş, 2014; Wilson, MacLeod, & Humphrey, 2008). Additionally, the scope and sample size of the aforementioned morphometric studies on Nile tilapia were limited to a few waterbodies. For instance, although Vreven et al. (1998) supplemented the morphometrics with Allozyme markers, their work was focused mainly on the West and North African (Egypt) populations and apart from Lake Edward, provided limited attention on the East African ones. Furthermore, albeit Seyoum and Kornfield (1992) revealed convincing results regarding the Ethiopian Nile tilapia identification, their small sample size ($n = 3$), might have been inadequate to allow solid deductions.

Despite the increased awareness on the importance of molecular data (Park & Moran, 1994), generally few studies have focused on the genetic diversity patterns of the East African Nile tilapia. For instance, Agnèse, Adépo-Gourène, Abban, and Fermon (1997); Vreven et al. (1998); Mwanja et al. (1996) studied the genetics of Nile tilapia populations using allozymes with restriction fragment length polymorphism (RFLP) on mitochondrial DNA (mtDNA), allozyme and random amplified polymorphic DNA (RAPD) respectively. However, these fingerprint methods have limited information content (Liu & Cordes, 2004; Miah et al., 2013). In this context, divergent morphotypes and genetic structure among the East African Nile tilapia populations might be predominant and not well documented.

Generally, the use of molecular genetic markers in identifying fish stocks has proven valuable for many countries in the management and conservation of organisms. Although, no single marker is fundamentally perfect, they vary in capabilities and thus their choice of placement may depend on the reliable information content derived from various analyses. The use of RAPD markers for example, has been criticized that their assays are usually unreproducible due to low annealing temperatures during PCR amplifications (Miah et al., 2013). Wirgin and Waldman (1994)

emphasize that, because of their irreproducibility and patterns of dominant inheritance, the applicability of RAPD markers in the fisheries science and other fields may be limited. In addition, Liu and Cordes (2004) reiterate that though band intensity may differ with the use of RAPD markers, variations in PCR efficiency renders scoring of band intensities problematic. Other studies regarding fisheries and aquaculture also reported that RAPD and other traditional markers like allozymes may never have the capacity to determine for instance population structure, stock identification, parentage and pedigree analyses, gene mapping etc. (Okumuş & Çiftci, 2003). It should also be noted that, most of the genetic markers utilized in the earlier works on Nile tilapia, are bound to around 10 loci which represents a small fraction of the genome (Hoffman et al., 2014). The low available molecular data on the East African Nile tilapia, may imply that that, genetically the cichlid in the region is not well studied, and hence may require the need of more robust loci and subsequent advanced genotyping platforms to infer informed conclusions on the populations.

1.3.1. Advances in morphometric and genetic markers

Advanced approaches in morphological studies such as geometric morphometrics have proven to be very powerful to catalogue populations (Breno, Leirs, & Van Dongen, 2011). Geometric morphometrics is defined as the study of form in two or three dimensional spaces which induces in-depth investigation of morphological modelling in organisms (Bookstein, 1982). Geometric morphometrics allow the visualization and quantification of morphological differences in individuals represented by deformations of landmarks (Lms) in a coordinate space (Zelditch, Swiderski, & Sheets, 2012). Several studies comparing traditional biometrics (morphometrics and meristics) and geometric morphometrics have revealed that the latter is more powerful and hence more reliable. For example, geometric morphometrics were more effective in detecting skull morphological variation of *Mastomys natalensis* (Rodentia: Muridae) than the traditional methods (Breno et al., 2011). Although geometric morphometric approaches are considered adequately informative, it is likely that the combination with highly variable molecular markers, e.g. microsatellites (Hedrick, 2001), may provide an even better diagnostic tool for investigating population demography (Lenstra et al., 2012). However, such an integrative analytical approach has not been comprehensively used for the East African Nile tilapia.

Microsatellites are vital markers for answering many biological questions (Sunnucks, 2000). Indeed, microsatellites are one of the most employed markers in population genetics as they are highly reproducible, polymorphic, co-dominant and widely distributed throughout the genome of many organisms (Chistiakov, Hellemans, & Volckaert, 2006). Microsatellites have also been credited as the best choice and ideal markers for aquaculture as they are informative in many

aspects such as genetic monitoring of stock changes, parentage and pedigree analysis for selective breeding, and general genome mapping (Mojekwu & Anumudu, 2013). Comparative analysis between RAPD and microsatellite markers indicated that the latter was more informative to describe the genetic structure of Common carp (*Cyprinus carpio*), and hence, recommended as the best choice for the assessment of fish culture stocks (Bártfai et al., 2003). Moreover, with the implementation of Next Generation Sequencing technologies (NGS) in microsatellites discovery (Curto et al., 2013; Deck et al., 2016), genotyping (De Barba et al., 2016), and the high throughput use of these markers became economically feasible.

Generally, NGS has fundamentally revolutionized sequence data collection where millions of amplicon reads can be generated at the same time rather than only 96 reads generated when using the traditional capillary-based sequencing techniques (Mardis, 2008; Vartia et al., 2016). NGS has simplified the process of sampling the whole genome and, consequently, the technology easily enables the uncovering of cryptic evolutionary responses (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013). Apart from this, NGS has popularized molecular genetic studies by allowing the use of wide range of genomic protocols, including RAD, ddRAD, target sequencing etc. which is collectively referred to as genotyping by sequence (GBS) (Davey et al., 2011; Narum et al., 2013; Vartia et al., 2016). GBS approaches, allow the targeted subset of the genome to be sequenced with NGS rather than the entire genome or even in non-model organisms hence greatly reducing costs and time (Narum et al., 2013). GBS approaches use a variety of genome downscaling techniques, such as restriction enzymes, DNA capture through probes, and amplification (Davey et al., 2011). One core merit of GBS in the evolutionary studies is the ability to generate large quantities of data sets which enhances the statistical power and allows a greater genome representation (Narum et al., 2013; Vartia et al., 2016). Many of these approaches focus majorly on single-nucleotide polymorphism (SNP), which is replacing traditional microsatellite fragment length genotyping in population genetics. GBS can also overcome some of the limitations of microsatellites such as the required expertise and intensive work during genotyping, null alleles, allele dropout and stuttering during PCR amplification etc. (Dakin & Avise, 2004; Estoup, Jarne, & Cornuet, 2002). In this context one outstanding advantage of microsatellites (SSR) over SNP chips is the generation of a high statistical power per locus due to the large number of alleles that can be recovered (Haas & Payseur, 2011; Vartia et al., 2016). Although many authors have quite often reported that SSR can be problematic during genotyping, the underlying challenges of these markers have been generally mitigated through GBS technologies (Vartia et al., 2016). For example, contrary to the conventional sequence length (amplicon length or allele size length), SSR-GBS eliminates the challenges of size-homoplasmy as genotyping incorporates the underlying

whole sequence information (whole amplicon information) (Pasqualotto, Denning, & Anderson, 2007).

In this thesis, one of the aims was to employ microsatellite genotyping by sequencing (SSR-GBS) to investigate the demographic composition of the East African Nile tilapia from native and non-native waterbodies, and fish farms. However, I also collected Nile tilapia samples beyond East Africa: from Ethiopia and Burkina Faso (West Africa). This was essential for comparison since for example, the Ethiopian fisheries subsector has not yet been affected by intense anthropogenic activities as the East African one. The following sections present and elaborate on the studied waterbodies, whereby each system is briefly described.

1.4. A brief overview of the waterbodies included in this thesis

The East African inland freshwater bodies include, Lakes Victoria, Turkana, Albert, Edward, George, Kyoga and Tanganyika as well as River Nile, but with many other numerous rivers, satellite or minor lakes and wetlands (Figure 1) (Ogutu-Ohwayo & Baliwa, 2006). Apart from Lake Victoria and Kyoga, most of the East African inland lakes are located in the Eastern and Western Rift Valleys. The largest lake in the Eastern Rift Valley is Lake Turkana. The rest of the major lakes are located in the Western Rift valley (Figure 1). It is believed that all the East African inland waterbodies used to be interconnected but later separated due the tectonic movements. For example, the Western Rift valley lakes which range from Tanganyika through Kivu and then Kigezi highland lakes (Bunyonyi, Mulehe, Kayumbu, Mutanda) to lakes Edward and Albert were once connected (Green, 2009). But, the Southern lakes (Kivu and Tanganyika), were disconnected from the Northern ones (Kigezi lakes, Edward, George and Albert) by the formation of the Virunga Volcanic mountains (Green, 2009). Although Lake Victoria is not located in either of the Rift Valleys, it was perhaps interconnected with the Western Rift Valley through the Katonga valley and other rivers that now drain into the lake (Green, 2009; Hughes & Hughes, 1992).

1.4.1. Lakes Kayumbu and Mulehe system

Lakes Kayumbu and Mulehe (Figure 1.1), including Lakes Mutanda, Bunyonyi and Chahafi, form a network of five mountainous waterbodies located in south-western Uganda. All these minor lakes are at an altitude above 1,800 m a.s.l. and are located at the foot of the volcanic Virunga mountains (Green, 2009; Hughes & Hughes, 1992; Tibinika et al., 2016). Lake Kayumbu (2.2km²) has a smaller surface area than Lake Mulehe (5km²). With the exception of Lake Chahafi and Kayumbu, which do not have outlets (Tibinika et al., 2016), the other lakes are interconnected and drain northwards into the Albertine Rift Valley via the Ruchuru River (Figure 2). Due to their altitude, these highland waterbodies are comparably cold and naturally unproductive in terms of fisheries

output and initially, they were fishless (Green, 1976). These minor waterbodies are probably one of the earliest lakes to be stocked in East Africa with various tilapiine and haplochromine fishes, during the 1920s (Green, 2009; Worthington, 1932). However, the fishery quickly became unsustainable due to overfishing and fish mortalities caused by breakdown of stratification (Green, 1976). Currently, the capture fisheries from these highland satellite lakes remains at a subsistence level and the local governments are contemplating for further restocking regimes (Tibihika, Barekye, & Byakora, 2015).



Figure 1 Illustration of the East African Great Lakes, also including some Ethiopian lakes. The dashed red lines mark the boundaries of the Eastern and Western Rift Valleys. Apart from Turkana, the other lakes in the Eastern Rift Valley are drawn shapes but rather not naturally occurring water body shapes (modified from Google maps)

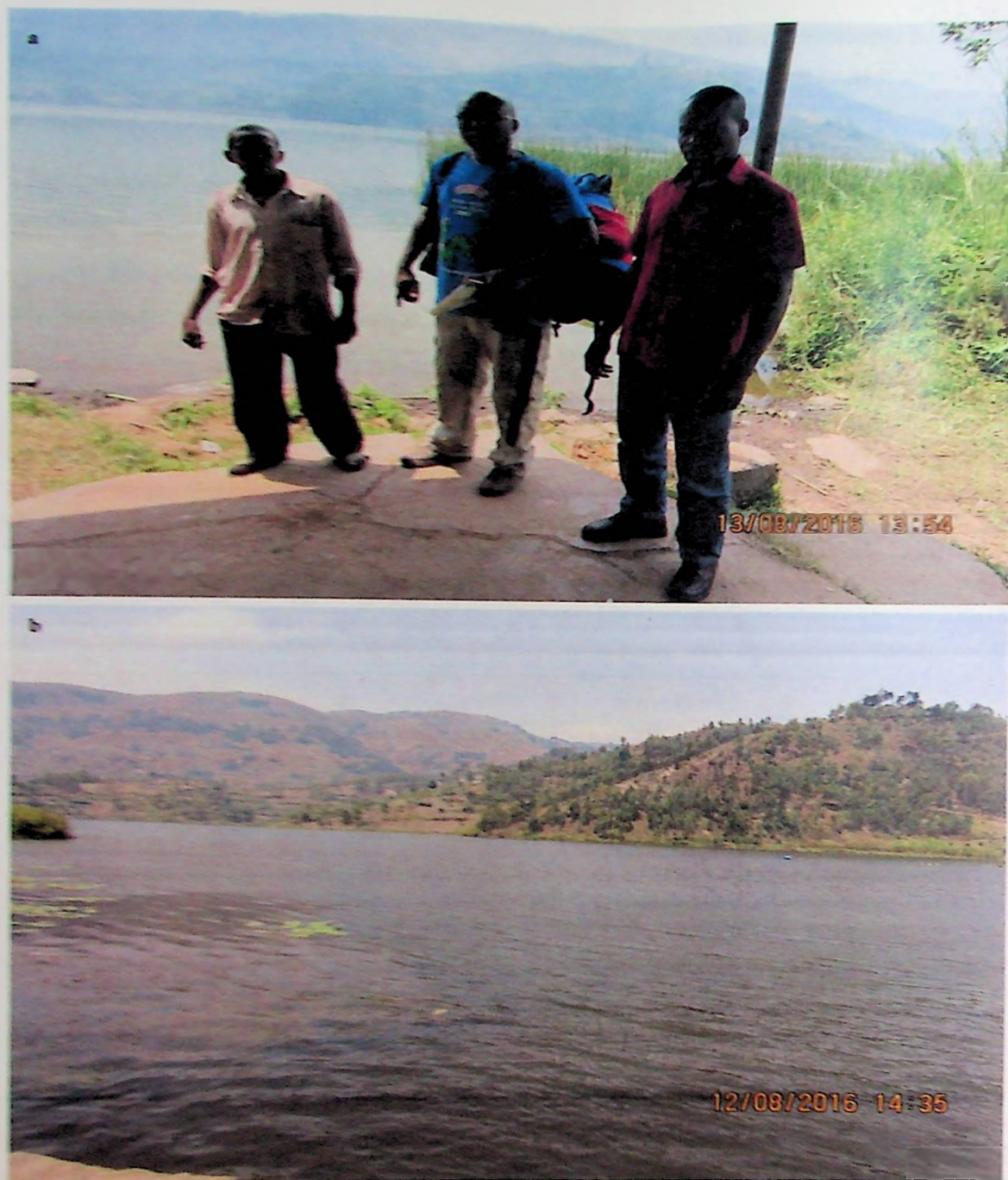


Figure 1.1 South-western Ugandan high-altitude lakes; a Lake Mulehe (from left to right: fisherman on Lake Mulehe, thesis author and Kisoro District Local Government Fisheries Officer), and b Lake Kayumbu (Author picture)

1.4.2. Lakes Edward and George system

Lakes Edward and George (Figure 1.2) lie within the Western Great Rift Valley and are interconnected by Kazinga Channel which stretches between these lakes for a distance estimated to be around 40 km (Green, 2009; Kamanyi, 1992). The only outflow from Lake George to Edward is via this Channel, but the flow velocity in Kazinga is close to zero, which indicates that water mixing between these aquatic systems is insignificant, allowing distinct limnological conditions between the lakes (Green, 2009; Greenwood, 1976). The sluggish water flow of Kazinga Channel between George and Edward is due to the fact that these lakes are almost at the same level ~ 915m asl (Bassa et al., 2015). With a shallow water depth of <2.4 m, Lake George is smaller (250 km²) of the two (Greenwood, 1976). Perhaps, because of shallowness, this explains why Lake George exhibits a high primary productivity (Dunn, 1972; Green, 2009; Greenwood, 1976). The high productivity in George supports various species including abundant amphibious Hippopotamus mammals (Figure 1.2b). Lake George receives most of its inflows from the Rwenzori mountain streams. On the other hand, Lake Edward is much bigger in area (2,325 km²) and deeper (112 m). Edward drains much of its water into Lake Albert via River Semliki (Figure 2). Unlike George, Edward is shared between Uganda and Democratic Republic of Congo (DRC), with the latter owning a big share (71%) (Bassa et al., 2015). These water bodies (Edward, Kazinga Channel and George) are reported to have been intoxicated with heavy metals from the famous Kilembe Copper mines (Uganda), despite the impact much more felt in George (Lwanga, Kansiime, Denny, & Scullion, 2003; Owor, Hartwig, Muwanga, Zachmann, & Pohl, 2007). Nonetheless, the three water bodies, support a rich fish fauna, including among others, *O. niloticus*, *Bagrus docmak*, *Protopterus aethiopicus*, *Clarias gariepinus*. Of these, *O. niloticus* has the greatest economic importance and dominates in catches (Bassa et al., 2015; Dunn, 1972; Kamanyi, 1992; Lowe-McConnell, 2009). However, overfishing has also been reported as one of the major constraints of this fishery (Lowe-McConnell, 2009). Moreover, no fish introductions into Lakes George, Edward and Kazinga Channel have been reported, implying that the ichthyofauna composition is entirely native in these systems.

1.4.3. Lake Albert system

Like the Lakes Edward, George and Kazinga Channel system, Lake Albert lies in the Great Western Rift Valley, but at a much lower altitude (619 masl) than the other Ugandan lakes (Green, 2009). With a surface area of 5,300 km², Lake Albert is the largest of the above-mentioned Albertine lakes. It is also shared between Uganda (54%) and the DRC (46%) (Bassa et al., 2015). The lake is 58 m less deep than Edward (Evans, 1997; Green, 2009) and is fed by several streams from the Rwenzori mountains, but primarily by River Semliki which drains from Lake Edward

(Figure 2). The water flow along River Semliki to Albert is characterized immense rapids and falls (Green, 2009). In the North East of Lake Albert, River Nile (here also called Victoria Nile) enters (Evans, 1997) and hence making this water body nearly a sink for much of the Ugandan waters. Albert Nile flows out of Lake Albert to Sudan in the North, where it becomes the White Nile (Figure 2). Similar to the Edward-George system, Nile tilapia is indigenous to Lake Albert (Figure 1.2b). Apart from *O. niloticus* and other tilapiines, Lake Albert is home to a variety of other fish species which do not occur in either Lake George or Edward. An example is the Nile perch (*Lates niloticus*), a voracious predator (Figure 1.3). Some other species, like *Alestes baremose*, *Hydrocynus forskahlii*, or *Brycinus nurse*, are endemic to Lake Albert (Green, 2009; Mbabazi et al., 2012).

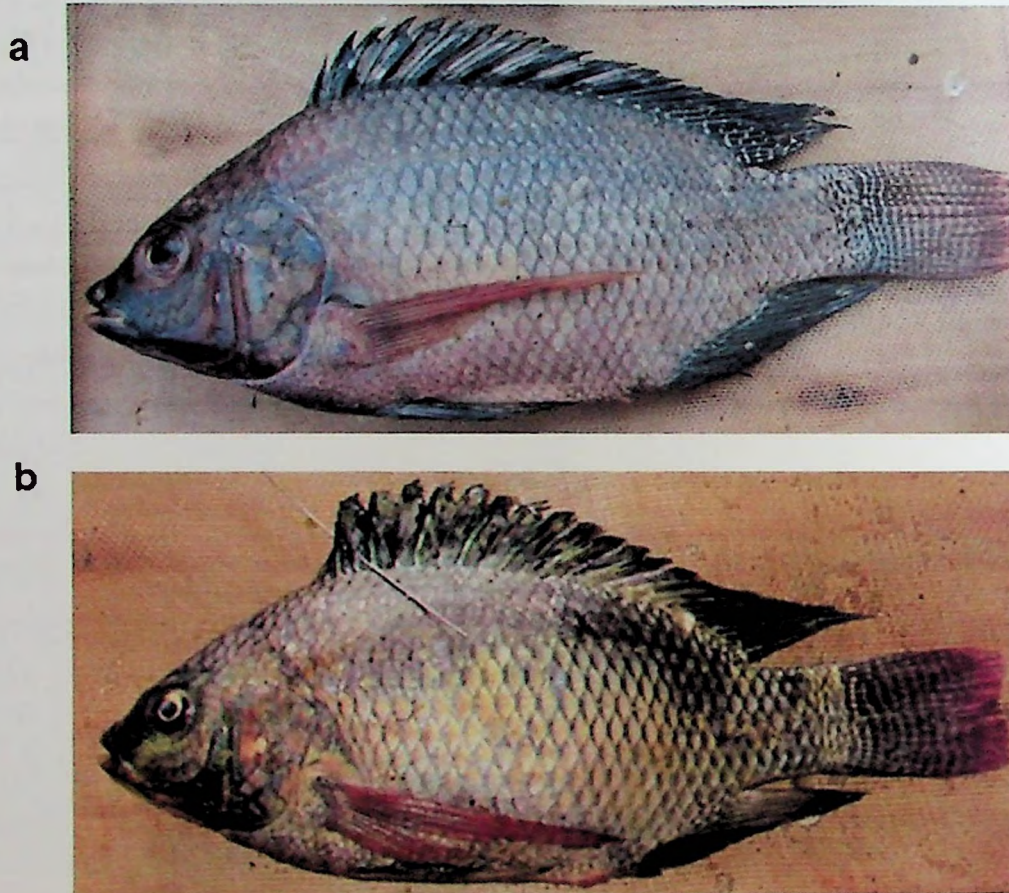


Figure 1.2a Indigenous *O. niloticus* from Lake Albert (a) and the introduce same species (b) in Lake Victoria (Author picture)

a



b

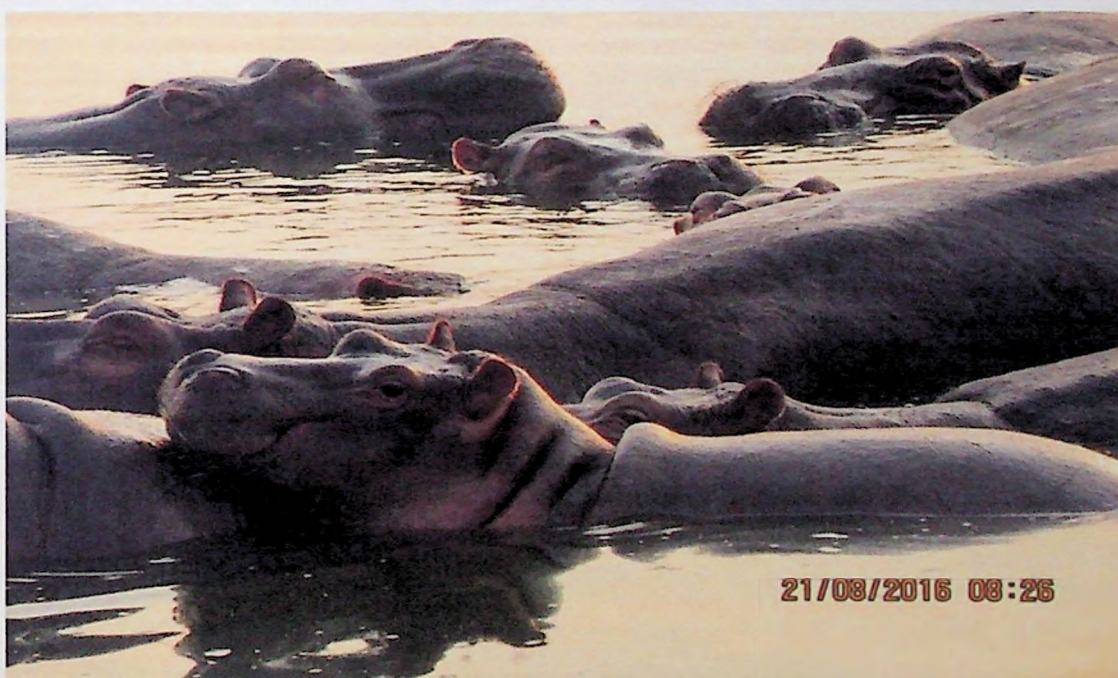


Figure 1.2b Impression of (a) Lake Edward (at Rwenshama landing site) and (b) Hippopotamuses at Lake George (Hamukungu landing site) (Author photo)



Figure 1.3 Nile perch (a) being weighed and (b) lying on ground, from Lake Albert. This fish species is native to Lake Albert and was transferred to the Victoria crescent region. The portrayed Nile perch weighed 67 kg at harvest (Author picture)

1.4.4. Lakes Victoria, Kyoga and River Nile system

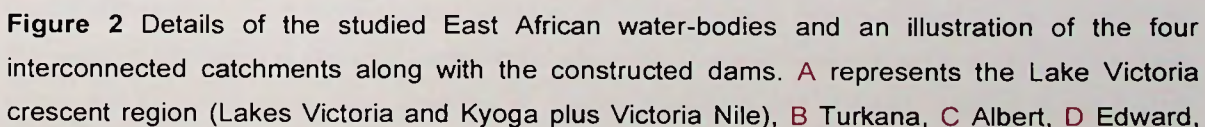
Covering a surface area of around 68, 800 km², Lake Victoria is the second largest freshwater lake in the World, but the largest tropical lake in the World (Odada et al., 2003; Verschuren et al., 2002). The lake is shared between the three East African Countries, Uganda (45%), Kenya (6%) and Tanzania (49%). Lake Victoria covering a catchment area of 184, 000 km², lies at an altitude of around 1134 m.a.s.l and it is generally rather shallow with a maximum depth of 85m, (Hughes & Hughes, 1992). Lake Victoria is considered as the source of River Nile, and it is in this river that Victoria drains much of the waters. From Lake Victoria, River Nile (Victoria Nile) flows into Lake Kyoga and then heads to Lake Albert (Figure 2). Lake Kyoga, at an altitude of 1033 masl, is a shallow lake (maximum depth 10.7m) and has a surface area of 2700 km² (Hughes & Hughes, 1992). Lake Kyoga lies centrally in Uganda and is not shared between the East Africa countries. The Lake Victoria crescent system (Victoria, Kyoga and River Nile) harboured a rich ichthyofauna, dominated by the haplochromine cichlids which were initially of little value for capture fisheries because of their small size (Ogutu-Ohwayo et al., 1997). For this matter, the traditional capture fisheries mainly targeted the indigenous tilapiine cichlids, in particular *O. esculentus* (Ngege fishery). However, other indigenous species like *O. variabilis* and *Labeo victorianus* formed another important fishery (Bairwa, 1992; Mwanja, 2000; Ogutu-Ohwayo, 1990; Ogutu-Ohwayo et al., 1997). Nevertheless, currently the fishery is dominated by alien species; *O. niloticus* and *L. niloticus* all assumed to have originated from Lake Albert (Bairwa, 1994; Ogutu-Ohwayo, 1990). However, at present more fish indirect introductions might be at large mainly through escapees from aquaculture production systems (Figure 1.4). It is believed that before the 1940s fish species between Victoria and Kyoga were interlinked via Victoria Nile (Dumont, 2009; Hughes & Hughes, 1992). However, since the construction of the Owen Falls (Nalubaale) hydroelectric power dam, (Figure 2), at the source of River Nile in the 1950s (Dumont, 2009; Hughes & Hughes, 1992), the migration of fishes from Victoria to Victoria Nile was likely blocked. Recently, two more hydropower dams (Bujagali and Karuma) have been constructed on the same stretch along Victoria Nile. These three dams (perhaps with the fourth one in the same River), therefore, might have further limited the connectivity among Nile tilapia populations, thus compromising the evolutionary genetic trajectory of the fish in this region. Moreover, Murchison falls on Victoria Nile (Figure 2), which is a natural barrier, inhibits fish interchange between the upper and lower Victoria Nile (Hughes & Hughes, 1992).



Figure 1.4 Cage aquaculture production systems planted in Lakes Victoria, Masese (a) in Jinja (author picture) and Albert, Butiaba (b) in Buliisa, (© FAO, Uganda)

1.4.5. Lake Turkana system

With a surface area of around 75,000 km² and a maximum depth of 115m, Lake Turkana (previously called Rudolf) is the fourth largest of the East African Great Lakes, but the largest water body in the Eastern Great Rift Valley (Kolding, 1992; Odada et al., 2003; Ogutu-Ohwayo et al., 1997). The lake lies in the arid North-western part of Kenya and in the South-western part of Ethiopia (Odada et al., 2003). The major inflow is River Omo, which supplies 90% of the incoming water and counterbalances water loss via evaporation as the lake does not have an outlet



Kazinga Channel and George and E (SUHL) South-western Ugandan high-altitude lakes, OD signifies Owen falls hydroelectric dam, BD Bujagali dam, KD Karuma dam, MF Murchison falls dam and RF Rapids and falls on River Semliki.



Figure 2.1 Impressions from Lake Turkana (a) and the catchment area. b shows gradual receding of water levels, c author in a laboratory at Turkana, and d the previously covered water area which currently only contains bear sand with human settlements.

1.4.6. The Ethiopian Lakes; Tana, Chamo, Hashenge and Ziway, and Burkina Faso-Loumbila reservoir

Historically, Ethiopia was not landlocked, but since the 1990s the country became non-coastal (Tesfaye, 2010). This transition may have contributed to increased fishing pressure on the freshwater fish species, given the discontinuity of marine waters. Nonetheless, the country is well endowed with rich inland aquatic resources with most of the Lakes lying in the Ethiopian Rift Valley apart from Tana and Hashenge (Figure 3). Although commercial fisheries is practiced in some Rift Valley water bodies, majorly the Ethiopian fishery is artisanal with even much of the catch

unreported (Tesfaye, 2010; Tesfaye & Wolff, 2014). In this thesis I studied four Ethiopian lakes, two in the Rift Valley (Chamo and Ziway) and two in the mountains (Tana and Hashenge) (Figure 3). The Ethiopian Rift Valley is part of the great East African Rift Valley (Tesfaye, 2010). Among these aforementioned lakes, Lake Tana, located at an altitude of 1,800 m a.s.l. and with an average depth of 8 m, is the largest of the Ethiopian lakes, with an area of 3,200 km² (Vijverberg, Dejen, Getahun, & Nagelkerke, 2012). Lake Tana is the source of Blue Nile which drains into the main Nile (White Nile) (Dumont, 2009). Lake Chamo, at an altitude of 1,233 m a.s.l. and a mean depth of 10 m, is the second largest of these systems with an area of 551 km². Lake Hashenge (2,440 m a.s.l.) and Ziway (1,636 m a.s.l.) are comparably small, with a mean area of 15 km² and 434 km², respectively (Vijverberg et al., 2012). It should be noted that, the mentioned Ethiopian lakes are not interconnected despite Ziway and Langano (un studied lake) linked by rivers through Lake Abijjata (Vijverberg et al., 2012). Although some reports indicate Nile tilapia transfers in Ethiopia (Tesfaye, 2010), the cichlid is native to all these lakes and serves as a major catch in all the water bodies (Tesfaye & Wolff, 2014). Nile perch (*Late niloticus*) is also present in Lake Chamo (Tesfaye & Wolff, 2014). Various anthropogenic activities including, water habitat destruction, hydroelectric power dams, irrigation, deforestation, overgrazing and overfishing inter alia, have been reported as major threats to the Ethiopian fisheries (Tesfaye, 2010; Vijverberg et al., 2012). Despite fish overexploitation which occurs in all over the Ethiopian Lakes, Lake Tana is reported to be underfished (Wudneh, 1998). In general, fish translocations in Ethiopia have not been widely reported in comparison to the status of East African Nile tilapia. However, a few reports indicate that non-native fish species were introduced to Ethiopia, including *Tilapia zillii*, *Tilapia rendalii*, *Salmo trutta*, *Salmo ghiardneri*, *Ctenopharyngodon Idella*, *Cyprinus carpio*, *Caracius caracius*, *Caracius auratus*, and *Gambusia holbrooki* (Getahun & Stiasny, 1998; Tesfaye, 2010; Wudneh, 1998). But, overall the source of the aforementioned exotic species and recipient habitats are likely undocumented or unknown.

Loumbila dam also referred to as Loumbila water reservoir is one of the biggest man made water bodies in Burkina Faso-West Africa (Hugueny & Kabré, 2008). The dam, with an area of 1500 ha and a mean depth of 6.5 m is located about 20km North of Burkina Faso capital city, Ouagadougou (Wagadugu) (Hugueny & Kabré, 2008). The dam was built in 1947 for mainly supplying water for agriculture, fishing, recreational and ranching (Ouédraogo, Chételat, & Amyot, 2015). Little is known about this dam in respect to the fisheries and fish species, but the reservoir faces numerous anthropogenic activities and heavy metal contaminations have been reported (Ouédraogo et al., 2015). As Loumbila dam is an artificial water body, Nile tilapia is likely to be stocked, but this is not documented.



Figure 3 Location of Burkina Faso and Ethiopia. The detailed Ethiopian map on the right shows the locations of the four Ethiopian lakes: Chamo, Ziway, Ashenge/Hashenge and Tana (Source: modified after Google maps).

1.5. Significance and general objective of the thesis

The rate of decline in fish biodiversity, associated with anthropogenic activities is a big concern especially in the developing world where human populations generally rely on fisheries as a cheap source of proteins. The East African inland freshwater fisheries are not exceptional as the resources are very vulnerable to overexploitation because of weak law enforcement measures (Garcia, 2003; Ogutu-Ohwayo & Baliirwa, 2006). Although the fishing industry already directly supports more than four million people, the increasing human population intensifies the already existing pressures on the fisheries, mainly attributed to policies which allow unregulated access to aquatic resources (Béné, 2006; Odongkara et al., 2005; Ogutu-Ohwayo & Baliirwa, 2006). In Uganda, for example, the fisheries sector was the second highest foreign exchange contributor after coffee, with annual returns equivalent to 147 million USD in 2005 (Basiita, 2016). However, this value has drastically declined over time because stocks are decreasing as a result of intensified fishing effort over years (Baliirwa et al., 2003). Options for addressing the threats to fish stocks in the East African freshwater fisheries have been based largely on the adoption of co-management strategies through creation of Beach Management Units (BMUs) (Nunan, Hara, &

Onyango, 2015; Obiero et al., 2015). Important to note though is that, all the previous attempts aimed at managing the East African fisheries sub-sector presented some failures (Obiero et al., 2015), a likelihood of misguided inadequate research (Ogutu-Ohwayo, 1990).

Since the original investigations of Trewavas, (1983), a number of studies have attempted to document the evolutionary biology and ecology of the East African Nile tilapia (Agnèse et al., 1997; Agnèse et al., 1998; Angienda et al., 2011; Balirwa, 1992; Mwanja, Booton, Kaufman, & Fuerst, 2008; Nyingi, De Vos, Aman, & Agnèse, 2009; Ogutu-Ohwayo, 1990). However, the knowledge on the genetic structure of the cichlid still remains insufficient as most of these studies might have been constrained by either the use of relatively uninformative loci (traditional genetic markers) or a small sample size (i.e., the limited number of populations studied). Additionally, some of the natural habitats for Nile tilapia (e.g., the upper Nile), different fish farms and the non-native satellite lakes are less known. Also, few or no studies have comprehensively studied the East African Nile tilapia in comparison to other populations outside the region, which is, however, a crucial step for acquiring in-depth understanding of the species.

The anthropogenic activities, in particular, Nile tilapia translocations, demands a careful assessment of the species through powerful state of the art methodologies. The expansion of aquaculture systems has also led to an increase of cage production systems which are rampant on many natural waterbodies, as well as fish ponds located in water peripheries – all of these pose further threats to local populations through escapees. The protection and maintenance of aquatic habitats is important to guarantee the sustainability of fish biodiversity. Nevertheless, the greatest driver in taxa conservation relies mostly on molecular genetic studies (Abdul Muneer & Gopalakrishnan, 2005; Allendorf & Luikart, 2007). However, to fully understand the demographic composition of the freshwater species in the region, it is necessary to combine geometric morphometrics approaches with molecular genetic markers, which would include a variety of benefits, such as the identification of distinct fishery stocks, the effective use of fish hatcheries and other breeding programs, or the detection of feral populations and the maintenance of population variability/diversity. Furthermore, understanding population pedigrees and phylogenetics may be a preliminary step to sourcing ideal strains for restocking programs for efforts aimed at lessening maladaptation and hence enhancing productivity, especially in the fragmented and naturally unproductive minor lakes (Allendorf & Luikart, 2007; Hallerman, 2003).

In this thesis, a concomitant application of powerful geometric morphometrics and molecular genetic tools also herein referred to as a multi-level analysis, was used to investigate the East African Nile tilapia populations. I studied both, natural and artificial populations of East African Nile

tilapia. The artificial populations include fish farms (pond aquaculture), whereas the other populations comprise Nile tilapia from natural waterbodies where they are either native or introduced. The studied populations, particularly the ones from natural waterbodies, form the major basins which represent the African Great Lakes.

This thesis aimed at unfolding a broader picture of the current demographic state of Nile tilapia populations. Considering the diverse waterbodies which the species inhabits, the overall objective of the thesis was gain better insights into the morphological state and genetic structure of the species in response to anthropogenic pressures in the African Great Lakes region. In this context, it was hypothesized that interspecific species admixture through fish translocations to different eco-geographical habitats might have influenced the alteration of both, morphological and genetic structure, among the Nile tilapia populations. Also, the physical connectivity or natural water intermixing might also explain the demographic relationships of East African Nile tilapia populations. In response to these predictions, the following subsequent chapters were organized against which research questions were formulated.

1.5.1. Organization of chapters and general research questions

The subsequent chapters of this thesis are compiled based on three studies, herein hierarchically presented as chapters 2, 3 and 4. Each chapter, follows the authors, year published, title and henceforth scientific research questions.

Chapter 2; Tibihika et al. (2018). Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus* L. 1758) in East Africa.

Scientific research questions:

Can condition factor (K) appropriately explain adaptability or performance of Nile tilapia populations?

To what extent are the introduced (non-native) Nile tilapia populations morphologically distant from their putative natives and source populations?

To what degree are the Lake Victoria metapopulations morphologically different wing to potential multiple introductions and or size of the system?

Chapter 3; Tibihika et al. (2018). Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*.

Scientific research questions:

Using next generation sequencing (NGS), can microsatellites be informative enough to delineate the Great Lakes Nile tilapia populations?

To what extent can genotyping by sequencing (SSR-GBS) be informative in defining the genetic structure of East African populations?

Chapter 4: Tibihika et al (Under revision). Genetic structure evidence of anthropogenic threats in the East African fisheries: the paradigm of Nile tilapia (*Oreochromis niloticus*, L. 1758)

Scientific research questions:

Does the genetic structure of the East African *O. niloticus* populations differ from those outside the region?

To what extent does the genetic structure of the East African *O. niloticus* populations translate to anthropogenic activities and geographical settings?

1.6. Thesis general structure and overview of the research approaches

Overall, the thesis structure begins with a general introduction (current chapter) then follows the various chapters including the research papers presented in article format as detailed in Figure 4. The thesis commences with a synthesis of the broad spectrum of the general information concerning the studied water bodies and previous work on Nile tilapia. I further show the relevance, significance and research questions for the thesis.

Chapter 2, entails advanced approaches for analyzing shapes or morphs of the East African Nile tilapia. This chapter elucidates the details of sampling/field excursions and subsequent digital pictures taken for geometric morphometrics. A specialized fabricated tripod stand (Figure 5) was used for fish photography which allowed pictures of the specimen to be taken from exactly the same angle, focal length and similar orientation. This was possible with the influence of the mesh cradle inserted between the tripod stands (Figure 5). The studied populations in this chapter include; Lakes Victoria (five sampling stations), Kyoga (two sampling stations), Edward (two sampling stations), George, Albert (two sampling stations), Kazinga Channel, Turkana, Mulehe, Kayumbu, River Nile (Victoria Nile), and three farms; Bagena (Kisoro-Uganda), Sindi (Kabale-Uganda) and Rwitabingi (Kamuli-Uganda). In total, 13 populations were studied. Morphological variations among the populations were analyzed with three programs; thin plate spline utility (TpsUtil), thin plate spline digitizer (TPsDig) and MorphoJ. TpsUtil and TpsDig programs were used to digitize fish images and generated x, y coordinate matrices which were then imported into MorphoJ program where Multivariate analyses were implemented. Results for morphological

divergence and condition factor of Nile tilapia populations are reported and discussed in this chapter.

Chapter 3, presents the details of microsatellite marker (SSR) development using next generation sequencing (NGS), which also entails the advanced genotyping platform; genotyping by sequencing (GBS) (Figure 6). The markers were developed from an Ethiopian Nile tilapia sample, which was genotyped and sequenced using the Illumina Miseq 300bp platform. Out of the 35 markers developed, 26 proved to be highly informative and were later tested on the four East African Nile tilapia as well as *Tilapia zillii* populations for cross-species amplification. This chapter further explains and justifies the strengths of NGS, in respect to microsatellite genotyping by sequencing platform (SSR-GBS). Here, SSR-GBS (whole amplicon/sequence information) was compared with the traditional amplicon/sequence length methods, regarding allele calling (Figure 6).

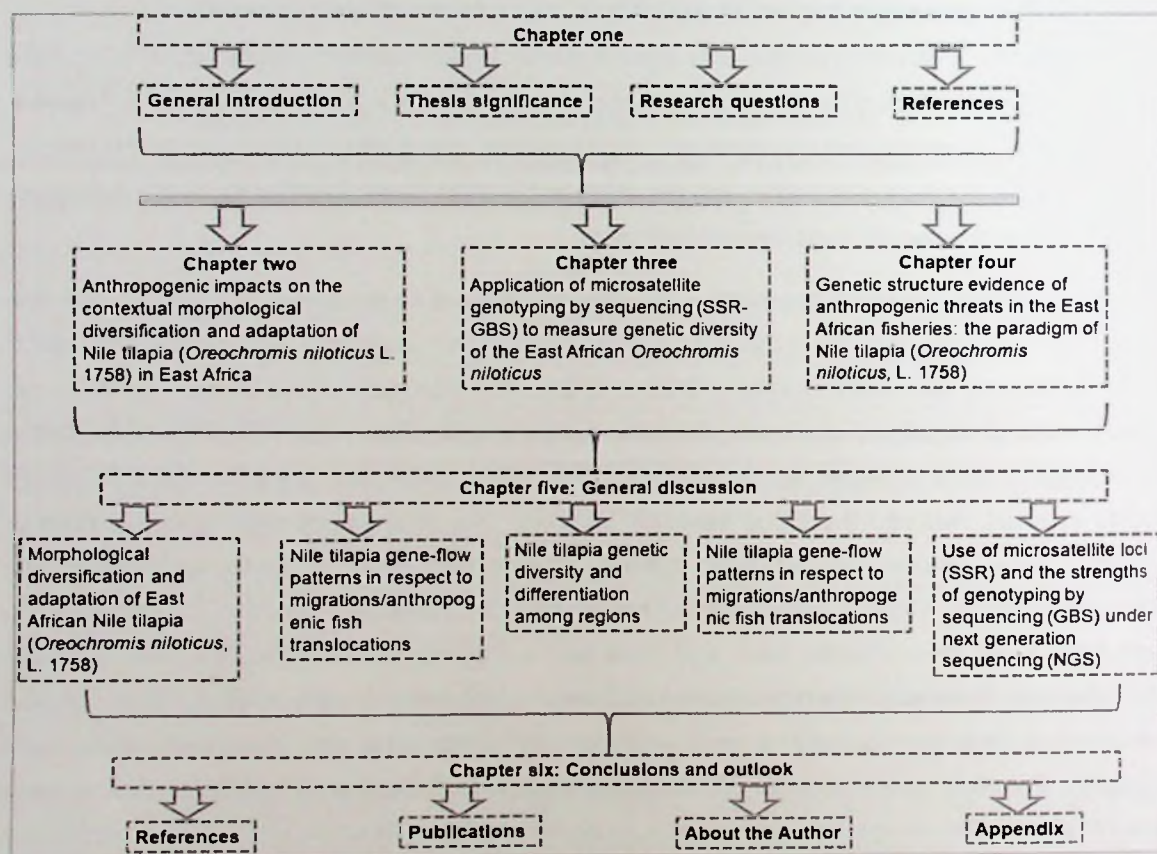


Figure 4 A schematic illustration of the thesis structure

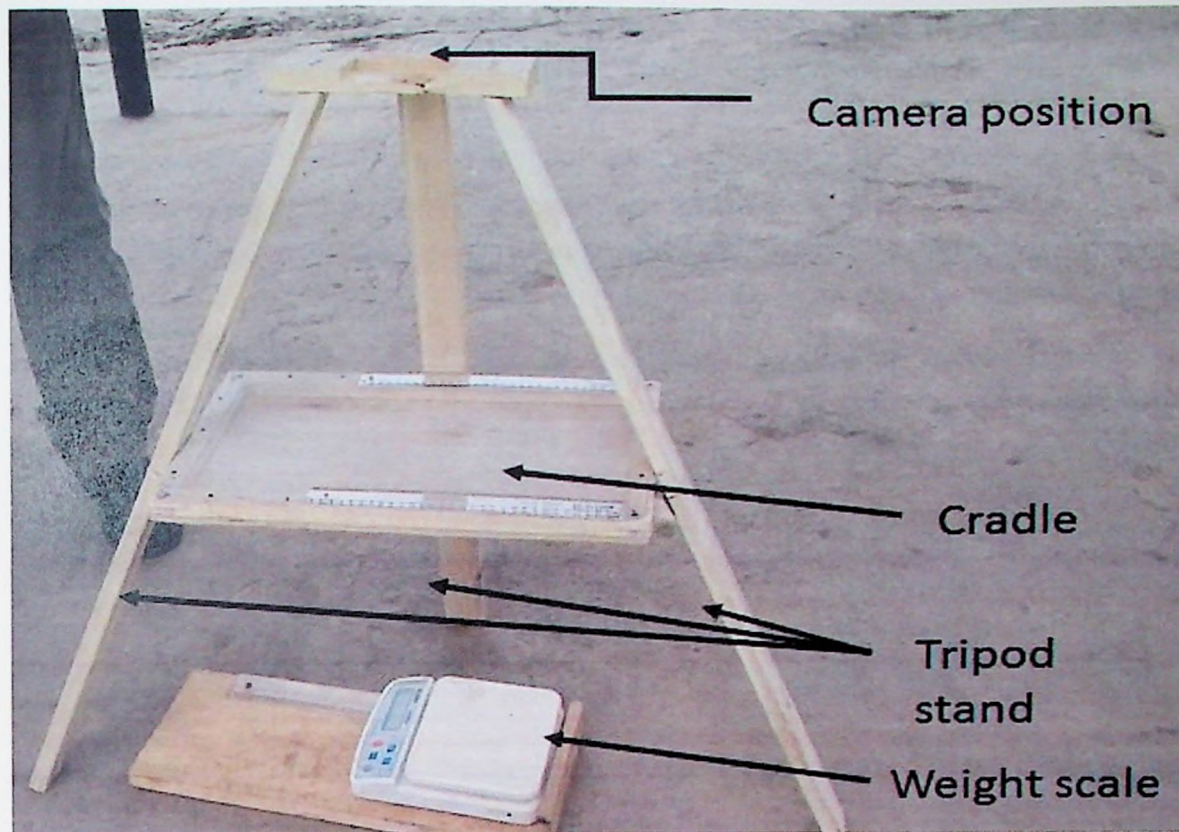


Figure 5 Tripod stands with a cradle and a digital weighing scale used for photography in the study

In chapter 4, additional four Nile tilapia populations from Ethiopia (lakes Ziway, Tana, Chamo and Hashenge) and one from Burkina Faso (West Africa) were amplified and analyzed together with the 13 East African populations (Figure 7). The 13 East African populations are analogous to those presented in chapter two. 26 microsatellite markers developed in chapter 3, were subsequently used to amplify these populations. However, 15 additional markers were designed and supplemented on the 26 markers earlier utilized in chapter three, to make a total of 41 which were then comprehensively employed in Chapter four. However, for quality check in downstream statistical analyses, one locus (Ti22) which proved to be under selection was dropped, leaving a total number of 40 loci. Genetic structure, diversity and differentiation among the populations within and outside East Africa are presented and discussed in this chapter. The research findings of chapter 2–4 are then discussed in chapter 5 which links the outputs to the literature and embeds the results in a practical (management) context. In this chapter five, the detailed information regarding the morphological and genetic structure of Nile tilapia is explained with view of the past

works. Specifically, the available information regarding the historical classification of Nile tilapia into subspecies using the traditional morphometrics and meristics in Africa in comparison to the present findings are discussed for literature enrichment. I further elaborate on the past anthropogenic activities on Nile tilapia importantly regarding the translocations of the cichlid in East African Great Lakes. Chapter 6, concludes the thesis. Here, a synthesis of the thesis is outlined and I end with a clear direction for future research perspectives and general recommendations for management strategies.

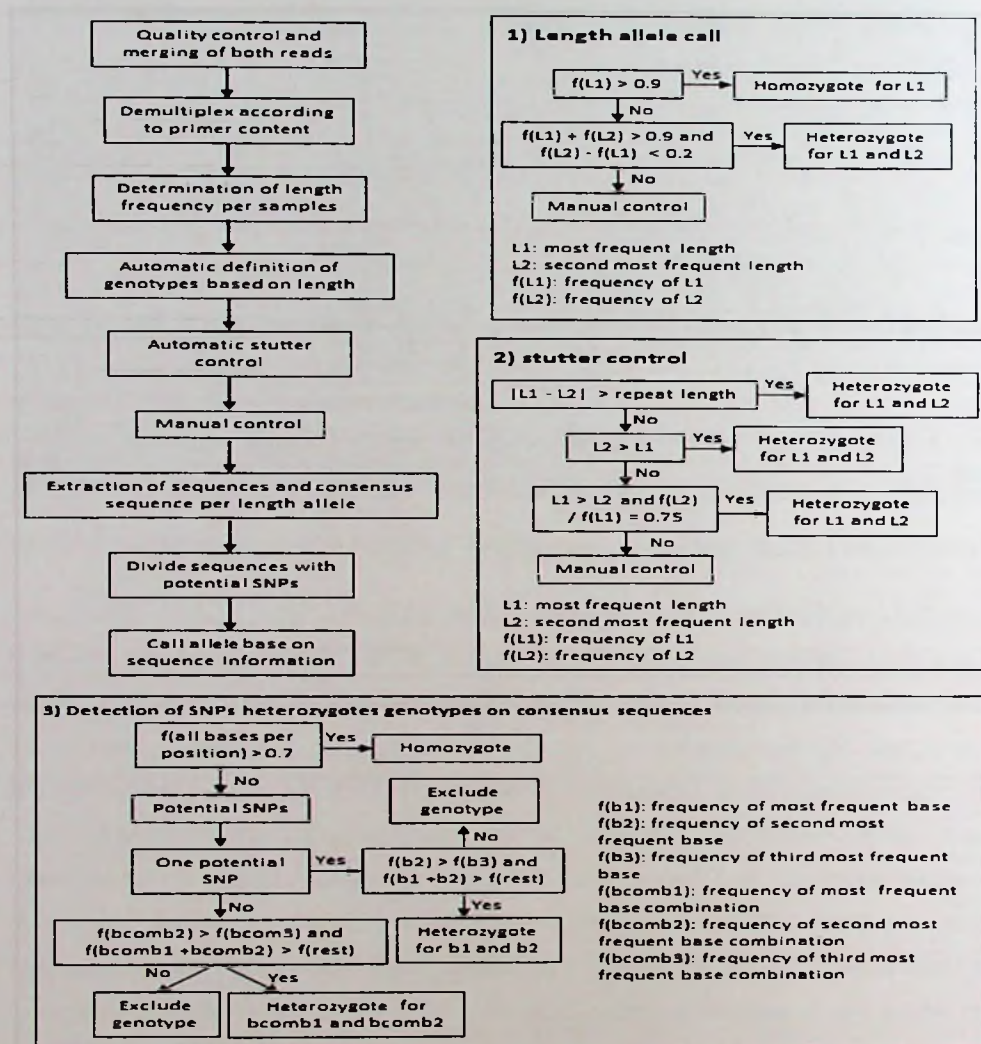


Figure 6 Schematic flow for amplicon analysis and genotyping by sequencing platform (GBS) under next generation sequencing (NGS), adopted from Curto et al. (2019)



Figure 7 Illustration of the various sampling locations per country. The black stars denote farms while the red ones are lakes. Two or more stars on a given water body indicates more sampling sites e.g. on Edward, Albert, Victoria and Kyoga. Note that Burkina Faso is not included in this map.

1.7. References

- Abdul Muneer, P., & Gopalakrishnan, A. (2005). *Molecular Genetic Characterization of Endemic Yellow Catfish, *Horabrus brachysoma* (Gunther)*. National bureau of fish genetic resources cochin unit.
- Agnèse, J.-F., Adépo-Gourène, B., Abban, E. K., & Fermon, Y. (1997). Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity*, 79(1), 88-96.
- Agnèse, J.-F., Adépo-Gourène, B., & Pouyaud, L. (1998). Natural hybridization in tilapias. *Genetics and Aquaculture in Africa*. 95-105.
- Allendorf, F. W., & Luikart, G. (2007). Conservation and the genetics of populations. *mammalia*, 2007, 189-197. doi:<https://doi.org/10.1515/MAMM.2007.038>
- Anderson, A. (1961). Further observations concerning the proposed introduction of Nile perch into Lake Victoria. *East African Agricultural and Forestry Journal*, 26(4), 195-201.
- Angienda, P. O., Lee, H. J., Elmer, K. R., Abila, R., Waindi, E. N., & Meyer, A. (2011). Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. *Conservation Genetics*, 12(1), 243-255. doi:<https://doi.org/10.1007/s10592-010-0136-2>
- Balirwa, J. S. (1992). The evolution of the fishery of *Oreochromis niloticus* (Pisces: Cichlidae) in Lake Victoria. *Hydrobiologia*, 232(1), 85-89. doi:<https://doi.org/10.1007/BF00014616>
- Balirwa, J. S. (1994). The biology, ecology, population parameters and the fishery of the Nile tilapia, *Oreochromis niloticus* (L.). *Aquatic Commons*, 26-38. doi:10.1007/BF00014616
- Balirwa, J. S. (1998). *Lake Victoria wetlands and the ecology of the Nile tilapia, Oreochromis niloticus Linne*. Rotterdam: Balkema.
- Balirwa, J. S., Chapman, C. A., Chapman, L. J., Cowx, I. G., Geheb, K., Kaufman, L., . . . Welcomme, R. L. (2003). Biodiversity and fishery sustainability in the Lake Victoria basin: an unexpected marriage? *BioScience*, 53(8), 703-715. doi:[https://doi.org/10.1641/0006-3568\(2003\)053\[0703:BAFSIT\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2003)053[0703:BAFSIT]2.0.CO;2)
- Bárfai, R., Egedi, S., Yue, G. H., Kovács, B., Urbányi, B., Tamás, G., . . . Orbán, L. (2003). Genetic analysis of two common carp broodstocks by RAPD and microsatellite markers. *Aquaculture*, 219(1-4), 157-167. doi:[https://doi.org/10.1016/S0044-8486\(02\)00571-9](https://doi.org/10.1016/S0044-8486(02)00571-9)
- Basiita, R. K. (2016). Biogeography of three commercially important African freshwater fishes: Lates niloticus, Bagrus docmak and Bagrus bayad. PhD thesis, James Cook University.
- Bassa, S., Taabu-Munyaho, A., Mbabazi, D., Nakiyende, H., Muhumuza, E., Amiina, R., . . . Ballirwa, J. (2015). Technical report of frame and catch assessment surveys of Lakes Edward and George; and the Kazinga Channel conducted in January and July 2011-2013.
- Béné, C. (2006). Small-scale fisheries: assessing their contribution to rural livelihoods in developing countries. *FAO fisheries circular*, 1008, 46.
- Beverton, R. (1959). *Report on the state of the Lake Victoria fisheries*. 40. Fisheries Laboratory, Lowestoft. (Cyclostyled.) Fisheries Laboratory.
- Biswas, A. K., Morita, T., Yoshizaki, G., Maita, M., & Takeuchi, T. (2005). Control of reproduction in Nile tilapia *Oreochromis niloticus* (L.) by photoperiod manipulation. *Aquaculture*, 243(1-4), 229-239. doi:<https://doi.org/10.1016/j.aquaculture.2004.10.008>
- Bookstein, F. L. (1982). Foundations of morphometrics. *Annual Review of Ecology and Systematics*, 13(1), 451-470.
- Breno, M., Leirs, H., & Van Dongen, S. (2011). Traditional and geometric morphometrics for studying skull morphology during growth in *Mastomys natalensis* (Rodentia: Muridae). *Journal of Mammalogy*, 92(6), 1395-1406. doi:<https://doi.org/10.1644/10-MAMM-A-331.1>
- Cadrin, S. X. (2000). Advances in morphometric identification of fishery stocks. *Reviews in Fish biology and Fisheries*, 10(1), 91-112. doi:<https://doi.org/10.1023/A:1008939104413>

- Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*, 255(1-4), 1-29. doi:<https://doi.org/10.1016/j.aquaculture.2005.11.031>
- Curto, M., Tembrock, L. R., Puppo, P., Nogueira, M., Simmons, M. P., & Meimberg, H. (2013). Evaluation of microsatellites of *Catha edulis* (qat; Celastraceae) identified using pyrosequencing. *Biochemical Systematics and Ecology*, 49, 1-9. doi:<https://doi.org/10.1016/j.bse.2013.02.002>
- Curto, M., Winter, S., Seiter, A., Schmid, L., Scheicher, K., Barthel, L. M., . . . Meimberg, H. (2019). Application of a SSR-GBS marker system on investigation of European Hedgehog species and their hybrid zone dynamics. *Ecology and Evolution*. doi:<https://doi.org/10.1002/ece3.4960>
- Dadzie, S. (1992). An overview of aquaculture in eastern Africa. *Hydrobiologia*, 232(1), 99-110. doi:<https://doi.org/10.1007/BF00014618>
- Dakin, E., & Avise, J. (2004). Microsatellite null alleles in parentage analysis. *Heredity*, 93(5), 504.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499.
- De Barba, M., Miquel, C., Lobréaux, S., Quenette, P., Swenson, J., & Taberlet, P. (2016). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*. doi:10.1111/1755-0998.12594
- Deck, L. M. G., Habel, J. C., Curto, M., Husemann, M., Sturm, S., Garitano-Zavala, Á., & Meimberg, H. (2016). New microsatellite markers for two sympatric tinamou species, the Ornate Tinamou (*Nothoprocta ornata*) and Darwin's Nothura (*Nothura darwini*). *Avian Biology Research*, 9(3), 139-146. doi:<https://doi.org/10.3184/175815515X14503747783157>
- Diana, J. S. (2009). Aquaculture production and biodiversity conservation. *BioScience*, 59(1), 27-38. doi:<https://doi.org/10.1525/bio.2009.59.1.7>
- Dumont, H. J. (2009). *The Nile: origin, environments, limnology and human use* (Vol. 89): Springer Science & Business Media.
- Dunkelman, T. J., Karson, J. A., & Rosendahl, B. R. (1988). Structural style of the Turkana rift, Kenya. *Geology*, 16(3), 258-261. doi:[https://doi.org/10.1130/0091-7613\(1988\)016<0258:SSOTTR>2.3.CO;2](https://doi.org/10.1130/0091-7613(1988)016<0258:SSOTTR>2.3.CO;2)
- Dunn, I. (1972). The commercial fishery of Lake George, Uganda (East Africa) <http://aquaticcommons.org/id/eprint/20533>. *African Journal of Tropical Hydrobiology and Fisheries*, 2(2), 109-120.
- Eknath, A., Dey, M., Rye, M., Gjerde, B., Abella, T., Sevilleja, R., . . . Bentsen, H. (1998). *Selective breeding of Nile tilapia for Asia*. Paper presented at the 6th World Congress on Genetics Applied to Livestock Production.
- Epifanio, J., & Philipp, D. (2000). Simulating the extinction of parental lineages from introgressive hybridization: the effects of fitness, initial proportions of parental taxa, and mate choice. *Reviews in Fish biology and Fisheries*, 10(3), 339-354. doi:<https://doi.org/10.1023/A:1016673331459>
- Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular ecology*, 11(9), 1591-1604. doi:<https://doi.org/10.1046/j.1365-294X.2002.01576.x>
- Evans, J. (1997). Spatial and seasonal distribution of phytoplankton in an African Rift Valley lake (L. Albert, Uganda, Zaire). *Hydrobiologia*, 354(1-3), 1-16. doi:<https://doi.org/10.1023/A:1003026415788>

- Garcia, S. (2003). Ecosystem approach to fisheries: issue, terminology, principles, institutional foundations, implementation and outlook. *FAO Fish Tech Paper*, 443, 1-71.
- George, T. T. (1995). The most recent nomenclature of tilapia species in Canada and the Sudan. *Bull. Aquacult. Assoc. Canada, Sp. Pub*(10), 33-37.
- Getabu, A., Tumwebaze, R., & MacLennan, D. N. (2003). Spatial distribution and temporal changes in the fish populations of Lake Victoria. *Aquatic Living Resources*, 16(3), 159-165. doi:[https://doi.org/10.1016/S0990-7440\(03\)00008-1](https://doi.org/10.1016/S0990-7440(03)00008-1)
- Getahun, A., & Stiasny, M. L. (1998). The freshwater biodiversity crisis: the case of the Ethiopian fish fauna. *SINET: Ethiopian Journal of Science*, 21(2), 207-230.
- Green, J. (1976). Changes in the zooplankton of lakes Mutanda, Bunyonyi and Mulehe (Uganda). *Freshwater Biology*, 6(5), 433-436. doi:<https://doi.org/10.1111/j.1365-2427.1976.tb01631.x>
- Green, J. (2009). Nilotic lakes of the Western Rift. . In: Dumont H.J. (eds) *The Nile. Monographiae Biologicae The Nile* (Vol. 89, pp. 263-286): Springer, Dordrecht.
- Greenwood, P. (1976). Lake George, Uganda. *Phil. Trans. R. Soc. Lond. B*, 274, 375-391. doi:<https://doi.org/10.1098/rstb.1976.0052>
- Haasl, R. J., & Payseur, B. A. (2011). Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity*, 106(1), 158.
- Hallerman, E. M. (2003). Population genetics: principles and applications for fisheries scientists: Hallerman, E.M. (ed.). *American Fisheries Society, Bethesda, MD (USA) eng*.
- Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105(S1), 795-809. doi:<https://doi.org/10.1093/jhered/esu033>
- Hedrick, P. W. (2001). Conservation genetics: where are we now? *Trends in Ecology and Evolution*, 16(11), 629-636. doi:10.1016/S0169-5347(01)02282-0
- Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A., . . . Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, 111(10), 3775-3780. doi:<https://doi.org/10.1073/pnas.1318945111>
- Hughes, R., & Hughes, J. (1992). *A Directory of African Wetlands*. (IUCN: Gland, Switzerland. United Nations Environment Programme: Nairobi. World Conservation Monitoring Centre: Cambridge.).
- Hugueny, B., & Kabré, G. B. (2008). Seasonal diet shift of the most important fish species in a Sahelo-Soudanian reservoir (Burkina Faso). *Journal of Fisheries and Aquatic Science*, 3(4), 240-251.
- Johnson, T. C., & Malala, J. O. (2009). Lake Turkana and its link to the Nile. In: Dumont H.J. (eds) *The Nile. Monographiae Biologicae*, vol 89 *The Nile* (pp. 287-304): Springer, Dordrecht.
- Kamanyi, J. (1992). Fish exploitation and changes in the fish community of Lake George, Uganda. <http://aquaticcommons.org/id/eprint/20563>.
- Kaufman, L. (1992). Catastrophic change in species-rich freshwater ecosystems. *BioScience*, 42(11), 846-858.
- Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in ecology & evolution*, 17(5), 230-241. doi:[https://doi.org/10.1016/S0169-5347\(02\)02489-8](https://doi.org/10.1016/S0169-5347(02)02489-8)
- Kirema Mukasa, C. (2012). *Regional fish trade in eastern and southern Africa. products and markets. A fish traders guide*. Retrieved from
- Kolding, J. (1992). A summary of Lake Turkana: an ever-changing mixed environment. *Mitt. Internat. Verein. Limnol*, 23, 25-35. doi:<https://doi.org/10.1080/05384680.1992.11904005>
- Langen, K., Schwarzer, J., Kullmann, H., Bakker, T. C., & Thünken, T. (2011). Microsatellite support for active inbreeding in a cichlid fish. *PLoS One*, 6(9), e24689. doi:org/10.1371/journal.pone.0024689

- Lenstra, J., Groeneveld, L., Eding, H., Kantanen, J., Williams, J., Taberlet, P., . . . Ciani, E. (2012). Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Animal Genetics*, 43(5), 483-502. doi:10.1111/j.1365-2052.2011.02309.x
- Liu, Z. J., & Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238(1-4), 1-37. doi:<https://doi.org/10.1016/j.aquaculture.2004.05.027>
- Lowe-McConnell, R. (2009). Fisheries and cichlid evolution in the African Great Lakes: progress and problems. *Freshwater Reviews*, 2(2), 131-151. doi:<https://doi.org/10.1608/FRJ-2.2.2>
- Lowe-McConnell, R. (1975). Fish communities in tropical freshwaters: their distribution, ecology and evolution. New York: Logman: Inc.
- Lwanga, M. S., Kansiime, F., Denny, P., & Scullion, J. (2003). Heavy metals in Lake George, Uganda, with relation to metal concentrations in tissues of common fish species. *Hydrobiologia*, 499(1-3), 83-93. doi:<https://doi.org/10.1023/A:1026347703129>
- Maderbacher, M., Bauer, C., Herler, J., Postl, L., Makasa, L., & Sturmbauer, C. (2008). Assessment of traditional versus geometric morphometrics for discriminating populations of the *Tropheus moorii* species complex (Teleostei: Cichlidae), a Lake Tanganyika model for allopatric speciation. *Journal of Zoological Systematics and Evolutionary Research*, 46(2), 153-161. doi:<https://doi.org/10.1111/j.1439-0469.2007.00447.x>
- Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in genetics*, 24(3), 133-141. doi:<https://doi.org/10.1016/j.tig.2007.12.007>
- Mbabazi, D., Taabu-Munyaho, A., Muhoozi, L., Nakiyende, H., Bassa, S., Muhunuza, E., . . . Balirwa, J. (2012). The past, present and projected scenarios in the Lake Albert and Albert Nile Fisheries: Implications for sustainable management. *Uganda Journal of Agricultural Sciences*, 13(2), 47-64.
- Miah, G., Rafii, M., Ismail, M., Puteh, A., Rahim, H., Islam, K., & Latif, M. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International journal of molecular sciences*, 14(11), 22499-22528. doi:<https://doi.org/10.3390/ijms141122499>
- Mojekwu, T., & Anumudu, C. (2013). Microsatellite markers in Aquaculture: Application in Fish population genetics. *IOSR J Environ Sci Toxicol Food Technol*, 5(4), 43-48.
- Mooney, H. A., & Cleland, E. E. (2001). The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences*, 98(10), 5446-5451. doi:10.1073/pnas.091093398
- Mwanja, W. W. (2000). *Genetic biodiversity and evolution of two principal fisheries species groups, the Labeine and Tilapiine, of Lake Victoria Region, East Africa*. The Ohio State University.
- Mwanja, W. W., Booton, G., Kaufman, L., Chandler, M., Fuerst, P., Donaldson, E., & MacKinlay, D. (1996). *Population and stock characterization of Lake Victoria. Tilapine fishes based on RAPD markers*. Paper presented at the Aquaculture Biotechnology Symposium Proceedings of the International Congress on the Biology of Fishes (ed. by EM Donaldson & DD Mackinlay).
- Mwanja, W. W., Booton, G. C., Kaufman, L., & Fuerst, P. A. (2008). 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*, 7(11). doi:<http://dx.doi.org/10.5897/AJB08.813>
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular ecology*, 22(11), 2841-2847. doi:<https://doi.org/10.1111/mec.12350>
- Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 1-10. doi:10.1111/j.1558-5646.1975.tb00807.x
- Njiru, M., Nzungi, P., Getabu, A., Wakwabi, E., Othina, A., Jembe, T., & Wekesa, S. (2007). Are fisheries management, measures in Lake Victoria successful? The case of Nile perch and

- Nile tilapia fishery. *African Journal of ecology*, 45(3), 315-323. doi:10.1111/j.1365-2028.2006.00712.x
- Njiru, M., Ojuok, J., Okeyo-Owuor, J., Muchiri, M., Ntiba, M., & Cowx, I. (2006). Some biological aspects and life history strategies of Nile tilapia *Oreochromis niloticus* (L.) in Lake Victoria, Kenya. *African Journal of Ecology*, 44(1), 30-37. doi:<https://doi.org/10.1111/j.1365-2028.2006.00610.x>
- Nunan, F., Hara, M., & Onyango, P. (2015). Institutions and co-management in east African inland and Malawi fisheries: A critical perspective. *World Development*, 70, 203-214. doi:<https://doi.org/10.1016/j.worlddev.2015.01.009>
- Nyingi, D., De Vos, L., Aman, R., & Agnèse, J.-F. (2009). Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758)(Cichlidae; Teleostei) in the Lobo Swamp (Kenya). *Aquaculture*, 297(1-4), 57-63. doi:<https://doi.org/10.1016/j.aquaculture.2009.09.017>
- Obiero, K. O., Abila, R. O., Njiru, M. J., Raburu, P. O., Achieng, A. O., Kundu, R., . . . Lawrence, T. (2015). The challenges of management: Recent experiences in implementing fisheries co-management in Lake Victoria, Kenya. *Lakes & Reservoirs: Research & Management*, 20(3), 139-154. doi:<https://doi.org/10.1111/lre.12095>
- Odada, E. O., Olago, D. O., Bugenyi, F., Kulindwa, K., Karimurumungu, J., West, K., . . . Achola, P. (2003). Environmental assessment of the East African rift valley lakes. *Aquatic sciences*, 65(3), 254-271. doi:<https://doi.org/10.1007/s00027-003-0638-9>
- Odongkara, K., Abila, R., & Onyango, P. (2005). *Distribution of economic benefits from the fisheries of Lake Victoria. The State of the Fisheries Resources of Lake Victoria and their Management*. Paper presented at the Proceedings of the Lake Victoria stakeholders conference: meeting the challenges to sustainability of the fisheries resources. Regional Stakeholders' Workshop. 24-25 February 2005, Entebbe, Uganda. ISBN: 9970-713-10-2 (Pp 124-131).
- Ogutu-Ohwayo, R. (1990). The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of introduced species, especially the Nile perch, *Lates niloticus*, and the Nile tilapia, *Oreochromis niloticus*. *Environmental Biology of Fishes*, 27(2), 81-96. doi:<https://doi.org/10.1007/BF00001938>
- Ogutu-Ohwayo, R., Hecky, R. E., Cohen, A. S., & Kaufman, L. (1997). Human impacts on the African Great Lakes. *Environmental Biology of Fishes*, 50(2), 117-131. doi:<https://doi.org/10.1023/A:1007320932349>
- Ogutu-Ohwayo, R. (1990). The reduction in fish species diversity in Lakes Victoria and Kyoga (East Africa) following human exploitation and introduction of non-native fishes. *Journal of Fish Biology*, 37, 207-208. doi:<https://doi.org/10.1111/j.1095-8649.1990.tb05040.x>
- Ogutu-Ohwayo, R., & Balirwa, J. S. (2006). Management challenges of freshwater fisheries in Africa. *Lakes & Reservoirs: Research & Management*, 11(4), 215-226. doi:<https://doi.org/10.1111/j.1440-1770.2006.00312.x>
- Okumuş, I., & Çiftçi, Y. (2003). Fish population genetics and molecular markers: II-molecular markers and their applications in fisheries and aquaculture. *Turkish Journal of Fisheries and Aquatic Sciences*, 3(1), 51-79.
- Otieno, H. O., & Awange, J. L. (2006). *Energy resources in East Africa: opportunities and challenges*: Springer Science & Business Media.
- Ouédraogo, O., Chételat, J., & Amyot, M. (2015). Bioaccumulation and trophic transfer of mercury and selenium in African sub-tropical fluvial reservoirs food webs (Burkina Faso). *PloS one*, 10(4), e0123048. doi:<https://doi.org/10.1371/journal.pone.0123048>
- Owor, M., Hartwig, T., Muwanga, A., Zachmann, D., & Pohl, W. (2007). Impact of tailings from the Kilembe copper mining district on Lake George, Uganda. *Environmental Geology*, 51(6), 1065-1075. doi:10.1007/s00254-006-0398-7

- Park, L. K., & Moran, P. (1994). Developments in molecular genetic techniques in fisheries. *Reviews in Fish Biology and Fisheries*, 4(3), 272-299. doi:10.1007/BF00042906
- Pasqualotto, A. C., Denning, D. W., & Anderson, M. J. (2007). A cautionary tale: lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. *Journal of Clinical Microbiology*, 45(2), 522-528. doi:<https://doi.org/10.1128/JCM.02136-06>
- Petrtyl, M., Kalous, L., & Memiş, D. (2014). Comparison of manual measurements and computer-assisted image analysis in fish morphometry. *Turkish Journal of Veterinary and Animal Sciences*, 38(1), 88-94. doi:10.3906/vet-1209-9
- Popma, T., & Masser, M. (1999). Tilapia, life history and biology-Southern Regional Aquaculture Center (SRAC) Publication, No. 283 (pp. 1-4).
- Pullin, R. S. (1988). Tilapia genetic resources for aquaculture. Proceedings of the Workshop on Tilapia Genetic Resources for Aquaculture, 23-24 March 1987. *The WorldFish Center Working Papers*.
- Rothuis, A., Turenhout, M., van Duijn, A., Roem, A., Rurangwa, E., Katunzi, E., . . . Kabagambe, J. B. (2014). *Aquaculture in East Africa: a regional approach*. Retrieved from
- Salzburger, W., Van Bocxlaer, B., & Cohen, A. S. (2014). Ecology and evolution of the African Great Lakes and their faunas. *Annual Review of Ecology, Evolution, and Systematics*, 45, 519-545. doi:<https://doi.org/10.1146/annurev-ecolsys-120213-091804>
- Scullion, J. (2007). *Inland fisheries co-management in East Africa. Lake Victoria Fisheries Organisation*.
- Seyoum, S., & Kornfield, I. (1992). Identification of the subspecies of *Oreochromis niloticus* (Pisces: Cichlidae) using restriction endonuclease analysis of mitochondrial DNA. *Aquaculture*, 102(1-2), 29-42.
- Spencer, C., Neigel, J., & Leberg, P. (2000). Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology*, 9(10), 1517-1528. doi:<https://doi.org/10.1046/j.1365-294x.2000.01031.x>
- Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, 15(5), 199-203. doi:org/10.1016/S0169-5347(00)01825-5
- Tanya, C., & Kumar, R. (2010). Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology*, 2010. doi:10.4236/abb.2010.14037
- Tesfaye, G. (2010). Impacts of Anthropogenic Activities on Fish Diversity of the Ethiopian Rift Valley Lakes: A Review.
- Tesfaye, G., & Wolff, M. (2014). The state of inland fisheries in Ethiopia: a synopsis with updated estimates of potential yield. *Ecohydrology & Hydrobiology*, 14(3), 200-219. doi:<https://doi.org/10.1016/j.ecohyd.2014.05.001>
- Tibihika, P. D., Barekye, A., & Byakora, E. (2015). Fish species composition, abundance and diversity of Minor Lakes in south western Uganda/Kigezi region. *International Journal of Science and Technology*, 4(5), 204-213.
- Tibihika, P. D., Curto, M., Dornstaeder-Schrammel, E., Winter, S., Alemayehu, E., Waidbacher, H., & Meimberg, H. (2018). Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*, 1-16. doi:<https://doi.org/10.1007/s10592-018-1136-x>
- Tibihika, P. D., Okello, W., Barekye, A., Mbabazi, D., Omony, J., & Kiggundu, V. (2016). Status of Kigezi minor Lakes: A limnological survey in the Lakes of Kisoro, Kabale and Rukungiri Districts. *International Journal of Water Resources and Environmental Engineering*, 8(5), 60-73. doi:<https://doi.org/10.5897/IJWREE2016.0645>
- Trewavas, E. (1983). *Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia*. London: British Museum (Natural History).

- Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., . . . Cross, T. F. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society open science*, 3(1), 150565. doi:<https://doi.org/10.1098/rsos.150565>
- Verschuren, D., Johnson, T. C., Kling, H. J., Edgington, D. N., Leavitt, P. R., Brown, E. T., . . . Hecky, R. E. (2002). History and timing of human impact on Lake Victoria, East Africa. *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1488), 289-294. doi:10.1098/rspb.2001.1850
- Vicente, I., & Fonseca-Alves, C. (2013). Impact of introduced Nile tilapia (*Oreochromis niloticus*) on non-native aquatic ecosystems. *Pakistan Journal of Biological Sciences*, 16(3), 121-126. doi:<https://doi.org/10.3923/pjbs.2013.121.126>
- Vijverberg, J., Dejen, E., Getahun, A., & Nagelkerke, L. A. (2012). The composition of fish communities of nine Ethiopian lakes along a north-south gradient: threats and possible solutions. *Animal Biology*, 62(3), 315-335. doi:<https://doi.org/10.1163/157075611X618246>
- Vreven, E., Teugels, G., Adepo-Gourene, B., & Agnese, J. (1998). Morphometric and allozyme variation in natural populations and cultured strains of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Belgian Journal of Zoology (Belgium)*, 128(1), 23.
- Wilson, L. A., MacLeod, N., & Humphrey, L. T. (2008). Morphometric criteria for sexing juvenile human skeletons using the ilium. *Journal of Forensic Sciences*, 53(2), 269-278. doi:<https://doi.org/10.1111/j.1556-4029.2008.00656.x>
- Wirgin, I. I., & Waldman, J. R. (1994). What DNA can do for you. *Fisheries*, 19(7), 16-27. doi:[https://doi.org/10.1577/1548-8446\(1994\)019<0016:WDCDFY>2.0.CO;2](https://doi.org/10.1577/1548-8446(1994)019<0016:WDCDFY>2.0.CO;2)
- Witte, F., Wanink, J., & Kische-Machumu, M. (2007). Species distinction and the biodiversity crisis in Lake Victoria. *Transactions of the American Fisheries Society*, 136(4), 1146-1159. doi:<https://doi.org/10.1577/T05-179.1>
- Worthington, E. B. (1932). A report on the fisheries of Uganda investigated by the Cambridge Expedition to the East African Lakes, 1930-1931. Chap. IV: Lake Nabugabo. Crown Agents.
- Wudneh, T. (1998). *Biology and management of fish stocks in Bahir dar Gulf, Lake Tana, Ethiopia. (Ph.D. thesis) Wageningen Agricultural University, The Netherlands 144 pp.*
- Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). A Practical Companion to Geometric Morphometrics for Biologists: Running analyses in freely-available software: Free download from <http://booksite.elsevier.com/9780123869036/> (18 June 2013).
- Zengeya, T. A., Booth, A. J., & Chimimba, C. T. (2015). Broad niche overlap between invasive Nile tilapia *Oreochromis niloticus* and indigenous congeners in southern Africa: Should we be concerned? *Entropy*, 17(7), 4959-4973. doi:10.3390/e17074959
- Zengeya, T. A., Robertson, M. P., Booth, A. J., & Chimimba, C. T. (2013). A qualitative ecological risk assessment of the invasive Nile tilapia, *Oreochromis niloticus* in a sub-tropical African river system (Limpopo River, South Africa). *Aquatic Conservation: Marine and Freshwater Ecosystems*, 23(1), 51-64. doi:<https://doi.org/10.1002/aqc.2258>

2. Chapter 2

1st Paper. Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa



Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa

Papius Dias Tibihika · Herwig Waidbacher · Charles Masembe · Manuel Curto · Stephen Sabatino · Esayas Alemayehu · Paul Meulenbroek · Peter Akoll · Harald Meimberg

Received: 28 May 2017 / Accepted: 4 December 2017 / Published online: 8 January 2018
 © The Author(s) 2018. This article is an open access publication

Abstract Nile tilapia occurs naturally in East Africa where it's an economically important species. Many of the natural populations of Nile tilapia have been affected by anthropogenic activities including translocations, associated with programmes aimed at enhancing capture fisheries and aquaculture productivity. Using geometric morphometric analyses, we tested the hypothesis that such anthropogenic activities have augmented admixture among natural populations of Nile tilapia and

influenced the geographical distribution of morphological variation within the species. Our expectation was that Nile tilapia shape divergent might be consistent with reportedly anthropogenic activities in nonnative environments. To test this hypothesis, we analyzed the shapes of 490 individuals from thirteen populations: three farms, six natives and four nonnative natural populations. Our analysis revealed that the most pronounced shape diversification was observed in seven populations: three nonnatives (Victoria, Kyoga and Sindi farm) and four natives (Albert, River Nile, George and Turkana). The features responsible for the observed morphotypes were mainly related to the orientation of the anterior region of the fish and may be due to diversifying selection in response to new environmental pressures (for nonnative populations), admixture or drift. Shape change in the nonnative high-altitude populations was unexpectedly conserved, suggesting recent introductions which may have not resulted in admixture or there was strong selection against change in the traits measured. On the other hand, the recorded morphotypic clusters explained the possible genetic link to their putative ancestral home. Our results were partially consistent with our prediction that the nonnative populations exhibited divergent morphotypes. We recommend further investigations with molecular genetics for follow-up of these findings.

P. D. Tibihika · M. Curto · E. Alemayehu · H. Meimberg
 Institute for Integrative Nature Conservation Research, University of Natural Resources and Life Sciences (BOKU) Vienna, Gregor Mendel-Straße 33, 1180 Wien, Austria

P. D. Tibihika (✉)
 National Agricultural Research Organization, Kachwekano Zonal Agricultural Research and Development Institute, P.O. Box 421, Kahale, Uganda
 e-mail: papius.tibihika@students.boku.ac.at
 e-mail: papiust@yahoo.com

H. Waidbacher · P. Meulenbroek
 Institute of Hydrobiology and Aquatic Ecosystems Management, University of Natural Resources and Life Sciences (BOKU) Vienna, Gregor-Mendel-Straße 33/DG, 1180 Wien, Austria

C. Masembe · P. Akoll
 Department of Zoology, Entomology and Fisheries Sciences, Makerere University Kampala, P.O. Box 7062, Kampala, Uganda

S. Sabatino
 CIBIO, Research Center in Biodiversity and Genetic Resources / InBio Associated Laboratory, University of Porto, Rua Padre Armando Quintas 7, 4485-661 Vairão, Portugal

Keywords Nile tilapia · Geometric morphometrics · MorphoJ · Populations · Condition factor · Morphological diversification

Introduction

Nile tilapia (Cichlid) is one of the many native East African (EA) freshwater species that have been greatly impacted by anthropogenic activities. Partly due to the ease in which tilapia can be farmed and thrive when introduced into new areas and its market demand, the species has now been spread well beyond its native ranges via human-mediated translocations associated with aquaculture (George 1995; Vreven et al. 1998; Eknath et al. 2007). The species is also stocked in many natural environments but overfishing has decimated native populations. The impact of these potential stressors on native Nile tilapia however, have not been well studied.

In EA, Nile tilapia is native to the Lakes George, Edward, Albert, Kivu, Tanganyika, Turkana, Baringo and to the River Nile (Trewavas 1983; Ogutu-Ohwayo 1990; Agnèse et al. 1997). The cichlid was introduced to the remaining large water bodies, lakes Victoria, Kyoga, Nabugabo and to numerous satellite lakes including various aquaculture setups (Ogutu-Ohwayo 1990; Boyd 2004). Generally, the main motive of Nile tilapia translocations in EA was to reinforce the fishery after native species had been overfished, and to stock water bodies that had been considered naturally unproductive (Bairwa 1994, 1992). Apparently, Nile tilapia in non-native water bodies since the 1950s, has become the commercially most important Tilapiine (Ogutu-Ohwayo 1990) in particular after many native fish species had declined due to competition as has been well studied in Lake Vitoria (Witte et al. 2007).

The success of stocking Nile tilapia in nonnative environments is attributed to several factors including its opportunistic omnivore, growth rate, and prolific breeding in artificial and natural ecosystems, the species strategies of predator avoidance and diversified habitats (Ogutu-Ohwayo 1990; Bairwa 1992; Eknath et al. 2007). The effectiveness of Nile tilapia to thrive in a broad range of habitats may also be reflected in the condition coefficient or factor of the species. Condition factor (K) of fish, also referred to as Fulton's condition factor, predicts the adaptability of the species by determining the status of the individual as weight at a given length (Ricker 1975). Adaptability might be one expression of a population's ability for morphological diversification that can be induced environmentally, either as a plastic or evolutionary response (Moezck et al. 2011). In extreme cases, morphology can diverge into distinct types

as evolutionary response to resource polymorphism (Robinson and Wilson 1994; Eklöv and Svanbäck 2005). Morphology can be influenced by a variety of environmental forces. This might not only occur in response to environmental pressures, but also adaptation to mimic against predation, defend territories and for foraging functions (Mekkawy and Mohammad 2011; El-Zaem et al. 2012). For example, Crucian carp, *Carassius carassius*, developed deeper body morphotypes to somewhat render predator's difficulty for handling (Scharnweber et al. 2013). Pumpkinseed sunfish (*Lepomis gibbosus*) have also diversified their shapes with increased body depths and dorsal fins in response to predator threats. (Januszkiewicz and Robinson 2007).

Nevertheless, divergence of phenotypic traits might be induced anthropogenically. When, for instance organisms are shifted from their natural habitats, as has been the case for Nile tilapia in EA, possible admixture may lead to shape changes when divergent gene pools are brought together and form recombinant individuals, as in form of hybridization (Sinama et al. 2013). Hybrids might form typical morphs divergent from their parent lines. For example, hybrid specimen from cyprinid fish species have been reported to show extreme body forms resulting from widening of the lower lip of their mouths contrary to their parental lines (Corse et al. 2012), as an example for transgressive segregation (Seehausen 2004; Sinama et al. 2013).

In Nile tilapia, hybridization with its congeneric species has been reported in Lake Edku, Egypt and in River Kafue, Zambia by Bakhoun et al. (2009); Deines et al. (2014) respectively. It is possible that in EA, Nile tilapia admixture might be complex, likely an intermingle of species from various sources including aquaculture, either intentional or accidental escapees. Despite the possibility of admixture leading to various species' morphs, the result might also pose other constraints to the populations. For example, Perrier et al. (2013) indicated loss of local adaptation and fitness reduction in wild Salmonids resulting from captive salmonids admixture. In cases where admixture leads to fitness reduction, natural reinforcements might result into displacement of the less fit characters to the periphery and hence the evolution of distinct morphotypes (Dodson et al. 2015). On the other hand, generated fit hybrids might form new populations that remain distinct from the ancestry lines, a condition which may lead to hybrid speciation (Mallet 2007). Therefore, critical studies linking anthropogenic influx on to species functionality

across environmental gradients might provide a central goal in understanding evolutionary biology of organisms. This might be essential for management and conservation of the resource.

In EA, Trewavas (1983) and Vreven et al. (1998), delineated Nile tilapia morphotypes employing traditional morphometrics (TM). This involves physical linear measurements of the fish anatomical parts like head length, body depth, total length, standard length, eye diameter, fin lengths, mouth gape length, among others. TM approaches are time consuming, biased/subjective and have been criticized as uninformative (Maderbacher et al. 2008; Wilson et al. 2008; Ponton et al. 2013; Petrtyl et al. 2014). Besides, most studies focused on West African water bodies and Egypt, with little attention to EA. Some approaches on fish morphometry like geometric morphometrics (GM) have appeared to yield good results (Breno et al. 2011). GM is defined as the study of form in two or three dimensional spaces which induce in-depth investigation of morphological modeling in organisms (Bookstein 1982). GM allows the visualization and quantification of morphological differences in individuals represented by deformations of landmarks (Lms) in a coordinate space (Costa et al. 2010; Park et al. 2013). The arrangement of morphological Lms as data source and shape variation is extracted by Procrustes superimposition (Klingenberg 2011). When Maderbacher et al. (2008); Park et al. (2013); Ponton et al. (2013); Petrtyl et al. (2014), studied shape variations in fish comparing TM and Geometric morphometrics (GM). TM methodologies were realized weak contrary to GM. In addition, Rohlf (2002) employed GM while studying shape variation in Bluefin tuna, and was successful. Furthermore, comparisons between TM and GM on *Mastomys natalensis* (Rodentia: *Muridae*) skull morphology during growth, revealed GM more effective than TM (Breno et al. 2011). Use of GM thus proves that its application might be powerful and hence a likelihood of unleashing essential tools for cataloguing Nile tilapia populations in EA.

So far GM on Nile tilapia in EA is only available but limited to a few lakes and swamps in Kenya but with a larger scale missing (Ndiwa et al. 2016). It should, moreover, be a good method to link morphotypes in native and nonnative ranges with potential admixture given the numerous translocations. Nile tilapia populations react also to environmental factors, abiotic (low temperatures as for high altitudes) and biotic interactions, e.g. Nile perch predator influence. Comprehensive

knowledge of the patterns of shape variation across various ranges by GM might be informative to determine such factors. The two principal advantages of the GM methodology include: 1) Geometric or shape information is extracted by means of Procrustes superimposition which filters variationally effects due to specimen size, position and orientation from the data (Klingenberg 2011). This procedure remains powerful, reliable and the most prevalent methodology that removes unnecessary variations in the data which would ultimately offer misleading information (Kerschbaumer and Stumbauer 2011). 2) the mathematical portrayal of different morphs can be rendered as graphical images illustrating where the shapes differ, in which direction and by how much (Zelditch et al. 2012).

Using GM, the objective of this study was to investigate the impact of anthropogenic activities on morphological divergence and potential adaptation of Nile tilapia populations from various environments. Given the strong evidence of the species translocations to various nonnative zones in EA, we hypothesize that possible interspecific species admixture and varying environments might have led to morph divergence among the populations in the region. We determined morphological differences using GM for native and nonnative populations, including fish from aquaculture and nonnative highland in respect to K. Investigating Nile tilapia K especially in nonnative high-altitude might be crucial as the species lay at the edge of its physiological range (Popma and Masser 1999). With this approach, we answer the following questions: 1) Is condition factor (K) appropriate to explain adaptability or performance of Nile tilapia populations? 2) To what extent are the introduced populations morphologically different to their putative native sources? 3) At what level are the Lake Victoria metapopulations morphologically distant owing to potential multiple introductions or to its size?

Materials and methods

Study areas/locations

Nile tilapia species were sampled from water body types: (1) Fish farms (Sindi, Bagena & Rwitabingi), (2) Lakes and Rivers with native populations (Albert, Edward, George, Turkana, Kazinga Channel & River Nile), (3) low altitude lakes with introduced populations (Kyoga & Victoria), and (4) high altitude Lakes with

introduced populations (Mulehe and Kayumbu; Fig. 1). All the native populations are at a low altitude despite variations (Table 1).

Brief overview of sampling locations/studied sites

The number of sampling locations/sites per Lake or River depended on the size of the water body, and were set after consulting experienced fishermen and local government fisheries enforcement officers about the type of variation that exists in each habitat. In Lake Victoria, largest lake in Africa in terms of surface area, five locations were sampled in contrast to two and or one for Lakes Albert, Edward, Kyoga and George, Turkana, Mulehe, Kayumbu, respectively (Fig. 1; Table 1). The strong proportion of Uganda sampling sites resulted from the fact that five of the great East African Lakes and River: Victoria, Albert, Kyoga, Edward, George and River Nile, respectively lie within the country (Byrnes

2010). However, it should be noted that most of these herein mentioned water bodies in the region are shared, apart from L. Turkana (Kenya), Kyoga and George (Uganda). Lakes Edward and Albert are shared between Uganda and Democratic Republic of Congo (Former Zaire) and L. Victoria is shared amongst the three East African countries (Uganda, Kenya and Tanzania).

Many of the lakes are connected by channels or rivers, e.g. L. Victoria and lakes Kyoga and Albert link via River Nile; L. George is associated to L. Edward via a large flowing water mass, Kazinga Channel and L. Edward to L. Albert through River Semliki (Green 2009; Byrnes 2010; Cockerton et al. 2013), (also see Fig. 1). From the South-Western Ugandan High Altitude Lakes (satellites), Mulehe and Kayumbu (Table 1; Fig. 1), Kayumbu seems not to have any outlet but L. Mulehe connects to another satellite, L. Mutanda in the south via Mucha (a small river) which links via River Kaku to L. Edward (Green 2009; Tibihika et al. 2016). Nile tilapia



Fig. 1 Locations of sampling points for both natural and artificial populations. Artificial populations herein also are referred to as fish farms, pond culture

Table 1 Detailed studied water bodies, locations, elevation and existing GPS coordinates for sites. The water bodies herein are also referred to as Nile tilapia populations, used in the current study

Water body	Sample location	Elevation (masl)	Population type	No.	Size (cm)	Wt (g)	UTM coordinates
Victoria	Kakungu	1136	NNL	37	30.9 ± 3	645.8 ± 252.2	N.00.18079 ^a E032.29332 ^a
Victoria	Masese	1136	NNL	30	16.8 ± 1.8	115.8 ± 42.7	N.00.43650 ^a E033.24081 ^a
Victoria	Gaba	1146	NNL	30	15.6 ± 2	91.3 ± 34.7	N.00.25819 ^a E032.63727 ^a
Victoria	Sibay	1129	NNL	30	21.3 ± 3.2	233.1 ± 106.6	N.00.86772 ^a E031.71332 ^a
Victoria	Kamu	1139	NNL	30	20.2 ± 4.7	226.2 ± 192.1	S.00.12747 ^a E031.93999 ^a
Albert	Nturoko	618	N	30	33.2 ± 3.6	784.4 ± 266.5	N.01.05206 ^a E030.53464 ^a
Albert	Kyehoooro	615	N	16	32.0 ± 3.9	755.7 ± 256.8	N.01.50990 ^a E030.93610 ^a
George	Humuhungu	916	N	40	20.2 ± 4.7	226.2 ± 192.1	S.00.01739 ^a E030.09694 ^a
Karungu C	Katu	915	N	30	25.0 ± 1.4	380.6 ± 41	S.00.12541 ^a E030.04744 ^a
Edward	Karungu	914	N	30	30.9 ± 1.9	668.9 ± 135.2	S.00.20783 ^a E029.89252 ^a
Edward	Rwenzu	908	N	31	25.4 ± 1.1	363.7 ± 46.2	S.00.40459 ^a E029.77283 ^a
Kyoga	Kibuye	1054	NNL	50	19.3 ± 1.8	187 ± 56.6	N.01.40028 ^a E032.57949 ^a
Kyoga	Bukanga	1045	NNL	30	24.6 ± 2.7	359.4 ± 124	N.01.43873 ^a E032.36809 ^a
River N	Kibuye	1062	N	30	23.4 ± 3.0	300.2 ± 110.8	N.01.18734 ^a E032.98865 ^a
Mulhe	Musenem	1801	NNH	31	21.8 ± 1.7	235.7 ± 59.7	S.01.21345 ^a E029.72668 ^a
Kayumbo	Raga	1910	NNH	34	17.2 ± 2.1	116.7 ± 38.1	S.01.34679 ^a E029.78448 ^a
Bagena	Bilim	1857	Farm	15	16.7 ± 4.2	104.7 ± 36	S.01.25617 ^a E029.73622 ^a
Sindi	Sindi farm	1733	Farm	30	20.3 ± 1.5	165.3 ± 28.3	S.01.17579 ^a E030.06198 ^a
Rwita	Rifim	1069	Farm	30	20.4 ± 1.5	156.7 ± 31.5	N.00.97116 ^a E033.13924 ^a
Turkana	Longoth	364	N	40	24.4 ± 3.4	308.9 ± 118.6	N.03.55617 ^a E035.91599 ^a

masl/meters above the sea level, No Number of samples, UTM/Universal Transverse Mercator coordinate system, Karungu C Karungu Channel, River N River Nile, Rwita Rwitabingi, Sibay Sango Bay, Kamu Kamuwanga, Katu Katungu, Rwenzu Rwenzuhamu, Raga Ragarumbe, Bilim Bagena farm, Rifim Rwitabingi farm, NNL Nonnative low altitude, N Native, NNH Nonnative high altitude

stocking in Lakes Mulhe and Kayumbo is reported to have taken place between 1920s and 1950s (ISSON 1960; Green 2009), with some reports from the Local Government archives indicating restocking continued up to the 80s. L. Turkana (previously called Rudolf), located in the Great Eastern Rift Valley is the largest freshwater body in Kenya (Kolding 1992). The Lake lies in the arid North-Western part of Kenya at an altitude of 364 m above the sea level, much lower than most of the other East African freshwater bodies (Table 1). Surprisingly, despite L. Turkana having much of its inlet from River Orso (originating from Ethiopian highlands), the Lake does not have any surface outlet and hence, the water balance is compensated for by evaporation, estimated at 2.3–2.8 m/yr. (Kolding 1992). Peripheries for most of these lakes are populous and user conflicts have amplified the demand for fisheries resources culminating into overfishing (Ogutu-Othman et al. 1997). Previous limnological studies regarding the herein mentioned E.A. water bodies have indicated increased eutrophication,

with oxygen levels and transparency (Secchi depth) decreasing, *Chlorophyll a* and siltation increasing, among others, all attributed to environment/catchment destruction particularly in Lake Victoria basin (Njiru et al. 2012). Additionally, translocations of fish species to nonnative homes, for instance a voracious predator, Nile perch (*Lates niloticus*) from Lake Albert to Victoria, Kyoga and Nabugabo, also have left majority of the cichlids depleted through predation pressure, a concern to the future sustainability of the resources (Ogutu-Othman et al. 1997). Therefore, it was thoughtful for sampling the above various water bodies for unfolding insights into the current form of Nile tilapia.

Fish sampling procedures

Samples were mainly collected by gill nets overnight apart from one sampling location, L. Victoria (Gaba) where a cast net was employed and a beach seine net on L. Turkana. Upon landing, specimens were inspected by

eye for possible morphological malformations. In cases where samples were malformed due to fishing artefacts, they were rejected and hence not utilized for subsequent analysis. Individual fish weight was taken on a portable digital weighing scale and the total length using a fish measuring board. The specimens were then tagged and photographed using a digital camera (Canon IXUS 275 HS, 12× optical zoom). Specimen photography was standardized using a wooden tripod stand with inserted mesh cradle for displaying fish. The mesh cradle prevents shape variations which would arise from differences in weight when lying on a plane surface (Muir et al. 2012). Digital images were always taken from the left side of each specimen displayed on a cradle at a focal length of 33 cm. Fish sampling/field was conducted in the dry season, Months of July to December 2016, and a total of 490 fish specimens were analyzed.

Fish coefficient of condition/condition factor

In general, the coefficient of condition or condition factor of an organism can be regarded as a measure of suitability in an environment. Condition factor is also regarded as a degree of the general physical health of fish based on fatness (Ogle 2013). Furthermore, a thin fish with purportedly low condition value might suggest muscle wasting resulting from starvation or less favorable environmental conditions (Froese 2006). Therefore, the condition factor can be an important indicator for fisheries management to estimate favorable environmental conditions for a fishery. In the current study, to have an insight in the over-all well-being of fish populations from different habitats which could be an indicator of adaptability and/or performance, we determined condition factor (K) from the expression: $K = \frac{100W}{L^3}$, where, W = weight of individual fish (g) and L = Total length of individual fish (cm) (Ricker 1975).

Landmark acquisition

Landmark programming was conducted by two Thin-plate-spline (Tps): Tps utility version 1.69 (TpsUtil) and Tps digitize version 1.40 (TpsDig) programs (Rohlf 2004; Zelditch et al. 2012). TpsUtil was used to build tps files from individual fish digital images. The tps files were then imported into TpsDig which was later used to digitize 10 Lms (Fig. 2) on each image and to generate a two dimensional x,y coordinates (Zelditch et al. 2012;

Park et al. 2013). Digital images were scaled along a ruler and all the 10 homologous Lms were digitized sequentially on the biological images (Fig. 2). The anatomical descriptions and location of each landmark are presented in Table 2 and Fig. 2. To minimize errors during Lms digitization, only one person performed the procedures. The Tps programs were freely downloaded from the site: <http://life.bio.sunysb.edu/ce/rohlf/software.html> (Zelditch et al. 2012).

Data analysis

Populations were pooled in different hierarchical levels (Ondhoro et al. 2016). One-way ANOVA was used to determine K variations among Nile tilapia populations using SAS for windows version 9.4. K values were taken as dependent variables and populations or population groups as explanatory variables. Morphological variation was analyzed using MorphoJ, vers. 1.06d (Klingenberg 2011). MorphoJ is an Integrated Morphometrics Package (IMP) for conducting multivariate analysis for shape variations. (http://www.flywings.org.uk/morphoj_page.htm; (Zelditch et al. 2012). Landmark coordinates generated by TpsDig were imported into MorphoJ and shape information was extracted by Procrustes Superimposition. The existence of outliers in the shape data was inspected by graphically plotting cumulative frequency against squared Mahalanobis distance implemented by MorphoJ (Klingenberg 2011). The superimposed landmark coordinates were then used to generate a covariance matrix as input for multivariate analyses: Principle Component Analysis (PCA), Canonical Variate Analysis (CVA), and nonmetric multidimensional scaling (NMDS). PCA was performed to investigate the main features of shape variation in samples and morpho-space with the axis describing highest amount of variation used as the first principle component (PC1), the next highest amount of variability becomes PC2 and so on, until variability becomes insignificant and therefore less important to describe the data (Jolliffe 2002). CVA was performed as an ordination type which maximizes the separation or discrimination of specified populations (Klingenberg 2011). CVA uses the assumption that the covariance structure within the groups is the same, thus utilizes the pooled within-group covariance matrix to separate populations (Klingenberg 2011; Zelditch et al. 2012). In case CVA violated the assumption, we conducted NMDS to validate results. With NMDS, a smaller number of axes are a priori chosen



Fig. 2 Representatives of Nile tilapia populations from EA illustrating digitized landmarks used in geometric morphometrics. A = Albert. E = Edward. G = George. KC = Kazinga Channel. RN =

River Nile. T = Turkana. Ky = Kyoga. VK = Victoria Kikuyu. M = Mulehe. Ka = Kayumbu. Rf = Rwtabingi farm. Sf = Sindi farm and Bf = Bagena farm

Table 2 Description of the landmark points, adopted and modified from: (Costa et al. 2010; Lorenz et al. 2014; Ndiwa et al. 2016; Peligro and Jurnawan 2016)

Landmark number	Location description
1	Upper tip of the snout/mouth
2	Orbit/eye center
3	Anterior insertion or origin of the dorsal fin
4	Posterior insertion or origin of the dorsal fin
5	Dorsal insertion of caudal fin
6	Ventral origin of caudal fin
7	Posterior insertion of anal fin
8	Dorsal base of pectoral fin or anterior origin of pectoral fin
9	Most posterior edge of operculum or apex of the opercular spine
10	Ventral edge of operculum

and all the data are fitted to those dimensions, implying unlimited graphical space, contrary to PCA or CVA. Furthermore, NMDS is less sensitive and makes fewer assumptions about the data and therefore suits a wide variety of data sets (Holland 2008). NMDS was performed using paleontological statistical software 'PAST' (Hammer et al. 2001); http://palaeo-electronica.org/2001_1/past/issue1_01.htm).

Results

Coefficient condition (K)

Generally, all populations' K values were between 1.5 and three (Fig. 3). Fish farms had on average lower K values than the wild populations. Over-all, among the

farms. Rwitabingi, indicated the least K values and significantly different from Sindi farm ($p < 0.05$) (Fig. 3a). Kazinga Channel, Victoria, George and Kyoga populations had equally high K values ($p > 0.05$). Native populations in Albert and Turkana indicated similar lower K values and not significantly different ($p > 0.05$), but significantly different ($p < 0.05$) from other natural populations (Fig. 3a). K values between nonnative high-altitude populations was not significantly different ($p > 0.05$), but surprisingly indicated higher values (greater than 2) and significantly different ($p < 0.05$) from Turkana, a native population (Fig. 3a). The subpopulations of *L. Victoria* showed K values greater than two each (Fig. 3b). One subpopulation, Kakyanga from the deeper waters, indicated a K value significant lower from the rest ($p < 0.05$; Fig. 3b). The remaining subpopulations (Gaba, Kamuwunga, Masese and Sango Bay) had relatively high K values on the same significance level.

Geometric morphometrics-all populations

The major trends of shape variation in all populations were derived from the Procrustes analysis as illustrated by PCA and the principle components (PCs) of the covariance matrix from which the first four PCs explained 60.8% of the total variance. Most of the populations overlapped to some extent along the first two PCs, but differentiation among populations was

observed on the on the graphs for PC1 and 2 (Figs. 4a, b and c). The first PC axis indicated that River Nile, Lakes Victoria, Albert, George, Kyoga, Turkana and Sindi farm morphotypes were considerably more variable than the rest of the populations (Figs. 4a and c). Interestingly, PC2 showed Rwitabingi farm and again River Nile, Albert, Sindi, Edward, Kyoga, Victoria with Kazinga Channel and Bagena more variable (Fig. 4b and c). PC2 indicated evidence of morphological shape differences within some populations (River Nile, Albert, Kyoga, Victoria and Sindi). Bagena and Kayumbu populations showed the least PC scores (Figs. 4a and c). In general, shape change was encountered mainly in the anterior region of the fish, with the orientation of mouth tip (landmark 1), dorsal fin region (Landmarks 3&4), origin/insertion of the caudal fin (landmarks 5 & 6) opercular spine, dorsal insertion of pectoral fin (landmarks 9,8 respectively) (Fig. 4 W1 and W2). Result of CVA was congruent to PCA. Mostly, populations which indicated corresponding PC scores, tended to cluster together (Fig. 5a and c). Across CV1, River Nile, Kyoga, Albert Victoria and Sindi populations showed relatively higher CV scores. Similarly, CV1 and CV2 separated the populations in total accounting for 61.64% of variation. CVA results showed three morphotypic clusters formed between populations: Victoria-Albert-Kyoga-River Nile (cluster 1); Kazinga Channel-Edward (cluster 2) and Bagena-Kayumbu (cluster 3) (Fig. 5a and c).

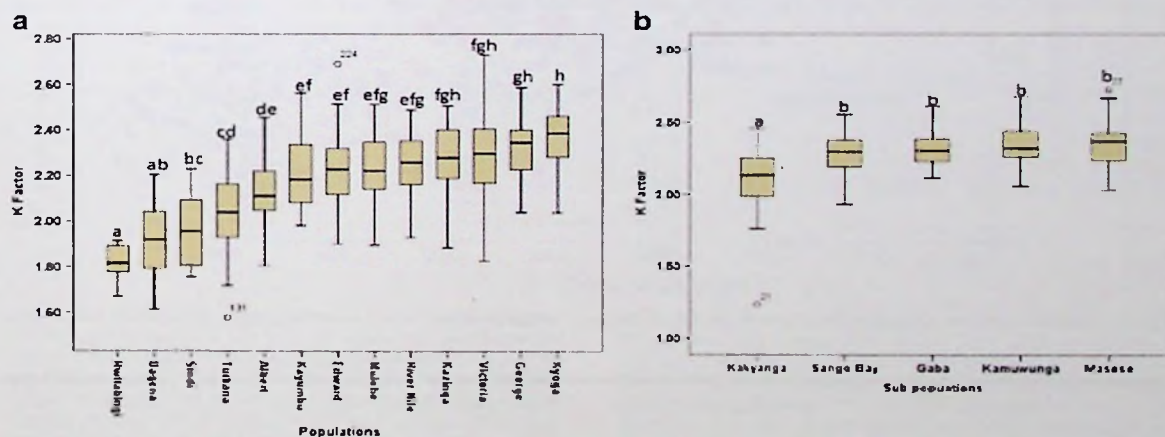


Fig. 3 Box plot graphs showing condition factor (K) variations among Nile tilapia populations. Graph a represents comparisons for all populations; fish farms (Bagena, Rwitabingi and Sindi).

nonnative high altitude (Mulhe and Kayumbu), nonnative low altitude (Victoria and Kyoga) and the rest natives. Graph b represents singled out *L. Victoria* subpopulations

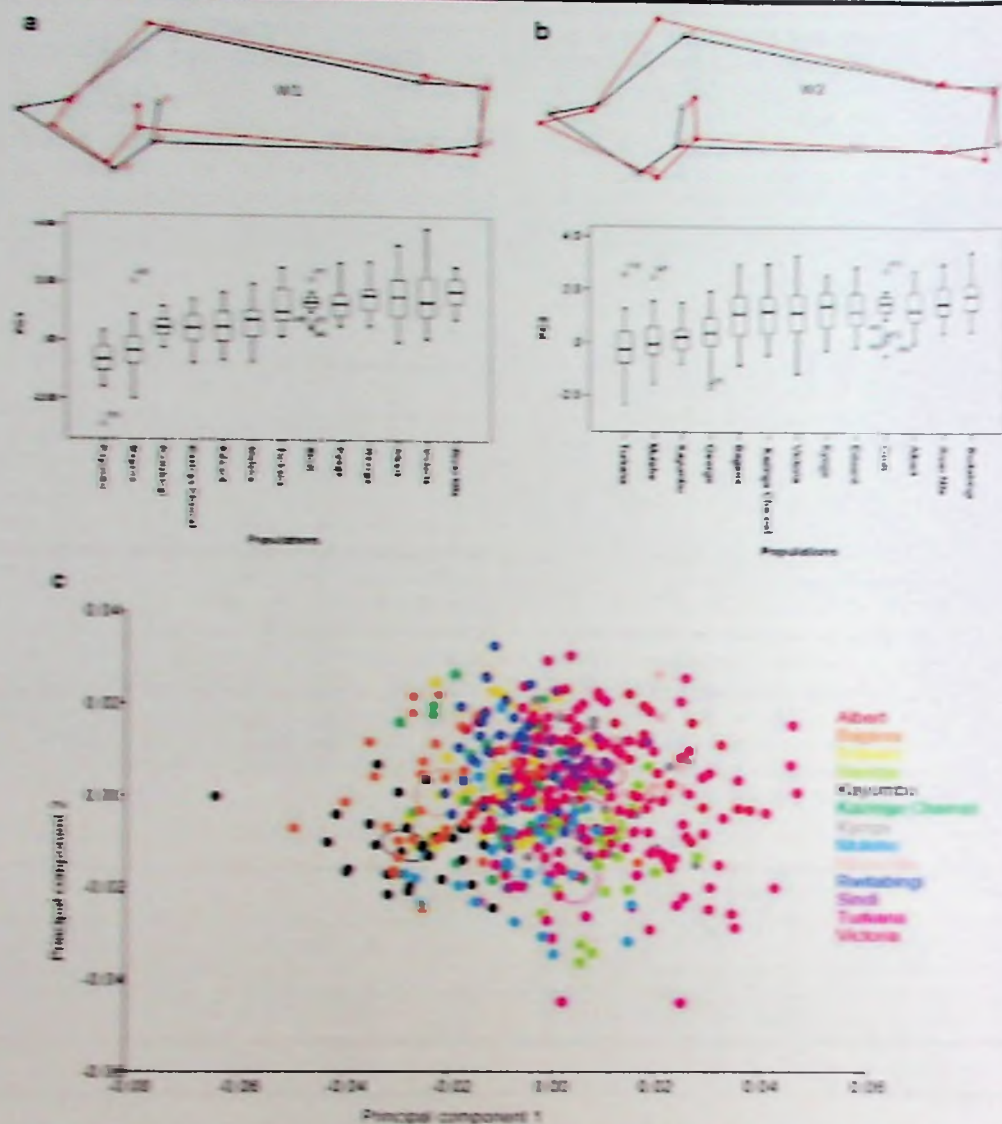


Fig. 6 Principle Component Analysis for all populations: a, b and c represent PC1, PC2 scores and scatter plot for pooled components in a larger space respectively. Circles in scatter plot represent 95% confidence ellipses drawn for means. W1 and W2 are wireframe graphs for PC1 and PC2 illustrating deformations for major features behind the observed shape variations, with blue and light blue wireframes indicating the shape change and ideal shape respectively. PCA graphics: a and b represent PC1 and PC2 scores respectively and c scatter plot for all pooled components or

t-tangent space. Circles in scatter graph represent 90% confidence ellipses drawn for means. W1 and W2 are wireframe graphs for PC1 and PC2 illustrating deformations for major features behind observed shape divergence. Blue wireframe graphs indicate the target shape (shape change) and light blue wireframe graphs indicate the starting shape (ideal shapes). The PC scores, indicated on box plot graphs, were exported from Minitab, standardized for clear visualization and plotted using S-Plus statistical program.

NMDS results were consistent with CVA, with Victoria, Albert, Kyoga, and River Nile most similar, indicated by the shortest distance on minimal spanning tree (Fig. 5d). Hereby Kyoga was closer to Albert than to Victoria (Fig. 5d). Other pairs of similarity were Edward and Kazinga Channel, as well as Kayumbu and Bagenia farm (Fig. 5d). The rest of the populations, generally demonstrated distal morphotypes (Fig. 5d).

Geometric morphometrics-*L. Victoria* metapopulations

The shape variation among the five sub populations in *L. Victoria* was found to be more pronounced for Sango Bay and Kakyanga as indicated by the first PC (Fig. 6a and c). Generally, Masese, Kamuwunga and Gaba were observed to be less morphologically variable (Fig. 6a and c). On the other hand, PC2 loaded the rest of the

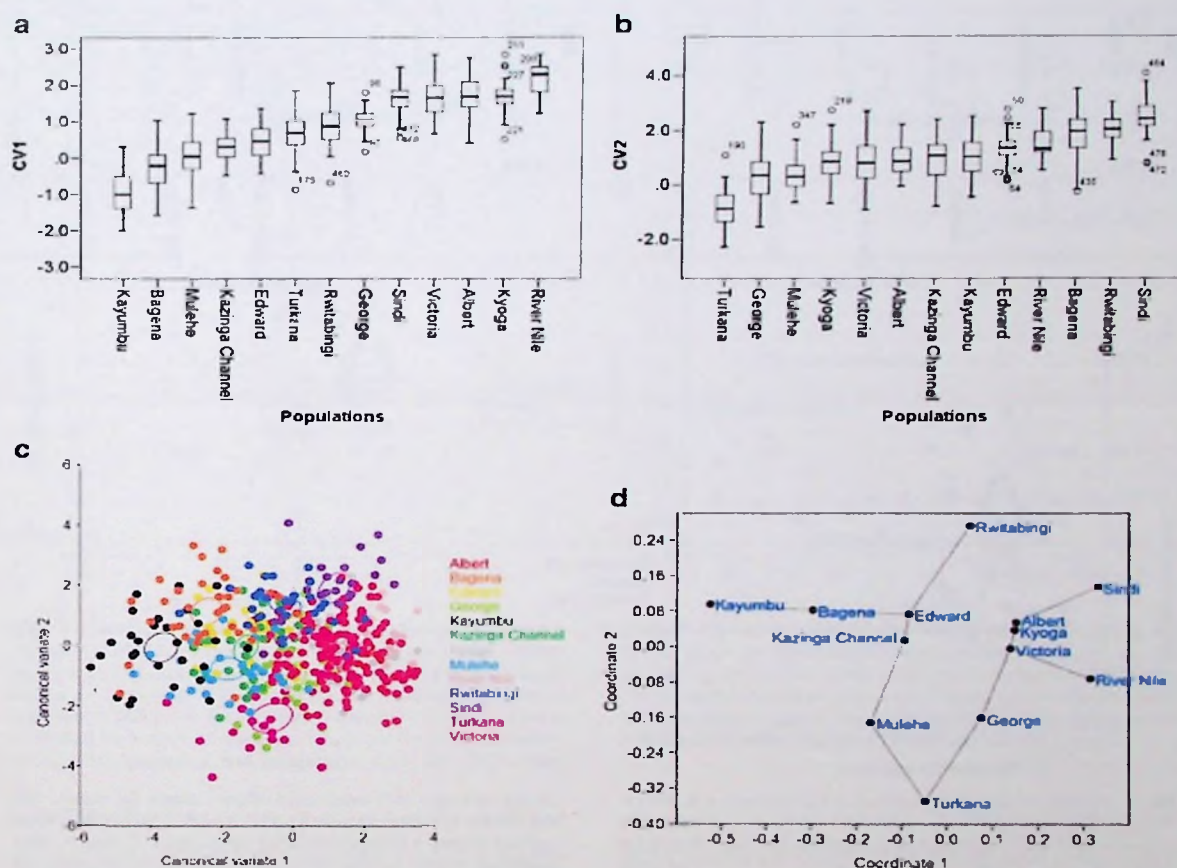


Fig. 5 Canonical Variate Analysis for all populations: **a**, **b** and **c** are CV1, CV2 scores and scatter plot for pooled variates on a tangent space. Circles in the scatter plot are confidence ellipses drawn for means at 90% probability, set at 10000 iterations. **d** represent NMDS plot for population morph closeness or distant explained by a minimal spanning tree set on Euclidean similarity index. CVA graphs: **a** and **b** are CV1 and CV2 respectively and **c** scatter plot for pooled variates on a tangent space. Circles in the

scatter graph are confidence ellipses drawn for means at 90% probability, set at 10000 iterations. **d** represent NMDS plot for population morph closeness or distant explained by a minimal spanning tree line set on Euclidean similarity index. The CV scores, indicated on box plot graphs, were imported from MorphoJ, standardized for clear visualization and plotted using SAS statistical program

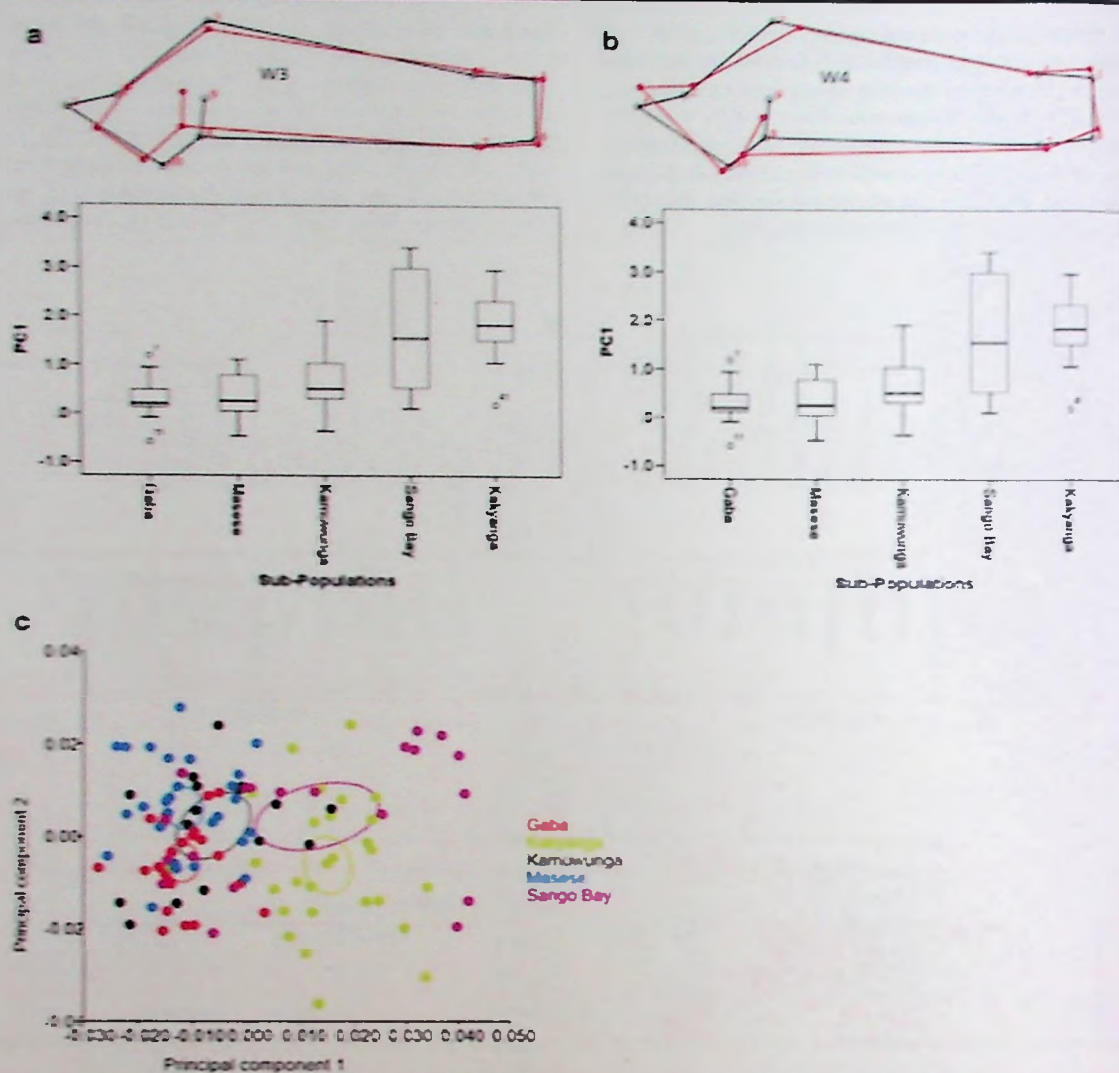


Fig. 6 Principle Component Analysis for *L. Victoria*: **a**, **b** and **c** are PC1, PC2 scores and scatter plot for the pooled components respectively. Circles represent 90% confidence ellipses drawn for means. W3 and W4 are wireframes for PC1 and PC2 illustrating deformations for major features behind observed shape divergent, with blue and light blue indicating the shape change and ideal shape respectively. PCA graphs: **a** and **b** represent PC1 and PC2 scores respectively and **c** scatter plot for the pooled components.

Circles represent 90% confidence ellipses drawn for means. W3 and W4 are wireframe graphs for PC1 and PC2 illustrating deformations for major features behind observed shape divergent. Blue wireframe graphs indicate the target shape (shape change) and light blue wireframe graphs indicate the starting shape (ideal shape). The PC scores for the graphs (a and b) were plotted similarly to Fig. 4

populations with Masese, Sango Bay and Kamuwunga scoring higher on the axis (Fig. 6b). The major features

for shape divergence were the same landmarks for the anterior region like the other populations (Fig. 6 W3

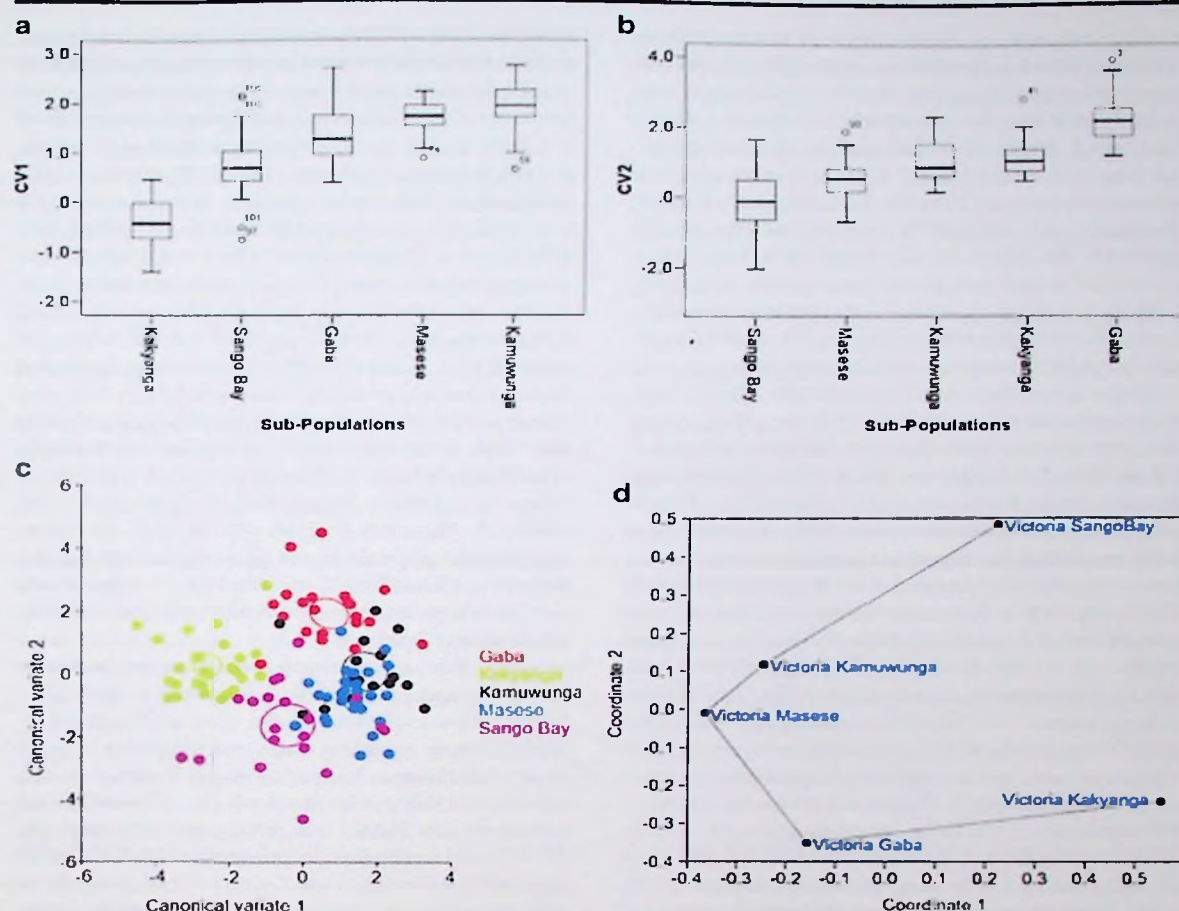


Fig. 7 Canonical Variate Analysis for *L. Victoria*: **a**, **b** and **c** represent CV1, CV2 scores and scatter plot for the pooled variates respectively. Circles in the scatter graph are confidence ellipses drawn for means at 90% probability, set at 10000 iterations. **d** represent NMDS plot for population morph closeness or distant explained by a minimal spanning tree set on Euclidean similarity index. CVA graphs: **a** and **b** represent CV1 and CV2 scores

and W4). The first three canonical variates extracted from the five sub populations explained 92% of the total variance, with 54.5% for CV1. Clear separation or discrimination of the sub populations was observed along CV1 axis where Kamuwunga-Masese were observed to form a cluster (Fig. 7c). Kakyanga, Sango Bay and Gaba sub populations were isolated on the CV axis. Similarly, analytical results from NMDS were consistent with CVA, and indicated

respectively and **c** scatter for the pooled variates. Circles in the scatter graph are confidence ellipses drawn for means at 90% probability and set at 10000 iterations. **d** represent NMDS plot for population morph closeness or distant explained by a minimal spanning tree line set on Euclidean similarity index. Graphs **a** and **b** were produced in the same way as in Fig. 5

close morphotypes between the subpopulation pair. Kamuwunga and Masese (Fig. 7d).

Discussion

Fisheries in developing countries are negatively impacted by poor enforcement and adherence to regulatory policies. One example of this failure in fisheries

management are the Great Lakes of Eastern Africa, which are vital for the rapidly increasing human population in the region, but are also biodiversity hotspots with a remarkable number of endemic fish species (Lowe-McConnell 2003). Over the past six decades the resources have been impacted majority by anthropogenic activities (Mugidde 1993; Ogutu-Obwayo et al. 1997; Wenzeluren et al. 2002). One of the most affected fish species in the region is Nile tilapia, which has had a significant impact following translocation to various habitats including aquaculture. The importance of the species in the region is significant and thoughtful attention is needed to ensure its future sustainability.

In the present study, K values for all the Nile tilapia populations were above 1.5 suggesting well-acclimated populations in all habitats. All the K values observed in this study were in the range suggested as indicative for fish in suitable habitat. The K value depicts robustness or suitability of a species in a habitat and primarily reflects the fish's health status, sexual maturity and the degree of feeding (Williams 2000). It is expressed as weight per unit length and a value of above 1 implies a higher weight or fat fish in relation to length (Datta et al. 2013). Therefore, K seems relevant for the fisheries managers to predict the performance of the subsector in differing environments. Our K results correspond to earlier studies: Balirwa (1998) reported a K value of 2.12 for Nile tilapia from L. Victoria, 1.95 from L. Albert and 2.15 from L. Kyoga, and Lowe-McConnell (1953), 2.1 from L. Turkana and 2.04 from L. George. However, the exact implications of such K values in this study are not entirely clear.

Condition factor K has been associated to other factors, including fish abundance and competition in ecosystems (Balirwa 1998; Bwanika et al. 2004; Otieno et al. 2013). For example, Balirwa (1998) reported that with declining fish populations in Lakes Kyoga and Victoria, less competition for food resources could have led to improvement of K values. The remarkable low K values observed under aquaculture might be linked to stress resulting from generally poor welfare of the fish under such conditions, like water quality, limited and low quality food, stock size and competition, among others. Stewart (1988) observed stress as a major factor underlying low reproduction in Nile tilapia nursery grounds around L. Turkana and dramatically negatively impacting on K. The relatively low K values from the artificial systems was not a

surprise, as it could be a symptom of poor management. It should nevertheless be noted that, albeit K is linked to nutritional factors/welfare, the proportionality in fish growth matters a lot when considering K. For instance, if a fish's weight increases proportionately with length, the fish becomes bigger but slender regardless of nutritional values from the environment. In this scenario, low K value might be recorded but could be misleading. The relatively low K values for L. Albert and Turkana populations together with Kakyunga subpopulation from L. Victoria observed with their realistic weights, could demonstrate this. More importantly, a fish might not actually grow to the desired table size despite improved K, i.e. it remains small relative to its length. This was found in Nile tilapia from the high-altitude populations (see Table 1, for sizes), where all the fish were typically small but had better K values than Albert and Turkana where comparably bigger fish weights dominated (Table 1). Thus, this suggests that K might be related to the actual physiology and genotype of the fish (Le Cren 1951). Therefore, K could be more meaningful and reliable under experimental conditions, but sampling artefacts may influence it.

Quite often, it is assumed that eco-geographical variation in a species morphology is related to local adaptation. In the current study, the most striking morphological pattern regarding shape variability was found in seven populations: Victoria, Kyoga, Turkana, Albert, George and River Nile plus Sindi farm. The remaining six populations did not reflect this pattern. Interestingly, PCA results were generally consistent with K variables, except Sindi farm. Because K is considered as one of the test parameters for environmental conditions for organisms, most likely the observed Nile tilapia shape variations could have played a bigger role. This therefore suggests that the observed K variables might have been influenced by morphological diversification in the species, a remarkable feature for adaptation especially in new habitats. The noticeable shape diversification among populations was to some extent unexpected, apart from lakes Victoria and Kyoga. Environmental forces can fuel shape changes in organisms especially when colonizing new niches (Seebacher et al. 2016). This might be the reason for the morphological diversity especially in Lake Victoria and Kyoga where the introduction of Nile tilapia coincides with the disappearance of the indigenous Nile tilapia congeneric species. Following Nile tilapia introduction into these ecosystems, the native species (*Oreochromis variabilis* and

Oreochromis esculentus), are reported to have vanished or disappeared with time. Several theories have been proposed for this, including competition for space or food and overfishing (Ogutu-Ohwayo 1990; Balirwa 1992; Lowe-McConnell 2010).

However, other reports have pinned possible intimate biotic interactions to have occurred. For example, Balirwa (1992) reported some form of hybridization in the Northern waters of *L. Victoria*, after observing phenotypic traits in some Tilapiine species identical to Nile tilapia. Relevant studies regarding cichlid gill morphological variation in *L. Victoria* have also been allied to hybridization (Witte et al. 2007). In addition, Tilapiine species (*Oreochromis esculentus* and *Oreochromis niloticus*) have also been reported to hybridize under experimental conditions (Welcomme 1966). Other GM studies regarding African cichlids (haplochromines) in lake Tanganyika reported pronounced shape variations in the F1 hybrids and concluded that hybridization among cichlids contribute significantly to morphological divergent (Stelkens et al. 2009; Kerschbaumer and Sturnbauer 2011). The outstanding morph divergent exhibited by Victoria and Kyoga populations might therefore be an expression of introgression between Nile tilapia and its native relatives which might be one possible explanation why the native Tilapiines vanished. Therefore, the morphology divergent observed could be the impacts of transgressive segregation effected during acclimatization, following introductions in the foreign homes, but this remains to be tested in the future studies.

Shape variation effects elaborated above, might as well be applicable to the River Nile morphotypes. The River Nile population was taken from the upper Murchison falls, between Victoria and Kyoga (Fig. 1). Nile tilapia from River Nile might be native below the herein mentioned barrier (Ogutu-Ohwayo 1990). However, it remains unknown if River Nile, upper falls, was stocked like in Lakes Victoria and Kyoga. However, due to drainage of water from Victoria to Kyoga via River Nile, there is undoubtedly a high probability for the species to have colonized this part of the river secondarily. The pronounced morphs observed in river might therefore result from similar forces like the introduced populations in other lakes. Also, the dynamic conditions in a river system might require an organism to develop anatomical features that can enhance its motion and endurance, particularly morphological traits like streamlined body shapes, with anterior body depths, that are

developed for steady swimming (Langerhans and Reznick 2010). The well-developed streamlined morphotypes although could aid maneuverability in high water currents, may as well serve the best purpose for food and or prey search in such habitats. Streamlined body form of morphology is also critical for maximizing thrust and minimizing drag forces while swimming (Langerhans and Reznick 2010). This is consistent with our current findings as illustrated by the wireframe graphs. In our results, the fish streamline was clearly indicated by shape changes mainly at anterior body depth and tapering at the caudal peduncle. Also, shape changes were manifested at the operculum perhaps for maximizing respiration during locomotion in water turbulences. These morph features might have also developed in response to evading predators particularly Nile perch in lakes Albert, Victoria, Kyoga and Turkana. It must be noted that, Nile tilapia in EA is reportedly the only Tilapiine species to coexist with Nile perch, from both native and introduced locations (Balirwa 1998). The long history of Nile tilapia to live and share ecosystems with the predator is linked to ecological separation (Balirwa 1998; Mwanja 2000). While Nile perch prefers the open waters, it is believed that Nile tilapia shifted their niche to shore waters to avoid predation pressure (Ogutu-Ohwayo 1990). Therefore, it is feasible that structural complexity in the shore areas with their littoral aquatic vegetation might have further exacerbated the observed morphological diversification.

The observed similarity of shapes based on NMDS and CVA between the populations of Kyoga-Victoria-River Nile and Albert suggested a possible common genetic link which might result from a common source of the translocations. Nile tilapia was introduced into Lakes Victoria and Kyoga from their putative native source, Albert (Mwanja 2000). However, other sources reported that the species might have also been introduced into lakes Victoria and Kyoga from Lakes Edward, Turkana and from pond aquaculture at Kajjansi (Uganda) (Welcomme 1966; Trewavas 1983; Balirwa 1992; Pullin et al. 1997). The multiple stocking of the cichlid in the region is likely complex which makes its in-depth examining a pressing concern. The most remarkable closer morphotypes between Kyoga and Albert as illustrated by NMDS might explain historical stocking interval variations. However, it remains unclear how the stockings in the lakes especially Victoria and Kyoga were effected, whether it was the same or varying periods. These results reflect those of Mwanja (2000)

who found Kyoga populations less genetically variable and associated it to more recent introductions in terms of history, contrary to Victoria metapopulation. Therefore, our NMDS and CVA results might be revealing the existing genetic intimacy between the introduced populations with their putative ancestral sources.

Similarity between populations of Kazinga Channel and Edward is likely to stem from their shared water channel. Fish from Kazinga Channel might be exchanged with L. Edward as the latter receives Kazinga Channel waters. Apart from L. George, the relatively minimal shape changes realized from the western rift valley populations, Edward and Kazinga Channel, suggests that these populations might be independent from anthropogenic influxes and/or disturbances. The nonnative high-altitude populations, Kayumbu and Mulche, showed similarly low morphological variability. Given that Nile tilapia is relatively a warm water species, one would undoubtedly expect the hostile environment, characterized by low water temperatures (Tibihika et al. 2016) to have significant role on the fish body form. This suggests that the populations might be of recent introduction and hence less variable morphotypes. However, the low variability in the marginal habitat might also result from lack of a native population that could induce variation via introgression (Tibihika et al. 2015), or by the higher selection pressure that initially might have reduced the introduced variation.

The noticeable shape changes in George population that differed from Kazinga Channel and Edward (closest neighbors), might be attributed to a bottleneck effect from overfishing and water contamination. Reduced fish numbers or demographic bottlenecked populations from overfishing with time are likely to be exposed to inbreeding and eventually lose genetic variation (Xenikoudakis et al. 2015). In addition to the eventualities of unfit generations, the inbred populations might further experience genetic drift which might influence the phenotypic traits and hence shape variations. Shao et al. (2007) reported significant morphological variation in inbred stocks of Chinese Minnow, *Gobio cypris rarus*. It should be noted that, although overfishing crosscuts in EA water bodies, significant fish overexploitation and accumulation of heavy metals (copper, zinc, cobalt and nickel) have been reported in L. George (Kamanyi 1996; Lowe-McConnell 2010) and (Swaibuh Lwanga et al. 2003) respectively. Heavy metals for example zinc and cadmium have been reported to have a pronounced effect on the morphology of fish like in

the Indian major carps (Rani and Gupta 2015). Therefore, it is most likely that, the observed morphs from L. George might be under the multiple influence of both overfishing and water toxicity. However, of a more important note, the shape divergent encountered in George might also be a result of admixture. Mwanja (2000) reported that, in the 1970s, while Nile tilapia fishery was booming on L. George, it prompted a fish processing plant on the lake. Later, overfishing (Lowe-McConnell 2010) triggered importation of Nile tilapia from L. Albert to reinforce the dwindled stocks in L. George which culminated into Albert strain planted into L. George. Perhaps the interaction between Albert and George strains might further be behind the unexpected and encountered morphological divergence in the Lake, albeit subject to future studies.

Our results concur with those of Mwanja (2000) who found Nile tilapia populations from Kyoga closer to Albert and those of George distant from Edward. On the other hand, we differ as Victoria population was observed to be distant morphologically from Edward. This could be reasonably related to the differential methodological approaches in the two studies.

Conclusions and recommendations

Our results demonstrated that all the populations exhibited adequate ranges for condition factor, suggesting that they are “healthy” populations. Nevertheless, we found K not recommendable for reliance when depicting fisheries status. Morphological divergence was manifested mainly in River Nile, Lakes Victoria, Kyoga, Albert, George and Turkana plus one fish farm. The pronounced shape divergence observed between these populations were suspected to population admixture and other environmental cues. The noted morph closeness in some populations might be an indicator of possible genetic nearness which could be explained by the common ancestral home. The findings from this study are vital for management and conservation of the species. It is most likely that Nile tilapia populations from Kazinga Channel and L. Edward are devoid of anthropogenic activities and thus probably the only existing pure lines of the species. These populations may also be recommended for selective breeding in efforts aimed at elevating aquaculture productivity in the region. This study however, provides a benchmark in the dynamics of the species and hence, we recommend a more explicit

consideration of follow-up of this work with molecular techniques while addressing genetic differentiation of these populations. Use of DNA bar-coding or Restriction site associated DNA (RAD) sequencing might allow detection of Nile tilapia diversification in the context of anthropogenic impacts. The future studies should broaden the scope of investigation for instance to include the populations below Murchison Falls, along River Nile and examine in-depth within population variations. Additionally, the same approach should also be true for all the great lakes, including other water bodies like Kivu, Bunyonyi and Baringo as the species variability might be at large.

Acknowledgements Open access funding provided by University of Natural Resources and Life Sciences Vienna (BOKU). The work was financed by the Austrian Partnership Programme in Higher Education and Research for Development - APPEAR, a programme of the Austrian Development Cooperation (ADC) and implemented by the Austrian Agency for International Cooperation in Education and Research (OeAD). Special thanks to Kachwekano Zonal Agricultural Research and Development Institute-National Agricultural Research Organization-Uganda and Makerere University-Kampala, Uganda for their great assistance during field activities. We are also grateful to diligent efforts provided by the Fisheries Officers from East African respective water bodies, more especially, a senior research officer, Mr. John Malala of L. Turkana (KEMFRI) for his memorable support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were observed and henceforth followed.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agnèse JF, Adépo-Gourène B, Abban EK, Fermon Y (1997) Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity* 79(1):88–96. <https://doi.org/10.1038/hdy.1997.126>
- Bakhoum SA, Sayed-Ahmed MA, Ragheb EA (2009) Genetic evidence for natural hybridization between Nile tilapia (*Oreochromis niloticus*; Linnaeus, 1758) and blue tilapia (*Oreochromis aureus*; Steindachner, 1864) in lake Edku, Egypt. *Glob Vet* 3(2):91–97
- Balirwa JS (1992) The evolution of the fishery of *Oreochromis niloticus* (Pisces: Cichlidae) in Lake Victoria. *Hydrobiologia* 232(1):85–89. <https://doi.org/10.1007/BF00014616>
- Balirwa JS (1994) The biology, ecology, population parameters and the fishery of the Nile tilapia, *Oreochromis niloticus* (L). 26–38. <https://doi.org/10.1007/BF00014616>
- Balirwa JS (1998) Lake Victoria wetlands and the ecology of the Nile tilapia, *Oreochromis niloticus* linn. Dissertation submitted in fulfillment of the requirements of the Board of Deans of Wageningen agricultural university and the academic Board of the International Institute for infrastructural, Hydraulic and Environmental Engineering for the Degree of DOCTOR. Balkema
- Bookstein FL (1982) Foundations of morphometrics. *Annu Rev Ecol Syst* 13(1):451–470. <https://doi.org/10.1146/annurev.es.13.110182.002315>
- Boyd CE (2004) Farm-level issues in aquaculture certification: Tilapia. Auburn, Alabama 36831. Report commissioned by WWF-US in 2004. Auburn University, Auburn
- Breno M, Leirs H, Van Dongen S (2011) Traditional and geometric morphometrics for studying skull morphology during growth in *Mastomys natalensis* (Rodentia: Muridae). *J Mammal* 92(6):1395–1406. <https://doi.org/10.1644/10-MAMM-A-331.1>
- Bwanika G, Makanga B, Kizito Y, Chapman L, Balirwa J (2004) Observations on the biology of Nile tilapia, *Oreochromis niloticus* L., in two Ugandan crater lakes. *Afr J Ecol* 42(s1): 93–101. <https://doi.org/10.1111/j.1600-0633.2006.00185.x>
- Byrnes R (2010) Uganda: A country study. GPO for the Library of Congress, Washington
- Cockerton H, Street-Perrott F, Leng M, Barker P, Horstwood M, Pashley V (2013) Stable-isotope (H, O, and Si) evidence for seasonal variations in hydrology and Si cycling from modern waters in the Nile Basin: implications for interpreting the quaternary record. *Quat Sci Rev* 66:4–21. <https://doi.org/10.1016/j.quascirev.2012.12.005>
- Corse E, Neve G, Sinama M, Pech N, Costedoat C, Chappaz R, Gilles A (2012) Plasticity of ontogenetic trajectories in cypriids: a source of evolutionary novelties. *Biol J Linn Soc* 106(2):342–355. <https://doi.org/10.1111/j.1095-8312.2012.01873.x>
- Costa C, Vandeputte M, Antonucci F, Buglione C, Menesatti P, Cenadelli S, Parati K, Chavanne H, Chatain B (2010) Genetic and environmental influences on shape variation in the European sea bass (*Dicentrarchus Labrax*). *Biol J Linn Soc* 101(2):427–436. <https://doi.org/10.1111/j.1095-8312.2010.01512.x>
- Datta SN, Kaur VI, Dhawan A, Jassal G (2013) Estimation of length-weight relationship and condition factor of spotted snakehead *Channa punctata* (Bloch) under different feeding regimes. *SpringerPlus* 2(1):436. <https://doi.org/10.1186/2193-1801-2-436>
- Deines AM, Bhole I, Katongo C, Feder JL, Lodge DM (2014) Hybridisation between native *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River.

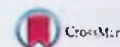
- Zambia. *Afr J Aquat Sci* 39(1):23–34. <https://doi.org/10.2989/16085914.2013.864965>
- Dodson JJ, Bourret A, Barrette MF, Turgeon J, Daigle G, Legault M, Lecomte F (2015) Intraspecific genetic admixture and the morphological diversification of an estuarine fish population complex. *PLoS One* 10(4):e0123172. <https://doi.org/10.1371/journal.pone.0123172>
- Eklöv P, Svanbäck R (2005) Predation risk influences adaptive morphological variation in fish populations. *Am Nat* 167(3): 440–452. <https://doi.org/10.1086/439544>
- Eknath AE, Bentsen HB, Ponzone RW, Rye M, Nguyen NH, Thodesen J, Gjerde B (2007) Genetic improvement of farmed tilapia: composition and genetic parameters of a synthetic base population of *Oreochromis niloticus* for selective breeding. *Aquaculture* 273(1):1–14. <https://doi.org/10.1016/j.aquaculture.2007.09.015>
- El-Zaeem SY, Ahmed MMM, Salama ME-S, El-Kader WNA (2012) Phylogenetic differentiation of wild and cultured Nile tilapia (*Oreochromis niloticus*) populations based on phenotype and genotype analysis. *Afr J Agric Res* 7(19): 2946–2954. <https://doi.org/10.5897/AJAR112170>
- Frøese R (2006) Cube law, condition factor and weight-length relationships: History, meta-analysis and recommendations. *J Appl Ichthyol* 22(4):241–253. <https://doi.org/10.1111/j.1439-0426.2006.00805.x>
- George TT (1995) The most recent nomenclature of tilapia species in Canada and the Sudan. *Bull Aquac Assoc Can. Sp Pub* 10: 33–37
- Green J (2009) Nilotic lakes of the western rift. *The Nile*:263–286. https://doi.org/10.1007/978-1-4020-9726-3_14
- Hammer O, Harper D, Ryan P (2001) PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electron* 4:1–9
- Holland SM (2008) Non-metric multidimensional scaling (MDS). Available at: <http://www.uga.edu/strata/software/pdts/mDSTutorial.pdf>
- ISSIN EAIC (1960) East African Fisheries Research Organization Annual Report 1954/1955. East Africa High Commission, Tanganyika
- Januszkiewicz AJ, Robinson BW (2007) Divergent walleye (*Sander vitreus*)-mediated inducible defenses in the centrarchid pumpkinseed sunfish (*Lepomis gibbosus*). *Biol J Linn Soc* 90(1):25–36. <https://doi.org/10.1111/j.1095-8312.2007.00708.x>
- Jolliffe I (2002) Principal Component Analysis and Factor Analysis. In: *Principal Component Analysis*. Springer Series in Statistics. Springer, New York
- Kamanyi J (1996) Management strategies for exploitation of Uganda Fisheries Resources. FIRI Mimeo, February, 1996. Fisheries Research Institute, Uganda
- Kerschbaumer M, Sturmbauer C (2011) The utility of geometric morphometrics to elucidate pathways of cichlid fish evolution. *Int J Evol Biol* 2011:1–8. <https://doi.org/10.4061/2011/290245>
- Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Mol Ecol Resour* 11(2):353–357. <https://doi.org/10.1111/j.1755-0998.2010.02924.x>
- Kolding J (1992) Population dynamics and life-history styles of Nile tilapia, *Oreochromis niloticus*, in Ferguson's Gulf, Lake Turkana, Kenya. *Environ Biol Fish* 37:25–46. <https://doi.org/10.1007/BF00000710>
- Langerhans RB, Reznick DN (2010) Ecology and evolution of swimming performance in fishes: predicting evolution with biomechanics. *Fish locomotion: an eco-ethological Perspective*:200–248
- Le Cren E (1951) The length-weight relationship and seasonal cycle in gonad weight and condition in the perch (*Perca fluviatilis*). *J Anim Ecol* 20:1–219. <https://doi.org/10.2307/1540>
- Lorenz O, Smith P, Coghill L (2014) Condition and morphometric changes in tilapia (*Oreochromis* sp.) after an eradication attempt in southern Louisiana. *NeoBiota* 20:49. <https://doi.org/10.3897/neobiota.20.506>
- Lowe-McConnell RH (1958) Observations on the biology of Tilapia Nilotica Linné in east African waters. *Rev Zool Bot Afr* 57:129–170
- Lowe-McConnell R (2003) Recent research in the African Great Lakes: fisheries, biodiversity and cichlid evolution. In: *Freshwater Forum*. pp 4–64
- Lowe-McConnell R (2010) Fisheries and cichlid evolution in the African Great Lakes: progress and problems. <https://doi.org/10.1608/RJ-2.2.2>
- Maderbacher M, Bauer C, Herler J, Postl L, Makasa L, Sturmbauer C (2008) Assessment of traditional versus geometric morphometrics for discriminating populations of the *Tropheus* Moorii species complex (Teleostei: Cichlidae), a Lake Tanganyika model for allopatric speciation. *J Zool Syst Evol Res* 46(2):153–161. <https://doi.org/10.1111/j.1439-0469.2007.00447.x>
- Mallet J (2007) Hybrid speciation. *Nature* 446(7133):279–283. <https://doi.org/10.1038/nature0570>
- Mekkawy IA, Mohammad AS (2011) Morphometrics and Meristics of the three Epinepheline species: *Cephalopholis Argus* (Bloch and Schneider, 1801), *Cephalopholis Miniata* (Forsskal, 1775) and *Variola Louti* (Forsskal, 1775) from the Red Sea, Egypt. *J Biol Sci* 1(11):10–21. <https://doi.org/10.3923/jbs.2011.10.21>
- Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin L, Nijhout HE, Abouheif E, Pliennig DW (2011) The role of developmental plasticity in evolutionary innovation. *Proc R Soc Lond B Biol Sci* 278(1719):2705–2713
- Mugidde R (1993) The increase in phytoplankton primary productivity and biomass in Lake Victoria (Uganda). *Verh Int Ver Limnol* 25:846–849
- Muir A, Vecsei P, Krueger C (2012) A perspective on perspectives: methods to reduce variation in shape analysis of digital images. *Trans Am Fish Soc* 141(4):1161–1170. <https://doi.org/10.1080/00028487.2012.685823>
- Mwanja W (2000) Genetic biodiversity and evolution of two principal fisheries species groups, the labeine and tilapine, of Lake Victoria, East Africa. Ph. D. thesis, Ohio State University, Cleveland, Ohio
- Ndiwa TC, Nyingi DW, Claude J, Agnèse J-F (2016) Morphological variations of wild populations of Nile tilapia (*Oreochromis niloticus*). *Environ Biol Fish* 99(5):473–485. <https://doi.org/10.1007/s10641-016-0492-y>
- Njiru M, Nyamweya C, Gichuki J, Mugidde R, Mumbo O, Witte F (2012) Increase in anoxia in Lake Victoria and its effects on the fishery. In: *Anoxia*. InTech
- Ogle DH (2013) FishR Vignette-Fish Condition and Relative Weights: AIFFD Chapter 10.1–12
- Ogutu-Ohwayo R (1990) The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of

- introduced species, especially the Nile perch, *Lates niloticus*, and the Nile tilapia, *Oreochromis niloticus*. Environ Biol Fish 27(2):81–96. <https://doi.org/10.1007/BF00001938>
- Ogutu-Ohwayo R, Hecky RE, Cohen AS, Kaufman L (1997) Human impacts on the African great lakes. Environ Biol Fish 50(2):117–131. <https://doi.org/10.1023/A:1007320932349>
- Ondhoro CC, Masembe C, Maes GE, Nkalubo NW, Walakira JK, Naluwaio J, Mwanja MT, Efitre J (2016) Condition factor, length–weight relationship, and the fishery of *Barbus altianalis* (Boulenger 1900) in lakes Victoria and Edward basins of Uganda. Environ Biol Fish:1–12. <https://doi.org/10.1007/s10641-016-0567-9>
- Otieno ON, Kitaka N, Njiru J (2013) Length-weight relationship, condition factor, length at first maturity and sex ratio of Nile tilapia, *Oreochromis niloticus* in Lake Naivasha, Kenya. Int J Fish Aquat Stud:67–72
- Park PJ, Aguirre WE, Spikes DA, Miyazaki JM (2013) Landmark-based geometric Morphometrics: what fish shapes can tell us about Fish evolution. Proceedings of the Association for Biology Laboratory Education 34:361–371
- Peligo VC, Jumawan JC (2016) Assessment of fluctuating asymmetry in the body shapes of Nile tilapia (*Oreochromis niloticus*) from Masao river, Butuan city, Philippines. AACL Bioflux 9(3):604–613
- Pernier C, Bagliniere JL, Evanno G (2013) Understanding admixture patterns in supplemented populations: a case study combining molecular analyses and temporally explicit simulations in Atlantic salmon. Evol Appl 6(2):218–230. <https://doi.org/10.1111/j.1752-4571.2012.00280.x>
- Pertyl M, Kalous L, Memis D (2014) Comparison of manual measurements and computer-assisted image analysis in fish morphometry. Turk J Vet Anim Sci 38(1):88–94. <https://doi.org/10.3906/vet-1209-9>
- Ponton D, Carassou L, Raillard S, Borsa P (2013) Geometric morphometrics as a tool for identifying emperor fish (Lethrinidae) larvae and juveniles. J Fish Biol 83(1):14–27. <https://doi.org/10.1111/jfb.12138>
- Popma T, Masser M (1999) Tilapia Life History and Biology. Southern Regional Aquaculture Centre. SRAC Publication No. 283, Stoneville
- Pullin R, Palomares M-L, Casal C, Dey M, Pauly D (1997) Environmental impacts of tilapias. In: tilapia aquaculture . 554–570
- Rani S, Gupta R (2015) Effect of heavy metals on the morphology and growth performance of Indian major carps. Ann Biol 31(1):117–121
- Ricker W (1975) Computation and interpretation of biological statistics of fish populations. Bull Fish Res Board Can 191: 1–382
- Robinson BW, Wilson DS (1994) Character release and displacement in fishes: a neglected literature. Am Nat 144(4):596–627. <https://doi.org/10.1086/285696>
- Rohlf FJ (2002) Geometric morphometrics and phylogeny. Morphol Shape Phylogeny:175–193. <https://doi.org/10.1201/9780203165171.ch9>
- Rohlf F (2004) Tpsdig (version 1.40). Department of Ecology and Evolution, State University of New York at Stony Brook, New York. <http://life.bio.sunysb.edu/ce/rohlf/software.html>
- Scharnweber K, Watanabe K, Syväranta J, Wanke T, Monaghan MT, Mehner T (2013) Effects of predation pressure and resource use on morphological divergence in omnivorous prey fish. BMC Evol Biol 13(1):132. <https://doi.org/10.1186/1471-2148-13-132>
- Seebacher F, Webster MM, James RS, Tallis J, Ward AJ (2016) Morphological differences between habitats are associated with physiological and behavioural trade-offs in stickleback (*Gasterosteus aculeatus*). Open Sci 3(6):160316. <https://doi.org/10.1098/rsos.160316>
- Seehausen O (2004) Hybridization and adaptive radiation. Trends Ecol Evol 19(4):198–207. <https://doi.org/10.1016/j.tree.2004.01.003>
- Shao Y, Wang J, Qiao Y, He Y, Cao W (2007) Morphological variability between wild populations and inbred stocks of a Chinese minnow, *Gobiocypris rarus*. Zool Sci 24(11):1094–1102. <https://doi.org/10.2108/zsj.24.1094>
- Sinama M, Gilles A, Costedoat C, Corse E, Olivier J-M, Chappaz R, Pech N (2013) Non-homogeneous combination of two porous genomes induces complex body shape trajectories in cyprinid hybrids. Front Zool 10(1):22. <https://doi.org/10.1186/1742-9994-10-22>
- Stelkens RB, Schmid C, Selz O, Seehausen O (2009) Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. BMC Evol Biol 9(1):283. <https://doi.org/10.1186/1471-2148-9-283>
- Stewart K (1988) Changes in condition and maturation of the *Oreochromis niloticus* L. population of Ferguson's gulf, Lake Turkana, Kenya. J Fish Biol 33(2):181–188. <https://doi.org/10.1111/j.1095-8649.1988.tb05461.x>
- Swaibuh Lwanga M, Kansime F, Denny P, Scullion J (2003) Heavy metals in Lake George, Uganda, with relation to metal concentrations in tissues of common fish species. Hydrobiologia 499(1):83–93. <https://doi.org/10.1023/A:1026347703129>
- Tibihika PDM, Barekye A, Bykora E (2015) Fish species composition, abundance and diversity of Minor Lakes in south western Uganda/Kigezi region. Int J Sci Technol 4(5):204–213
- Tibihika PDM, Okello W, Barekye A, Mbabazi D, Omony J, Kiggundu V (2016) Status of Kigezi minor lakes: a limnological survey in the lakes of Kisoro, Kabale and Rukungiri districts. Int J Water Resour and Environ Eng 8(5):60–73
- Trewavas E (1983) Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*. British Museum (Natural History), London. <https://doi.org/10.5962/bhl.title.123198>
- Verschuren D, Johnson TC, Kling HJ, Edgington DN, Leavitt PR, Brown ET, Talbot MR, Hecky RE (2002) History and timing of human impact on Lake Victoria, East Africa. Proc R Soc Lond B Biol Sci 269(1488):289–294. <https://doi.org/10.1098/rspb.2001.1850>
- Vreven E, Teugels G, Adepo-Gourene B, Agnese J (1998) Morphometric and allozyme variation in natural populations and cultured strains of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). Belg J Zool 128(1):23–34
- Welcomme R (1966) Recent changes in the stocks of tilapia in Lake Victoria. Nature 212(5057):52–54. <https://doi.org/10.1038/212052a0>
- Williams J (2000) The coefficient of condition of fish. Chapter 13 in Schneider, James C.(ed.) 2000. Manual of fisheries survey methods II: with periodic updates. Michigan Department of Natural Resources Fisheries Special Report 25
- Wilson LA, MacLeod N, Humphrey LT (2008) Morphometric criteria for sexing juvenile human skeletons using the ilium.

- J Forensic Sci 53(2):269–278. <https://doi.org/10.1111/j.1556-4029.2008.00656.x>
- Witte F, Wamink J, Kashe-Machumu M (2007) Species distinction and the biodiversity crisis in Lake Victoria. *Trans Am Fish Soc* 136(4):1146–1159. <https://doi.org/10.1577/T05-179.1>
- Nemikoudakis G, Ernmark E, Tison JL, Waits L, Kindberg J, Swenson J, Dalen L (2015) Consequences of a demographic bottleneck on genetic structure and variation in the Scandinavian brown bear. *Mol Ecol* 24(13):3441–3454. <https://doi.org/10.1111/mec.13239>
- Zelditch ML, Swiderski DL, Sheets HD (2012) A Practical Companion to Geometric Morphometrics for Biologists: Running analyses in freely-available software. Free download from <https://booksite.elsevier.com/9780123869030/> (18 June 2013)

3. Chapter 3

2nd paper. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*



Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*

Papius Dias Tibihika^{1,3} · Manuel Curto¹ · Eva Domstauder-Schrammel¹ · Silvia Winter^{1,5} · Esayas Alemayehu^{2,4} · Herwig Waldbacher² · Harald Meimberg¹

Received: 19 July 2018 / Accepted: 10 December 2018
 © The Author(s) 2018

Abstract

Microsatellites play an important role when investigating population and ecological genetics, although high effort in development and genotyping constitute a technical constraint and remains a major bottleneck. Here we use a microsatellite genotyping approach utilizing sequences of amplicons for allele calling (SSR-GBS) based on Illumina that requires less effort and time. The approach consist of development of highly polymorphic loci, sequencing of multiplexed PCR amplified microsatellites on an Illumina Miseq PE 300 platform and bioinformatic treatment of the sequenced data using custom scripts. The procedure allows automation in allele calling, which can be more reliably replicated and thereby removes biases that might prevent concatenation of datasets from different analyses. Additionally, the methodology enhances information content in the sequenced data beyond the traditional amplicon length (AL) approaches. Using 26 newly developed microsatellite markers and SSR-GBS we investigate the population genetic assessment of anthropogenically altered populations of East African Nile tilapia to show the potential of this genotyping approach. More precisely, we compare genotypic data generated considering AL and whole amplicon information (WAI). We found that genotypes based on WAI are not only able to recover a higher number of alleles but also a more detailed genetic structure pattern. We discuss the capability and importance of WAI allele calling and show perspectives for implementation in the future conservation genetic studies. More specifically, we demonstrate how the current markers and techniques might contribute useful information for studies concerning resources sustainable exploitation and conservation using the East African Nile tilapia.

Keywords Nile tilapia · Microsatellite markers · Amplicon sequencing · Amplicon length · Genotyping pipeline · Next generation sequencing

Introduction

Microsatellites/simple sequence repeats (SSR) or short tandem repeats (STRs) have proved reliable as ideal genetic markers in population genetic studies (Guichoux et al. 2011; Selkoe and Toonen 2006; Schlotterer et al. 1991). This is due to their important attributes including, but not limited

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10592-018-1136-x>) contains supplementary material, which is available to authorized users.

✉ Papius Dias Tibihika
 papiust@yahoo.com; papius.tibihika@students.boku.ac.at

¹ Institute for Integrative Nature Conservation Research, University of Natural Resources and Life Sciences, Vienna (BOKU), Gregor Mendel-Strasse 33, 1180 Vienna, Austria

² Institute for Hydrobiology and Aquatic Ecosystems Management, University of Natural Resources and Life Sciences, Vienna (BOKU), Gregor-Mendel-Strasse 33/DG, 1180 Vienna, Austria

³ Kachwekano Zonal Agricultural Research and Development Institute, National Agricultural Research Organization, P. O. Box 421, Kabale, Uganda

⁴ National Fishery and Aquatic Life Research Centre, P.O. Box 64, Sebeta, Ethiopia

⁵ Division of Plant Protection, University of Natural Resources and Life Sciences, Vienna (BOKU), Gregor Mendel-Strasse 33, 1180 Vienna, Austria

Published online: 17 December 2018

Springer

to, their ubiquity in the genome, co-dominant inheritance, high polymorphism and reproducibility (Sundaray et al. 2016; Selkoe and Toonen 2006). Although there is a growing set of other markers and sequencing techniques, SSRs are still one of the leading genotyping options based on their methodological simplicity. Additionally, the use of these markers is further rendered reliable following the genesis of next-generation sequencing (NGS) technologies and the identification of numerous loci in non-model species accompanied by reduced costs (Guichoux et al. 2011; Curto et al. 2013). Nevertheless, many of the previous genetic studies on Nile Tilapia, have utilised finger print methods (Mwanja et al. 1996, 2008; Agnèse et al. 1997; Seyoum and Kornfield 1992). These methods have limited information content (e.g. Liu and Cordes 2004; Miah et al. 2013; Okumuş and Çiftçi 2003; Guichoux et al. 2011). Traditionally developed SSR had been employed, but frequently with low number of loci investigated and the resulting low amount of statistical power, (e.g. Bezaul et al. 2012; Lee and Kocher 1996; Hassanien and Gilbey 2005; Tariq Ezaz et al. 2004).

Normally, SSRs are genotyped by using the fragment length variation, through capillary electrophoresis. In most cases, due to the high mutation rates of the repetition motifs, this is sufficient to generate adequate amounts of variation at the intraspecific level. However, considering the amplicon length (AL) variation alone might miss out other useful polymorphic information. By genotyping microsatellite markers via high throughput sequencing methods such as Illumina, it is possible to compare AL and whole amplicon information (WAI) allele calling procedures, and ascertain variations in the amount of genetic information generated (De Barba et al. 2017). Other advantages of using high throughput sequencing for microsatellite genotyping include, the automation and replicability with low cost of the genotyping process (De Barba et al. 2017; Vartia et al. 2016; Zhan et al. 2017). Using available bioinformatic tools, one can either automatize or at least significantly decrease the level of artefacts in allele calling and thereby elevating information content of the marker. This facilitates the reproducibility of the genotyping process, in comparison to the traditional electropherogram length frequency (AL) procedures (Farrell et al. 2016; De Barba et al. 2017). In the current study, we expect this method (WAI allele calling), which we refer to as SSR-GBS (Simple Sequence Repeats-Genotyping By-Sequencing), to become more informative and subsequently widely adopted as an important tool for genotyping. We test this assumption using the East Africa Nile tilapia as a case study based on their historical dynamics (Balirwa 1992; Kaufman 1992; Ogutu-Ohwayo 1990).

The East African Nile tilapia (*Oreochromis niloticus*, L. 1758) is an important species for augmenting aquaculture and capture fisheries. The cichlid is native to lakes Albert, George, Edward, Turkana, and Baringo among others, as

well as the lower Nile River (Ogutu-Ohwayo 1990). However, human activities, might have had an effect on the native populations through the introduction of potentially maladaptive traits through translocations and promoting bottlenecks by overfishing. Additionally, the activities like transplantations of Nile tilapia into numerous non-native habitats including lakes Victoria, Kyoga, Nabugabo, many satellite water bodies and other systems, like aquacultural set-ups (Welcomme 1966; Balirwa 1992) might have altered the species' genetic structure through admixture promoted by hybridization with closely related indigenous species. Broadly, the contrast between native and non-native Nile tilapia stocks in the context of anthropogenic influence remains under-documented (Mwanja 2000). Although Nile tilapia fishery is vital for the livelihoods in East Africa, management of this important resource against detrimental anthropogenic activities is unfortunately inadequate due to insufficient research (Njiru et al. 2007; Ogutu-Ohwayo 1990). Therefore, the use of more informative microsatellite datasets, might contribute to efficiently characterise the populations (Hedrick 2001; Oliveira et al. 2006).

Here, we investigate the potential of the SSR-GBS approach by describing the genetic variability and structure of East African Nile tilapia. This study is a proof of concept testing the ability of SSR-GBS in producing genotype data with a higher statistical power than the traditional SSR amplicon length genotyping due to its potential in recovering SNPs associated with SSR motif variation. This provides a higher number of alleles and reduces the length homoplasy effect that is characteristic to this marker system. We demonstrate this by comparing the importance and strengths of the two allele calling methods (WAI and the AL, similar to the traditional electropherogram methods). As a side product of this work, we develop a set of SSR markers for Nile tilapia and test their applicability on other tilapiine species.

Materials and methods

Sample collection

The East Africa (Uganda) Nile tilapia samples were collected monthly overnight between July and December 2016, as part of commercial catches of local fishermen using the East Africa member states recommended 127 mm of gill nets. The sampling sites were defined upon advice from law enforcement fisheries officers, together with local fishermen and sampling sites recorded using a Global Positioning System (GPS). On landing, muscle tissue samples were taken immediately, preserved in absolute ethanol and later stored under -20°C , except during transportation. The species are not protected, but rather fished commercially, and sampling was performed by members of a government

institution (National Agricultural Research Organization of Uganda), hence no other permission was required. Apart from one non-native population (Lake Victoria), the other three populations (Lakes Albert and George and River Nile) are native. In total, we collected 107 Ugandan individuals (Table 1). All animal rights were observed during the field excursions. Fish was obtained as part of commercial fishing operations by local fishermen and killed in this process.

For SSR development, we used a single Ethiopian Nile tilapia sample (Lake Ziway), which was accessed through a selective breeding project at the Institute for Integrative Nature Conservation Research-BOKU, Vienna. From this sample, fresh fin clips were taken, implying that there was no fish sacrifice. The Ethiopian sample was later used for low coverage total DNA shot gun sequencing for SSR marker discovery prior to East African sampling. We also sampled Ethiopian *Tilapia zillii* for cross-species amplification tests.

DNA extraction

A piece of ethanol-preserved tissue muscle (approximately 0.1 g) was digested overnight in 500 µl lysis buffer (2% SDS, 2% PVP 40, 250 mM NaCl, 200 mM Tris-HCl, 5 mM EDTA, pH8) containing 200 ng of proteinase K enzyme. Genomic DNA extraction was carried out using magnetic beads (MagSi-DNA beadsMagnaMedics) and a magnetic separator SL-MagSep96 (Steinbrenner, Germany) following a modified MagSi-DNA Vegetal kit protocol. DNA binding was carried out by mixing 17 µl of beads with 500 µl each of binding buffer and clear lysate in a 2 ml 96 well plate. Bound DNA beads were then washed twice in 80% 600 µl cold ethanol. Later, DNA was eluted twice by adding 50 µl (first elution) and 70 µl (second elution) of elution buffer (65 °C 10 mM Tris-HCl, pH 8.3), following the above-mentioned kit protocol. The quality of the extracted DNA was inspected on 0.8% agarose gel.

SSR discovery

High quality extracted genomic DNA from the Ethiopian sample was sent for library preparation on the Illumina MiSeq paired-end (PE) 300 sequencing platform (San Diego, USA) as described in Shendure and Ji (2008); Cas- toe et al. (2012). Both library preparation and sequencing

were done at the Genomics Service Unit, Ludwig-Maximilian's-Universität München, Germany. Sequences generated by Illumina Miseq were quality checked using FASTQC (Andrews and FastQC 2010) and trimmed for the removal of adapter sequences and low quality regions (Phred <20) using CUTADAPT vers. 0.11.1 (Martin 2011). Forward and reverse reads were merged using PEAR version 0.9.4 (Zhang et al. 2014) considering only minimum overlaps of 15 bp with a p-value below 0.01 for the highest observed expected alignment scores. Later, the sequences were screened for microsatellite motifs (from di to penta nucleotide repeats) using the SSR_pipeline program (Miller et al. 2013). Here, we considered sequences with at least 10 repeats for 2mers, eight for 3mers, and six for 4/5mers. A total of 6,724 SSR motif reads were revealed comprising 4,629 2mers, 818 3mers, 868 4mers and 409 5mers. For subsequent primer design, sequences of equal or greater than 350 bp long and microsatellites with flanking regions longer than 30 bp were considered. This length was chosen to facilitate detection and elimination of primer dimer and other low molecular weight artefacts from the specific amplification products using the washing method described below. Considering the inclusion of 60 bp oligonucleotides in the PCR as a multiplex assay, such artefacts are difficult to suppress. Artefacts can be unequivocally detected using gel electrophoresis and discarded using magnetic beads. In addition, longer sequence information is higher chances of recovering extra information on the flanking sites that may contribute to the increase on the number of alleles. Raw reads were submitted to Sequence Read Archive (SRA) database with the reference number SRX3398501.

Primer design for amplicon sequencing

Specific PCR primers were designed in Geneious software version 10.3 (Kearse et al. 2012) using the default Primer3 program (Untergasser et al. 2012). Manual primer3 adjustments were set at 55 °C for optimal primer melting temperature, with a GC content in the range of 20–50–80, optimal oligo length between 18–20–23 bp, and amplification product size between 350 and 450 bp. We designed primers in a way that the complete primer motif would be included in the first or last 300 bp of the amplicon being able to be covered completely by one of MiSeq's paired reads. This prevents

Table 1 Nile tilapia sample sources from East African freshwater bodies in Uganda

Population	Location	Habitat type	No.	Latitude	Longitude	Elevation
Albert	Ntoroko	Native	23	01.05206N	030.53464E	618
George	Hamukungu	Native	35	00.01739S	030.08698E	916
River Nile	Kibuye	Native	24	01.18734N	032.96865E	1062
Victoria	Kamuwunga	Non-native	25	00.12747N	031.93999E	1139

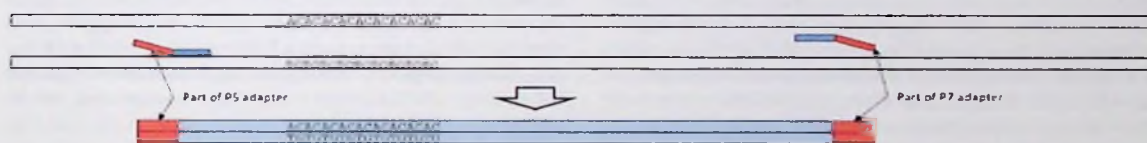
Elevation = meters above the sea level and No. = number of samples

that the repetitive unit is part of the overlap of the paired reads which would lead to difficulties with the merging step in the bioinformatics pipeline. Here, 48 primer pairs were designed. We decided to develop primers from own shot gun sequences instead of published Nile tilapia genomic information to increase the likelihood that markers fit on East African populations. However, the availability of a Nile tilapia genome from the GenBank accession number MKQE01000000, allowed us to screen the entire sequences containing microsatellites used for primer design for potentially duplicated markers, using BLASTn (Basic Local Alignment Search Tool for nucleotides). BLASTn outputs provide a list of pairwise alignment matches and sequence hits above which a statistical threshold is displayed (Xiong 2006). In the current study, BLASTn-aligned sequences were selected if the E-values were zero. The E-value is comparable to the probability value p , as the least value suggests a lower likelihood that the database matches are a result of random chance, but rather the database matches display a significant similarity (Xiong 2006). Only primers that originated from sequences showing single matches on the genome were considered because they were more likely to represent single copy regions. Of the 48 primer sequences initially mapped on the genome, 13 primers were found more than once in the genome and subsequently discarded, thus leaving 35 primer pairs. Although higher numbers of

microsatellite markers can provide robust population genetics results (Capote et al. 2012; Ryman et al. 2006), generally a number of microsatellite loci in the range of 8–20 are considered adequately informative to determine genetic structure between populations (Arthofer et al. 2018; Vartia et al. 2014; Koskinen et al. 2004). In the current study, the initial 48 primer pairs resulted in a number sufficient to test genetic structure patterns in east African Nile Tilapia. Nevertheless, more markers can be easily added with the procedure and resources presented here.

For Illumina sequencing, primers were extended by part of the Illumina adapters P5 (TCCTTCCCTACACGACGC TCTTCCGATCT) and P7 (CTGGAGTTCAGACGTGTG CTCTTCCGATCT) at the 5' end of the primer forward and reverse, respectively (Fig. 1). These correspond to the Illumina sequencing primers and served as a linker for the second PCR where the remaining parts of the adapters are added. This procedure was conducted using primers containing all the components necessary for Illumina sequencing. In this second (index) PCR, for each sample, we used a novel combination of two different indexes of 8 bp. P5-(AAT GATACGGCGACCACCGAGATCTACAC[Index]ACACTC TTTCCCTACACGACG) and P7-(CAAGCAGAAGACGGC ATACGAGAT[Index]GTGACTGGAGTTCAGACGTGT). This was vital for allowing the pooling of a large sample size in the down-stream analysis. After the second PCR,

1st PCR



2nd PCR

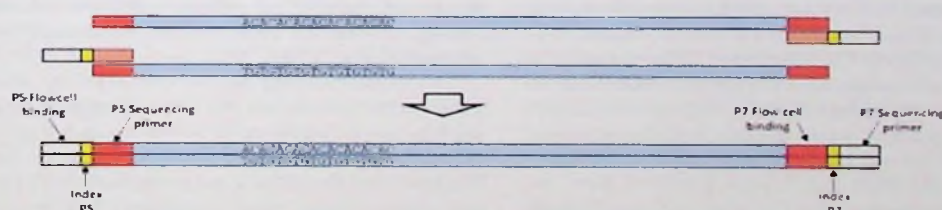


Fig. 1 Amplicon library preparation approach. We used two-step PCR procedures: In the first PCR, the microsatellite was amplified with locus specific primers (blue) and extended with part of the Illumina adapter (orange). This motif serves as linker for the second PCR where the remaining parts of the Illumina adapter were included. The

primers in the second PCR from 5' to 3' are constituted by: a region complementary to the linker corresponding to the Illumina sequencing primer (orange), Index information used for sample identification after pooling (yellow), and Illumina flow cell binding motif (white). (Color figure online)

the resulting amplicons had following parts from 5' to 3' (Fig. 1): (1) P5 motif for flow cell hybridization, (2) index 1 of 8 bp, (3) P5one sequencing primer, (4) specific forward primer, (5) target DNA for sequencing; specific reverse primer; (6) P7 sequencing primer, (7) index 2 of 8 bp, and (8) P7 motif for flow cell hybridization.

SSR primer testing

To ascertain the applicability and usability of the developed primers, we first amplified them in singleplex reactions to tested two scenarios: (1) transferability of the developed primers on East African Nile tilapia, and (2) cross-species amplification of *T. zillii*. The amplification success rate of the candidate loci during PCR reactions was tested by assaying a Nile tilapia sample from Uganda. For cross-species amplification, only two genomic DNA samples for *T. zillii* were tested on a panel of 35 SSR loci. PCR reactions were conducted in a 10 µl total volume. All primer pairs were tested using the QIAGEN Multiplex PCR Master Mix (Qiagen, CA, U.S.A) kit. PCR reaction volume during Nile tilapia amplification was composed of 5 µl Master mix, 4 µl primer mix and 1 µl genomic DNA. Primer mix was a combination of 1 µl Reverse primer + 1 µl Forward primer (100 µM each), plus 98 µl of water. Finally, the cycler reaction mixtures were performed based on the following PCR profiles: initialisation at 95 °C for 15 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 60 s, elongation at 72 °C for 60 s and final extension at 72 °C for 10 min. The success of the PCR products was tested by electrophoresis on 1.8% agarose gel. Here, 33 primer pairs were successfully identified in specific PCR products, which subsequently were used for the multiplex PCR approach on Ugandan Nile tilapia populations. Successful markers for cross-species amplification were based on the PCR singleplex gel products based on two replications.

PCR multiplex and Illumina sequencing

All 33 gel-screened primers were combined in a single multiplex reaction. PCR reactions were carried out in a 10 µl total volume containing 5 µl Master mix, 2.5 µl water, 0.5 µl primer mix (1 µM each) and 2 µl genomic DNA. Thermal cycler profiles were analogous to the single-plex PCR. The resulting PCR products were purified using magnetic bead procedures, following slight modifications from Agencourt AMPure XP PCR Purification protocol. Here, we mixed 4 µl PCR products with 2.86 µl of AMPure XP beads (Beckman Coulter, Inc, Brea, CA) and incubated for five minutes at room temperature. Bound DNA beads were captured by an inverted magnetic bead extraction device, VP 407-AM-N (V&P SCIENTIFIC, INC) and washed twice in 200 µl of 80% ethanol for 45 s. Later, the beads were dried

at room temperature for five minutes and eluted in 17 µl of elution buffer (65°C 10 mM Tris-HCl, pH 8.3).

The second (index) PCR was performed in a total reaction volume of 10 µl, containing 5 µl master mix, 2 µl each of index primer (1 µM) P5 and P7, and 1 µl of template purified PCR products. PCR was run with the following thermocycler conditions: 95 °C for 15 min, followed by 10 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C for 60 s, elongation at 72 °C for 60 s and final extension at 72 °C for five minutes. Finally, all the PCR products were pooled and sent for PE 300 bp sequencing in an Illumina MiSeq at the Genomics Service Unit at Ludwig Maximilian Universität, München, Germany. The samples used in this work occupied 11% of the MiSeq run.

Sequence analysis and SSR-GBS genotyping

Reads from Illumina were quality controlled and merged as described in "SSR discovery". Overlap was only possible because SSR motifs were covered completely by one of the paired reads. The resulting sequences should start with the forward primer and end with the reverse and we used this criterion to de-multiplex the sequences according to primer content, creating one fastq files per sample and locus using script 1 (Supplementary material Table S4). This script looks for mismatches between the amplification primers from the beginning (forward) and the end (reverse) of the sequences. Only reads with a mismatch to both primers below two base pairs were considered. From this step the allele calling was performed in two steps: first using the AL, which resembles the traditional SSR genotyping, and then by considering possible SNP variation within alleles of the same length recovering the whole amplicon sequence information (WAI). After each step, a codominant matrix was produced allowing for a comparison between data that would be produced by tradition SSR genotyping to the one from SSR-GBS. Allele calling based on AL incorporates length variation at the repetition motif plus possible indels in the flanking regions. All types of length variation and SNPs are used as information with WAI because two amplicon sequences were only considered as the same alleles if they were equal.

To call alleles using AL, we calculated the length distribution per sample and per locus using script 2 producing files containing the number of times each length occurs per marker and sample (Supplementary material Table S4). Then we used the script 3 to automatically call the alleles and plot histograms based on sequence length resembling the chromatograms obtained in traditional SSR genotyping (Supplementary material Table S4). Only genotypes with a minimum depth of 10 reads were considered. Automatic allele calling considered homozygous genotypes if there was a length with a frequency equal or above to 90% of the total

number of reads. Heterozygous genotypes were assumed if the frequency of two lengths was above 90% of the reads and if the frequency of both lengths did not differ by more than 20%. In case these criteria were not matched, the genotypes were marked for manual control. Nevertheless, all possible genotypes were manually controlled with the aid of the produced histograms.

For WAI allele calling, the sequences with the same length of the alleles defined in the AL step were extracted and used to produce a 70% consensus sequence per length class. The extraction of sequences per length allele was done using the script 4 and the consensus sequence the script 5. In the 70% consensus, the positions with the most common nucleotide of frequency below 70% were coded with the ambiguous base "N" and considered as potential heterozygous SNPs. These sequences were divided into two files based on the two most frequent nucleotides for that position using the script 6. In case more than one potential SNP was found, these positions were considered as linked and the two most frequent nucleotide combinations were recovered. We observed, that chimeric sequences could occur between alleles that differed by more than one SNP. This causes the occurrence of sequence states intermediate between the alleles. In each case these intermediate states were less frequent and could be unambiguously resolved visually, either by comparing them to other alleles in the sampling, or by including the sequence length as additional information. However, these occurrences were rare. Only in a few cases which were not considered due to low overall read counts more than two similarly frequent nucleotide combinations were found. Similarly, the two most frequent combinations between a SNP and length signal was called as allele in samples that were heterozygous with AL. WAI allele calling was finally done using scrip 7 where a number was attributed to each unique sequence (allele) and saving this information in a codominant matrix. All scripts are available in GitHub (<https://github.com/mcurto/SSR-GBS-pipeline>) and detailed description of script 1–script 7 is given in supplementary table S4. The sequence analyses resulted in 26 loci with genotypes for most of the samples that were used for further statistical assessment. Raw reads can be found in the SRA database with the references SRX3398667 to SRX3398776.

Statistical analyses

Descriptive population genetics analyses for SSR loci were determined using various programs. The software Micro-Checker version 2.2.0.3 (Van Oosterhout et al. 2004), was used to estimate the presence of null alleles, evidence of allele drop-out, or stuttering during PCR amplification. Test for deviations from Hardy–Weinberg Equilibrium (HWE) and calculations of the fixation index (Fis) were performed in GenePop version 4.6.9 (Rousset 2008). Markov chain

parameters for all tests in GenePop were run at 10,000 dememorizations, 100 batches and 5000 iterations per batch and Fis values were recorded (Weir and Cockerham 1984). Fis was specifically determined to assess the type of HWE on the populations, in aspects of excess or deficiency heterozygosity (Dorak 2014). Here, positive or negative Fis values indicate excess homozygosity or excess heterozygosity (outbreeding) respectively (Dorak 2014). Observed heterozygosity (Ho), expected heterozygosity (He) and loci polymorphic information content (PIC) were determined using Cervus software version 3.0.7 (Kalinowski et al. 2007). Allelic richness and number of alleles per locus were calculated with Fstat program version 2.9.3.3 (Goudet 2001). To further assess the extent of informativeness and hence usability of the developed markers, we tested the genetic structure and principle coordinate analysis (PCoA) on the four Nile tilapia populations using STRUCTURE version 2.3.4 (Hubisz et al. 2009) and GenAlex version 6.5 (Peakall and Smouse 2006), respectively. STRUCTURE classifies populations by genetically allocating them into groups whose individuals share similar patterns of variation (Porrás-Hurtado et al. 2013). The program further is rendered useful as it can identify subpopulations of the whole population by maximizing HWE linkage within potential subpopulations (Porrás-Hurtado et al. 2013). STRUCTURE was set at 100,000 burn-in period and the application of Markov Chain Monte Carlo (MCMC) was run at 500,000 replications, with each cluster (K) assigned to 10 iterations. STRUCTURE default settings for the admixture model and allele frequencies correlated were implemented. For inference to the K that best suits the data, we ran STRUCTURE HARVESTER. Here, the program collates STRUCTURE results and validates multiple K values for optimal detection and thereby depicts the best K value from tens or hundreds of iterations (Earl and vonHoldt 2012), as indicated in the supplementary material Table S1, Fig. S1 and Fig. S2. Similarly, for presenting informative genetic STRUCTURE outputs, we ran the CLUMPAK clustering Markov package pipeline across the K values for summation and graphical representation of the results obtained from STRUCTURE (Kopelman et al. 2015). From these analyses, we present and compare the results regarding: PIC, number of alleles (Na), allelic richness (Ar), HWE per population, Fixation index (Fis), PCoA, and STRUCTURE based on the two allele calling methods, AL and WAI.

Results

From the 2,404,293 paired reads generated for primer design, 6,724 contained SSR motifs and complied with our quality criteria. Out of the 35 developed SSR primer pairs, only 26 successfully amplified most of the Ugandan

Nile tilapia samples used (Table 2). Both 2mers and 3mers dominated (each eight in number), followed by 4mers and 5mers; each seven in numbers. For SSR-GBS, we generated a total number of 4,783,118 paired reads, of which 1,738,315 passed our quality control procedure, yielding an average number of 181 reads per marker and individual. In total, 12% of the called alleles were flagged for manual control, between 0 and 73% per locus. However, only in four loci most genotypes could not be determined (39–73%), mainly because of a frequently occurring unspecific product or length polymorphisms out of frame of the microsatellite motif and thus easy to correct. Without these markers only about 3.5% were flagged.

Not all of the developed markers were very informative based on the genotyping results (Table 2; Fig. 2). For instance, two loci, Ti27 and Ti28, exhibited two alleles, which were the lowest number of alleles indicated by WAI, but with AL showing only one allele for locus Ti27 (Table 2). The same loci, Ti27 and Ti28, indicated similarly the least allelic richness of less than 2 considering both allele calling methods (Table 2). In general, the rest of the loci indicated allelic numbers and richness ranging from 3 to 56 and from 1.8 to 13.9 respectively (Table 2; Fig. 2a). In total for the whole dataset, AL resulted in 270 alleles compared to 407 alleles using WAI (Table 2). Similarly, WAI yielded a total of 157 allelic richness contrary to AL with a total of 140 (Table 2). The analysis consistently recovered higher values for number of alleles when we used whole sequence information rather than sequence length for a specific population per locus (Fig. 3). For instance WAI generated, for Lake Albert, 212 alleles compared to 157 for AL; for Lake George, 222 compared to 187 for AL; for River Nile, 129 compared to 110 for AL; and for Victoria, 207 compared to 170 for AL (Fig. 3; Supplementary material Table S3). Number of alleles based on the two allele calling methods was congruent to allelic richness where WAI presented higher values than AL throughout the loci (Supplementary material Table S2).

The PIC per locus generally varied from 0.03 to 0.94 with WAI and 0.00 to 0.93 based on AL (Table 2; Fig. 2b). The WAI procedure displayed 18 loci with a polymorphic information content (PIC) ≥ 0.50 , contrary to the AL with 14 loci (Table 2; Fig. 2b). WAI exhibited four loci with a PIC > 0.25 and < 0.50 and four loci with a PIC < 0.25 . Contrary, AL revealed five loci with a PIC > 0.25 and < 0.50 and seven loci with a PIC < 0.25 (Table 2; Fig. 2b). Based on WAI, the expected and observed heterozygosity ranged from 0.00 to 0.94 and 0.00 to 0.89, respectively (Table 2, Table S8). On average, WAI recovered, respectively, an H_o and an H_e of 0.51 and 0.61, while AL an H_o and an H_e of 0.46 and 0.51, respectively (Table 2). For cross-species amplification, a total number of 22 loci exhibited positive amplification on *T. zillii*, albeit four indicated weak gel bands. Interestingly,

four loci (Ti10, Ti11, Ti21 and Ti25) that initially showed negative results following the Nile tilapia amplification reportedly tested positive with *T. zillii* (Table 2).

Test for Hardy–Weinberg Equilibrium (HWE) for WAI and AL within the four Nile tilapia populations showed in four markers consistent deviations in all populations (Ti, 22, Ti29, Ti31, Ti35). Congruence between results of tests for HWE and of Micro-Checker indicate that this stems from the presence of null alleles (Table 3 and Table S5–7). This is also supported by Fis values, indicating deviations from HWE due to heterozygosity deficiency (excess homozygosity) for almost all loci. When all samples considered, the populations GN and RNK showed deviations at additional 4 and 3 loci respectively, of which 4 occurred with both allele call methods and three only with WAI. Populations of Lake Albert and Lake Victoria indicated the majority of loci deviating from HWE (additional 7 or 12 loci respectively). 11 of the deviations occurred with both allele call methods, 5 only with WAI and 2 only with AL. The high amount of deviations from HWE in these populations is consistent with the subpopulation structure indicated in the STRUCTURE analysis.

In the STRUCTURE analysis using STRUCTURE HARVESTER based on WAI and AL, $K=5$ and $K=2$ were, respectively, the best K values for the populations (Fig. 4; Fig. S1; Fig. S2; Table S1). Results from $K=2$ up to $K=5$ are shown, to include also the population number used in this study (Fig. 4). K values from 6 to 10 were uninformative and therefore were not considered based on STRUCTURE HARVESTER outputs (Table S1, Fig. S1; Fig. S2). In comparison, STRUCTURE results in more ambiguous clustering for AL than WAI (Fig. 4b). STRUCTURE results indicated a separation among all populations only for WAI (Fig. 4). At $K=5$, the separation between the River Nile and Lake Albert populations for WAI was not evident in the AL results. Moreover, substructure within the Lake Albert and Victoria populations is indicated at $K=4$ and 5 by assigning single individuals within one of the lakes to other clusters (AN18 for Lake Albert, and VKM 10 and 13 for Lake Victoria). Generally, STRUCTURE analyses were consistent with PCoA results where a slight division between the populations is indicated (Fig. 5). Here, Lake Albert is connected to the Nile which is reflected also in the analysis. Albert and the Nile populations are assigned to the same cluster. Lake George individuals are assigned to their own cluster at $K=2$ to $K=5$. Lake Victoria populations seem less strongly divided from Lake Albert than George (Fig. 5).

Excluding the individuals in Lake Albert and Victoria populations that were assigned to different clusters, a test for HWE and for null alleles using Micro-Checker, indicated no pronounced deviations from HWE with no differences between the allele call methods (Table 3; Table S6). The only significant signals were for the above-mentioned four

Table 2 Information of the final set of 26 SSR markers, concerning primer sequence, repetition motif, amplicon length, variation, species cross-amplification

Loci	Primer sequence (5'-3')	R. motif	ASR	PIC	PI-C*	Na	Na*	Ho	Ho*	He	He*	Ar	Ar*	CS
Ti12	F:GCCACCAAAATATTCGTGT R:CCAATGTCCTCTCTGAA	(TAC)12	377-416	0.39	0.06	10	5	0.31	0.05	0.44	0.06	3.4	1.8	--
Ti16	F:CAGACGTAGCGATAATCT R:GAACACATCCATTTCCAC	(AAC)10	377-428	0.16	0.16	8	8	0.13	0.14	0.16	0.17	2.5	2.9	**
Ti18	F:CTGAAGTCTCTGAGATT R:CATGTCTTGGCAGCTTA	(AC)15	400-402	0.16	0.07	7	6	0.13	0.04	0.16	0.07	2.3	1.8	**
Ti34	F:GCTTACAGTACATTGTGTC R:CTGATGAGAAAACAGACGC	(TCTCT)10	375-415	0.90	0.86	26	12	0.84	0.80	0.91	0.88	10.5	8.8	**
Ti15	F:GCTGTGATCATCTGGAGAA R:AGGATCTAGAACCTCCAACC	(TGC)10	394-403	0.10	0.08	6	4	0.07	0.05	0.10	0.08	2	2.0	**
Ti2	F:TCTGGGCTAACACACAG R:AAGGTGTCAACAGTTTAGG	(CA)18	402-434	0.77	0.79	22	17	0.75	0.86	0.80	0.81	7.4	8.5	**
Ti29	F:AGGTCAAGATCAAGCAGTT R:CATCAACATAATTCAGTGTGA	(TAAA)10	345-355	0.52	0.24	12	12	0.20	0.11	0.60	0.25	4.1	4.0	**
Ti1	F:TTATCAGTGTGACGTCTT R:GTTTGGCTGCTACACATTC	(TG)10	399-407	0.50	0.50	8	6	0.48	0.51	0.52	0.52	3.5	3.5	**
Ti18	F:AGCAAGTGATAGACACTG R:TACATAGACATGACGTTTC	(ATCT)8	397-461	0.92	0.88	33	17	0.89	0.95	0.93	0.90	11.7	10.1	**
Ti4	F:TGTGCAGATAGANTAGCCC R:GAACGAGAAAATGTTGGTGT	(GT)18	403-415	0.78	0.77	11	9	0.82	0.84	0.81	0.81	6.1	6.0	--
Ti13	F:AATCCGTGAGCTGCAGATAG R:GCTGATTAACACAAAGTTGG	(ATG)10	394-403	0.38	0.38	5	6	0.27	0.29	0.40	0.40	3.7	4.1	--
Ti26	F:ATTGCTTCATCCCTTGAGTT R:ACACCGAAACCTAATGACA	(ACAA)10	411-427	0.64	0.63	10	10	0.56	0.58	0.68	0.68	5	5.5	--
Ti35	F:TCAACCAACACTCTCTTT R:AAACTAAGTCCAGCTCATGA	(AAAAG)14	364-424	0.90	0.83	25	17	0.53	0.59	0.91	0.85	11	9.1	**
Ti28	F:GTCTTGGGATTGAGATCA R:CGGAGGTTTCTCTCTTTAA	(ATTCA)8	384-384	0.26	0.01	2	2	0.26	0.01	0.30	0.01	2	1.1	*
Ti17	F:AACTGAGAGAGACCTTGG R:ATCATCTTCTCTACTGCTT	(GAA)21	364-391	0.71	0.75	17	13	0.62	0.62	0.74	0.73	6.3	6.4	*
Ti5	F:AAGGAGATGATCAGGACAC R:AGACCTCCACTGTGATCTTA	(CA)10	401-419	0.53	0.25	9	7	0.50	0.21	0.56	0.26	4.5	3.5	--
Ti33	F:GCTTATGCTGTATGAGTT R:CGACTTCGTGTGTGATTTGG	(TTCAA)6	382-422	0.82	0.82	9	8	0.75	0.76	0.85	0.85	6.5	6.7	--
Ti31	F:GAACATCCACAGAGCCA R:AGGCTTCTTACAGTTGGATG	(CTAAT)7	302-380	0.75	0.73	13	9	0.20	0.34	0.78	0.77	6.3	5.8	--
Ti22	F:ACTGACCAAGTGTCTTGTAT R:AACTGCTGTGTGAATTTG	(CTAT)20	303-475	0.94	0.93	56	44	0.59	0.65	0.95	0.94	13.9	15.4	*
Ti7	F:TCTTTGTGACAGACTGTGT R:ACTCTGCTTTTAGTCAATCA	(AC)17	404-424	0.73	0.75	14	13	0.92	0.96	0.76	0.78	6.8	7.7	**

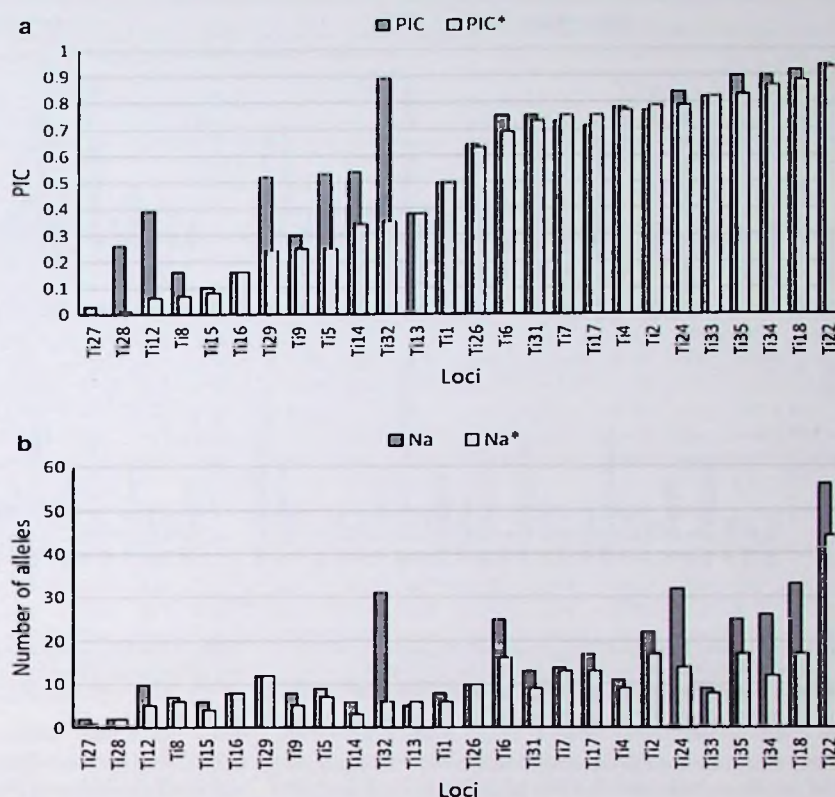
Table 2 (continued)

Loci	Primer sequence (5'-3')	R. motif	ASR	PIC	PIC*	N _a	N _a *	H _o	H _o *	H _e	H _e *	A _r	A _r *	CS
Tf9	F:CTCAGTGCAGAGCCAAA R:CTTGGCAATCAAAAGAACAA	(AC)10	409-415	0.30	0.25	8	5	0.27	0.24	0.32	0.26	3.6	3.4	**
Ti14	F:TCCTTAAATATGCCACCA R:TAGTCTTTTATGGCTCTGG	(TAA)19	368-374	0.54	0.34	6	3	0.43	0.28	0.58	0.38	4.3	6.0	**
Ti27	F:CTGCTTTCTTGATGGGA R:ATGCACAAATTTAAGGCC	(TTT)6	391-391	0.03	0.00	2	1	0.03	0.00	0.03	0.00	1.3	1.0	--
Ti6	F:CAGCTCTCATGAACACTTGA R:ACCCATAAATCACACAGTC	(GA)23	384-418	0.75	0.69	25	16	0.97	0.98	0.78	0.73	7.1	6.5	**
Ti32	F:CAGGAATGGCTCCAAAATG R:TTGTAGCTAGGAATCAGTGC	(AAAAT)7	330-350	0.89	0.35	31	6	0.88	0.36	0.90	0.39	10.7	3.4	**
Ti24	F:ACTGACAAACATAAGACATATGC R:CACAGTTTGAATCCACCATC	(TTAC)9	403-455	0.84	0.79	32	14	0.78	0.73	0.85	0.81	10.7	8.5	*
Ti10*	F:CCATTGATTAGTGAACACC R:ACATGAATGCCAAATGACAG	(TTA)9	354-403											**
Ti11*	F:CATTTGCGATGACCTTAAGT R:CCCTGCTTTTATGCTACTT	(TAA)12	384-452											**
Ti21*	F:CAAACTAAAGCTTCCCCC R:GGATGTTTTCAGTGTGATGT	(TAGA)11	384-491											**
Ti25*	F:TTCAACTCAACACAGTTGCT R:TCCATCAAGATGAATATAGTGT	(ATCT)6	417-420											**
Total				15.1	12.8	407	267	13.1	11.9	15.8	13.4	157.2	143.5	
Ave				0.58	0.49	15.7	10.3	0.51	0.46	0.61	0.51	6.1	5.5	

Variability values were estimated using all four Ugandan Nile tilapia populations

F forward, R reverse, R. motif repeat motif, ASR allele size range (bp), PIC WAI polymorphic information content, PIC* AL polymorphic information content, N_a and N_a* number of alleles for WAI and AL, respectively, A_r and A_r*=allelic richness for WAI and AL, respectively, H_o* and H_e* observed and expected heterozygosity for WAI, H_o and H_e observed and expected heterozygosity for AL, Superscript n loci which tested negative with Nile tilapia amplification, CS cross-species amplification success, *strong PCR products, **weak, -- very weak, and Ave average

Fig. 2 Comparison between allele calling methods: whole sequence information (dark grey) and amplicon length (white). This was performed using number of alleles (Na) and polymorphic information content (PIC) for each of the 26 SSR loci. a, b represent PIC and Na per locus respectively



markers for all populations, besides that only three to four markers showed deviations in single populations. Of a total of 12 occurrences, 8 applied to both methods, and 3 only to WAI and 1 only to AL.

Discussion

With the genesis of high throughput sequencing methods, the classical genotyping approaches can be modified to recover a high level of information and to increase automation. This decreases the necessity for technical expertise in SSR genotyping that remains one difficulty despite the widespread use of the markers worldwide (Sunnucks 2000; De Barba et al. 2017). Constraints in allele call for SSR genotyping include the occurrence of null alleles, stutter bands as result of the slippage artefact in PCR reaction or allele dropout caused by amplification or fragment measurement (Vartia et al. 2016). In this study, we compared the information usually recovered from traditional electropherogram (AL)

frequency distribution for scoring microsatellite genotypes with WAI for allele scoring/calling procedures. To a greater extent, the two allele calling procedures varied in generating the number of alleles, allelic richness, PIC, and genetic structure, of the studied Nile tilapia populations. Here, WAI approach consistently indicated to be more informative in aspects of cataloguing these herein mentioned genetic characteristics in contrast to the AL frequency procedure. However, this was unsurprising because unlike AL procedure, WAI approach systematically surveys the entire sequence hence summarising variability from the repetition motif and SNPs in the flanking regions. In our findings, the high level of variation also resulted in a genetic structure where every population could be characterized by its own cluster, which was not possible when AL was considered. One obstacle in the traditional microsatellite genotyping approach is attributed to genotyping artefacts and biases which originate from equipment specifications and different laboratory methods. With SSR-GBS, because the entire sequence information is utilised, this may no longer be a problem. This approach

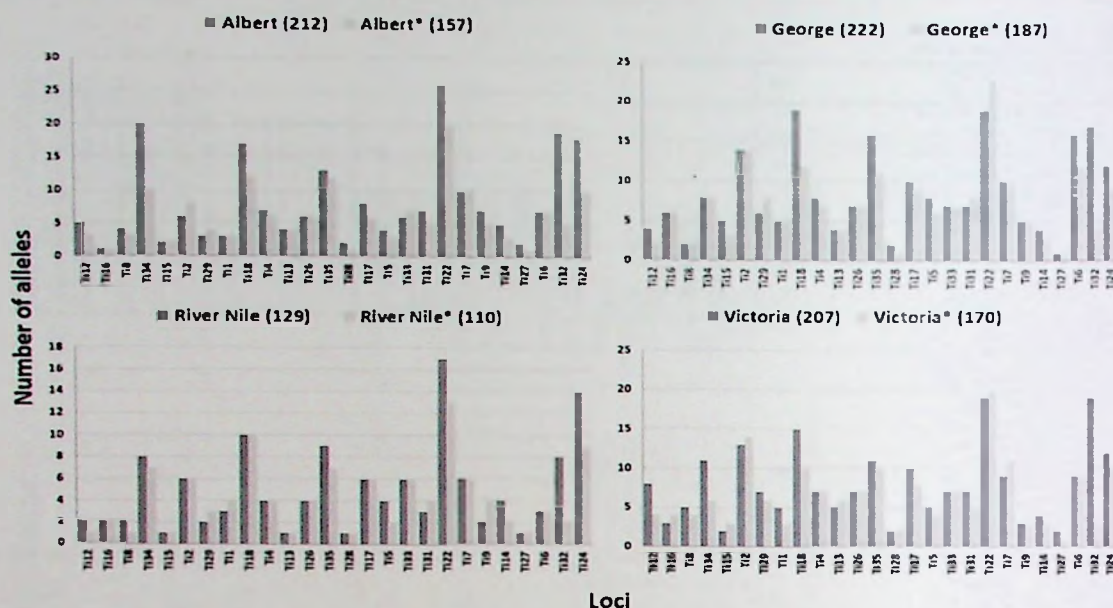


Fig. 3 Comparison between allele calling methods based on number of alleles exhibited per population for each of the 26 SSR loci. Whole amplicon information (dark grey) and amplicon length (light grey).

Values in brackets indicate total number of alleles generated from the two allele calling procedures

(WAI) will allow for the creation of reliable and informative SSR genotype databases that can be used for meta-analytical studies and long-term projects with the contribution of several research groups. For the reasons stated above, it is expected that SSR-GBS will become a standard practice in conservation genetics.

Despite a few loci indicating PIC values of less than 0.25, most of the developed markers were polymorphic and informative. PIC values were congruent with the expected heterozygosity and allelic numbers/richness among the populations, translating into their capability and reliability for molecular genetic analyses. In genetic marker development, heterozygosity and PIC are seen as important attributes to estimate information content of loci and their applicability in down-stream analysis (Nagy et al. 2012; Chesnokov and Artemyeva 2015). In the current study, our analysis indicated 18 loci with PIC values above 0.5 and 4 loci with PIC values below 0.5 but above 0.25. Here, the informative loci summed up to 22. According to Botstein et al. (1980), highly informative markers should have a PIC value of greater than 0.5, reasonably informative markers a PIC of greater than 0.25 and slightly informative markers of PIC of less than 0.25. Therefore, even though 18 loci exhibited a PIC greater than 0.5, it is likely that the remaining loci (4) with a PIC

greater than 0.25 are also useful if combined with the former during genotyping (Sunnucks 2000). Zane et al. (2002) showed that the use of several loci can increase the sensitivity of population genetic analyses and by using SSR-GBS the number of loci can easily be increased.

From the 26 tested loci, four loci deviated from HWE for most studied populations. Positive Fis and Micro-Checker results indicated that this might be caused by allele null alleles, a known limitation of amplification based genotyping approaches where one of the copies in a heterozygote individual is failed to be detected and genotyped as homozygote (Selkoe and Toonen 2006). WAI did recover more information and might have allowed to observe undetected subpopulation structure within the lakes investigated. Based on STRUCTURE results this was the case of the lakes Albert and Victoria, which harbour populations with individuals assigned to multiple clusters. In accordance to this, is the finding that a higher number of loci deviate from HWE in these populations. Given the translocation and artificial stocking background especially of lake Victoria, substructure might be caused within these populations due to human mediated admixture between distinct gene pools (Tibihika et al. 2018). Further studies will focus on how Nile tilapia stocks were affected by this anthropogenic activity.

Table 3 Tests for deviations of Hardy–Weinberg Equilibrium (HWE) per population and locus for both allele calling methods without considering the samples contributing to subpopulation structure in Lakes Albert and Victoria

Populations	AN				GH				RNK				VKM			
	p ¹	p*	Fis ¹	Fis*	p ¹	p*	Fis ¹	Fis*	p ¹	p*	Fis ¹	Fis*	p ¹	p*	Fis ¹	Fis*
Ti12	0.02	–	0.48	–	0.10	–	–0.08	–	0.01	–	0.79	–	0.03	1.00	0.32	–0.02
Ti16	–	–	–	–	0.69	0.69	–0.02	–0.02	–	–	–	–	–	–	–	–
Ti8	0.25	–	0.34	–	–	–	–	–	1.00	–	–0.10	–	–	–	–	–
Ti34	0.79	0.02	–0.02	0.12	0.62	0.61	–0.02	–0.02	0.75	0.79	–0.03	–0.04	0.36	0.57	0.06	–0.03
Ti15	–	–	–	–	1.00	1.00	–0.04	–0.02	–	–	–	–	–	–	–	–
Ti2	0.33	0.92	0.11	–0.12	0.60	0.95	–0.05	–0.09	0.34	0.94	0.08	–0.20	0.62	0.74	0.01	–0.01
Ti29	0.00	1.00	0.83	–0.03	0.00	0.00	0.59	0.58	0.08	0.02	0.38	0.50	0.00	0.00	0.73	0.82
Ti1	0.83	0.87	–0.11	–0.14	0.04	0.04	0.06	0.07	0.21	0.44	0.21	0.07	0.43	0.83	0.07	–0.11
Ti18	0.14	0.87	0.06	–0.09	0.89	0.88	–0.04	–0.05	0.01	0.86	0.12	–0.10	0.93	1.00	–0.08	–0.14
Ti4	0.19	0.63	0.09	–0.05	0.96	0.96	–0.19	–0.19	0.86	0.79	–0.22	–0.17	0.91	0.93	–0.10	–0.11
Ti13	0.07	0.07	0.32	0.32	0.14	0.14	0.23	0.23	–	–	–	–	0.36	1.00	0.03	–0.17
Ti26	0.85	1.00	–0.16	–0.28	0.26	0.27	0.06	0.06	0.00	0.00	0.34	0.34	0.00	0.00	0.43	0.36
Ti35	0.00	0.01	0.61	0.23	0.00	0.06	0.26	0.15	0.00	0.06	0.45	0.30	0.00	0.05	0.38	0.28
Ti28	1.00	–	–0.02	–	0.97	–	–0.25	–	–	–	–	–	0.21	–	0.35	–
Ti17	0.08	0.09	0.27	0.32	0.12	0.13	0.03	0.03	0.66	0.66	0.01	0.01	0.02	0.04	0.26	0.25
Ti5	0.91	1.00	–0.16	–0.06	0.20	0.23	0.19	0.13	0.07	0.21	0.18	0.35	0.80	1.00	–0.13	–0.13
Ti33	0.21	0.21	0.13	0.12	0.49	0.50	0.01	0.01	0.07	0.07	0.05	0.05	0.59	0.62	–0.08	–0.09
Ti31	0.00	0.00	0.73	0.51	0.00	0.00	0.78	0.54	0.01	0.00	0.78	0.60	0.00	0.00	0.67	0.54
Ti22	0.00	0.00	0.25	0.23	0.00	0.00	0.68	0.44	0.00	0.03	0.27	0.20	0.00	0.00	0.37	0.30
Ti7	0.98	1.00	–0.13	–0.18	1.00	1.00	–0.22	–0.22	0.99	1.00	–0.32	–0.41	1.00	1.00	–0.38	–0.40
Ti9	0.33	1.00	0.12	–0.10	0.02	0.04	0.22	0.19	–	1.00	–	–0.07	1.00	1.00	–0.05	–0.02
Ti14	0.00	0.05	0.40	0.45	0.01	0.01	0.20	0.43	0.13	0.43	0.21	0.17	0.03	1.00	0.09	–0.24
Ti27	–	–	–	–	–	–	–	–	–	–	–	–	1.00	–	–0.02	–
Ti6	1.00	1.00	–0.74	–0.75	0.88	0.91	–0.10	–0.13	1.00	1.00	–0.78	–0.85	1.00	1.00	–0.53	–0.54
Ti32	0.25	0.10	–0.02	–0.01	0.01	0.14	0.02	0.07	0.73	0.60	–0.01	0.06	1.00	0.41	–0.07	0.03
Ti24	0.92	0.53	–0.05	0.00	0.48	0.53	0.05	0.03	0.59	0.26	0.00	0.07	0.34	0.22	0.06	0.09
NDL	6.00	4.00			8.00	6.00			6.00	4.00			8.00	5.00		
NLwi	3.00	6.00			2.00	3.00			6.00	7.00			3.00	5.00		

AN Albert; GH George; RNK Nile; VKM Victoria. NDL number of deviated loci from HWE, NLwi number of loci without information

p¹ probability value for WAI allele calling, p* probability value for AL allele calling, Fis¹ and Fis* fixation indices for WAI and AL respectively

The genotyping based on WAI showed slightly more loci deviating from HWE than with AL (3 vs 1 after exclusion of individuals that cause substructure within populations). Since the WAI genotyping is able to detect a higher number of alleles, such effects might become more likely using this method.

Cross-species amplification results were unexpected given the genetic distance between Nile tilapia and *Tilapia zillii* (Elghohashy et al. 2005). A total of 18 loci strongly amplified *T. zillii* genomic DNA. Four loci (Ti10, Ti11, Ti21 and Ti25), which showed negative results for Ugandan Nile tilapia amplification, were positive with *T. zillii*. This suggests the existence of allele drop out in Ugandan Nile tilapia, while the individual from Ethiopia shared alleles with *T. zillii*. The capacity of SSR markers to cross amplify species might also be locus-dependent (Gen-Hua et al. 2010),

although species' phylogenetic relationships or homoplasy could play a role (Bezault et al. 2012). The pattern of cross amplification between the species might here be explained by factors like outbreeding/hybridisation or admixed populations (Barbara et al. 2007).

Conclusions

The development of SSR-GBS has enhanced microsatellite analysis. The Illumina approach genotyped SSR loci successfully, demonstrated polymorphism and their usability in down-stream applications such genetic diversity and structure. We expect that also other analyses like pedigree analysis, gene flow and hybridisation, genetic linkage maps, etc. can be facilitated by the method. SSR genotyping that

Fig. 4 Genetic structure bar plots for the four East African (Uganda) Nile tilapia populations based in cluster assignment probability calculated in STRUCTURE for the best values of $K=5$ and $K=2$. **a** represents Nile tilapia genetic structures inferred from whole amplicon information and **b** from amplicon length allele calling methods

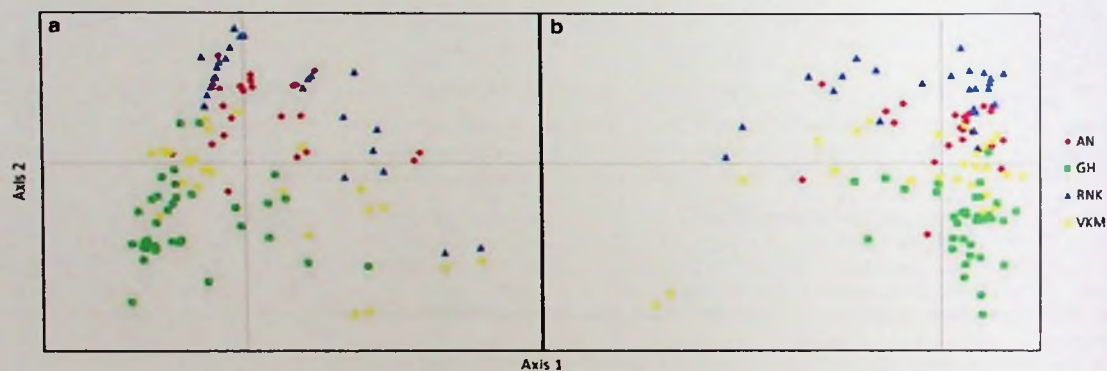
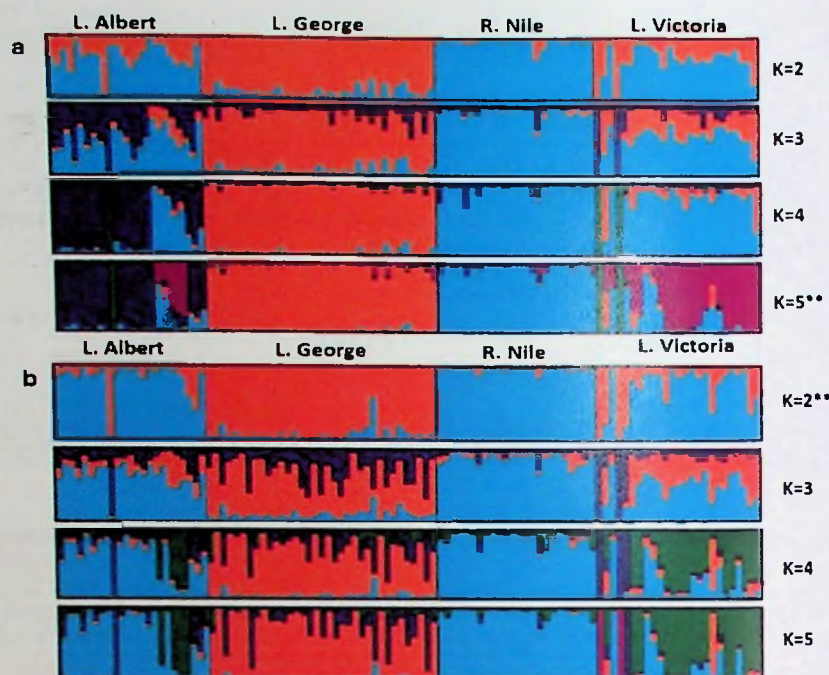


Fig. 5 Principle coordinate analysis (PCoA), depicting genetic similarities/variabilities between the four East African (Uganda) Nile tilapia populations. **a** represents PCoA results from whole amplicon information and **b** from amplicon length allele calling methods

uses automated WAI for allele calling greatly reduced bias and artefacts, and thereby yielded more information than the commonly employed traditional AL. The reproducibility of the method could be vital for Nile tilapia breeding programs, because genotyping from multiple generations across

a long period of time are directly comparable. STRUCTURE, PCoA, PIC and H_e results suggest that the majority of the presented 26 loci, particularly the 22 polymorphic loci, are highly informative for comprehensively studying the Nile tilapia population patterns. In addition, this number

can be easily increased using this approach. The findings from the current study should aid in a more integrative and comprehensive approach towards restoration, management and conservation of the East African Nile tilapia and related species. Future studies should consider the application of SSR-GBS for Nile tilapia congeners; *Oreochromis variabilis* and *Oreochromis esculentus* (Ngege), as the species have been recorded as endangered. Knowledge about the genetic make of the herein studied tilapiines might be useful for restoration and subsequent conservation.

Acknowledgements Open access funding provided by University of Natural Resources and Life Sciences Vienna (BOKU). This work was funded by the Austrian Partnership Programme in Higher Education and Research for Development (APPEAR), a programme of the Austrian Development Cooperation (ADC) and implemented by the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH). The bioinformatics pipeline was supported by research funds from the Department of Integrative Biology and Biodiversity Research-University of Natural Resources and Life Sciences Vienna (BOKU). We are also indebted to Kachwekano Zonal Agricultural Research and Development Institute-National Agricultural Research Organization-Uganda for their great assistance during field activities.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All international, national or institutional guidelines for the care and use of animals were observed and adhered to. We also declare that the current work and manuscript have not been submitted to more than one journal for simultaneous consideration and have not been published elsewhere.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agnès JF, Adépo-Gourène B, Abban EK, Fermon Y (1997) Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity* 79(1):88–96. <https://doi.org/10.1038/hdy.1997.126>
- Andrews S, FastQC A (2010) A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Arthofer W, Heussler C, Krapf P, Schlick-Steiner BC, Steiner FM (2018) Identifying the minimum number of microsatellite loci needed to assess population genetic structure: a case study in fly culturing. *Fly* 12(1):13–22. <https://doi.org/10.1080/19336934.2017.1396400>
- Balirwa JS (1992) The evolution of the fishery of *Oreochromis niloticus* (Pisces: Cichlidae) in Lake Victoria. *Hydrobiologia* 232(1):85–89. <https://doi.org/10.1007/BF00014616>
- Barbosa T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Mol Ecol* 16(18):3759–3767. <https://doi.org/10.1111/j.1365-294X.2007.03439.x>
- Bezaul E, Rognon X, Gharbi K, Baroiller J-F, Chevassus B (2012) Microsatellites cross-species amplification across some African cichlids. *Int J Evol Biol* 2012:1–7. <https://doi.org/10.1155/2012/870935>
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32(3):314–331
- Capote N, Pastrana AM, Aguado A, Sánchez-Torres P (2012) Molecular tools for detection of plant pathogenic fungi and fungicide resistance. In: Cumagun CJR (ed.) *Plant pathology*. InTech, Rijeka. <https://doi.org/10.5772/38011>
- Castoe TA, Poole AW, de Koning AJ, Jones KL, Tomback DF, Oyler-McCance SJ, Fike JA, Lance SL, Streicher JW, Smith EN (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE* 7(2):e30953. <https://doi.org/10.1371/journal.pone.0136465>
- Chesnokov YV, Artemyeva A (2015) Evaluation of the measure of polymorphism information of genetic diversity. *Agric Biol* 50(5):571–578. <https://doi.org/10.15389/agrobology.2015.5.571eng>
- Curto MA, Tembrock LR, Puppo P, Nogueira M, Simmons MP, Meimberg H (2013) Evaluation of microsatellites of *Catha edulis* (qat; Celastraceae) identified using pyrosequencing. *Biochem Syst Ecol* 49:1–9. <https://doi.org/10.1016/j.bse.2013.02.002>
- De Barba M, Miquel C, Lobréaux S, Quenette P, Swenson J, Taberlet P (2017) High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Mol Ecol Resour* 17(3):492–507. <https://doi.org/10.1111/1755-0998.12594>
- Dorak MT (2014) Basic population genetics. <http://www.dorak.info/genetics/popgen.html>
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Elghobashy HA, Ramadan AA, Ibrahim IH, Rashed MA (2005) Phylogenetic relationships for some *Tilapia* species using electrophoresis. *Egypt J Exp Biol (Zool)* 1:57–63
- Farrell ED, Carlsson JE, Carlsson J (2016) Next Gen Pop Gen: implementing a high-throughput approach to population genetics in boarfish (*Capros aper*). *Open Sci* 3(12):160651. doi:1098/rsos.160651
- Gen-Hua Y, Kovacs B, Orban L (2010) A new problem with cross-species amplification of microsatellites: generation of non-homologous products. *Zool Res* 31(2):131–140. <https://doi.org/10.3724/SPJ.1141.2010.02131>
- Goudet J (2001) FSTAT: a program to estimate and test gene diversities and fixation indices, version 2.9.3. <http://www.unil.ch/izea/software/fstat.html>
- Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepointevin C, Malausa T, Revardel E, Salin F (2011) Current trends in microsatellite genotyping. *Mol Ecol Resour* 11(4):591–611. <https://doi.org/10.1111/j.1755-0998.2011.03014.x>
- Hassanien HA, Gilbey J (2005) Genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*) revealed by DNA microsatellites. *Aquac Res* 36(14):1450–1457
- Hedrick PW (2001) Conservation genetics: where are we now? *Trends Ecol Evol* 16(11):629–636. [https://doi.org/10.1016/S0169-5347\(01\)02282-0](https://doi.org/10.1016/S0169-5347(01)02282-0)

- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9(5):1322–1332. <https://doi.org/10.1111/j.1755-0998.2009.02591.x>
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16(5):1099–1106. <https://doi.org/10.1111/j.1365-294X.2007.03089.x>
- Kaufman L (1992) Catastrophic change in species-rich freshwater ecosystems. *Bioscience* 42(11):846–858. <https://doi.org/10.2307/1312084>
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15(5):1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Koskinen MT, Hirvonen H, LANDRY PA, Primmer CR (2004) The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. *Heredity* 141(1):61–67. <https://doi.org/10.1111/j.1601-5223.2004.01804.x>
- Lee WJ, Kocher T (1996) Microsatellite DNA markers for genetic mapping in *Oreochromis niloticus*. *J Fish Biol* 49(1):169–171. <https://doi.org/10.1111/j.1095-8649.1996.tb00014.x>
- Liu Z, Cordes J (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238(1):1–37. <https://doi.org/10.1016/j.aquaculture.2004.05.027>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17(1):10–12. <https://doi.org/10.14806/ej.17.1.200>
- Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam KN, Latif MA (2013) A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *Int J Mol Sci* 14(11):22499–22528. <https://doi.org/10.3390/ijms141122499>
- Miller MP, Knaus BJ, Mullins TD, Haig SM (2013) SSR_pipeline: a bioinformatic infrastructure for identifying microsatellites from paired-end Illumina high-throughput DNA sequencing data. *J Hered* 104(6):881–885. <https://doi.org/10.1093/jhered/est056>
- Mwanja W (2000) Genetic biodiversity and evolution of two principal fisheries species groups, the labeine and tilapiine, of Lake Victoria, East Africa. PhD Thesis, Ohio State University, Cleveland
- Mwanja W, Booton G, Kaufman L, Chandler M, Fuerst P, Donaldson E, MacKinlay D (1996) Population and stock characterization of Lake Victoria Tilapia fishes based on RAPD markers. In: Donakson EM, MacKinlay DD (eds) *Aquaculture biotechnology symposium proceedings of the international congress on the biology of fishes*. American Fisheries Society, Washington, pp 115–124
- Mwanja W, Booton GC, Kaufman L, Fuerst PA (2008) 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. *Afr J Biotechnol* 7(11):1769–1773. <https://doi.org/10.5897/AJB08.813>
- Nagy S, Poczai P, Cernák I, Gorji AM, Hegedűs G, Tállai J (2012) PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. *Biochem Genet* 50(9–10):670–672. <https://doi.org/10.1007/s10528-012-9509-1>
- Njiru M, Nzungi P, Getabu A, Wakwabi E, Othina A, Jembe T, Wekesa S (2007) Are fisheries management measures in Lake Victoria successful? The case of Nile perch and Nile tilapia fishery. *Afr J Ecol* 45(3):315–323. <https://doi.org/10.1111/j.1365-2028.2006.00712.x>
- Ogutu-Ohwayo R (1990) The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of introduced species, especially the Nile perch, *Lates niloticus*, and the Nile tilapia, *Oreochromis niloticus*. *Environ Biol Fishes* 27(2):81–96. <https://doi.org/10.1007/BF00001938>
- Okumuş İ, Çiftçi Y (2003) Fish population genetics and molecular markers: II-molecular markers and their applications in fisheries and aquaculture. *Turk J Fish Aquat Sci* 3(1):51–79
- Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira MLC (2006) Origin, evolution and genome distribution of microsatellites. *Genet Mol Biol* 29(2):294–307. <https://doi.org/10.1590/S1415-47572006000200018>
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Resour* 6(1):288–295. <https://doi.org/10.1111/j.1471-1471.2005.01155.x>
- Porrás-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. *Front Genet*. <https://doi.org/10.3389/fgene.2013.00098>
- Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 8(1):103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Ryman N, Palm S, André C, Carvalho GR, Dahlgren TG, Jorde PE, Laikre L, Larsson LC, Palmé A, Ruzzante DE (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Mol Ecol* 15(8):2031–2045. <https://doi.org/10.1111/j.1365-294X.2006.02839.x>
- Schlotterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* 354(6348):63
- Schoebel C, Brødbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M, Řičánová Š, Segelbacher G (2013) Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. *J Evol Biol* 26(3):600–611. <https://doi.org/10.1111/jeb.12077>
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett* 9(5):615–629. <https://doi.org/10.1111/j.1461-0248.2006.00889.x>
- Seyoum S, Kornfield I (1992) Identification of the subspecies of *Oreochromis niloticus* (Pisces: Cichlidae) using restriction endonuclease analysis of mitochondrial DNA. *Aquaculture* 102(1–2):29–42
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26(10):1135
- Sundaray JK, Raval KD, Chakrapani V, Swain P, Kumar D, Ninawe AS, Nandi S, Jayasankar P (2016) Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. *Aquacult Int* 24(4):1089–1102. <https://doi.org/10.1007/s10499-016-9973-4>
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends Ecol Evol* 15(5):199–203. [https://doi.org/10.1016/S0169-5347\(00\)01825-5](https://doi.org/10.1016/S0169-5347(00)01825-5)
- Tariq Ezaz M, Sayeed S, McAndrew BJ, Penman DJ (2004) Use of microsatellite loci and AFLP markers to verify gynogenesis and clonal lines in Nile tilapia *Oreochromis niloticus* L. *Aquac Res* 35(15):1472–1481. <https://doi.org/10.1111/j.1365-2109.2004.01175.x>
- Tibihika PD, Waidbacher H, Masembe C, Curto M, Sabatino S, Alemayehu E, Meulenbroek P, Akoll P, Meimberg H (2018) Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa. *Environ Biol Fishes* 101(3):363–381. <https://doi.org/10.1007/s10641-017-0704-0>
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces.

- Nucleic Acids Res 40(15):e115–e115. <https://doi.org/10.1093/nar/gks596>
- Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4(3):535–538. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
- Vartia S, Collins PC, Cross TF, Fitzgerald RD, Gauthier DT, McGinnity P, Mirimin L, Carlsson J (2014) Multiplexing with three-primer PCR for rapid and economical microsatellite validation. *Hereditas* 151(2–3):43–54. <https://doi.org/10.1111/hrd2.00044>
- Vartia S, Villanueva-Cañas JL, Finarelli J, Farrell ED, Collins PC, Hughes GM, Carlsson JE, Gauthier DT, McGinnity P, Cross TF (2016) A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *R Soc Open Sci* 3(1):150565. <https://doi.org/10.1098/rsos.150565>
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Welcomme R (1966) Recent changes in the stocks of Tilapia in Lake Victoria. *Nature* 212:52–54. <https://doi.org/10.1038/212052a0>
- Xiong J (2006) Essential bioinformatics. Cambridge University Press, New York
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11(1):1–16. <https://doi.org/10.1046/j.0962-1083.2001.01418.x>
- Zhan L, Paterson IG, Fraser BA, Watson B, Bradbury IR, Nadukalam Ravindran P, Reznick D, Beiko RG, Bentzen P (2017) MEGASAT: automated inference of microsatellite genotypes from sequence data. *Mol Ecol Resour* 17(2):247–256. <https://doi.org/10.1111/1755-0998.12561>
- Zhang L, Wang S, Yin S, Hong S, Kim KP, Kleckner N (2014) Topoisomerase II mediates meiotic crossover interference. *Nature* 511(7511):551–556. <https://doi.org/10.1038/nature13442>

4. Chapter 4

3rd paper. Genetic structure evidence of anthropogenic threats in the East African fisheries: the paradigm of Nile tilapia (*Oreochromis niloticus*)

Genetic structure evidence of anthropogenic threats in the East African fisheries: the paradigm of Nile tilapia (*Oreochromis niloticus*, L. 1758)

Short running title: population genetics of *O. niloticus* in East Africa

Papius Dias Tibihika^{†, §}, Manuel Curto[†], Esayas Alemayehu^{‡, ¶}, Herwig Waidbacher[‡], Charles Masembe[¶], Peter Akoll[¶], Harald Meimberg[†]

[†] Institute for Integrative Nature Conservation Research, University of Natural Resources and Life Sciences Vienna (BOKU), Gregor Mendel Straße 33, 1180 Wien, Austria. papius.tibihika@students.boku.ac.at, manuel.curto@boku.ac.at, meimberg@boku.ac.at

[‡] Institute for Hydrobiology and Aquatic Ecosystems Management, University of Natural Resources and Life Sciences Vienna (BOKU), Gregor Mendel Straße 33/DG, 1180 Wien, Austria. esayas.negash@students.boku.ac.at, herwig.waidbacher@boku.ac.at

[§] National Agricultural Research Organization, Kachwekano Zonal Agricultural Research and Development Institute, P.O. Box 421, Kabale, Uganda.

[¶] Department of Zoology, Entomology and Fisheries Sciences-Makerere University Kampala, P. O. Box 7062, Kampala, Uganda. cmasembe@cns.mak.ac.ug

[¶] National Fishery and Aquatic Life Research Centre, Sebeta, P.O. Box 64, Addis Ababa, Ethiopia. esayas.negash@students.boku.ac.at

[§] Corresponding author; Papius Dias Tibihika

papiust@yahoo.com, papius.tibihika@students.boku.ac.at

Tel. +4368 864 971870

Abstract

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*) threatens this species. Using 40 microsatellite loci with genotyping by sequencing 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. Genetic structure analyses revealed four East African *O. niloticus* groups which were consistently supported by F_{st} results. Moreover, the East African populations genetically differed from their Ethiopian and West African counterparts. Moreover, *O. niloticus* from Lake Turkana was observed isolated, while the Ugandan populations were rather closely related, albeit geographical location and anthropogenic activities influenced some genetic clusters. On the other hand, one Ethiopian population was found to be genetically distinct. Signals of genetic erosion were detected in Ethiopian and some East African populations being likely a consequence of either bottlenecks associated with fragmentation or founder effects. *O. niloticus* in Lakes Tana, Victoria and the cluster of Lakes Edward, George as well as Kazinga Channel, may require further exploration for taxonomic recognition. Apart from Lake Tana, the distinctiveness of the cichlid in Lake Victoria and the western Rift Valley Lakes, have not been reported before. Finally, we show that anthropogenic activities have partly contributed to a significant change in the genotypic composition of *O. niloticus* which may compromise the sustainability of its exploitation.

Keywords: cichlids, admixture, fish translocations, genetic structure, geneflow, bottleneck.

Introduction

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 (Agnèse, Adépo-Gourène, Abban, & Fermon, 1997; Trewavas, 1983). For more than nine decades, this species has been intentionally dispersed worldwide, in particular for aquaculture and restocking programmes (Canonico, Arthington, McCrary, & Thieme, 2005; Trewavas, 1983). 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 (Agnèse, Adépo-Gourène, & Pouyaud, 1998; Worthington, 1932). For example, Lake Bunyonyi was stocked with individuals from Lake Edward in 1920 (Worthington, 1932).

In the 1950s, several tilapiine species were stocked into Lake Victoria, Nabugabo and Kyoga to counteract the decline of native fish species (Baliwa, 1992; Kaufman, 1992; Ogutu-Ohwayo, 1990; Welcomme, 1966). The introduced species (*O. niloticus*, *O. leucosticus*, *Tilapia zillii* and *Tilapia melanopleura*) were all suspected to originate from Lake Albert (Baliwa, 1992; Ogutu-Ohwayo, 1990; Welcomme, 1966). However, some introductions might have also originated from Lake Edward (Agnèse et al., 1998; Baliwa, 1992). Following these introductions, the indigenous fish species, such as *O. variabilis* and Ngege (*O. esculentus*), significantly declined (Baliwa, 1992; Mwanja, 2000; Ogutu-Ohwayo, 1990). 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*) (Kaufman, 1992; Ogutu-Ohwayo, 1990). But one additional factor might have been hybridization with the introduced *O. niloticus* (Baliwa, 1992; Mwanja, 2000; Welcomme, 1966). The expanded distribution of *O. Niloticus* in East Africa complicates 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 changed into different populations (Bairwa, 1992; Pullin, 1988; Welcomme, 1966). The loss of indigenous *O. mossambicus* due to hybridization with the introduced *O. niloticus* has been reported in South Africa (Zengeya, Booth, & Chimimba, 2015). The situation in East Africa may have worsened with the recent boom of fish hatcheries and aquaculture production systems (Rothuis et al., 2014). 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 (Nunan, Hara, & Onyango, 2015). 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 (Allendorf & Luikart, 2007). Therefore, in particular for the East African *O. niloticus*, as the species especially were 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 (1983) described seven subspecies of *O. niloticus* in Africa and associated them to different regions or lakes. Nevertheless, this classification was contradicted as misleading (Seyoum & Kornfield, 1992; Vreven, Teugels, Adepo-Gourene, & Agnese, 1998). For example, Vreven et al. (1998) used morphometric analyses accompanied with allozyme markers and found that, contrary to findings from Trewavas (1983), 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 (1983). Other earlier genetic studies concerning *O. niloticus* in East Africa are reflected by Agnèse et al. (1997); Mwanja et al. (1996); Mwanja, Booton, Kaufman, and Fuerst (2008). Agnèse et al. (1997) employed allozymes and restriction fragment length

polymorphism (RFLP) of mitochondrial DNA (mtDNA) and Mwanja et al. (1996); Mwanja et al. (2008) randomly amplified polymorphic DNA (RAPD). Interestingly, Agnèse et al. (1997) 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 (1983). Furthermore, Agnèse et al. (1997) indicated that *O. niloticus* in Lake Tana is clustered together with Lake Edward and the Kenyan Lake Turkana system, which is contrary to the findings from Seyoum and Kornfield (1992). These findings are conflicting probably because of the different methodological approaches used, that comprised different information content (Guichoux et al., 2011; Hedrick, 2001; Liu & Cordes, 2004; Miah et al., 2013; Okumuş & Çiftci, 2003; Wirgin & Waldman, 1994). 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 (Wirgin & Waldman, 1994). The lack of methodologies with high discriminating power therefore suggests that the genetic structure patterns of the East African *O. niloticus* may be 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) (Tibihika, Curto, et al., 2018). Albeit recent studies using SSR have also focused on *O. niloticus* (Angienda et al., 2011; Ndiwa, Nyingi, & Agnese, 2014; Nyingi, De Vos, Aman, & Agnèse, 2009), 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 (Frankham, Briscoe, & Ballou, 2002). SSR-GBS approaches are useful in getting rid size

homoplasy, which is one of the constraints characterised by the traditional amplicon length procedures (Narum, Buerkle, Davey, Miller, & Hohenlohe, 2013; Vartia et al., 2016). Here, we explicitly investigate the genetic structure of *O. niloticus*, where we also compare the species to populations from Ethiopia and West Africa (Burkina Faso), representing the Sub-Saharan African Great Lakes. With this approach, the broad objective of our study was to investigate the impact of anthropogenic activities on the *O. niloticus*' gene pool. We hypothesized that the anthropogenic activities have contributed to the genetic divergent and differentiation of *O. niloticus*, particularly in environments where the species was introduced. We further test the hypothesis that, given interspecies translocations, the East African *O. niloticus* are genetically divergent from the ones outside the region (i.e., Ethiopia and Burkina Faso). 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?

Materials and Methods

Sampling/study areas

Following the earlier sampling design from Tibihika, Waidbacher, et al. (2018), 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. A total of 664 samples were collected from 18 waterbodies during several field excursions in the year, 2016. From each single fish, a muscle tissue (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 as 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) (Tibihika, Waidbacher, et al., 2018). 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).

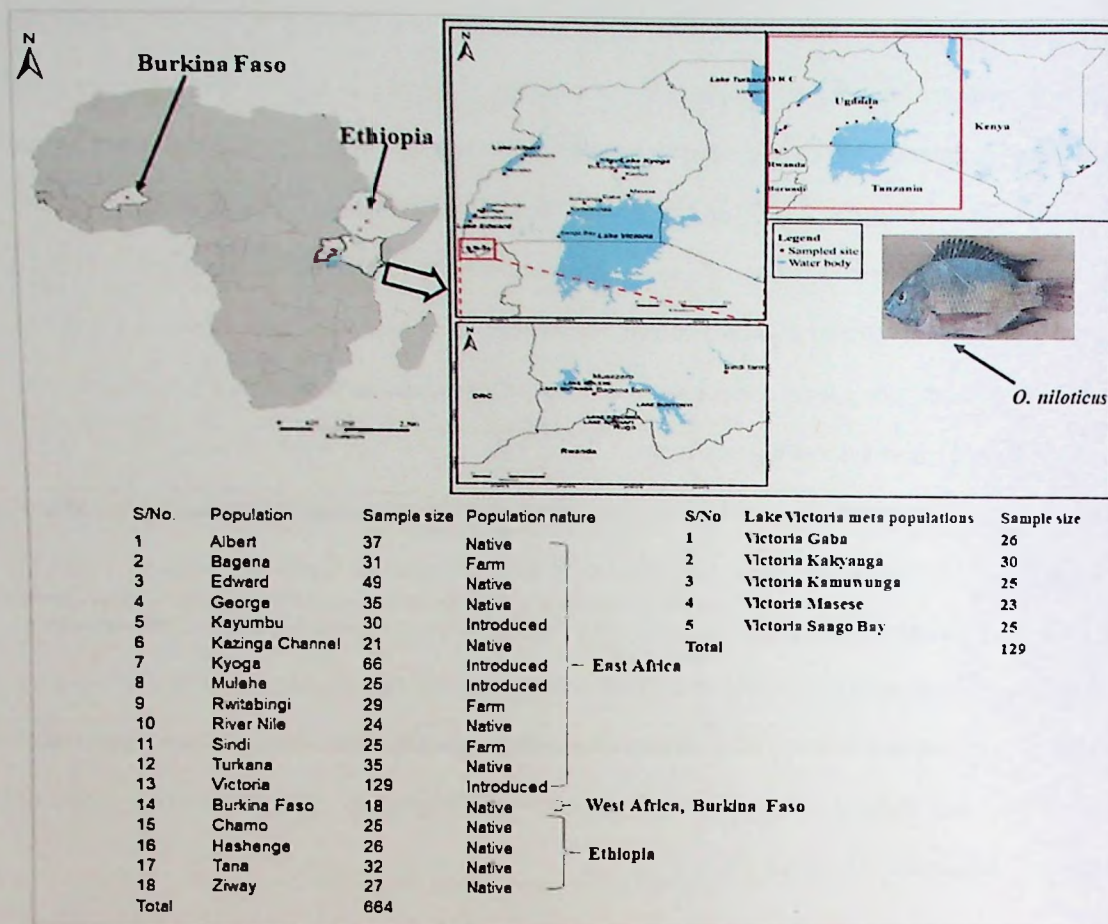


Figure 1 Sample sources and sizes. Natives are indigenous *O. niloticus* populations, non-natives are introduced, and farms are the pond culture systems. The top panel shows a map for the sampling locations with a focus on East Africa, Ethiopia and Burkina Faso. The bottom panel shows the details of the sampling sites and total number of individuals collected (Left) and Lake Victoria sub-population (right).

Genotyping

Genomic DNA extraction followed the modified MagSi-DNA Vegetal kit protocol, using magnetic beads (MagSi-DNA beads MagnaMedics) and a magnetic separator, SL-MagSep96 (Steinbrenner, Germany). For a detailed description, see Tibihika, Curto, et al. (2018). We used microsatellite markers (Tibihika, Curto, et al., 2018), to which we added 15 extra primers (Table 1). The SSR primers were designed and tested analogous to Tibihika, Curto, et al. (2018), using the same shot

gun sequencing data present in the short-read archive (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 (Tibihika, Curto, et al., 2018). 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. 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 $\geq 50\%$ were excluded, leaving a total number of 40 markers (supplementary materials, Table S1)

Table 1 15 new primer pairs developed in the present study. Other 26 tested primers developed by (Tibihika, Curto, et al., 2018) can be found in the supplementary materials 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	TGCAGCTGCTCCTGTTTAA	TTGTGCACGTGGACATGTTG	(AAAC)11	381-471
Ti43	ATTGCCATCACCAGGAACCA	TGCTAGCCCAGAGCATTGA	(GAATA)6	425-478
Ti44	TGCTCCTGACTCAGCATCAC	GCAGCACTCTGACATGAAGC	(GAAAA)6	419-469
Ti49	TCGAAGTAGCGTGGAACCT	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	TTTCTTGCCAGCAAAACAGT	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	TGAGCATTTGACCCAGCAT	(AGGA)6	429-461
Ti60	GAGCCGCCATAGTGCACTT	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 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 (Peakall & Smouse, 2006). Genetic similarity between populations was evaluated by plotting neighbor network based on Nei's genetic distance using program, SplitsTree4 version, 4.14.8 (Huson & Bryant, 2005). We also constructed UPGMA dendrograms for making inferences on the hierarchical clustering between populations using PAUP* version 4.0a (Swofford, 2002). Neighbor networks and UPGMA dendrogram were conducted with 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 (Hubisz, Falush, Stephens, & Pritchard, 2009). STRUCTURE clusters individuals into hypothetical populations through optimization of Hardy-Weinberg equilibrium (Porras-Hurtado et al., 2013). 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 (Evanno, Regnaut, & Goudet, 2005), 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 (Earl & vonHoldt, 2012) using the delta K (ΔK) statistic, which is the second order rate of change ($\ln P(D)$) across successive K values (Chaluvadi, Khanam, Aly, & Bennetzen, 2014; Evanno et al., 2005). 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 (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) 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 number of migrants per generation, were determined as proxy estimates of geneflow 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 (Wilson & Rannala, 2003). Here, the program was run for 200, 000,000 iterations, discarding the first 100,000,000 generations and sampling every 1000th generation (Curto, Puppo, Kratschmer, & Meimberg, 2017). Only results with a fraction of above 0.01 individual migrants between populations per generation at 95% confident interval were considered as vital migrants. Recent migration rates were used because most of the fish translocations in the region, ostensibly 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 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 (Excoffier & Lischer, 2010). Ho, He, Na and PIC per locus were determined through Cervus version 3.0.7 (Kalinowski, Taper, & Marshall, 2007). Ar was analyzed using the rare-fraction algorithm implemented in Hp-rare program (Kalinowski, 2005). 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 (Excoffier, Estoup, & Cornuet, 2005). 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 (Peakall & Smouse, 2006).

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 supplementary materials, 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 (H_e) values greater than 0.5. Polymorphic Information Content (PIC) was generally congruent to H_e , with 78% of loci indicating values of greater than 0.5 (supplementary materials 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. Over all, most of the nonnative populations (including farms) showed longer branches than the natives (Figure 2.2).

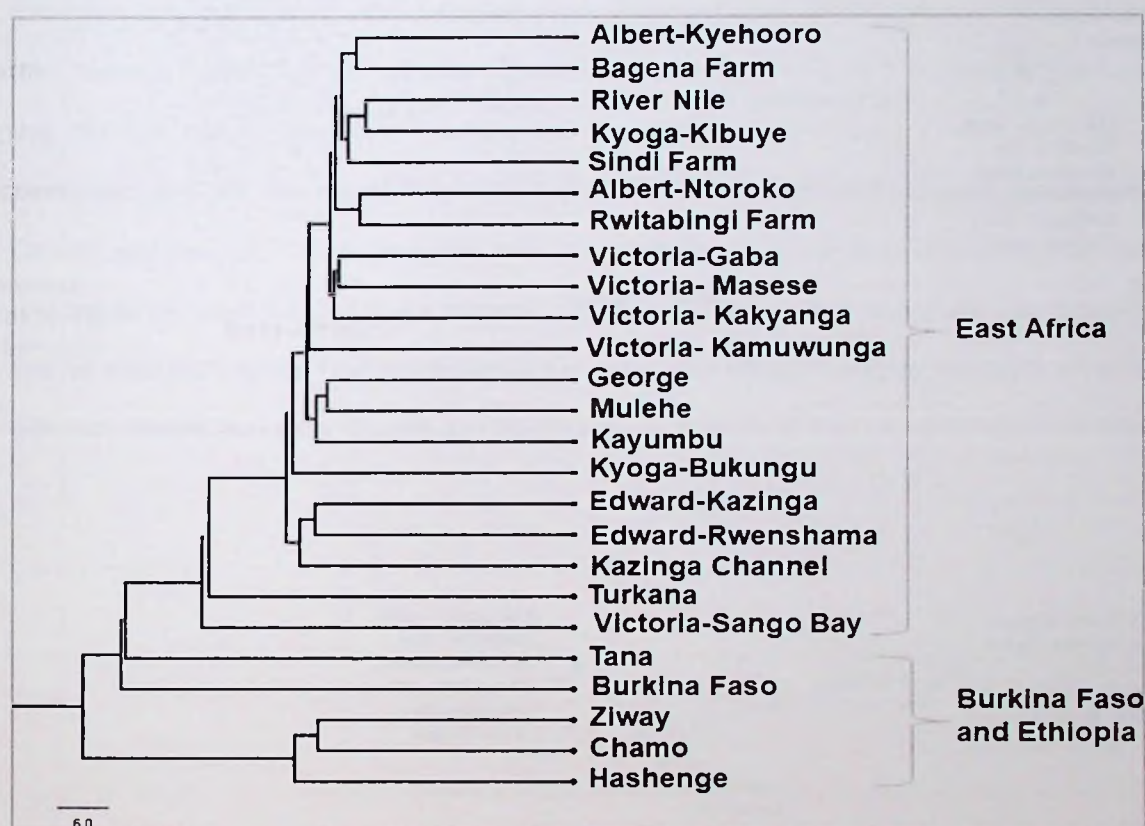


Figure 2.1 Genetic structure of *O. niloticus* populations based on UPGMA dendrogram

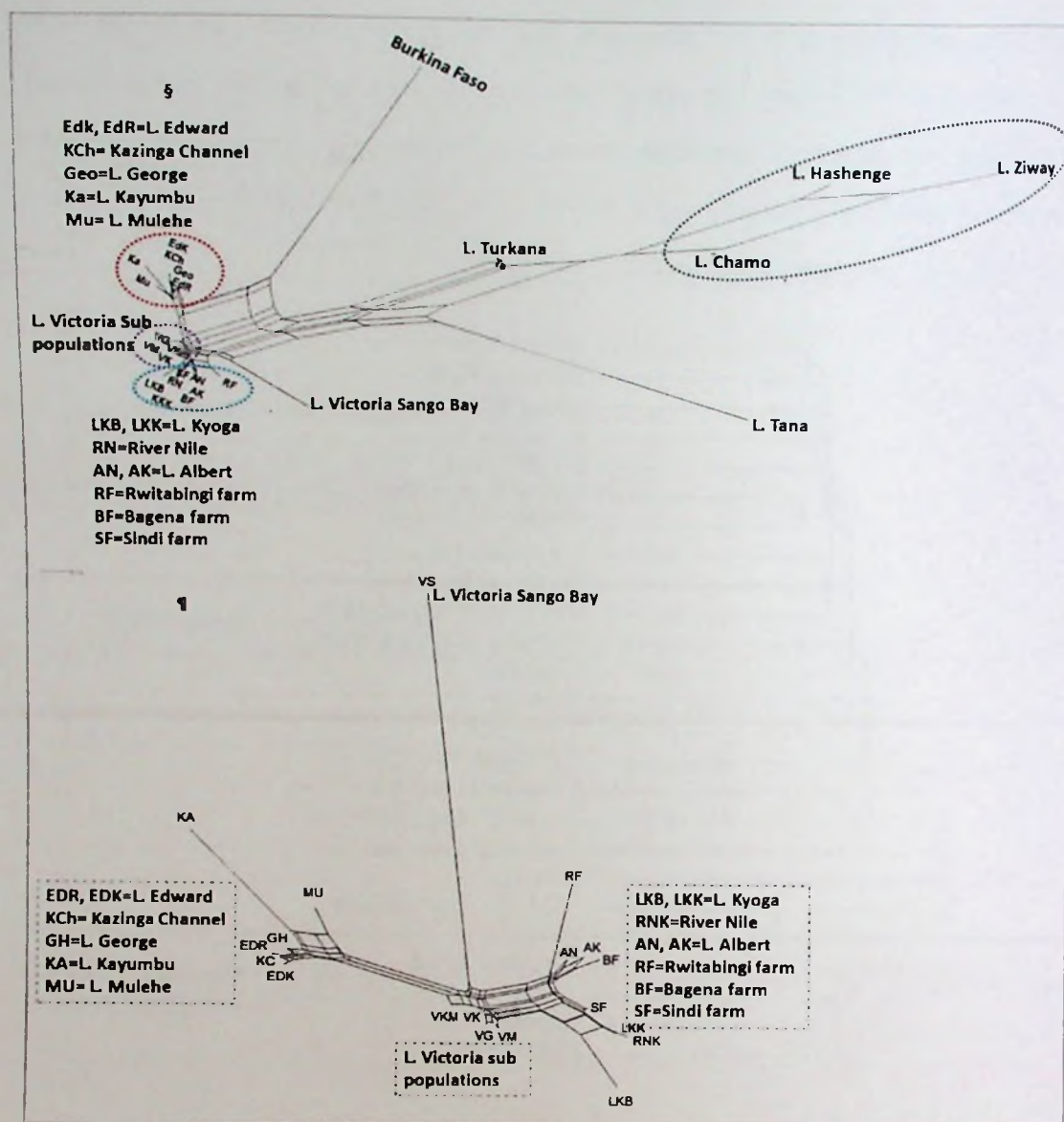
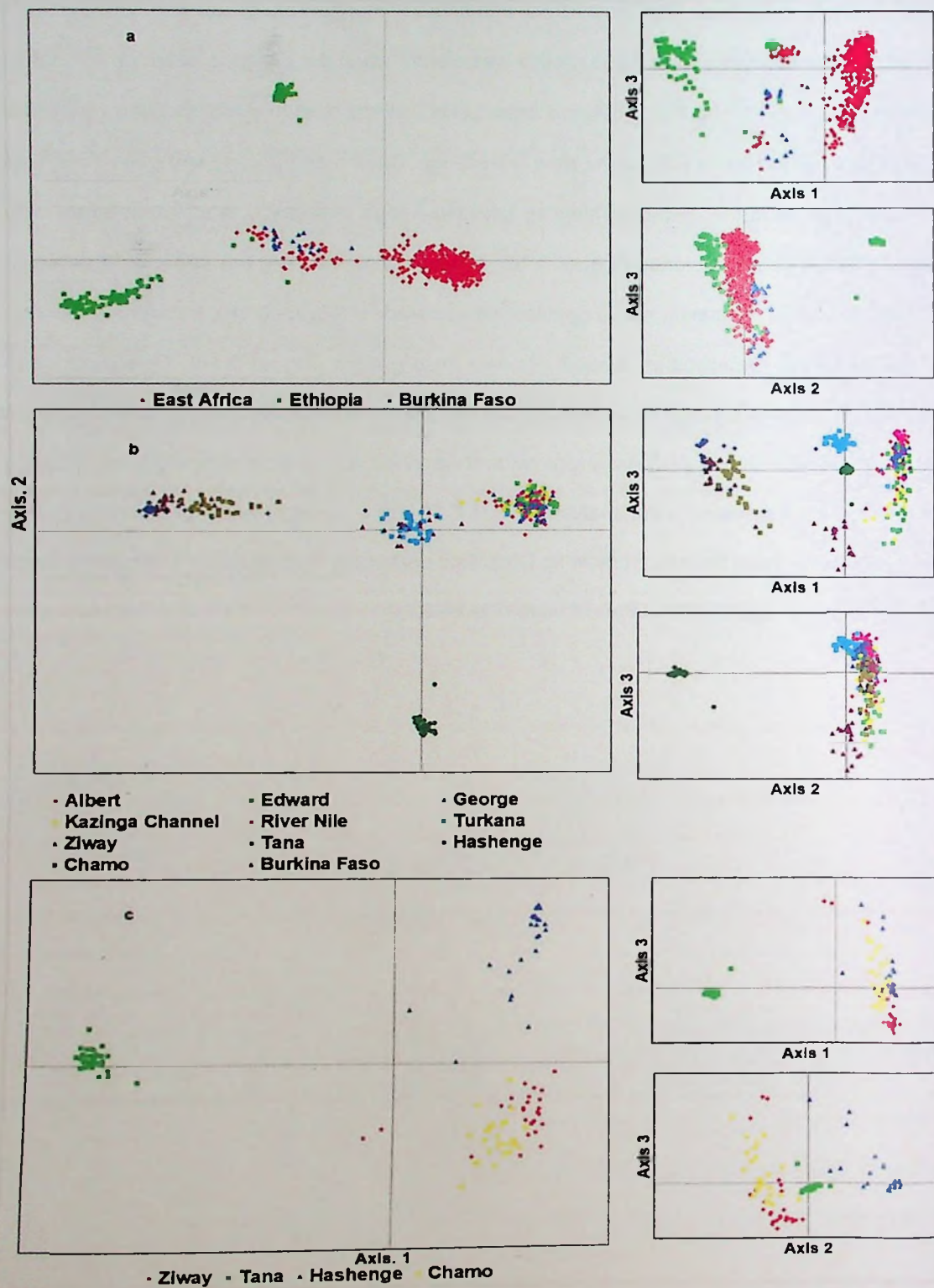


Figure 2.2 Genetic structure based on unrooted network tree illustrating population relationships based on distance. § represents a tree for all the populations and ¶ for only the Ugandan populations. Dotted oval and rectangular shapes depict closely related genetic groups.

Genetic distance between individuals 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. A further group containing lake Hashenge individuals was found when only Ethiopian individuals were plotted (Figure 3.1c). Amongst East African, the separation between lake Turkana and the remaining native populations was evident (Figure 3.1d). Individuals found in the Ugandan native waterbodies divided into two main groups (Figure 3.1e). 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 3.1f). Here, some individuals from Sango Bay formed a separate group from the remaining Ugandan individuals. Substructure within same lakes was only evident for Lakes Victoria and Kyoga (Figure 3.2).



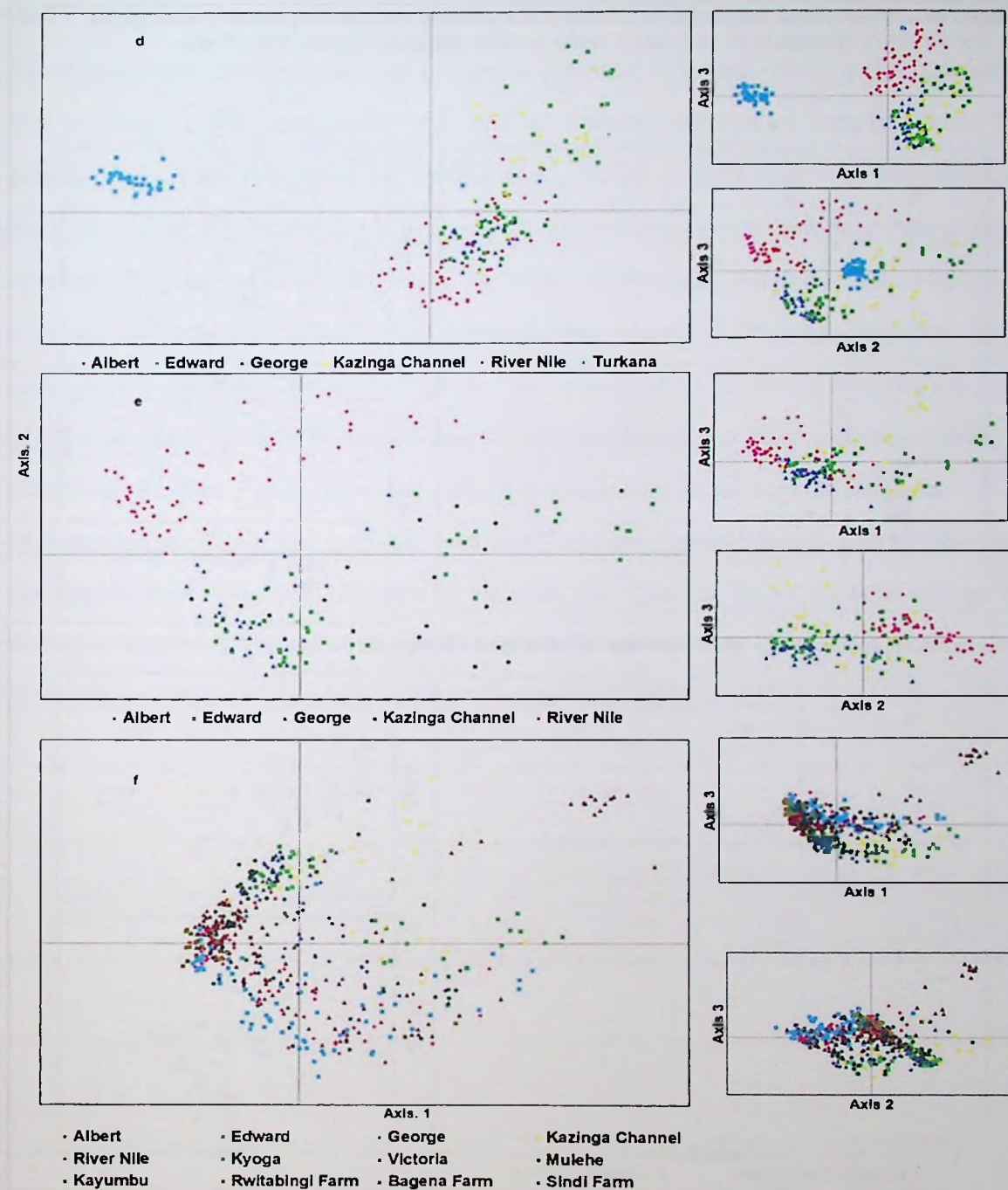


Figure 3.1 Genetic scatter plots of *O. niloticus* exhibited by Principal Coordinate Analysis (PCoA). a populations per region, b all native populations, c Ethiopian populations, d East African natives,

e Ugandan natives, and f Ugandan populations including non-native, natives and farms. PCoA was constructed in respect to un-biased Nei's genetic distance among individuals

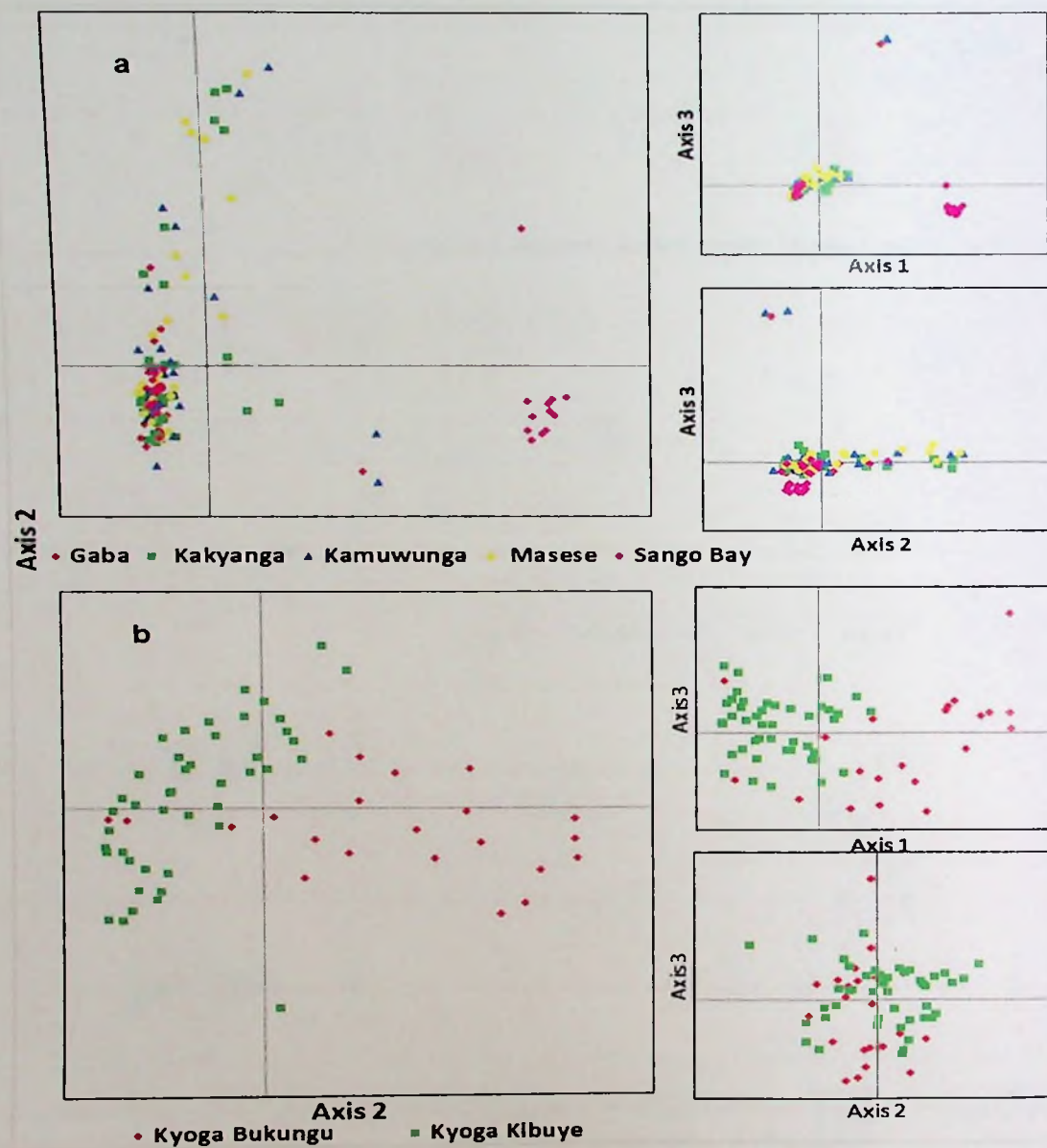


Figure 2.2 Genetic structure based on PCoA within Lakes Victoria (a) and Kyoga (b) populations

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 (supplementary materials, 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).

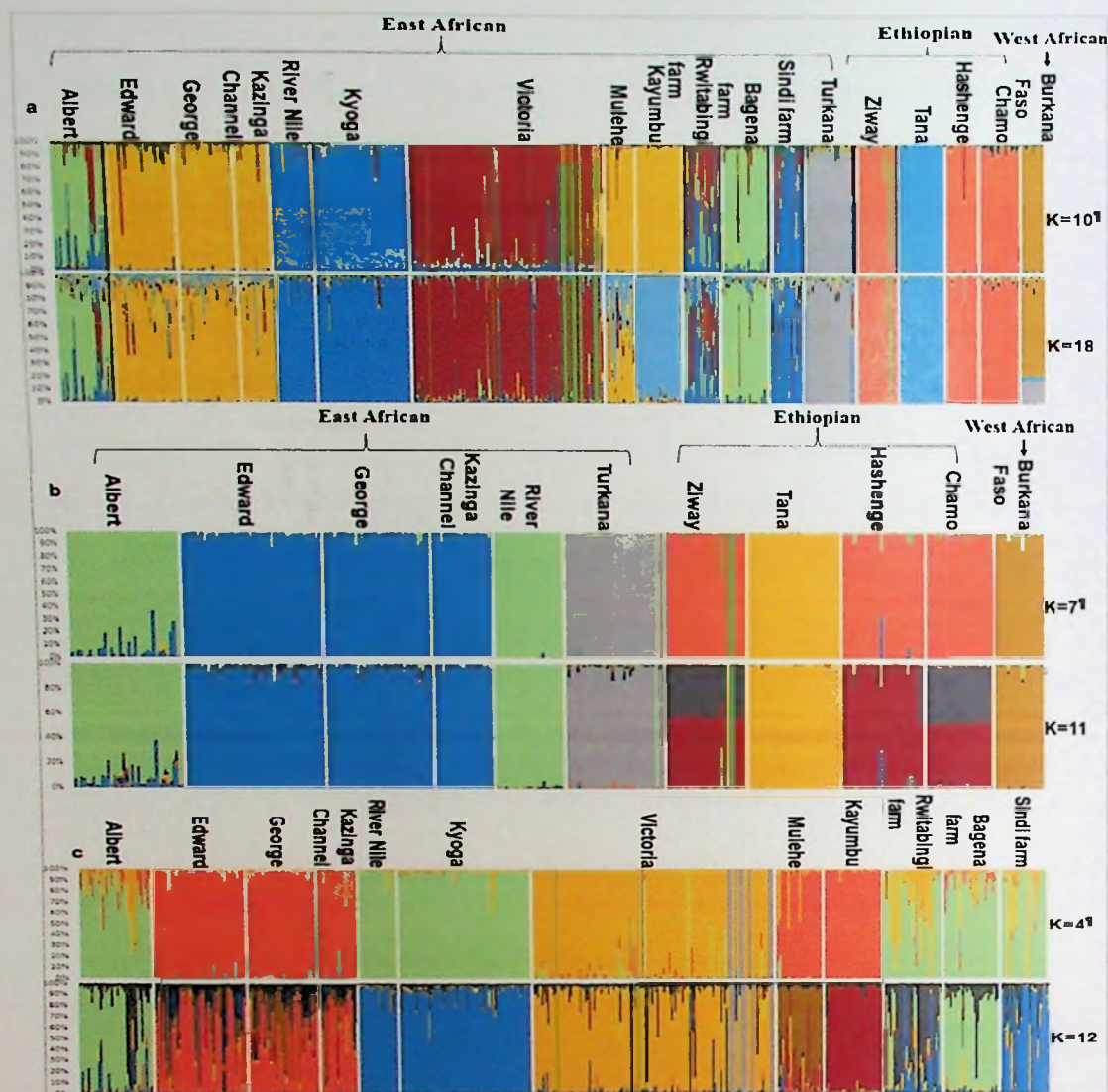


Figure 4 Bayesian clustering for genetic assignments of *O. niloticus* populations. **a** represents all populations, **b** all natives, and **c** all Ugandan populations including natives, non-natives and farms. Ks with a superscript symbol (†) indicates the optimal K value based on STRUCTURE HARVESTER analyses.

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 population 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 (Supplementary material Table S3).

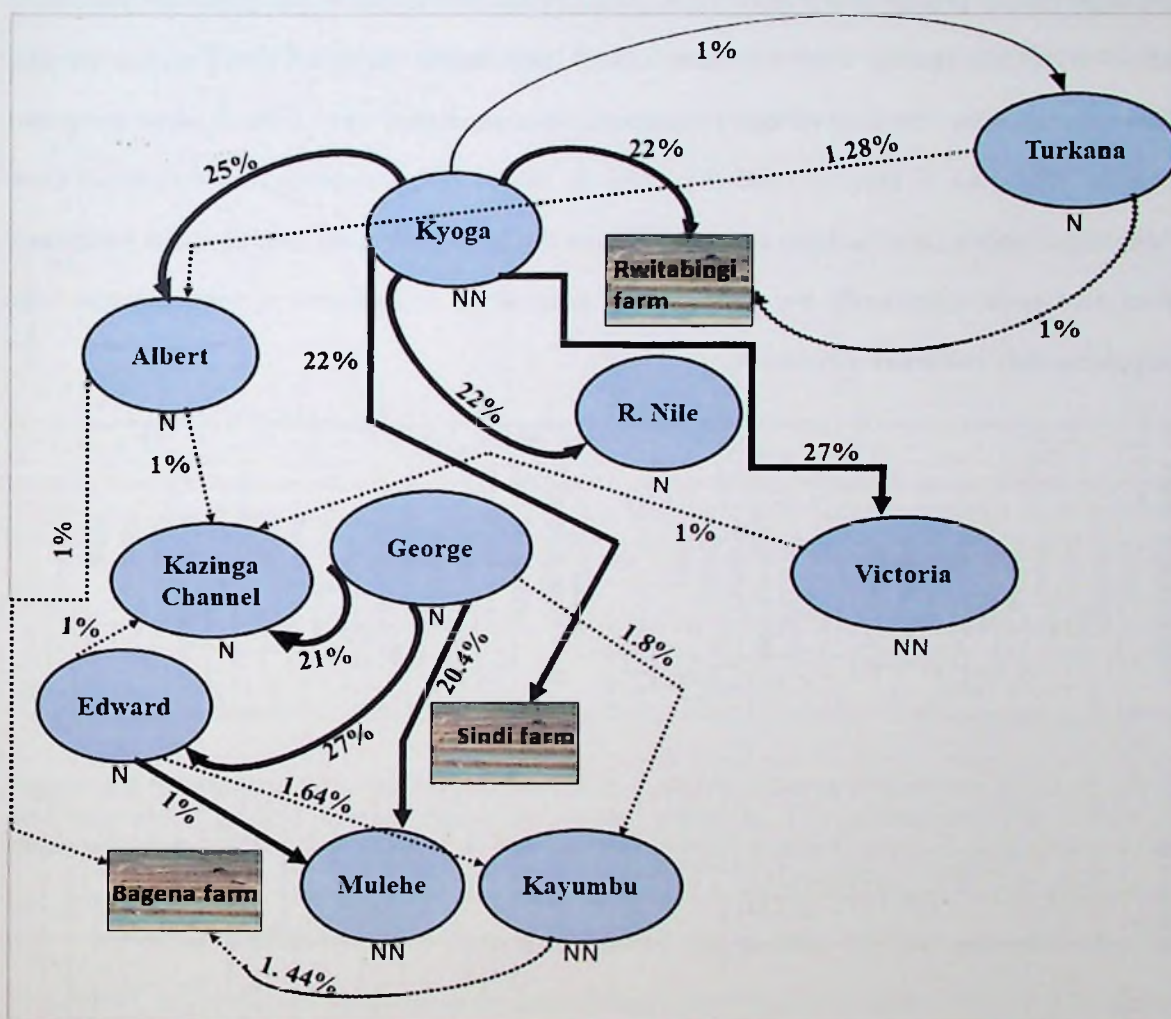


Figure 5 Bayesian inference of recent migratory rates for the 13 East African *O. niloticus* populations. Oval light blue and rectangular shapes indicate natural and farm populations, respectively. The arrows contain percentage values showing the direction and magnitude of gene flow. Darker and thick 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

Genetic differentiation, diversity and isolation by distance

Genetic differentiation was consistent with the STRUCTURE results. For instance, the F_{st} 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 F_{st} 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 (supplementary materials, Figure S4).

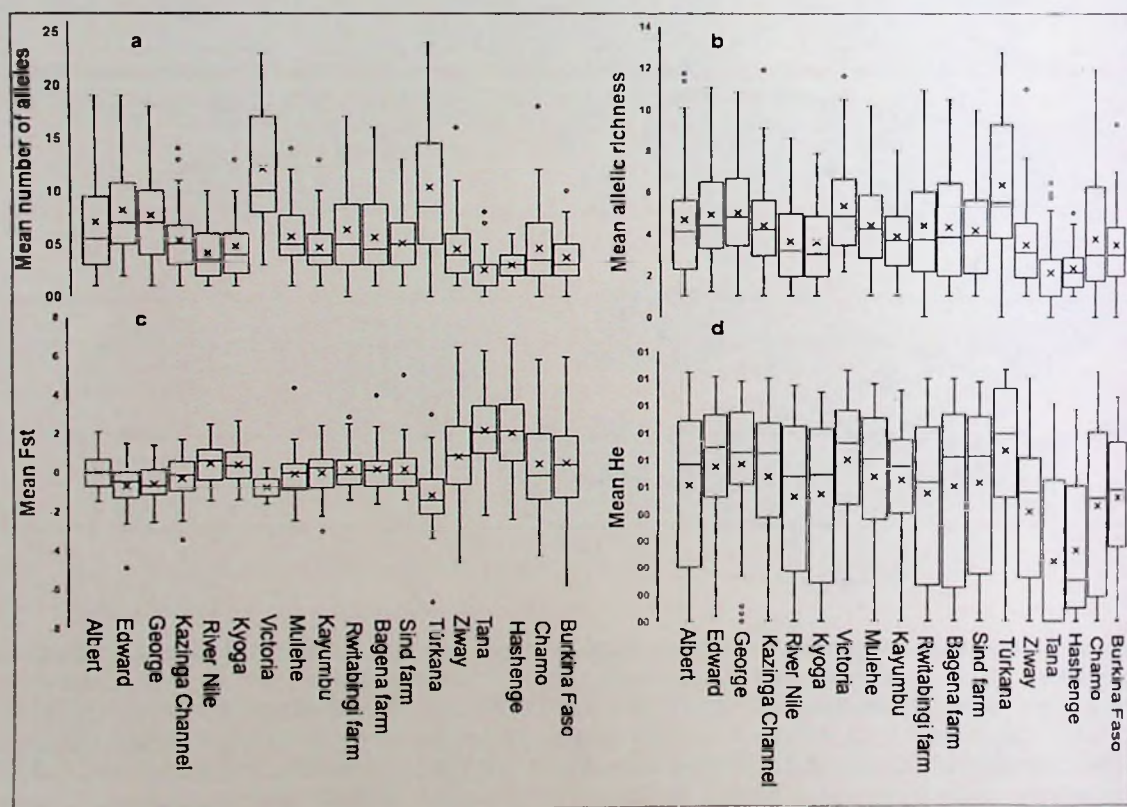


Figure 6.1 Genetic diversity and differentiation indices. **a** number of alleles, **b** allelic richness, **c** fixation index (F_{st}) and **d** expected heterozygosity. The genetic diversity indices are mean values over loci

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).

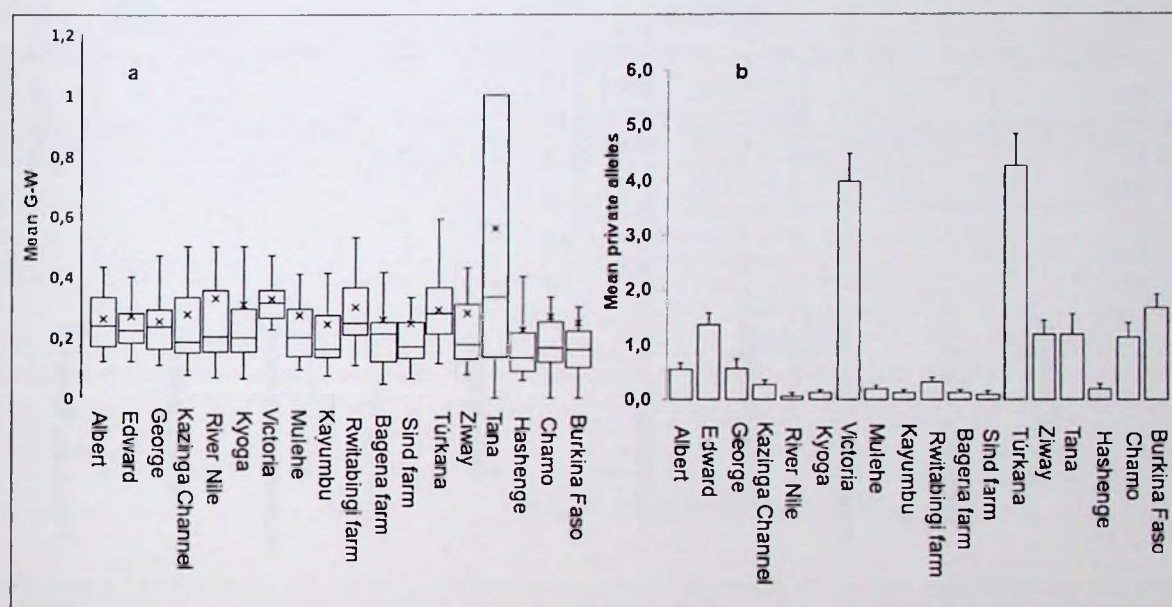


Figure 6.2 Estimations of population bottleneck derived from Garza-Williamson Index (G-W) (**a**) and measure of genetic diversity based on private alleles (**b**). The genetic diversity indices are mean values over loci

When we partitioned Lake Victoria to assess the genetic diversity patterns within the ecosystem, generally one sub-population was distinguished from the others (Figure 6.3). 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 6.3).

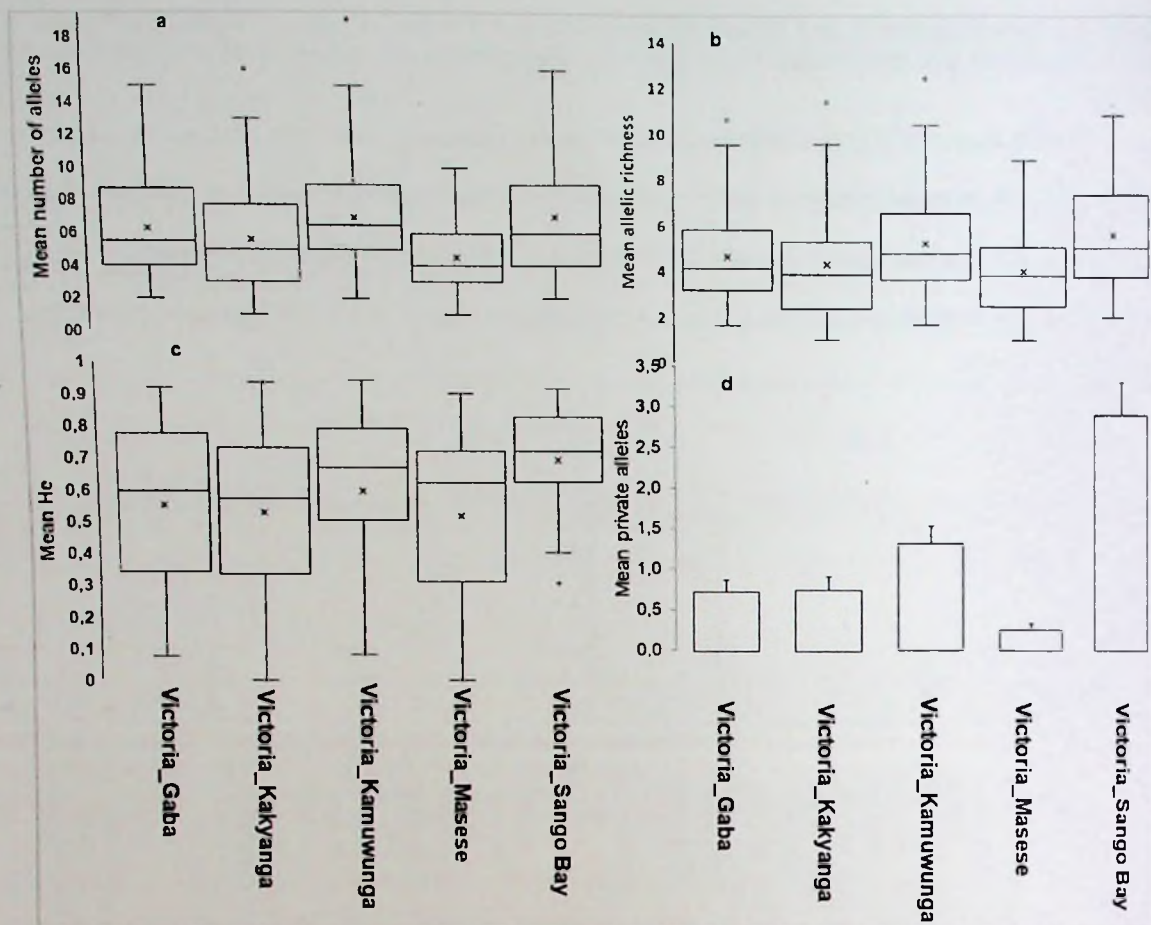


Figure 6.3 Genetic diversity of Lake Victoria within population. a number of alleles, b allelic richness c expected heterozygosity, and d private alleles

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‡).

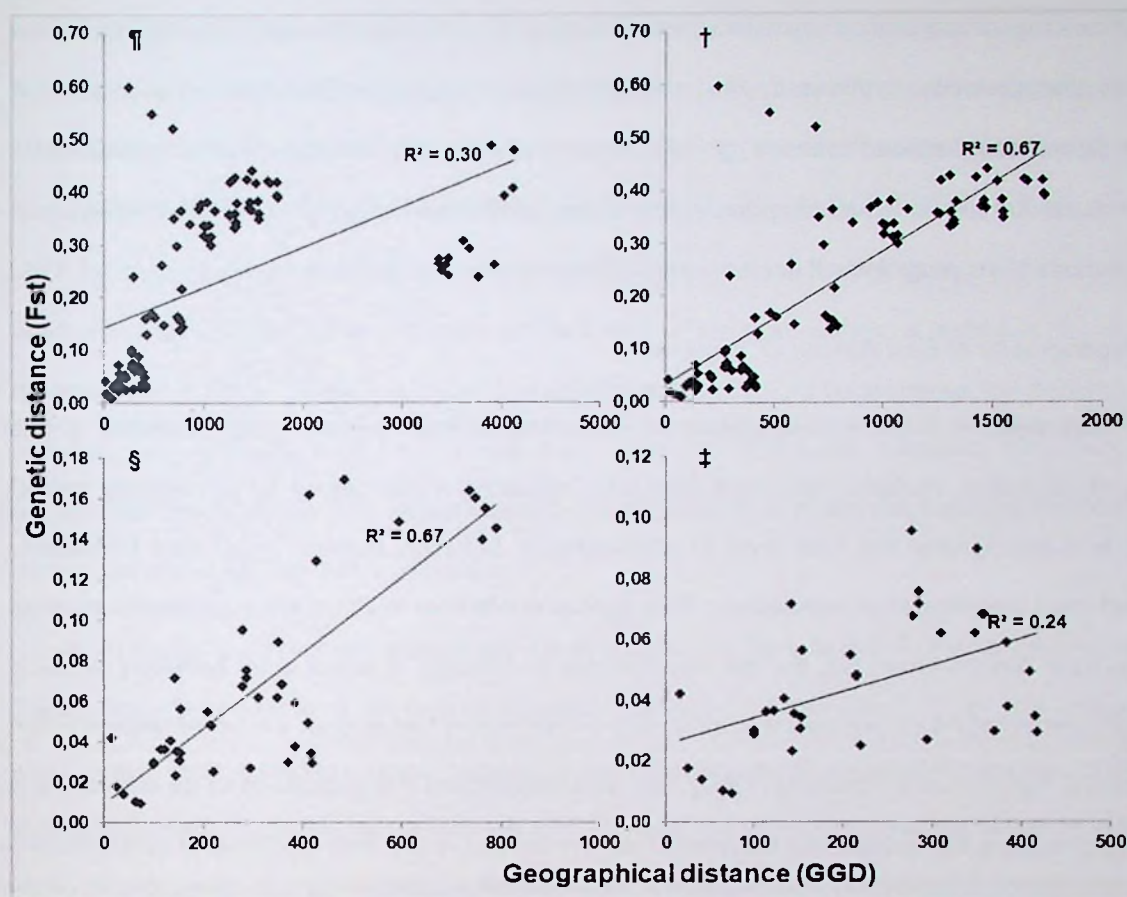


Figure 7 Mantel tests for correlations between genetic distance (Fst) and geographical distance (GGD in km) for *O. niloticus* populations. ¶ represents isolation by distance (IBD) between all populations, † all populations without Burkina Faso, § East African, and ‡ only Ugandan populations

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 (Welcomme et al., 2010). One aspect of the anthropogenic threats is the change or alteration of the natural genetic structure of fish stocks through admixture (Rhymer & Simberloff, 1996; Todesco et al., 2016). To understand 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 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 to identify 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 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 (1992) with allozymes found this population to be different and thus contradicting the subspecies treatment from Trewavas (1983). This high level of differentiation was confirmed in concurrence to past studies and so the species concept is suggested.

Lake Tana lies in the Ethiopian mountains and is isolated from the Lakes in the Rift valley (Vijverberg, Dejen, Getahun, & Nagelkerke, 2012). This might explain the high degree of differentiation of this lake because of 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 (Tsegay, 2017). 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.

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 (Odada et al., 2003). Trewavas (1983) 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 isolation. The fact that the East African arid Lake naturally does not seem to have any potential contact with other populations, it explains the current observed remarkable genetic diversity. One additional factor might be introgression perhaps from anthropogenic activities or influx of geneflow via 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, three main groups can be described: 1) the western Rift Valley water bodies (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 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. 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 (Dumont, 2009). The strong rapids and falls present in this river (Dumont, 2009; Ogutu-Ohwayo, 1990), 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 (Tibihika, Waidbacher, et al., 2018), but do not concur with past studies (Mwanja et al., 2008; Trewavas, 1983). This incongruity might be associated with different methodological approaches utilized between the earlier studies and the current. For example, Trewavas (1983) using morphometric and meristics treated *O. niloticus* from the Edward-George system and Albert as one subspecies (*O. niloticus eduardianus*). However, use of traditional morphometrics and counts have been

strongly labelled as weak in identifying species due to the fact that the characters involved in the method overlap among many subspecies (Seyoum & Kornfield, 1992). Similarly, while we used SSR-GBS techniques, Mwanja et al. (2008) employed random amplified polymorphic DNA (RAPD) markers, which due to their dominant nature reduce result reliability (Wirgin & Waldman, 1994). For the 2nd and 3rd Uganda groups, see the next section.

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 (Western Rift Valley and Lake Kyoga) contributed to stocking of different water bodies. Apparently, *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 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 waterflow. The main reason is the natural occurrence of Murchison Falls on River Nile that acts as a barrier between the systems (Dumont, 2009; Ogutu-Ohwayo, 1990). For this matter, the genetic similarity between River Nile, Lakes Kyoga and Albert may translate to the descendants from the latter waterbody during stocking regimes as reported by Ogutu-Ohwayo (1990).

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 geneflow from Kyoga to Rwitabingi and all these farms showed to be admixed with other populations including Lake Victoria clusters. Apart from farms, evidences of admixture

were 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 tree, 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 (1983) suggests that introductions into Lake Victoria may have originated from Lake Edward, with other authors suggesting multiple sources (Baliwa, 1992; Fuerst, Mwanja, & Kaufman, 2000; Kaufman, 1992; Pullin, 1988), 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 a 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. The fact that, the indigenous congeners were extirpated following the explosion of *O. niloticus* in Lake Victoria (Bairwa, 1992; Bairwa et al., 2003; Kaufman, 1992; Ogutu-Ohwayo, 1990), the observed genetic structure of the cichlid may have undergone genetic swamping via introgressive hybridization. However, in this study we cannot directly test for hybridization since we did not include samples of *O. niloticus* congeners. However, hybridization involving *O. niloticus* and other tilapiines has been reported to be relatively frequent and it needs to be considered (Angienda et al., 2011; Deines, Bbole, Katongo, Feder, & Lodge, 2014; Shechonge et al., 2018; Welcomme, 1966).

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 (Benvenuto et al., 2012; Dodson et al., 2015), admixture is usually reported to have negative consequences (Allendorf, Leary, Spruell, & Wenburg, 2001; Rhymer & Simberloff, 1996). In more drastic scenarios, hybridization can result in genomic incompatibilities contributing to a fast reduction of population fitness (Roberts, Gray, West, & Ayre, 2010). Alternatively, the hybrids may potentially exhibit more fitness and subsequently extirpate the parental lines (Deines et al., 2014). 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.

Evidences for the second process of admixture were 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 in respect to translocations do 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 evidences of bottleneck according to the G-W estimations. In a previous study Mwanja et al. (2008) 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 (Allendorf, England, Luikart, Ritchie, & Ryman, 2008). This might be the case for these populations, 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 analysis, especially when additional sampling is conducted in sections of the lower Nile. In Lake Kyoga we detected population subdivisions and so restrictions to gene-flow in the whole system might encourage founder effects which may ultimately lead to genetic erosion. Although nearly all the populations equally indicated to be bottlenecked, the high to moderate genetic diversity exhibited in the other East African populations might be an indicator of recent population expansion or linked to earlier stages of genetic erosion.

Implications for management and outlook

Overall, we found evidences 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 a long term this may have negative effects on populations fitness due to outbreeding depression and genetic swamping. Thus, management measures should inhibit any form spread of fish in the aquatic ecosystems. The Western or Albertine Rift Valley lakes (Edward-George) may be ideal brood stock sources for subsequent breeding programs and aquaculture, as these systems seem un admixed from external sources. To avoid influx of feral populations and other infrastructure developments, proper environmental impact assessment should be prioritized before implementation. We also found evidences 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 genetic structure in populations from Lake Tana, Victoria and one genetic cluster in the western Rift Valley, that may require further exploration for taxonomic recognition. Moreover, this study shows the importance of molecular genetic markers in particular use of SSR-GBS in cataloguing 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 congeners for a more comprehensive picture.

Acknowledgements

We are grateful to funding from 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). We are also 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.

Author contributions

P. D. T and partly with E. A. conducted field excursions. H. M provided laboratory space/analytical tools and research focus. P. D. T., M. C and H. M were involved in research design, laboratory work, bioinformatics and data analysis. P. D. T wrote the original draft of the manuscript, which was further implemented by M. C and H. M. H. W and C. M made substantial manuscript reviews. All co-authors read and approved the final version of the manuscript. The authors declare that they have no conflict of interest.

References

- Agnèse, J.-F., Adépo-Gourène, B., Abban, E. K., & Fermon, Y. (1997). Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity*, 79(1), 88-96.
- Agnèse, J.-F., Adépo-Gourène, B., & Pouyaud, L. (1998). Natural hybridization in tilapias. *Genetics and Aquaculture in Africa*, 95-105.
- Allendorf, F. W., England, P. R., Luikart, G., Ritchie, P. A., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in ecology & evolution*, 23(6), 327-337. doi:https://doi.org/10.1016/j.tree.2008.02.008
- Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: setting conservation guidelines. *Trends in ecology & evolution*, 16(11), 613-622.
- Allendorf, F. W., & Luikart, G. (2007). Conservation and the genetics of populations. *mammalia*, 2007, 189-197. doi:10.1515/MAMM.2007.038
- Angienda, P. O., Lee, H. J., Elmer, K. R., Abila, R., Waindi, E. N., & Meyer, A. (2011). Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. *Conservation Genetics*, 12(1), 243-255. doi:https://doi.org/10.1007/s10592-010-0136-2
- Baliwa, J. S. (1992). The evolution of the fishery of *Oreochromis niloticus* (Pisces: Cichlidae) in Lake Victoria. *Hydrobiologia*, 232(1), 85-89. doi:https://doi.org/10.1007/BF00014616
- Baliwa, J. S., Chapman, C. A., Chapman, L. J., Cowx, I. G., Geheb, K., Kaufman, L., . . . Welcomme, R. L. (2003). Biodiversity and fishery sustainability in the Lake Victoria basin: an unexpected marriage? *BioScience*, 53(8), 703-715. doi:https://doi.org/10.1641/0006-3568(2003)053[0703:BAFSIT]2.0.CO;2
- Benvenuto, C., Tabone, E., Vercken, E., Sorbier, N., Colombel, E., Warot, S., . . . Ris, N. (2012). Intraspecific variability in the parasitoid wasp *Trichogramma chilonis*: can we predict the outcome of hybridization? *Evolutionary applications*, 5(5), 498-510.
- Canonico, G. C., Arthington, A., McCrary, J. K., & Thieme, M. L. (2005). The effects of introduced tilapias on native biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 15(5), 463-483. doi:https://doi.org/10.1002/aqc.699
- Chaluvadi, S. R., Khanam, S., Aly, M. A., & Bennetzen, J. L. (2014). Genetic diversity and population structure of native and introduced date palm (*Phoenix dactylifera*) germplasm in the United Arab Emirates. *Tropical plant biology*, 7(1), 30-41. doi:https://doi.org/10.1007/s12042-014-9135-7
- Curto, M., Puppo, P., Kratschmer, S., & Meimberg, H. (2017). 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*, 17(198), 1-16. doi:DOI 10.1186/s12862-017-1031-y
- Deines, A., Bbole, I., Katongo, C., Feder, J., & Lodge, D. (2014). Hybridisation between native *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River, Zambia. *African Journal of Aquatic Science*, 39(1), 23-34. doi:https://doi.org/10.2989/16085914.2013.864965
- Dodson, J. J., Bourret, A., Barrette, M. F., Turgeon, J., Daigle, G., Legault, M., & Lecomte, F. (2015). Intraspecific genetic admixture and the morphological diversification of an estuarine fish population complex. *PloS one*, 10(4), e0123172. doi:https://doi.org/10.1371/journal.pone.0123172
- Dumont, H. J. (2009). *The Nile: origin, environments, limnology and human use* (Vol. 89): Springer Science & Business Media.
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, 4(2), 359-361. doi:org/10.1007/s12686-011-9548-7

- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., Estoup, A., & Cornuet, J.-M. (2005). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, 169(3), 1727-1738. doi:org/10.1534/genetics.104.036236
- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567. doi:org/10.1111/j.1755-0998.2010.02847.x
- Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). *Introduction to conservation genetics*: Cambridge university press.
- Fuerst, P., Mwanja, W., & Kaufman, L. (2000). *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*. Paper presented at the Tilapia Aquaculture in the 21st Century. Proceedings from the Fifth International Symposium on Tilapia in Aquaculture.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., . . . Salin, F. (2011). Current trends in microsatellite genotyping. *Molecular ecology resources*, 11(4), 591-611. doi:https://doi.org/10.1111/j.1755-0998.2011.03014.x
- Hedrick, P. W. (2001). Conservation genetics: where are we now? *Trends in Ecology and Evolution*, 16(11), 629-636. doi:10.1016/S0169-5347(01)02282-0
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population Structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332. doi:10.1111/j.1755-0998.2009.02591.x
- Huson, D. H., & Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23(2), 254-267. doi:https://doi.org/10.1093/molbev/msj030
- Kalinowski, S. T. (2005). Hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Resources*, 5(1), 187-189.
- Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16(5), 1099-1106. doi:10.1111/j.1365-294X.2007.03089.x
- Kaufman, L. (1992). Catastrophic change in species-rich freshwater ecosystems. *BioScience*, 42(11), 846-858.
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179-1191. doi:10.1111/1755-0998.12387
- Liu, Z., & Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238(1), 1-37. doi:10.1016/j.aquaculture.2004.05.027
- Miah, G., Rafii, M. Y., Ismail, M. R., Puteh, A. B., Rahim, H. A., Islam, K. N., & Latif, M. A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International journal of molecular sciences*, 14(11), 22499-22528. doi:10.3390/ijms141122499
- Mwanja, W. W. (2000). *Genetic biodiversity and evolution of two principal fisheries species groups, the Labeine and Tilapiine, of Lake Victoria Region, East Africa*. The Ohio State University.
- Mwanja, W. W., Booton, G., Kaufman, L., Chandler, M., Fuerst, P., Donaldson, E., & MacKinlay, D. (1996). *Population and stock characterization of Lake Victoria. Tilapine fishes based on RAPD markers*. Paper presented at the Aquaculture Biotechnology Symposium Proceedings of the International Congress on the Biology of Fishes (ed, by EM Donaldson & DD Mackinlay).

- Mwanja, W. W., Booton, G. C., Kaufman, L., & Fuerst, P. A. (2008). 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*, 7(11). doi:<http://dx.doi.org/10.5897/AJB08.813>
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular ecology*, 22(11), 2841-2847. doi:<https://doi.org/10.1111/mec.12350>
- Ndiwa, T. C., Nyingi, D. W., & Agnese, J.-F. (2014). An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Lobo drainage, Kenya. *PLoS One*, 9(9), 1-15.
- Nunan, F., Hara, M., & Onyango, P. (2015). Institutions and co-management in east African inland and Malawi fisheries: A critical perspective. *World Development*, 70, 203-214. doi:<https://doi.org/10.1016/j.worlddev.2015.01.009>
- Nyingi, D., De Vos, L., Aman, R., & Agnès, J.-F. (2009). 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*, 297(1-4), 57-63. doi:<https://doi.org/10.1016/j.aquaculture.2009.09.017>
- Odada, E. O., Olago, D. O., Bugenyi, F., Kulindwa, K., Karimurumungu, J., West, K., . . . Achola, P. (2003). Environmental assessment of the East African rift valley lakes. *Aquatic sciences*, 65(3), 254-271. doi:<https://doi.org/10.1007/s00027-003-0638-9>
- Ogutu-Ohwayo, R. (1990). The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of introduced species, especially the Nile perch, *Lates niloticus*, and the Nile tilapia, *Oreochromis niloticus*. *Environmental Biology of Fishes*, 27(2), 81-96. doi:<https://doi.org/10.1007/BF00001938>
- Ogutu-Ohwayo, R. (1990). The reduction in fish species diversity in Lakes Victoria and Kyoga (East Africa) following human exploitation and introduction of non-native fishes. *Journal of Fish Biology*, 37, 207-208. doi:<https://doi.org/10.1111/j.1095-8649.1990.tb05040.x>
- Okumuş, İ., & Çiftçi, Y. (2003). Fish population genetics and molecular markers: II-molecular markers and their applications in fisheries and aquaculture. *Turkish Journal of Fisheries and Aquatic Sciences*, 3(1), 51-79.
- Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources*, 6(1), 288-295. doi:[10.1111/j.1471-8286.2005.01155.x](https://doi.org/10.1111/j.1471-8286.2005.01155.x)
- Porrás-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in genetics*, 4. doi:[10.3389/fgene.2013.00098](https://doi.org/10.3389/fgene.2013.00098)
- Pullin, R. S. (1988). Tilapia genetic resources for aquaculture. Proceedings of the Workshop on Tilapia Genetic Resources for Aquaculture, 23-24 March 1987. *The WorldFish Center Working Papers*.
- Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual review of ecology and systematics*, 27(1), 83-109.
- Roberts, D. G., Gray, C. A., West, R. J., & Ayre, D. J. (2010). Marine genetic swamping: hybrids replace an obligately estuarine fish. *Molecular ecology*, 19(3), 508-520. doi:<https://doi.org/10.1111/j.1365-294X.2009.04501.x>
- Rothuis, A., Turenhout, M., van Duijn, A., Roem, A., Rurangwa, E., Katunzi, E., . . . Kabagambe, J. B. (2014). *Aquaculture in East Africa: a regional approach*. Retrieved from
- Seyoum, S., & Kornfield, I. (1992). Identification of the subspecies of *Oreochromis niloticus* (Pisces: Cichlidae) using restriction endonuclease analysis of mitochondrial DNA. *Aquaculture*, 102(1-2), 29-42.

- Shechonge, A., Ngatunga, B. P., Tamatamah, R., Bradbeer, S. J., Harrington, J., Ford, A. G., . . . Genner, M. J. (2018). Losing cichlid fish biodiversity: genetic and morphological homogenization of tilapia following colonization by introduced species. *Conservation Genetics*, 19(5), 1199-1209. doi:<https://doi.org/10.1007/s10592-018-1088-1>
- Swofford, D. L. (2002). Phylogenetic analysis using parsimony (* and other methods) PAUP* Version 4.0a. : Sinauer Associates, Sunderland, Massachusetts.
- Tibihika, P. D., Curto, M., Dornstauder-Schrammel, E., Winter, S., Alemayehu, E., Waidbacher, H., & Meimberg, H. (2018). Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*, 1-16. doi:<https://doi.org/10.1007/s10592-018-1136-x>
- Tibihika, P. D., Waidbacher, H., Masembe, C., Curto, M., Sabatino, S., Alemayehu, E., . . . Meimberg, H. (2018). Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa. *Environmental Biology of Fishes*, 101(3), 363-381. doi:<https://doi.org/10.1007/s10641-017-0704-0>
- Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., . . . Bock, D. G. (2016). Hybridization and extinction. *Evolutionary Applications*, 9(7), 892-908. doi:<https://doi.org/10.1111/eva.12367>
- Trewavas, E. (1983). *Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia*. London: British Museum (Natural History).
- Tsegay, T. (2017). Characteristics and Status of the Highland Lake Hashenge Fishery, Tigray, Northern Ethiopia. *International Journal of Aquaculture and Fishery Sciences*, 3(4), 71-76. doi:<https://doi.org/10.17352/2455-8400.000032>
- Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., . . . Cross, T. F. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society open science*, 3(1), 150565. doi:<https://doi.org/10.1098/rsos.150565>
- Vijverberg, J., Dejen, E., Getahun, A., & Nagelkerke, L. A. (2012). The composition of fish communities of nine Ethiopian lakes along a north-south gradient: threats and possible solutions. *Animal Biology*, 62(3), 315-335. doi:<https://doi.org/10.1163/157075611X618246>
- Vreven, E., Teugels, G., Adepo-Gourene, B., & Agnese, J. (1998). Morphometric and allozyme variation in natural populations and cultured strains of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Belgian Journal of Zoology (Belgium)*, 128(1), 23.
- Welcomme, R. (1966). Recent changes in the stocks of Tilapia in Lake Victoria. *Nature*, 212(5057), 52. doi:<https://doi.org/10.1038/212052a0>
- Welcomme, R., Cowx, I. G., Coates, D., Béné, C., Funge-Smith, S., Halls, A., & Lorenzen, K. (2010). Inland capture fisheries. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554), 2881-2896. doi:<https://doi.org/10.1098/rstb.2010.0168>
- Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163(3), 1177-1191.
- Wirgin, I. I., & Waldman, J. R. (1994). What DNA can do for you. *Fisheries*, 19(7), 16-27. doi:[https://doi.org/10.1577/1548-8446\(1994\)019<0016:WDCDFY>2.0.CO;2](https://doi.org/10.1577/1548-8446(1994)019<0016:WDCDFY>2.0.CO;2)
- Worthington, E. B. (1932). A report on the fisheries of Uganda investigated by the Cambridge Expedition to the East African Lakes, 1930-31.
- Zengeya, T. A., Booth, A. J., & Chimimba, C. T. (2015). Broad niche overlap between invasive Nile tilapia *Oreochromis niloticus* and indigenous congeners in southern Africa: Should we be concerned? *Entropy*, 17(7), 4959-4973. doi:10.3390/e17074959

5. Chapter 5

5.1 General discussion

Advanced geometric morphometrics was used in combination with microsatellite genotyping by sequencing (SSR-GBS), under next generation sequencing (NGS) based techniques, for exploring the phenotypic and genetic structure of Nile tilapia in East Africa. The studied species was particularly of interest as it inhabits different environments, many of which are heavily influenced by various anthropogenic activities. Principally, focus was directed on three environmental types occupied by the species; native, non-native and farms from East Africa, with the inclusion of the Ethiopian and Burkina Faso populations.

Sustainable exploitation of fish species is a global concern and maintaining biodiversity is vital for conservation. Nonetheless, the fisheries subsector in the developing world are of more worry as they are immensely negatively impacted on due to poor law enforcements and weak regulatory policies. Additionally, the continued decline of fish stocks is due to inclusion of management options and by-laws which were developed by colonialists and are outdated in the modern fisheries (Ogutu-Ohwayo & Baliwa, 2006). The multiple influx of anthropogenic activities in the East African Great Lakes presents a good example of fish dilemma in the region, despite the subsector being a vital source of cheap proteins to the rapidly shooting riparian human populations (Lowe-McConnell, 2009). The open-access-fisheries policy in the sub-Saharan region, triggers a leeway for resource over utilization or mis-use. Although Nile tilapia is currently commercially exploited in all over East African countries, the anthropogenic activities (fish translocation, habitat destruction, overfishing, pollution etc.) involved with this species pose a big threat for the species' sustainability. Given that overfishing may be regulated through co-management and enforcement of by-laws (Nunan, Hara, & Onyango, 2015), the problem of mixed species may be complex and difficult to manage/conserves for future generations. Even though restocking can be an important intervention to manage and reinforce overfished stocks or improve the naturally unproductive/fishless water bodies, the ultimate results from this intervention may hinder the evolutionary genetic integrity of the indigenous populations via admixture and or competition (McClure et al., 2008; Valiquette, Perrier, Thibault, & Bematchez, 2014). The East African *O. niloticus* presents a paradigm of threatened species underlying anthropogenic induced activities, accrued from translocations. For instance, the stocking of Nile tilapia and Nile perch into lakes Victoria, Nabugabo, Kyoga and additionally for the former species into other numerous minor lakes (Ogutu-Ohwayo, 1990). Generally, the stocking of Nile tilapia and Nile perch into Lake Victoria and Kyoga improved greatly fish productivity, in particular, the large marketable table sized fish, which used not to be the case

with the abundant small-sized haplochromines (Kaufman, 1992). However, the transfer of non-native species were at a cost of the indigenous species' disappearance through mutual biotic interactions (Ogutu-Ohwayo, 1990; Balirwa, 1992; Balirwa et al, 2003). Apparently, the genetic structure of Nile tilapia is likely to have been affected following introductions and subsequent admixture with the congeners. In this general discussion, I elucidate the over-all demography of the East African Nile tilapia and point-out recommendations for relevant future work.

5.1.1 Morphological divergence and local adaptation of the East African Nile tilapia (*Oreochromis niloticus*, L. 1758).

In chapter 2, geometric morphometric analysis was applied to study the shape variation of the East African Nile tilapia. This analytical approach was considered, based on two major merits; the method is generally able to retain the geometry or shape of the organism in question based on coordinate data which is ideally impossible for the traditional methods that involve linear measurements like, lengths, depths etc. In traditional methods, there is always a challenge of variable overlap which renders the method problematic to depict shape changes in organisms. Unlike the traditional approaches, geometric analysis is generally less prone to biases during analysis considering the specimen size or weight, which is possible through Procrustes superimposition implemented in MorphoJ software (Breno, Leirs, & Van Dongen, 2011; Klingenberg, 2013; Maderbacher et al., 2008).

Like other organisms, fish are susceptible to both natural and anthropogenic environmental alterations which may ultimately impact the discrepancy in their phenotypic and genotypic traits amongst various populations. For any successful organism to exploit heterogeneous environments, it will depend on its ability to respond to varying conditions through phenotypic plasticity (Corse et al., 2012; Nyboer & Chapman, 2013; Ruehl, Shervette, & Dewitt, 2011). In this case, phenotypic plasticity triggered by environmental cues will enable a given genotype to produce varying phenotypes via ontogenic modifications (Corse et al., 2012). However, unconditionally an organism can still buffer against adverse environmental pressures/stressors and remain in a stable state to produce genetically predetermined phenotypes, a condition referred to as developmental stability (Carpentero and Tabugo, 2014; Clarke, 1998). But organisms in new environments may reflect reduced developmental stability (developmental instability) because of their inability to maintain a stable condition against environmental disturbances (Carpentero and Tabugo, 2014; Almeida, Almodóvar, Nicola, & Elvira, 2008). In this scenario, gradual divergent selective pressures from the inhabited environments can drive plasticity within and between populations (Nyboer & Chapman, 2013). Critical studies linking ecosystem exploitation and

functionality across environmental gradients may provide a central goal in understanding the evolutionary biology of organisms. Therefore, investigating Nile tilapia populations in the context of shape changes from various habitats can be vital for understanding the demography of the cichlid. For instance, if components of organisms are reproductively isolated, phenotypic variations among populations may be unfolded due to underlying differences in environmental and biotic stressors (Nyboer & Chapman, 2013). Also, genotypic composition of these populations can drift apart and consequently have an impact in their morphology.

The East African Nile tilapia presents a scenario of a species that has undergone isolations during stockings in natural water bodies and aquaculture production systems. In principle, organisms can colonize new habitats naturally like through river connections and this may eventually encourage varying morphological traits during local adaptation or drift. For example, in East Africa, the interconnection of lakes via rivers may ideally have impacts on the species' phenotypes but perhaps not significantly different amongst organisms in the same catchment. Theoretically, this is because analogous abiotic and biotic factors in the habitats may be maintained over generations which may have similar selective pressures on the inhabited organisms. The interconnection of water bodies generally suggest that the systems may not differ in both physicochemical and biological conditions which explains why the organisms may not as well diverge between the systems. Additionally, the fact that the connected water bodies present high chances of geneflow, as supported by the current genetic data, it shows that these populations are less likely to drift apart and maintain similar genotypes.

However, it should be noted that when species are anthropogenically dispersed further from their putative ancestral homes like in the case of the East African Nile tilapia, new selective environmental pressures and unfamiliar indigenous organisms may pronouncedly affect the aliens. Here, the introduced organisms will definitely encounter new stringent environmental pressures which induce the species to local adaptation and consequent morphological variability (Digo et al., 2015). In this same trend, morphological divergence can further be exacerbated by admixture where hybrids may evolve divergent morphotypes from the parental lines (Valiquette, Perrier, Thibault, & Bernatchez, 2014). Thus, it was expected that shape divergence was conspicuously in the waterbodies where Nile tilapia was introduced, e.g., in Lakes Victoria and Kyoga. However, as shown in chapter two, this was not entirely the case. Although morphological divergence was revealed pronouncedly in non-native water bodies; Victoria and Kyoga, similar morphotypes were also observed in some native environments like lakes Turkana, George and Albert as well as Victoria Nile. Moreover, shape divergent among the Lake Victoria sub-populations was evidenced

particularly for Sango Bay. These findings were partly concordant with my hypothesis. A possible explanation is that Nile tilapia might have encountered similar environmental pressures in waterbodies where it is non-native as in the ones where it is native. In this respect, the environments (e.g. Edward and Kazinga Channel) where the species exhibited relatively low shape changes could be different from the above waterbodies and perhaps with less stressing parameters. Similarly, the low shape change exhibited in the southwestern Ugandan high-altitude lakes was suspected to be due to limited time to accumulate shape divergence as the species' introduction happened rather recently. But the aspect of admixture with the congeners in the non-native ranges might have contributed to the species' shape change as well as in the natives.

The shape divergent within Lake Victoria sub-populations was a clear suspect of admixture of Nile tilapia with the indigenous species. For example, following Nile tilapia introduction from Lake Albert in the 1950s, the indigenous congeners (*O. variabilis* and *O. esculentus*), virtually disappeared from Victoria and Kyoga at a time when the biomass of the alien species was increasing (Bairwa et al, 2003; Ogutu-Ohwayo, 1990; Bairwa, 1992). It is suspected that the indigenous tilapiines became extinct due to hybridization with the introduced Nile tilapia, although competition could have been another factor (Ogutu-Ohwayo, 1990; Bairwa, 1992). Although Nile tilapia is native to Lake Albert, shape changes in this population might have been attributed to the effects of aquaculture production facilities, in particular, cage farming on the lake (Mbowa, Odokonyero, & Munyaho, 2016). Cage farming on Lake Albert could have influenced effects analogous to non-native environments as stocked fish frequently escape from holding systems into the surrounding waterbody where they then interbreed with natural populations. In reality, a combination of many factors might have contributed to shape change in the East African Nile tilapia. For example, the noted shape divergence in Lake Turkana might have been intensified because of isolations and by the adverse environmental conditions, such as water level decline exacerbated by extreme high temperatures (Seebacher, Webster, James, Tallis, & Ward, 2016). The clear morphological divergence in Lakes Victoria, Albert, Kyoga, Turkana and the upper Nile (Victoria Nile) may have also been influenced by the intense predation pressure from Nile perch. Predation risk avoidance can cause morphological change in fish in situations when the prey must seek refugia in habitats with lower risk, but at a cost of decreasing foraging efficiency (Eklöv & Svanbäck, 2005).

In the subsequent sub-section of this general discussion, the relevant genetic markers were developed and thereby utilized to delineate the genetic diversity and differentiation of the East African *O. niloticus*. Also details of fish translocations in East Africa are explained in light of Nile tilapia genetic diversity.

5.1.2 Genetic diversity and differentiation of Nile tilapia among regions

In chapter 4, the genetic structure of the East African *O. niloticus* was explored with 40 SSR markers using GBS techniques. Results indicated that the East African populations were genetically diverse in comparison to, for example, the Ethiopian populations. Moreover, the F_{st} and genetic isolation by distance (IBD) analyses showed that the East African populations were more homogenous but distant from the Ethiopians. In the same analysis, the Ethiopian populations showed generally two divisions, with Tana evidently distinct and genetically distant from the cluster of, Ziway, Chamo and Hashenge. But in general, for all the analyses, the genetic structure corresponded with multiple divisions both at local and regional levels. However, high levels of admixture were observed mainly among the East African populations, which may be partly caused by fish translocation practices for stocking purposes (discussed below). Admixture levels dramatically declined when only the indigenous populations were analysed separately, an indication that the introduced populations were more admixed as would be expected.

In all cases, the genetic diversity for a given population may be inclined on to fragmentation and admixture, among other factors. For example, the low genetic diversity indicated by Burkina Faso and Ethiopian populations may suggest strong effects of genetic drift leading to genetic erosion, resulting from populations fragmentation. This is associated to the probable absence or very limited potential of the contemporary dispersal of geneflow among the Ethiopian lakes (Vijverberg, Dejen, Getahun, & Nagelkerke, 2012). Apparently, all the studied Ethiopian lakes are not interconnected apart from the link between Lake Ziway and another unstudied Lake, Langano through Lake Abijata (Tudorancea & Taylor, 2002; Vijverberg et al., 2012). However, fragmentation among the Ethiopian populations is more pronounced in the mountainous Lakes Tana and Hashenge (Vijverberg et al., 2012) which clearly explains the distinctness of Nile tilapia in the former water body. This is contrary to the Rift Valley lakes Chamo and Ziway of which their vicinity and similar climatic conditions (Tesfaye, 2010; Vijverberg et al., 2012) may contribute to their observed cluster. But the closely relatedness of the mountainous Lake Hashenge to the Rift Valley Lakes Chamo and Ziway may be associated to anthropogenic activities. Apparently, Lake Hashenge is reported to have been stocked with alien *O. niloticus* following mass mortalities of the native species in the waterbody (Tsegay, 2017). However, the seed source for stocking in Hashenge is not reported which might have originated from the Rift Valley lakes, hence explaining the present results.

Overall, the current results contradict those of Trewavas (1983), who is perhaps the first scholar to comprehensively study the African freshwater cichlids and particularly the tilapiines (Noakes, 1994). Trewavas (1983), categorized African *O. niloticus* into seven subspecies including; *O.*

niloticus niloticus (*O. n. niloticus*), *O. niloticus eduardianus* (*O. n. eduardianus*), *O. niloticus cancellatus* (*O. n. cancellatus*), *O. niloticus filoa* (*O. n. filoa*), *O. niloticus vulcani* (*O. n. vulcani*), *O. niloticus baringoensis* (*O. n. baringoensis*) and *O. niloticus sugutae* (*O. n. sugutae*). The author further characterised and associated these groups geographically (Figure 8). For example, according to Trewavas (1983), *O. n. niloticus* is native to Yorkon River near the Mediterranean Sea, Nile below Albert Nile to the Delta, Lake Chad, Rivers Niger, Benue, Gambia, Volta and Senegal, *O. n. eduardianus* is native to Lakes George, Albert, Edward, Kivu, Tanganyinka and the Ruzizi River, *O. n. cancellatus* is native to Ethiopian Lakes Tana, Ziway, Omo River system and the Awash River, *O. n. filoa* is also native to the Ethiopian Hot Alkaline Springs of the Awash River System, *O. n. vulcani* is native to Lake Turkana and Loiengalani stream in Kenya, *O. n. baringoensis* is native to Lake Baringo, Kenya and *O. n. sugutae* naturally belong in Kapedo river and its warm alkaline springs of Kenya (Figure 8). While in this thesis, Albert looked rather undifferentiated from Victoria Nile and Lake Kyoga, but typically distant from the rest of the western Rift Valley lakes, Trewavas (1983) found that Lakes Edward-George system was similar to Albert (Uganda) and that Lake Tana is connected to other Ethiopian populations. Nonetheless, the current results regarding the distinctiveness on Nile tilapia from Lake Turkana concur with the findings of Trewavas (1983), who equally found the populations unique and considered the species in there to be *O. n. vulcani*. However, the work of Trewavas (1983) has not only differed from some of the current results, but also to other authors. For example, Vreven, Adépo-Gourène, Agnèse, and Teugels (1998) using similar morphometrics and accompanied with Allozyme genetic markers found the Lake Edward strain was closely related to the lower Nile (Egypt). Trewavas (1983) in contrast concluded that, the lower Nile strain in Egypt belonged to the West African strains collectively referred to as *O. n. niloticus*. However, the current results concerning the distinctiveness of Ethiopian Tana concur with the findings from Seyoum and Kornfield (1992). These authors employed restriction endonuclease mitochondrial DNA (mtDNA) to investigate the Ethiopian populations and similarly found Nile tilapia from Lake Tana genetically distinct and in fact they ruled out the name *O. n. cancellatus* which was previously assigned by Trewavas (1983). Therefore, the observed distinctness of the Ethiopian Nile tilapia in Lake Tana and the Ugandan Edward-George cluster may warrant future explorations for taxonomic recognitions.

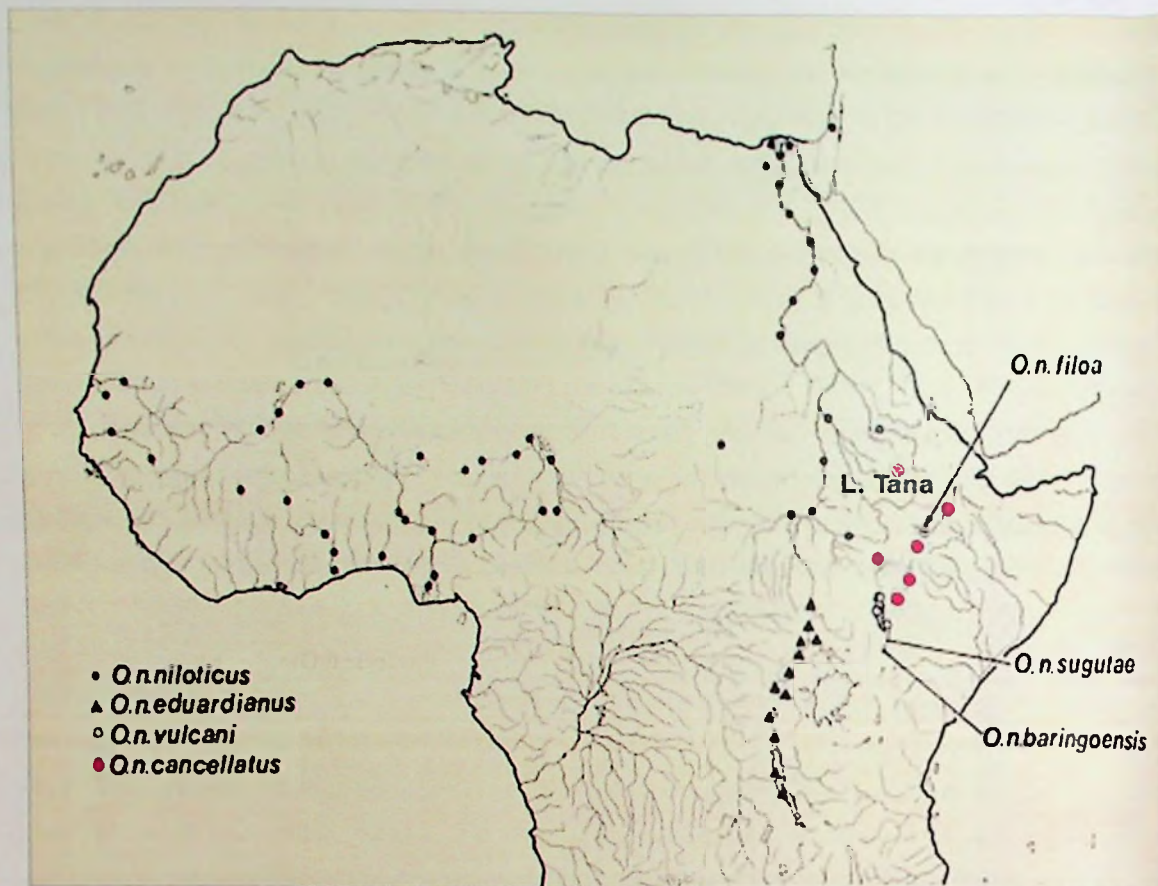


Figure 8 Classification of seven subspecies of *Oreochromis niloticus* and their natural geographical ranges (source: Trewavas, 1983, with minor modifications).

5.1.3 Nile tilapia geneflow patterns delineated by migration rates/anthropogenic fish translocations

The high genetic diversity and geneflow observed mainly in the East African populations are thought to be a consequence of re-stocking practices which may have partly promoted hybridization between *O. niloticus* and the native congeners. Perhaps, this may explain why the indigenous and close relatives of Nile tilapia disappeared from Lake Victoria basin following the introduction of the alien cichlids through genetic swamping.

The present results regarding admixture was more markedly in the non-native populations especially Lake Victoria consistent with high numbers of private alleles. Based on the current genetic results, admixture of *O. niloticus* populations in East African may be attributed to three main anthropogenic induced 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 especially in the non-native environment. In this context, the first and third anthropogenic mediated processes, may explain partly the genetic variation found in Lake Victoria. Although *O. niloticus* in Lake Victoria is generally isolated, in the distance neighbor network tree, the population occupied an intermediate position between the western Rift Valley lakes (the genetic cluster of Lakes Edward-George as well as Kazinga Channel) and another Ugandan cluster of Albert-Victoria Nile and Kyoga. Thus, it is clearly possible that multiple stockings might have contributed to the gene-pool indicated by Lake Victoria. This clearly explains why Lake Victoria (Ugandan waters) clustered independently from for example Lake Kyoga, despite both being non-native and reported to have received stockings from the same putative source (Ogutu-Ohwayo, 1990; Baliwa, 1992). In this case, Indeed Lake Kyoga might have received stocking seed from a single source. Apparently, the isolation of Lake Victoria may be linked to the past multiple stockings and subsequent genetic swamping, accrued from introgressive hybridisation from various lineages. Additionally, while the subdivisions in Victoria were expected to unfold low genetic diversity perhaps resulting from founder effects, this was not the case. Consistently, Victoria sub populations were genetically heterogeneous. In contrast, the meta-populations in Kyoga were less divergent and analogous to the upper Nile (Victoria Nile). The meta-populations in Lake Victoria might also emanate from the multiple reported introductions/stockings of Nile tilapia into Victoria ecosystem. On the other hand, considering that most populations appeared to be undergoing a bottleneck (mostly detected in Kyoga and Victoria Nile), the population in Lake Kyoga and Victoria Nile might be suffering from the effects of overfishing with subsequent founder effects. Population subdivision and related low genetic diversity among fish species have been associated partly with the impacts of over-fishing (Allendorf, England, Luikart, Ritchie, & Ryman, 2008). But the bottleneck effects in River Nile, might also be highly linked to the dams that were constructed along the river (Ugandan side, upper Nile), which might be confining the species. However, this can be subjected to further analysis especially when additional samples are taken below the Nile at Murchison Falls. Nevertheless, the observed genetic traits of Nile tilapia from the East African water bodies are thought to be from mainly anthropogenic induced activities which could have impacted on both the alien and recipient populations.

The idea of fish relocations in East Africa started rather earlier, even when the native tilapiine fishery (particularly the famous Ngege fishery) was still booming (Graham, 1929; Worthington, 1932) explained by Anderson (1961). The initial plan was to introduce Nile perch in Lake Victoria. This was congruent with the high abundance of the cichlid haplochromines in both inshore and

offshores of L. Victoria, which were supposed to be converted into useful and high market demand marketable table size fish through predation. However, Nile tilapia is reported to have been introduced earlier in the 1950s and later Nile perch in the 1960s, but against the advice from Fryer (1960) who had predicted a possible tragedy on the indigenous fishes of Lake Victoria. Even before this time, earlier ideas for fish introductions in the 1920s were still criticized by (Graham, 1929). Though both species; Nile perch and Nile tilapia, are to blame for the collapse of the native species, the former was mainly responsible for the decline of haplochromines (Kaufman, 1992). Albeit overfishing led to early declines in the preferred native tilapiines, the most pronounced decline of these species occurred in the 1980s when the stocks of Nile tilapia exploded (Bairwa, 1992). Consequently, this situation raised many questions and concerns, as most scientists and other stakeholders blamed Nile tilapia as the causative agent (Bairwa, 1992; Kaufman, 1992; Mwanja, 2000; Ogutu-Ohwayo, 1990). It is likely that, based on arguments and advice against alien species from Fryer (1960), there was no consensus on how to conduct formal species introductions into the Lake Victoria region (see Table 1). In this context, what is perplexing is the incongruity in the historical reporting and it is not always easy to trace the putative source populations of Nile tilapia.

According to Welcomme (1966); Agnèse, Adépo-Gourène, and Pouyaud (1998), Nile tilapia was first stocked in Lake Bunyonyi (South-western Ugandan highland minor lake) in 1936 from L. Edward and later to Lake Victoria in the 1950s via Koki lakes (Table 1). But Agnèse, Adépo-Gourène, and Pouyaud (1998), explain that, some tilapiines (*Tilapia nigra*) was also introduced from Lake Naivasha (Kenya) into Lake Bunyonyi (see also table 1). Here it is unclear as to whether the stockings from Lake Bunyonyi are the ones which ended up into Victoria. In 1966, further massive stocking campaigns were conducted on the Ugandan side of Lake Victoria, whereby the brood stock source was unknown (Bairwa, 1992; Ogutu-Ohwayo, 1990). Trewavas (1983) reported that, in the 1950s, the species was stocked several times into Lakes Victoria and Kyoga, and that other introductions might have come from Lake Turkana. Other reports indicated that the stocking material came from the ponds at Kajjansi (Entebbe Road-Uganda) (Pullin, 1988; Trewavas, 1983). In general, Trewavas, (1983) and Bairwa (1992) reported numerous sources of Nile tilapia into Lakes Victoria and Kyoga. In contrast, Mwanja (2000) reported that, intentional Nile tilapia stocking in L. Kyoga was from a single source, L. Albert as opposed to multiple stockings in L. Victoria (Table 1). Although (Pullin, 1988) also indicated possible stocking from ponds at Kajjansi and Lake Turkana, the author suggests that Nile tilapia and *O. leucostictus* were mistakenly introduced when *Tilapia zillii* was translocated from Lake Albert into Lake Victoria. Pullin (1988) also reported that, in 1967, Nile tilapia was observed in Mwanza Prisons' culture ponds near the

shores of Lake Victoria (Tanzanian side), thereby suggesting other possible sources of feral populations.

The massive and multiple introductions of Nile tilapia into particularly Lake Victoria may be linked to the observed subdivisions indicated by both geometric morphometrics and molecular data. Additionally, high genetic diversity and private alleles observed in Lake Victoria may be in line with multiple stockings or hybridization. However, the uniqueness showed by Nile tilapia from Lake Victoria despite the population being non-native may signify that apparently the population is speciated via introgressive hybridization. This is in line with, Balirwa (1992); Welcome, (1966) who observed some unique phenotypic characters in the Northern Lake Victoria waters and suspected them to hybrids between Nile tilapia and *O. variabilis*. Based on the current results, the distinctiveness of Nile tilapia in Lake Victoria requires further exploration, particularly at mitochondrial genome level for subsequent taxonomic recognition.

Table 1. Historical translocations of tilapiines in East African water bodies

Fish source	Species	Year	Destination	Citation
L. Edward	<i>O. niloticus</i>	1927	L. Bunyonyi	Agnèse et al 1998
L. Naivasha	<i>O. spilurus nigra</i>	1932	L. Bunyonyi	Agnèse et al 1998
L. Edward	<i>O. niloticus</i>	1950s	L. Victoria & Kyoga	Balirwa 1992 Kaufman 1992
P. Kajjansi U	<i>O. niloticus</i> fry	1956 & 1958	L. Victoria	Balirwa 1992
L. Albert	<i>Tilapia zillii</i>	1951 & 1962	L. Victoria	Balirwa 1992
L. Albert	<i>O. leucostictus</i>	1951 & 1962	L. Victoria	Balirwa 1992
Zambia	<i>Tilapia rendalli</i>	1952 & 1962	L. Victoria	Balirwa 1992
**	Thousands of fry	1961 & 1966	L. Victoria	Balirwa 1992
L. Albert	<i>O. niloticus</i>	1950s & 1960s	L. Victoria	Ogutu-Ohwayo 1990
L. Albert	<i>Tilapia zillii</i>	1950s & 1960s	L. Victoria	Ogutu-Ohwayo 1991
L. Albert	<i>O. leucostictus</i>	1950s & 1960s	L. Victoria	Ogutu-Ohwayo 1992
L. Albert	<i>O. niloticus</i>	-	L. Victoria & Kyoga	Pullin 1988
L. Albert	<i>Tilapia zillii</i>	-	L. Victoria	Pullin 1988
L. Albert	<i>O. leucostictus</i>	-	L. Victoria	Pullin 1988
P. Kajjansi U	<i>O. niloticus</i>	-	L. Victoria	Pullin 1988
L. Turkana	<i>O. niloticus</i>	-	L. Victoria	Pullin 1988
L. Edward	<i>O. niloticus</i>	1924	L. Bunyonyi	Trewavas 1983
L. Turkana	<i>O. niloticus</i>	1951	P. Kajjansi U	Trewavas 1983
L. Albert	<i>Tilapia zillii</i>	1951	L. Victoria	Trewavas 1983

L. Albert	<i>O. leucostictus</i>	-	L. Victoria	Trewavas 1983
L. Edward	<i>O. niloticus</i>	1936	L. Bunyonyi	Welcome 1966
L. Edward	<i>O. niloticus</i>	1954	Koki lakes U	Welcome 1967
L. Albert	<i>O. niloticus</i>	1960s	L. Victoria & Kyoga	Welcome 1967; Lowe-McConnel 1975; Mwanja 2000

**=unknown source, P=ponds and U= Uganda

5.1.4 Nile tilapia genetic structure in the context of ecogeographical location

The genetic structure of the East African populations indicated to be consistent with geographical location and corresponded to water catchment interconnections. For example, the genetic link of George to Edward via Kazinga Channel and the connection between Victoria Nile with Kyoga. This was ideally expected based on the water connections in these systems. Moreover, the Edward-George cluster was unveiled to be connected to the South-western Ugandan high-altitude lakes which clearly explains the possible putative sources of stocking to the latter populations. Similarly, the distinctiveness of *O. niloticus* in Lake Turkana may be associated to its nature of isolation which prevents any possibility of geneflow from gene pools of other populations.

However, the low genetic diversity results from upper River Nile (Victoria Nile) and the non-native Kyoga were generally unexpected. But it is likely that the natural water connections between George and Edward via Kazinga Channel generally maintains the genetic diversity moderately in these systems for generations. This corresponds to explanations from earlier reports that, under natural population connectivity patterns, the evolutionary genetic potential of organisms may be kept to relatively higher levels than in the local populations (Palstra & Ruzzante, 2011). On the other hand, ideally, Lake Albert would be clustered to the Edward, Kazinga Channel and George cluster via River Semliki, which was not possible with genetic structure analyses. However, it appears that the strong rapids on R. Semliki falls (Dumont, 2009; Ogutu-Ohwayo, 1990) interrupt geneflow between these systems which maintain Albert apart. Similarly, the unexpected absence of genetic connectivity between Lake Victoria and Victoria Nile, given the former being the source to the latter, may stem from the obstruction at Owen Falls Hydro Electric Power Dam between these systems (Dumont, 2009). But the observed genetic link of Lake Albert, to the cluster of Victoria Nile and Kyoga in Uganda, may be a result of the geneflow influenced by anthropogenic activities, and rather not the naturally connectivity between the cluster and the Albertine waterbody. This is because the presence of Murchison Falls on Victoria Nile acts as a natural barricade against the possibility of fish migrations to downstream which conventionally allows un connectivity of fish species between the upper Nile and Albert (Dumont, 2009). The cluster between these systems

may be elucidated by *O. niloticus* introduction from Albert to directly Victoria Nile in the 1950s as reported by Ogutu-Ohwayo (1990). Ogutu-Ohwayo (1990) further clarifies that, the Nile tilapia strain below the Murchison falls in Victoria Nile may be native contrary to the non-native one in the upper Murchison falls. These results tally with those of Mwanja, (2000) who argued that, Nile tilapia was directly introduced from a single source, Albert to Lake Kyoga system (and perhaps in Victoria Nile), which contrasts to multiple sources in Lake Victoria. Therefore, the interaction of many gene pools from various Nile tilapia strains together with the indigenous congenics, may explain the current observed distinctiveness of the cichlid in Lake Victoria, compared to for example Lake Kyoga.

Expectedly, the non-native Lake Victoria population in general was found to be genetically divergent consistent with high private alleles. Although the population revealed numerous gene flow, only a few individuals from Lake Victoria appeared to cluster with other populations. The fact that Victoria is non-native, a pronounced cluster with the possible putative source of stocking would be expected. But rather, this was not the case and it could be a clear indicator that Nile tilapia in Lake Victoria might have undergone genetic swamping, as a result of introgressive hybridisation with the native congenics. It should be noted that, although several reports indicate hybrids as commendable traits for adaptation and subsequent performance in diversified environments, in respect to heterosis (Dodson et al., 2015; Dowling et al., 2016), on the other side, the consequences may be impractical and unsustainable over generations (Valiquette, Perrier, Thibault, & Bernatchez, 2014). For instance, consequences of admixture may lead to formation of transgressive segregation phenotypes which may be positive, and in extreme cases negative in terms of fitness. In this context, transgressive segregation may influence extinction either via parental line displacement, or gradual extinction through outbreeding depression (Rhymer & Simberloff, 1996; Dodson et al., 2015; Rieseberg, Archer, & Wayne, 1999). Moreover, erosion in genetic diversity and rapid loss in numbers of some wild fish populations have been condemned on to hybridization e.g. in Salmon (*Salmo salar*) (Bourret, O'reilly, Carr, Berg, & Bernatchez, 2011). Similarly, disappearance or extinction of South African indigenous *O. mossambicus* has been reported to the causal effects of introgression with nonnative *O. niloticus* (Zengeya, Robertson, Booth, & Chimimba, 2013). In East Africa, the fact that *O. niloticus* has been massively cultured over years, with some aquaculture production systems on or around various water bodies (Blow & Leonard, 2007), escapees from these systems into natural water bodies could have had dramatic effects with indigenous species. This is in addition to the past direct restocking regimes that were implemented in these water bodies. It should be noted that, conclusive deductions on introgression

is necessary with inclusion of species closely related to Nile tilapia in the analyses and such studies are ongoing but out of scope of this thesis.

The current results differed from other previous molecular studies on Nile tilapia in the region. For instance, in East Africa, Agnèse, Adépo-Gourène, Abban, and Fermon (1997); Vreven et al. (1998); and Mwanja, Booton, Kaufman, and Fuerst (2008) studied the evolutionary biology of Nile tilapia using allozymes with restriction fragment length polymorphism (RFLP) of mitochondrial DNA (mtDNA), allozyme and random amplified polymorphic DNA (RAPD) respectively. Agnèse et al. (1997) studied only three East African lakes; Edward (Uganda), Baringo and Turkana (Kenya), the rest of the samples were Ethiopian and West African. On the other hand, Mwanja et al. (2008) studied most of the major bodies in Uganda; Victoria, Kyoga, Edward, George, Albert and several satellites in central Uganda. Mwanja et al. (2008) showed that Nile tilapia populations from lakes Victoria and Edward were alike and they concluded that Lake Victoria population was most likely from Edward instead of Lake Albert as previously reported. Additionally, these authors noted that populations from George and Albert were similar than would be for George to Edward and attributed it to stocks planted from Albert to George. However, our results concerning the similarity of Lake Albert and Kyoga concur with those of Mwanja et al. (2008). These authors explain that, whereas both Lake Victoria and Kyoga are non-native, the introduction of Nile tilapia into the latter water body was much more recent than in Lake Victoria which may relate to the observed genetic variations in these systems. Agnèse et al. (1997) reported that, Nile tilapia populations were clustered into three groups; i) West African populations (Lakes; Senegal, Niger, Volta, and Chad drainages), ii) Ethiopian Rift Valley populations (Lakes; Ziway, Awasa, Koka and Sodre hot springs in the Awash River, and iii) Nile drainage populations; River Nile, Lakes Tana and Edward and Kenyan Rift Valley populations; Turkana, Baringo and River Suguta). Interestingly, findings from Agnèse et al. (1997) also deviated from both the studies of Trewavas (1983) and Seyoum and Kornfield (1992). While Seyoum and Kornfield (1992) for example, indicated that Nile tilapia in Lake Tana is distinct, Agnèse et al. (1997) clustered the population to Lake Edward and the Turkana system. Trewavas (1983) associated Nile tilapia from the lower River Nile (Egypt) to West African strains, but Agnèse et al. (1997) linked Nile tilapia from River Nile with the Edward and Turkana system.

Albeit the aforementioned previous studies provide important contributions in understanding the demography of Nile tilapia in Africa, it is likely that the used methodological tools might have been constrained by either low resolving power or precision accompanied with limited narrow genome regions (Narum et al., 2013) in characterizing the populations and hence the observed

discrepancies. Moreover, some of the earlier used markers like RAPD, have been labelled as loci with patterns of dominance inheritance which reduce result reliability (Wirgin & Waldman, 1994). The present patterns of genetic structure of Nile tilapia in East African based on SSR-GBS show the impacts of anthropogenic activities, which might not sustain the populations. Proper guidelines and environmental impact assessment for fish translocations and infrastructure development should be adhered to when deemed necessary for aquatic environments. This is important for avoiding the disruption of the natural genetic structure of the fish species through propelled gene flow between species or confinements with related founder effects.

5.1.5 Use of microsatellite loci (SSR) and the strengths of genotyping by sequencing (GBS) under next generation sequencing (NGS)

Since the 1980s, SSR markers have been extensively employed in many biological fields, including population genetics, conservation genetics, parentage, molecular breeding etc on various organisms because of their proved robustness (Oliveira, Pádua, Zucchi, Vencovsky, & Vieira, 2006; Vartia et al., 2014). In principle, albeit SSR are utilized widely, they are not without bottlenecks, including labor intensive, time consuming, costs involved in genotyping and low marker yield. But this is more realized specifically when considering the traditional SSR development techniques that involve enriching and cloning procedures and genotyping based on fragment length analysis (Vartia et al., 2014; Zane, Bargelloni, & Patarnello, 2002). In the current work, 26 SSR markers were developed and genotyped using NGS with the aim of testing the genetic Structure of East African *O. niloticus*. Such an approach can overcome some of the limitations associated with microsatellites. This was done by comparing the traditional SSR sequence length analysis with genotyping by sequencing approaches.

One intriguing question would be, why the available/existing developed SSR markers from molecular data were not used? Or why not utilize the markers developed for other generically similar species (cross-species amplification) or even neutral markers? The development of new markers was partially necessary to acquire longer amplicon lengths to the maximum of 300 bp with traditional genotyping. This way more sequencing information can be recovered increasing the change of recovering SNP variation in the flanking region and consequently increasing the number of alleles. However, the main reason was to avoid consequences of high numbers of null alleles which result from cross-species amplification. It is indicated that even when cross-species amplification is successful, generally lower levels of variability may be exhibited than the previous species to which the markers were developed (Primmer, Møller, & Ellegren, 1996). The traditionally developed loci normally result into low number of loci and may be less taken into consideration

primarily in studies that involve undifferentiated populations because large numbers of SSR markers are required (Zane et al., 2002). With the use of next generation sequencing (NGS), the discovery of SSR markers has proved gradually more economical and faster especially on non-modal organisms (Curto et al., 2013; Vartia et al., 2014).

In this thesis, the use of SSR markers based on genotyping by sequencing (GBS) provided a greater advantage over the traditional amplicon length in resolving issues like size homoplasy (Vartia et al., 2016). The genotyping approach also revealed potentially the veiled genotypes in the amplicon/whole sequence information contrary to the conventional sequence length. Moreover, the SSR-GBS approach exposed more highly polymorphic loci and were all generally informative when the genetic structure of the East African *O. niloticus* populations was tested contrary to the traditional amplicon length (AL) method (Tibihika et al., 2018). The robustness of SSR-GBS was recommended for future studies involving evolutionary studies of cichlids particularly the East African *O. niloticus*. This is because the SSR-GBS analytical platform allows for high reproducibility and the genotyping of many individuals/populations may be very useful for monitoring stocks like the case of East African Nile tilapia.

Microsatellite markers with NGS under genotyping by sequencing (SSR-GBS) might be a choice by evolutionary biologists because of its promises as a common genotyping platform in future molecular studies (Narum et al., 2013). In comparison to other genotyping approaches and marker systems, GBS has proved beyond doubts in population genomics by unfolding thousands of SNPs throughout the genome that are prone to evolutionary trajectories including selection, recombination, mutation and genetic variability, inter alia (Narum et al., 2013). Moreover, GBS has proved robust at enhancing precision of structural inferences by increasing the number of neutral markers during genotyping (Narum et al., 2013).

6. Chapter 6

6.1. Conclusions and outlook

Overall, the combined geometrics and molecular genetics methodological tools in this thesis revealed to be powerful in delineating the current state of the East African *O. niloticus* populations from various environments. Using geometric morphometrics, it was difficult to differentiate the populations as they generally overlapped. However, populations from Victoria, Kyoga, Albert, Turkana, River Nile and George were observed morphologically divergent. In this analysis, two clusters, 1) River Nile, Albert, Kyoga and Victoria and 2) Kazinga Channel and Edward, were recorded. On the other hand, the shape variation in Lake Victoria sub-populations was revealed to be more pronounced in Sango Bay and Kakyanga. The rest of the sub-populations formed a cluster. I showed that application of SSR-GBS was a more powerful alternative to traditional fragment analysis by not only recovering a higher level of variation on Nile tilapia populations but also a more detailed genetic structure pattern. For this reason, this method was subsequently employed for analysing the genetic diversity and differentiation of all the East African, Ethiopian and Burkina Faso populations. Apparently, it was indicated that the Ethiopian Tana is distinct from all other Nile Tilapia populations studied which may require further exploration for taxonomic recognition. I also found the genetic structure of the East African population admixed, despite some isolations, but consistent with Geometric morphometrics. Genetic diversity was remarkable in non-native Victoria, though Kyoga unveiled to be less diversified. Despite geometric analyses indicating George divergent from Kazinga Channel and Edward, the molecular analyses indicated a cluster among these populations as one system. Genetic molecular analyses further showed Nile tilapia in Lake Albert is distant from Lake Victoria, unlike to Lake Kyoga and River Nile (upper Nile) which formed another cluster. The slight discrepancy between geometrics and molecular genetic analyses in this thesis, is linked to obvious different tools, which was expected. This is because, traits observed at molecular level may not necessarily be reflected in phenotypes which actually explains the importance of molecular data. The south-western Ugandan high-altitude lakes, Mulehe and Kayumbu were found associated to Edward, Kazinga Channel and George, which suggests their putative source of stocking. We could not trace the clear putative source of native *O. niloticus* into Lake Victoria which is an indicator that Nile tilapia in this water body might have undergone genetic swamping, accrued from introgressive hybridisation with the native congeners. Largely, all the populations were indicated somewhat bottlenecked but with severe indicators observed in River Nile and Kyoga consistent with low genetic diversity, which is a major challenge for management. While the challenges in the upper Nile might be related to bottleneck effects perhaps arising from the combined effects of constructed dams and overfishing, Lake

Kyoga maybe be solely overfished. In all systems, apart from the Edward-George cluster, it is possible that anthropogenic threats are enormous which calls for proper management interventions. Aquaculture activities, especially cage farming may be discouraged from the major naturally productive East African water bodies, but rather be set in the satellite/minor water bodies which are somewhat characterized by discontinuities. In general, both chapters in this thesis have revealed that the East African *O. niloticus* populations are admixed, deduced by geometric morphometrics and microsatellite genotyping by sequencing. The applied combination of morphometrics and molecular genetic tools in analyzing the current state of the East African *O. niloticus*, clearly indicate the relevance of the approach for providing adequate information for management and conservation. Albeit these approaches have shown to be strong, additional follow-up of these findings with mitochondrial genome analyses may sound vital (Mohanta et al., 2016; Zhao, Gao, & Lu, 2015). I am currently optimizing the findings from this thesis with inclusion of samples from the lower Nile (below Murchison Falls), Kivu (Rwanda), Tanganyika and Baringo for a broader picture. The analyses will consider genetic introgression and phylogeny through mitochondrial DNA amplification for broadly understanding the extent of biotic interactions of *O. niloticus* with congeners and putative sources of *O. niloticus*.

6.2. References

- Agnèse, J. F., Adépo-Gourène, B., Abban, E. K., & Fermon, Y. (1997). Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity*, 79(1), 88.
- Allendorf, F. W., England, P. R., Luikart, G., Ritchie, P. A., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends in ecology & evolution*, 23(6), 327-337. doi:<https://doi.org/10.1016/j.tree.2008.02.008>
- Almeida, D., Almodóvar, A., Nicola, G., & Elvira, B. (2008). Fluctuating asymmetry, abnormalities and parasitism as indicators of environmental stress in cultured stocks of goldfish and carp. *Aquaculture*, 279(1-4), 120-125. doi:<https://doi.org/10.1016/j.aquaculture.2008.04.003>
- Anderson, A. (1961). Further observations concerning the proposed introduction of Nile perch into Lake Victoria. *East African Agricultural and Forestry Journal*, 26(4), 195-201.
- Balirwa, J. S. (1992). The evolution of the fishery of *Oreochromis niloticus* (Pisces: Cichlidae) in Lake Victoria. *Hydrobiologia*, 232(1), 85-89. doi:10.1007/BF00014616
- Balirwa, J. S., Chapman, C. A., Chapman, L. J., Cowx, I. G., Geheb, K., Kaufman, L., . . . Welcomme, R. L. (2003). Biodiversity and fishery sustainability in the Lake Victoria basin: an unexpected marriage? *BioScience*, 53(8), 703-715. doi:[https://doi.org/10.1641/0006-3568\(2003\)053\[0703:BAFSIT\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2003)053[0703:BAFSIT]2.0.CO;2)
- Blow, P., & Leonard, S. (2007). A review of cage aquaculture: sub-Saharan Africa. *FAO Fisheries Technical Paper*, 498, 191.
- Bourret, V., O'reilly, P., Carr, J., Berg, P., & Bernatchez, L. (2011). Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity*, 106(3), 500. doi:<https://doi.org/10.1038/hdy.2010.165>
- Carpentero, E. R., & Tabugo, S. R. M. (2014). Determining developmental instability via fluctuating asymmetry in the shell shape of Arctica islandica Linn. 1767 (ocean quahog). *European Journal of Zoological Research*, 3(3), 1-7.
- Clarke, G. M. (1998). The genetic basis of developmental stability. IV. Individual and population asymmetry parameters. *Heredity*, 80(5), 553.
- Corse, E., Neve, G., Sinama, M., Pech, N., Costedoat, C., Chappaz, R., & ANDRE, G. (2012). Plasticity of ontogenetic trajectories in cyprinids: a source of evolutionary novelties. *Biological Journal of the Linnean Society*, 106(2), 342-355. doi:<https://doi.org/10.1111/j.1095-8312.2012.01873.x>
- Curto, M. A., Tembrock, L. R., Puppo, P., Nogueira, M., Simmons, M. P., & Meimberg, H. (2013). Evaluation of microsatellites of *Catha edulis* (qat; Celastraceae) identified using pyrosequencing. *Biochemical systematics and ecology*, 49, 1-9. doi:<https://doi.org/10.1016/j.bse.2013.02.002>
- Dodson, J. J., Bourret, A., Barrette, M. F., Turgeon, J., Daigle, G., Legault, M., & Lecomte, F. (2015). Intraspecific genetic admixture and the morphological diversification of an estuarine fish population complex. *PloS one*, 10(4), e0123172. doi:<https://doi.org/10.1371/journal.pone.0123172>
- Dowling, T. E., Markle, D. F., Tranah, G. J., Carson, E. W., Wagman, D. W., & May, B. P. (2016). Introgressive hybridization and the evolution of lake-adapted Catostomid fishes. *PloS one*, 11(3), e0149884. doi: <https://doi.org/10.1371/journal.pone.0149884>
- Dumont, H. J. (2009). *The Nile: origin, environments, limnology and human use* (Vol. 89): Springer Science & Business Media.
- Eklöv, P., & Svanbäck, R. (2005). Predation risk influences adaptive morphological variation in fish populations. *The American Naturalist*, 167(3), 440-452. doi:<https://doi.org/10.1086/499544>

- Fryer, G. (1960). Concerning the Proposed Introduction of Nile Perch into Lake Victoria. *The East African Agricultural Journal*, 25(4), 267-270. doi:https://doi.org/10.1080/03670074.1960.11665278
- Graham, M. (1929). *The Victoria Nyanza and its Fisheries: A Report on the Fishing Survey of Lake Victoria, 1927-1928, and Appendices*: Crown agents for the Colonies.
- Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, 11, 117-122. doi:https://doi.org/10.1111/j.1755-0998.2010.02967.x
- Kaufman, L. (1992). Catastrophic change in species-rich freshwater ecosystems. *BioScience*, 42(11), 846-858. doi:10.2307/1312084
- Lowe-McConnell, R. (2009). Fisheries and cichlid evolution in the African Great Lakes: progress and problems. *Freshwater Reviews*, 2(2), 131-151. doi:https://doi.org/10.1608/FRJ-2.2.2
- Mohanta, S. K., Swain, S. K., Das, S. P., Bit, A., Das, G., Pradhan, S., . . . Das, P. (2016). Complete mitochondrial genome sequence of *E. suratensis* revealed by next generation sequencing. *Mitochondrial DNA Part B*, 1(1), 746-747. doi:https://doi.org/10.1080/23802359.2016.1176877
- Moore, S. S., Whan, V., Davis, G. P., Byrne, K., Hetzel, D. J., & Preston, N. (1999). The development and application of genetic markers for the Kuruma prawn *Penaeus japonicus*. *Aquaculture*, 173(1-4), 19-32. doi:https://doi.org/10.1016/S0044-8486(98)00461-X
- Mwanja, W. (2000). *Genetic biodiversity and evolution of two principal fisheries species groups, the labeine and tilapiine, of Lake Victoria, East Africa*. Ph. D. thesis, Ohio State University, Cleveland, Ohio.
- Mwanja, W., Booton, G. C., Kaufman, L., & Fuerst, P. A. (2008). 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*, 7(11), 1769-1773. doi:10.5897/AJB08.813
- Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular ecology*, 22(11), 2841-2847. doi:https://doi.org/10.1111/mec.12350
- Noakes, D. L. (1994). The life and work of Ethelwynn Trewavas: beyond the focus on tilapiine cichlids. *Environmental Biology of Fishes*, 41(1-4), 33-50. doi:https://doi.org/10.1007/BF00023799
- Nyboer, E. A., & Chapman, L. J. (2013). Ontogenetic shifts in phenotype-environment associations in Nile perch, *Lates niloticus* (Perciformes: Latidae) from Lake Nabugabo, Uganda. *Biological Journal of the Linnean Society*, 110(2), 449-465. doi:https://doi.org/10.1111/bij.12122
- Ogutu-Ohwayo, R. (1990). The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of introduced species, especially the Nile perch, *Lates niloticus*, and the Nile tilapia, *Oreochromis niloticus*. *Environmental Biology of Fishes*, 27(2), 81-96. doi:https://doi.org/10.1007/BF00001938
- Ogutu-Ohwayo, R., & Balirwa, J. S. (2006). Management challenges of freshwater fisheries in Africa. *Lakes & Reservoirs: Research & Management*, 11(4), 215-226. doi:https://doi.org/10.1111/j.1440-1770.2006.00312.x
- Ogutu-Ohwayo, R., Hecky, R. E., Cohen, A. S., & Kaufman, L. (1997). Human impacts on the African Great Lakes. *Environmental Biology of Fishes*, 50(2), 117-131. doi:https://doi.org/10.1023/A:1007320932349
- Oliveira, E. J., Pádua, J. G., Zucchi, M. I., Vencovsky, R., & Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29(2), 294-307. doi:http://dx.doi.org/10.1590/S1415-47572006000200018

- Palstra, F., & Ruzzante, D. E. (2011). Demographic and genetic factors shaping contemporary metapopulation effective size and its empirical estimation in salmonid fish. *Heredity*, 107(5), 444.
- Pojas, R. G., & Tabugo, S. R. (2015). Fluctuating asymmetry of parasite infested and non-infested *Sardinella* sp. from Misamis Oriental, Philippines. *Aquaculture, Aquarium, Conservation & Legislation*, 8(1), 7-14.
- Pullin, R. S. (1988). Tilapia genetic resources for aquaculture. Proceedings of the Workshop on Tilapia Genetic Resources for Aquaculture, 23-24 March 1987. *The WorldFish Center Working Papers*.
- Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual review of ecology and systematics*, 27(1), 83-109.
- Ruehl, C. B., Shervette, V., & Dewitt, T. J. (2011). Replicated shape variation between simple and complex habitats in two estuarine fishes. *Biological Journal of the Linnean Society*, 103(1), 147-158. doi:<https://doi.org/10.1111/j.1095-8312.2011.01626.x>
- Seebacher, F., Webster, M. M., James, R. S., Tallis, J., & Ward, A. J. (2016). Morphological differences between habitats are associated with physiological and behavioural trade-offs in stickleback (*Gasterosteus aculeatus*). *Royal Society open science*, 3(6), 160316. doi:<https://doi.org/10.1098/rsos.160316>
- Seyoum, S., & Kornfield, I. (1992). Identification of the subspecies of *Oreochromis niloticus* (Pisces: Cichlidae) using restriction endonuclease analysis of mitochondrial DNA. *Aquaculture*, 102(1-2), 29-42.
- Tesfaye, G. (2010). Impacts of Anthropogenic Activities on Fish Diversity of the Ethiopian Rift Valley Lakes: A Review.
- Tsegay, T. (2017). Characteristics and Status of the Highland Lake Hashenge Fishery, Tigray, Northern Ethiopia. *International Journal of Aquaculture and Fishery Sciences*, 3(4), 71-76. doi:<https://doi.org/10.17352/2455-8400.000032>
- Tibihika, P. D., Curto, M., Dornstaeder-Schrammel, E., Winter, S., Alemayehu, E., Waidbacher, H., & Meimberg, H. (2018). Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*, 1-16. doi:<https://doi.org/10.1007/s10592-018-1136-x>
- Tibihika, P. D., Waidbacher, H., Masembe, C., Curto, M., Sabatino, S., Alemayehu, E., . . . Meimberg, H. (2018). Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa. *Environmental Biology of Fishes*, 101(3), 363-381. doi:<https://doi.org/10.1007/s10641-017-0704-0>
- Trewavas, E. (1983). *Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia*: British Museum (Natural History).
- Tudorancea, C., & Taylor, W. D. (2002). *Ethiopian rift valley lakes*: Backhuys.
- Valiquette, E., Perrier, C., Thibault, I., & Bernatchez, L. (2014). Loss of genetic integrity in wild lake trout populations following stocking: insights from an exhaustive study of 72 lakes from Québec, Canada. *Evolutionary Applications*, 7(6), 625-644. doi:<https://doi.org/10.1111/eva.12160>
- Vartia, S., Collins, P. C., Cross, T. F., Fitzgerald, R. D., Gauthier, D. T., McGinnity, P., . . . Carlsson, J. (2014). Multiplexing with three-primer PCR for rapid and economical microsatellite validation. *Heredity*, 151(2-3), 43-54. doi:<https://doi.org/10.1111/hrd2.00044>
- Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., . . . Cross, T. F. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society open science*, 3(1), 150565. doi:<https://doi.org/10.1098/rsos.150565>

- Vijverberg, J., Dejen, E., Getahun, A., & Nagelkerke, L. A. (2012). The composition of fish communities of nine Ethiopian lakes along a north-south gradient: threats and possible solutions. *Animal Biology*, 62(3), 315-335. doi:<https://doi.org/10.1163/157075611X618246>
- Vreven, E. J., Adépo-Gourène, B., Agnèse, J.-F., & Teugels, G. (1998). Morphometric and allozyme variation in natural populations and cultured strains of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Belgian Journal of Zoology*, 128(1), 23-34.
- Welcomme, R. (1966). Recent changes in the stocks of Tilapia in Lake Victoria. *Nature*, 212, 52-54. doi:[10.1038/212052a0](https://doi.org/10.1038/212052a0)
- Wirgin, I. I., & Waldman, J. R. (1994). What DNA can do for you. *Fisheries*, 19(7), 16-27. doi:[https://doi.org/10.1577/1548-8446\(1994\)019<0016:WDCDFY>2.0.CO;2](https://doi.org/10.1577/1548-8446(1994)019<0016:WDCDFY>2.0.CO;2)
- Worthington, E. B. (1932). A report on the fisheries of Uganda investigated by the Cambridge Expedition to the East African Lakes, 1930-1931. Chap. IV: Lake Nabugabo. Crown Agents.
- Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11(1), 1-16. doi:<https://doi.org/10.1046/j.0962-1083.2001.01418.x>
- Zengeya, T. A., Robertson, M. P., Booth, A. J., & Chimimba, C. T. (2013). A qualitative ecological risk assessment of the invasive Nile tilapia, *Oreochromis niloticus* in a sub-tropical African river system (Limpopo River, South Africa). *Aquatic Conservation: Marine and Freshwater Ecosystems*, 23(1), 51-64. doi:<https://doi.org/10.1002/aqc.2258>
- Zhao, L., Gao, T., & Lu, W. (2015). Complete mitochondrial DNA sequence of the endangered fish (*Bahaba taipingensis*): Mitogenome characterization and phylogenetic implications. *ZooKeys*(546), 181. doi:[10.3897/zookeys.546.5964](https://doi.org/10.3897/zookeys.546.5964)

Publications

- Tibihika, P. D., Curto, M., Dornstaeder-Schrammel, E., Winter, S., Alemayehu, E., Waidbacher, H., & Meimberg, H. (2018). Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*, 1-16. <https://doi.org/10.1007/s10592-018-1136-x>
- Tibihika, P. D., Waidbacher, H., Masembe, C., Curto, M., Sabatino, S., Alemayehu, E., . . . Meimberg, H. (2018). Anthropogenic impacts on the contextual morphological diversification and adaptation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East Africa. *Environmental Biology of Fishes*, 101(3), 363-381. <https://doi.org/10.1007/s10641-017-0704-0>
- Tibihika P. D., Curto M, Alemayehu E, Waidbacher H, Masembe C, Meimberg H Molecular genetic evidence of anthropogenic threats in the East African fisheries: the example of *Oreochromis niloticus* (**internal review**)
- Kwikiriza, G., Barekye, A., Aheisibwe, A., Byakora, E., & Tibihika, P. D. (2017). Comparative Growth Performance and Proximate Nutrient Composition of Three Local Strains of Nile Tilapia (*Oreochromis Niloticus* L.) Collected From Different Locations in Uganda. *Fisheries and Aquaculture Journal*, 8(3), 1-6. <https://doi:10.4172/2150-3508.1000226>
- Tibihika, P. D., Okello, W., Barekye, A., Mbabazi, D., Omony, J., & Kiggundu, V. (2016). Status of Kigezi minor Lakes: A limnological survey in the Lakes of Kisoro, Kabale and Rukungiri Districts. *International Journal of Water Resources and Environmental Engineering*, 8(5), 60-73. <https://doi.org/10.5897/IJWREE2016.0645>
- Kwikiriza, G., Tibihika, P. D, Barekye, A., Beingana, A., & Orina, P. (2016). Performance Of African Catfish (*Clarias gariepinus*) Clarridae Fry Fed On Live Rotifers (*Brachionus Calyciflorus*), Formulated Diet And A Mixture Of Rotifers And Formulated Diet. *International Journal of Fisheries and Aquatic Studies*, 4(6), 11-15.
- Tibihika, P. D., Barekye, A., & Byakora, E. (2015a). Critical analysis of the growth performance of three fish species; Mirror carp (*Cyprinus carpio*), Nile tilapia (*Oreochromis niloticus*) and Sharp toothed African catfish (*Clarias gariepinus*), cultured in earthen ponds in Kigezi region South-western Uganda. *International Journal of Fisheries and Aquatic Studies*, 2(3), 67-73.
- Tibihika, P. D., Barekye, A., & Byakora, E. (2015b). Fish species composition, abundance and diversity of Minor Lakes in south-western Uganda/Kigezi region. *International Journal of Science and Technology*, 4(5), 204-213.
- Tibihika, P. D. (2013). Fish farming dynamics in South-western Highlands Agro-Ecological Zone, SWHAEZ, baseline survey in Kabale, Kisoro, Rukungiri and Kanungu Districts. *International Journal of Fisheries and Aquatic Studies*, 1(5), 182-205.

About the author

Mr. Papius Dias Tibihika was born on 26th August 1976 in Rubuguri Parish, Kisoro District, South-western Uganda, Uganda. Papius is married to Ms. Kakama Doreen and the



have four children; Ayebare Valerie, Amutuheire Emmy, Mercy Leila Ampaire and Aheisibwe Keneddy. In 2007, Papius obtained a bachelor's degree in Fisheries and Aquaculture of Makerere University, Kampala-Uganda. After graduation, Papius started working in various organisations including, research assistant at Kajjansi Aquaculture Research and Development Centre, Kampala, Fisheries officer at Kabale Local Government, Uganda. In 2009, Papius received a scholarship from the Netherlands Fellowship Programme (NUFFIC) and started to pursue a Master of Science in Aquaculture and Fisheries from Wageningen University and Research Centre, the Netherlands. After graduation in 2011, Papius started working as a research officer and later a programme leader, livestock and fisheries research programme, at Kachwekano Zonal Agricultural Research and Development Institute (KAZARDI), under National Agricultural Research Organisation, Kabale Uganda. In 2015, Papius obtained another placement/job and started working as a Programme Assistant for Livestock and Fisheries under the Association for Strengthening Agricultural Research in Eastern and Central Africa (ASARECA). In 2016, Papius went back to KAZARDI and started working as a Research Officer under Aquaculture and Fisheries from where he embarked on a PhD project having received funding/a scholarship grant from Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH) under Austrian Partnership Programme in Higher Education and Research for Development (APPEAR). His PhD project was enrolled at the University of Natural Resources and Life Sciences Vienna (BOKU), Austria, at the Institute for Integrative Nature Conservation Research for three years. Papius intends to continue with his post-doc at BOKU, but at the same time working with his mother institution at KAZARDI, in Uganda.

Papius Dias Tibihika can be contacted via papiust@yahoo.com

Appendix

Supplementary materials for chapter three

Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*

Journal: Conservation Genetics

Table S1 Descriptions of the STRUCTURE HARVESTER output tables depicting the best K based on Evanno's methods. The highlighted rows show the largest value in Delta K corresponding to the best K. Tables a and b represent, respectively, to STRUCTURE HARVESTER outputs for whole amplicon information and amplicon length allele calling approaches.

a

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
1	10	-8032.99	1.261789	—	—	—
2	10	-7698.09	2.499533	334.9	193.25	77.31443
3	10	-7556.44	4.870136	141.65	428.19	87.92157
4	10	-7842.98	1169.358	-286.54	947.33	0.810129
5	10	-7182.19	4.146873	660.79	789.98	190.5002
6	10	-7311.38	54.32016	-129.19	169.93	3.128305
7	10	-7270.64	49.80047	40.74	142.36	2.858607
8	10	-7372.26	75.48525	-101.62	2.92	0.038683
9	10	-7476.8	94.61481	-104.54	20.28	0.214343
10	10	-7561.06	119.941	-84.26	—	—

b

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
1	10	-6685.770000	1.095496	—	—	—
2	10	-6401.870000	1.910236	283.900000	254.910000	133.444273

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
3	10	-6372.880000	4.443422	28.990000	79.790000	17.956881
4	10	-6264.100000	4.616877	108.780000	18.890000	4.091510
5	10	-6174.210000	38.198733	89.890000	62.720000	1.641939
6	10	-6147.040000	65.745557	27.170000	16.470000	0.250511
7	10	-6103.400000	5.450382	43.640000	266.470000	48.890149
8	10	-6326.230000	138.313204	-222.830000	77.310000	0.558949
9	10	-6471.750000	119.782463	-145.520000	64.690000	0.540062
10	10	-6552.580000	75.890913	-80.830000	—	—

Table S2 Allelic richness for the studied population based on 26 SSR loci. Figures with or without superscript * indicate allelic richness from amplicon length and whole amplicon information allele calling methods respectively

Loci	Albert		George		River Nile		Victoria	
Ti12	3.24	2.38*	3.04	1.37*	1.94	1.00*	4.64	1.81*
Ti16	1.00	1.00*	3.425	3.87*	1.417	1.542*	2.26	2.93*
Ti8	2.92	2.18*	1.29	1.38*	1.94	1.00*	2.85	1.84*
Ti34	12.73	8.93*	6.43	6.84*	6.67	6.40*	8.74	8.77*
Ti15	1.46	1.59*	2.60	2.00*	1.00	1.00*	1.88	1.98*
Ti2	5.07	6.50*	8.24	9.31*	4.84	5.37*	9.23	8.52*
Ti29	2.75	2.77*	4.36	4.53*	2.00	2.13*	4.81	3.98*
Ti1	2.45	2.57*	3.84	4.08*	2.85	3.53*	3.81	3.53*
Ti18	11.1	9.72*	11.05	9.35*	8.62	8.77*	10.41	10.08*
Ti4	5.56	5.79*	5.32	5.49*	3.95	3.99*	6.28	5.97*
Ti13	2.81	3.20*	3.71	3.84*	1.00	1.00*	4.46	4.08*
Ti26	4.85	5.19*	4.70	5.14*	3.69	3.88*	5.89	5.46*
Ti35	10.63	9.69*	9.24	7.58*	7.96	6.88*	8.55	9.05*
Ti28	1.69	1.00*	2.00	1.00*	1.00	1.00*	1.95	1.12*
Ti17	5.38	5.19*	6.43	7.03*	5.31	5.52*	6.22	6.37*
Ti5	3.34	2.57*	4.91	3.69*	3.36	1.99*	4.28	3.47*
Ti33	5.14	6.00*	5.67	6.02*	4.97	5.30*	6.54	6.74*

Ti31	5.82	4.72*	5.30	6.26*	3.00	4.00*	5.97	5.83*
Ti22	14.49	14.95*	12.56	14.91*	10.94	10.45*	12.06	15.36*
Ti7	7.79	8.56*	6.92	7.68*	4.51	4.98*	6.49	7.71*
Ti9	5.12	4.15*	3.50	3.79*	1.77	3.72*	2.56	3.37*
Ti14	4.71	2.99*	3.28	2.37*	3.73	2.00*	3.55	2.91*
Ti27	1.44	1.00*	1.00	1.00*	1.00	1.00*	1.67	1.00*
Ti6	4.38	4.96*	8.60	8.45*	2.67	2.80*	5.61	6.52*
Ti32	12.22	4.06*	9.86	3.69*	5.84	2.00*	12.43	3.36*
Ti24	11.31	8.44*	7.04	6.83*	10.22	7.87*	7.74	8.46*
Total	149.42	130.10	144.28	137.51	106.20	99.10	150.87	140.20

Table S3 Number of alleles per population based on 26 SSR loci. Figures with or without superscript * indicate number of alleles from amplicon length and whole amplicon information allele calling methods respectively

Loci	Albert		George		River Nile		Victoria	
Ti12	5	3*	4	2*	2	1*	8	4*
Ti16	1	1*	6	6*	2	2*	3	4*
Ti8	4	3*	2	2*	2	1*	5	4*
Ti34	20	10*	8	8*	8	7*	11	6*
Ti15	2	2*	5	3*	1	1*	2	3*
Ti2	6	8*	14	14*	6	6*	13	14*
Ti29	3	4*	6	8*	2	3*	7	6*
Ti1	3	3*	5	5*	3	4*	5	3*
Ti18	17	12*	19	12*	10	10*	15	10*
Ti4	7	6*	8	7*	4	4*	7	7*
Ti13	4	4*	4	4*	1	1*	5	6*
Ti26	6	6*	7	7*	4	4*	7	7*
Ti35	13	12*	16	11*	9	7*	11	10*
Ti28	2	1*	2	1*	1	1*	2	2*
Ti17	8	6*	10	9*	6	6*	10	8*
Ti5	4	3*	8	6*	4	2*	5	4*
Ti33	6	7*	7	7*	6	6*	7	7*
Ti31	7	5*	7	8*	3	4*	7	5*
Ti22	26	20*	19	23*	17	13*	19	20*
Ti7	10	10*	10	10*	6	6*	9	11*
Ti9	7	5*	5	5*	2	4*	3	3*
Ti14	5	3*	4	3*	4	2*	4	3*
Ti27	2	1*	1	1*	1	1*	2	1*
Ti6	7	7*	16	12*	3	3*	9	9*

Ti32	19	5*	17	4*	8	2*	19	3*
Ti24	18	10*	12	9*	14	9*	12	10*
Total	212	157	222	187	129	110	207	170

Table S4. Descriptions and procedures used in the in-house-scripts. We used the scripts one after the other and information concerning how this was done can be found in the material and methods.

Script number 1

File name: primer_demultiplex.py

Requirements: python 2 or 3; Biopython

Description: This script demultiplexes merged fastq files according to primer content. The outputs are one fastq file per sample and locus. This script allows for a user specified maximum number of mismatches between the primer and the reads. Only reads with a mismatch to both primers below to the defined are kept. In this case they are saved in a separate file. Moreover, sequences below a certain length can be excluded.

How to run: python extract_reads_correct_primer_merged.py [1] [2] [3] [4] [5]

[1] Directory containing input fastq files. Files should be named without the underscored character (ex: Sample1.fastq)

[2] File containing primer information: this should be a tab separated text file containing one locus per line with the following information:

maker name[TAB]sequence primer forward[TAB] sequence reverse primer

Markers should be named in the following way: MarkerName_RepetitionMotif (ex: HH1_AT)

[3] Maximum number of mismatches

[4] Minimum sequence length

[5] Directory to save output files. Files names are save in the following format:

RepetitionMotif_SampleName_MarkerName.fasq (ex: Sample1_HH1_AT.fastq)

Script number 2**File name:** CountLengths.sh**Requirements:** Unix system

Description: Per fastq file it counts the number of occurrences of each sequence length present. It outputs this information in a space separated text file being the first column the length and the second the number of occurrences. Example:

```
417  14
422  18
418  276
423  282
```

How to run: sh CountLengths.sh [1] [2]

[1] Directory containing input demultiplexed fastq files. Files should be as described in the output from script 1: RepetitionMotif_SampleName_MarkerName.fastq (ex: Sample1_HH1_AT.fastq)

[2] Directory to save output files. Output file is saved in the following format:

MarkerName_RepetitionMotif_SampleName statistics (ex: HH1_AT_Sample1_.statistics)

Script number 3**File name:** Rscript_Markerlength_develop_Color.R

Requirements: R version 3.4.1 or newer, packages reshape and reshape2 (Wickham 2007) will be installed automatically by the R script.

Description: This script uses sequence length counts to call alleles and define genotypes. A homozygote genotype is considered if the relative read count of the most abundant allele is larger than the user defined alpha value (e.g. 0.7). A heterozygote genotype is considered when the sum of the relative abundance of two most abundant read lengths exceeds the user defined alpha value

- (i) and if the difference in length of the potential alleles is larger than one time the repeat motif length.
- (ii) and if the two most abundant potential alleles only differ in one repeat size length, the less abundant allele is considered true if it is longer than the most abundant allele.
- (iii) and if the second largest potential allele is just one time smaller than the repeat motif length, it is only considered if its relative abundance equals 0.75 of the most abundant allele.
- (iv) in addition, we also considered point mutations which lead to non-integer multiples of the repeat length, if their relative abundance was at least 0.6 of the most abundant allele.

If these criteria are not met the sample is considered for manual control. Genotypes are saved in a comma separate file and length histograms are plotted in a pdf file highlighting the called alleles with different colors: blue for the homozygous allele, green for the alleles in heterozygous samples and black for the other sequence lengths.

How to run: `Rscript --vanilla Rscript_Markerlength_develop_Color.R [1] [2] [3] [4] [5] [6] [7] [8]`

- [1] Directory containing length counts. This is the output from script 2.
- [2] Minimum count of the total reads for a genotype to be considered.
- [3] Path to directory where pdf file with length counts histograms will be saved.
- [4] Alpha threshold for defining homo- and heterozygotes genotypes.
- [5] Path to Csv matrix file containing genotypes of markers of the respective samples
- [6] Path to Csv file containing three columns with sample and marker name and respective allele
- [7] Minimum allele size length for plotting the results
- [8] Maximum allele size length for plotting the results

References

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.

Script number 4

File name: extract_alleles.py

Requirements: python 2 or 3; Biopython.

Description: It extracts all sequences with the same length of the alleles saved in the csv file produced by script number 2 and saves them in a fasta file per sample and allele.

How to run: python extract_alleles_of_a_certain_length_v2.py [1] [2] [3]

[1] Coma separated text file containing genotypes. This file should contain a header containing the marker information. The next lines it contains the genotype information. The first column should contain the sample names while the remaining the genotype information. Two columns per markers should be added allowing for heterozygote genotypes. Missing data should be coded as 0. Example:

samplename	HH1_T					
	A	HH1_TA	HH2_TGT	HH2_TGT	HH3_AAAC	HH4_AAAC
Sample1	425	425	429	426	418	418
Sample2	423	425	429	429	0	0
Sample3	0	0	429	429	414	414
Sample4	427	427	426	426	418	418

[2] Directory containing input fastq files. These are the output from script 1.

[3] Directory to save output fasta files. These files are named in the following format:

MarkerName_SampleName_AI_SequenceLength.fasta (ex: HH1_AT_Sample1_AI_425.fasta)

Script number 5

File name: get_consensus_and_freq.py

Requirements: python 2 or 3.

Description: It produces a consensus sequence per file keeping bases above a certain similarity threshold. For positions where this is not met the script outputs a "N".

How to run: python get_consensus_and_freq.py [1] [2] [3]

[1] Directory containing input fasta extracted based on length genotype information (output from script 4)

[2] Directory to save the consensus files. Files will be named in the following format: MarkerName_SampleName_AI_SequenceLength_C Numer Of Sequences Used_Consensus Threshold.fasta (ex: HH1_AT_Sample1_AI_425_C881_70.fasta)

[3] Similarity of frequency threshold in an integer form. The value of 0.7 will do a 70% consensus.

Script 6

File name: correct_allele_sequence.py

Requirements: python 2 or 3

Description: In case a sequence has an ambiguous base ("N") after the consensus and it is homozygote based on amplicon length (AL), it divides the sequence into two new ones. The Ns are corrected based on the frequency that the N position shows up in the reads extracted from script 4 taking the two most frequent nucleotide combinations. In case of heterozygote genotype based on AL the sequence is not divided but only corrected with the most frequent nucleotide information.

How to run: python correct_allele_sequence.py [1] [2] [3] [4]

[1] fasta file containing all consensus sequences from one marker. File name should be MarkerName.fasta (ex: HH1_AT.fasta). Sequences should be named in the following format:

MarkerName_SampleName_AI_SequenceLength_CNumerOfSequencesUsed_ConsensusThreshold (ex: >HH1_AT_Sample1_AI_425_C881_70)

[2] Directory containing sequences extracted based on length genotypes (output from script 4)

[3] Minimum number of sequence counts required for an allele to be considered

[4] Name of the output fasta file

Script number 7

File name: call_alleles_from_fasta.py

Requirements: python 2 or 3.

Description: Uses the haplotypes obtained from the SNP correction process and converts them into allele's numbers. If a haplotype can be assigned to more than one allele it is saved as missing data. The results are saved in tab separated text file in the format of a codominant matrix (*matrix.txt). Allele's numbers and which haplotypes they correspond to are saved in file ending with *allele_list.txt with the following format:

Marker 1

Allele 1: Haplotype

Allele 2: Haplotype

...

Marker 2

...

How to run: python call_alleles_from_fasta.py [1] [2] [3] [4] [5]

[1] Directory containing haplotypes per locus in fasta format (output from script 7)

[2] Prefix common to all input files (ex: HH in marker HH1)

[3] List of samples names to be considered. This should be a text file with one sample name per line)

[4] Prefix that should be used to save output files

[5] Minimum number of sequences required to for a haplotype to be consider

Table S5 Results of Micro-checker tests, including all samples, for the frequency of null alleles per population and locus based on the two allele calling methods: whole amplicon information and amplicon length.

Popns	AN				GH				RNK				VKM¶			
Locus	Np¶	Np*	Oo¶	Oo*	Np¶	Np*	Oo¶	Oo*	Np¶	Np*	Oo¶	Oo*	Np¶	Np*	Oo¶	Oo*
Ti12	yes	yes	0.13	0.07	no	no	-0.06	-0.01	yes	no	0.22	0.00	yes	no	0.15	-0.06
Ti16	yes	yes	0.09	0.09	no	no	-0.01	-0.01	no	no	-0.02	-0.02	yes	yes	0.23	0.20
Ti8	yes	yes	0.11	0.14	no	no	-0.01	-0.01	no	no	-0.11	0.00	yes	yes	0.18	0.20
Ti34	yes	yes	0.04	0.04	no	no	-0.02	-0.02	no	no	-0.02	-0.03	no	no	0.05	0.00
Ti15	yes	yes	0.10	0.10	no	no	-0.09	-0.04	no	no	0.00	0.00	yes	yes	0.24	0.22
Ti2	No	no	0.03	-0.03	no	no	-0.04	-0.05	no	no	0.03	-0.13	no	no	0.02	0.00
Ti29	yes	yes	0.30	0.21	yes	yes	0.24	0.23	no	yes	0.17	0.16	yes	yes	0.30	0.28
Ti1	No	no	0.05	0.01	no	no	0.01	0.02	no	no	0.10	0.04	no	no	0.06	-0.05
Ti18	No	no	0.02	-0.03	no	no	-0.03	-0.03	no	no	0.04	-0.07	no	no	-0.04	-0.08
Ti4	No	no	-0.01	-0.02	no	no	-0.12	-0.12	no	no	-0.14	-0.11	no	no	0.01	0.00
Ti13	yes	yes	0.17	0.15	no	no	0.13	0.13	no	no	0.00	0.00	no	no	0.09	0.01
Ti26	yes	yes	0.08	0.06	no	no	0.02	0.02	no	no	0.13	0.13	yes	yes	0.22	0.19
Ti35	yes	yes	0.21	0.16	yes	no	0.13	0.07	yes	yes	0.21	0.15	yes	yes	0.20	0.16
Ti28	No	no	0.07	0.00	no	no	-0.14	0.00	no	no	0.00	0.00	no	no	0.12	-0.02
Ti17	yes	yes	0.08	0.08	no	no	0.01	0.00	no	no	0.01	0.01	yes	yes	0.15	0.12
Ti5	no	yes	0.05	0.09	no	no	0.12	0.09	no	no	0.07	0.12	no	no	0.00	0.09
Ti33	yes	yes	0.05	0.05	no	no	0.00	0.00	no	no	-0.01	-0.01	no	no	-0.06	-0.07
Ti31	yes	yes	0.35	0.26	yes	yes	0.37	0.25	yes	yes	0.29	0.26	yes	yes	0.33	0.26
Ti22	yes	yes	0.19	0.15	yes	yes	0.33	0.21	yes	yes	0.13	0.09	yes	yes	0.16	0.13
Ti7	no	no	-0.13	-0.15	no	no	-0.14	-0.14	no	no	-0.21	-0.27	no	no	-0.29	-0.25
Ti9	no	no	0.07	0.02	no	no	0.10	0.09	no	no	-0.04	-0.14	no	no	-0.10	-0.06
Ti14	yes	yes	0.11	0.12	no	yes	0.07	0.17	no	no	0.11	0.07	no	no	0.04	-0.25
Ti27	no	no	-0.01	0.00	no	no	0.00	0.00	no	no	0.00	0.00	no	no	-0.04	0.00
Ti6	no	no	-0.18	-0.23	no	no	-0.06	-0.08	no	no	-0.69	-0.75	no	no	-0.31	-0.32
Ti32	no	no	0.01	0.01	no	no	-0.01	0.00	no	no	0.00	0.02	no	no	-0.02	0.09

Ti24	yes	yes	0.04	0.05	no	no	0.03	0.02	no	no	-0.01	0.02	no	no	0.08	0.09
NL no	15	16			4	4			4	4			10	9		

NLnull=number of loci with null alleles, NP=Null allele present, ¶=Whole amplicon information, *=Amplicon length, Oo=null allele frequency based Van

Oosterhout et al. (2004), AN=Albert, GH=George, RNK=River Nile and VKM=Victoria

Table S6 Results of Micro-checker tests, excluding the three samples contributing for subpopulation structure from Lakes Albert and Victoria, for the frequency of null alleles per population and locus based on the two allele calling methods.

Populations	AN				GH				RNK				VKM			
Locus	NP [†]	NP*	Oos [†]	Oos*	NP [†]	NP*	Oos [†]	Oos*	NP [†]	NP*	Oos [†]	Oos*	NP [†]	NP*	Oos [†]	Oos*
Ti12	yes	no	0.14	-0.02	no	no	-0.05	-0.01	no	no	0.16	0.00	no	no	0.13	-0.07
Ti16	no	no	0.00	0.00	no	no	-0.01	-0.01	no	no	-0.02	-0.02	no	no	0.00	0.00
Ti8	no	no	0.12	0.00	no	no	-0.02	-0.02	no	no	-0.12	0.00	no	no	-0.02	0.00
Ti34	no	no	-0.02	0.05	no	no	-0.01	-0.01	no	no	-0.03	-0.03	no	no	0.02	-0.03
Ti15	no	no	-0.02	-0.02	no	no	-0.09	-0.05	no	no	0.00	0.00	no	no	0.00	0.00
Ti2	no	no	0.05	-0.07	no	no	-0.05	-0.06	no	no	0.04	-0.13	no	no	0.00	-0.01
Ti29	yes	no	0.29	-0.07	yes	yes	0.26	0.25	no	yes	0.15	0.17	yes	yes	0.31	0.28
Ti1	no	no	-0.08	-0.10	no	no	0.03	0.03	no	no	0.09	0.04	no	no	0.03	-0.09
Ti18	no	no	0.02	-0.06	no	no	-0.03	-0.03	no	no	0.05	-0.07	no	no	-0.06	-0.09
Ti4	no	no	0.03	-0.05	no	no	-0.12	-0.12	no	no	-0.13	-0.10	no	no	-0.05	-0.06
Ti13	no	no	0.14	0.14	no	no	0.12	0.12	no	no	0.00	0.00	no	no	-0.03	-0.26
Ti26	no	no	-0.12	-0.20	no	no	0.03	0.03	no	yes	0.13	0.13	yes	yes	0.20	0.18
Ti35	yes	yes	0.29	0.11	yes	no	0.13	0.08	yes	yes	0.23	0.16	yes	yes	0.19	0.15
Ti28	no	no	-0.05	0.00	no	no	-0.13	0.00	no	no	0.00	0.00	no	no	0.12	0.00
Ti17	no	no	0.12	0.15	no	no	0.01	0.01	no	no	0.00	0.00	no	yes	0.12	0.14
Ti5	no	no	-0.11	-0.10	no	no	0.09	0.09	no	no	0.06	0.12	no	no	-0.10	-0.18
Ti33	no	no	0.05	0.05	no	no	0.01	0.01	no	no	0.00	0.00	no	no	-0.06	-0.06
Ti31	yes	yes	0.33	0.23	yes	yes	0.36	0.25	yes	yes	0.29	0.26	yes	yes	0.30	0.26
Ti22	yes	yes	0.12	0.11	yes	yes	0.35	0.22	yes	yes	0.13	0.10	yes	yes	0.17	0.13
Ti7	no	no	-0.08	-0.11	no	no	-0.14	-0.14	no	no	-0.19	-0.25	no	no	-0.48	-0.55
Ti9	no	no	0.05	-0.18	no	no	0.10	0.09	no	no	-0.04	-0.15	no	no	-0.10	-0.06
Ti14	yes	yes	0.18	0.21	no	yes	0.08	0.17	no	no	0.13	0.06	no	no	0.04	-0.23
Ti27	no	no	-0.02	0.00	no	no	0.00	0.00	no	no	0.00	0.00	no	no	-0.05	0.00
Ti6	no	no	-0.54	-0.55	no	no	-0.07	-0.08	no	no	-0.68	-0.75	no	no	-0.57	-0.57
Ti32	no	no	-0.03	-0.05	no	no	0.00	0.01	no	no	-0.02	0.02	no	no	-0.05	0.07
Ti24	no	no	-0.04	-0.01	no	no	0.03	0.03	no	no	-0.01	0.02	no	no	0.02	0.04
NLnull	6	4			4	4			3	5			5	6		

NLnull=number of loci with null alleles, NP=Null allele present, ¶=Whole amplicon information, *=Amplicon length, Oo=null allele frequency based Van Oosterhout et al. (2004), AN=Albert, GH=George; RNK=River Nile and VKM=Victoria

Table S7 Tests for deviations of the Hardy-Weinberg Equilibrium (HWE) for heterozygosity deficit presented for two allele calling methods per population and locus. HWE probability (p) and Fis values with superscript '¶' and '*' indicate allele calling procedure based on whole amplicon information and amplicon length respectively, with all the samples included.

Populations	AN				GH				RNK				VKM			
locus	P¶	P*	Fis¶	Fis*	P¶	P*	Fis¶	Fis*	P¶	P*	Fis¶	Fis*	P¶	P*	Fis¶	Fis*
Ti12	0.00	0.02	0.65	0.66	0.10	-	-0.08	-	0.01	-	0.79	-	0.02	1.00	0.33	-0.02
Ti16	-	-	-	-	0.69	0.69	-0.02	-0.02	-	-	-	-	0.00	0.00	0.74	0.58
Ti8	0.05	0.02	0.37	0.50	-	-	-	-	1.00	-	-0.10	-	0.03	0.02	0.38	0.49
Ti34	0.10	0.00	0.03	0.17	0.62	0.63	-0.02	-0.02	0.74	0.81	-0.03	-0.04	0.03	0.33	0.11	0.02
Ti15	-	-	-	-	1.00	1.00	-0.04	-0.02	-	-	-	-	0.00	0.00	1.00	0.79
Ti2	0.01	0.08	0.18	-0.05	0.62	0.95	-0.05	-0.09	0.34	0.94	0.08	-0.20	0.55	0.70	0.04	0.00
Ti29	0.00	1.00	0.87	-0.02	0.00	0.00	0.59	0.58	0.08	0.02	0.38	0.50	0.00	0.02	0.64	0.67
Ti1	0.62	0.69	0.01	-0.03	0.04	0.04	0.06	0.07	0.21	0.44	0.21	0.07	0.32	0.74	0.12	-0.08
Ti18	0.13	0.87	0.06	-0.08	0.91	0.86	-0.04	-0.05	0.01	0.87	0.12	-0.10	0.90	1.00	-0.06	-0.13
Ti4	0.19	0.66	0.08	-0.05	0.95	0.96	-0.19	-0.19	0.86	0.79	-0.22	-0.17	0.01	0.00	0.04	0.02
Ti13	0.05	0.05	0.37	0.37	0.14	0.14	0.23	0.23	-	-	-	-	0.01	0.06	0.26	0.11
Ti26	0.84	1.00	-0.16	-0.28	0.26	0.26	0.06	0.06	0.00	0.00	0.34	0.34	0.00	0.00	0.45	0.39
Ti35	0.00	0.01	0.57	0.26	0.00	0.06	0.26	0.15	0.00	0.06	0.45	0.30	0.00	0.00	0.40	0.32
Ti28	1.00	-	-0.02	-	0.97	-	-0.25	-	-	-	-	-	0.21	-	0.35	-
Ti17	0.02	0.03	0.33	0.37	0.12	0.13	0.03	0.03	0.66	0.66	0.01	0.01	0.00	0.00	0.32	0.28
Ti5	0.89	1.00	-0.15	-0.06	0.20	0.22	0.19	0.13	0.07	0.21	0.18	0.35	0.02	0.00	0.08	0.26
Ti33	0.01	0.01	0.20	0.19	0.50	0.49	0.01	0.01	0.07	0.07	0.05	0.05	0.58	0.62	-0.08	-0.09
Ti31	0.00	0.00	0.68	0.47	0.00	0.00	0.78	0.54	0.01	0.00	0.78	0.60	0.00	0.00	0.71	0.57
Ti22	0.00	0.00	0.24	0.22	0.00	0.00	0.68	0.44	0.00	0.03	0.27	0.20	0.00	0.00	0.34	0.28
Ti7	0.97	1.00	-0.12	-0.17	1.00	1.00	-0.22	-0.22	0.99	1.00	-0.32	-0.41	1.00	1.00	-0.32	-0.32
Ti9	0.33	1.00	0.09	-0.10	0.02	0.04	0.22	0.19	-	1.00	-	-0.07	1.00	1.00	-0.05	-0.02
Ti14	0.00	0.04	0.41	0.46	0.01	0.01	0.20	0.43	0.13	0.43	0.21	0.17	0.03	1.00	0.09	-0.24
Ti27	-	-	-	-	-	-	-	-	-	-	-	-	1.00	-	-0.02	-
Ti6	1.00	1.00	-0.63	-0.64	0.86	0.92	-0.10	-0.13	1.00	1.00	-0.78	-0.85	1.00	1.00	-0.40	-0.42
Ti32	0.18	0.09	-0.02	-0.04	0.01	0.14	0.02	0.07	0.70	0.60	-0.01	0.06	0.25	0.22	-0.02	0.16
Ti24	0.93	0.51	-0.06	-0.01	0.47	0.54	0.05	0.03	0.65	0.24	0.00	0.07	0.01	0.01	0.18	0.21
NDL	9.00	9.00			8.00	6.00			6.00	3.00			16.00	12.00		
NLwi	3.00	4.00			2.00	4.00			6.00	7.00			0.00	2.00		

NDL= number of deviated loci from HWE, NLwi=Number of loci without information

Table S8 Genetic diversity (Ho and He) per population and locus based on two allele calling procedures. The superscript '†' and '*' represent, WAI and AL respectively.

Popns	AN				GH				RNK				VKM			
Locus	Ho [†]	Ho [*]	He [†]	He [*]	Ho [†]	Ho [*]	He [†]	He [*]	Ho [†]	Ho [*]	He [†]	He [*]	Ho [†]	Ho [*]	He [†]	He [*]
Ti12	0.09	0.04	0.24	0.12	0.60	0.03	0.55	0.03	0.04	0.00	0.19	0.00	0.36	0.12	0.52	0.12
Ti16	0.00	0.00	0.00	0.00	0.32	0.32	0.31	0.31	0.04	0.04	0.04	0.04	0.04	0.08	0.16	0.19
Ti8	0.19	0.05	0.29	0.09	0.03	0.03	0.03	0.03	0.21	0.00	0.19	0.00	0.12	0.08	0.19	0.15
Ti34	0.91	0.74	0.92	0.87	0.83	0.83	0.80	0.80	0.82	0.83	0.78	0.78	0.78	0.79	0.86	0.79
Ti15	0.05	0.05	0.04	0.04	0.18	0.09	0.17	0.09	0.00	0.00	0.00	0.00	0.00	0.04	0.15	0.18
Ti2	0.62	0.83	0.73	0.77	0.85	0.88	0.80	0.80	0.68	0.91	0.72	0.74	0.78	0.83	0.79	0.82
Ti29	0.05	0.14	0.37	0.13	0.21	0.14	0.49	0.33	0.32	0.04	0.50	0.08	0.21	0.12	0.56	0.35
Ti1	0.45	0.48	0.45	0.45	0.53	0.51	0.55	0.54	0.36	0.45	0.45	0.48	0.52	0.58	0.58	0.53
Ti18	0.86	0.95	0.90	0.87	0.94	0.91	0.89	0.86	0.78	0.95	0.85	0.85	0.94	1.00	0.87	0.86
Ti4	0.68	0.78	0.73	0.73	0.91	0.91	0.76	0.76	0.86	0.83	0.70	0.69	0.78	0.79	0.80	0.79
Ti13	0.13	0.13	0.20	0.20	0.46	0.46	0.58	0.58	0.00	0.00	0.00	0.00	0.39	0.46	0.52	0.50
Ti26	0.77	0.83	0.65	0.64	0.65	0.65	0.68	0.68	0.36	0.36	0.53	0.53	0.41	0.45	0.72	0.73
Ti35	0.40	0.64	0.88	0.84	0.63	0.61	0.84	0.70	0.47	0.53	0.80	0.73	0.52	0.55	0.85	0.78
Ti28	0.09	0.00	0.08	0.00	0.63	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.13	0.04	0.19	0.04
Ti17	0.48	0.43	0.69	0.67	0.74	0.74	0.75	0.75	0.75	0.75	0.74	0.74	0.46	0.48	0.65	0.65
Ti5	0.52	0.19	0.45	0.18	0.41	0.24	0.50	0.27	0.50	0.13	0.59	0.19	0.58	0.28	0.62	0.37
Ti33	0.62	0.64	0.75	0.76	0.76	0.76	0.76	0.76	0.71	0.71	0.73	0.73	0.91	0.91	0.82	0.82
Ti31	0.25	0.39	0.73	0.70	0.18	0.38	0.78	0.79	0.10	0.23	0.41	0.54	0.23	0.32	0.76	0.72
Ti22	0.73	0.73	0.93	0.91	0.30	0.52	0.89	0.90	0.67	0.71	0.89	0.87	0.63	0.68	0.92	0.92
Ti7	0.95	1.00	0.83	0.84	0.94	0.94	0.76	0.76	0.85	0.91	0.63	0.64	0.90	1.00	0.68	0.75
Ti9	0.50	0.33	0.53	0.30	0.26	0.25	0.32	0.30	0.08	0.27	0.07	0.24	0.20	0.12	0.18	0.11
Ti14	0.39	0.26	0.65	0.47	0.49	0.20	0.60	0.34	0.36	0.23	0.45	0.27	0.45	0.45	0.49	0.36
Ti27	0.04	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.08	0.00
Ti6	1.00	1.00	0.61	0.60	0.94	0.94	0.85	0.83	0.96	1.00	0.54	0.54	1.00	1.00	0.70	0.70
Ti32	0.95	0.48	0.92	0.45	0.86	0.40	0.87	0.42	0.73	0.29	0.70	0.30	0.96	0.28	0.92	0.33
Ti24	0.95	0.86	0.88	0.83	0.68	0.68	0.70	0.69	0.91	0.79	0.89	0.83	0.65	0.63	0.77	0.77
Ave	0.49	0.46	0.56	0.48	0.55	0.48	0.61	0.51	0.44	0.42	0.48	0.42	0.50	0.47	0.59	0.51
NL <0.45	11.00	13.00	8.00	10.00	9.00	12.00	5.00	11.00	13.00	14.00	9.00	12.00	11.00	11.00	6.00	11.00

Ave=average gene diversity (Ho & He), NL <0.45=number of loci per population with gene diversity (Ho & He) of less than 0.45, Popns=populations

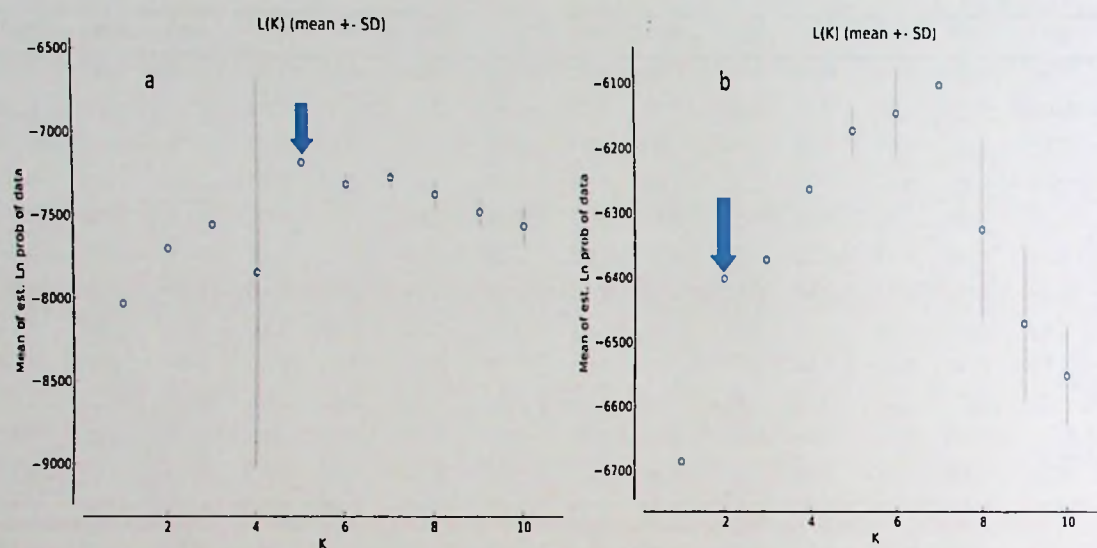


Figure S1 Mean Likelihood $L(K)$ and variance per K values from STRUCTURE summarized in STRUCTURE HARVESTER. The arrow represents the best K value: $K=5$. When the best K value approaches, $L(K)$ plateaus or increases slightly up to a high variance (Rosenberg et al. 2001). Figures a and b show STRUCTURE HARVESTER outputs from whole sequence information allele and amplicon length allele calling methods respectively

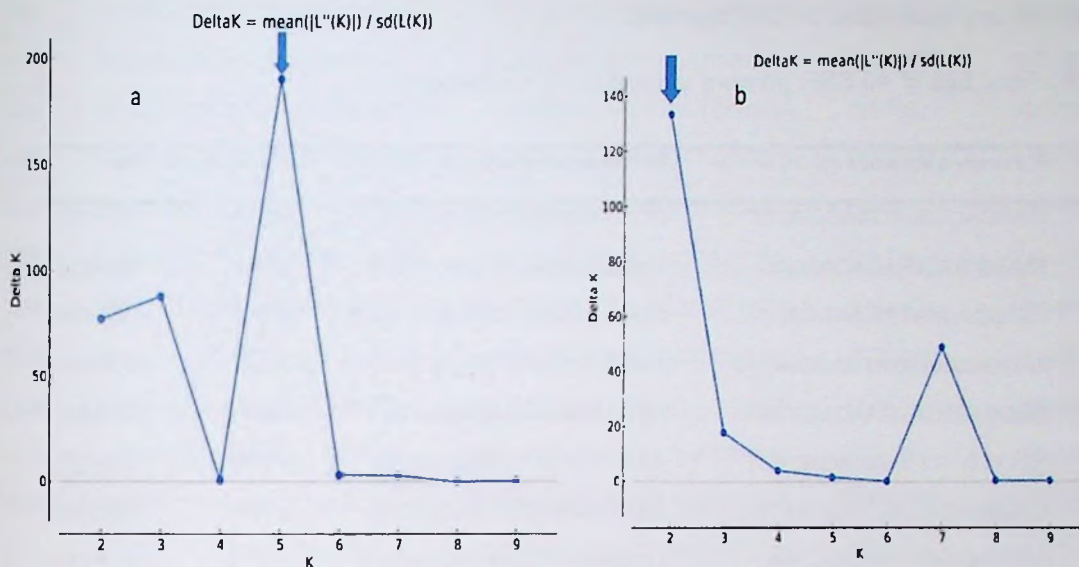


Figure S2 STRUCTURE HARVESTER analysis for the detection of the best K like based on Evanno et al. (2005). The best K is denoted by the highest Delta K value (arrow). The value is calculated based on the second order rate of change of the likelihood K (Evanno et al. 2005). Figures a and b show STRUCTURE HARVESTER outputs from whole sequence information allele and amplicon length allele calling methods respectively.

Supplementary materials for chapter four

Table S1 Final List of 40 SSR primers utilised in PCR reactions

Locus	F: Primer sequence (5'-3')	R: Primer sequence (5'-3')	Repeat motif	Asr
Ti1	TTATCACTGCTGAACGTCTT	GTTTTGGCTGCTACACATTC	(TG)10	393-411
Ti2	TTCTGGGCTAACACACAAG	AAGGTGTCACACAGTTTAGG	(CA)18	396-438
Ti4	TGTGCAGAAATAGATAGCCC	GAAAGGAAAAATGTTGGTGGT	(GT)18	384-436
Ti5	AAGGAGGATGATCAGGACAC	AGACCTCCACTGTGATCTTA	(CA)10	401-423
Ti6	CAGCTCTCATGAACACTTGA	ACCCATAAATCACACCAGTC	(GA)23	384-437
Ti7	TCTTTGTGTCAGAACTGTGT	ACTCTGCTTTTAGCCAATCA	(AC)17	385-430
Ti8	CTGAAGTCCTGCTGAGATTT	CATTGTTCTTGGCACCTCTA	(AC)15	400-446
Ti9	CTCAGTGACGAAGCCAAA	CCTGGCAATCAAAAGAACAA	(AC)10	409-419
Ti12	GCCACCAAAATATTCGTGTT	CCATGTTCTGTCTCCTTGAA	(TAC)12	377-431
Ti13	AATCCGTTAGCTGCAGATAG	GCTGATTAAACACAAAGTTGG	(ATG)10	369-427
Ti14	TCCCTAAATATGCCACCAA	TAGTGCTTTAATGGCTCTGG	(TAA)19	340-408
Ti15	GCTGTGATCATCTGGAGAAA	AGGATCTAGAACCTCCAACC	(TGC)10	385-406
Ti16	CAGACGTAGGCGATAAATCT	GAACACATCCATTTCCACAC	(AAC)10	368-428
Ti17	AACTGAAGAAGAAGCCTTGG	ATCATCTTCCTCTACTGCCT	(GAA)21	361-430
Ti18	AGCAAGTGAGATAAGCACTG	TACATAGCAGTGCAGTTTGC	(ATCT)8	373-461
Ti24	ACTGACAACATAAAGACATATGC	CACAGTTTGAATCCACCATC	(TTAC)9	403-483
Ti26	ATTGCTTCATCCCTTGAGTT	ACACGGAAAACCTAATGACA	(ACAA)10	402-475
Ti27	CTGTCTTTCTTGATGTGGGA	ATGCACAAATTTTAAGGGCC	(TTTG)6	391-405
Ti28	TGTCTTGGGATTTGAGATCA	CGGAGGTTTCTTCCTGTAA	(ATTCA)8	307-409
Ti29	AGGTCAAGATCAAGCAGTTT	CATCAACATAATTCAGTGTGGA	(TAAAA)10	321-400
Ti31	GAAACTATCCACAGAAGCCA	AGGCTTCTTACAGTTGGATG	(CTAAT)7	302-395
Ti32	CAGGAAATGGCTCCAAAATG	TTGTAGCTAGGAATCAGTGC	(AAAAT)7	325-361
Ti33	GCTTATGGCTGTATGGAGTT	CGACTTCTGTTGTGATTTGG	(TTCAA)6	382-427
Ti34	GCTTACAGTACATTGTGTGC	CTGATGAGAAAAACAGACGC	(TCTCT)10	355-448
Ti35	TCAACCACAAACTCCTCTTT	AAACTAAGTGCAGCTCATGA	(AAAAG)14	364-434
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	TCGAAGTAGCGTGGAACCT	ACAACAACAACAGGTCGGGA	(TGT)8	395-403
Ti50	CCTGTGACAGACTGGTGACC	ACACTGATGCGGTTTACGGT	(ATGG)7	442-517
Ti51	TGCTAAACGCCAGCTGATGA	TTACCACACGATGTCGCAGG	(TGT)8	401-428
Ti52	GAGAAACGTCCAGTGGCAGA	TTTCGATCTGCTGCCCCCTT	(TAT)8	373-429
Ti54	TTTCTTGCCAGCAAAACAGT	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	GAGCCGCCATAGTGCTCACTT	CCTGCTCTCACTCAAAGAGGG	(ATCC)7	473-516
Ti61	GCTACACAGGAAAGCAGAGC	ACTCAATGCTGGACGTGACC	(TGGA)6	474-501

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

Table S2 Variation of the final set of 40 SSR loci. Loci Ti1–Ti35 were taken from Tibihika et al. (28) while loci Ti39–Ti61 were developed in the current study

Loci	Na	Ho	He	Ht	PIC
Ti1	17	0.377	0.571	0.601	0.500
Ti2	54	0.704	0.854	0.864	0.844
Ti4	22	0.757	0.842	0.862	0.823
Ti5	19	0.438	0.644	0.671	0.614
Ti6	50	0.866	0.841	0.879	0.828
Ti7	35	0.843	0.820	0.833	0.801
Ti8	29	0.197	0.387	0.469	0.375
Ti9	10	0.156	0.597	0.589	0.558
Ti12	36	0.306	0.552	0.633	0.527
Ti13	12	0.224	0.449	0.485	0.431
Ti14	18	0.341	0.618	0.708	0.591
Ti15	14	0.138	0.417	0.461	0.381
Ti16	30	0.126	0.423	0.507	0.407
Ti17	64	0.661	0.820	0.861	0.804
Ti18	76	0.652	0.951	0.950	0.948
Ti24	68	0.654	0.879	0.900	0.873
Ti26	20	0.480	0.735	0.734	0.696

Ti27	7	0.060	0.325	0.390	0.310
Ti28	15	0.201	0.386	0.420	0.361
Ti29	39	0.230	0.718	0.808	0.675
Ti31	24	0.191	0.830	0.828	0.811
Ti32	84	0.938	0.931	0.929	0.926
Ti33	18	0.606	0.832	0.823	0.809
Ti34	79	0.811	0.940	0.946	0.936
Ti35	56	0.562	0.951	0.951	0.948
Ti39	33	0.384	0.644	0.686	0.615
Ti41	50	0.539	0.880	0.904	0.871
Ti43	61	0.364	0.654	0.729	0.647
Ti44	29	0.600	0.839	0.850	0.821
Ti49	24	0.303	0.381	0.427	0.370
Ti50	22	0.434	0.616	0.695	0.584
Ti51	40	0.545	0.759	0.789	0.737
Ti52	14	0.271	0.694	0.776	0.651
Ti54	47	0.677	0.920	0.932	0.914
Ti55	29	0.728	0.905	0.902	0.897
Ti56	24	0.200	0.747	0.784	0.721
Ti57	16	0.054	0.401	0.476	0.370
Ti59	28	0.596	0.803	0.835	0.778
Ti60	22	0.531	0.737	0.740	0.697
Ti61	17	0.127	0.259	0.307	0.250
Total	1,352		80%>0.5	82.5>0.5	78%>0.5

Na = number of alleles, Ho = observed heterozygosity, He = expected heterozygosity, Ht = total expected heterozygosity, PIC = Polymorphic Information Content.

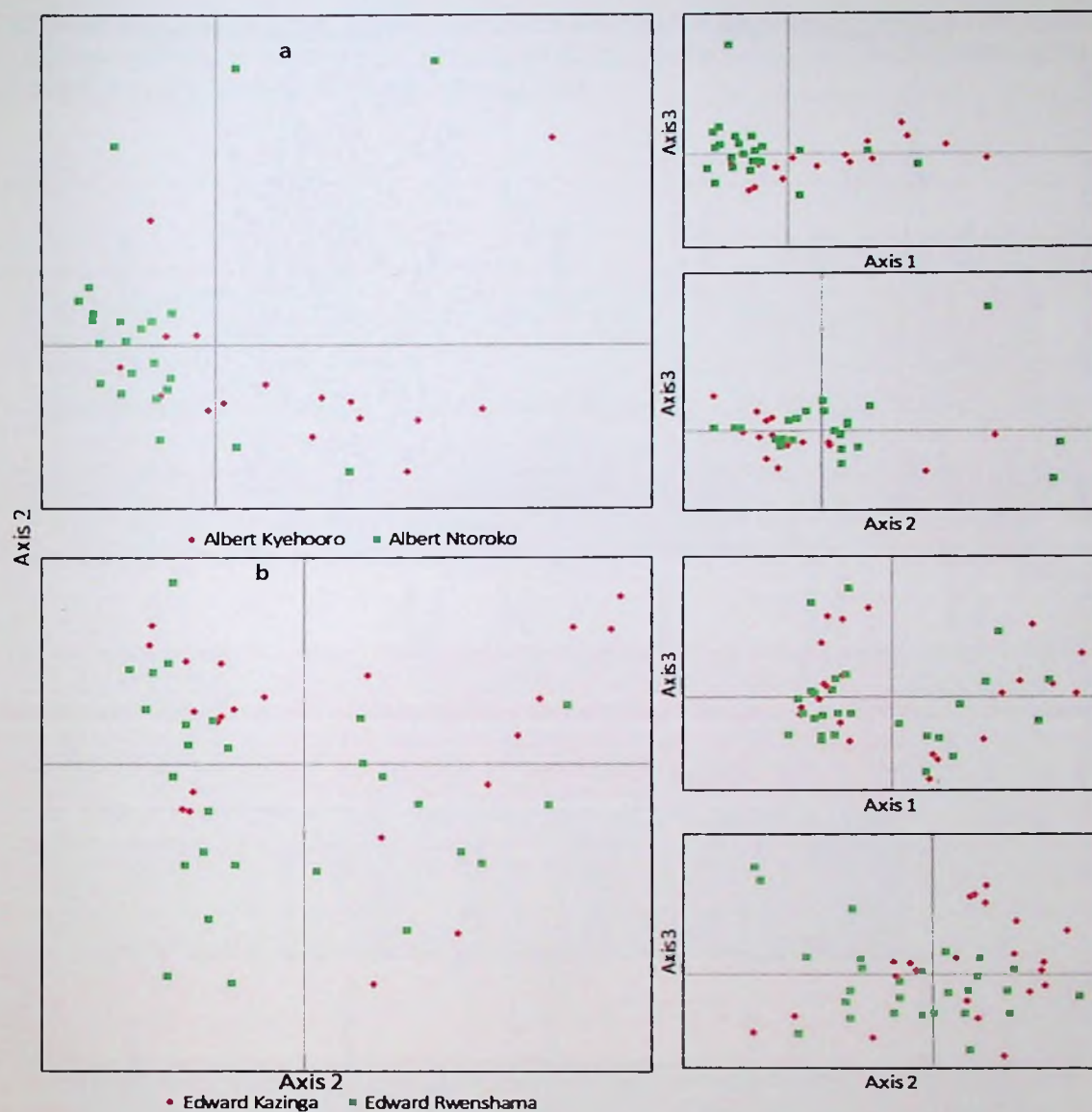


Figure S1 Genetic structure based on PCoA within the East African Western Rift Valley lakes, Albert (a) and Edward (b) metapopulations

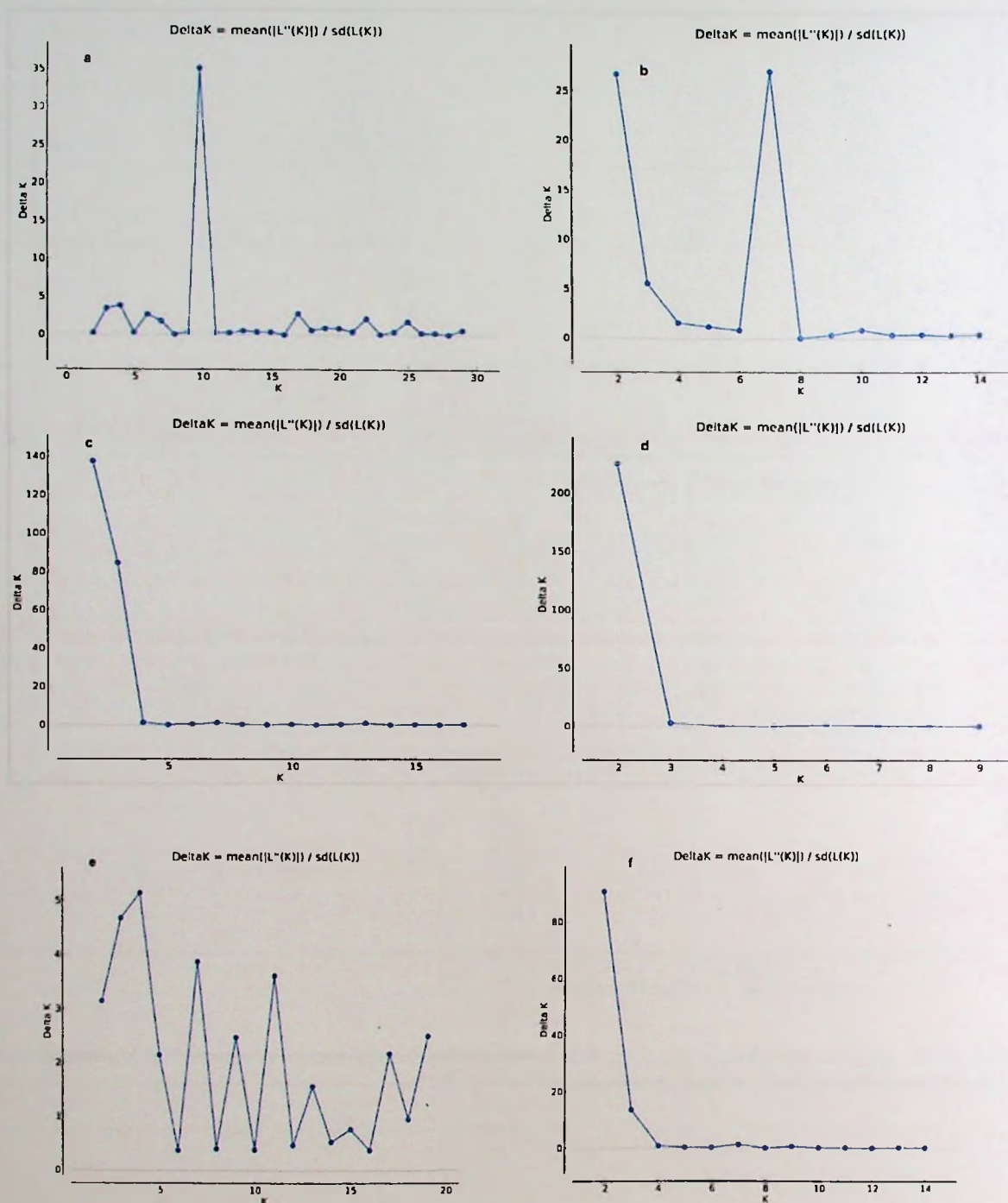


Figure S2 STRUCTURE HARVESTER analyses for depicting the optimal K values, which is derived from STRUCTURE results. a represents the best K for all populations (K=10), b best K for

all native populations ($K=7$), c best K for East African native populations ($K=2$), d best K for only Ugandan native populations ($K=2$), e best K for all Ugandan populations including farms ($K=4$) and f best K for Lake Victoria within populations ($K=2$)

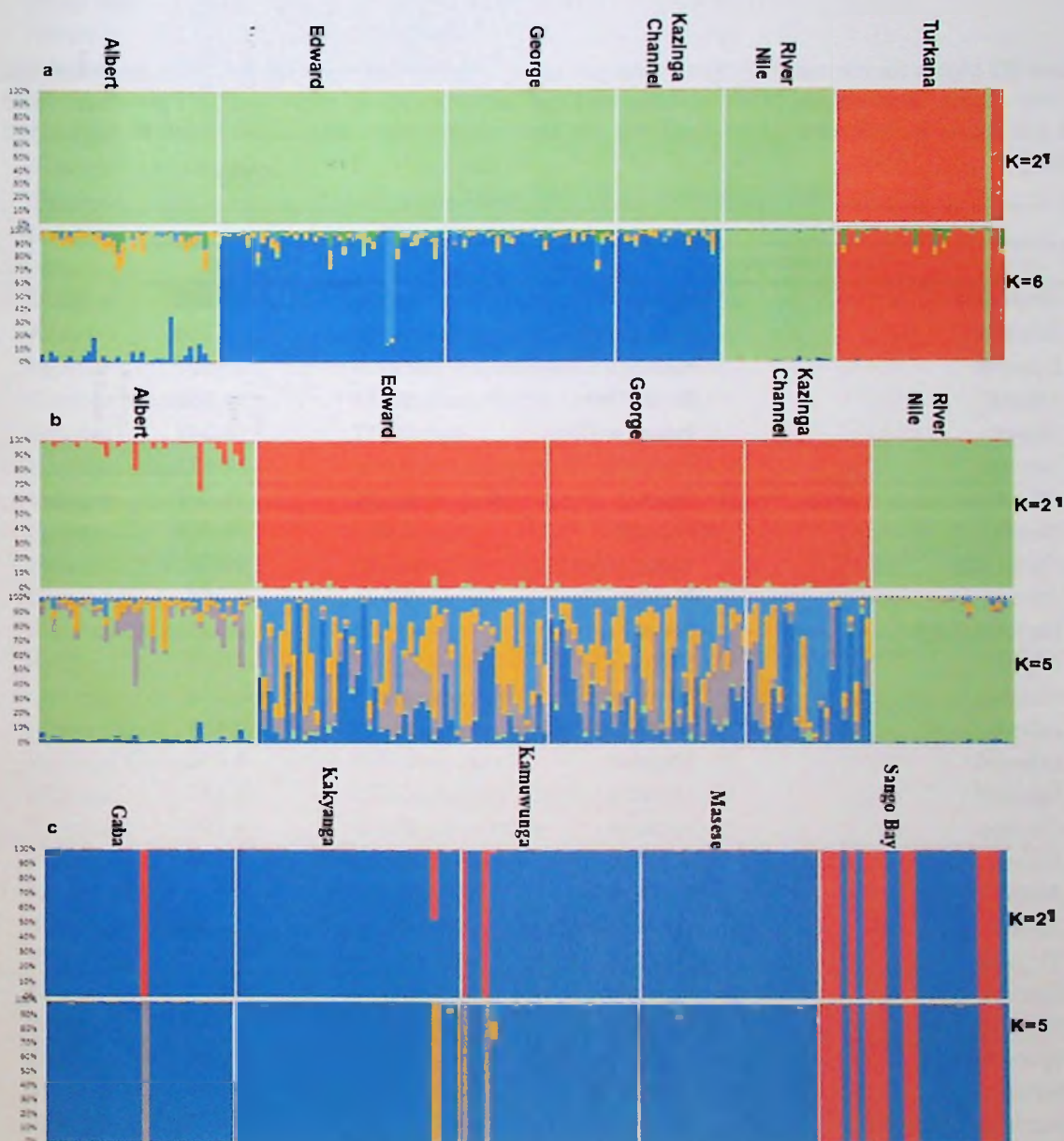


Figure S3 Bayesian clustering for genetic assignments of *O. niloticus* populations. **a** represents East African native, **b** Only Uganda native populations and **c** Lake Victoria subpopulations. K values with a superscript symbol ([†]) indicate the optimal K value based on STRUCTURE HARVESTER analyses.

Table S3 Matrix for the number of migrants per generation between population pairs as evidenced by Fst values. Bold values (>10) indicate vital number of migrants explained by Fst values: Note the variations in boldness which explains the level of important migrations between population pairs

Population pair		Fst	Nm
River Nile	Kyoga	0.009	26.495
Edward	George	0.010	23.754
Edward	Kazinga Channel	0.014	17.627
Kyoga	Sindi Farm	0.017	14.628
Albert	Bagenia Farm	0.017	14.341
George	Kazinga Channel	0.018	13.735
Albert	Sindi Farm	0.021	11.447
George	Mulehe	0.023	10.466
River Nile	Sindi Farm	0.025	9.878
Albert	Kyoga	0.025	9.604
Bagenia Farm	Sindi Farm	0.027	9.178
Albert	Victoria	0.027	8.972
Victoria	Sindi Farm	0.028	8.764
Albert	Rwitabingi Farm	0.029	8.431
Edward	Mulehe	0.029	8.418
Edward	Victoria	0.029	8.242
George	Victoria	0.030	8.129
Kyoga	Victoria	0.030	8.034
Albert	River Nile	0.031	7.837
Kyoga	Bagenia Farm	0.032	7.485
Victoria	Bagenia Farm	0.034	7.206
George	Kayumbu	0.035	6.965
Victoria	Mulehe	0.035	6.950
River Nile	Victoria	0.036	6.745
Edward	Kayumbu	0.036	6.649
Kazinga Channel	Mulehe	0.037	6.581
Rwitabingi Farm	Sindi Farm	0.037	6.568
River Nile	Bagenia Farm	0.038	6.398
Kazinga Channel	Victoria	0.038	6.319

Kyoga	Rwitabingi Farm	0.038	6.304
Victoria	Rwitabingi Farm	0.039	6.130
Rwitabingi Farm	Bagena Farm	0.040	5.967
Kazinga Channel	Kayumbu	0.041	5.904
Mulehe	Kayumbu	0.042	5.674
River Nile	Rwitabingi Farm	0.048	5.239
Albert	Edward	0.048	4.964
Albert	George	0.049	4.899
Victoria	Kayumbu	0.049	4.802
George	Bagena Farm	0.050	4.754
Edward	Bagena Farm	0.052	4.565
George	Sindi Farm	0.052	4.515
Edward	Sindi Farm	0.054	4.399
Albert	Kazinga Channel	0.055	4.288
Mulehe	Bagena Farm	0.055	4.266
Albert	Mulehe	0.056	4.185
Mulehe	Sindi Farm	0.057	4.160
Kazinga Channel	Bagena Farm	0.058	4.082
Edward	Kyoga	0.059	3.980
Kazinga Channel	Sindi Farm	0.061	3.826
Edward	River Nile	0.062	3.774
George	Kyoga	0.062	3.773
Edward	Rwitabingi Farm	0.068	3.444
George	River Nile	0.068	3.438
Kazinga Channel	Kyoga	0.068	3.405
Kyoga	Mulehe	0.068	3.405
George	Rwitabingi Farm	0.069	3.378
Albert	Kayumbu	0.072	3.243
Kazinga Channel	River Nile	0.072	3.232
Mulehe	Rwitabingi Farm	0.072	3.211
Kayumbu	Bagena Farm	0.073	3.184
Kayumbu	Sindi Farm	0.075	3.084
River Nile	Mulehe	0.076	3.057
Kazinga Channel	Rwitabingi Farm	0.076	3.035
Kayumbu	Rwitabingi Farm	0.089	2.561
Kyoga	Kayumbu	0.089	2.544
River Nile	Kayumbu	0.096	2.365
Ziway	Chamo	0.101	2.223
Victoria	Turkana	0.130	1.680
George	Turkana	0.140	1.541
Edward	Turkana	0.145	1.472
Albert	Turkana	0.148	1.435
Sindi Farm	Turkana	0.153	1.384

Bagena Farm	Turkana	0.153	1.379
Kazinga Channel	Turkana	0.155	1.363
Rwitabingi Farm	Turkana	0.157	1.342
Mulehe	Turkana	0.160	1.311
Kyoga	Turkana	0.162	1.295
Ziway	Hashenge	0.162	1.294
Kayumbu	Turkana	0.164	1.277
River Nile	Turkana	0.170	1.225
Hashenge	Chamo	0.217	0.902
Victoria	Burkina Faso	0.238	0.802
Turkana	Chamo	0.240	0.791
George	Burkina Faso	0.243	0.781
Edward	Burkina Faso	0.251	0.748
Turkana	Burkina Faso	0.261	0.707
Kazinga Channel	Burkina Faso	0.261	0.707
Turkana	Ziway	0.263	0.701
Mulehe	Burkina Faso	0.264	0.696
Kayumbu	Burkina Faso	0.273	0.667
Albert	Burkina Faso	0.275	0.659
Bagena Farm	Burkina Faso	0.279	0.645
Sindi Farm	Burkina Faso	0.287	0.620
Kyoga	Burkina Faso	0.293	0.603
Victoria	Chamo	0.298	0.590
Rwitabingi Farm	Burkina Faso	0.300	0.584
Turkana	Hashenge	0.300	0.583
River Nile	Burkina Faso	0.307	0.564
George	Chamo	0.312	0.551
Edward	Chamo	0.316	0.540
Victoria	Ziway	0.317	0.538
Sindi Farm	Chamo	0.331	0.506
George	Ziway	0.333	0.501
Mulehe	Chamo	0.335	0.497
Edward	Ziway	0.336	0.494
Kazinga Channel	Chamo	0.337	0.491
Kayumbu	Chamo	0.338	0.489
Albert	Chamo	0.339	0.487
Victoria	Tana	0.343	0.478
Bagena Farm	Chamo	0.345	0.475
George	Tana	0.348	0.469
Kyoga	Chamo	0.350	0.465
Rwitabingi Farm	Chamo	0.352	0.461
Mulehe	Ziway	0.356	0.452
Sindi Farm	Ziway	0.358	0.449

Kayumbu	Ziway	0.358	0.446
Edward	Tana	0.359	0.446
Kazinga Channel	Ziway	0.359	0.446
Albert	Ziway	0.364	0.436
River Nile	Chamo	0.364	0.436
Victoria	Hashenge	0.367	0.432
Turkana	Tana	0.368	0.429
Bagena Farm	Ziway	0.371	0.425
Kayumbu	Tana	0.371	0.424
Kyoga	Ziway	0.377	0.413
Rwitabingi Farm	Ziway	0.378	0.412
Albert	Tana	0.379	0.409
Mulehe	Tana	0.381	0.406
Bagena Farm	Tana	0.382	0.404
River Nile	Ziway	0.382	0.404
Kazinga Channel	Tana	0.384	0.401
Sindi Farm	Tana	0.387	0.395
Edward	Hashenge	0.391	0.389
Chamo	Burkina Faso	0.392	0.387
George	Hashenge	0.393	0.385
Rwitabingi Farm	Tana	0.401	0.374
Ziway	Burkina Faso	0.406	0.365
Sindi Farm	Hashenge	0.410	0.359
Mulehe	Hashenge	0.418	0.349
Albert	Hashenge	0.418	0.348
Kazinga Channel	Hashenge	0.419	0.347
Kyoga	Tana	0.420	0.345
Kayumbu	Hashenge	0.421	0.344
Kyoga	Hashenge	0.424	0.340
Bagena Farm	Hashenge	0.425	0.339
Rwitabingi Farm	Hashenge	0.426	0.337
River Nile	Tana	0.428	0.334
River Nile	Hashenge	0.441	0.317
Hashenge	Burkina Faso	0.480	0.271
Tana	Burkina Faso	0.487	0.263
Tana	Chamo	0.521	0.230
Ziway	Tana	0.548	0.206
Tana	Hashenge	0.599	0.167

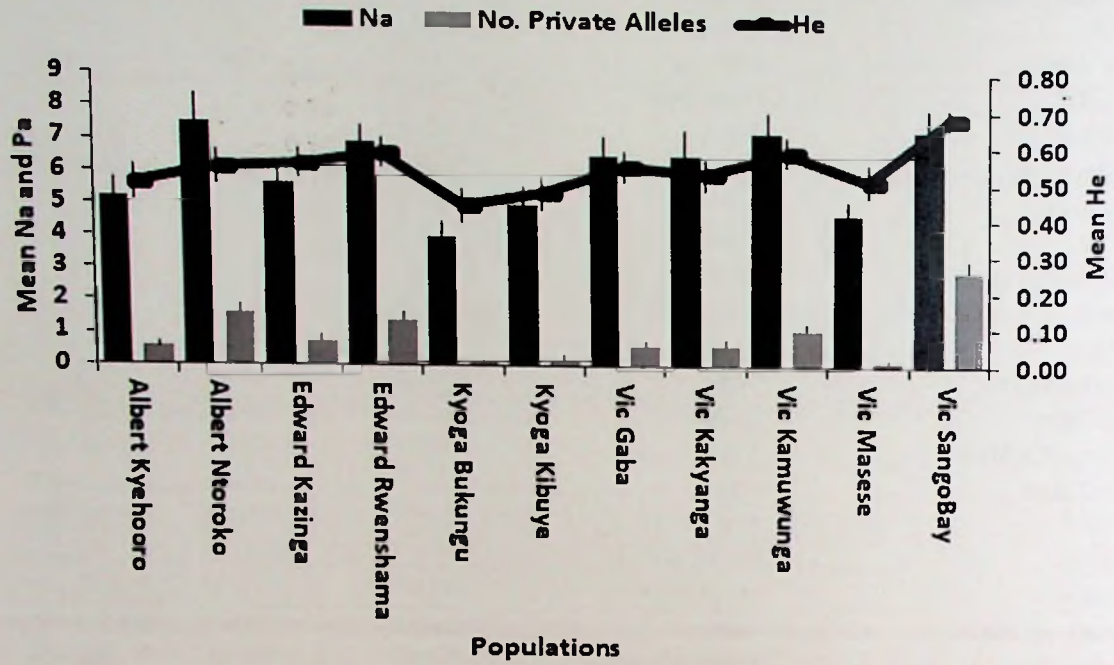


Figure S4 Within population genetic diversity of Albert, Edward, Kyoga and Victoria. Vic=Victoria, Na=number of alleles, Pa=Private alleles and He=expected heterozygosity