GENETIC STRUCTURE AND BIOGEOGRAPHY OF THREE COMMERCIALLY IMPORTANT AFRICAN FRESHWATER FISHES; *LATES NILOTICUS, BAGRUS DOCMAK* AND *BAGRUS BAYAD*

Thesis submitted by

Rose Komugisha Basiita

Bsc (Hons), MSc.

for the degree of Doctor of Philosophy

In the College of Science and Engineering

James Cook University

Australia

STATEMENT OF ACCESS

I, the undersigned author of this thesis, understand that James Cook University will make it available for use within the University Library, in other approved libraries and, via the Australian Digital Thesis network to allow access to users.

I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and I do not wish to place any further restriction on access to this work

Rose Komugisha Basiita

August 2016

DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in a text and a list of references is given.

Rose Komugisha Basiita

August 2016

…….

Professor Dean Jerry

(Primary Supervisor)

…….

A/Professor Kyall Zenger

(Secondary Supervisor)

CONTRIBUTION OF OTHERS

I was involved in the conceptualisation, experimental design and writing of all work presented in this thesis. I implemented all laboratory experiments including extractions, checks for quality, quantity and standardisation of DNA, carried out all polymerase chain reactions and PCR product clean ups prior to genotyping and sequencing at commercial laboratories which are duly acknowledged in all publications and this thesis. I was responsible for the data storage, management, integrity, analysis and interpretation. This included sequence trimming, alignment, genotype and allele scoring and checking for errors using various programs as outlined in the methods section for each chapter

Publications

Chapter	Publication and conference output on which Chapter is based	Contribution by others
Chapter 2	**Basiita, RK[1&2]**, Zenger, KR[1], Mwanja, MT[2] & Jerry, DR[1] (2015). Development of genome-wide microsatellite genetic resources in a commercially important African freshwater fish species–the Nile perch, *Lates niloticus.* Brief note: Animal Genetics.	Professor Dean Jerry was involved in the conception of this work, supervised all my data analysis and writing. He also read and reviewed manuscript prior to its submission to Animal Genetics. Associate Professor Kyall Zenger was involved in the supervision of data analysis and editing the manuscript. Dr. Matthew Mwanja provided most of the Nile perch samples and he also contributed to writing and editing of the manuscript.
	Basiita, Rose K.[1&2], Zenger, Kyall R.[1], Jones, David B. [1] & Jerry, Dean R[1]. (2015). Microsatellite discovery in a freshwater Siluroid, *Bagrus docmak* and their utility in a closely related species", Published under Microsatellites records volume 7, issue in Conservation Genetics Resources.	Professor Dean Jerry was involved in the conception of this work, supervised all my writing and data analysis and also read and reviewed manuscript. Associate Professor Kyall Zenger was involved in the supervision of the data analyses and editing the manuscript. Dr David Jones contributed expertise in laboratory and data analysis, as well as the editing of the manuscript.
	Basiita, RK[1&2], Zenger KR[1] and Jerry DR[1] 2015. *Novel Genetic Resources in an Iconic African Freshwater Fish, the Nile Perch, Lates niloticus*. International Symposium on Genetics	Professor Dean Jerry and Associate Professor Kyall Zenger supervised the data analysis and writing of this poster paper presented at the International Symposium on Genetics in Aquaculture, held in Spain.

	in Aquaculture (ISGA), 21-27 June 2015, Santiago de Compostela, Spain.	
Chapter 3	**RK. Basiita**[1&2], KR. Zenger[1] and DR. Jerry[1] (2014). *Genetic diversity and structure of Bagrus docmak from the Albertine rift and Lake Victoria basin, East Africa.* Oral Paper presented at the World Aquaculture Society Conference held in Adelaide, Australia, June 2014.	Professor Dean Jerry and Associate Professor Kyall Zenger supervised the data analysis and writing of this oral paper presented at the World Aquaculture Society conference in Adelaide, Australia.
	Basiita, RK[1&2], Zenger KR[1], and Jerry DR[1] 2016. Populations genetically rifting within a complex geological system: The case of strong genetic structure and low genetic diversity in the migratory freshwater catfish, *Bagrus docmak*, in East Africa. Submitted and is under Review in the Journal of Ecology and Evolution; Manuscript ID: ECE-2016-07-00677.	Professor Dean Jerry and Associate Professor Kyall Zenger supervised the data analysis and writing of this manuscript. They both read through and edited the manuscript priort o its summission to the Journal of Ecology and Evolution
Chapter 4	**Basiita, RK**[1&2], Zenger, KR[1], Mwanja, MT[2] &, Jerry DR[1] (2016). Restricted gene flow and strong genetic structure in Nile Perch, *Lates niloticus,* from African freshwater rivers and lakes. Manuscript prepared for submission to PLOS ONE.	Professor Dean Jerry was involved in the conception of this work, supervised all my writing and data analysis and also read and edited the chapter. Associate Professor Kyall Zenger was involved in the supervision of the data analyses and editing the chapter.
Chapter 5	**Basiita, RK**[1&2], Zenger, KR[1] & Jerry, DR[1] (2016). Historical and contemporary genetic structuring revealed in the African freshwater Bagrid catfish, *Bagrus bayad.* Manuscript prepared for submission to the Journal of Marine and Fresh water.	Professor Dean Jerry was involved in the conception of the work and supervised the execution and writing of this chapter. He read through all the drafts of the chapter. Associate Professor Kyall Zenger supervised the data analysis and read through the drafts of this chapter.

Affiliations

[1]Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Townsville QLD, Australia

[2]National Agricultural Research Organisation, National Fisheries Resources Research Institute, Aquaculture Research and Development Centre, Kajjansi, Uganda

ACKNOWLEDGEMENTS

I thank Professor Dean Jerry for guiding me throughout my entire PhD, but more so for believing in me. Your positive attitude and assurance kept me going in this journey and I am so grateful to have you as my primary supervisor, mentor and team leader. I am indebted to your constant reminder to have the big picture in mind, a principle that helped me to see the end more clearly from the beginning. Your quick response to written drafts was amazing, for instance, I would send you a draft on Friday hoping to take the weekend off, and you always amazed me with very quick responses …such as "hey Rose, while I was having coffee this afternoon I went over your draft and..……here it is". We, as your students are the envy of the University!

Thanks to Associate Professor Kyall Zenger, my secondary supervisor for being meticulous and detailed, always helping to fill in the questions that Dean left hanging, your assistance in data interpretation is highly appreciated. You and Dean make such a great team of supervisors and I can only say that I am honored to have been one of your students.

I would like to thank the NaFIRRI, Kajjansi staff; Constantine Ondhoro, Godfrey Kityo, Cassius Aruho, Martin Sserwada Philip Lwezaula, Sophie Hamba and Matthew Mwanja for the assistance in the collection of the fish tissue samples across Uganda. Professor Daye and Dr. Mouhamedene Seye are also recognized for the support and advice during sample collection from the Senegal River. It was a pleasure to work with you during my field work visit to Senegal. Dr. Babatunde and Dr. Eric Lugada for the support towards sample acquisition in Nigeria; you braved through the huddles of sample collection and transportation in a country where it almost became impossible for me to acquire samples.

Dr. David Jones and Ms Tansyn Noble are highly appreciated for the warm friendship, technical assistance and advice in the Lab, especially in regards to the initial primer design and testing and later on during the analysis phase of my PhD. Both of you made me feel welcome to Australia and also to belong. Dr. Erica Todd and Hoc Tan Dao are also thanked for their assistance in mitochondrial DNA analyses. Special thanks to Monal Lal, who read through earlier versions of Chapters; 3 and 5 of this thesis, as well as his invaluable input towards some figures in the thesis. Dr Greg Maes for all assistance received in interpreting Pop gen data analyses in my third year of the PhD. Also thanks to my lovely mentees and volunteers in MEEL, Ivany Argueta and Sandra Infante, who were involved in some of the DNA extractions in the laboratory.

My fellow students who provoked meaningful discussions throughout our PhD lives, thanks comrades this far we have all come, Monal Lal, Qyuen Quyen Banh, Shannon Keldjson, Alyssa Budd, Hoc Dan Tao, Giana Bastos Gomes, Ashton Gainsford, Maria Nayfa, Diana Pazmnio, Ali Rahzagi, Gobi, Geoffrey Collins, Donald McKnight, Eike Steinig, Jeff Colins, David Vaughan and Jarrod Guppy. I would also like to thank Ms Kristin Nunn for Organising Dean and making sure that supervisory meetings are arranged as well as all other administrative logistics. The post docs; Roger Huerlimann, David Jones, Jose Domingos, Heather Villeuix, Rebecca Lawton and Erica Todd. Thanks for trekking the journey before us, your guidance was highly appreciated. I cannot forget the immeasurable experience achieved while working with the rest of the eDNA team; Tansyn Noble, Heather Robson, Agnes Le Port, Prof. Damien Burrows, Roger Huerlimann, Julie Goldsbury, Prof. Colin Simpfendorfer and Prof. Dean Jerry, I am glad to have been and still feel part of this team.

The laboratory managers at MEEL, Dr Carolyn Smith-Keune and Ms Heather Welladsen, thank you for all the lab support during my PhD. As a student rep, I had the honour of working with the MEEL committee chaired by Lynne Van Herwerden, thanks for trusting me with such a huge, but very exciting responsibility. Thanks to Alyssa for taking on the mantle after me on the committee, I am pretty sure you will enjoy and benefit immensely from it.

Funding for this research was partially provided by AusAID through the CRC budget (sample collection in Senegal), ATAAS project in NARO (sample collection in Uganda) and the GRS scheme at JCU through which I had the honor of winning two grants (one research grant and one travel grant) during my candidature. AusAID and JCU funded my tuition fees for my candidature and extension respectively. All my stipend was provided for by AusAID under the ADS scholarship.

Thanks to NARO for releasing me from work to study and the AWARD fraternity for the inspiration to pursue my PhD studies. AWARD fellows and mentors, thank you for your great leadership contribution that stirred me to pursue my career goals as a driver of change in agricultural research and development. Profs Justus Rutaisire and Fred Bugenyi and Drs John Balilwa, Ogutu-Ohwayo, Fred Makumbi, Noah Kiwanuka, Gladys Bwanika, Vincent Muwanika and Charles Masembe, thank you for your individual input at the initiation of my PhD.

Thanks to my dear Pastors and spiritual leaders- Samson Basiita, Gerald Nyasulu, Sam Mukabi, Dan Kajubi and Williams Beshisha.

My only biological sister Irene Nuwagaba Kasirye, you are such a blessing along with whom I have acquired several other spiritual sisters; Phoebe Mugerwa, Ulemu Nyasulu, Rachel Kazum, Rachel Mwebia, Bunmi Malau-Aduli, Prisca Eboh, Titilayo, Pauline

Taurai, Miriam, Brenda Makumbi, Salley Kemei, Grace Othieno, Gertrude Atukunda, Margret Aanyu, Kate Ouma, Gofaene Mpusetsang, Peace Kankwasa, Josephine Kiwanuka, Dorcas Kibirige and the late Peruth Beshisha for always being there whenever I needed emotional and spiritual support. My older brothers to whom I looked up to; Dr. Herbert Turyagyenda and Musinguzi Richard. Thank you for setting the pace for me. I chose not to deviate from the high standards of integrity and success you two have set before me. To Dr and Mrs Seka Kiyingi, thank you for being a great support throughout this PhD journey.

My parents; Dad (the late Wilson G. Nsasiirwe) and Mum (Mrs Rachel Nsasiirwe) for always encouraging me to move forward in life and not be hindered by circumstances. It is with sadness though that I lost my dear dad during this PhD trek, I will miss you, but your words of encouragement and trust will always remain dear in my heart. As for mum, one thing I am always assured of is a prayer of favour and grace daily. I am privileged and proud to have had you both as parents.

My dear husband Pr. Samson Basiita for the deep relational love, care and always being there for me through thick and thin, without your loving support and patience, this PhD would have been impossible. Thank you for trusting and believing in me. My prayer is that you will realise the full potential of your dreams and I pledge to reciprocate it in your ministry journey and career path to the highest you have ever dreamed.

To my little heroes, thank you very much Elizabeth and Elijah Basiita for your interest and involvement in my PhD throughout my candidature. Elionora too is appreciated for always having a love heart drawing for me whenever I picked her up from the day care, thank you for your unwavering love. Our three children added a spackle and laughter to the journey

Greatest of all, to my Jehovah, the GREAT I AM, the one with the final say, thank you so much for being a loving father throughout this Journey and bringing in my life all the angels above to accomplish the purpose for which you created me. What can I say to my Lord, except, “THANK YOU LORD”.

DEDICATION

I dedicate this thesis to our children, Elizabeth Tirzah Mirembe Basiita, Elijah Mukisa Basiita and Elionora Mulungi Basiita. I pray this will be an inspiration to all your pursuits in your future academic lives. Remember, always you can do even greater things than this!

SUMMARY

The complex geology and hydrological evolution of Africa has resulted in a landscape characterised by some of the world's largest freshwater lakes and rivers. These freshwater habitats contain diverse fish fauna from which commercial inland fisheries support millions of people in Africa and other parts of the world through fish exports. Despite the extreme importance of African fisheries, versus increased threats that have led to their stagnation and decline, fisheries conservation and management based on genetic knowledge of stock structure has remained poorly studied and understood in Africa. It is clear though, that biogeography of fish fauna will often follow patterns of stock structure along gradients of interconnected, or fragmented habitats. Understanding the genetic structure and diversity of commercial fish species along these gradients in Africa is vital in identifying management units (MUs) for sustainable management and potential aquaculture development of aquatic species.

The widely distributed bagrid catfishes (*Bagrus docmak* and *B. bayad*) and the Nile perch, *Lates niloticus*, have been identified as model species that could be used to not only understand the biogeography of the African aquatic fauna, but are also important commercial fisheries and candidate species for aquaculture development. These fish species are indigenous to Africa and valued for food, as well as important components of recreational-based fisheries. These three species have been overexploited in their natural environments, with the Nile perch also in decline within environments where it has been translocated.

The overarching objective of this thesis was to develop and utilise new genetic resources to understand the genetic structure and biogeography of these three important African freshwater fishes and further understanding on how genetic population structure has been shaped by both the geology and anthropological (i.e., translocations), especially in the

Great Lakes Region of Africa. The knowledge acquired will be useful for the management and conservation of existing fisheries and aquaculture development programmes.

Firstly, to address the lack of genetic resources for important commercial fish species in Africa, high throughput Next Generation Sequencing (NGS) was used to generate novel genomic resources for two commercially important fish species, *L. niloticus* and *B. docmak*. Roche 454 GS-FLX shotgun sequencing was used to generate over 160,000 sequence fragments for *B. docmak,* from which hundreds of microsatellites markers with primers were *in silico* mined. Of these a subset of 20 novel microsatellite loci were tested in the laboratory from which 15 markers successfully amplified and were characterized in *B docmak* individuals from Lake Albert in East Africa (mean allelic richness (N_A) of 4.5, mean H_o, and H_e, of 0.535 ± 0.37 and 0.605 ± 0.028, respectively). These 15 loci also were found to cross-amplify in a sister species, *Bagrus bayad*. For the Nile perch, next generation sequencing on Roche 454 GS-FLX instrumentation yielded ~122,000 sequences, which have been deposited at in NCBI (accession# SRP051308), and allowed the design of primers for 285 *in silico* perfect microsatellite loci. Thirty-one microsatellite loci were tested for polymorphism. Additionally, 16 microsatellites previously characterized for *Lates calcarifer* (a sister species) were tested in *L. niloticus*. The successful development of the microsatellite markers *de novo* in these species were used to resolve the stock structure of these species in the ensuing data chapters of the thesis. Furthermore, although not evaluated herein, these markers will also have utility in parentage analysis in future breeding programs for aquaculture development.

Whilst there is some understanding of how macroevolutionary drivers have shaped teleost speciation in East Africa, there is a paucity of research into how the same biogeographical factors have affected microevolution (ie evolution of populations within a species). To

address this deficiency, population genetic diversity, demography and structure were investigated in the widely distributed and migratory (potamodromous) African teleost species, *Bagrus docmak*. Samples were acquired from five geographical locations in East Africa, within two major drainage basins; the Albertine Rift and Lake Victoria Basin. Individuals (N = 175) were genotyped at 12 microsatellite loci and 93 individuals sequenced at the mitochondrial DNA control region. Results suggested populations from Lakes Edward and Victoria had undergone a severe historic bottleneck resulting in very low nucleotide diversity ($\pi = 0.004$ and 0.006 respectively) and negatively significant *Fu* values (-3.769 and-5.049, at $P < 0.05$). High genetic structuring between drainages was detected at both historical (mtDNA ϕ_{ST} = 0.62, P = 0.000) and contemporary (microsatellite F_{ST} = 0.46, P = 0.000) levels. Patterns of low genetic diversity and strong population structure revealed are consistent with speciation patterns that have been linked to the complex biogeography of East Africa, suggesting that these biogeographical features have operated as both macro- and micro-evolutionary forces in the formation of the East African teleost fauna.

Thirdly, in order to understand the population genetic structure of Nile perch, an iconic species of high commercial importance, both in the species' native range and where it has been translocated, Nile perch tissue samples were acquired from two West (Senegal River and Lake Kainji on the Niger River) and four East African (Lakes Albert, Kyoga, Victoria and Turkana) locations. Nineteen polymorphic microsatellite loci were used to study the genetic variation among populations across regions (West and East Africa), as well as between native and introduced populations within East Africa. Results revealed strong and significant genetic structuring among populations across the sampled distribution (divergence across regions, F_{CT} = 0.26, P =.0.000). STRUCTURE analysis at a broad scale revealed K.=.2 clusters, whereby Nile perch from West Africa were

assigned to one genetic cluster, while all individuals from East Africa regardless of whether native or introduced populations were assigned to another genetic cluster. Analyses at a regional scale revealed further structuring of up to K = 3 genetic clusters in East African Nile perch. Lower genetic diversity based on analysis of allelic richness (A_R) was observed for the two translocated populations of Lake Kyoga ($A_R = 3.61$) and Lake Victoria ($A_R = 3.52$), compared to Nile perch populations from their putative parental populations of Lakes Albert ($A_R = 4.12$,) and Turkana ($A_R = 4.43$). The lower genetic diversity in the translocated populations may be an indication of previous bottlenecks and may also indicate a difficulty for these populations to persist and adapt to climatic changes and anthropogenic pressures that are currently present in the East African region. On the contrary, except for Nile perch from Lake Turkana, the Nile perch populations examined indicated signs of homozygote excess with positive and significant F_{IS} values ($P < 0.05$), a critical finding useful in identifying genetically diverse aquaculture foundational stocks. Finally, the distinct genetic clusters identified in the current study between the West and East African Nile perch have been maintained by presence of biogeographic barriers and restricted gene flow between the two regions and, as such, the two genetic groupings should be possibly managed separately as two fishery stocks.

Lastly, the biogeographic relationships of a congener catfish, *B. bayad,* among drainages of Lake Albert in East Africa, Lake Kainji on the Niger River and the Senegal River in West Africa were also explored in the current study. Present and historical connectivity of *B. bayad* from these three drainages were investigated using 223 bp of the mitochondrial DNA d-loop region and nine polymorphic microsatellites in order to define management and conservation units for the species. Mitochondrial DNA via haplotype networks and phylogenetic tree analysis revealed two haplotypic clades, whereby the West African populations were divergent from the East African population. Overall,

haplotype diversity ($H_d \pm SD = 0.87 \pm 0.03$) was high across all sampled *B. bayad* individuals. Comparably higher genetic diversity was observed in the *B. bayad* individuals from Lakes Albert and Kainji, whilst the lowest diversity was from those individuals sampled from the Senegal River ($H_d \pm SD = 0.27 \pm 0.11$). Bayesian analyses using nine microsatellites identified three genetic clusters (K = 3), grouping *B .bayad* by drainage. Thus, whereas the mitochondrial results highlight the existence of two evolutionary significant units, the microsatellite data seems to resolve the stock structure of the species into three contemporary genetic groupings. Therefore, based on these findings, three fishery management units are recommended to take into account the three genetic populations that appear to correspond to the drainages under study. Separate management units for identified genetic groupings will be important in drawing strategies specific to each drainage system, as varying levels of genetic diversity were also identified.

Overall this thesis provides the first comprehensive set of species-specific genomic resources developed for the freshwater African bagrids (*B. docmak* and *B. bayad*) and the Nile Perch, *L. niloticus*. Using the genomic resources developed, the three freshwater species examined, showed strong phylogeographic and genetic structure patterns that enabled the successful identification of management units (MUs). The identification of these management units will be useful in the formulation of future conservation strategies and identification of stocks suitable for aquaculture development for these species. The molecular data generated herein is central to the management of freshwater fisheries in Africa, especially West and East Africa (that have been identified each as a unique genetic group) and contributes to the understanding of microevolutionary forces in shaping the evolution of the African freshwater fish fauna.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

General Introduction

1.1 Drainage evolution in Africa

Africa is the second largest continent and possesses many of the largest and longest rivers on the planet. Despite the importance that rivers and freshwater environments play in Africa, hydrological evolution in Africa remains largely enigmatic, with limited studies examining the evolution of river formation across the broader continent (Goudie et al., 1992, Stewart, 2001). Those studies that have examined drainage evolution have primarily focused on the Eastern, Central and Southern African drainages, with even fewer studies focusing on the evolution of West African rivers (Issawi and McCauley, 1992, Stewart, 2009, Machado-Schiaffino et al., 2015, Hugueny and Lévêque, 1994, Hugueny, 1989, Hamilton, 1982). What is evident is that the current major drainage networks that exist in Africa (such as the Niger, Senegal, Congo, Chad and the Nile Rivers) have arisen from complex geological processes, including tectonic uplifting, continental drifting, shearing and warping (Stewart, 2009, Stewart, 2001, Stankiewicz and de Wit, 2006, Haddon and McCarthy, 2005).

The present day hydrology of Africa consists of discontinuous and fragmented river and lacustrine water systems. This is opposed to what was once likely to be a continuous habitat as indicated by the historically uniform fauna that existed for instance during the Holocene, a time characterized by high humidity with interconnecting waterways. (Stankiewicz and de Wit, 2006, Stewart, 2009, Stewart, 2001, Petit-Maire, 1989, Petit-Maire et al., 1983). During this Holocene period, diverse aquatic fauna, including fish (e.g. catfish and the Nile perch, *Lates niloticus*) occupied present day fossil sites in the Nile valley, as well as northern Africa; an occurrence that has been linked to the presence of permanent water pathways that enabled the exchange of fauna across drainages

(Stewart, 2009, Petit-Maire et al., 1983). This connectivity through permanent water ways possibly explains the uniformity in faunal assemblages in these drainages(Hugueny and Lévêque, 1994). As the geology of the African continent has continued to evolve into its present form these major river drainages have become separated by significant biogeographic barriers, including expanses of deserts and mountains; nevertheless, a large number of species are still shared among these drainages (Stewart, 2009).

Few studies have attempted to explore drainage evolution in Africa using patterns in present day aquatic fauna. Those studies that have been conducted mainly concentrated on the aquatic faunas of the Nile, Congo and Chad River systems and associated lakes (Stewart, 2009, Stankiewicz and de Wit, 2006, Daniels et al., 2015, Danley et al., 2012, Odada and Olago, 2006). For example, a recent and broader study based around freshwater crabs from the Family Potamonautidae revealed a historical connectivity of drainages in Africa. In this study, specimens from the Nilo-Sudan and East African coast were found to be closely related to specimens from upper Guinea in West Africa (Daniels et al., 2015). However, gaps still exist in the understanding of evolution of the aquatic fauna using fossil records, since the identification of fossils within the river basins have been limited to genus and rarely to species (Greenwood, 1966, Stewart, 2009). Additionally, the rarity of fossils in some areas, such as West Africa, which has a very poor fossil record, makes it difficult to understand the evolution of these drainages from a historical perspective (Lévêque, 1997).

1.2 Geological evolution in Africa

Complex processes, such as uplift, volcanism, desert formation and tectonic rifting are responsible for the creation of Africa's distinct biogeographic features e.g. waterfalls and cascades; catchment divides and major mountain ranges (Wood and Guth, 2013,

Haddon and McCarthy, 2005, Rahel, 2007). The presence of these biogeographic barriers is likely to have impeded gene flow, leading to population differentiation and species formation. In addition, there has been substantive anthropogenic mediated movement in Africa of aquatic fauna (i.e., Nile perch) which have shaped the biodiversity in the region and directly impacted the genetic diversity and structure of the aquatic freshwater teleost fauna (Rahel, 2007, Mwanja et al., 2012, Hauser et al., 1998).

Within East Africa, tectonic rifting and uplifting are major processes that have been responsible for the formation of several water bodies, as well as the larger East African Rift (EAR) valley system. The water bodies include all EAR lakes (such as Albert, Turkana, Edward, George, Tanganyika and Malawi), Kazinga Channel and the Rivers Nile, Kagera, Katonga, Semliki (see Figure 1.1). Other major rivers that were affected by the regional uplift, were the Congo and Zambezi river systems (Roberts et al., 2012). The largest of the Great Lakes, Lake Victoria, lies outside the EAR and, as such, various processes have been proposed for the formation of Lake Victoria, including geological down warping; however most evidence points to tectonic uplifting associated with rifting around the craton margins causing ponding as the primary force responsible for its formation (Rach, 1992, Ominde, 1971).

Essentially, the EAR comprises of the Eastern and the Western side of the rift (also referred to as the Albertine Rift). Evidence suggests that the eastern side of the rift valley was formed earlier than the western side of the rift (Wolfenden et al., 2004, Cohen et al., 1993b). Recent studies on the EAR and its effects on the geology and evolution of species in Africa have revealed that the uplift of Eastern Africa was more widespread and synchronous between the Eastern and Western rifts of the EAR system than had been previously perceived (Roberts et al., 2012). All the above processes led to formation of

complex and fragmented landscapes and habitats and these, coupled with climatic changes, have directly impacted the evolution of fauna in Africa (Roberts et al., 2012).

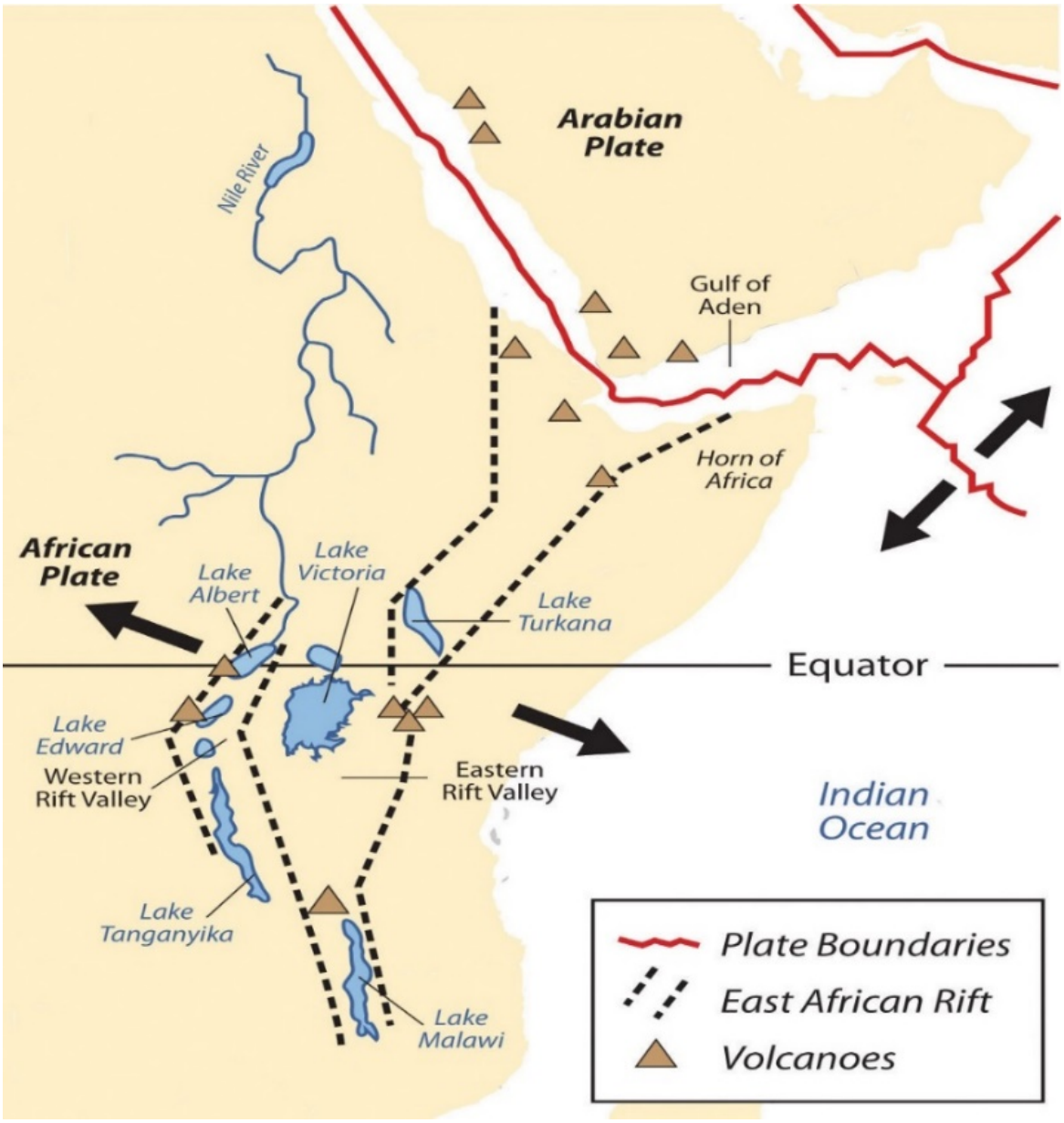


Figure 1.1: The East African Rift Valley and the major Lakes and rivers that reside within the Rift Valley. Source: USGS, http://pubs.usgs.gov/gip/dynamic/East_Africa.html

During the late Neogene (8 to 2 million years ago), the climate of East Africa was considered quite dynamic, oscillating between wet and dry intervals that have consequently impacted water levels in the Great Lakes Region (Cohen et al., 1997, Cohen

et al., 1993a, Johnson et al., 1996, Johnson et al., 1987). These climatic changes in the geological complex have been associated with speciation and dispersal in higher order taxa, including hominoids (Sepulchre et al., 2006). Freshwater fish fauna exchange among Lakes Edward, Turkana and the Nile River during wet periods of the Holocene have also been reported (Stewart, 2009). Similarly, during one of the arid periods around 12,000 to 18,000 years ago, Lake Victoria underwent total desiccation, a process that is partly responsible for biodiversity changes of freshwater teleost fauna in the region (Elmer et al., 2009, Seehausen, 2002). It is not well understood how these arid and humid climates interacted together with the present biogeographic barriers and connections to maintain or restrict gene flow among most freshwater teleosts.

While there is some degree of understanding of how macroevolutionary drivers (ie. at the species level or above) have shaped teleost speciation in Africa, there are gaps in research on how the same biogeographical factors have affected microevolution (ie. evolution within and among populations) (Carvalho and Hauser, 1995). Within the African freshwater fauna a significant amount of research has been devoted to understanding the evolution and endemism of large groups of fishes such as cichlids. The focus on the above species level research has been mainly due to the difficulty in confidently identifying discrete populations (Carvalho and Hauser, 1995). In this regard, the cichlids of Africa present a rare, but good example, where the adaptive radiation of fish in the Great Lakes Region (including Lake Victoria, Tanganyika and Malawi) has been well documented (Danley et al., 2012, Demenocal, 2004, Elmer et al., 2009, Johnson et al., 1996, Kocher, 2004, Lowe-McConnell, 2009, Meyer, 1993, Takahashi and Koblmüller, 2011). However, few other studies have focused on evolution at below the species level posing a challenge in drawing management plans to protect unique and cryptic populations that may exist.

1.3 State and importance of inland freshwater fisheries

The alarming rate of decline of freshwater biodiversity (and associated exploited fisheries) is a concern owing to the compartmentalisation of freshwater habitats, where often low levels of drainage interconnectivity are evident (Ricciardi and Rasmussen, 1999). Inland fisheries are exceptionally vulnerable, as often each drainage system will possess unique fauna that cannot be replaced via immigration from other drainages. According to the FAO (2014), most countries in sub-Saharan Africa depend on inland fisheries (Figure 1.2). For instance the fisheries sector supports up to 4 million people in East Africa. In Uganda, as a commodity, fisheries is the second highest foreign exchange earner next to coffee, whereby the earnings from the fisheries sector reached a peak of USD 147 million in 2005. Since this time, however, earnings from fishing in Uganda have been on the decline as a result of dwindling and overexploited wild stocks (Mkumbo and Marshall, 2015, Jaabi and Rasiah, 2014).

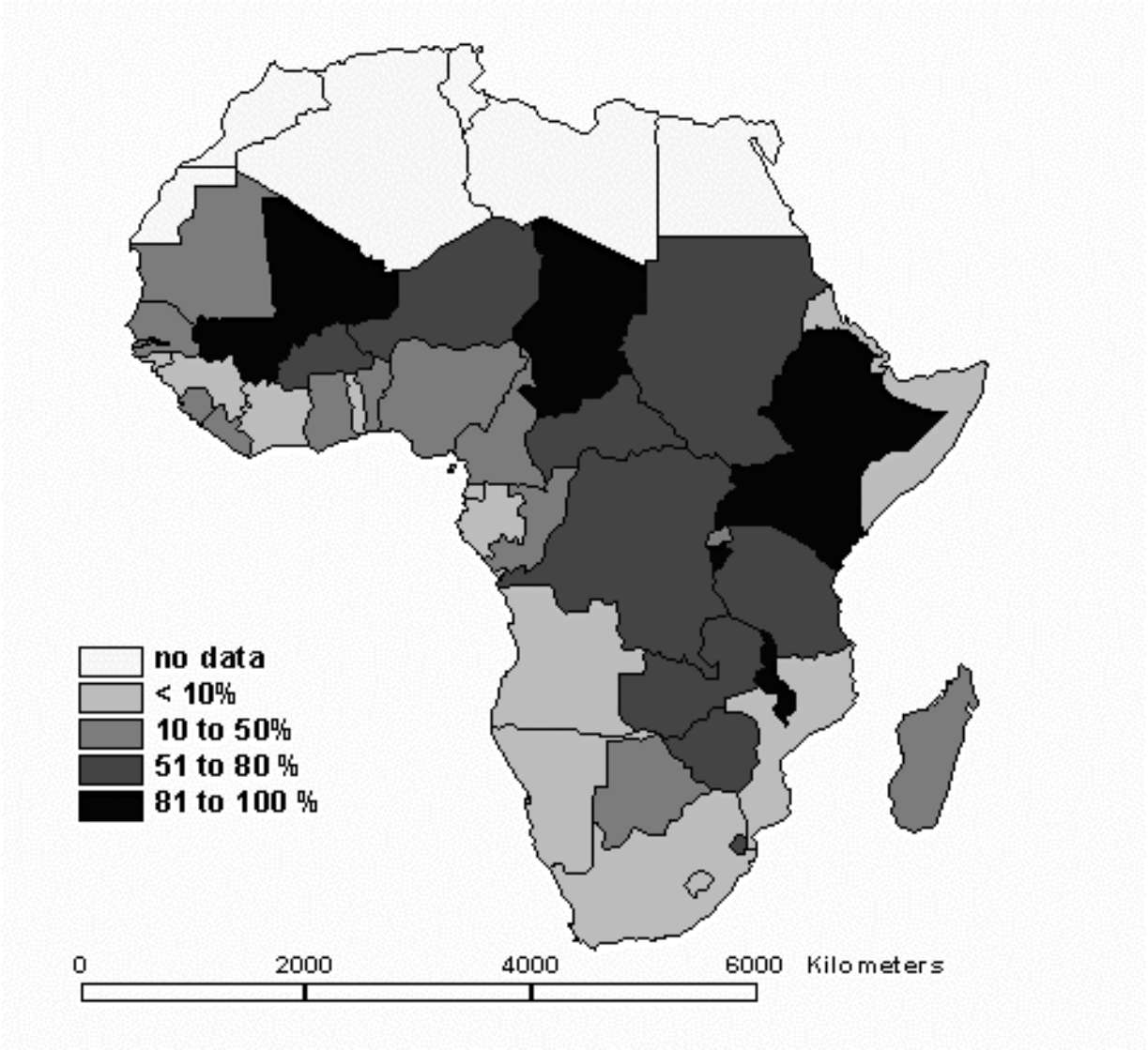


Figure 1.2: Inland fisheries production as a proportion of per capita supply of fish in sub-Saharan Africa Source: FAO 1996

Options to address the decline in fisheries include both the sustainable management of available fisheries resources and the development of aquaculture to try and alleviate pressure on the wild resource (Stotz, 2000). Important to note though is that previous attempts to address the issue of declining fisheries have presented problems given that such attempts were not guided by adequate research information (Ogutu-Ohwayo, 1990b, Pringle, 2005). For example, like elsewhere, traditional interventions have included, but have not been limited to, translocation of species, introduction of aquaculture (practiced in artificial and natural water systems) and restocking natural water bodies (Moehl and

Machena, 2000, Dadzie, 1992, Pringle, 2005). Unfortunately, poor results have stemmed from such well-intended interventions largely because of knowledge gaps on the targeted species life history, breeding biology and genetic diversity.

Although there has been an increased number of fish species farmed worldwide (FAO 2014), aquaculture in Africa faces the challenge of currently only comprising a limited number of species; hence there is still a heavy reliance on wild fisheries stocks. Successful aquaculture development, especially for new aquaculture species, depends on the reliance of healthy populations in the wild to draw stocks initially for breeding. As such it is not only important to know the state of fisheries stocks in their natural habitats for their sustainable management, but also for aquaculture development.

1.4 Status of aquaculture and species development in Africa

The decline of wild capture fisheries has indirectly propelled the development of aquaculture as an alternative to providing animal protein, and consequently, alleviating to some degree pressure on wild capture fisheries (FAO 2006, 2014). Moreover, aquaculture is an economically viable venture for commercial development and responsible aquaculture is potentially sustainable.

Despite the commercial and nutritional importance coupled with a positive policy environment, aquaculture production in the African region is still quite low compared to other regions such as Asia. The industry is based on a very limited number of fish species; the two major cultured species are the relatively low value African catfish, *Clarius gariepinus* and Nile tilapia, *Oreochromis niloticus*. Harnessing the genetic potential of existing aquaculture species through selective breeding could increase current aquaculture production levels; however, diversification to include other species of high economic value has also been viewed as one of the approaches to boost and further

increase aquaculture production in Africa (FAO, 2006). Although several species have been identified in Africa as potential commercial aquaculture species (including the Ssemutundu, *Bagrus docmak,* black Nile catfish *B. bayad,* Nile perch*, Lates niloticus* and Silver fish, *Rastrineobola argentia*), the persistence of the current limited number of cultured species could be attributed to inadequate knowledge in aspects of genetics, breeding biology, husbandry and nutritional requirements of other candidate species (Basiita et al., 2011).

1.5 Identification of study species

The bagrid catfishes (*B. docmak* and *B. bayad*) and the Nile perch (*Lates niloticus*) have been identified as model species that could be used to not only understand biogeography, but also are important commercial fisheries species and candidates for aquaculture development in Africa. These species have high economic importance and market acceptability, in addition to their wide distribution across the continent (Alhassan and Ansu-Darko, 2011, Anja et al., 2009, Aruho et al., 2013, El-Drawany and Elnagar, 2015, Khallaf, 1988, Mwanja et al., 2014a, Okach and Dadzie, 1988, Basiita et al., 2011, Basiita et al., 2015, Coulter, 1976). The species have been reported as being under threat in East Africa, largely due to overexploitation, climatic change and other associated anthropogenic activities (Balirwa et al., 2003, Bundy and Pitcher, 1995, Goudswaard and Witte, 1997, Hughes, 1992, Mkumbo and Marshall, 2015).

Anthropogenic activities have also impacted the natural distribution of freshwater fish fauna in Africa (Carvalho and Hauser, 1995). One of the best examples of a fish species that has been translocated causing shifts in aquatic fish diversity is that of Nile perch. This species which is native to the Senegal, Niger and Nile rivers , and some of the EAR Lakes (Albert and Turkana), has been subject to multiple human mediated translocations

into other East African lakes (Lakes Kyoga, Victoria and Nabugabo) (Balirwa et al., 2003, Basiita et al., 2011, Goudswaard and Witte, 1997, Ogutu-Ohwayo, 1993, Pringle, 2005). The Nile perch has become fully established in these three lakes, though attempts to introduce the species into the Kagera River and Lake Kijanebalola were futile (Pringle, 2005, Mwanja, 2013, Mwanja et al., 2012). However, following successful establishment of the Nile perch, especially in Lake Victoria, several haplochromine cichlid species have been reported to have disappeared (Hart and Pitcher, 2012). Additionally, catfish species such as the *B. docmak* have been affected in that their stocks have declined as a result of direct competition and predation from the Nile perch (Goudswaard and Witte, 1997, Kudhongania et al., 1992, Ogutu-Ohwayo, 1993). Moreover, it is also likely that the human translocations have altered the genetic history of the Nile perch.

1.6 Genetics in fisheries and aquaculture management

Whereas traditional fisheries stock assessment methods provide valuable data for managing fishery resources, additional information on the genetic diversity and connectivity of populations of a particular species can help provide a more comprehensive picture in regards to defining management units and stock structure (Griffiths et al., 2010, Hart and Pitcher, 2012). From a fisheries perspective, identifying genetic stocks in decline due to historic bottlenecks, fishing pressure, habitat degradation and climatic changes, is vital in conservation. This is because small populations have a reduced ability to adapt to abrupt environmental changes and calamities that may lead to sudden extinction of populations. Studies have provided evidence where harvesting pressure on fish stocks has led to loss of genetic diversity (Smith et al., 1991). Thus, the use of molecular genetic tools cannot be underrated in the timely identification of management units of concern and provide means to redeem populations that have shown evidence of initial genetic declines, or bottlenecks.

There are a number of ways of measuring genetic diversity among populations, ranging from measurement of phenotypic traits, chromosome typing, allozymes, mitochondrial DNA, and nuclear DNA markers. Nuclear DNA markers have been particularly useful for resolving genetic information that is important to aquaculture and fisheries (Brown & Epifanio, 2003). Among nuclear DNA markers, microsatellites are highly polymorphic and quite useful for studies of genetic differentiation, genetic structure and progeny testing (Miller and Kapuscinski, 1996, Triantafyllidis et al., 2002, Okumus and Cifti, 2003, Corujo et al., 2004, Chistiakov et al., 2006, Abdul-Muneer, 2014). Although microsatellites provide an important genetic tool for management of a given species (Melo et al., 2011), they have not been developed and characterized for *B. docmak* nor *B. bayad* and only limited markers developed from the barramundi, *L. calcarifer,* a sister species to *L. niloticus* have been applied in genetic studies of Nile perch.

1.7 Thesis Overview

The overall aim of this PhD was to develop genetic resources to better understand the genetic population structure and diversity of three commercially important freshwater fish species in Africa; the Ssemutundu (Silver catfish; *Bagrus docmac),* the Bayad catfish (*Bagrus bayad)* and the Nile perch (*Lates niloticus)*, species currently being considered as potential aquaculture candidates.

Chapter 2- Describes the generation of species-specific *de novo* genomic resources in *B. docmak* and *L niloticus* using a high throughput Next Generation Sequencing (NGS) approach. Roche 454 GS-FLX next generation sequencing was used to generate thousands of sequence fragments from which microsatellites markers were *in silico* mined. The chapter has been subsequently published as primer notes and microsatellite

records in the journals of *Animal Genetics* and Conservation Genetics Resources, respectively.

Chapter 3- This chapter used mitochondrial DNA markers in addition to microsatellite markers (developed in chapter 2) to determine the population genetic diversity, demography and structure in the widely distributed and migratory (potamodromous) African teleost species, *Bagrus docmak*. Samples acquired from five geographical locations in East Africa within two major drainage basins; the Albertine Rift and Lake Victoria basin were used to identify the extent of genetic differentiation and diversity between the two drainages.

Chapter 4- This chapter explored the restricted gene flow and strong genetic structure in Nile Perch, *L. niloticus*, from African freshwater rivers and lakes. Nile perch, which is an economically important species, has had its range extended to include new environments such as Lake Victoria where the species has now been established. Understanding the genetic structure and diversity of the Nile perch in both native and introduced environments provides an opportunity to manage sustainably the resource which is so highly exploited, especially in East Africa.

Chapter 5- Comprises work done on the *Bagrus bayad* catfish, a congener of *B. docmak.* The study utilizes mitochondrial DNA analyses and cross amplified nuclear microsatellites developed in chapter 2 to understand the historical and contemporary genetic structure of *B. bayad* from three drainage systems in West and East Africa. This work defines the possible management units based on genetically descrete populations and varied genetic diversity of the species identified across the sampled range.

Finally, Chapter 6 is a synopsis of the major thesis findings from the preceding data chapters; two to five. Highlighted also in this chapter, is the future direction of research in regards to the genetic resources and their utilisation in the conservation and management of African freshwater fish fauna.

1.8 Publications and international conference papers derived from this PhD project

Chapter 2

Basiita, RK, Zenger KR, Mwanja MT and Jerry DR. 2015. Development of genome-wide microsatellite genetic resources in a commercially important African freshwater fish species – the Nile perch, *Lates niloticus*. ***Animal Genetics*** Volume, 46, issue 3. DOI: 10.1111/age.12283.

Basiita RK, Bruggemann JH, Cai N, et al. 2016 Erratum to: Microsatellite records 7. ***Conservation Genetics Resources***. March 2016, Volume 8, Issue 1, pp 85-87. Doi: 10.1007/s12686-015-0515-6.

Basiita, RK, Zenger KR and Jerry DR 2015. Novel genetic resources in an iconic African freshwater fish, the Nile Perch, *Lates niloticus*. International Symposium on Genetics in Aquaculture (ISGA), 21-27 June 2015, Santiago de Compostela, Spain.

Chapter 3

Basiita, RK, Zenger KR, and Jerry DR 2016. Populations genetically rifting within a complex geological system: The case of strong genetic structure and low genetic diversity in the migratory freshwater catfish, *Bagrus docmak*, in East Africa. Submitted and is under Review in the **Journal of Ecology and Evolution**; Manuscript ID: ECE-2016-07-00677.

Basiita, RK, Zenger KR, and Jerry DR 2014. Genetic Diversity and Structure of *Bagrus docmak* from the Albertine Rift and Lake Victoria Basin. World Aquaculture Society (WAS) international conference held June 2014, Adelaide Australia

Chapter 4

Basiita, RK, Zenger KR, Mwanja MT and Jerry DR (2016). Restricted gene flow and strong genetic structure in Nile perch, *Lates niloticus*, from African freshwater rivers and lakes. Manuscript prepared for submission to the **Journal of Fish and Fisheries**

Chapter 5

Basiita, RK, Zenger, KR & Jerry, DR (2016). Historical and contemporary genetic structuring revealed in the African freshwater Bagrid catfish, *Bagrus bayad.* Manuscript prepared for submission to the **Journal of Marine and Freshwater**.**Research**

CHAPTER 2

Development of microsatellite genetic resources for commercial freshwater fish in Africa; Semutundu, *Bagrus docmak* and the Nile Perch, *Lates niloticus*

Abstract

Genomic resources have revolutionised scientific research worldwide, but these resources are still limited in availability and use in many freshwater fish. In order to address the lack of genetic resources for important commercial fish species in Africa, high throughput Next Generation Sequencing (NGS) was used to generate novel genetic marker loci for two commercially important fish species, *Lates niloticus* and *Bagrus docmak*. Roche 454 GS-FLX shotgun sequencing was used to generate over 160,000 sequence fragments for *B. docmak,* from which hundreds of microsatellite markers with primers were *in silico* mined. A subset of 20 novel microsatellite loci were tested in the laboratory from which 15 markers successfully amplified and were characterized in *B docmak* individuals from Lake Albert in East Africa (mean allelic richness (N_A) of 4.5, mean $H_{o,}$ and $H_{e,}$ of 0.535 ± 0.37 and 0.605 ± 0.028, respectively). These 15 loci also were found to cross-amplify in a sister species, *Bagrus bayad*. For the Nile perch, next generation sequencing on Roche 454 GS-FLX instrumentation yielded ~122,000 sequences and allowed the design of primers for 285 *in silico* perfect microsatellite loci. Thirty-one microsatellite loci were tested for ease of genotyping and polymorphism. Twelve markers were prioritized based on consistency in amplification. Additionally, 16 microsatellites previously characterized for *Lates calcarifer* (a sister species) were tested in *L. niloticus*.

The successful development of microsatellite markers *de novo* within these species permitted the investigation of stock structure in the ensuing data chapters of the thesis.

Furthermore, although not evaluated herein, these markers will have utility in parentage analysis in future breeding programs for aquaculture development.

2.1 Development of genetic resources for *Bagrus docmak*

2.1.1 Introduction

Bagrus docmak (Forsskal, 1775), or Semutundu, is a Siluroid catfish widely distributed in African freshwater rivers including the Nile, Chad, Niger, Volta, and Senegal, as well as many of Africa's large lakes (i.e. Lakes Victoria, Edward, George, Albert, Turkana and Tanganyika). Regional declines in East Africa have been reported for the species, mainly as a result of over fishing and competition from the introduced Nile perch, *Lates niloticus* (Lévêque, 1997). *Bagrus docmak* has been displaced from open waters of the lakes and is currently restricted to mouths of large rivers (Goudswaard and Witte, 1997). Farming *B. docmak* is considered a viable option to reduce the fishing pressure on wild populations; however, attempts to domesticate the fish have been unsuccessful. Prior to this research project, there were no genetic resources, or information on the species' genetic stock structure to guide conservation planning strategies of wild populations, or to identify appropriate broodstock source populations for restocking, or farming. Accordingly, the first informative microsatellite DNA markers for *B. docmak* were developed for utility in studies aimed at resolving its natural stock structure.

2.1.2 Methods

DNA from 12 wild *B. docmak* individuals (from Lake Victoria (n=3) and the Nile River (n=9)) was extracted using a modified CTAB protocol (Wilson, 1990). The DNAwas then pooled in equimolar concentrations before Roche 454 FLX sequencing (1/8th of a plate) at the Australian Genome Research Facility in Brisbane. Sequencing returned 160,232 raw reads that have been deposited into Genbank (accession # PRJNA276164). Using iQDD (Meglécz et al., 2010), data mining of sequence reads yielded 212 *in silico* microsatellite loci (165 dinucleotides, 25 trinucleotides and 22 tetranucleotides) where

primers could be designed (Basiita et al., 2016). Primers for 20 of these loci were synthesised and indirectly labelled with an M13 tag (Shimizu et al., 2002, Schuelke, 2000) and screened for polymorphism using a TYPE IT (QIAGEN) PCR microsatellite kit. Fifteen loci that amplified consistently were further directly labelled and characterised in a *B. docmak* population (N = 16) from Lake Albert, East Africa. The markers were also trialled for cross amplification in a closely related taxa, *B. bayad* (N = 5). Optimised PCR conditions were characterised by an initial denaturation at 95 ^{o}C for 5 min, 6 cycles of 95 ^{o}C for 30 s (denaturation)/59 °C for 90 s (annealing)/72 °C for 30 s (extension), 10 cycles for each reducing annealing temperatures of 57 ^{o}C, 55 ^{o}C and 53 ^{o}C, prior to a final extension at 60 ^{o}C for 30 min.

Microsatellite genotyping was performed on an ABI-3730 instrument (Applied Biosystems) using a 5-standard dye system (6-FAM, VIC, NED, PET and LIZ GS-500 size standard) at the Georgia Genomics Facility, USA. Alleles were scored using Genemarker 2.4 (Softgenetics) and checked for genotyping errors and null alleles in Microchecker 2.2.3 (Van Oosterhout et al., 2004). Diversity indices were computed in GenAlex 6.5 (Peakall and Smouse, 2012).

2.1.3 Results

Characteristics of the 15 loci are presented in Table 2.1. The loci displayed a mean allelic richness, $N_{a,}$ of 4.5 (range 2-7 alleles per locus), and mean observed and expected heterozygosity, $H_{o,}$ and $H_{e,}$ of 0.54 ± 0.37 and 0.61 ± 0.03, respectively, except for locus *BD16* that was monomorphic in the Lake Albert population. Null alleles were also suggested at locus *BD07*. All markers tested were found to be polymorphic in the sister species *B. bayad* (Table 2.1). Consistency in amplification of the polymorphic microsatellite loci illustrates their potential utility in conservation genetic studies, parentage assignment and maintaining pedigree data in breeding programs.

Table 2. 1 Details of microsatellite loci isolated from *Bagrus docmak* and cross amplified in *Bagrus bayad*

Locus	Primer sequence	T_M (°C)	Motif	Characteristics of loci in *B. docmak* individuals						Cross species amplification in *B. bayad*		
				N	N_a	H_o	H_e	*PIC*	Size range (bp)	Amplification success	N_a	Size range (bp)
Bd04	F: TGTGGACCAAGAGACAGGTG R: AATGAACAAGGCAGGTGATG	59	$(AGAT)^{18}$	16	3	0.563	0.646	0.571	200-208	5/5	5	197-229
Bd18	F: ATGGGGAGGAAAAGTGGAG R: CCTGAGTGCATTGCTCATGG	61	$(AC)^{15}$	16	2	0.563	0.451	0.349	100-102	5/5	4	125-133
Bd01	F:TTGCCAATCCTGATGACACTC R:TAAAGCTGGGCAACTGATCC	60	$(TTCT)^{15}$	15	4	0.333	0.589	0.528	203-219	5/5	5	190-206
Bd02	F:TGTGCTCTGACCCCTACCTC R:GGGTATCGCATCCCAGATAG	60	$(AGAT)^{17}$	16	6	0.563	0.541	0.516	110-130	5/5	4	110-126
Bd12	F: CCGACCATCTCAAATACAAGTC R: CTCTTCCCCAAGGCTATTCC	60	$(AAT)^{18}$	16	3	0.688	0.508	0.428	237-258	5/5	2	236-239
Bd09	F: ACTGTTCCCATGAAGTTGGG R: TGGTCAACTTTAGATGTGCAGC	60	$(ATT)^{19}$	16	6	0.563	0.734	0.702	223-238	5/5	2	240-243
Bd06	F: TTCTGAAGCCCAAAGTAGACG R: GCCCACACTATTGACACAGG	59	$(GATA)^{16}$	16	7	0.625	0.609	0.582	171-199	5/5	4	203-215
Bd20	F: TCCTGGAGACCAAGACCAAG R: TGCAGGTTAAGAATGGAGGC	60	$(CA)^{11}$	16	4	0.688	0.672	0.612	156-168	5/5	2	148-150
Bd05	F: GCTGGCAACATGCAGTAATC R: CAGCATTTCATTGCTATGTGC	59	$(ATAC)^{15}$	13	4	0.250	0.719	0.667	136-172	5/5	2	125-133
BD07	F: GAGCACACGAAACATTGCAG R: TTGTAGATTCCCTTTGGGATG	60	$(GATA)^{15}$	4	5	0.538	0.678	0.638	125-157	4/5	4	134-158
Bd16	F: GCAATCGCACTCTTGTTATCG R: TAGTAGCGCACCCAGGAAAC	61	$(ATT)^{13}$	13	1	-	-	-	83	5/5	3	83-89
BD08	F: TTACCTCACACTCTGGGGTTG R: GGTAAAGGTTTACACTGTGGGG	60	$(ATCT)^{16}$	13	3	0.615	0.544	0.484	179-187	5/5	3	179-187
BD03	F:CCTGCAGGAGTTTGTTTGTG R: CGTGCCATAGGCATTTATCC	60	$(TAGA)^{15}$	16	3	0.438	0.568	0.482	159-191	5/5	6	179-203
BD14	F: CTTTAATGACACTGCGCTGC R: CTCAAAGCGCTTGAAGTGG	60	$(TAT)^{17}$	16	6	0.688	0.781	0.748	218-233	5/5	3	225-241
Bd10	F: GTCCCACGGACTGAAAAGTG R:TCAACTTCTTAGCACAAAATCAGAC	60	$(TTA)^{15}$	16	7	0.375	0.426	0.412	266-287	5/5	4	265-277

T_M, optimal primer melting temperature; *N,* sample size; N_a, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity ; *PIC,* polymorphic

2.2 Development of genetic resources for the Nile perch, Lates niloticus

2.2.1 Introduction

Lates niloticus, commonly referred to as the Nile perch*,* is a freshwater fish species with marked economic and food security importance where it occurs in Africa. Additionally, a large proportion of Nile perch wild catch, mainly from East Africa worth about $US 350 million annually, is also exported to North America, Europe and Australia, and further emphasises the importance and contribution of the Nile perch as a viable animal protein source globally (Josupeit, 2006, Mkumbo and Marshall, 2015). However, in the recent past the species has been reported to be on the decline in many parts of its distribution, particularly Lake Victoria where the species has been introduced and the highest quantity of exports are derived (Josupeit, 2006, Mwanja et al., 2013, Mwanja et al., 2012, Mkumbo and Marshall, 2015). Secondly, like many species in Africa, there is limited information to provide an account of their population and genetic status to allow for management in natural and introduced environments. The limited genetic information is in part due to the unavailability of sufficiently powerful genetic resources that can resolve differences in population structure of Nile perch across the species range in Africa.

Although a limited number of genetic markers have been previously used for this species (i.e. allozymes and nine non-species specific microsatellite loci), these markers have had limited resolution, encumbering solid conclusions with regard to the genetic structure of the species (Mwanja et al., 2014b, Mwanja and Mwanja, 2008, Hauser et al., 1998). In particular, to date there are no species-specific microsatellite markers that have been published for *L. niloticus*. Currently the highest number of microsatellite markers that

have been used to attempt and resolve population genetic structure in Nile perch is nine loci developed for *L calcarifer* (*Lca*) (Mwanja et al., 2012, Mwanja and Mwanja, 2008, Mwanja et al., 2014b, Lusweti, 2013). Increasing the number of genetic markers useful for genetics studies in Nile perch is fundamental for the species' conservation and future aquaculture development through breeding.

2.2.2 Sample acquisition, DNA preparation and Roche 454 Sequencing

Lates niloticus were acquired from wild populations. Muscle tissue clips were excised from fish and preserved in DMSO salt preservation solution (Dawson et al., 1998) prior to DNA extraction using a modified CTAB protocol (Wilson, 1990). DNA was extracted from six individuals of *L. niloticus* (one from Lake Victoria and one from Lake Kyoga in Uganda, East Africa, two from the Senegal River in Senegal, West Africa, and two from Lake Turkana in Kenya, East Africa). The DNA quality and quantity were determined using both 0.8% agarose gel electrophoresis and spectrophotometry (Nanodrop spectrophotometer, ND-1000), respectively. DNA from these samples was then pooled in equal concentrations to constitute a final total concentration of ~40ng/µl. Pooled DNA samples were sent to the Australian Genome Research Facility (AGRF) in Brisbane for Roche 454 GS-FLX sequencing using Titanium chemistry (1/8th of a plate).

2.2.3 Roche 454 sequencing results and data mining to identify microsatellite loci in *Lates niloticus*

In excess of 122,000 sequences with an average length of 322 bp and a total of 39.42 Mbp of DNA sequence data with a GC content of 40.97% was obtained. All the sequences were deposited at the NCBI (accession # SRP051308). Data mining was undertaken to

identify and design microsatellite primers from the raw sequences using IQDD (Meglécz et al., 2010) and MSATcommander (Abdelkrim et al., 2009). An inbuilt search for perfect and compound dinucleotide, tri-nucleotide and tetra-nucleotide microsatellite repeats (≥10) using the program MSATcommander (Faircloth, 2008) resulted in identification of 8311 di-, 2386 tri- and 649 tetra-nucleotide *in silico* microsatellites. Additional stringent screening for only perfect repeats (of >10 for tetra- and tri-nucleotides and >15 for di-nucleotides) was then conducted in the program IQDD, (Meglécz et al., 2010) using the parameters; amplicon length of 100–400 bp, primer length set at 20 bp (range 18–27), optimal primer T_M 60 ^{o}C (range 57–63 ^{o}C), primer GC content range of 40–70% and at least one G/C clamp at the 3′ end of primers.

Primers were designed in the flanking regions of 285 microsatellites (Appendix 2.1). More stringent parameters were further used to filter the best microsatellites with flanking regions, the additional parameters included: sequence quality (range of 15-40, average of 34 for left primer and 25 for right primer) of the portion from which the primer pair was derived, primer pair penalty (<5), T_M (58-61^{o}C) and distance from the end of the sequence (>20 bp). Thirty-one loci isolated (10 tetra, 11 tri and 10 dinucleotide repeats) with primers were synthesised and further validated with PCR for amplification success.

2.2.4 Polymerase chain reaction (PCR) conditions and genotyping

Microsatellite markers were tested for amplification via polymerase chain reaction using a QIAGEN microsatellite PCR kit (TYPE IT) following the manufacturer's protocol. Although the initial testing was performed in singleplex reactions, the indirect fluorescent labelling with M13 allowed for co-loading of samples from up to four primers, minimising the cost of genotyping. Further optimisation was achieved using a step-down

PCR taking into account the downstream use of primers in multiplex reactions (primer multiplex combinations were modifications from those designed by Multiplex Manager 1.0 (Holleley and Geerts, 2009) to ensure minimal costs of laboratory consumables and commercial genotyping services. The optimised PCR program was characterised by an initial denaturation at 95 ^{o}C for 5 min, followed by 6 cycles of 95 ^{o}C for 30 s (denaturation)/59 °C for 90 s (annealing)/72 °C for 30 s (extension) then 10 cycles each of reducing annealing temperatures of 57 ^{o}C, 55 ^{o}C and 53 ^{o}C, prior to a 30 min final extension at 60 ^{o}C. Using this optimised PCR program, 14 of the 31 loci initially amplified but 12 were prioritised since they consistently amplified and consequently were characterised in 38 individuals from Lake Victoria (Table 2.2).

Additionally, 16 microsatellite markers originally developed for *L. calcarifer* (*Lca* markers) (Zhu et al., 2006) were tested in the same 38 *L. niloticus* individuals from Lake Victoria and optimised with the same PCR program developed for the *L. niloticus* (*Ln*) markers above. The markers were optimised in two multiplex reactions of nine and seven markers, respectively. A positive control containing DNA from barramundi (original species for which the *Lca* markers were designed) was included for all experiments using the *Lca* markers.

Genotyping of all PCR product was performed on an ABI-3730 platform (Applied Biosystems) using a 5 ABI standard dye system (6-FAM, VIC, NED, PET and LIZ GS-500 size standard) at the Georgia Genomics Facility, USA.

2.2.5 Characterising microsatellites in *Lates niloticus*

Alleles for each marker pair were scored in Genemarker 2.4 (Softgenetics) and checked for null alleles and other possible genotyping errors in Microchecker 2.2.3 (Van

Oosterhout et al., 2004). Genetic diversity characteristics of the *Ln* and *Lca* markers in *L. niloticus* (Table 2.2 and 2.3, respectively) were computed in GenAlex 6.5 (Peakall and Smouse, 2012, Peakall and Smouse, 2006) and Cervus version 3.0 (Kalinowski et al., 2007).

Twelve, *Lates niloticus* (*Ln*) markers were characterized with a mean allelic richness, N_A, of 4.6 and a range of 2-10 alleles per locus (Table 2.2). Levels of expected heterozygosity were relatively high in most loci (mean H_e = 0.47 and a range 0.1-0.8). All *Ln* markers exhibited no evidence of stuttering and/or large allelic dropouts. Loci *Ln31, Ln09 and Ln05* showed a possibility of null alleles.

Lates calcarifer (*Lca*) markers when characterized in *L. niloticus* had a slightly lower mean allelic richness (N_a of 3.8, range = 2-7) and a slightly higher expected mean heterozygosity (mean H_e = 0.48, range = 0.02-0.9). All *Lca* markers amplified in *L. niloticus* did not deviate from Hardy Weinberg equilibrium (all probabilities > 0.05), except for locus *Lca 40*.

Table 2. 2 Microsatellite loci isolated through Roche 454 GS-FLX sequencing and validated for *Lates niloticus*

Locus	Repeat Motif	Primer sequence (5'-3')	Primer concentration (μM)	TM (°C)	*N*	*Ho*	H_e	N_a	*PIC*	Allele size
*Ln*23[ab]	(TCA)10	F-CACAAGTGTAACCAGTGCGG R-TCTATTAACGGGAGCGGTTG	0.2	60	37	0.135	0.128	2	0.118	227-244
*Ln*29[a]	(AC)17	F-TGCCAGGAGCTCAGTAACAC R-TGTCAAGTAGGCATGTGGTG	0.2	59	37	0.486	0.541	3	0.428	174-182
*Ln*31[c]	(AC)15	F-GGTCACAGGTCACTGCTTTG R-CACTGAGCAAAGACGAATGC	0.2	59	38	0.270	0.426	5	0.389	220-230
*Ln*05[b]	(CTGT)11	F-ACCCTCCACTGTGTTTCCTC R-AGGCTTGGTCTCCAGGATG	0.2	59	38	0.028	0.126	4	0.105	107-130
*Ln*09[c]	(TGA)15	F-CCCACAATACTATGAGGGTGC R-TCCAAATTTTGTCTTAAACCTGC	0.2	59	35	0.086	0.131	3	0.131	232-238
*Ln*15 [ab]	(GT)18	F-GTTGTGACGATATTGCGTGG R-CACGGTCTGTTGCATTATGG	0.2	60	37	0.73	0.683	7	0.625	257-293
*Ln*16 [b]	(ATGG)10	F-CAATGTCAGCTGGGATAGGC R-CCGGTTCTTGTCAATGTCAC	0.2	61	35	0.353	0.137	3	0.428	134-154
*Ln*17 [a]	(CAGA)10	F-AGCAGCATCAGTTCAACCAC R-CCTGGTCATTTTCCACACTG	0.2	59	37	0.865	0.847	10	0.865	91-147
*Ln*19 [a]	(AGG)11	F-TGTCTGTGTTCAGCCCTCAG R-TGATCCATTAGCCAGCAGTG	0.2	60	37	0.486	0.459	3	0.4	139-154
*Ln*02 [a]	(ATCT)18	F-ATTGGCCCTCATTCAGTACG R-AACATTAACTGTGGGCCTCTG	0.2	60	37	0.595	0.736	5	0.680	166-182
*Ln*10[c]	(ATA)15	F-ACGGTATACAACAGCAGCCC R-AAGCGGTGACCTCCATAGTC	0.2	60	36	0.629	0.659	4	0.594	227-239
*Ln*11[c]	(TTC)13	F-CTCACACGTCCTGTTTGCTG R-TCAGAGGAAGTCGTGCAGTG	0.2	61	36	0.389	0.363	6	0.340	174-219

Ln, L. niloticus; TM, optimal primer melting temperature; *N,* sample size; N_a, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; *PIC,* polymorphic content; *a, b and c*, PCR multiplexes to which the markers belong following optimised PCR conditions.

Table 2.3: Characterisation of microsatellites validated to amplify in *Lates niloticus* that were originally designed for *Lates calcarifer*

Locus	Source	Motif		Observed allele size range (bp) in *L.calcarifer*	Observed allele size range (bp) in *L.niloticus* (current study)	Number of alleles in *L.niloticus* (current study)	H_O	H_e	PIC
Lca03	(Wang et al. 2006)	$(CA)_{14}$	0.1	216- 239	226-228	3	0.108	0.105	0.1
Lca16	(Yue *et al.* 2009)	$(CA)_{13}$	0.2	243–279	not scorable	questionable alleles	-	-	-
**Lca40	(Wang et al. 2006)	$(GT)_{15}$	0.2	122-141	147-151	5	0.250	0.653	0.583
Lca57	(Zhu et al. 2006)	$(GT)_{24}$	0.2	207-222	128-134	4	0.500	0.712	0.599
Lca154	(Wang et al. 2006)	$(TG)_{8}$	0.1	144-154	162-178	4	0.784	0.752	0.698
Lca178	(Wang et al. 2006)	$(GA)_{9}$	0.2	434-446	423-434	3	0.541	0.482	0.428
Lca371	(Wang et al. 2006)	$(CA)_{10}$	0.2	375-387	not scorable	failed to amplify	-	-	-
Lca08	(Zhu et al. 2006)	$(GA)_{21}$	0.2	253-260	248-250	2	0.429	0.507	0.375
Lca20	(Zhu et al. 2006)	$(CA)_{21}$	0.8	129-142	128-134	4	0.861	0.644	0.561
Lca21	(Wang et al. 2006)	$(CA)_{13}$	0.8	183-195	158- 160	2	0.029	0.029	0.029
Lca58	(Zhu et al. 2006)	$(GT)_{18}$	0.2	402-459	416-438	6	0.389	0.385	0.361
Lca64	(Zhu et al. 2006)	$(AC)_{21}$	0.2	279-313	270-274	3	0.083	0.108	0.104
Lca69	(Zhu et al. 2006)	$(GT)_{14}$	1.3	356-361	376-380	3	0.001	0.711	0.563
Lca70	(Zhu et al. 2006)	$(CAG)_{17}$	0.6	297-313	301-305	2	0.556	0.505	0.374
Lca74	(Zhu et al. 2006)	$(CA)_{13}$	1.3	163-170	170-182	7	0.833	0.831	0.794
Lca98	(Zhu et al. 2006)	$(TG)_{14}$	1.2	189-213	201-223	6	0.806	0.702	0.632

H_o, observed heterozygosity; H_e, expected heterozygosity; *PIC,* polymorphic content. ****Deviation from Hardy Weinberg Equilibrium at loci Lca40 (<0.001).*
Source: Table published in Basiita et al. (2015).

2.3 General conclusion

The development *de novo* of genomic resources through next generation sequencing and the subsequent use of these resources has rapidly advanced knowledge in aquatic research, particularly that related to population genetic surveys, breeding programs and conservation efforts (Ekblom and Galindo, 2010, McCormack et al., 2013, Liu et al., 2011, Hale et al., 2009, Abdelkrim et al., 2009, Kang et al., 2012). However, the availability and utility of such resources has been limited among African freshwater fishes. The development in the present chapter of a large number of microsatellites for three commercially important species now allows comprehensive population genetic audits to be conducted to understand stock structure to inform management plans, along with increased understanding of how biogeography has shaped the African aquatic fauna (Chapters 3, 4 & 5). Although not formally evaluated in the present thesis, the markers developed will also have application to the conduct of genetic audits and parentage determination in future breeding programs for these species.

CHAPTER 3

Populations genetically rifting within a complex geological system: The case of strong structure and low genetic diversity in the migratory freshwater catfish, *Bagrus docmak*, in East Africa

Abstract

The complex geological history of East Africa has been a driving factor in the rapid evolution of teleost biodiversity. Whilst there is some understanding of how macroevolutionary drivers have shaped teleost speciation in East Africa, there is a paucity of research into how the same biogeographical factors have affected microevolution. To address this deficiency, population genetic diversity, demography and structure were investigated in a widely distributed and migratory (potamodromous) African teleost species, *Bagrus docmak*. Samples were acquired from five geographical locations in East Africa within two major drainage basins; the Albertine Rift and Lake Victoria Basin. Individuals (N = 175) were genotyped at 12 microsatellite loci and 93 individuals sequenced at the mitochondrial DNA control region. Results suggested populations from Lakes Edward and Victoria had undergone a severe historic bottleneck resulting in very low nucleotide diversity (π = 0.004 and 0.006, respectively) and negatively significant *Fu* values (-3.769 and -5.049, at $P < 0.05$). Heterozygosity deficiencies and restricted effective population size (N_{eLD}) suggested contemporary exposure of these populations to stress, consistent with reports of the species decline in East Africa. High genetic structuring between drainages was detected at both historical (ϕ_{ST} = 0.62 for mtDNA, P = 0.000) and contemporary (microsatellite F_{ST} = 0.46, P = 0.000) levels. Patterns of low genetic diversity and strong population structure revealed are consistent with speciation patterns that have been linked to the complex biogeography of

East Africa, suggesting that these biogeographical features have operated as both macro- and micro-evolutionary forces in the formation of the East African teleost fauna.

3.1 Introduction

East African freshwater systems possess a diverse teleost fauna shaped by a complex geological history, including large-scale tectonic movements, volcanic activity, and significant uplifting (Danley et al., 2012, Verheyen et al., 2003, Sturmbauer et al., 2001). The East African Rift (EAR) Valley, which was formed by tectonic uplift, is the major geological structure that has forged the hydrographical network in Africa (Pinton et al., 2013, Giddelo et al., 2002) and created various freshwater habitats. Evolutionary and geological processes such as fragmentation, hydrological connectivity, river reversal and desiccation, among others (Danley et al., 2012, Johnson et al., 1996, Russell and Johnson, 2001), were responsible for the creation and maintenance of these habitats in which substantial numbers of aquatic taxa were isolated.

Freshwater lakes and rivers within the EAR (ie. Lake Edward, Lake Albert, Lake George, Lake Tanganyika, Lake Malawi), and the largest tropical freshwater body, Lake Victoria (which lies outside the EAR), are habitats to one of the world's most biologically diverse aquatic faunas. For instance, prior to the introduction of the Nile perch, *Lates niloticus*, into Lake Victoria, the lake had between 350 - 600 endemic cichlids. (Helfman, 2007, Turner et al., 2001). Geological evidence suggests a former connection of Lake Edward to Lake Victoria by late Pleistocene rivers, which were subsequently truncated by uplifting, causing river reversal and a break in connectivity of lake systems (Lévêque, 1997). Presently, connectivity of these lakes is restricted to Lake Edward, which is connected to both Lake George and Lake Edward (via the Kazinga Channel and Semliki River, respectively) (Figure 3.1).The Lake Edward-George system provides a biogeographic confluence between the Victorian and Albertine freshwater fauna (Thieme et al., 2005). Despite this hydrological

connectivity between lakes in the EAR through rivers and channels, biogeographic barriers are evident including the Semliki rapids and falls, which descends 300 metres from Lake Edward to Lake Albert (Lowe-McConnell, 2009, Lowe-McConnell, 1993), although Greenwood (1966) considers the Semliki rapids as inefficient barriers. To the right of Lake Albert are the Murchison Falls along the Victoria Nile River, another biogeographical barrier separating the Albertine rift system (includes Lakes Edward, George and Albert) from Lake Victoria. This barrier has been documented as an effective obstacle in maintaining for instance the Nile perch stocks in Lake Albert beyond which they were not able to migrate upwards into Lakes Kyoga and Victoria (Hopson, 1972, Basiita et al., 2011).

Africa's freshwater systems are degrading at a very high rate with over 80 species listed as critically endangered, 116 species endangered and up to 103 threatened (IUCN, 2002; Thieme *et al*. 2005). As elsewhere in Africa, the unique East African teleost fauna in both riverine and lacustrine freshwater habitats are currently under threat due to natural and anthropogenic pressures, with many species experiencing rapid population declines. Relative to other environments, biodiversity declines are at their highest in freshwater lacustrine bodies owing to the level of compartmentalization that naturally exists among these large water systems (Ricciardi and Rasmussen, 1999). The natural divides of freshwater bodies over evolutionary timescales seemingly deem these divides as important in regards to defining management units for biodiversity. Any interventions such as aquaculture developments, restocking existing water bodies, zoning fishing and breeding grounds that are geared towards augmenting the declines (resulting from natural and anthropogenic pressures), need to take into account evolutionary significant units in the region.

Although numerous studies have looked at aquatic fish diversity, composition and endemism, to understand biogeographic processes in East Africa (Craig, 1992, Elmer et al., 2009, Sato et al., 2003, Pinton et al., 2013), very few studies in fish have looked at genetic signatures left by biogeographical processes below the level of species (i.e. among populations). Unravelling antecedent genetic signatures among populations may not only help understand the processes that have led to their evolution and adaption in recent timescales, but more importantly, will assist with the identification of genetically divergent populations and/or management units that can be integrated into the formation of contemporary management plans.

Ssemutundu (*Bagrus docmak*) is a freshwater catfish with a widespread distribution in African freshwater rivers, including the Nile, Chad, Niger, Volta, and Senegal Rivers. It is also found in Lake Victoria, as well as the Rift Valley Lakes Edward, George, Albert, Tanganyika, Malawi and Turkana (Golubtsov et al., 1995, Greenwood, 1966, Aruho et al., 2013, Mwanja et al., 2014a, Goossens, 2015). The fish is potamodromous migrating from lakes to rivers during rainy seasons and from deep waters to shallow sandy bottoms to spawn (Thieme et al., 2005, Chapman et al., 2012). *Bagrus docmak* is an important species currently commercially fished from the wild, but also is a species with high aquaculture potential, largely due to its attractive attributes including size, taste, flesh quality and overall commercial importance (Alhassan and Ansu-Darko, 2011, Aruho et al., 2013, Mwanja et al., 2014a). The fish used to be listed as threatened by the IUCN, and currently the overall population status where it occurs is not known, but appeals are made to provide information on the species population status (IUCN, 2010). Additionally, natural populations of this species are in decline and are under threat, especially in East Africa where the species has

become very rare (IUCN 2006; IUCN 2010). Population declines of Ssemutundu are a result of mainly anthropogenic pressures, such as introductions of exotic species (i.e. Nile perch, *Lates niloticus*) and habitat degradation (Hauser et al., 1998, Kudhongania et al., 1992, Ogutu-Ohwayo, 1990b, Ogutu-Ohwayo, 1993, Olowo and Chapman, 1999, Chapman and Chapman, 2003, Chapman et al., 2007, Dickson et al., 2012). Elucidation of *B docmak's* population genetic structure is crucial for the species' future management given its high conservation and commercial importance. Additionally, the species' widespread distribution throughout Africa, along with its potamodromous migratory life-history, identifies it as an ideal candidate to examine how the complex geological history of East Africa shapes evolution of the fish fauna biodiversity at below the species-level.

3.2 Methods

3.2.1Study area

Samples of *B. docmak* were collected from five water systems across the species' distribution in East Africa (from Lakes Albert, Edward and Victoria; and Rivers Victoria Nile and the Kazinga Channel) (Figure 3.1). Lake Albert and Edward are located in the western arm of the East African Rift Valley (referred to as the Albertine rift), whilst Lake Victoria is outside.

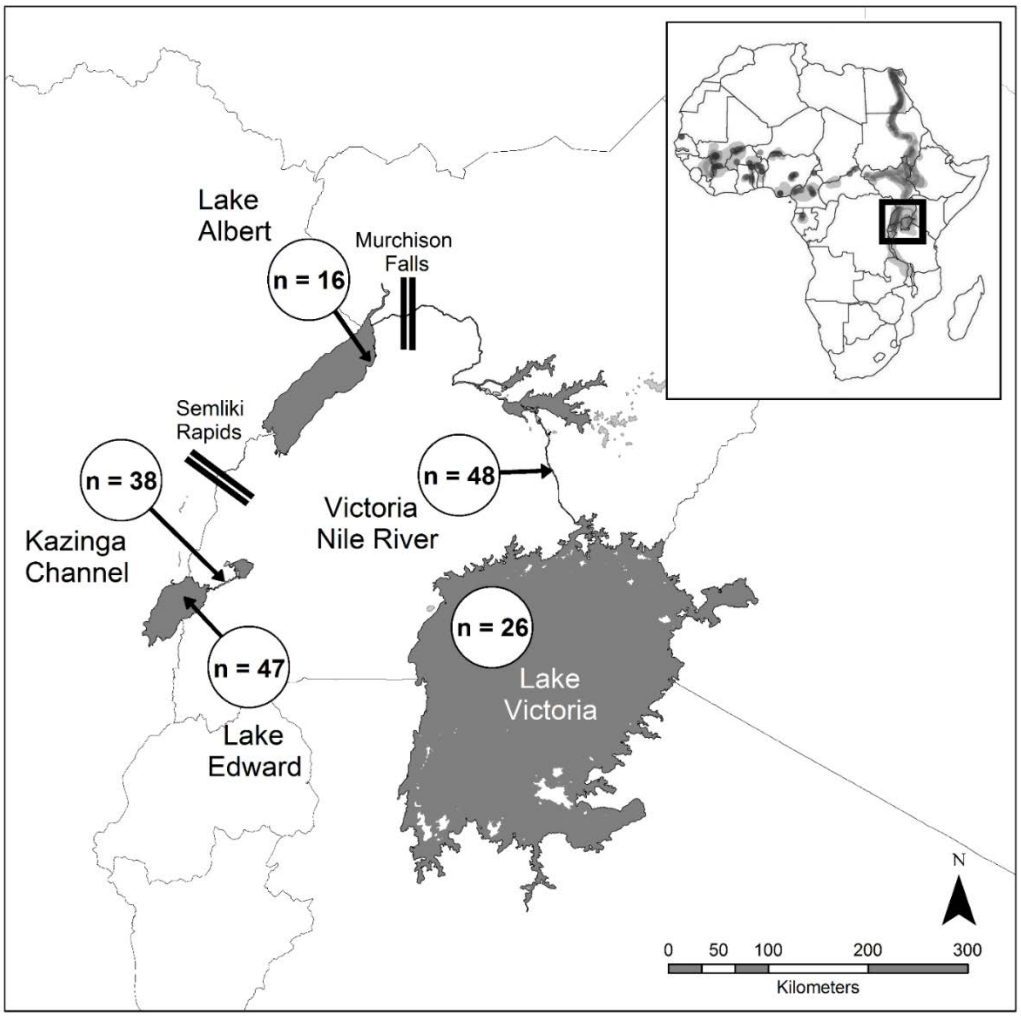


Figure 3.1: Map showing lakes and rivers (sampling locations within East Africa) and number of individuals of *Bagrus docmak* collected from each location.The insert at the top right corner denotes the distribution of *B. docmak* across Africa.

3.2.2 Sample collection and laboratory procedures

Fin clips of *B. docmak* were obtained from commercial fishers at each of the five sampling locations (Figure 3.1). All fin clips were preserved in 20 % dimethyl sulfoxide (DMSO) saturated with sodium chloride salt (Dawson et al., 1998, Amos and Hoelzel, 1991) and transported to the Molecular Ecology and Evolution Laboratory (MEEL) in Townsville, Australia, where they were stored at -20 ˚C until extraction. DNA extractions and polymerase chain reaction (PCR) assays were also carried out at MEEL.

Total genomic DNA was extracted using a modified CTAB protocol (Wilson, 1990) and later using the Bioline Isolate II Genomic DNA kit. The CTAB protocol involved a digestion step using a CTAB buffer with 200 µg of Proteinase K incubated at 55 ˚C for 2 hr, followed by a 24:1 chloroform: isoamyl alcohol purification (700 µl of chloroform–isoamyl and centrifugation at 13,200 rpm for 20 min), and an ethanol precipitation (2.5x 100 % EtOH: 1:10x 5 M NaAce and centrifugation at 13,200 rpm for 30 min, 1x 70 % EtOH at 13,200 rpm for 20 min). Genomic DNA was re-suspended in 25 µl of 1X TE (10 mM Tris–HCl, 1 mM EDTA, pH 8.0). DNA quality was estimated based on 0.8 % agarose gel electrophoresis and quantity was assessed using a ND-1000 Spectrophotometer (Nano-Drop® Technologies). For some inhibited samples that failed to amplify during PCR, DNA was re-extracted using a column based Bioline Isolate II Genomic DNA kit following the manufacturer's protocol. Briefly the protocol involved a pre-lysis stage using 180 µl Lysis buffer with 25 µl of protease K incubated at 56 ºC. 200 µl of G -3 lysis buffer was added and further incubated for 10 min at 70 ºC. 210 µl of 100 % EtOH were then added to alter the buffer conditions prior to two (GW1 and GW2) buffer washes. The eluted DNA was stored at -80 ºC prior to downstream PCR.

DNA from 175 individuals were genotyped at 12 polymorphic microsatellite loci, Bd04, Bd18, Bd01, Bd02, Bd12, Bd09, Bd06, Bd20, Bd05, Bd03, Bd14 and Bd10 (Appendix 3.2) (Basiita *et al*. in Goossens 2015). Forward primers were fluorescently labeled using the 5-dye system (6-FAM, VIC, NED, PET and LIZ GS-500 size standard) and all reverse primers were pigtailed (Brownstein et al., 1996) to ensure consistent amplification and minimize stuttering. A total of 20 µl PCR reactions were run on a Biorad C1000 thermocycler under the following conditions: an initial denaturation at 95 °C for 5 min, 6 cycles of 95 °C for 30 s (denaturation)/59 °C for 90 s (annealing)/72 °C for 30 s (extension), 10 cycles at reduced annealing temperatures of 57 °C, 55 °C and 53 °C, prior to a final extension at 60 °C for 30 min (Basiita *et al.* in (Goossens, 2015)). Visualisation of PCR product was performed on an ABI-3730 instrument (Applied Biosystems) using a 5-standard dye system (6-FAM, VIC, NED, PET and LIZ GS-500 size standard) at the Georgia Genomics Facility, USA. Alleles were scored using Genemarker 2.4 (Softgenetics), and checked for genotyping errors and null alleles in Microchecker 2.2.3 (Van Oosterhout et al., 2004).

The mitochondrial control region (D-loop) was amplified in 93 individuals from the five locations. Initially, catfish oligonucleotide primers, MT16498H and L19 (Chenoweth and Hughes, 1997) were used to amplify the D-loop mitochondrional region in order to obtain *B. docmak* sequences that were then used to design more robust species-specific primer pairs. Primers specific to *Bagrus spp*. were designed using a free online software, Primer3 (Rozen and Skaletsky, 2000). A forward (BagDF2) - TTGAGGGTTGGTGGTTTCTT and reverse (BagDR2) - AAACTATTTTCTGTAAATGCATAAT primer pair were designed and tested for specificity in *B. docmak* via PCR. A total of 25 µl reaction volume was used; 2.5 µl 10X buffer, $MgCl_2$ 1.5 mM, dNTPs 0.2 mM, *B. docmak* control region primers; BagDF2 and

BagDR2 (at a final concentration of 0.2 µM for each primer) and 1 µl DNA (5 ng/µl); PCR cycling conditions comprising of an initial denaturation of 94 °C for 5 min, then 30 cycles of 94 °C 30 s (denaturation)/ 50 °C for 30 s (annealing)/72 °C for 30 s and a final extension of 72 °C for 5 min. Visualization for amplification was achieved via electrophoresis using a 1.5 % agarose gel. PCR product was cleaned with Sephadex G-50 (GE Healthcare UK) columns prior to sending approximately 5 ng of each sample for Sanger sequencing at the Australian Genome Research Facility (AGRF), in Brisbane, Australia. Sequencing was performed of both forward and reverse directions using the designed *B. docmak* specific primers.

3.2.3 Mitochondrial DNA sequence editing and alignment

Individual D-loop sequences were aligned and consensus sequences generatedin GENEIOUS 8.02 (Biomatters Ltd). Following generation and alignment of all consensus sequences, a 400 bp size region was trimmed and exported to MEGA 5.2 (Tamura et al., 2011). The best substitution model for analysing sequence divergence and population structure among the different populations was selected as T92+G (Nei and Kumar, 2000, Tamura et al., 2011) using MEGA 5.2 according to Schwarz (1978).

3.2.4 Genetic diversity: descriptive statistics

Analyses to determine the number of alleles, observed and expected heterozygosities (H_o and H_e respectively), inbreeding coefficient (F_{IS}) and conformation to Hardy Weinberg Equilibrium (HWE) at 12 microsatellites loci were performed in GenAlex (Peakall and Smouse, 2012) and Arlequin 3.5 (Excoffier and Lischer, 2010). All significance levels were corrected for multiple tests using a two-step false discovery rate (FDR) correction (Benjamini and Hochberg, 1995, Benjamini et al., 2006) set at a maximum of 0.001. The F_{IS} was used as

a measure of inbreeding and/or population sub-division. Linkage disquilibrium method was used for estimating the effective population size (N_{eLD}) and was implemented in an updated softfaware program, *Ne*Estimator version 2 (Do et al., 2014). Input files were converted for use between programs using the free software, PGDspider 2.0.5 (Lischer and Excoffier, 2012).

All microsatellite loci were polymorphic in all populations sampled, except for locus BD18 that was monomorphic in the Victoria Nile River and BD20 in the Lake Victoria populations. Micro-checker 2.2.3 was used to detect genotyping errors, presence of null alleles and allelic dropouts (Van Oosterhout et al., 2004, Morin et al., 2009). Diversity indices calculated for the mitochondrial data included; haplotype number (n), haplotype diversity (H_d) and nucleotide diversity (π). These parameters were used as a measure of genetic divergence at the mitochondrial d-loop region within and among the sampled populations. All computations were implemented in DnaSP (Librado and Rozas, 2009) and Arlequin 3.5 (Excoffier and Lischer, 2010).

3.2.5 Demographic history

Microsatellite allele frequencies were used to assess the recent demographic history of *B. docmak*. All three microsatellite mutation models (SMM, IAM and TPM 7:3 ratio) were used to test for signatures of population reductions for each of the five populations using Bottleneck 1.2.02 software (Selkoe & Toonen 2006; Cornuet & Luikart 1996; Piry et al 1999).

Historical demography was investigated at the mitochondrial d-loop by estimating *Fu's* F statistic for each of the five locations (Fu, 1997). Neutrality tests to estimate *Fu's* F statistic

were carried out in Arlequin (Excoffier and Lischer, 2010). Furthermore, mismatch analysis distribution was performed for the demographic analysis in which pairwise difference distributions and the frequency of segregating sites were analyzed in DnaSP (Librado and Rozas, 2009, Rozas et al., 2003).

3.2.6 Population structure

Genetic population structure was investigated at the d-loop region of the mtDNA and at 12 polymorphic microsatellite loci. Using both data sets, analysis of molecular variance (AMOVA) and pairwise comparisons between locations was completed in Arlequin 3.5 (Excoffier and Lischer, 2010). For microsatellite data, the AMOVA was based on allelic frequencies and for mtDNA sequence data the AMOVA was based on a genetic distance matrix of pairwise differences between pairs of populations. Significance was estimated at 10,000 permutations.

Furthermore, using mitochondrial sequence data organised in DnaSP (Librado and Rozas, 2009) and exported to Network 4.611 (Flexus Technology, 2012), the population genetic structure and geographical distribution of haplotypes was visualized. Calculation and drawing of a minimum spanning network based on haplotype distribution in sampled individuals from all the five populations was completed in Network4.611 (Flexus Technology, 2012).

Bayesian clustering analysis conducted in the program, STRUCTURE HARVESTER (Earl and VonHoldt, 2012) was used to assign individuals from the five locations into distinct genetic clusters. The analysis was based on microsatellite data at 12 polymorphic loci and runs were conducted on putative populations (K) set from 1-10 iterations with 10,000 burn-

ins followed by 100,000 Markov-Chain Monte Carlo (MCMC) steps for each run. Potential clusters were determined based on an MCMC approach, both with and without a priori definition of structure, and also assuming independent frequencies of alleles. Using the web-based Structure Harvester software (Earl and vonHoldt, 2012), the best K value was then selected. Additionally a multivariate method, the Discriminant Analysis of Principal components (DAPC), was performed in the R package adegenet v1.4.2 (Jombart, 2008), to explore a finer scale structure of the populations based on a two-step procedure. Firstly the genetic data is transformed using a Principal Component Analysis (PCA) and then clusters are identified by Discriminant Analysis (DA) without assuming panmixia (Jombart et al., 2010, Jombart, 2008).

3.3 Results

3.3.1 Genetic diversity

Mitochondrial DNA d-loop variation among 93 *B. docmak* individuals from the five locations revealed high haplotype diversity, H_d, with a narrow range across populations (H_d range of 0.698 for Kazinga Channel to 0.857 for Lake Albert population). The nucleotide diversity was highest in individuals from Lake Albert and the Victoria Nile river, both with $\pi = 0.010$, and least in Lakes Edward and Victoria, $\pi = 0.004$ and 0.006 respectively (Table 3.1). Overall there were 25 distinct haplotypes, with 31 polymorphic sites and a total of 33 mutations (Appendix 3.1).

Analysis of microsatellite data, revealed that all populations were in Hardy Weinberg equilibrium and there was no detection of large allelic dropouts. Null alleles were not detected in the Lake Victoria population; however they were suggested at locus BD02 for individuals sampled from Victoria Nile River, BD01 in the Lake Albert and Lake Edward populations, and BD20 in the Kazinga Channel population. There was no locus that consistently exhibited null alleles, and as such all microsatellite loci were polymorphic across populations with number of alleles up to seven alleles per locus and overall mean allelic diversity $N_a \pm SE = 4.05 \pm 0.19$ (Table 3.1). The allelic diversity was highest in Lake Albert (mean $N_a \pm SE = 4.67 \pm 0.5$) and lowest in the Lake Victoria population (mean $N_a \pm SE = 3.25 \pm 0.49$). Similarly observed and expected heterozygosities were highest in the Lake Albert population (mean $H_o \pm$ SE of 0.55 ± 0.034 and $H_e \pm$ SE of 0.66 ± 0.03) and lowest in the Lake Victoria population (mean $H_o = 0.199$ and $H_e = 0.213$) (Table 3.1). Bottleneck analysis under the SMM and TPM models revealed significant heterozygosity deficiencies

(with P values < 0.005) for *B. docmak* populations from Lake Victoria and Edward. These two populations additionally displayed an L-shaped allelic distribution characteristic of populations undergoing expansion following contractions (data not shown). The effective population size, $N_{eLD,}$ for Lake Victoria, Lake Edward and the Nile River were finitely restricted (Table 3.1). On the contrary, the results showed that the sampled populations were mating randomly with the F_{IS} indices generally low and showing no significant deviation from zero (P >0.05), with the exception of Lake Victoria having the highest F_{IS} of 0.11 (P = 0.054), but still being statistically insignificant (Table 3.2).

3.3.2 Demographic history

Mismatch distribution analyses showed varied demographic histories for *B. docmak* from the five locations. *Fu's* F statistics were negative in all populations and highly significant for *B. docmak* individuals from Lakes Victoria, Edward and the Kazinga channel (see Table 3.1).

Overall, all populations sampled exhibited heterozygosity deficiencies, except for the Lake Albert population that did not show significant deviations from a stable population (P = 0.37, SMM and P = 0.168, under the IAM model). The Lake Albert population exhibited characteristics of a stable population with three loci showing heterozygosity deficiency and nine loci with heterozygosity excess. Contrary to Lake Albert, results indicated that the Lake Victoria population exhibited a normal L-shaped curve characteristic of a young and expanding population. This is further confirmed by a significantly high heterozygosity deficiency at 10 loci (P = 0.000 under the SMM model). Additionally, relative to all populations, Lake Victoria displayed the highest F_{IS} index of 0.01 (Table 3.2) although it was not significant (P > 0.05).

Table 3.1: Microsatellite and mitochondrial DNA diversity indices for *Bagrus docmak* from five freshwater systems in East Africa

Location	Microsatellite data					Mitochondrial data			
	N	N_a (range)	H_o	H_e	N_{eLD} (95 % CI)	*N*	π	H_d	*Fu's*
†Lake Victoria	26	3.25 (1-6)	0.199 (0.00 - 0615)	0.213 (0.000 -.706)	18.2 (6.7-127.4)	13	0.006	0.846	-3.769**
†Victoria Nile	47	3.33 (1-5)	0.225 (0.00 - 0.391)	0.238 (0.000 - 0.48)	58.2 (22.9- 6545)	18	0.010	0.750	-1.461
Lake Albert	16	4.67 (2-7)	0.552 (0.333 - 0.688)	0.66 (0.426 - 0.781)	∞ (111.6 - ∞)	16	0.010	0.857	-1.606
Lake Edward	48	4.42 (2-6)	0.385 (0.126 - 0.542)	0.381 (0.119 - 0.556)	31.3 (18.8 - 60.3)	21	0.004	0.732	-5.049***
Kazinga Channel	38	4.58 (2-6)	0.382 (0.147 - 0.579)	0.38 (0.190 - 0.586)	978.5 (71.1 - ∞)	29	0.009	0.698	-6.033**

Values for microsatellite statistics are means over all loci (range) for each location: *N*, sample size, *Na*, mean number of alleles, H_o, mean observed heterozygosity, H_e, mean expected heterozygosity, N_{eLD} , effective population size as estimated by linkage disequilibrium method (at 95% confidence interval), π, nucleotide diversity and H_d, haplotype diversity. Populations marked with † exhibited monomorphism at one of the twelve loci investigated. Asterisks* on *Fu's* statistic denote the level of significance.

Table 3.2: Inbreeding Coefficient, F_{IS}, and genetic population bottleneck tests for *Bagrus docmak* from five East African locations

Location	FIS (P value)	Bottleneck test: IAM	Bottleneck test: SMM	Bottleneck test TPM
Lake Victoria	0.110(0.0540)	0.007	0.001	0.002
Victoria Nile River	0.0723 (0.0928)	0.007	0.001	0.001
Lake Albert	0.057 (0.1951)	0.339	0.910	0.677
Lake Edward	-0.057 (0.9158)	0.204	0.002	0.043
Kazinga Channel	0.017 (0.3650)	0.064	0.001	0.002

The bottleneck probabilities reported for IAM, SMM and TPM models above were Wilcoxon probability 2-tail tests for heterozygosity deficiency and excess as implemented in the program Bottleneck (Cornuet & Luikart 1996; Piry et al 1999)

Similary, using the linkage disquilibrium method for estimating the effective population size (N_{eLD}) in *Ne*Estimator, Lake Albert showed an infinite N_{eLD} with infinite number of breeders, followed by the Kazinga channel which also exhibited a relatively high N_e (N_e = 998, range 71.1 - ∞). Populations from Lake Victoria and Lake Edward had restricted effective population sizes, N_{eLD} of 18.2 (6.7 - 127.4) and 31 (18.8 - 60) respectively (Table 3.1). The N_{eLD} estimates are consistent with results from the bottleneck test, as well as the mitochondrial results that showed high nucleotide diversity for the Lake Albert population and relatively low nucleotide diversity for *B. docmak* from Lakes Victoria and Edward (Table 3.1).

3.3.3 Population structure

MtDNA d-loop analysis for *B. docmak* populations revealed highly significant structuring between populations across the Albertine Rift valley and the Lake Victoria basin populations (ϕ_{ST} = 0.62; P = 0.000). Pairwise ϕ_{STs} between populations across the two regions ranged from 0.63 – 0.82 (P = 0.000) which were up to seven-fold higher in comparison to the within region population ϕ_{ST} pairs (range of 0.01 to 0.10). Moderate to weak structuring also existed between populations within basins, especially for water systems that were geographically proximate, or directly connected with no major biogeographical barriers. For instance, pairwise ϕ_{ST} values between "Kazinga Channel - Lake Albert" and "Lake Edward - Lake Albert" showed moderately low ϕ_{STs} of 0.05 (P = 0.032) and 0.10 (P = 0.000), respectively (Table 3.3). In both cases there was relatively weak genetic structuring among these populations situated within the Albertine Rift valley. There was no significant structuring between Lake Edward and the Kazinga channel populations (P = 0.125), which is not surprising given their direct geographical connectivity. Similarly, there was weak and

insignificant structuring (P = 0.125) between the Victoria Nile River samples and the Lake Victoria population.

Table 3.3: Comparison of *Bagrus docmak* population pairwise ϕ_{ST} based on the mitochondrial DNA data with ϕ_{ST} values below diagonal and corresponding P values (10000 permutations) above diagonal

Population	Lake Victoria	Victoria Nile River	Lake Albert	Lake Edward	Kazinga Channel
Lake Victoria	-	0.369	0.000	0.000	0.000
Victoria Nile River	0.01	-	0.000	0.000	0.000
Lake Albert	0.72	0.63	-	0.002	0.033
Lake Edward	0.82	0.74	0.10	-	0.125
Kazinga Channel	0.73	0.66	0.05	0.016	-

Overall fixation index, ϕ_{ST} was 0.62 at P = 0.000

Table 3.4 Pairwise F_{ST} among *B. docmak* populations from East Africa using microsatellite data. Below diagonal, F_{ST} values, above diagonal significance level corrected for FDR at 0.001

Population	Lake Victoria	Victoria Nile	Lake Albert	Lake Edward	Kazinga Channel
Lake Victoria	-	0.001	0.000	0.000	0.000
Victoria Nile	0.05	-	0.000	0.000	0.000
Lake Albert	0.36	0.38	-	0.000	0.000
Lake Edward	0.52	0.53	0.15	-	0.029
Kazinga Channel	0.54	0.55	0.16	0.01	-

Results of population structure based on microsatellite analysis mirrored the mitochondrial DNA ϕ_{ST} patterns. Here, population pairwise F_{ST} values were in the range of 0.01 to 0.55, with significant structuring between populations across the Albertine (Lakes Edward and Albert and the Kazinga Channel) and Lake Victoria basin (Lake Victoria and Victoria Nile river populations (Table 3.4). Additionally there was weak, but significant F_{ST} between Lake Victoria and the Nile river populations, F_{ST} 0.05 (P = 0.001).

Bayesian STRUCTURE analysis revealed two distinct genetic clusters based on the highest probability likelihood achieved of K = 2. One genetic cluster predominantly comprised of

the Lake Victoria basin populations (i.e. from the Nile River and Lake Victoria) and the second genetic cluster mainly populations from the Albertine Rift Valley (i.e. Lake Edward and the Kazinga Channel). However, individuals from Lake Albert showed an admixture comprising of both genealogies from the Albertine rift and Lake Victoria basin (Figure 3.2). Further analysis with DAPC revealed three clusters (Figure 3.2), grouping individuals from Lake Albert as a discrete population. Lake Edward and the Kazinga channel grouped into one cluster and equally the Lake Victoria and Nile River individuals into another distinctive cluster.

The haplotype network based on mtDNA data for *B. docmak* was characterised by two centrally shared haplotypes, H-2 and H-9, (one from each of the two regions separated by up to four mutational events; the Albertine rift and the Lake Victoria basin) and 23 other smaller haplotypes that were branching out in the periphery. The most common haplotype was shared by 33 individuals restricted to the Albertine rift (Lake Edward, Lake Albert and the Kazinga channel) and representing up to 35 % of the all sampled individuals. The second most frequent haplotype had 16 individuals restricted to the Nile River and Lake Victoria. From these two major haplotypes span other haplotypes with lower frequencies (Figure 3.3). These results suggest the presence of two fairly distinct genetic groups spanning the five geographical locations sampled.

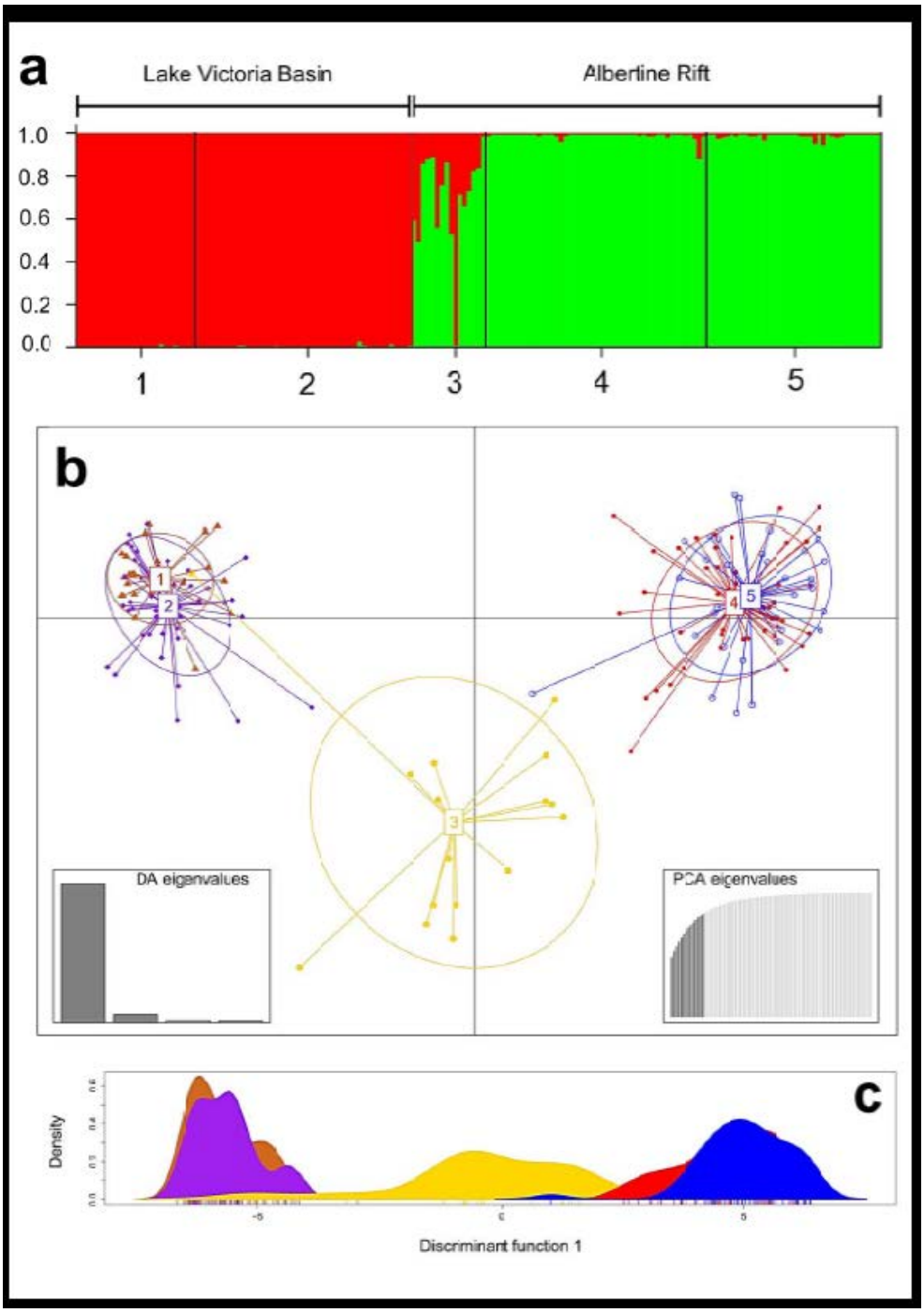


Figure 3.2: STRUCTURE bar plot of Bagrus docmak populations depicting two genetic clusters (K = 2) as per the genome ancestry assignment. The X axis represents individuals from the five locations: Lake Victoria (1), Victoria Nile River, 3- Lake Albert, 4- Lake Edward and 5- Kazinga Channel), assigned to two major stocks. These were; Victoria basin populations (red) and Albertine Rift populations (Green) except for the apparently admixed Lake Albert population that showed two ancestral assignments (Red and Green). b): Scatterplot showing Discriminant Analysis of Principal Components (DAPC) created using the R package. c) DAPC density plot based on the most variable component. Further separation of the Lake Albert (A) population from other Albertine rift populations (i.e. Lake Edward (E) and Kazinga channel (K)) and also a clear separation from Lake Victoria (V) and the Victoria Nile River (N) populations.

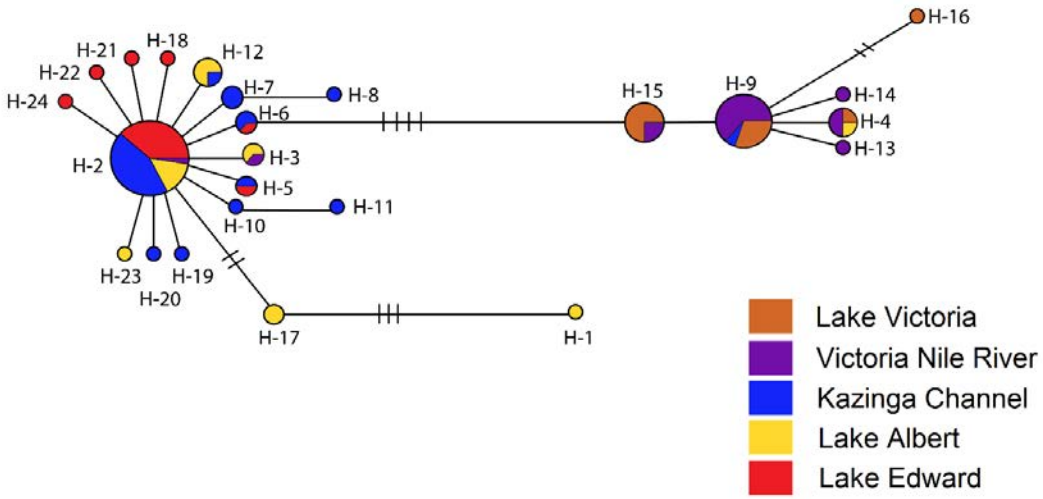


Figure 3.3: Haplotype network for *Bagrus docmak* from five geographical locations as drawn in Network 4.6.1. Each circle denotes a single haplotype whose size is proportional to the frequency of the haplotype. The colours represent the geographical source of the haplotype. Each branch indicates a single mutational event except where indicated by lines that correspond to the total number of mutations.

3.4 Discussion

3.4.1 Genetic diversity, population structure and connectivity of *Bagrus docmak*

The complex geological history of East Africa has acted as a biogeographical evolutionary driver of the diverse freshwater fish species fauna seen in the region (Danley et al., 2012, Verheyen et al., 2003, Sturmbauer et al., 2001). Similarly, genetic analyses of *B. docmak* populations from lacustrine and riverine habitats of the EAR show evidence for vicariance leading to genetic divergence at the level of populations. Both mtDNA and nuclear markers highlight the presence of significant genetic population structure among *B. docmak* populations within and outside the EAR (F_{ST} of 0.46 for microsatellites and ϕ_{ST} of 0.62 for mitochondrial DNA data, P value = 0.000), with two major genetic stocks identified: one within the Albertine rift and the other within the Lake Victoria basin. Mitochondrial DNA analyses show that the separation of stocks has an ancient origin with a high degree of haplotype sorting evident, with just two major haplotypes and restricted mutations of up to four base pair changes between these two major haplotypes. The near complete haplotype sorting and reduction in diversity suggests a severe vicariant and historic bottlenecking event among catfish populations (Figure 3.3). The split lineage and highly truncated haplotype sharing between Lake Edward and Lake Victoria individuals is noteworthy, emphasising the divergence of these two lakes situated in the Albertine rift valley and Lake Victoria basin drainages, respectively. The divergence of populations of this potamodromous freshwater catfish in the EAR lakes and Lake Victoria is consistent with high degree of endemism and radiations among the cichlids that have been well documented in African Great Lakes, including Lake Victoria (Bracken et al., 2015, Johnson et al., 2000, McGlynn et al., 2013, Russell and Johnson, 2001, Seehausen, 2002). Additional evidence to the Pleistocene events

of river reversal is the complete desiccation of Lake Victoria following a long drought period (Johnson, 1996).

Despite similar geological history of formation by tectonic movements and their connectivity through the Semliki River, moderate, but significant structuring was detected between *B. docmak* populations of Lake Edward and Lake Albert using both mitochondrial and microsatellite data. The divergence of *B. docmak* populations between these two lakes is a likely the result of the biogeographical barrier of the Semliki rapids (Thieme et al., 2005) separating the Lake Albert population from the Lake Edward population. This finding of structuring are consistent with earlier findings, wwhere unique fish fauna in Lake Albert have been identified that no longer exist in Lake Edward and similarly this has been attributed to the break in connectivity between the two lake systems by the wide section of rapids (Thieme et al., 2005). This biogeographic barrier plays a significant role in the reduction of gene flow between Lake Edward and Lake Albert and hence the detection of a genetic break between Lake Edward and Lake Albert *B.docmak* populations.

Correspondently, the weak and insignificant F_{ST} values obtained between lake - river systems (i.e. Lake Edward - Kazinga Channel River and Lake Victoria - Victoria Nile River) is consistent with the direct connection of these lake - river systems which have no known biogeographic barriers. In addition, Ssemutundu being potamodromous (Manyala et al., 2005, Chapman et al., 2012), implies that gene flow between lakes and river systems will be enhanced where there are no barriers, but inhibition of gene flow will be accelerated in the presence of geological features that interrupt migration and consequently restrict gene flow. Connectivity would foster the exchange of genes of populations between the lacustrine and riverine habitats.

3.4.2 Historical and contemporary genetic signatures

Understanding biogeography using genetics is important to determine patterns influencing distribution of geographically distant populations. Evidence of sequence and haplotype divergence revealed that *B. docmak* populations (Lake Victoria, Edward and the Kazinga Channel) underwent severe contractions as a result of biogeographical influences within the East African region. Generally mitochondrial haplotypes phylogenetic branches were very short indicating that clades had evolved fairly recently with just up to four mutations between the two major lineages. It also suggest a short evolutionary time since the common ancestor of the two genetic clades. Additional evidence of low diversity in the mitochondrial data reveals that the populations from Lakes Victoria and Edward have been restricted to a very small size over a few thousands of years. High haplotype diversity coupled with low nucleotide diversity in teleosts, as was observed in the Lake Victoria and Edward *B docmak* populations (Table 3.1), have been associated with recent population expansions following bottleneck and low effective population sizes (Grant and Bowen, 1998, Yang et al., 2012). Evidence of a historical bottleneck and a possible recent colonisation was also confirmed by the negative and significant *Fu's* statistic obtained in the Lake Victoria, Lake Edward and Kazinga channel populations (Fu, 1997). The results are consistent with reports of mass extinctions in these lakes and recolonization events, particularly in Lakes Edward, George and Victoria (Beadle, 1974, Thieme et al., 2005). It should be noted that long droughts characterised the East African region during the late Pleistocene coupled with the tectonic uplifts that were responsible for the drying off and river reversal between the lacustrine environments (Day et al., 2013). *Bagrus docmak* being potamodromous was most likely susceptible given its migratory lifehistory.

Corroborative evidence was obtained using the microsatellite data in which recent population expansions following contractions were revealed by the significant heterozygosity deficiencies in all the populations, except for Lake Albert. Restricted N_{eLD} estimates realized especially in Lakes Edward and Victoria further confirm contemporary pressures in these lake systems in contrast to Lake Albert populations that showed an infinite N_{eLD} estimate (Table 3.1) and hence a higher degree of stability. The low N_{eLD}, relatively high inbreeding coefficient, and significant heterozygosity deficiencies in the Lake Victoria and Edward populations, further indicate that these populations are constrained and have not had enough time to recover from the severe historic bottleneck. As pointed out in earlier studies the *Bagrus docmak* population in Lake Victoria is constrained by anthropogenic pressures such as heavy fishing pressure, but its decline is more probably linked to the introduction of the Nile perch that has been clearly documented in the region (Hauser et al., 1998, Kudhongania et al., 1992, Ogutu-Ohwayo, 1990b, Ogutu-Ohwayo, 1993, Goudswaard and Witte, 1997). Evidence highlights that Ssemutundu was formally a dominate and higher-order carnivorous fish in the Lake Victoria basin, but it currently faces direct competition for prey and is also preyed upon by the introduced Nile perch (Goudswaard and Witte, 1997).

Furthermore the present study reveals a recent expansion amidst high, but insignificant F_{IS}, thus rejecting the hypothesis of possible inbreeding. It is therefore predicted that the large surface area of Lake Victoria (68,000 Km^2) provides an advantage of a large habitat in which the species will survive if proper management strategies are put in place. Conversely, Lake Edward, which is a small (2235 km^2) EAR lake, has been considered hydrologically and chemically sensitive to climatic changes as it lies in the intersection of the Indian and

Atlantic air masses (Russell and Johnson, 2001). The truncated gene flow between Lakes Edward, Victoria and Albert due to geological and biogeographic barriers largely exposes the Lake Edward population to risk, especially in the event where the population will be exposed to stochastic events (such as disease outbreaks) and anthropogenic pressures (including heavy fishing and habitat degradation).

The exception of Lake Albert and the Victoria Nile River in showing negative, but non-significant neutrality tests, reveals the possibility of a long-term demographic stability. The microsatellite data confirms the stability of the Lake Albert system following a balanced heterozygosity consistent with a stable population as assessed by the bottleneck test. Furthermore, the Lake Albert population was characterised by an infinite N_{eLD} estimate coupled with the highest level of heterozygosity and number of private alleles. This implies that new alleles are being generated, but are not being lost due to genetic drift. The stability evidenced in the Lake Albert populations suggests presence of undisturbed habitats within the EAR; Lake Albert that has also been referred to as an abyss with deep waters compared to the shallow Lake Victoria (Mwanja et al 2014).

The findings in the current study do not in any way warrant populations of *B. docmak* from Lake Albert to be neglected; however they provide a baseline of the current genetic status for the species following past historical processes (Johnson et al., 2000, Johnson et al., 1996, Russell and Johnson, 2001) and call for the sustainable management and conservation of the species. Amidst current anthropogenic pressures including aquaculture establishments (Dickson et al., 2012) in the lake, as well as the proposed oil extraction within the Albertine region (Vokes, 2012, Kathman and Shannon, 2011) that could accelerate habitat destruction and loss of genetic biodiversity, a genetic baseline is needed to monitor and design

appropriate management measures for the lake and the region at large.

Information provided by the current study will facilitate a comprehensive management of Ssemutundu and related taxa for sustainable harvest in the wild and/or culture under captive conditions. The strategy should consider river and lake regulations and their associated developments such as damming, hydropower generation, fishing, and transport, among others. Replenishment of fish stocks through restocking and aquaculture need to take into account the current and historic genetic diversity and population structure of the identified stocks. These stocks potentially translate into majorly two evolutionary significant units (evidence from mitochondrial data) and two management units (based on STRUCTURE analysis). These findings are important in that, appropriate conservation and management strategies depend on our ability to correctly assign genetically distinct populations (Latch et al., 2006).

3.5 Conclusion

The low genetic diversity and strong population structure patterns revealed in Ssemutundu are consistent with patterns observed in higher-order species composition studies and that have been linked to the complex biogeography of East Africa. The results in the current chapter support the hypothesis of (Danley et al., 2012, Pinton et al., 2013) who hypothesize that a linkage of paleohydrological changes in a geological context has been a major cause of diversification of the freshwater teleost fauna in East Africa; more precisely, the East African Rift system and associated historical events appears as a strong driver of freshwater diversification and evolution.

Strategically, the management of aquatic fauna in the East African region would initially consisder taking into account the two major management units identified i.e. the Albertine rift valley region and the Lake Victoria basin. However, further research is required for the Lake Albert population which was singled out as a discrete group at a finer scale with DAPC analysis to make a total of apparently three management units comprising Lake Edward and Kazinga Channel cluster, Lake Victoria – Nile River and Lake Albert populations.

CHAPTER 4

Restricted gene flow and strong genetic structure in Nile Perch, *Lates niloticus*, from African freshwater rivers and lakes

Abstract

Geological evolution of the African continent has been subject to complex processes including uplift, volcanism, desert formation and tectonic rifting. This complex geology has created substantial biogeographical barriers, and coupled with anthropogenic introductions of freshwater fishes, has had an impact on the genetic diversity, connectivity and sub-structuring of the teleost fauna. Nile perch, *Lates niloticus*, is an iconic fish in Africa and is of high commercial importance, both in the species' native range and where it has been translocated. However, the species is in decline and there is a need to understand its population genetic structure to facilitate sustainable management of the fishery and aquaculture development.

Nile perch tissue samples were acquired from two West (Senegal River and Lake Kainji on the Niger River) and four East African (Lakes; Albert, Kyoga, Victoria and Turkana) locations. Nineteen polymorphic microsatellite loci were used to study the genetic variation among populations across regions (West and East Africa), as well as between native and introduced environments within East Africa. Results revealed strong and significant genetic structuring among populations across the sampled distribution (divergence across regions, F_{CT} = 0.26, P = 0.000). STRUCTURE analysis at a broad scale revealed K =.2 clusters, the West African individuals were assigned to one cluster, while all individuals from the East African region, regardless of whether native or introduced, were assigned to another cluster. Analysis at a regional scale revealed further structuring of up to K = 3 genetic clusters in East

African Nile perch. Significantly ($P < 0.05$) lower genetic diversity based on analysis of allelic richness (A_R) was obtained for the two translocated populations of Lake Kyoga ($A_R = 3.61$) and Lake Victoria ($A_R = 3.52$), compared to Nile perch populations from their putative origins of Lakes Albert ($A_R = 4.12$) and Turkana ($A_R = 4.43$). The lower genetic diversity in the translocated populations may be an indication of previous bottlenecks and may also indicate a difficulty for these populations to persist and adapt to climatic changes and anthropogenic pressures that are currently present in the East African region. On the contrary, except for Nile perch from Lake Turkana, all Nile perch populations examined indicated signs of heterozygote excess with positive and significant F_{IS} values ($P < 0.05$), a critical finding useful in identifying genetically diverse aquaculture foundational stocks. Finally, the distinct genetic clusters identified in the current study between the West and East African Nile perch have been maintained by presence of biogeographic barriers and restricted gene flow between the two regions and as such the two genetic groupings should be managed separately.

4.1 Introduction

Advances in the knowledge-base of African freshwater fisheries have been made over the last five decades (Odada et al., 2003, Odada and Olago, 2006, Lowe-McConnell, 2009). Like elsewhere in the world, these advances in knowledge have been driven mainly by threats that are facing the sustainable exploitation of fisheries resources. For African fisheries these threats include increases in urbanisation linked to human population growth, exotic fish introductions, overfishing, and sedimentation / pollution as a result of changes in land utilisation. The impacts of these threats are particularly noticable in the freshwater lake basins of East Africa (Lowe-McConnell, 2009, Luck, 2007, Butchart et al., 2010). Most freshwater fish research has focused on the evolution of one group of African fishes, the haplochromine cichlids (Cichlidae) in the Great Lakes region. Several mutually contradictory reports have characterised the haplochromine cichlids' high level of endemism and population demise as a result of human-induced impacts (Danley et al., 2012, Elmer et al., 2009, Johnson et al., 2000, Guo and Thompson, 1992, Ogutu-Ohwayo, 1993, Seehausen, 2002). However, many other African fish species remain data deficient in the face of declining fisheries. This poses a serious threat to the economic development of African communities that largely depend on fish in these lakes and rivers for their livelihoods (Achieng, 1990, Balirwa et al., 2003, Ogutu-Ohwayo, 1990a, Mkumbo and Marshall, 2015).

The genus *Lates* comprises fish from 11 species which occur in both brackish and freshwater. These species are widely distributed across Africa, Asia and Australia (Froese, 2009). *Lates niloticus,* the Nile Perch*,* is a freshwater species endemic to African rivers and lakes including the Nile, Chad, Senegal, Niger and Congo River basins. This species also naturally occurs in the East African Rift (EAR) Lakes of Tanganyika, Albert and Turkana. However, more

recently the range of Nile perch was expanded through anthropogenic introductions about 50-60 years ago by British colonialists, to include Lakes Nabugabo, Kyoga and Victoria (Hauser et al., 1998, Pringle, 2005, Balirwa et al., 2003, Kudhongania et al., 1992).

Stocking reports for Nile perch are contradictory, but generally indicate that the species was introduced into Lakes Victoria and Kyoga in the 1950's and 1960's (number of stocked individuals ranging from 8-585), where the species has now become established (Balirwa et al., 2003, Basiita et al., 2011, Kenchington, 1939, Kudhongania et al., 1992, Mwanja et al., 2012, Ogutu-Ohwayo, 1993, Ogutu-Ohwayo, 1994, Pringle, 2005). Nile perch introduced into Lake Kyoga putatively originated from Lake Albert, whilst fish introduced into Lake Victoria originated from both Lakes Albert and Turkana in the western and eastern arms of the EAR respectively (Pringle, 2005). Introduced populations have become important to the food security and economic development of communities, with Nile perch introduced into Lake Victoria, for example, being the major commercial freshwater fishery in Africa which supports directly and indirectly up to four million people in East Africa (Mkumbo and Marshall, 2015). The fishery from Lake Victoria, contributes up to USD 350 million (from EU exports alone) to the GDP of the three East African countries (Uganda, Kenya and Tanzania) (Mkumbo and Marshall, 2015). Nevertheless, there is disagreement among biologists, conservations and economists with regard to the social, environmental and economic impacts of Nile perch introduction to these lakes (Reynolds and Greboval, 1988, Kudhongania et al., 1992, Ogutu-Ohwayo, 1994, Pringle, 2005, Bundy and Pitcher, 1995).

While anecdotal observations and official records provide a historical understanding of the pattern of introduction of Nile perch into various lakes within the East African region (Pringle, 2005), there is still some controversy and limited knowledge of the actual

biogeography of the Nile perch in Africa. Information on the genetic structure of Nile perch in Africa at a broad-scale, has been limited to mitochondrial DNA (mtDNA) analyses in which a single West African population (from the Senegal River), and six East African populations were sampled to assess the evolutionary history of the species using mtDNA variation (Mwanja et al., 2013). In this study, two historical genetic groupings between East and West Africa were identified, with no shared haplotypes between these two genetic groups (Mwanja et al., 2013). Within East African populations, mtDNA and nine microsatellites developed from a sister species, *L. calcarifer*, showed at least two genetic subpopulations of Nile perch (Mwanja et al., 2014b, Mwanja and Mwanja, 2008, Mwanja et al., 2012, Mwanja, 2013). Fragmented genetic studies limited by the scale of sampling and use of few microsatellite loci, may compromise concrete deductions on the genetic structure and diversity of Nile perch in Africa. Therefore, a holistic Nile perch study to genetically characterise the species across its range in Africa and the clarification of the present genetic stocks using sufficient robust markers is critical.

Genetic characterisation of fish populations can provide useful management information about the probable historical and contemporary sizes, as well as compositions of populations, movement patterns and gene flow (both natural and human mediated) and identification of unique genetic stocks that need to be managed separately (Mychajliw and Harrison, 2014). Additionally from aquaculture and fishery restocking perspectives, a lack of information on stock origin of broodfish of a target species, including population diversity and relatedness, can limit the success of targeted selective breeding programs and augmentation of overexploited wild stocks. The current study investigated the population genetic structure of the Nile perch in six populations from Western and Eastern Africa. Unravelling the genetic

structure of Nile perch will be useful in understanding how the interactions between biological, physical and geological processes have shaped populations of the species in its natural and introduced range. Information on the broad-scale genetic structure of Nile perch will be critical in achieving successful aquaculture establishments, conservation and management of the wild fisheries.

4.2 Methods

4.2.1 Study Area

Nile perch were sampled over a wide geographical range in Western and Eastern Africa to capture patterns of both biogeographical and human mediated gene flow of the species in both native and introduced environments. In Western Africa, sample locations included the Senegal River (n = 18) and Lake Kainji, Niger River (n = 30). Eastern African samples were acquired from the species native habitats of Lakes Albert (n = 48) and Turkana (n = 18), as well as from introduced populations from Lakes Kyoga (n = 38) and Victoria (n = 38) (Figure 4.1). Important to note is that this study was limited to Western and Eastern African populations (Figure 4.1), as samples could not be obtained from north of Lake Albert (Albert Nile) to the lower parts of the Nile where the habitat gets brackish towards the Mediterranean sea.

4.2.1.1 Senegal River basin

The Senegal River basin which drains into the Atlantic Ocean, is located in Western Africa and covers 1.6 % (483,181 km^2) of the African continent spreading over four countries (Senegal, Mauritania, Mali and Guinea). The Nile perch samples for this study were captured below Dagana, at Richard-Toll, where the Senegal enters its delta with the Diama dam. The dam was constructed in 1985 and currently acts as a low water storage dam and salt-water barrier near the mouth of the Senegal River (Hargreaves et al., 1985).

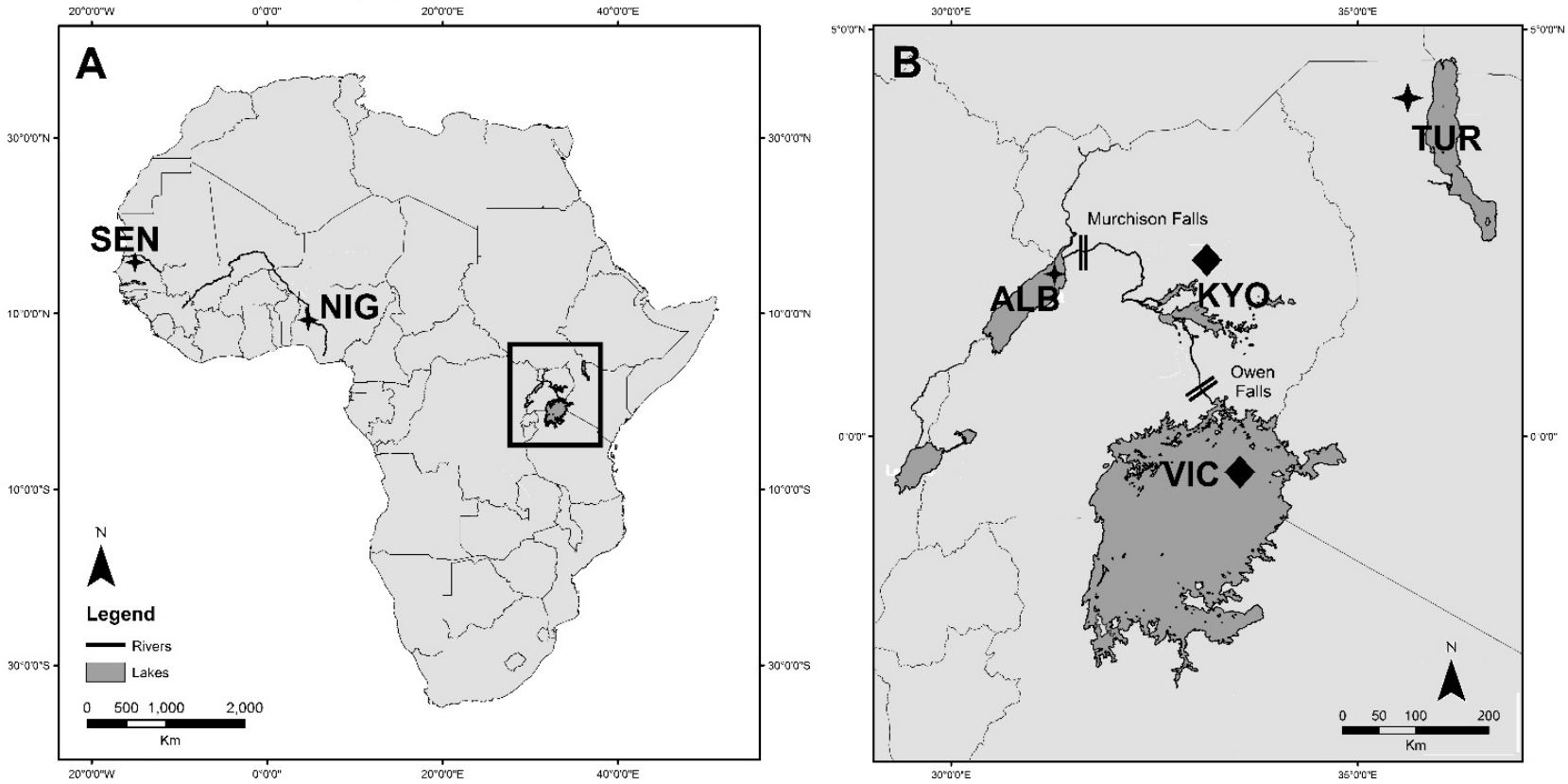


Figure 4.1: Maps A & B showing Nile perch sampling locations. Stars denote where populations of Nile perch were sampled from their natural range and the diamonds denote locations sampled which comprise translocated populations. A: Map of Africa showing the water bodies from which samples of L. niloticus were obtained; the Senegal River (SEN) and Lake Kainji on the Niger River (NIG), both in West Africa, while the black box denotes the area in East Africa from which East African Nile perch samples were obtained. B- Excerpt showing details of sampling locations in East Africa; Lake Albert (ALB), Lake Kyoga (KYO, Lake Victoria (VIC) and Lake Turkana (TUR).

Fishing and Agriculture are the major economic activities in the Senegal River basin, with overfishing of Nile perch, a native species to the drainage, evident given its high economic importance in the region. (Ronal James Harrison Church, accessed on 26/03/2016 at http://www.britannica.com/place/Senegal-River)

4.2.1.2 Lake Kainji on River Niger

Lake Kainji located in north western Nigeria is the largest man-made lake in Western Africa formed in 1968 by damming part of the Niger River. The lake covers an area of up to 1,300 km^2. Lake Kainji (from which the samples were obtained in the current study) is important and extensively used for fishing and irrigation in the region (Balogun, 1986, Ita, 1978).

4.2.1.3 Lake Albert

Lake Albert is located at the tip of the eastern arm of the EAR. Lake Albert drains directly into the White Nile, to which it is believed to be the main regulating water source for the Nile responsible for making the Nile a permanent river as opposed to a seasonal river (Beuning et al., 1997, Holden, 1967, McGlynn et al., 2013, Pickford et al., 1993). The lake is connected to Lake Kyoga to the east by the Victoria Nile and connected to Lake Edward in the south by the Semliki River.

4.2.1.4 Lake Kyoga

Lake Kyoga is centrally located in Uganda. Lake Kyoga is connected to Lakes Victoria and Albert by the Victoria Nile River; however, despite the connection are separated by a series of cascading falls, the Murchison falls located just before the Victoria Nile drains into Lake

Albert. At the top of the falls, the Victoria Nile narrows to a width of about seven meters and drops by a height of about 400 feet. Geologically the Murchison falls were formed as a result of riverine retreat from a buried fault line and superposition from deeply lateralised rift surface sediments (Bishop, 1965). Similarly, to the south of the Victoria Nile is the Owen falls and Rippon falls separating Lake Kyoga from Lake Victoria. The falls on the northern and southern parts of the Victoria Nile River act as biogeographic barriers across the three water bodies restricting the free movement of fish species, including the Nile perch (Mwanja et al., 2012).

4.2.1.5 Lake Victoria

Lake Victoria is the largest tropical freshwater lake in the world and is home for the majority of Africa's diverse haplochromine species. Currently three commercial species are heavily fished in the lake; *L. niloticus* (Nile perch), *Orechromis niloticus* (Nile tilapia) and *Rastrineobola argentea* (Silver fish). Although introduced into Lake Victoria, Nile perch is the primary fished species for both local consumption and export (Pringle, 2005). Due to heavy fishing pressure Nile perch populations have sharply declined in the lake over the last decade and are becoming threatened (Balirwa et al., 2003, Mkumbo and Marshall, 2015).

4.2.1.6 Lake Turkana

Turkana is a large holomictic endorheic rift valley lake located in the eastern arm of the EAR; the lake is generally considered to be understudied, but biologically and environmentally stable (Kolding, 1995, Beadle, 1981). Historically, there is evidence that Lake Turkana in the western arm of the EAR was linked until recently (7000 years ago) to the Nile river system

(Beadle, 1981, Dgebuadze et al., 1994). The lake is central to the livelihoods in the arid northern Kenya and southern Ethiopia providing both fisheries and water resources.

4.2.2 Sample preservation

Samples from all the six sampling locations were obtained from commercial line fishers, except for the samples from Lake Victoria which were captured by trawling. From each individual fish, a fin clip was obtained from the caudal fin and preserved in 20 % dimethyl sulfoxide (DMSO) saturated with sodium chloride salt (Dawson et al., 1998, Amos and Hoelzel, 1991), or in 70% ethanol. All fin clips were transferred in fresh DMSO salt solution prior to being shipped to the Molecular Ecology and Evolution Laboratory (MEEL) in Townsville, Australia, where they were stored at -20 $^{\circ}$C prior to DNA extraction.

4.2.3 DNA extraction, Polymerase Chain Reaction (PCR) and genotyping

Total genomic DNA was extracted using the Bioline Isolate II Genomic DNA kit (Bioline) following the manufacturer's protocol. Briefly; the protocol involved a pre-lysis stage using 180 µl Lysis buffer with 25 µl of protease K incubated at 56 ^{o}C for 1 hr. A second lysis buffer, G -3 (200 µl) was added and further incubated at 70 ^{o}C for 10 min. In order to alter the buffer conditions, 210 µl of 100 % EtOH were then added prior to two (GW1 and GW2) buffer washes. DNA was then eluted using a warm (70 ^{o}C) elution buffer and later the eluted DNA was stored at -80 ^{o}C prior to downstream PCR. DNA quality and quantity was assessed using a ND-1000 Spectrophotometer (Nano-Drop® Technologies).

4.2.3.2 Polymerase Chain Reaction

DNA from 192 individuals were genotyped at 19 polymorphic microsatellite loci: 10 markers were species specific (used in two suites; NP3 and NP1B each of four and six markers respectively) and nine markers (in a single suite, P1 suite) were from a sister species; *L. calcarifer* (Basiita et al., 2015, Zhu et al., 2006). For all markers, the forward primers were fluorescently labeled using the 5- dye system (6-FAM, VIC, NED, PET and LIZ GS-500 size standard) and reverse primers were pigtailed (Brownstein et al., 1996) to ensure consistent amplification and minimize stuttering. A total volume of 10 µl PCR reactions were run on a Biorad C1000 thermocycler under the following conditions: an initial denaturation at 95 °C for 5 min, 6 cycles of 95 °C for 30 s (denaturation)/59 °C for 90 s (annealing)/72 °C for 30 s (extension), 10 cycles at reduced annealing temperatures of 57 °C, 55 °C and 53 °C, prior to a final extension at 60 °C for 30 min.

4.2.3.3 Genotyping

Visualisation of PCR product was performed on an ABI-3730 instrument (Applied Biosystems) using a 5-standard dye system (6-FAM, VIC, NED, PET and LIZ GS-500 size standard) at the Georgia Genomics Facility, USA. Alleles were scored using Genemarker 2.4 (Softgenetics) and checked for genotyping errors, stuttering and null alleles in Microchecker 2.2.3 (Van Oosterhout et al., 2004). Successful PCR amplification was achieved in all individuals except for two that failed to amplify at several loci. These two individuals were consequently removed in all downstream analyses.

4.2.4 Genetic diversity among populations

Deviations from Hardy-Weinberg equilibrium (HWE) and locus linkage equilibrium were evaluated using GenAlex 6.41 (Peakall and Smouse, 2012) and Genepop on the web v.4.2 (Rousset, 2008) respectively. In order to identify genetically diverse populations, diversity indices including expected heterozygosity (H_e), observed heterozygosity (H_o), mean number of alleles (N_a), allelic richness (AR), rare and private alleles were calculated in GenAlEx 6.41 (Peakall and Smouse, 2006). Additionally, allelic richness and rare alleles, while adjusting for small sample size, were calculated in FSTAT and HP Rare Version 1.1, respectively (Goudet, 2002, Kalinowski, 2004). Furthermore, the Wilcoxon test was performed between population pairs of diversity indices including private alleles for native and translocated populations in order to test for reductions in genetic diversity that may occur following translocation of populations from their native to new environments. All populations were tested for deviations in heterozygosity using the inbreeding coefficient, F_{IS}, calculated in Genetix vs 4.05 (Belkhir et al., 2004).

A heterozygosity excess test for recent demographic history for the Nile perch was executed in bottleneck 1.2.02 (Cornuet & Luikart 1996; Piry et al 1999) for each of the six populations. Although the SMM is consensually the accepted mutation model for microsatellites assuming that they evolve by addition or subtraction of one or more repeat units, the infinite stepwise model (IAM) was additionally considered in the analyses given that sometimes there could be violations to the SMM and as such the IAM becomes a sensitive model in such scenarios (Selkoe and Toonen, 2006). As such the two phase model (TPM) in the proportion of 7:3 (IAM: SMM) was used and differences tested using the Wilcoxon test. The allele frequency mode shift was also explored in order to determine the populations that have undergone a

bottleneck or recent population reduction. Additionally, effective population sizes (N_e) for each population, were estimated using the linkage disequilibrium method as implemented in NeEstimator vs 2 (Do et al., 2014).

4.2.5 Genetic differentiation and population structure

Standard genetic tests were performed to examine the amount of genetic differentiation among the six sites sampled. Using an analysis of molecular variance (AMOVA) the level of genetic variance partitioned among groups (East and West African populations), among populations, individuals within populations and that within individuals was determined. A subsequent AMOVA test was then performed, excluding samples from the two West African locations in order to test for the levels of genetic variance in the four East African populations. Global F_{ST} testing was used to determine whether significant genetic differentiation existed between populations. Following the significant overall F_{ST} values obtained (at $P < 0.001$), pairwise genetic differentiation tests were performed on all the six populations after Weir and Cockerham (1984). AMOVA and pairwise F_{ST} estimates were all calculated in Arlequin 3.5 (Excoffier and Lischer, 2010) and Genepop (Rousset, 2008), with estimates of population differentiation derived from 10,000 to 16,000 permutations of the data. In order to correct for multiple comparisons and the increased probability of committing a type 1 error, P-values for pairwise F_{ST} were corrected using Bonferroni correction at $P < 0.001$ (Benjamini and Hochberg, 1995).

In order to test whether Nile perch populations conform to an isolation by distance (IBD) model of gene flow, genetic distance (pairwise F_{ST}) and geographic distance (km) matrices were constructed and correlations analysed using a Mantel's test (1000 permutations) in

GenAlex 6.41 (Peakall and Smouse, 2012). Two IBD analyses were performed, one including all the six populations from West and East Africa, and the second analysis, taking into account only East African Nile perch populations.

The Bayesian assignment test implemented in STRUCTURE 2.3.3 (Pritchard et al., 2000) was used to determine the number of inferred genetic clusters and to test the proportion of genetic admixture among all the six populations. The implementation of individual-based clustering in this program creates clusters of individuals within which the Hardy Weinberg equilibrium is maximised and linkage disequilibrium minimised, thus assigning individuals into gene pools, or breeding groups (Ball et al., 2011). In order to evaluate the number of genetic clusters, analyses involved testing up to 10 possible clusters (from K = 1 to K = 10), followed by 10 iterations using the admixture model with the assumption of correlated allele frequencies and no location prior information provided. Similar tests of the admixture model were carried out, but this time assuming that the allele frequencies are independent; except for this modification all other input parameters in STRUCTURE analyses were left the same (i.e. a burn-in period of 100,000 iterations and 100,000 MCMC). The most likely number of distinct genetic clusters (K) was determined using STRUCTURE HARVESTER (Earl and VonHoldt, 2012), which uses the methods described by Evanno et al. (2005).

Further visual representation of gene flow and relative migration among populations was performed using divMigrate analyses in the *R* diveRsity package (Keenan et al., 2013). The visual network derived depicts arrows that indicate the direction and magnitude of gene flow between populations.

4.3 Results

4.3.1 Genetic diversity indices and differentiation

Microchecker did not detect any evidence of scoring errors due to stuttering and large allelic dropouts at all loci examined. Linkage disequilibrium was observed in only 0.05% (8 out of 172) of loci across all populations, following Bonferroni correction set at $P < 0.001$ (Appendix 4.1& 4.2). However, deviations from HWE were observed at three loci among the translocated Nile perch populations of Lakes Kyoga and Victoria from East Africa. These loci were; *Ln11* (in samples from Lake Kyoga), *Ln23* (samples from Lake Victoria) and *Ln31* (samples from lakes Kyoga and Victoria). These deviations from HWE (at $P < 0.001$ after Bonferroni correction) may suggest a detection of non-random mating within the two introduced populations.

Most loci were polymorphic across populations, except for *Ln09* and *Ln23*, which were monomorphic in the Lake Turkana population, and *Lca21* which was also monomorphic in the Lake Kainji (Niger) population. The mean number of alleles (N_a) per locus for all six populations was 5.13 ± 0.02 (mean ± SE) and number of alleles in the polymorphic loci ranged from 2-14 alleles (Appendix 4.4d). Among East African Nile perch populations, private alleles were highest in native Nile perch populations, but low in the translocated populations of Lakes; Kyoga and Victoria (Figure 4.2).

The allelic richness (A_R) across all 19 loci in the six populations ranged from 2.37 at *Lca21* to 9.54 at locus *Ln17* (Table 4.1 and Appendix 4.4a). The A_R of native populations across the two regions of west and East Africa were comparable with a narrow range ($A_R = 4.19 \pm 0.40$ to $A_R = 4.88 \pm 0.58$ for Lake Albert in East Africa and Lake Kainji in West Africa

respectively). Analysis of A_R as a measure of genetic diversity found a significantly ($P < 0.05$) lower A_R among individuals from introduced (Lakes; Kyoga and Victoria) populations than in individuals from the putative origins (Lakes Albert and Turkana) in East Africa (Table 4.1 and Appendix 4.4a & b). Similarly the private alleles were lowest in the translocated Kyoga and Lake Victoria populations and highest in native Lake Albert and Turkana populations of Nile perch (Figure 4.2 and Appendix 4.4e). A similar trend was observed for the rare alleles where there was a statistical, difference between the introduced and native Nile perch populations within East African Nile perch (Appendix 4.4f).

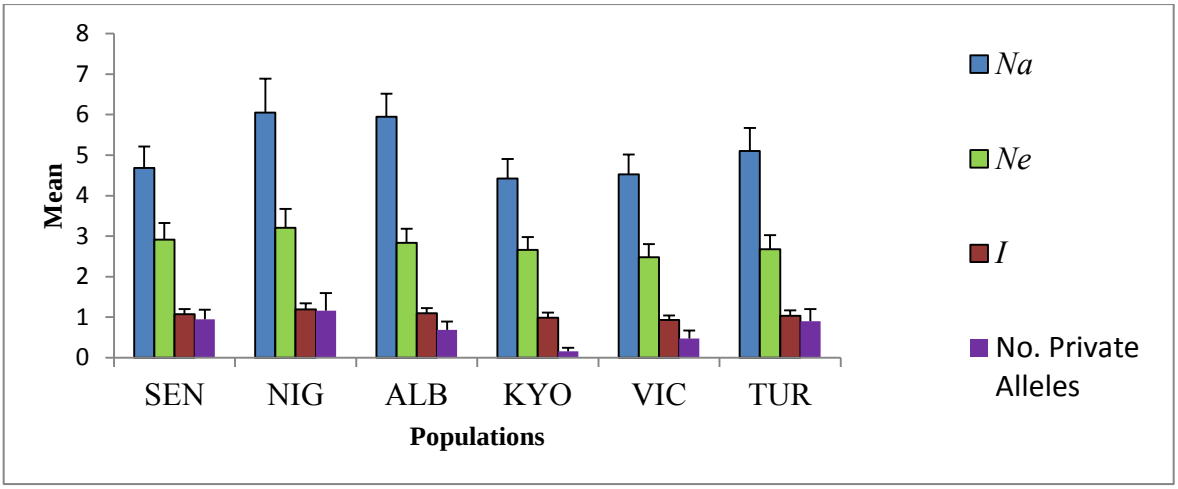


Figure 4.2: Mean allelic patterns in Nile perch plotted against each population. The Y – axis, shows the mean number of alleles (*Na*) , *Ne*, mean number of effective alleles and Mean number of private alleles ,while the X-axis shows the populations ; SEN - Senegal, NIG - Lake Kainji on the Niger River, ALB- Lake Albert, KYO- Lake Kyoga, VIC- Lake Victoria and TUR- Lake Turkana. Error bars represent standard error (SE).

The overall mean expected and observed heterozygosities (Mean ± SE) were 0.51 ± 0.03 and 0.52 ± 0.02, respectively across all populations. There were no large variations in heterozygosity observed amongst populations, although Lake Albert showed the lowest mean H_o (0.45 ± 0.05) and Lake Victoria the lowest H_e (0.48 ± 0.06). All diversity indices, including the mean expected and observed heterozygosity, for each population across all loci are listed in Table 4.1 and Appendix 4.4.

Lake Turkana exhibited the only negative mean F_{IS} value of – 0.029, while Lake Albert had the highest positive F_{IS} of 0.148. The high F_{IS} within Lake Albert could be a result of sub-structuring of the Nile perch within the lake, and this is consistent with the admixture observed in this Albert population as per the STRUCTURE bar plot. Results of the bottleneck tests did not detect significant deviations in allele frequencies, and in all instances an L-shaped curve was obtained (results not shown). However, with the Wilcoxon test, significant heterozygosity deficiencies were identified within the native populations of Nile perch from

Lake Albert and Turkana (P = 0.004) under the IMM model and for Lake Kyoga a heterozygote excess was identified under the IAM model (P = 0.03). The effective population size (*Ne*) at the lowest frequency of 0.05 ranged from (*Ne* = 20.3; 95% CI 15-27.5) in Lake Kainji to (*Ne* = 253; 95% CI 38.4-∞) in the Senegal River. Effective population size, *Ne,* in East African populations ranged from *Ne* = 39.6; 95% CI 28.8-58.8) in Lake Albert to *Ne* = 79.1; 95% CI 45.5-213) in Lake Victoria. Except for the Senegal population, in all other populations the *Ne* was finite and restricted.

Table 4.1: Microsatellite DNA statistics for Lates niloticus from six populations in West and East Africa

Location	*N*	No. of alleles	N_a ± SE (range)	A_R ± SE (range)	H_o ± SE (range)	H_e ± SE (range)	F_{IS}
Senegal River	18	89	4.68 ± 0.53 (2-9)	4.31 ± 0.47 (2-9)	0.56 ± 0.66 (0.14 - 0.94)	0.54 ±0.05 (0.18–0.87)	0.009*
Lake Kainji -Niger River	30	115	6.05 ± 0.84 (1-14)	4.88 ± 0.58 (1-11)	0.52 ± 0.07 (0.00 - 0.92)	0.55 ±0.06 (0.00-0.87)	0.078*
Lake Albert	48	113	5.95 ± 0.57 (3-12)	4.19 ± 0.40 (2-9)	0.45 ± 0.05 (0.09- 0.78)	0.54 ±0.05 (0.09-0.86)	0.148*
Lake Kyoga	38	84	4.42 ± 0.49 (2-8)	3.61 ± 0.35 (1-7)	0.51 ± 0.06 (0.03- 0.87)	0.51 ±0.06 (0.03 - 0.82)	0.019*
Lake Victoria	38	86	4.53 ± 0.49 (2-10)	3.52 ± 0.35 (1-8)	0.50 ± 0.06 (0.03 - 0.87)	0.48 ±0.06 (0.03 - 0.84)	0.002*
Lake Turkana	18	97	5.11 ± 0.57 (1-9)	4.43 ± 0.458 (1-8)	0.54 ± 0.08 (0.00-1.00)	0.51 ±0.06 (0.00-0.83)	-0.029*

Values for microsatellite statistics are means over all loci (range) for each location: *N*, sample size, N_a, mean number of alleles, A_R, allelic richness, H_o, mean observed heterozygosity, H_e, mean expected heterozygosity, F_{IS}, inbreeding coefficient (all F_{IS} values with * were significant at $P< 0.05$)

4.3.2 Genetic differentiation and structure

Nile perch found to show strong genetic structure, especially between West and East African populations (Global F_{ST} of 0.26, P < 001). Genetic variation among the six populations was partitioned into various sources, with total genetic variance predominantly due to within individuals (70.2%) and lowest was between groups (0.3%). Pairwise F_{ST} between populations across the West and East African region were very high and significant (F_{ST} range of 0.33 to 0.42, P<0.001), although the F_{ST} were relatively shallow among populations within groups (F_{ST} range of 0.07 to 0.20, significant P < 0.001) (Table 4.2, Appendix 4.3).

Table 4.2: Estimates of pairwise F_{ST} between six Nile perch populations. Below diagonal is F_{ST} and above significance (sequential Bonferroni corrected P-value ≤0.001) denoted by an * symbol above diagonal.

	Senegal River	Kainji (Niger River)	Lake Albert	Lake Kyoga	Lake Victoria	Lake Turkana
Senegal River	-	*	*	*	*	*
Kainji (Niger River)	0.14	-	*	*	*	*
Lake Albert	0.38	0.34	-	*	*	*
Lake Kyoga	0.39	0.35	0.13	-	*	*
Lake Victoria	0.42	0.40	0.08	0.17	-	*
Lake Turkana	0.42	0.40	0.13	0.21	0.18	-

The high and significant F_{ST} indicate major differences in allele frequencies between the East and West African populations. The highest pairwise genetic differentiation was between the Senegal River and the two most geographically distant East African populations of Lake Turkana and Lake Victoria (F_{ST} = 0.42, P < 0.001), which were the most geographically distantly sampled populations. The lowest pairwise F_{ST} was between Lake Victoria and Albert (F_{ST} = 0.08), although it was still significant (P = 0. 000 after sequential Bonferroni correction).

Furthermore, the Mantel's test revealed a positive and significant correlation (R^2 = 0.86, P = 0.01) (Figure 4.3). This significance seemed to have been driven by the strong differentiation of

West African samples from the East African samples in the analysis, since at a within East African regional scale there was a very low and non-significant correlation ($R^2 = 0.165$, P = 0.2)

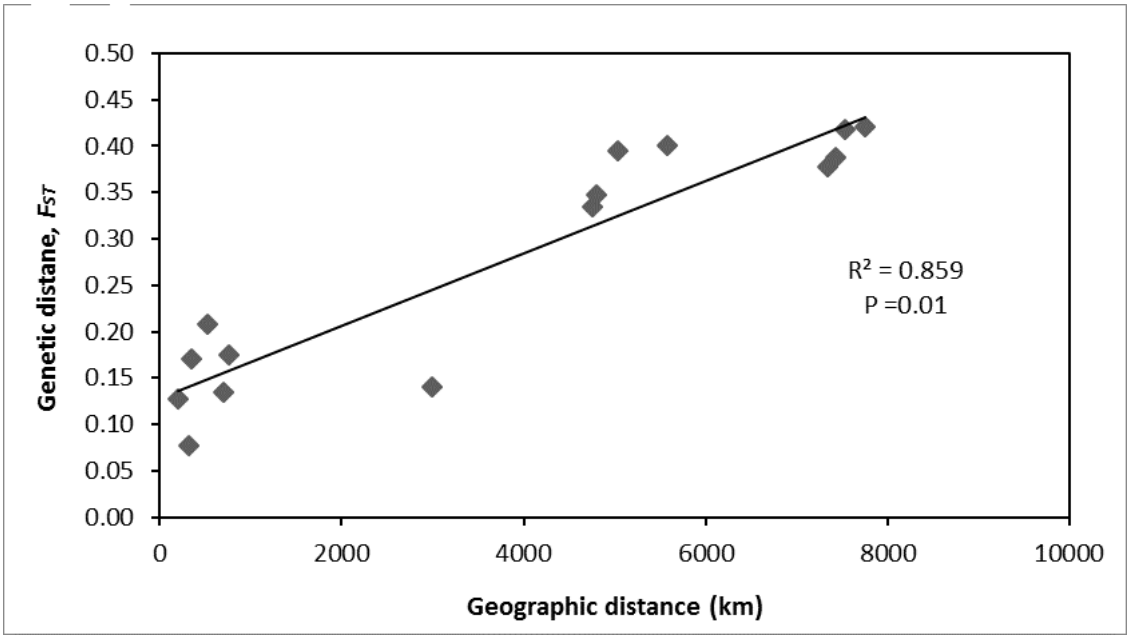


Figure 4.3: Mantels test comparing geographic (km) and genetic (F_{ST}) distance of Nile perch, *Lates niloticus*, populations from six locations in West and East Africa (n = 190).

Bayesian STRUCTURE analysis confirmed strong partitioning of genetic variance among Nile perch populations with two distinct genetic clusters (K = 2) identified. The individuals from West Africa (Senegal and Lake Kainji on Niger River) were assigned to one genetic group (green cluster) and the East African populations (Lakes Albert, Kyoga, Victoria and Turkana) mainly to another (red) (Figure 4.4), although in a few individuals some sharing of genetic ancestry was evident. All individuals from the Senegal River were only assigned to the green genetic cluster (West) and no admixture was observed in this population unlike the population from Lake Kainji on the Niger River that consisted of a few individuals with an admixed green and red genome ancestry.

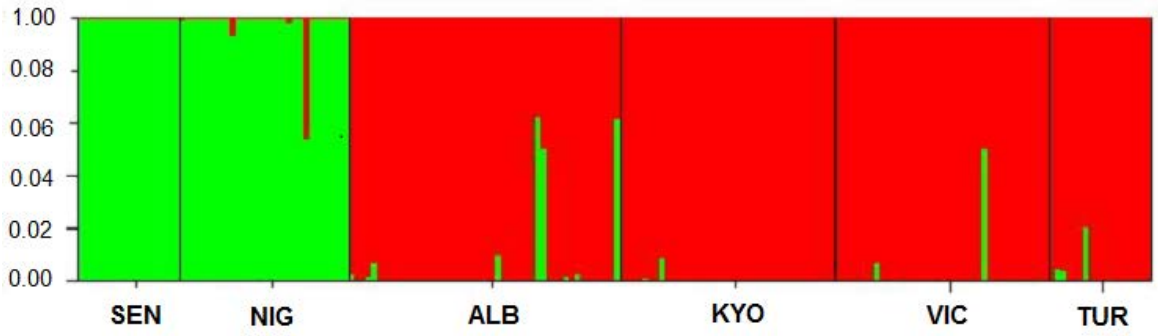


Figure 4.4: STRUCTURE bar plot showing the genetic assignment of West (Green) and East (red) African Nile perch, *Lates niloticus*, populations. The X-axis represents the sampling locations SEN- Senegal, NIG- Lake Kainji on the Niger River, ALB- Lake Albert, KYO- Lake Kyoga, VIC- Lake Victoria and TUR- Lake Turkana.

4.3.3 Gene flow and relative migration among western and eastern Nile perch populations

Results showed distinctive connectivity, reflecting the directional and relative strengths of migration between the six sampled populations of Nile perch. The Nile perch from Senegal River and Lake Kainji clustered into one genetic group and all the other four populations into another major genetic group. The two major genetic groups exhibited in the divMigrate relative gene flow network diagram (Figure 4.5), did mirror the STRUCTURE assignment genetic clusters (Figure 4.4). Here, restricted gene flow was detected between the West and East African populations. Additionally the gene flow between the two West African populations was low compared to the gene flow among the East African populations.

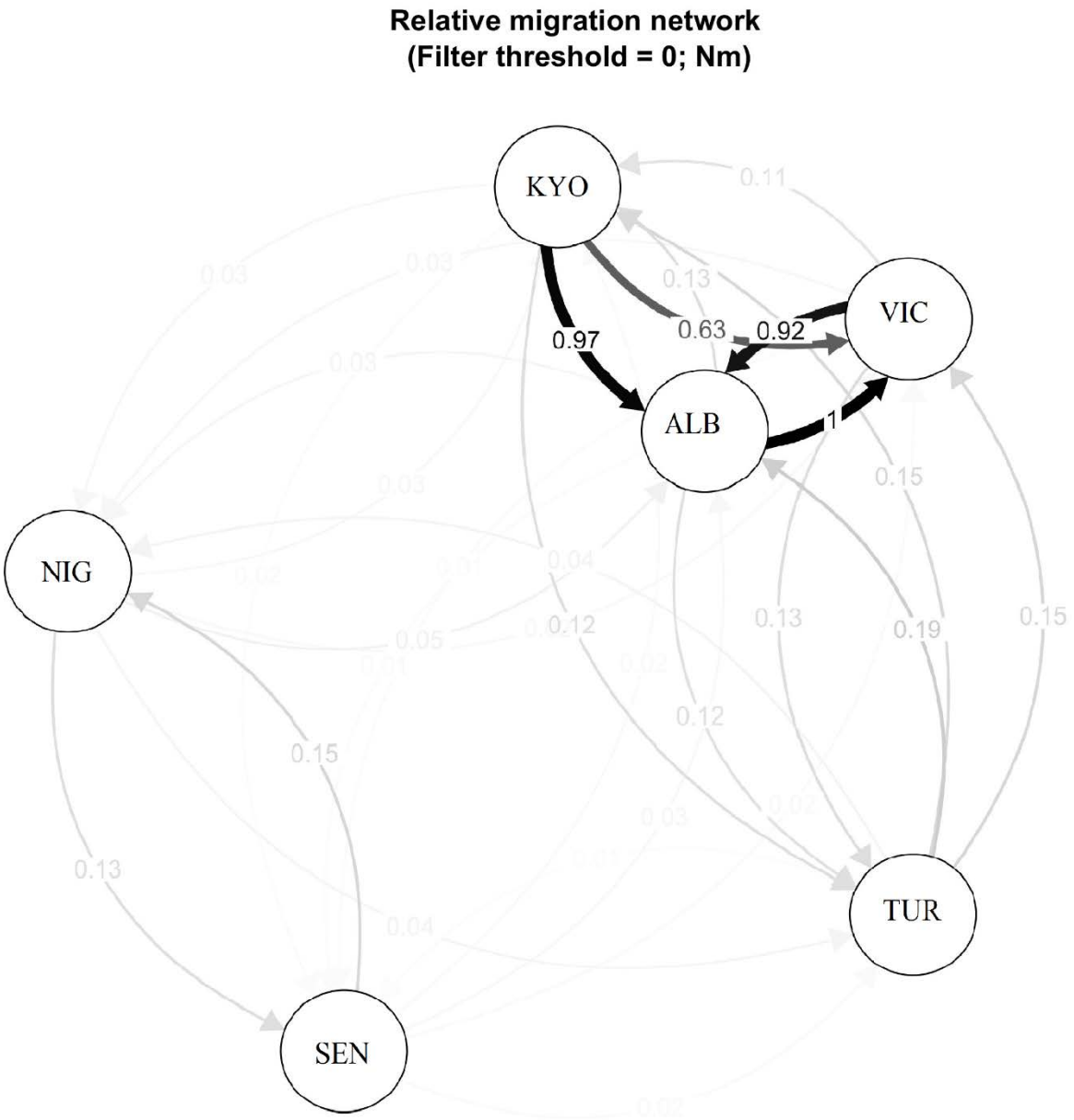


Figure 4.5: Relative migration network in which the circles represent sample locations; - SEN- Senegal, NIG - Nigeria, ALB - Albert, KYO - Kyoga, VIC -Victoria and TUR -Turkana. The arrows contain edge values showing the direction and magnitude of migration levels, with the darker arrows showing stronger migration. The figure reveals stronger gene flow between populations ALB- Albert (putative population) and VIC- Lake Victoria (translocated population originally from mainly Lake Albert) albeit reduced gene flow between Lake Albert and Lake Kyoga (another translocated population also from Lake Albert).

Similarly, there is a genetic sub-group exhibited among three (Lakes; Albert, Kyoga and Victoria) of the four East African populations shown by the high gene flow between these populations. However, restricted gene flow to and fro Lake Turkana population to the rest of

the East African populations; Turkana is appropriately excluded by geographical distance and historical connectivity. In a way, the relative migration network did show genetic separation of the populations relative to distance but more so was a surrogate of historical Nile perch translocations that have characterised the species from its native (Lakes Albert and Turkana) to the new introduced (Lake Kyoga and Victoria) environments.

4.3.4 Population sub-structuring within East African Nile perch populations

Regional STRUCTURE analysis also revealed strong genetic structuring among the EAR lake populations, with three distinct genetic clusters evident (Fig 4.6); the Lake Turkana cluster (green), Lake Kyoga (red), and Lake Victoria (blue). Lake Albert individuals were mainly assigned to the same genetic grouping as Lake Victoria fish, although Nile perch in Lake Albert exhibited evidence of increased genetic admixture with fish from Lakes Kyoga and Turkana.

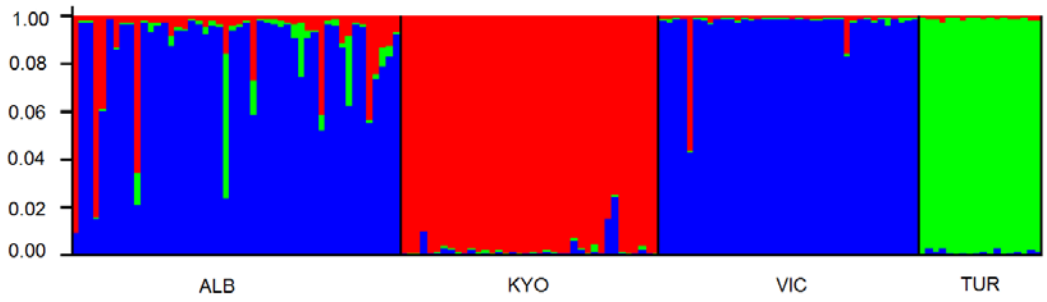


Figure 4.6: STRUCTURE (Pritchard et al., 2000) bar plot for Nile perch stock of East Africa populations; ALB - Lake Albert, KYO - Kyoga, VIC -Victoria and TUR - Turkana. The bar plot shows the proportion of genome ancestry assigned to three genetic clusters represented by blue, red and green colours. Lake Victoria and Albert individuals were assigned to one major cluster (blue) and Kyoga (red), while Turkana to another separate cluster (green).

4.4 Discussion

The Nile perch in Africa was partitioned into two discrete genetic groupings that may require different management strategies; one from West Africa and the other in East Africa. Further structuring within East Africa was revealed, with three discrete populations at a regional scale. The strong genetic structure is a function of restricted levels of gene flow between populations across biogeographical barriers, and particularly for the Nile perch in East Africa, the structuring may also partially be a product of historical anthropogenic introductions. The categorical evidence of decreased genetic diversity exemplified within the translocated populations of Lake Victoria and Kyoga, suggests introgression and loss of genetic diversity resulting from anthropogenic activities. The two translocated Nile perch populations of Lake Kyoga and Lake Victoria are genetically distinct, and ought to be treated separately for sustainable management of the now dwindling fishery in the two new habitats. However, for an aquaculture breeding program within the East African region, all three distinct populations should be included at the start of the breeding program so as to maximise the advantages of diverse genetic variability given that the F_{ST} values within regions were shallow, albeit significant (P = 0.000).

4.4.1 Broad scale genetic differentiation and structure

The Nile perch in Africa's freshwater rivers and lakes was highly differentiated with significant (P < 0.000) global F_{ST} of 0.26 and strong genetic structuring which is comparable to other freshwater species with similar dispersal and habitat usage (Ward, 2006). Because freshwater habitats are typically discrete (unless connected by river systems), populations occupying such habitats tend to display strong genetic structuring due to isolation and

restricted gene flow. Nile perch in Africa was confirmed to be partitioned into two discrete genetic groupings; one from West Africa and the other one in East Africa. Separate genetic groupings among populations can only occur following restricted gene flow over time (Rutz et al., 2012). Therefore, the existence of the two major Nile perch groups confirmed in the current study using high resolution nuclear markers clearly indicates the separation of these genetic lineages over a long period of time.

The results from the current study showing the two genetic groupings were consistent with the limited genetic work based on a single study carried out in Nile perch from the broader Africa region using a single mitochondrial marker (d-loop region). That study identified two genetic groupings of the Nile perch in Africa, with no shared haplotypes between them (Mwanja et al., 2013). The detection of strong genetic differentiation (at both the mtDNA and nuclear DNA markers) between the West and East African populations of Nile perch indicates that the Nile perch, like other species (e.g. African buffalo, *Syncerus caffer,* and Nile tilapia *Oreochromis niloticus)* in Africa, exhibits a genetic break between the populations of Western and Eastern Africa (Smitz et al., 2013, Agnese et al., 1997).

The significant isolation by distance revealed by the Mantel's test in the current investigation further suggests limited dispersal of the species between the two regions of West and East Africa. The result is consistent with findings in other aquatic species, where patterns of isolation by distance have been found to arise from limited dispersal that decreases with increased distance (Bohonak, 2002, Hardy and Vekemans, 1999, Jensen et al., 2005, Meirmans, 2012, Slatkin, 1993, Sokal and Wartenberg, 1983, Wright, 1943). Strong genetic differentiation is found to be common over a larger geographical area as opposed to minimal or no structuring at all between populations in close geographic proximity, although recent

findings reveal a discrepancy between dispersal and genetic structuring (Arnoux et al., 2014, Meirmans, 2012). Thus, whereas the IBD seemed to be playing a role in the separation of the two major genetic groupings of West and East Africa, further analysis at a regional scale revealed that the Nile perch in East Africa does not conform to IBD.

4.4.2 Genetic differentiation and structure in East Africa

The genetic structure in East Africa is complex and the current study unraveled up to three differentiated genetic clusters which appeared to be grouped according to lakes. Except for Lakes Victoria and Albert that were relatively undifferentiated from each other, there was comparatively deeper differentiation observed amongst all the other lakes in the East African region (all with significant F_{ST}, P<0.000). Strong genetic structuring among the East African populations was evident where Lakes Albert and Victoria clustered together (Group 1); and Lake Kyoga with Lake Turkana each were assigned to their own genetic cluster (Group 2 and 3 respectively). Results of the current study showing three genetic groupings differs from Mwanja et al. (2014b), who identified only two genetic groupings within the East African lakes. How the three genetic groupings have evolved in the face of biogeographic events and geological structures is still open to interpretation, but these three genetically distinct Nile perch populations identified for the first time in the current study may be treated as different management units (based on their high level of differentiation and reduced connectivity) for the sustainable exploitation of the species.

It is hypothesized that the disruption of natural gene flow following human introductions of Nile perch into new habitats and the complex geological history in the East African region are some of the key factors driving the genetic structure and evolution of this and other

species in the region. Previous studies in other freshwater species in the region have demonstrated that the complex geological history in East Africa is a major driver of diversification at both species and population levels (Sturmbauer et al., 2001, Verheyen et al., 2003, Danley et al., 2012). However, the structuring within the East African Nile perch populations seems to be in agreement to some degree with reports of the anthropogenic introductions of the Nile perch into Lake Victoria and Kyoga, with parent introduced Nile perch stocks derived from Lake Albert and a few individuals from Lake Turkana (Pringle, 2005).

4.4.3 Genetic structure following fish introductions in the East African freshwater lakes

In the current study, strong genetic differentiation was detected among native and translocated Nile perch. As mentioned earlier, the major translocations in the region involved the stocking of Nile perch from Lakes Albert and Turkana into Lakes Victoria and Kyoga (Pringle, 2005, Ogutu-Ohwayo, 1990b, Ogutu-Ohwayo, 1990a). Although a portion of Nile perch individuals introduced into Lake Victoria were putatively acquired from Lake Turkana, results from the current investigation showed that the Nile perch in Lake Victoria exhibited a very limited genetic signature from Lake Turkana, but had a strong signal from Lake Albert. The very weak signal from Lake Turkana seems to be consistent with the historical stocking records, where only eight individuals were introduced in Lake Victoria at Kisumu Port (Pringle 2005).

In addition to the Nile perch individuals from Lake Victoria exhibiting a strong genetic signal from the Lake Albert population, there was also evidence of strong directional gene flow

obtained for the species between these two lakes (Figures; 4.5 & 4.6). This could be due to the fact that; 1) multiple introductions of Nile perch from Lake Albert were made into Lake Victoria (Pringle, 2005) and 2) the niches in Lake Victoria are highly variable, and some of these habitats could still mimic to a certain degree those available in the source habitat of Lake Albert. Lake Victoria has varied habitats that harbour Nile perch, while displacing other species [habitats include shallow and deep water characterised by a gradient of hypoxic, chronic hypoxic and anoxic conditions (Chapman et al., 2007, Chapman et al., 1995, Olowo and Chapman, 1999)]. The genetic convergence of Nile perch from Lake Victoria and Albert is slightly contradictory to an earlier report that was based on morphology, indicating that the form of Nile perch in Lake Victoria differed from the form in the putative parental origins of Lake Albert and Turkana (Harrison, 1991). The results in the current study clearly show that the Nile perch in Lake Victoria is still genetically similar to the Nile perch in Lake Albert, with shallow but significant F_{ST} (P = 0.003), although these populations have become divergent from Lake Turkana individuals ((F_{ST} = 0.18, P = 0.003) Table 4.2, Figure 4.5) which lies in the eastern side of the EAR (Harrison, 1991).

Nile perch from Lake Kyoga was differentiated from that of Lake Albert, a finding which is quite interesting since the foundation stocks into Lake Kyoga were derived from Lake Albert (Figure 4.6). This is possibly due to low founder numbers that may have survived following stocking and possibly most genotypes to the red genetic ancestry (Figure 4.6) may have survived and consequently propagated. Another possibility could be the uncertainties usually associated with actual facts on fish introductions, in which case the historical record of introduction from Lake Albert may be questionable, and it is likely that the stock could have been sourced from elsewhere, probably in the upper Nile. Unfortunately in the current

investigation, there were no samples acquired from the Nile river system to ascertain this possibility, and as such we propose that future genetic work should involve samples from the Nile River (Pringle, 2005).

4.4.4 Genetic diversity

Globally, introductions of non-native fish species into freshwater ecosystems are a major concern. This is especially so in the African Great Lakes region, where fish translocations have supposedly resulted in disruptions to aquatic biodiversity and overall community structure. Moreover, the translocated populations may go through bottlenecks that are revealed through reduced genetic diversity. Low genetic diversity has been found to impact negatively on the viability of populations posing concerns for conservation as populations may fail to adapt and persist in the event of harsh ecological and environmental pressures (Milot et al., 2007, Christie et al., 2012).

The current study showed evidence of reduced genetic diversity in the translocated Nile perch populations of Lakes Victoria and Kyoga, compared to the native populations from which the founder stocks were derived. The higher genetic diversity in the native Nile perch populations of Lakes Albert and Turkana, indicates that these populations have a higher potential to adapt and persist over time in their native habitats, than would be the introduced populations (Greenbaum et al., 2014, Caballero and García-Dorado, 2013, Christie et al., 2012). The findings in the current investigation are contrary to earlier findings in the same species where the native populations showed lower genetic diversity (allelic diversity) than the translocated populations (Mwanja et al., 2012). However, data in the present study should be viewed as being more robust, as the genetic differences were statistically significant and

population sample sizes were taken into account with regards to diversity indices. Furthermore the patterns observed in the current study for the allelic diversity are consistent with the general hypothesis of loss of genetic diversity due to founder effects within translocated populations (e.g. in the speckled dace cyprinid fish, *Rhinichthys osculus* which was introduced into Van Duzen River in California (Kinziger et al., 2011).

The situation of reduced genetic diversity in the introduced Nile perch is exacerbated by pressures that the Nile perch could be experiencing in the new environments (of Lakes Kyoga and Victoria) in comparison to their putative original populations (from Lakes; Albert and Turkana). These include; heavy fishing pressure coupled with climate and ecological pressures which may in part be responsible for the current declines of Nile perch in Lakes; Kyoga and Victoria. Results from the present study are consistent with reports that have shown declining numbers of Nile perch in Lakes Kyoga and Victoria despite the increase in fishing effort of the Nile perch amidst climate and ecological pressures (Basiita et al., 2011, Basiita et al., 2015, Mkumbo and Marshall, 2015, Mwanja et al., 2012, Ogutu-Ohwayo, 1988, Taabu-Munyaho et al., 2014, Kudhongania et al., 1992). Nile perch declines were first detected in Lake Kyoga as early as the 1980's, and predictions of such declines speculated then for Lake Victoria have now been realised (Ogutu-Ohwayo, 1988).

Although reduced genetic diversity was detected among the translocated populations at the allelic level, the genetic bottleneck was not severe enough to be detected by the bottleneck analysis (Luikart et al., 1998, Luikart and Cornuet, 1998). This non-detection was most probably due to the multiple introductions that generally could have counter balanced the founder effects that were not severe enough to be detected by the heterozygosity method employed under bottleneck analysis.

4.4.5 Management implications

Knowledge of genetic structure and relationships revealed in the present study among Nile perch populations will be useful for the effective management and conservation of the species in both its native and introduced environments. Separate management strategies of Nile perch populations from West and East Africa (for both aquaculture development and fisheries restocking) should be considered, following the genetic divergence and differentiation that has occurred over evolutionary time between these two groups. Following restricted gene flow, the two genetic groupings have evolved in isolation, and consequently may have accrued adaptive differences, and any translocations of Nile perch should be carefully considered across the regions of West and East Africa.

Programs aimed at restocking these lakes and rivers should seek founder stocks independently within each of the two regions of West and East Africa. As discussed previously, there was complex structuring within East African Nile perch where three distinct genetic stocks were identified, and these seemed to be linked to lakes. Thus, within East Africa, the Nile perch from Lake Albert and Lake Victoria should be treated as a single group, Lake Kyoga as another and Turkana separately. Movement of the Nile perch across lake basins should be discouraged in order to prevent possible genetic introgression.

Furthermore, from an aquaculture perspective, the recognition of the three genetic groupings identified within East Africa, provides a large genetic base upon which genetic diversity can be maximized through potentially mixing the three distinct genetic populations at the start of a Nile perch breeding program for the region. However, it is important to note that even though the populations are distinct, F_{ST} is relatively shallow and could potentially be mixed

together without raising concerns of out breeding that may include a reduction in fitness (Ward, 2006). The three genetic groups may hold a broad range of standing genetic variation, which aquaculture efforts (through breeding programmes) can capitalise on to develop an artificial base population. A similar approach was utilized in the development of the GIFT strain of Tilapia, where individuals from various sources of up to eight strains were mixed together to provide a foundation for a genetic improvement program (Eknath et al., 1998, Eknath et al., 1993).

4.5 Conclusion

Nile perch in Africa is partitioned into two discrete genetic groupings; one comprising populations from West Africa and the other populations in East Africa. Further structuring within East Africa was revealed, with three discrete populations requiring different management strategies at a regional scale. Although, strong genetic structure is a function of restricted levels of gene flow between populations across biogeographical barriers, in the current study, results indicated that it may also partially be a product of historical anthropogenic introductions. The categorical evidence of decreased genetic diversity exemplified the persistence of founder effects within the translocated populations of Lakes Victoria and Kyoga. The two translocated Nile perch populations of Lakes Kyoga and Victoria were distinct, and ought to be treated separately for sustainable management of the now dwindling fishery in the two new habitats. The use of more robust markers in the current study clearly resolved the genetic structure of Nile perch in East Africa, and the three populations identified will be useful in forming an artificial and diverse base population for selective breeding of the Nile perch. Therefore, findings and information provided by the current study will be critical and useful in achieving successful and comprehensive

aquaculture establishments, as well as conservation and management of the wild fisheries for the freshwater Nile perch.

CHAPTER 5

Historical and contemporary genetic structuring revealed in the African freshwater Bagrid catfish, *Bagrus bayad.*

Abstract

Bagrids are important as food fish in both Asia and Africa. *Bagrus bayad*, a freshwater Bagrid widely distributed in Africa, remains poorly researched and there is no information on the genetic population structure and differentiation of the species. The biogeographic relationships of *B. bayad* among drainages of Lake Albert in East Africa, Lake Kainji on the Niger River and the Senegal River in West Africa were explored in the current study. Present and historical connectivity of *B. bayad* from these three drainages were investigated using 223 bp of the mitochondrial DNA d-loop region and nine polymorphic microsatellites in order to define management and conservation units for the species. Mitochondrial DNA via haplotype networks and phylogenetic tree analysis revealed two haplotypic clades, whereby the West African *B. bayad* populations were divergent from the East African population. Overall, haplotype diversity, $H_d \pm SD = 0.87 \pm 0.028$, and nucleotide diversity, $\pi = 0.029$, was high across all sampled *B. bayad* individuals. Comparably higher genetic diversity was observed in the *B. bayad* individuals from Lake Albert and Lake Kainji, whilst the lowest diversity was from those individuals sampled from the Senegal population ($H_d \pm SD = 0.27 \pm 0.11$). Bayesian analyses using nine microsatellites furthermore, identified three genetic clusters (K =.3), grouping *B .bayad* by drainage. Thus, whereas the mitochondrial results highlight the existence of two evolutionary significant units, the microsatellite data resolved the stock structure of the species into three contemporary genetic groupings. Therefore, based on these findings, three management units are recommended to take into account the three genetic *B. bayad* populations that appear to correspond to the drainages under study.

Separate management units for identified genetic groupings will be important in drawing strategies specific to each drainage, as varying levels of genetic diversity were also identified.

5.1 Introduction

Africa is endowed with a rich biodiversity. This biodiversity has evolved amidst a complex and dynamic geological system, coupled with a changing climate. Unfortunately, due to an increasing number of threats, most reports highlight a stagnation, or decline in biodiversity in various ecosystems (Butchart et al., 2010, Chapman et al., 2007, Helfman, 2007, Lévêque, 1997, Luck, 2007, Sala et al., 2000). Therefore understanding the presence, or absence of distinct genetic structure and differentiation is key in the management and conservation of African biodiversity.

Aquatic fauna in both freshwater and marine environments have experienced an alarming rate of diversity decline. The drivers of biodiversity loss are common in both environments and include losses due to invasive or exotic species introductions, over exploitation, pollution, changing climate and the rapid degradation of habitats (Kokko and López-Sepulcre, 2006, Ogutu-Ohwayo, 1990b, Ogutu-Ohwayo, 1990a, Rodríguez et al., 2005). Unfortunately, in Africa, from a genetics perspective, very few freshwater species have received attention and non-model species (eg. non-cichlids) have remained data deficient and neglected, largely due to lack of required resources such as genomic resources to start research *de novo*.

Bagrus bayad is a demersal freshwater fish species that inhabits lakes, swamps and rivers and that is highly exploited in the North, Central, West and Eastern countries of Africa (Alhassan and Ansu-Darko, 2011, Anja et al., 2009, Aruho et al., 2013, Diogo et al., 1999, Goudswaard and Witte, 1997, Khallaf and Authman, 1992, Hart and Pitcher, 2012). The distribution of *B. bayad* is relatively extensive in Africa's freshwater lakes (such as Lakes Chad, Albert and Turkana) and rivers (including the Nile, Niger, Senegal and Congo Rivers)

(Greenwood, 1966, Hart and Pitcher, 2012, Khallaf and Authman, 1992). *Bagrus bayad,* like its congener *B. docmak*, is a potamadromous siluroid catfish, migrating upstream mainly for spawning purposes (Riede, 2004). The two species co-exist in some parts of their range, including the Nile, Senegal and Niger Rivers, as well as in the rift valley lakes of Lake Albert and Turkana (Mohamed, 2016). Although these bagrids co-exist, they are easily distinguishable morphologically; *B. bayad* differs from *B. docmak* in that the head is narrower and the lobes of the caudal fin are prolonged into two long filaments, whereas in the later only the upper lobe culminates into a long filament (Sandon, 1950, Greenwood, 1966, Mohamed, 2016).

Despite its commercial importance and widespread distribution, scant information exists on aspects of distribution, morphology, reproductive biology and feeding habits of *B. bayad* (El-Drawany and Elnagar, 2015, Mohamed, 2016, Tsadu et al., 2014). The species remains poorly researched and there is no information found on the population genetic structure and differentiation of *B. bayad* across its natural species' range. Therefore the aim of this chapter was to contribute to the gap in knowledge on the population genetics of the species, an aspect that is important in management, conservation and aquaculture development of *B. bayad* in Africa.

5.2 Methods

5.2.1 Sampling Locations

Bagrus bayad samples were collected from three major water bodies in Africa; Lake Albert (28 samples) in East Africa, Senegal River (42) and Lake Kainji (28) on the Niger River in West Africa. The three sample locations are among the five sample locations in Chapter 4 (Figure 4.1), and a brief summarry of just the three locations relevant to this study is given below;

Senegal River basin

The Senegal River basin drains into the Atlantic Ocean and covers 1.6 % (483,181 km^2) of the West African continent spreading over four countries (Senegal, Mauritania, Mali and Guinea). The basin covers over 36.5% of the total area of Senegal, with the river forming the border between Senegal and Mauritania (FAO document accessed on 26-03-2016 at http://www.fao.org/docrep/w4347e/w4347e0h.htm). Below Dagana, at Richard-Toll where the *B. bayad* samples for the current study were captured, the Senegal River enters its delta with the Diama dam which was constructed in 1985 and currently acts as a low water storage dam and salt-water barrier near the mouth of the Senegal River (Hargreaves et al., 1985).

Lake Kainji on Niger River

Lake Kainji, located in north-western Nigeria is the largest man-made lake in Western Africa and was formed in 1968 by damming part of the Niger River. The lake covers an area of up to 1,300 km^2. Lake Kainji (from which the samples were obtained in the current study) is

important and extensively used for fishing and irrigation in the region (Balogun, 1986, Ita, 1978).

Lake Albert

Lake Albert is located at the tip of the eastern arm of the East African Rift (EAR). Lake Albert drains directly into the White Nile, to which it is believed to be the main regulating water source for the Nile and responsible for making the Nile a permanent river, as opposed to a seasonal river (Beuning et al., 1997, Holden, 1967, McGlynn et al., 2013, Pickford et al., 1993). The lake is connected to Lake Kyoga to the east by the Victoria Nile and connected to Lake Edward in the south by the Semliki River.

5.2.2 Sample preservation and DNA extraction

From each individual fish, a fin clip was cut from the caudal fin and preserved in 20% dimethyl sulfoxide (DMSO) saturated with sodium chloride salt (Dawson et al., 1998, Amos and Hoelzel, 1991), or in 70% ethanol. All fin clips were transferred in fresh DMSO salt prior to being shipped to the Molecular Ecology and Evolution Laboratory (MEEL) in Townsville, Australia, where they were stored at -20 $^{\circ}$C prior to DNA extraction.

Total genomic DNA was extracted using the Bioline Isolate II Genomic DNA kit (Bioline) following the manufacturer's protocol. The protocol involved a pre-lysis stage using 180 µl Lysis buffer with 25 µl of protease K incubated at 56 ^{o}C for 1 hr. A second lysis buffer, G-3 (200 µl) was added and further incubated at 70 ^{o}C for 10 min. In order to alter the buffer conditions, 210 µl of 100 % EtOH were then added prior to two (GW1 and GW2) buffer washes. DNA was then eluted using a warm (70 ^{o}C) elution buffer and later the eluted DNA

was stored at -20 ºC prior to downstream PCR. DNA quality and quantity was assessed using a ND-1000 Spectrophotometer (Nano-Drop® Technologies).

5.2.3 Polymerase chain reaction (PCR)

DNA from 84 individuals was amplified via PCR and sequenced at the mitochondrial d-loop region using Sanger sequencing. As described in Chapter 3 of this thesis, primers specific to *Bagrus spp*., were designed using a free online software, Primer3 (Rozen and Skaletsky, 2000). A forward (BagDF2) - TTGAGGGTTGGTGGTTTCTT and reverse (BagDR2) - AAACTATTTTCTGTAAATGCATAAT) primer pair were tested for specificity in *B. bayad* via PCR; 25 µl PCR reactions volume were used; 2.5 µl 10X buffer, $MgCl_2$ 1.5 mM, dNTPs 0.2 mM, *Bagrus* control region primers; BagDF2 and BagDR2 (at a final concentration of 0.2 µM for each primer) and 1 µl DNA (5 ng/ul); PCR cycling conditions comprised an initial denaturation of 94 °C for 5 min, then 30 cycles of 94 °C for 30 s (denaturation)/50 °C for 30 s (annealing)/72 °C for 30 s, followed by a final extension step at of 72 °C for 5 min. Visualization of PCR products (approximately 400-450 bp length) was achieved via electrophoresis using a 1.5 % agarose gel stained with gel green. PCR products were then cleaned with Sephadex G-50 columns (GE Healthcare UK) prior to approximately 5 ng of each sample being sent for Sanger sequencing at the Georgia Genomic Facility, USA. Both forward and reverse primer directions were sequenced for each individual using the designed *Bagrus* specific primers. Approximately a 400 bp sequence was returned for each individual and all sequences were visually edited and aligned in Geneious vs 8.1, prior to being exported to MEGA (version 7) for downstream analyses (Drummond et al., 2009, Kumar et al., 2016). A model test was performed in MEGA (version 7), and the best model (T92+G, where G = 0.46) that suited the *B. bayad* sequence data set, was selected for all analyses.

DNA from 98 individuals were also genotyped at 13 polymorphic microsatellite loci that were initially designed for *B. docmak* and tested for cross amplification in *B. bayad* (details of primer development and cross amplification are given in Chapter 2). For all markers, the forward primers were fluorescently labeled using the 5- dye system (6-FAM, VIC, NED, PET and LIZ GS-500 size standard) and reverse primers were pigtailed (Brownstein et al., 1996) to ensure consistent amplification and minimize stuttering. A total volume of 10 µl PCR reactions were run on a Biorad C1000 thermocycler under the following conditions: an initial denaturation at 95 °C for 5 min, 6 cycles of 95 °C for 30 s (denaturation)/59 °C for 90 s (annealing)/72 °C for 30 s (extension), 10 cycles at reduced annealing temperatures of 57 °C, 55 °C and 53 °C, prior to a final extension at 60 °C for 30 min.

Visualisation of PCR products was performed on an ABI-3730 instrument (Applied Biosystems) at the Georgia Genomics Facility, USA. Alleles were scored using Genemarker 2.4 (Softgenetics) and checked for genotyping errors, stuttering large allelic dropouts and presence of null alleles in Microchecker 2.2.3 (Van Oosterhout et al., 2004). Of the 13 loci, nine polymorphic markers were selected for downstream use in analyses based on minimal occurrence of null alleles and missing data at these loci. The nine loci selected were *BD04, Bd18, Bd01, Bd02, Bd12, Bd09, Bd05, Bd08 and BD14*. Individuals with missing data of up to 40% were also removed. Ninety-five catfish samples were retained for downstream analysis after quality control procedures were conducted (n = 28 for Lake Albert, n = 25 for Lake Kainji and n = 42 for the Senegal River).

5.2.5 Genetic differentiation, phylogeography and population structure

Genetic structure was explored using an analysis of molecular variance (AMOVA) for both mitochondrial DNA and microsatellite data, while a Bayesian analysis was used to infer

probable number of genetic clusters present for *B. bayad* for microsatellite DNA data only. Using a hierarchical AMOVA analysis, the level of genetic variance partitioned among populations, individuals within populations and that within individuals was determined. Global F_{ST} and ϕ_{ST} testing was also performed for both microsatellite and mitochondrial DNAdata sets respectively in Arlequin 3.5 (Excoffier and Lischer, 2010) to determine whether significant genetic differentiation existed between populations. Following the significant overall F_{ST} and ϕ_{ST} values obtained (at $P < 0.001$), pairwise genetic differentiation tests were performed on all the three populations. Significance was estimated based on 16,000 permutations.

Furthermore, using mitochondrial sequence data organised in DnaSP (Librado and Rozas, 2009) and exported to Network 4.611 (Flexus Technology, 2012), the phylogeography and geographical distribution of haplotypes was visualized. Calculation and drawing of a minimum spanning network based on haplotype distribution in sampled individuals from the three populations was completed in Network 4.611 (Flexus Technology, 2012). A phylogeny reconstruction analysis based on a maximum likelihood (ML) tree with 1000 bootstraps was performed in MEGA, version 7 (Kumar et al., 2016). The default initial tree for the ML was drawn by applying the neighbor joining (NJ/BioNJ) algorithms to a matrix of pairwise distances estimated using a maximum Composite likelihood (MCL) approach and selecting the topology with superior log likelihood value. The phylogenetic tree was drawn to scale, with branch lengths measured based on the number of substitutions.

Bayesian assignment tests for microsatellite data was implemented in STRUCTURE 2.3.3 (Pritchard *et al.,* 2000) in order to determine the number of inferred genetic clusters and to test the proportion of genetic admixture among the three sampled *B. bayad* populations. The

implementation of individual-based clustering in this program creates clusters of individuals within which the Hardy-Weinberg equilibrium is maximised and linkage disequilibrium minimised, thus assigning individuals into gene pools, or breeding groups (Ball et al., 2011). In order to evaluate the number of genetic clusters, analyses involved testing up to 10 possible clusters (from K = 1 to K = 10), followed by 10 iterations using the admixture model with the assumption of correlated allele frequencies and no location prior information provided. Similar tests of the admixture model were carried out, but this time assuming that the allele frequencies are independent; except for this modification all other input parameters in STRUCTURE analyses were left the same (i.e. a burn-in period of 100,000 iterations and 100,000 MCMC). The most likely number of distinct genetic clusters (K) was determined using STRUCTURE HARVESTER (Earl and VonHoldt, 2012), which uses the methods described by Evanno et al. (2005).

5.3 Results

5.3.1 Mitochondrial DNA

A 400 bp mitochondrial DNA d-loop region amplified from 84 *B. bayad* individuals was aligned and trimmed in Geneious 8.1 prior to being exported to MEGA (version 7) where two very short sequences were further removed. A 223 bp sequence without ambiguities and gaps, was used in the remaining 82 sequences for subsequent analysis using the best model (T92+G, where G = 0.46) for the data as determined by MEGA7 (Appendix 5.1).

5.3.1.1 Genetic diversity and historic demography of *Bagrus bayad*

Mitochondrial DNA d-loop variation among the 82 *B. bayad* individuals from the three locations (Lake Albert, Lake Kainji and Senegal River) revealed 22 haplotypes, with a total of 31 mutations. The haplotypes were unequally distributed among the sampled locations, Lake Albert had the highest number of haplotypes (H_n = 11) and the lowest (H_n = 4) was found in the Senegal River population (Table 5.1). Overall, haplotype diversity, $H_d \pm SD$ = 0.87 ± 0.028 was high across all sampled *B. bayad* individuals. Comparably higher genetic diversity (based on indices such as H_n, H_d and π shown in Table 5.1) was observed in the *B. bayad* individuals from Lake Albert and Lake Kainji in Nigeria, whilst the lowest diversity in individuals was from those sampled from the Senegal River (Table 5.1).

Table 5.1: Mitochondrial DNA diversity indices and neutrality test results for 82 mtDNA d-loop sequences of *Bagrus bayad* from three freshwater systems in Africa

Location	*N*	H_n	$\pi \pm SD$	$H_d \pm SD$	*D (P value)*	*Fu (P value)*	*K*
ALB	32	11	0.020 ± 0.01	0.85 ± 0.04	-0.74 (0.234)	0.16 (0.566)	3.99
NIG	22	8	0.010 ± 0.00	0.87 ± 0.04	-0.14 (0.486)	-2.64 (0.057)	2.1
SEN	28	4	0.001 ± 0.00	0.27 ± 0.11	-1.53 (0.024)*	-2.61 (0.002)**	0.28

Locations ALB, Lake Albert, NIG, Lake Kainji on the Niger River and SEN, Senegal River. *N*, number of sampled individuals, *H*, number of haplotypes, π, Nucleotide diversity, H_d, haplotype diversity. *D*, Tajiman's *D* statistic, Fu, is *Fu*'s F statistic and *k*, average number of nucleotide differences. The diversity indices were all calculated in DnaSP except for *D* and *Fu*'s values that were implemented in Arlequin assuming the T92 +G model, where G = 0.46.

Neutrality tests performed on the *B. bayad* individuals from the three sampled drainages were indicative of signatures of possible population expansion, as indicated by Tajima's *D* -test statistics. However analysis of *Fu's F* statistic further confirmed the West African (Lake Kainji and the Senegal River) populations as having undergone a possible historical bottleneck given the negative *Fu*'s values obtained for these two populations (Table 5.1, Figure 5.1). Of the two West African populations, stronger support for a demographic expansion model was observed in the Senegal population that had a negative and statistically significant Tajima's *D* and *Fu'*s *F* value (*D* = -1.53 P= 0.024 and *Fu* = -2.61, P = 0.002). The results for the Senegal River are corroborative with the lower genetic diversity obtained from almost all diversity measures from this population including the number of haplotypes, haplotype diversity and nucleotide diversity. Conversely, Lake Albert was depicted as having a more stable demography as shown by a bimodal and slightly ragged graph (Figure 5.1), in addition to the evidence shown by the positive and non-significant *Fu's F* statistic obtained following the neutrality tests (Table 5.1).

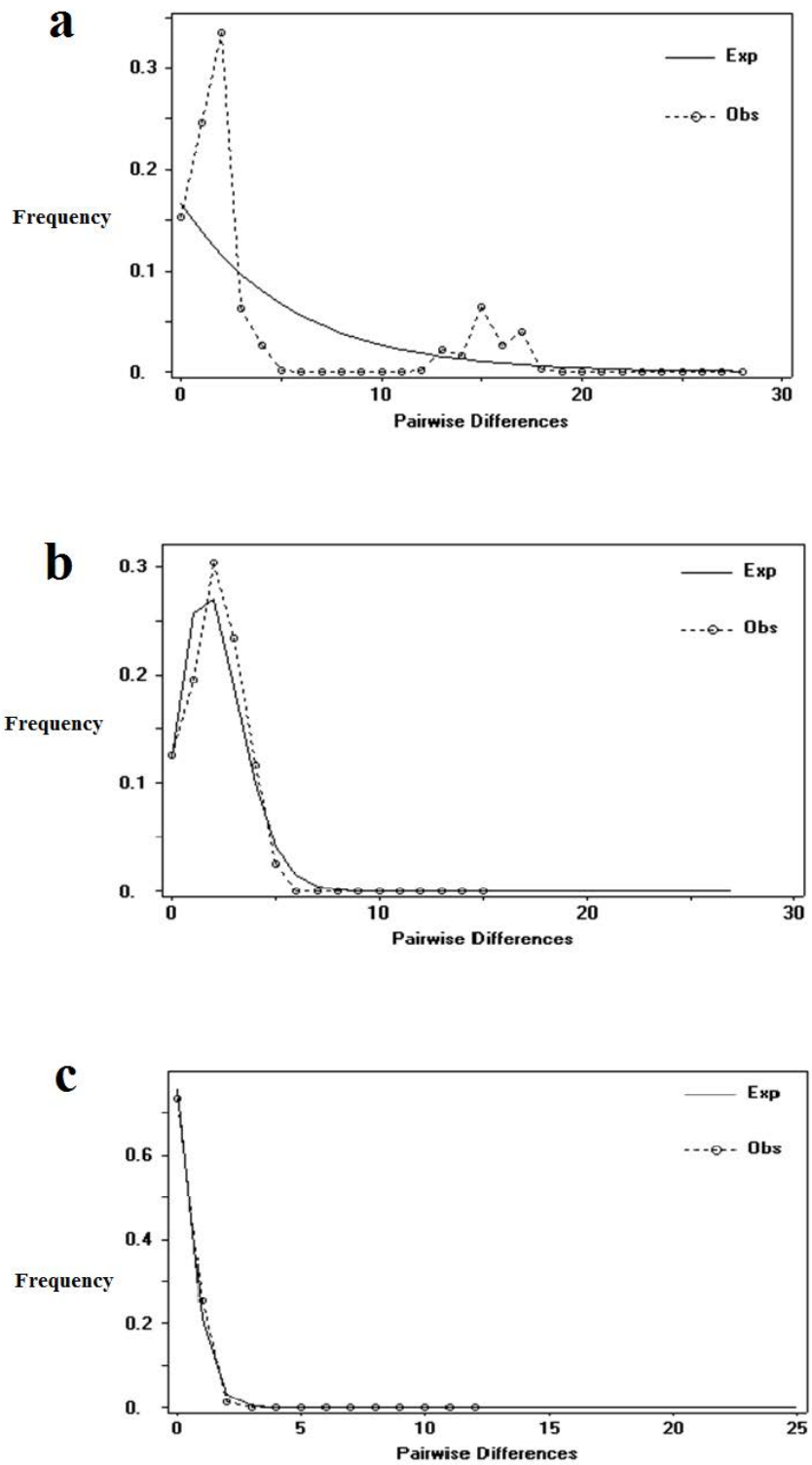


Figure 5.1: Mitochondrial DNA mismatch distribution for a) Lake Albert, b) Lake Kainji and c) Senegal River populations based on sudden demographic expansion. Exp, expected and Obs, observed values shown by the solid and dashed line on the graph respectively.

5.3.1.2 Genetic structure and differentiation within the Mitochondrial DNA d-loop region

Hierarchical AMOVA analysis in Arlequin revealed strong genetic differentiation among groups (East and West) of *B. bayad* individuals. Partitioning of mtDNA genetic variance indicated that the highest source of variance was that present among Eastern and Western African bayad catfish (Table 5.2), although these differences were not significant at $P < 0.05$. On the contrary, the variances due to among locations within regions, and that due to within sample locations were relatively low, but highly significant (percent variation = 10%, ϕ_{SC} =0.35 and 18.4%, ϕ_{ST} = 0.82, respectively; P = 0.000). Downstream pairwise comparisons equally indicated high levels of genetic differentiation (as evidenced by high ϕ_{ST}) between populations of Lake Albert, Lake Kianji and the Senegal River (Table 5.3).

Table 5.2: Hierarchical Analysis of Molecular Variance (AMOVA) for *Bagrus bayad* mitochondrial d-loop region data

Source variation	Component of variance	Percentage variation	Fixation index	P- value
Grouped data East And West African drainage systems				
Among regions	Va	71.55	ϕ_{CT} = 0.72	0.333
Among locations within regions	Vb	10.02	ϕ_{SC} = 0.35	0.000
Within sample locations	Vc	18.43	ϕ_{ST} = 0.82	0.000
Ungrouped data				
Among sample locations	Va	76.99	ϕ_{ST} = 0.76	0.000
Within sample locations	Vb	23.01	n/a	n/a

Table 5.3: Pairwise ϕ_{ST} comparisons between *Bagrus bayad* populations of Lake Albert, Lake Kainji and the Senegal River based on 82 mitochondrial DNA d-loop sequences. Below diagonal are F_{ST} values and above their significance value.

Population	ALB	NIG	SEN
ALB	-	0.000	0.000
NIG	0.56	-	0.000
SEN	0.76	0.81	-

ALB, Lake Albert in East Africa, NIG, Lake Kainji on the Niger River and SEN, Senegal River in West Africa (estimates were performed in Arlequin 3.5 based on 16,000 permutations).

The mtDNA haplotype network confirmed the existence of historical structuring and differentiation among the populations sampled (Figure 5.2). The network highlighted two main haplotypic clades (Albert clade and West African clade), separated by up to eight mutations including an intermediary additional non-sampled haplotype (Figure 5.2). Further evidence from the Maximum likelihood (ML) tree confirmed the existence of two clades, one with the East African population of Lake Albert and other the Western populations from Lake Kainji and the Senegal River (Figures 5.3).Thus, there was evidence of haplotype sorting and minimal haplotype sharing exhibited amongst the three sampled populations across drainages. There was no single haplotype shared between the West African populations (Lake Kainji and the Senegal River) and the East African individuals from Lake Albert (Figure 5.2 & 5.3 and Table 5.4). Within the West African sampled individuals, only a single haplotype was shared between the Senegal River and Lake Kainji in Nigeria (Figure 5.2). This shared haplotype between the two West African populations was the highest represented haplotype being found in 26 individuals sampled (two individuals from Lake Kainji and 24 from the Senegal River)

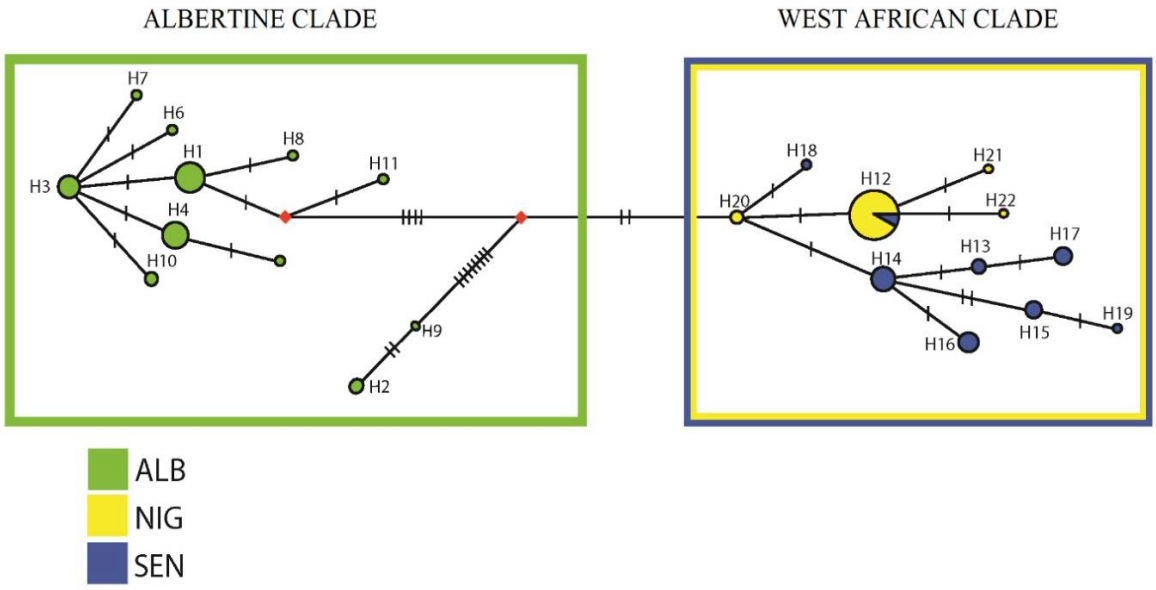


Figure 5.2: Minimum spanning haplotype network depicting mitochondrial d-loop region relationships from *Bagrus bayad* individuals sampled from each geographic location; ALB, Lake Albert (green) drainage from East Africa, SEN, Senegal River (yellow) and NIG, Lake Kainji (blue) from West A\

rica. Circles denote haplotypes and the size of each circle corresponds to the number of individuals sharing that haplotype. Circles are connected to one another by a single mutation and the cross bars indicate any additional mutational step between the haplotypes. Note: the two red diamonds denote haplotypes that were not sampled in the current study, but were predicted by the program to exist.

Table 5.4: Distribution of *Bagrus bayad* haplotype frequencies across sampled drainages in East and West Africa

			Haplotype frequency distribution among sample locations			
Haplotype Name	Sequence Name	Sampling site/Location	Lake Albert	Lake Kainji	Senegal River	Total
H1	ALB-BB0074	Aboko	9	-	-	9
H2	ALB-BB0066	Murchison falls	2	-	-	2
H3	ALB-BB0067	Aboko	5	-	-	5
H4	ALB-BB0072	Aboko	8	-	-	8
H5	ALB-BB0081	Aboko	1	-	-	1
H6	ALB-BB0090	Wanseko	1	-	-	1
H7	ALB-BB0092	Wanseko	1	-	-	1
H8	ALB-BB0093	Wanseko	1	-	-	1
H9	ALB-BB0094	Wanseko	1	-	-	1
H10	ALB-BBNTK01	Ntoroko	2	-	-	2
H11	ALB-BBNTK05	Ntoroko	1	-	-	1
H12	NIG-BB27	Monai - Kainji	-	2	24	26
H13	NIG-BB01	Monai - Kainji	-	2	-	2
H14	NIG-BB02	Monai - Kainji	-	6	-	6
H15	NIG-BB03	Monai - Kainji	-	3	-	3
H16	NIG-BB04	Monai - Kainji	-	4	-	4
H17	NIG-BB06	Monai - Kainji	-	3	-	3
H18	NIG-BB15	Monai - Kainji	-	1	-	1
H19	NIG-BB21	Monai - Kainji	-	1	-	1
H20	SEN-BB0004	Richard Toll	-	-	2	2
H21	SEN-BB0022	Richard Toll	-	-	1	1
H22	SEN-BB0027	Richard Toll	-	-	1	1
Total			**32**	**22**	**28**	**82**

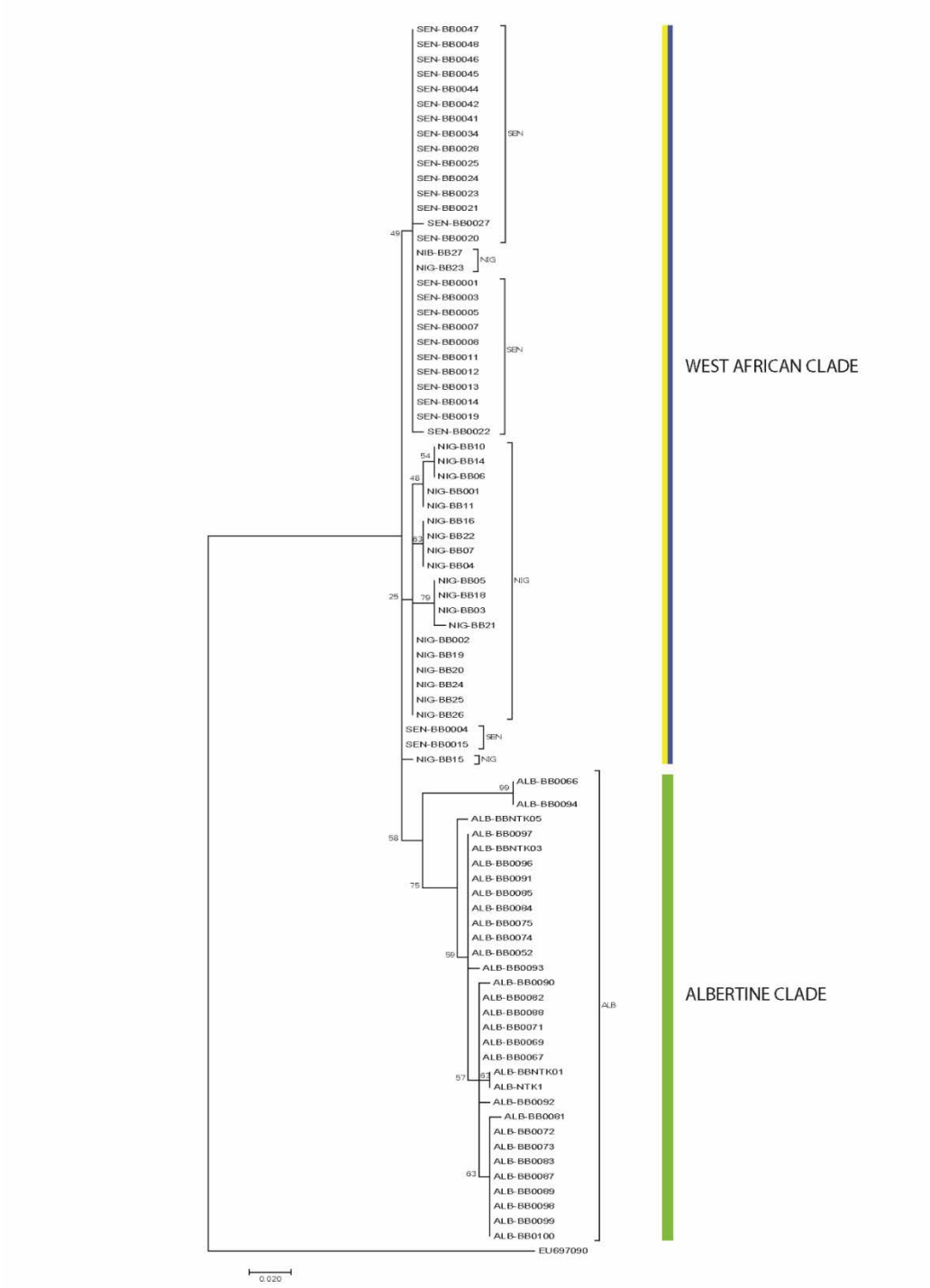


Figure 5.3 *Bagrus bayad* phylogenetic tree as inferred by Maximum likelihood (ML). The analysis involved 82 nucleotide sequences of *B bayad*. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). Two major clades are inferred, one clade consisting of a monophyletic group of individuals from ALB (Lake Albert) in East Africa and the second major clade, a West African clade, consisting of mixed haplotypes from NIG (Lake Kainji) and SEN (Senegal).

5.3.2 Microsatellite DNA

5.3.2.1 Genetic diversity at nine microsatellite DNA loci

When assessing the microsatellite data, Microchecker did not detect any evidence of scoring errors due to stuttering and/or large allelic dropouts at any of the loci examined. However, for the Senegal River population, null alleles were suggested at most loci genotyped (i.e., *Bd04, Bd01, Bd02, Bd12, Bd09, Bd20, Bd05, Bd16 and Bd08*). Deviations from Hardy-Weinberg Equilibrium (HWE) were detected at locus *Bd05* in all populations and additionally, *Bd04*, for Lake Albert and *Bd08* for the Senegal River populations. Analyses were conducted including and excluding the locus *Bd05* (unpublished data). Removal of this locus did not impact on results and thus all analyses were presented with this locus included.

Based on most diversity indices (including, *Ar,* H_o, and H_e), Lake Albert and Lake Kainji showed the highest genetic diversity, whilst Senegal River bayad catfish exhibited the lowest genetic diversity (Table 5.5). The total number of alleles detected across all populations at the nine loci was 126 alleles, with Lake Albert having the highest number of alleles (77 alleles), followed by Lake Kainji on the Niger River (64 alleles). Senegal had the lowest number of alleles genotyped (62 alleles) (Appendix 5.3). Although, allelic richness (*Ar*) takes into account the differences in sample size, the Senegal River still had the lowest *Ar* of 5.8 ± 1.2 and Lake Albert (*Ar* = 7.6 ± 1.2) had the highest, thus the *Ar* results corroborated the trend observed in the other genetic diversity indices among the populations (Table 5.5, Appendix 5.4, Figure 5.4). These microsatellite results were congruent with the mitochondrial DNA results in which the Senegal River population emerges as the population exhibiting the lowest within-location genetic diversity.

Table 5.5: Microsatellite DNA diversity indices from nine markers for 95 *Bagrus bayad* from three freshwater systems in Africa

Location	*N*	$N_a \pm SE$	$AR \pm SE$	$H_o \pm SE$	$H_e \pm SE$	F_{IS} (IC 95%)
Lake Albert	28	8.6 ±1.6	7.6 ± 1.2	0.62±0.04	0.72 ± 0.03	0.15 (0.04-0.20)
Lake Kainji	25	7.1 ±0.9	6.6 ±0.7	0.67 ± 0.05	0.70 ± 0.03	0.07 (-0.03 -0.0.12)
Senegal River	42	6.9 ±1.2	5.8 ± 1.1	0.42 ± 0.09	0.53 ± 0.11	0.21 (0.01-0.29)

N, number of individuals, N_a, mean number of alleles, H_o, mean observed heterozygosity, H_e, mean expected heterozygosity F_{IS}, inbreeding coefficient and IC, confidence interval.

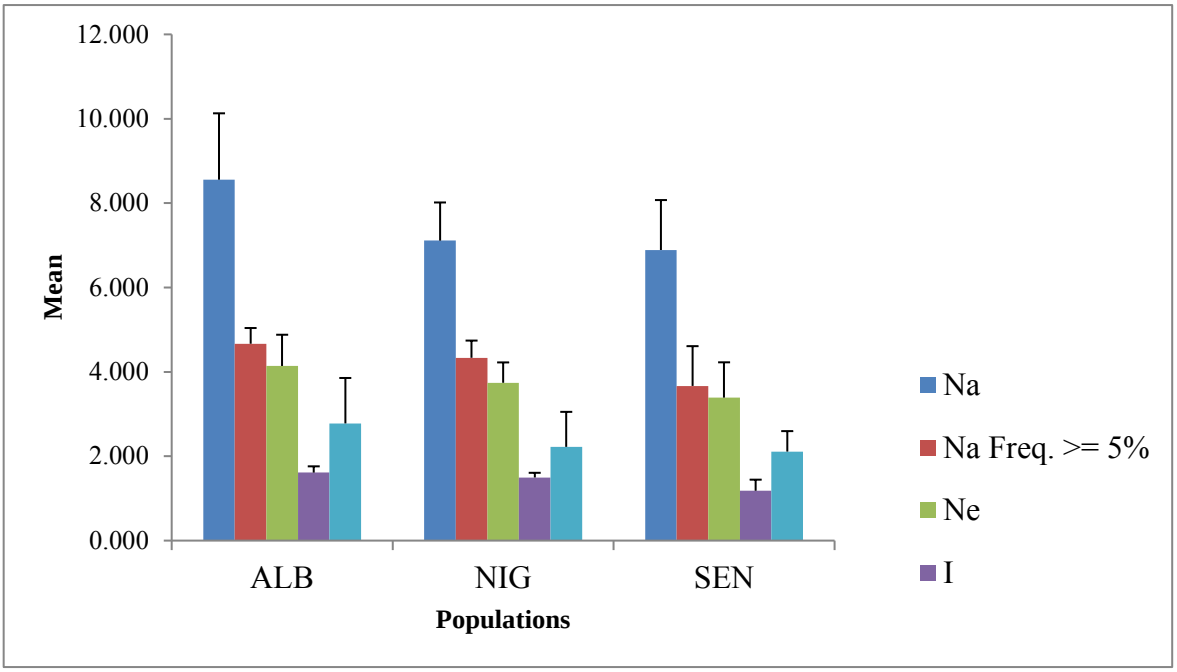


Figure 5.4: Allelic patterns observed across the sampled *Bagrus bayad* populations of ALB- Lake Albert, NIG- Lake Kainji and SEN- the Senegal River. Na mean number of alleles and Ne, the effective mean number of alleles. The variation in data is shown by the error bars (standard error, SE).

5.3.2.2 Population genetic structure and differentiation

Strong genetic structuring was detected among *B. bayad* from the three drainages. Using the admixture model, as implemented in STRUCTURE 2.1 (Pritchard et al., 2000), and regardless of assuming independent or correlated allele frequencies, three distinct genetic clusters (k = 3) were identified in *B. bayad.* This shows that *B. bayad* in Africa comprises *at least* three independent genetic stocks (Figure 5.5).

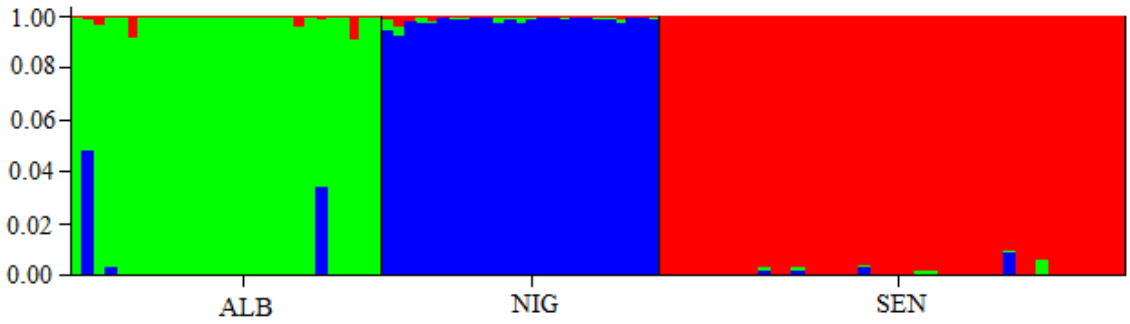


Figure 5.5: STRUCTURE bar plot showing the proportion of genome ancestry for 95 *Bagrus bayad* individuals sampled from three drainages (ALB-Lake Albert, NIG-Lake Kainji on the Niger River and SEN-Senegal River) based on analysis of nine microsatellite loci. The individuals were assigned to the three genetic clusters, each cluster corresponding to a colour; green, blue and red. Individuals from Lake Albert were assigned to the green cluster, while individuals from Lake Kainji and Senegal River were predominately assigned to the blue and red clusters, respectively.

AMOVA results of grouped microsatellite data (into East and West African populations) revealed that the least source of variation was that among groups, with genetic structuring between the two regions being insignificant (percentage variation = -3.36%, F_{CT} = -0.03 and P = 0.665) which was contrary to the mtDNA data where variation levels were very high, but also not significant (71.6 %, P = 0.33). The percent variation due to among populations within groups was more than threefold higher for the microsatellite data (percentage variation = 35.46%, F_{SC} = 0.34 and P = 0.000) compared to the mtDNA data (10%, P = 0.006) (Tables

5.6 & 5.2, respectively). The highest source of variation was due to that within individuals (F_{IT} = 0.36, P = 0.000).

Using ungrouped microsatellite data, *B. bayad* were found to exhibit strong genetic differentiation at the microsatellite DNA loci among populations from the three drainages of Lake Albert, Lake Kainji and the Senegal River (Global F_{ST} = 0.32, P = 0.000). Further pairwise comparisons between populations revealed significant genetic differentiation, with F_{ST} ranging from F_{ST} = 0.23 (P = 0.000) between Lakes Albert and Kainji to F_{ST} = 0.37 (P = 0.000) between Lake Kainji and the Senegal River. The significantly high pairwise F_{ST} estimates again show that the different drainages each constitute an independent genetic population unit.

Table 5.6: Hierarchical Analysis of Molecular Variance (AMOVA) for *Bagrus bayad* microsatellite data

Source variation	Component of Variance	Percentage Variation	Fixation index	P- Value
Grouped data East And West African drainage systems				
Among groups	Va	-.3.32	F_{CT} = -0.033	0.665
Among populations within groups	Vb	35.4	F_{SC} = 0.340	0.000
Among individuals Within populations	Vc	4.84	F_{IS} = 0.071	0.002
Within individuals	Vd	63.02	F_{IT} = 0.37	0.000
Ungrouped data				
Among populations	Va	32.94	F_{ST} = 0.33	0.000
Within populations	Vb	4.84	F_{IS} = 0.07	0.002

Table 5.7: Microsatellite pairwise F_{ST} values (for 95 *B. bayad* individuals) between the three sampled major waterbodies in Africa, Lake Albert (ALB) in East Africa, the Senegal River (SEN) and Lake Kainji (NIG) on the Niger River in West Africa (estimates were performed in Arlequin 3.5 at 16,000 permutations). Below diagonal is F_{ST} and above significance.

Population	ALB	NIG	SEN
ALB	-	0.00	0.00
NIG	0.23	-	0.00
SEN	0.36	0.37	-

5.4 Discussion

5.4.1 Genetic structure and differentiation

The present study examined the genetic diversity and structure of *B. bayad* in three major freshwater localities; Lake Albert in East Africa, Lake Kainji on the Niger River, and the Senegal River in West Africa. The region from where fish were sampled is characterised by known biogeographical barriers, such as the EAR, and has also been known to have undergone extreme fluctuations in the climate during the late Pleistocene and early Holocene (Olaka et al., 2010, Ferguson and Harbott, 1972). These biogeographic barriers and a dynamic climate have been suggested as major drivers of evolution in present day teleost biodiversity in Africa (Lévêque, 1997, Elmer et al., 2009, Seehausen, 2002, Seehausen, 2006).

Analysis of mitochondrial DNA d-loop sequences in the current study highlighted two genetically distinct stocks of *B. bayad* between East and West Africa. Evidence for haplotype sorting between the West and East African *B. bayad* was also indicative of long-term historical barriers to gene flow among these populations. This pattern of genetic differentiation appears to be concordant with that observed in the Nile perch (Chapter 4) between the two regions. Additionally a similar pattern has been suggested in two other aquatic associated African species (e.g. African buffalo, *Syncerus caffer,* and Nile tilapia *Oreochromis niloticus)*, further confirming the presence of a significant biogeographic driver of genetic structuring between Western and Eastern Africa aquatic (and terrestrial) populations (Smitz et al., 2013, Agnese et al., 1997).

Within the West African clade, evidence of historical genetic connectivity of Lake Kainji and the Senegal River populations was observed in the current study, through a shared major

haplotype. Corroborative evidence was also shown by all sequences from the two sample locations belonging to one major phylogenetic clade, confirming historical genetic connectivity of these two populations in West Africa. This genetic connectivity could be congruent with geological data that has shown historical hydrological connectivity of African rivers through permanent waterways during the humid Holocene period (Stankiewicz and de Wit, 2006, Stewart, 2009, Stewart, 2001, Petit-Maire, 1989, Petit-Maire et al., 1983). In particular, the Niger River basin has been identified in the West African region to have experienced major floods during this humid Holocene period and it is thus speculated that faunal exchange and increased gene flow between the river systems in West Africa may have occurred (Fontes et al., 1991).

At the contemporary microsatellite DNA level, further separation of up to three genetic groupings among the sampled *B bayad* was revealed, with each genetic group corresponding to catfish within a discrete freshwater drainage. Many freshwater species have been shown to exhibit very limited natural dispersal among drainages due to the natural boundaries that often separate these drainages (Hughes et al., 1999). Therefore the findings of substantial genetic differentiation among drainages of the freshwater *B. bayad* in the present study are consistent with other biogeographical studies that have often linked freshwater drainage boundaries as being strong and effective barriers, limiting dispersal and gene flow (Linder et al., 2012, Montoya-Burgos, 2003, Rahel, 2007, Seehausen, 2002, Todd, 2013).

5.4.2 Genetic diversity and demography

The higher genetic diversity observed in *B. bayad* in the lotic environments of Lake Albert in East Africa and Lake Kainji in West Africa may suggest that these environments are more stable compared to the lentic Senegal River drainage that had low genetic diversity. The

contrasting demographic history of Lake Albert from the other two west African populations in which the *B. bayad* from this lake showed a bi-modal near ragged mismatch distribution further affirms and is congruent with characteristics of stable populations (Figure 5.1)(Lawton et al., 2011)

Considerable and consistently low genetic diversity was revealed within the Senegal River population, which indicates that this population may have undergone a severe historical bottleneck given the negative and significant *D* and *Fu's F* statistic obtained in mtDNA analyses. It is possible that the Senegal River is recovering from this historic bottleneck which probably was a result of long severe drought periods that have been reported to have impacted the Gambia and Senegal rivers, among other freshwater bodies in Africa, about 18,000 years ago (Cushing et al., 2006, Roberts, 1975, Seehausen, 2002, Elmer et al., 2009). The flow of the Senegal River is believed to have been greatly reduced, or even to have ceased during this time and evidence from other species suggest a recolonization of the rivers during times of flooding (Cushing et al., 2006).

The evidence for strong genetic structuring and reduction in mitochondrial DNA diversity among *B. bayad* from the Senegal River is sufficient to consider this population as a discrete management and conservation unit. This is largely because low genetic diversity often impacts negatively on small populations as they may fail to adapt, or even persist, under harsh ecological and environmental pressures (Milot et al., 2007, Christie et al., 2012).

5.4.3 Management implications

Similar to other commercial teleost fauna, management of *B bayad* populations in the natural environment will require routine auditing and findings from the current study provide a

baseline upon which future monitoring can be conducted to manage and conserve the species accordingly. The revealed genetic separation between the *B. bayad* populations of West and East Africa, coupled with varied genetic diversity, warrants that these two regions to be recognised as separate genetic groups for management at a continental scale. This is because the two genetic groupings appear to have evolved in isolation, and consequently, may have independently accrued adaptive differences. Therefore, any human mediated movements of *B. bayad* across the two regions will require careful consideration and needs to be discouraged.

Secondly, when considering breeding programs aimed at restocking, founder stocks should be sought independently within each of the water bodies (i.e. from Lake Albert, Lake Kainji and the Senegal River). This is because three distinct genetic stocks were identified using microsatellite DNA markers, with genetic groupings restricted to catfish sampled within each drainage. Thus, *B. bayad* from Lake Albert, Lake Kainji and the Senegal should be treated separately as management units. Patterns of strong genetic differentiation at such a geographic scale often implies restricted gene flow and limited dispersal among the drainages. Therefore, intentional movement of *B bayad* across drainages should be discouraged in order to prevent possible genetic introgression. Previous translocations of freshwater fishes have raised concerns and such translocations have negatively impacted native populations, a situation that can be avoided with appropriate scientific knowledge as now gerneated in this study (Eldridge and Naish, 2007, Madeira et al., 2005, Palsbøll et al., 2007, Allendorf et al., 2001).

Conversely from an aquaculture perspective and given that *B bayad* is a candidate species for farming in Africa, this chapter highlights the presence of discrete genetic stocks that can

be exploited when starting a breeding program. Mixing together these genetically distinct populations provides an opportunity to develop a breeding program with a diverse genetic base and that maximises available genetic diversity. A similar approach was utilized in the development of the GIFT strain of Nile tilapia, where individuals from various sources of up to eight strains were mixed together to form a composite foundation population for this genetic improvement program (Eknath et al., 1998, Eknath et al., 1993).

4.4.3 Conclusion

The current study has confirmed that the natural freshwater boundaries of *B. bayad* constrain dispersal and gene flow of the species leading to strong genetic structuring. The strong genetic structuring and varied genetic diversity observed among the populations of *B. bayad* linked to their respective drainages suggest that each population should be treated as a separate management unit for the species' conservation and sustainable exploitation. However, it is important to note that the current sampling was limited, both in geographic coverage and sample size, given the wider distribution of *B. bayad* across Africa. Therefore, more sampling to include other locations in order to identify boundaries and further clearly refine management units than have been identified presently will benefit this fishery at a continental scale.

Finally, from an aquaculture perspective, the unique genetic stocks coupled with varying levels of genetic diversity identified may be used to the advantage of the industry by developing a diverse genetic base population for breeding which maximises genetic diversity.

CHAPTER 6

6.0 General Discussion

Work presented in this PhD gives an account of the molecular genetic resources developed to explore evolutionary histories, genetic population structure and diversity of three widely distributed, commercially important African freshwater fish species. The species chosen (*Bagrus docmak*, *B. bayad* and *Lates niloticus*) belong to two divergent genera (*Bagrus* and *Lates*) and are important foodfish, as well as recreationally fished species in different parts of Africa. Within the complex geological and historically dynamic climate that characterizes the African continent, these ecologically divergent freshwater taxa are found to co-exist in some parts of their range. Their use in inferring microevolutionary biogeographic patterns throughout this thesis demonstrates a common sensitivity to the historical and contemporary hydrological systems and changes that have occurred over time and shaped their population genetic structure.

6.1 Molecular markers and their use in Africa's freshwater fisheries management and conservation

The use of genomic resources has rapidly advanced knowledge in population genetics, breeding and conservation. Although genomic resources and their rapid development using next generation sequencing (NGS) have revolutionized research (Ekblom and Galindo, 2010, McCormack et al., 2013, Liu et al., 2011, Hale et al., 2009, Abdelkrim et al., 2009), their utility is still limited when applied to African freshwater fishes. This is not inexplicable, as studies have shown that just a small proportion of approximately 0.2 % of fish species in the world have significant genomic resources useful in giving insights into environmental and ecological adaptions (Oleksiak, 2010). Therefore, development of genetic resources among

natural fish populations has the great potential to revolutionise research for both captive and natural fish populations of widely distributed taxa. In this thesis (Chapter 2), using the 454 Roche GS-FLX next generation sequencing platform, a number of genetic tools were tested and reliably amplified in three widely distributed freshwater fish taxa in Africa. Thus, although the purpose of developing these genetic tools was for inference of the genetic structure and biogeography of the target fish species (Chapters 3 to 5), the robustness of the microsatellites for both *Bagrus* and *Lates* spp. will be valuable resources for further research in these species. Essentially, the genetics tools will be a powerful resource to not only continue to understand the genetics of *Bagrus* and *Lates* spp., but they may also offer the opportunity to be cross-amplified in closely related taxa providing genomic resources for presently other uncharacterized species.

The most common inferences on Africa's freshwater biogeography have been based on ecological observations of primarily species presence, absence, distribution, as well as endemism among the lakes and rivers in the region. However, when investigated from a genetics perspective, the few genetic studies that have been carried out in the region have primarily used mtDNA to make phylogeographic and biogeographic inferences (Koblmüller et al., 2005, Nagl et al., 2001, Mwanja et al., 2013, Crozier, 1990, Giddelo et al., 2002, Mwita and Nkwengulila, 2008). Using this approach alone has received criticism, with several reviews indicating that mtDNA inferences are limited when evaluated on their own to provide fine scale genetic resolution (Crozier, 1990, Rogers et al., 1996, Godinho et al., 2008). Thus, although mtDNA is informative for phylogeographic studies, combining mtDNA with information from nuclear (multi-allelic) loci can provide much finer resolution of historical and contemporary patterns than provided by mtDNA alone (Godinho et al., 2008, Edwards

and Bensch, 2009). For the first time (in Chapters 2, 3 and 5), species-specific nuclear DNA markers for the African bagrid species were developed and utilized in combination with the mtDNA d-loop region to successfully infer the population genetic structure and diversity of these species in Africa.

6.2 Outcomes and limitations of the present study versus future directions of research

The work in the current thesis has explored explicitly the development and use of microsatellite markers in non-model taxa to resolve population genetics structure and diversity among populations of *Bagus* and *Lates* in Africa. The approach taken in the present study was the implementation of genetic tools to estimate parameters useful in management and conservation, such as identification of unique populations, effective population sizes, population growth and other diversity and demography parameters (Angeloni et al., 2012, Charlesworth and Willis, 2009). Consequently management units for the wild fisheries management and also stocks suitable for aquaculture development have been identified.

Furthermore, although not evaluated herein, microsatellite markers developed (in Chapter 2) will have utility in parentage analysis in future breeding programs for commercial aquaculture development. Particularly the developed genomic resources in the current project can be used to maintain pedigree data and also to assign parentage of offspring to their respective parents, which was unachievable in the life of this project.

Meanwhile, the development of detailed genetic linkage maps useful in locating useful genetic traits and genes of interest important for selective breeding has been made easier and cheaper through next generation sequencing (NGS) (McCormack et al., 2013, Zhang and Sun, 2005, DeFaveri et al., 2013, Hubert et al., Moen et al., 2008, He et al., 2003, Wang et

al., 2008). Thus, although microsatellite genetic resources developed in the current study (Chapter 2) are and will remain useful in maintaining pedigree data, larger numbers of genome-wide markers i.e. SNPs from NGS, will be more powerful in transforming the aquaculture industry in Africa. SNP technology has been already widely embraced and some benefits accrued in Africa already from the use of these genetic markers, especially in more the advanced plant production sector (Morris et al., 2013, Mace et al., 2013). The high resolution of SNPs in comparison to the traditional markers (such as microsatellites) in genome-wide mapping has been emphasized elsewhere in conservation (Angeloni et al., 2012) and herein, recommended for fisheries and aquaculture development practices in Africa. Thus future linkage maps for the species should consider use of SNPs to identify genes of interest and, as such, will go a long way in maximizing genetic gains in fisheries and aquaculture. This is achievable given that the variation of SNP as markers has the power to reveal both functional and neutral genetic variation (Angeloni et al., 2012).

The species (*B. docmak, B bayad* and *L.niloticus*) utilized in the current PhD study are not only widely and co-distributed in parts of their range across sub-Saharan Africa, but are also key to the fisheries sector and aquaculture development (Greenwood, 1966, Mkumbo and Marshall, 2015). Although the species are currently exploited from the wild, futile attempts to close their breeding cycle in captivity has not yet relieved fishing pressure on the natural populations as would be desired (Naylor et al., 2000, Pruginin, 1965). Therefore at present, the management of fisheries resources in Africa that is heavily exploited through capture fisheries (and commercial scale aquaculture in the near future), will benefit from the thorough understanding of genetic structure, diversity and population connectivity provided in the current body of work.

Distinct genetic stocks with varied genetic diversity among populations determined in the present study define management units and boundaries suitable for sustainable exploitation and restocking of depleted and/or threatened populations. Within West and East Africa, sustainable fisheries management of the natural populations of *Bagrus* and *Lates* spp. can now be achieved coupled with viable and responsible aquaculture stocks for the specific species given the population genetic data obtained in the present study. Responsible translocations are also feasible within the confines of genetic boundaries determined to boost wild and captive populations. Although the present study provides this useful and concordant information for two of the species' (*B. bayad* and *L. niloticus*) genetic stock structure, clearly separating the West and East Africa populations, it should be noted that limited sample locations and number of individuals were obtained in some parts of these species' total distribution. The limited sampling locations may have constrained fine scale inferences on population structure along the species' entire distribution gradient (L Pruett and Winker, 2008). Therefore, future studies should utilize finer scale sampling to identify where genetic boundaries occur in order to define more clear management (MUs) and evolutionary significant units (ESUs) than what the present PhD study was able to achieve. Inclusion of additional locations where these species are distributed will require large-scale collaborations in various countries, but will greatly benefit and contribute to the overall management of the species in the wild at continental and regional levels. Furthermore, suitable population genetic stocks will also be ably isolated for commercial aquaculture development in view of potential breeding programs for the continent.

6.3 Conclusion

Overall this thesis provides the first comprehensive set of species-specific genomic resources developed for two freshwater African bagrids (*B. docmak* and *B. bayad*) and the Nile Perch, *L. niloticus*. Using the genomic resources developed, the three freshwater species examined showed strong phylogoegraphic structure and genetic structural patterns that enabled the successful identification of management units (MUs) potentially useful in the conservation of the bagrids and Nile perch. The molecular data generated herein is central to the management of freshwater fisheries in Africa, especially those in West and East Africa (that have been identified each as a unique genetic group) and contributes to the understanding of microevolutionary forces in shaping the evolution of the African freshwater fish fauna. Future work should aim at large scale sampling to include the greater Central African drainages in order to identify if there are any gradients of geographical genetic structuring from the West to the East Coast of Africa.

BIBLIOGRAPHY

ABDELKRIM, J., ROBERTSON, B. C., STANTON, J. A. L. & GEMMELL, N. J. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques,* 46**,** 185-192.

ABDUL-MUNEER, P. 2014. Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. *Genetics Research International,* 2014.

ACHIENG, A. 1990. The impact of the introduction of Nile perch, Lates niloticus (L.) on the fisheries of Lake Victoria. *Journal of Fish Biology,* 37**,** 17-23.

AGNESE, J. F., ADEPO-GOURENCE, B., ABBAN, E. & FERMON, Y. 1997. Genetic differentiation among natural populations of the nile tilapia, Oreochromis niloticus (Teleosti, cichlidae). *Heredity,* 79**,** 88-96.

ALHASSAN, E. & ANSU-DARKO, M. 2011. Food and Feeding Habits of a Potential Aquaculture Candidate, the Black Nile Catfish, Bagrus Bajad in the Golinga Reservoir. *Australian Journal of Basic & Applied Sciences,* 5.

ALLENDORF, F. W., LEARY, R. F., SPRUELL, P. & WENBURG, J. K. 2001. The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution,* 16**,** 613-622.

AMOS, B. & HOELZEL, A. R. 1991. *Long-term preservation of whale skin for DNA analysis*.

ANGELONI, F., WAGEMAKER, N., VERGEER, P. & OUBORG, J. 2012. Genomic toolboxes for conservation biologists. *Evolutionary Applications,* 5**,** 130-143.

ANJA, H., DADEBO, E. & MENGISTOU, S. 2009. Reproductive biology and condition factor of the catfish Bagrus docmak (Forsskål)(Pisces: Bagridae) in Lake Chamo, Ethiopia. *SINET: Ethiopian Journal of Science,* 32**,** 29-36.

ARNOUX, E., ERAUD, C., NAVARRO, N., TOUGARD, C., THOMAS, A., CAVALLO, F., VETTER, N., FAIVRE, B. & GARNIER, S. 2014. Morphology and genetics reveal an intriguing pattern of differentiation at a very small geographic scale in a bird species, the forest thrush Turdus lherminieri. *Heredity,* 113**,** 514-525.

ARUHO, C., BASIITA, R., KAHWA, D., BWANIKA, G. & RUTAISIRE, J. 2013. Reproductive biology of Bagrus docmak in the Victoria Nile, Uganda. *African Journal of Aquatic Science,* 38**,** 263-271.

BALIRWA, J. S., CHAPMAN, C. A., CHAPMAN, L. J., COWX, I. G., GEHEB, K., KAUFMAN, L., LOWE-MCCONNELL, R. H., SEEHAUSEN, O., WANINK, J. H., WELCOMME, R. L. & WITTE, F. 2003. Biodiversity and Fishery Sustainability in the Lake Victoria Basin: An Unexpected Marriage? *BioScience,* 53**,** 703-715.

BALL, M. C., FINNEGAN, L. A., NETTE, T., BRODERS, H. G. & WILSON, P. J. 2011. Wildlife forensics: "Supervised" assignment testing can complicate the association of suspect cases to source populations. *Forensic Science International: Genetics,* 5**,** 50-56.

BALOGUN, J. 1986. Fish distribution in Kainji lake, Nigeria. *Journal of Fish Biology,* 29**,** 489-498.

BASIITA, R. K., ARUHO, C., KAHWA, D., NYATIA, E., BUGENYI, F. W. & RUTAISIRE, J. 2011. Differentiated gonochorism in Nile perch Lates niloticus from Lake Victoria, Uganda. *African Journal of Aquatic Science,* 36**,** 89-96.

BASIITA, R. K., BRUGGEMANN, J. H., CAI, N., CÁLIZ-CAMPAL, C., CHEN, C., CHEN, J., CIZEK, L., CORDERO, P. J., DAWSON, D. A. & DING, Y. 2016. Erratum to: Microsatellite records for volume 7, issue 4. *Conservation Genetics Resources***,** 1-3.

BASIITA, R. K., ZENGER, K. R., MWANJA, M. T. & JERRY, D. R. 2015. Development of genome-wide microsatellite genetic resources in a commercially important African freshwater fish species–the Nile perch, Lates niloticus. *Animal Genetics,* 46.

BEADLE, L. 1981. The inland waters of Africa. An introduction to tropical limnology. Longman, London.

BEADLE, L. C. 1974. *The inland waters of tropical Africa. An introduction to tropical limnology*, Longman Group Ltd, Publishers, 74 Grosvenor Street, London W. 1.

BELKHIR, K., BORSA, P., CHIKHI, L., RAUFASTE, N. & CATCH, F. 2004. Genetix 4.05. 2. *University of Montpellier II, Laboratoire Génome et Populations, Montpellier, France***,** 55-65.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 289-300.

BENJAMINI, Y., KRIEGER, A. M. & YEKUTIELI, D. 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika,* 93**,** 491-507.

BEUNING, K. R., TALBOT, M. R. & KELTS, K. 1997. A revised 30,000-year paleoclimatic and paleohydrologic history of Lake Albert, East Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology,* 136**,** 259-279.

BISHOP, W. 1965. Quaternary geology and geomorphology in the Albertine Rift Valley, Uganda. *Geological Society of America Special Papers,* 84**,** 293-322.

BOHONAK, A. 2002. IBD (isolation by distance): a program for analyses of isolation by distance. *Journal of Heredity,* 93**,** 153-154.

BRACKEN, F. S., HOELZEL, A. R., HUME, J. B. & LUCAS, M. C. 2015. Contrasting population genetic structure among freshwater-resident and anadromous lampreys: the role

of demographic history, differential dispersal and anthropogenic barriers to movement. *Molecular Ecology,* 24**,** 1188-1204.

BROWNSTEIN, M. J., CARPTEN, J. D. & SMITH, J. R. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques,* 20**,** 1004-6, 1008-10.

BUNDY, A. & PITCHER, T. J. 1995. An analysis of species changes in Lake Victoria: did the Nile perch act alone? *The impact of species changes in African lakes.* Springer.

BUTCHART, S. H., WALPOLE, M., COLLEN, B., VAN STRIEN, A., SCHARLEMANN, J. P., ALMOND, R. E., BAILLIE, J. E., BOMHARD, B., BROWN, C. & BRUNO, J. 2010. Global biodiversity: indicators of recent declines. *Science,* 328**,** 1164-1168.

CABALLERO, A. & GARCÍA-DORADO, A. 2013. Allelic diversity and its implications for the rate of adaptation. *Genetics,* 195**,** 1373-1384.

CARVALHO, G. R. & HAUSER, L. 1995. Genetic impacts of fish introductions: a perspective on African lakes. *The Impact of Species Changes in African Lakes.* Springer.

CHAPMAN, B., SKOV, C., HULTHÈN, K., BRODERSEN, J., NILSSON, P. A., HANSSON, L. A. & BRÖNMARK, C. 2012. Partial migration in fishes: definitions, methodologies and taxonomic distribution. *Journal of Fish Biology,* 81**,** 479-499.

CHAPMAN, L. & CHAPMAN, C. 2003. Fishes of the African rain forests: Emerging and potential threats to a littleknown fauna. *Conservation, ecology, and management of African freshwaters. Univ. Press of Florida***,** 176-209.

CHAPMAN, L., CHAPMAN, C., KAUFMAN, L., WITTE, F. & BALIRWA, J. 2007. Biodiversity conservation in African inland waters: lessons of the Lake Victoria region. *Proceedings - International Association of Theoretical and Apllied Limnology,* 30**,** 16.

CHAPMAN, L. J., KAUFMAN, L. S., CHAPMAN, C. A. & MCKENZIE, F. E. 1995. Hypoxia tolerance in twelve species of East African cichlids: potential for low oxygen refugia in Lake Victoria. *Conservation Biology,* 9**,** 1274-1288.

CHARLESWORTH, D. & WILLIS, J. H. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics,* 10**,** 783-796.

CHENOWETH, S. F. & HUGHES, J. M. 1997. Genetic population structure of the catadromous Perciform: Macquaria novemaculeata (Percichthyidae). *Journal of Fish Biology,* 50**,** 721-733.

CHISTIAKOV, A. D., HELLEMANS, B. & VOLCKAERT, A. M. F. 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture,* 255**,** 1-29.

CHRISTIE, M. R., MARINE, M. L., FRENCH, R. A. & BLOUIN, M. S. 2012. Genetic adaptation to captivity can occur in a single generation. *Proceedings of the National Academy of Sciences,* 109**,** 238-242.

COHEN, A., LEZZAR, K.-E., TIERCELIN, J.-J. & SOREGHAN, M. 1997. New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. *Basin Research,* 9**,** 107-132.

COHEN, A. S., SOREGHAN, M. J. & SCHOLZ, C. A. 1993a. Estimating the age of formation of lakes – an example from Lake Tanganyika, East-African Rift System. *Geology,* 21.

COHEN, A. S., SOREGHAN, M. J. & SCHOLZ, C. A. 1993b. Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. *Geology,* 21**,** 511-514.

CORUJO, M., BLANCO, G., VAZQUEZ, E. & SANCHEZ, J. A. 2004. Genetic structure of northwestern Spanish brown trout (Salmo trutta L.) populations, differences between microsatellite and allozyme loci. *Hereditas,* 141**,** 258-71.

COULTER, G. 1976. The biology of Lates species (Nile perch) in Lake Tanganyika, and the status of the pelagic fishery for Lates species and Luciolates stappersii (Blgr.). *Journal of Fish Biology,* 9**,** 235-259.

CRAIG, J. F. 1992. Human-induced changes in the composition of fish communities in the African Great Lakes. *Reviews in Fish Biology and Fisheries,* 2**,** 93-124.

CROZIER, R. 1990. From population genetics to phylogeny: uses and limits of mitochondrial DNA. *Australian Systematic Botany,* 3**,** 111-124.

CUSHING, C. E., CUMMINS, K. W. & MINSHALL, G. W. 2006. *River and stream ecosystems of the world*, Univ of California Press.

DADZIE, S. 1992. An overview of aquaculture in eastern Africa. *Hydrobiologia,* 232**,** 99-110.

DANIELS, S. R., PHIRI, E. E., KLAUS, S., ALBRECHT, C. & CUMBERLIDGE, N. 2015. Multilocus phylogeny of the Afrotropical freshwater crab fauna reveals historical drainage connectivity and transoceanic dispersal since the Eocene. *Systematic biology***,** syv011.

DANLEY, P. D., HUSEMANN, M., DING, B., DIPIETRO, L. M., BEVERLY, E. J. & PEPPE, D. J. 2012. The impact of the geologic history and paleoclimate on the diversification of East African cichlids. *International Journal of Evolutionary Biology,* 2012.

DAWSON, M. N., RASKOFF, K. A. & JACOBS, D. K. 1998. Field preservation of marine invertebrate tissue for DNA analyses. *Molecular Marine Biology and Biotechnology,* 7**,** 145-152.

DAY, J. J., PEART, C. R., BROWN, K. J., FRIEL, J. P., BILLS, R. & MORITZ, T. 2013. Continental diversification of an African catfish radiation (Mochokidae: Synodontis). *Systematic Biology,* 62**,** Pp. 351-365.

DEFAVERI, J., VIITANIEMI, H., LEDER, E. & MERILÄ, J. 2013. Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular Ecology Resources,* 13**,** 377-392.

DEMENOCAL, P. B. 2004. African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth and Planetary Science Letters,* 220**,** 3-24.

DGEBUADZE, Y. Y., GOLUBTSOV, A., MIKHEEV, V. & MINA, M. 1994. Four fish species new to the Omo-Turkana basin, with comments on the distribution ofNemacheilus abyssinicus (Cypriniformes: Balitoridae) in Ethiopia. *Hydrobiologia,* 286**,** 125-128.

DICKSON, M., JAGWE, J., LONGLEY, C. & DALSGARD, J. 2012. Uganda aquaculture value chains: Strategic planning mission report.

DIOGO, R., VANDEWALLE, P. & CHARDON, M. 1999. Morphological Description of the Cephalic Region of Bagrus Docmac, With a Reflection On Bagridae (Teleostei: Siluriformes) Autapomorphies. *Netherlands Journal of Zoology,* 49**,** 207-232.

DO, C., WAPLES, R. S., PEEL, D., MACBETH, G., TILLETT, B. J. & OVENDEN, J. R. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources,* 14**,** 209-214.

DRUMMOND, A., ASHTON, B., CHEUNG, M., HELED, J., KEARSE, M., MOIR, R., STONES-HAVAS, S., THIERER, T. & WILSON, A. 2009. Geneious v4. 7. *Biomatters, Ltd., Auckland, New Zealand*.

EARL, D. & VONHOLDT, B. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources,* 4**,** 359-361.

EDWARDS, S. & BENSCH, S. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology,* 18**,** 2930-2933.

EKBLOM, R. & GALINDO, J. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity,* 107**,** 1-15.

EKNATH, A., DEY, M., RYE, M., GJERDE, B., ABELLA, T., SEVILLEJA, R., TAYAMEN, M., REYES, R. & BENTSEN, H. Selective breeding of Nile tilapia for Asia. 6th World Congress on Genetics Applied to Livestock Production, 1998. University of New England Armidale, Australia, 89-96.

EKNATH, A. E., TAYAMEN, M. M., PALADA-DE VERA, M. S., DANTING, J. C., REYES, R. A., DIONISIO, E. E., CAPILI, J. B., BOLIVAR, H. L., ABELLA, T. A. & CIRCA, A. V. 1993. Genetic improvement of farmed tilapias: the growth performance of eight strains of Oreochromis niloticus tested in different farm environments. *Aquaculture,* 111**,** 171-188.

EL-DRAWANY, M. & ELNAGAR, W. 2015. Growth, Food and Feeding Habits of Bagrus bayad and Bagrus docmac Inhibiting Muess Channel, Sharkia Province, Egypt. *Journal of Aquaculture Research & Development,* 2015.

ELDRIDGE, W. H. & NAISH, K. A. 2007. Long-term effects of translocation and release numbers on fine-scale population structure among coho salmon (Oncorhynchus kisutch). *Molecular Ecology,* 16**,** 2407-2421.

ELMER, K. R., REGGIO, C., WIRTH, T., VERHEYEN, E., SALZBURGER, W., MEYER, A. & WAKE, D. B. 2009. Pleistocene Desiccation in East Africa Bottlenecked but Did Not Extirpate the Adaptive Radiation of Lake Victoria Haplochromine Cichlid Fishes. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 13404-13409.

EVANNO, G., REGNAUT, S. & GOUDET, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology,* 14**,** 2611-2620.

EXCOFFIER, L. & LISCHER, H. 2010. S., Schneider. 2010. Arlequin suite ver. 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* 10**,** 464-567.

FAIRCLOTH, B. C. 2008. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources,* 8**,** 92-94.

FAO 2006. State of world aquaculture. *FAO Fisheries Technical Paper.* Rome.

FERGUSON, A. & HARBOTT, B. 1972. Geographical, physical and chemical aspects of Lake Turkana. *Lake Turkana: A report on the findings of the Lake Turkana Project,* 1975**,** 1-107.

FONTES, J. C., ANDREWS, J. N., EDMUNDS, W. M., GUERRE, A. & TRAVI, Y. 1991. Paleorecharge by the Niger river (Mali) deduced from groundwater geochemistry. *Water Resources Research,* 27**,** 199-214.

FROESE, R. A. D. P. 2009. FishBase.World Wide Web electronic publication

Froese, R. and D. Pauly.

FU, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics,* 147**,** 915-925.

GIDDELO, C. S., ARNDT, A. D. & VOLCKAERT, F. A. M. 2002. Impact of rifting and hydrography on the genetic structure of Clarias gariepinus in eastern Africa. *Journal of Fish Biology,* 60**,** 1252-1266.

GODINHO, R., CRESPO, E. & FERRAND, N. 2008. The limits of mtDNA phylogeography: complex patterns of population history in a highly structured Iberian lizard are only revealed by the use of nuclear markers. *Molecular Ecology,* 17**,** 4670-4683.

GOLUBTSOV, A., DARKOVE, A., DGEBYADZE, Y. & MINA, M. 1995. An artificial key to fish species of Gambela region. Joint. Ethio-Rassian Biological Expedition. Artistic Printing Press, Addis Ababa.

GOOSSENS, B. 2015. Microsatellite records for volume 7, issue 4. *Conservation Genetics Resources,* 7**,** 917-944.

GOUDET, J. 2002. Fstat 2.9. 3.2. *URL: http://www2. unil. ch/popgen/softwares/fstat. htm*.

GOUDIE, C. A., SIMCO, B. A., DAVIS, K. B. & PARKER, N. C. 1992. Reproductive Performance of Pigmented and Albino Female Channel Catfish Induced to Spawn with HCG or Ovaprim. *Journal of the World Aquaculture Society,* 23**,** 138-145.

GOUDSWAARD, K. & WITTE, F. 1997. The catfish fauna of Lake Victoria after the Nile perch upsurge. *Environmental Biology of Fishes,* 49**,** 21-43.

GRANT, W. & BOWEN, B. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity,* 89**,** 415-426.

GREENBAUM, G., TEMPLETON, A. R., ZARMI, Y. & BAR-DAVID, S. 2014. Allelic Richness following Population Founding Events–A Stochastic Modeling Framework Incorporating Gene Flow and Genetic Drift. *PloS one,* 9**,** e115203.

GREENWOOD, P. H. 1966. *Fishes of Uganda,* Kampala, The Uganda Society.

GRIFFITHS, A. M., MACHADO-SCHIAFFINO, G., DILLANE, E., COUGHLAN, J., HORREO, J. L., BOWKETT, A. E., MINTING, P., TOMS, S., ROCHE, W. & GARGAN, P. 2010. Genetic stock identification of Atlantic salmon (Salmo salar) populations in the southern part of the European range. *BMC Genetics,* 11**,** 1.

GUO, S. W. & THOMPSON, E. A. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics***,** 361-372.

HADDON, I. & MCCARTHY, T. 2005. The Mesozoic–Cenozoic interior sag basins of Central Africa: The Late-Cretaceous–Cenozoic Kalahari and Okavango basins. *Journal of African Earth Sciences,* 43**,** 316-333.

HALE, M., MCCORMICK, C., JACKSON, J. & DEWOODY, J. 2009. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (Acipenser fulvescens): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics,* 10**,** 203.

HAMILTON, A. C. 1982. *Environmental history of East Africa: a study of the Quaternary*, Academic press London.

HARDY, O. J. & VEKEMANS, X. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity,* 83**,** 145-154.

HARGREAVES, G. L., HARGREAVES, G. H. & RILEY, J. P. 1985. Irrigation water requirements for Senegal River basin. *Journal of Irrigation and Drainage Engineering,* 111**,** 265-275.

HARRISON, K. 1991. The taxonomy of East African Nile perch, Lates spp.(Perciformes, Centropomidae). *Journal of Fish Biology,* 38**,** 175-186.

HART, P. & PITCHER, T. 2012. *The impact of species changes in African lakes*, Springer Science & Business Media.

HAUSER, L., CARVALHO, G., PITCHER, T. J. & OGUTU-OHWAYO, R. 1998. Genetic affinities of an introduced predator: Nile perch in Lake Victoria, East Africa. *Molecular Ecology,* 7**,** 849-857.

HE, C., CHEN, L., SIMMONS, M., LI, P., KIM, S. & LIU, Z. 2003. Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Animal Genetics,* 34**,** 445 - 448.

HELFMAN, G. S. 2007. *Fish conservation: a guide to understanding and restoring global aquatic biodiversity and fishery resources*, Island Press.

HOLDEN, M. 1967. The systematics of the genus Lates (Teleosti: Centropomidae) in Lake Albert, East Africa. *Journal of Zoology,* 151**,** 329-342.

HOLLELEY, C. E. & GEERTS, P. G. 2009. Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *Biotechniques,* 46**,** 511-517.

HOPSON, A. J. 1972. A study of the nile perch (Lates niloticus (L), Pisces Centropomide) in Lake Chad.

HUBERT, S., HIGGINS, B., BORZA, T. & BOWMAN, S. Development of a SNP resource and a genetic linkage map for Atlantic cod (Gadus morhua). *BMC Genomics,* 11**,** 191.

HUGHES, J., PONNIAH, M., HURWOOD, D., CHENOWETH, S. & ARTHINGTON, A. 1999. Strong genetic structuring in a habitat specialist, the Oxleyan Pygmy Perch Nannoperca oxleyana. *Heredity,* 83**,** 5-14.

HUGHES, N. 1992. Growth and reproduction of the Nile perch, Lates niloticus, an introduced predator, in the Nyanza Gulf, Lake Victoria, East Africa. *Environmental Biology of Fishes,* 33**,** 299-305.

HUGUENY, B. 1989. West African rivers as biogeographic islands: species richness of fish communities. *Oecologia,* 79**,** 236-243.

HUGUENY, B. & LÉVÊQUE, C. 1994. Freshwater fish zoogeography in west Africa: faunal similarities between river basins. *Environmental Biology of Fishes,* 39**,** 365-380.

ISSAWI, B. & MCCAULEY, J. F. 1992. The Cenozoic rivers of Egypt: the Nile problem. *The followers of Horus, Studies dedicated to Michael Allen Hoffman, Oxbow Monograph,* 20**,** 121-138.

ITA, E. 1978. An analysis of fish distribution in Kainji lake Nigeria. *Hydrobiologia,* 58**,** 233-244.

JAABI, S. A. & RASIAH, R. 2014. Assessing the importance of fish exports in the economies of Uganda and the Gambia: A supply-side constraint analysis.

JENSEN, J. L., BOHONAK, A. J. & KELLEY, S. T. 2005. Isolation by distance, web service. *BMC Genetics,* 6**,** 13.

JOHNSON, T. C., HALFMAN, J. D., ROSENDAHL, B. R. & LISTER, G. S. 1987. Climatic and tectonic effects on sedimentation in a rift-valley lake: Evidence from high-resolution seismic profiles, Lake Turkana, Kenya. *Geological Society of America Bulletin,* 98**,** 439-447.

JOHNSON, T. C., KELTS, K. & ODADA, E. 2000. The Holocene History of Lake Victoria. *Ambio,* 29**,** 2-11.

JOHNSON, T. C., SCHOLZ, C. A., TALBOT, M. R. & KELTS, K. 1996. Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science,* 273**,** 1091.

JOMBART, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* 24**,** 1403-1405.

JOMBART, T., DEVILLARD, S. & BALLOUX, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics,* 11**,** 94.

JOSUPEIT, H. 2006. *The Market for Nile Perch*, FAO.

KALINOWSKI, S. T. 2004. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics,* 5**,** 539-543.

KALINOWSKI, S. T. 2005. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes,* 5**,** 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**,** 1099-1106.

KANG, J.-H., PARK, J.-Y. & JO, H.-S. 2012. Rapid development of microsatellite markers with 454 pyrosequencing in a vulnerable fish, the mottled skate, Raja pulchra. *International Journal of Molecular Sciences,* 13**,** 7199-7211.

KATHMAN, J. & SHANNON, M. 2011. Oil extraction and the potential for domestic instability in Uganda. *African Studies Quarterly,* 12**,** 23.

KEENAN, K., MCGINNITY, P., CROSS, T. F., CROZIER, W. W. & PRODÖHL, P. A. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution,* 4**,** 782-788.

KENCHINGTON, F. Observations on the Nile Perch (Lates niloticus) in the Sudan. Proceedings of the Zoological Society of London, 1939. Wiley Online Library, 157-168.

KHALLAF, E. A. 1988. Effect of impoundment on the growth and fishery of Bagrus docmac in lake Nasser. *Regulated Rivers: Research & Management,* 2**,** 547-552.

KHALLAF, E. A. & AUTHMAN, M. N. 1992. Changes in diet, prey size and feeding habit in Bagrus bayad, and possible interactions with B. docmac in a Nile canal. *Environmental Biology of Fishes,* 34**,** 425-431.

KINZIGER, A. P., NAKAMOTO, R. J., ANDERSON, E. C. & HARVEY, B. C. 2011. Small founding number and low genetic diversity in an introduced species exhibiting limited invasion success (speckled dace, Rhinichthys osculus). *Ecology and evolution,* 1**,** 73-84.

KOBLMÜLLER, S., DUFTNER, N., KATONGO, C., PHIRI, H. & C, S. 2005. Ancient divergence in bathypelagic Lake Tanganyika deepwater cichlids: Mitochondrial phylogeny of the tribe Bathybatini. *Journal Molecular Evolution,* 60.

KOCHER, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics,* 5**,** 288-298.

KOKKO, H. & LÓPEZ-SEPULCRE, A. 2006. From individual dispersal to species ranges: perspectives for a changing world. *Science,* 313**,** 789-791.

KOLDING, J. 1995. Changes in species composition and abundance of fish populations in Lake Turkana, Kenya. *The Impact of Species Changes in African Lakes.* Springer.

KUDHONGANIA, A., TWONGO, T. & OGUTU-OHWAYO, G. 1992. Impact of the Nile perch on the fisheries of Lakes Victoria and Kyoga. *Hydrobiologia,* 232**,** 1-10.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution,* 33 1870-1874.

L PRUETT, C. & WINKER, K. 2008. The effects of sample size on population genetic diversity estimates in song sparrows Melospiza melodia. *Journal of Avian Biology,* 39**,** 252-256.

LATCH, E. K., DHARMARAJAN, G., GLAUBITZ, J. C. & RHODES JR, O. E. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics,* 7**,** 295-302.

LAWTON, R. J., MESSMER, V., PRATCHETT, M. S. & BAY, L. K. 2011. High gene flow across large geographic scales reduces extinction risk for a highly specialised coral feeding butterflyfish. *Molecular ecology,* 20**,** 3584-3598.

LÉVÊQUE, C. 1997. *Biodiversity Dynamics and Conservation: The Fresh-water Fish of Tropical Africa*, Cambridge University Press.

LIBRADO, P. & ROZAS, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25**,** 1451-1452.

LINDER, H. P., DE KLERK, H. M., BORN, J., BURGESS, N. D., FJELDSÅ, J. & RAHBEK, C. 2012. The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography,* 39**,** 1189-1205.

LISCHER, H. & EXCOFFIER, L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics,* 28**,** 298-299.

LIU, S., ZHOU, Z., LU, J., SUN, F., WANG, S., LIU, H., JIANG, Y., KUCUKTAS, H., KALTENBOECK, L., PEATMAN, E. & LIU, Z. 2011. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics,* 12**,** 53.

LOWE-MCCONNELL, R. 2009. Fisheries and cichlid evolution in the African Great Lakes: progress and problems. *Freshwater Reviews,* 2**,** 131-151.

LOWE-MCCONNELL, R. H. 1993. Fish Faunas of the African Great Lakes: Origins, Diversity, and Vulnerability

Faunas de peces de los Grandes Lagos Africanos: Orígenes, diversidad y vulnerabilidad. *Conservation Biology,* 7**,** 634-643.

LUCK, G. W. 2007. A review of the relationships between human population density and biodiversity. *Biological Reviews,* 82**,** 607-645.

LUIKART, G., ALLENDORF, F., CORNUET, J. & SHERWIN, W. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity,* 89**,** 238-247.

LUIKART, G. & CORNUET, J. M. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology,* 12**,** 228-237.

LUSWETI, D. 2013. *Genetic Variation of the Lake Victoria Nile Perch.* PhD Dissertation, University of Nairobi.

MACE, E. S., TAI, S., GILDING, E. K., LI, Y., PRENTIS, P. J., BIAN, L., CAMPBELL, B. C., HU, W., INNES, D. J. & HAN, X. 2013. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications,* 4.

MACHADO-SCHIAFFINO, G., KAUTT, A. F., KUSCHE, H. & MEYER, A. 2015. Parallel evolution in Ugandan crater lakes: repeated evolution of limnetic body shapes in haplochromine cichlid fish. *BMC Evolutionary Biology,* 15**,** 9.

MADEIRA, M. J., GÓMEZ-MOLINER, B. J. & BARBÉ, A. M. 2005. Genetic introgression on freshwater fish populations caused by restocking programmes. *Biological Invasions,* 7**,** 117-125.

MANYALA, J., BOLO, J., ONYANGO, S. & RAMBIRI, P. 2005. Indigenous knowledge and baseline data survey on fish breeding areas and seasons in Lake Victoria, Kenya. *Knowledge and Experiences gained from Managing the Lake Victoria Ecosystem.***,** 529-551.

MCCORMACK, J. E., HIRD, S. M., ZELLMER, A. J., CARSTENS, B. C. & BRUMFIELD, R. T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution,* 66**,** 526-538.

MCGLYNN, G., MOONEY, S. & TAYLOR, D. 2013. Palaeoecological evidence for Holocene environmental change from the Virunga volcanoes in the Albertine Rift, central Africa. *Quaternary Science Reviews,* 61**,** 32-46.

MEGLÉCZ, E., COSTEDOAT, C., DUBUT, V., GILLES, A., MALAUSA, T., PECH, N. & MARTIN, J.-F. 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics,* 26**,** 403-404.

MEIRMANS, P. G. 2012. The trouble with isolation by distance. *Molecular Ecology,* 21**,** 2839-2846.

MELO, D. A. M., DA SILVA, B. R. A., VARELA, S. E., SAMPAIO, I. & TAGLIARO, H. C. 2011. Development and Characterization of ten microsatellite markers for population studies of native Brazilian oyster *Crassostrea gasar*. . *Conservation Genetics Resources*

MEYER, A. 1993. Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology & Evolution,* 8**,** 279-284.

MILLER, L. M. & KAPUSCINSKI, A. R. 1996. Microsatellite DNA markers reveal new levels of genetic variation in northern pike. *Trans. Am. Fish. Soc,* 125**,** 971–977.

MILOT, E., WEIMERSKIRCH, H., DUCHESNE, P. & BERNATCHEZ, L. 2007. Surviving with low genetic diversity: the case of albatrosses. *Proceedings of the Royal Society of London B: Biological Sciences,* 274**,** 779-787.

MKUMBO, O. C. & MARSHALL, B. E. 2015. The Nile perch fishery of Lake Victoria: current status and management challenges. *Fisheries Management and Ecology,* 22**,** 56-63.

MOEHL, J. & MACHENA, C. 2000. African Aquaculture: A Regional Summary with Emphasis on Sub-Saharan Africa.

MOEN, T., HAYES, B., BARANSKI, M., BERG, P., KJOGLUM, S., KOOP, B., DAVIDSON, W., OMHOLT, S. & LIEN, S. 2008. A linkage map of the Atlantic salmon (Salmo salar) based on EST-derived SNP markers. *BMC Genomics,* 9**,** 223.

MOHAMED, E. 2016. Morphological revision of the Nile catfish Bagrus bayad (Forsskal, 1775) and Bagrus docmac (Forsskal, 1775)(Pisces: Bagridae).

MONTOYA-BURGOS, J. I. 2003. Historical biogeography of the catfish genus Hypostomus (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. *Molecular Ecology,* 12**,** 1855-1867.

MORIN, P. A., LEDUC, R. G., ARCHER, F. I., MARTIEN, K. K., HUEBINGER, R., BICKHAM, J. W. & TAYLOR, B. L. 2009. Significant deviations from Hardy–Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources,* 9**,** 498-504.

MORRIS, G. P., RAMU, P., DESHPANDE, S. P., HASH, C. T., SHAH, T., UPADHYAYA, H. D., RIERA-LIZARAZU, O., BROWN, P. J., ACHARYA, C. B. & MITCHELL, S. E. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences,* 110**,** 453-458.

MWANJA, M., ARUHO, C., NAMULAWA, V., DDUNGU, R., ONDHORO, C. & BASIITA, R. 2014a. Morphometric variation among Bagrus docmak (Ssemutundu) of the Ugandan major water bodies. *Journal of Fisheries and Aquaculture,* 5**,** 167-172.

MWANJA, M., MUWANIKA, V., MASEMBE, C., NYAKAANA, S. & MWANJA, W. 2013. Evolutionary history of Nile perch Lates sp. inferred from mitochondrial DNA variation analyses. *Zoological Studies,* 52**,** 1-12.

MWANJA, M. T., MUWANIKA, V. B., MASEMBE, C., MWANJA, W. W. & NYAKAANA, S. 2014b. Microsatellite DNA analyses reveal population subdivisions among the recently introduced Nile perch (Lates spp.) in Lake Victoria. *African Zoology,* 49**,** 181-194.

MWANJA, M. T. & MWANJA, W. W. 2008. Preliminary results of the analysis of the population structure of the Nile perch (Lates niloticus linnaeus, 1758) of Lake Victoria using microsatellite markers. *African Journal of Biotechnology,* 7.

MWANJA, M. T., MWANJA, W. W., MUWANIKA, V., MASEMBE, C. & NYAKAANA, S. 2012. Genetic evidence of successful establishment of the Nile perch (Lates spp. L.) in East African lakes and implications for management. *Management,* 3**,** 77-88.

MWANJA, T. M. 2013. *Evolutionary history of the Nile Perch in its native habitat-the nilotic system versus its new haven–the Lake Victoria region waters.* Makerere University.

MWITA, C. J. & NKWENGULILA, G. 2008. Molecular phylogeny of the clariid fishes of Lake Victoria, Tanzania, inferred from cytochrome b DNA sequences. *Journal of Fish Biology,* 73**,** 1139-1148.

MYCHAJLIW, A. M. & HARRISON, R. G. 2014. Genetics Reveal the Origin and Timing of a Cryptic Insular Introduction of Muskrats in North America. *PloS One,* 9**,** e111856.

NAGL, S., MAYER, W. E., SAMONTE, I. E., MCANDREW, B. J. & KLEIN, J. 2001. Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. *Molecular Phylogenetics Evolution,* 20.

NAYLOR, R. L., GOLDBURG, R. J., PRIMAVERA, J. H., KAUTSKY, N., BEVERIDGE, M. C., CLAY, J., FOLKE, C., LUBCHENCO, J., MOONEY, H. & TROELL, M. 2000. Effect of aquaculture on world fish supplies. *Nature,* 405**,** 1017-1024.

NEI, M. & KUMAR, S. 2000. *Molecular evolution and phylogenetics*, Oxford University Press.

ODADA, E. O. & OLAGO, D. O. 2006. *The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity*, Springer Science & Business Media.

ODADA, E. O., OLAGO, D. O., BUGENYI, F., KULINDWA, K., KARIMUMURYANGO, J., WEST, K., NTIBA, M., WANDIGA, S., ALOO-OBUDHO, P. & ACHOLA, P. 2003. Environmental assessment of the East African Rift Valley lakes. *Aquatic Sciences,* 65**,** 254-271.

OGUTU-OHWAYO, R. 1988. Reproductive potential of the Nile perch, Lates niloticus L. and the establishment of the species in Lakes Kyoga and Victoria (East Africa). *Hydrobiologia,* 162**,** 193-200.

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

OGUTU-OHWAYO, R. 1990b. 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.

OGUTU-OHWAYO, R. 1993. The Effects of Predation by Nile Perch, Lates niloticus L., on the Fish of Lake Nabugabo, with Suggestions for Conservation of Endangered Endemic Cichlids. *Conservation Biology,* 7**,** 701-711.

OGUTU-OHWAYO, R. 1994. *Adjustments in fish stocks and in life history characteristics of the Nile perch, Lates niloticus L. in lakes Victoria, Kyoga and Nabugabo.* PhD Thesis, University of Manitoba.

OKACH, J. O. & DADZIE, S. 1988. The food, feeding habits and distribution of a siluroid catfish, Bagrus docmac (Forsskal), in the Kenya waters of Lake Victoria. *Journal of Fish Biology,* 32**,** 85-94.

OKUMUS, I. & CIFTI, 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**,** 51-79.

OLAKA, L. A., ODADA, E. O., TRAUTH, M. H. & OLAGO, D. O. 2010. The sensitivity of East African rift lakes to climate fluctuations. *Journal of Paleolimnology,* 44**,** 629-644.

OLEKSIAK, M. F. 2010. Genomic approaches with natural fish populations. *Journal of Fish Biology,* 76**,** 1067-1093.

OLOWO, J. & CHAPMAN, L. 1999. Trophic shifts in predatory catfishes following the introduction of Nile perch into Lake Victoria. *African Journal of Ecology,* 37**,** 457-470.

OMINDE, S. H. 1971. Studies in East African geography and development.

PALSBØLL, P. J., BERUBE, M. & ALLENDORF, F. W. 2007. Identification of management units using population genetic data. *Trends in Ecology & Evolution,* 22**,** 11-16.

PEAKALL, R. & SMOUSE, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes,* 6**,** 288-295.

PEAKALL, R. & SMOUSE, P. E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics,* 28**,** 2537-2539.

PETIT-MAIRE, N. 1989. Interglacial environments in presently hyperarid Sahara: palaeoclimatic implications. *Paleoclimatology and Paleometeorology: modern and past patterns of global atmospheric transport.* Springer.

PETIT-MAIRE, N., CELLES, J., COMMELIN, D., DELIBRIAS, G. & RAIMBAULT, M. 1983. The Sahara in northern Mali: man and his environment between 10,000 and 3500 years bp.(Preliminary results). *African Archaeological Review,* 1**,** 105-125.

PICKFORD, M., SENUT, B. & HADOTO, D. P. M. 1993. *Geology and Palaeobiology of the Albertine Rift Valley, Uganda-Zaire: Geology*, CIFEG.

PINTON, A., AGNÈSE, J.-F., PAUGY, D. & OTERO, O. 2013. A large-scale phylogeny of Synodontis (Mochokidae, Siluriformes) reveals the influence of geological events on continental diversity during the Cenozoic. *Molecular Phylogenetics and Evolution,* 66**,** 1027-1040.

PRINGLE, R. M. 2005. The Origins of the Nile Perch in Lake Victoria. *BioScience,* 55**,** 780-787.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.

PRUGININ, Y. 1965. Experimental fish culture project in Uganda: Report to the Government of Uganda. Rome: FAO.

RACH, N. M. 1992. Tectonic Controls on the Lake Victoria Basin. *Basement Tectonics 7.* Springer.

RAHEL, F. J. 2007. Biogeographic barriers, connectivity and homogenization of freshwater faunas: it's a small world after all. *Freshwater Biology,* 52**,** 696-710.

REYNOLDS, J. E. & GREBOVAL, D. 1988. *Socio-economic effects of the evolution of Nile perch fisheries in Lake Victoria: a review*, Food & Agriculture Org.

RICCIARDI, A. & RASMUSSEN, J. B. 1999. Extinction rates of North American freshwater fauna. *Conservation Biology,* 13**,** 1220-1222.

RIEDE, K. 2004. *Global register of migratory species: from global to regional scales: final report of the R&D-Projekt 808 05 081*, Federal Agency for Nature Conservation.

ROBERTS, E. M., STEVENS, N., O'CONNOR, P., DIRKS, P., GOTTFRIED, M. D., CLYDE, W., ARMSTRONG, R., KEMP, A. & HEMMING, S. 2012. Initiation of the

western branch of the East African Rift coeval with the eastern branch. *Nature Geoscience,* 5**,** 289-294.

ROBERTS, R. J. 1975. Melanin-containing cells of teleost fish and their relation to disease. *The Pathology of Fishes***,** 399-428.

RODRÍGUEZ, C., BÉCARES, E., FERNÁNDEZ-ALÁEZ, M. & FERNÁNDEZ-ALÁEZ, C. 2005. Loss of diversity and degradation of wetlands as a result of introducing exotic crayfish. *Biological Invasions,* 7**,** 75-85.

ROGERS, A. R., FRALEY, A. E., BAMSHAD, M. J., WATKINS, W. S. & JORDE, L. B. 1996. Mitochondrial mismatch analysis is insensitive to the mutational process. *Molecular Biology and Evolution,* 13**,** 895-902.

ROUSSET, F. 2008. '007: a complete re-implementation of the GENEPOP software for windows and linux. *Molecular Ecology Resources,* 8**,** 103–106.

ROZAS, J., SÁNCHEZ-DELBARRIO, J. C., MESSEGUER, X. & ROZAS, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics,* 19**,** 2496-2497.

ROZEN, S. & SKALETSKY, H. 2000. Primer3 on the www for general users and for biologist programmers. *Methods of Molecular Biology,* 132**,** 365 - 386.

RUSSELL, J. & JOHNSON, T. The Mid-to Late Holocene History of Lake Edward, Uganda and its Implications for Hydrologic Change in the African Tropics. AGU Fall Meeting Abstracts, 2001. 04.

RUTZ, C., RYDER, T. & FLEISCHER, R. 2012. Restricted gene flow and fine-scale population structuring in tool using New Caledonian crows. *Naturwissenschaften,* 99**,** 313-320.

SALA, O. E., CHAPIN, F. S., ARMESTO, J. J., BERLOW, E., BLOOMFIELD, J., DIRZO, R., HUBER-SANWALD, E., HUENNEKE, L. F., JACKSON, R. B. & KINZIG, A. 2000. Global biodiversity scenarios for the year 2100. *Science,* 287**,** 1770-1774.

SANDON, H. 1950. An illustrated guide to the freshwater fishes of the Sudan.

SATO, A., TAKEZAKI, N., TICHY, H., FIGUEROA, F., MAYER, W. E. & KLEIN, J. 2003. Origin and speciation of haplochromine fishes in East African crater lakes investigated by the analysis of their mtDNA, Mhc genes, and SINEs. *Molecular Biology and Evolution,* 20**,** 1448-1462.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology,* 18**,** 233-234.

SEEHAUSEN, O. 2002. Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history for its cichlid species flock. *Proceedings of the Royal Society of London.Series B: Biological Sciences,* 269**,** 491-497.

SEEHAUSEN, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proc R Soc Lond B,* 273.

SELKOE, K. A. & TOONEN, R. J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters,* 9**,** 615-629.

SEPULCHRE, P., RAMSTEIN, G., FLUTEAU, F., SCHUSTER, M., TIERCELIN, J.-J. & BRUNET, M. 2006. Tectonic uplift and Eastern Africa aridification. *Science,* 313**,** 1419-1423.

SHIMIZU, M., KOSAKA, N., SHIMADA, T., NAGAHATA, T., IWASAKI, H., NAGAI, H., SHIBA, T. & EMI, M. 2002. Universal fluorescent labeling (UFL) method for automated microsatellite analysis. *DNA Research,* 9**,** 173-178.

SLATKIN, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution***,** 264-279.

SMITH, P. J., FRANCIS, R. I. C. C. & MCVEAGH, M. 1991. Loss of genetic diversity due to fishing pressure. *Fisheries Research,* 10**,** 309-316.

SMITZ, N., BERTHOULY, C., CORNÉLIS, D., HELLER, R., VAN HOOFT, P., CHARDONNET, P., CARON, A., PRINS, H., VAN VUUREN, B. J. & DE IONGH, H. 2013. Pan-African genetic structure in the African buffalo (Syncerus caffer): investigating intraspecific divergence. *PLoS One,* 8**,** e56235.

SOKAL, R. R. & WARTENBERG, D. E. 1983. A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics,* 105**,** 219-237.

STANKIEWICZ, J. & DE WIT, M. J. 2006. A proposed drainage evolution model for Central Africa—Did the Congo flow east? *Journal of African Earth Sciences,* 44**,** 75-84.

STEWART, K. 2009. Fossil Fish from the Nile River and Its Southern Basins. *In:* DUMONT, H. (ed.) *The Nile.* Springer Netherlands.

STEWART, K. M. 2001. The freshwater fish of Neogene Africa (Miocene–Pleistocene): systematics and biogeography. *Fish and Fisheries,* 2**,** 177-230.

STOTZ, W. 2000. When aquaculture restores and replaces an overfished stock: is the conservation of the species assured? The case of the scallop Argopecten purpuratus in northern Chile. *Aquaculture International,* 8**,** 237-247.

STURMBAUER, C., BARIC, S., SALZBURGER, W., RÜBER, L. & VERHEYEN, E. 2001. Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution,* 18**,** 144-154.

TAABU-MUNYAHO, A., NYAMWEYA, C. S., SITOKI, L., KAYANDA, R., EVERSON, I. & MARTEINSDÓTTIR, G. 2014. Spatial and temporal variation in the distribution and density of pelagic fish species in Lake Victoria, East Africa. *Aquatic Ecosystem Health & Management,* 17**,** 52-61.

TAKAHASHI, T. & KOBLMÜLLER, S. 2011. The adaptive radiation of cichlid fish in Lake Tanganyika: a morphological perspective. *International Journal of Evolutionary Biology,* 2011.

TAMURA, K., PETERSON, D., PETERSON, N., STECHER, G., NEI, M. & KUMAR, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution,* 28**,** 2731-2739.

THIEME, M. L., ABELL, R., BURGESS, N., LEHNER, B., DINERSTEIN, E. & OLSON, D. 2005. *Freshwater ecoregions of Africa and Madagascar: a conservation assessment*, Island Press.

TODD, E. V. 2013. *Evolutionary biogeography of Australian riverine turtles: Elseya spp. and Emydura macquarii krefftii.* PhD Thesis, James Cook University.

TRIANTAFYLLIDIS, A., KRIEG, F., COTTIN, C., ABATZOPOULOS, T. J., TRIANTAPHYLLIDIS, C. & GUYOMARD, R. 2002. Genetic structure and

phylogeography of European catfish (Silurus glanis) populations. *Molecular Ecology,* 11**,** 1039-1055.

TSADU, S., LAMAI, S., YISA, T. & IBRAHIM, S. 2014. Gonad Development and Histology in Bagrus Bayad Cultured in Outdoor Earthen Ponds.

TURNER, G. F., SEEHAUSEN, O., KNIGHT, M. E., ALLENDER, C. J. & ROBINSON, R. L. 2001. How many species of cichlid fishes are there in African lakes? *Molecular Ecology,* 10**,** 793-806.

VAN OOSTERHOUT, C., HUTCHINSON, W. F., WILLS, D. P. & SHIPLEY, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes,* 4**,** 535-538.

VERHEYEN, E., SALZBURGER, W., SNOEKS, J. & MEYER, A. 2003. Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science,* 300**,** 325-329.

VOKES, R. 2012. The politics of oil in Uganda. *African Affairs***,** ads017.

WANG, S., SHA, Z., SONSTEGARD, T., LIU, H., XU, P., SOMRIDHIVEJ, B., PEATMAN, E., KUCUKTAS, H. & LIU, Z. 2008. Quality assessment parameters for EST-derived SNPs from catfish. *BMC Genomics,* 9**,** 450.

WARD, R. D. 2006. The importance of identifying spatial population structure in restocking and stock enhancement programmes. *Fisheries Research,* 80**,** 9-18.

WEIR, B. S. & COCKERHAM, C. C. 1984. Estimating F-statistics for the analysis of population structure. *evolution***,** 1358-1370.

WILSON, K. 1990. Preparation of genomic DNA from bacteria. *In:* AUSUBEL, F. M., BRENT, R., KINGSTON, R. E., MOORE, D. D., SEIDMAN, J. G., SMITH, J. A. & STRUHL, K. (eds.) *Current protocols in molecular biology.* New York: Wiley Interscience.

WOLFENDEN, E., EBINGER, C., YIRGU, G., DEINO, A. & AYALEW, D. 2004. Evolution of the northern Main Ethiopian rift: birth of a triple junction. *Earth and Planetary Science Letters,* 224**,** 213-228.

WOOD, J. & GUTH, A. 2013. East Africa's Great Rift Valley: A Complex Rift System. *Geology. com*.

WRIGHT, S. 1943. Isolation by distance. *Genetics,* 28**,** 114.

YANG, G., XIAO, M., YU, Y. & XU, S. 2012. Genetic variation at mtDNA and microsatellite loci in Chinese longsnout catfish (Leiocassis longirostris). *Molecular Biology Reports,* 39**,** 4605-4617.

ZHANG, K. & SUN, F. 2005. Assessing the power of tag SNPs in the mapping of quantitative trait loci (QTL) with extremal and random samples. *BMC Genet,* 6**,** 51.

ZHU, Z., WANG, C., LO, L., FENG, F., LIN, G. & YUE, G. 2006. Isolation, characterization, and linkage analyses of 74 novel microsatellites in Barramundi (Lates calcarifer). *Genome,* 49**,** 969-976.

APPENDICES

Appendix 2.1: Microsatellite loci identified *in silico* within the Nile Perch 454 sequence database. *Ln01-30* microsatellites were synthesised and tested in the laboratory

Locus ID	SEQ_CODE	Motif	Length of Repeat	Motif Type	Estimated Amplicon Size	Primer Left Sequence	Primer_Left_TM (°C)	Primer Right Sequence	Primer Right TM (°C)	Primer_Left_GC_Percent	Primer_Right_GC_Percent
Ln01	IJMZ0LK03FIFLH_0	ATCT	22	Tetranucleotide	207	TGTCCGCATGTCGAAGT	60	AGGAGGAGAGGATATG	59	50	52
Ln02	IJMZ0LK03GVMSJ_0	ATCT	18	Tetranucleotide	191	ATTGGCCCTCATTCAGTA	60	AACATTAACTGTGGGCC	59	50	48
Ln03	IJMZ0LK03HDLLH_0	TAGA	16	Tetranucleotide	112	CACAGGCAGAAAACAGA	57	TGTGTCATGTGTCCATGT	60	50	50
Ln04	IJMZ0LK03FNJAG_0	GATA	12	Tetranucleotide	164	TTGTTTGAGTTGTTGCGT	61	TTGTATGCTGCTTAGAA	60	45	43
Ln05	IJMZ0LK03GA9A5_0	CTGT	11	Tetranucleotide	132	ACCCTCCACTGTGTTTCC	59	TAGGCTTGGTCTCCAGG	60	55	55
Ln06	IJMZ0LK03FOYBM_0	AGAT	10	Tetranucleotide	300	GGGGTTTCGTAGCATTA	60	TGCGTTTACAACACGAG	60	48	50
Ln07	IJMZ0LK03GKC19_0	TCTA	10	Tetranucleotide	251	TGGCTTTGCTTCTTACTT	59	ATAGGGCTTCCTAAGAC	60	43	55
Ln08	IJMZ0LK03HAMAQ_0	TAA	19	Trinucleotide	119	GGACTGCAAGCAGATTA	60	TCATGCACCTCTTCTGCA	60	55	50
Ln09	IJMZ0LK03GJMYA_0	TGA	15	Trinucleotide	243	CCCACAATACTATGAGG	59	TCCAAATTTTGTCTTAAA	60	52	35
Ln10	IJMZ0LK03FYZYZ_0	ATA	15	Trinucleotide	234	ACGGTATACAACAGCAG	60	AAGCGGTGACCTCCATA	59	55	55
Ln11	IJMZ0LK03HAZCP_0	TTC	13	Trinucleotide	197	CTCACACGTCCTGTTTGC	61	TCAGAGGAAGTCGTGCA	60	55	55
Ln12	IJMZ0LK03F1WBX_0	AC	21	Dinucleotide	135	TGGGTGAAAGAGATGGG	61	CATTATGCATGCCTCGTT	60	50	45
Ln13	IJMZ0LK03GN8BZ_0	CA	21	Dinucleotide	103	ATGTGACCTCGCAACTT	59	TAAATACCCTCGCGACT	58	50	50
Ln14	IJMZ0LK03HBXO1_0	AC	19	Dinucleotide	147	CTGGCTCAGGCCTTATGT	60	ACACCTTCTCCTCCTCCC	60	55	60
Ln15	IJMZ0LK03GWPZD_0	GT	18	Dinucleotide	254	GTTGTGACGATATTGCGT	60	CACGGTCTGTTGCATTAT	60	50	50
Ln16	IJMZ0LK03GNGBJ_0	ATGG	10	Tetranucleotide	146	CAATGTCAGCTGGGATA	61	CCGGTTCTTGTCAATGTC	59	55	50
Ln17	IJMZ0LK03F6M9A_0	CAGA	10	Tetranucleotide	106	AGCAGCATCAGTTCAAC	59	CCTGGTCATTTTCCACAC	59	50	50
Ln18	IJMZ0LK03GMF0Y_0	GGAT	8	Tetranucleotide	100	AGCTGTACATGCCACCA	60	CCCACGTCCACACCTCT	59	58	63
Ln19	IJMZ0LK03GAC55_0	AGG	11	Trinucleotide	136	TGTCTGTGTTCAGCCCTC	60	TGATCCATTAGCCAGCA	60	55	50
Ln20	IJMZ0LK03G8HCN_0	TAT	11	Trinucleotide	134	TCTGTCTCTGCATCACCT	60	CCAGGTAAACCCCATGA	58	55	48
Ln21	IJMZ0LK03FYL93_0	TAA	11	Trinucleotide	127	TGGGATCATCTCATCTCA	61	ACATATGCAGCACCCCA	61	48	55
Ln22	IJMZ0LK03HBON4_0	GCA	11	Trinucleotide	105	AAGATGCTGCATGTCAG	60	GCGGATGTAAAGAGGAG	60	50	55
Ln23	IJMZ0LK03GM41G_0	TCA	10	Trinucleotide	225	CACAAGTGTAACCAGTG	60	TCTATTAACGGGAGCGG	60	55	50
Ln24	IJMZ0LK03G8Z96_0	ATT	10	Trinucleotide	188	GAAAGCTGCAGCATATT	61	TTGCTACTGAGTGAGCT	59	50	55
Ln25	IJMZ0LK03FWC78_0	CTG	10	Trinucleotide	171	AGGGTACCCACTGCTAC	59	AAACCAGGACTGCACCA	60	60	50
Ln26	IJMZ0LK03HFWHI_0	GT	18	Dinucleotide	163	TCTCCTCAGAGCGTGAT	60	CCTCAAATAGTGGACGT	60	55	46
Ln27	IJMZ0LK03GJ7EU_0	AC	18	Dinucleotide	110	CCGCATCAATTTAAAAC	58	TGTTTTGGTTGATGGTTG	58	40	40
Ln28	IJMZ0LK03FP57W_0	CA	18	Dinucleotide	108	TGATAATCAGCGTGATA	60	GTTGGTACGTGGGTGGG	60	39	60
Ln29	IJMZ0LK03GGYQB_0	AC	17	Dinucleotide	179	TGCCAGGAGCTCAGTAA	59	TGTCAAGTAGGCATGTG	58	55	50
Ln30	IJMZ0LK03GN190_0	TG	16	Dinucleotide	124	ACTTCATGCATGTGCAG	60	GCATCGTCTTCTTCCTCT	60	50	52
Ln31	IJMZ0LK03F8MTE_0	AC	15	Dinucleotide	219	GGTCACAGGTCACTGCT	59	CACTGAGCAAAGACGAA	60	55	50

Ln32	cons387_5_A_0	AC	21	Dinucleotide	201	ATACCAGGCGATCAATG	60	GTTCAGCCTGGTGATCA	59	50	55
Ln33	cons594_2_A_0	CTAT	21	Tetranucleotide	208	AGATCACATGCATTACA	60	ATCCTCCACTCACAATG	59	45	50
Ln34	IJMZ0LK03FTJND_0	AC	18	Dinucleotide	106	CTATGGCTCGCCTGGAT	60	CTATTAAAGTAGAGGTA	57	61	40
Ln35	IJMZ0LK03HAHH3_0	CA	18	Dinucleotide	115	TGACTGCTTCTGTCTGAG	59	CAAGTAAGACAATAGAG	60	52	37
Ln36	cons186_2_A_0	TG	17	Dinucleotide	129	CTAATTAGCTGCCCTGCT	60	CACATACGTATACTGAG	58	55	50
Ln37	cons449_2_A_0	AC	17	Dinucleotide	176	GGTGGATGGGTTGTGTTT	60	CCCAGATTATCGACCCA	60	50	55
Ln38	cons545_2_A_0	CA	17	Dinucleotide	104	GCAACCTTCATCAGGAC	61	CAAGATGAGTAAGGGGA	59	55	50
Ln39	cons625_2_A_0	CA	17	Dinucleotide	110	CACCTGGACACAATACG	60	GTTTGTTGTGTGGTGGAT	60	55	50
Ln40	IJMZ0LK03F75PQ_0	AC	17	Dinucleotide	133	ATGAGATATCTGCCGAT	61	ACATACCACTCTGTACT	57	50	41
Ln41	IJMZ0LK03GWI9Z_0	AC	17	Dinucleotide	103	TGTCTGTTTACATGCCAT	60	CCACAGTAGATCTTTGG	60	43	50
Ln42	cons332_2_A_0	TG	16	Dinucleotide	157	AACGCAAATGGCTAAAT	60	CGTGTGCTCACTTGCACT	59	40	58
Ln43	IJMZ0LK03FT3ON_0	AC	16	Dinucleotide	124	GAGCAGAGCAGTGTAGC	60	TGTCCCTCCTGCCTGTAT	60	60	55
Ln44	IJMZ0LK03GH40F_0	TG	16	Dinucleotide	123	TTTTGATTGCAGAATGTG	60	CATCAGTCTGTTCAGTGT	60	40	37
Ln45	cons111_3_A_0	CA	15	Dinucleotide	112	TCATGACAGCCTCCAAA	60	GTTGTGTCATGAACCTG	58	50	48
Ln46	cons618_2_A_0	AC	15	Dinucleotide	194	CCTCGGCTGTCCTGACTA	60	GGCTGAATCTGTGGGAA	60	60	50
Ln47	IJMZ0LK03F3Z6P_0	AC	15	Dinucleotide	182	ATGTGAGCCTGGCACTT	60	ATAAACAGCGTGTGAGT	60	55	50
Ln48	IJMZ0LK03F45ET_0	CA	15	Dinucleotide	109	GAGCAAGATCTATCAAA	58	GCCATCCAAGACCATCT	60	33	55
Ln49	IJMZ0LK03F7IKM_0	AC	15	Dinucleotide	100	GCAAGCTCAAGAGCCAG	60	GGACGATGACCTCTGAC	60	61	60
Ln50	IJMZ0LK03FJU2U_0	TG	15	Dinucleotide	125	CTTGACTGATGCAGGCA	60	TAAATACTCAGCAGGCG	59	50	50
Ln51	IJMZ0LK03FPLIL_0	CA	15	Dinucleotide	100	ATTTACACTCGAGCTTGG	59	AAACAGCAGGAGTGGA	60	50	55
Ln52	IJMZ0LK03FVA3C_0	CA	15	Dinucleotide	126	GGCACCACGGTCAGTCA	61	TGACATTTTGTCAGACT	58	67	38
Ln53	IJMZ0LK03FVZBI_0	TG	15	Dinucleotide	153	GAAACGAAGACTGGACC	60	CTGTACATTTCCCCACG	61	48	55
Ln54	IJMZ0LK03FW81D_0	GT	15	Dinucleotide	100	AGATAGGTCACCAACAC	60	CGGGTATTGGTAAGAGG	59	52	55
Ln55	IJMZ0LK03G06Z5_0	TG	15	Dinucleotide	175	ACTTTACCTGTGCTCGCT	60	ACTACTTCTCCTCCCCAC	59	55	60
Ln56	IJMZ0LK03GDEPC_0	AC	15	Dinucleotide	172	ATTTCCGGTGCAAACATT	60	TTAACTGCAAATTACCC	59	40	43
Ln57	IJMZ0LK03GK5Q5_0	AC	15	Dinucleotide	122	TCCTTTCCTCTCGCACTT	60	TGCCTCTCCTAGCCTCCT	59	50	60
Ln58	IJMZ0LK03HBHG7_0	AC	15	Dinucleotide	102	AGCTTCAGGGCCTGTCTT	61	TAGTGCCAGCTCTGTCC	59	58	55
Ln59	cons202_2_A_0	AC	14	Dinucleotide	163	TGTCAGCTGGCAGATTT	59	CTCACGTGGAGTCTGAG	60	48	60
Ln60	cons579_2_A_0	GA	14	Dinucleotide	109	TGTACTTTACTGCACCGA	59	TCAGTGGGCTGCACTAT	60	43	55
Ln61	IJMZ0LK03F3T5E_0	AC	14	Dinucleotide	147	GGCAACAAGAAATGGTT	60	AATAGGCGTTAAAGGGT	59	45	48
Ln62	IJMZ0LK03F6B1B_0	AC	14	Dinucleotide	174	GTCACCAATCAACCAGG	60	AGCCAGAGGAGGTCAGC	61	55	63
Ln63	IJMZ0LK03FJYHR_0	AC	14	Dinucleotide	222	ATTCAATATTTCATCTCC	57	GCTGCATGACAAGCTCA	60	32	55
Ln64	IJMZ0LK03FKWD6_0	GT	14	Dinucleotide	123	GGCTATGGGTTTAGGTA	61	TACACTGCCCCACTGTA	61	55	60
Ln65	IJMZ0LK03FLROM_0	AT	14	Dinucleotide	188	TGATATCAATGACAGCA	60	GACAATGTTGCCTTCAG	59	43	50
Ln66	IJMZ0LK03G7ML7_0	AC	14	Dinucleotide	182	GATGCTGACATCAGGGT	59	TACATCCACCCTACCAC	59	55	55
Ln67	IJMZ0LK03GAVNU_0	GT	14	Dinucleotide	134	TGATTACACAACTAAAA	59	AGAGGAAGACCTGCACA	62	39	60
Ln68	IJMZ0LK03GHOGM_0	AC	14	Dinucleotide	165	TTTCATTACGGGAAACCT	60	TGTCTCCGTCCACATTAA	60	45	48
Ln69	IJMZ0LK03GM4CC_0	GT	14	Dinucleotide	145	CTCAGCTCTCCGTGTGCT	60	ATAGCTATGCACGGTCC	60	63	55
Ln70	IJMZ0LK03GP2UP_0	AC	14	Dinucleotide	190	ATAAACACGGACAGGAA	60	GGGAGGGAGATGTGTGT	60	50	55
Ln71	IJMZ0LK03GTAG6_0	AC	14	Dinucleotide	115	ACATGAGAGGATGAGGG	60	GTCACCACATTGACCTG	60	55	55
Ln72	IJMZ0LK03GUQ3V_0	GT	14	Dinucleotide	313	GTCTGTGACATGAACAC	60	TCAGTTACAGCAGCCTG	60	55	55
Ln73	cons211_2_A_0	CA	13	Dinucleotide	154	CCTGAGCTACGGTCTTCC	60	ACTCAGCCCTGCACACT	61	60	55
Ln74	cons278_2_A_0	AC	13	Dinucleotide	172	ATTTCCCCTGACCTCCAT	60	CGAAGAAACAAGTATAT	58	50	37

Ln75	cons314_3_A_0	AC	13	Dinucleotide	109	CTAGCTGAAGAGCGTGT	59	GCGAGCAGAGACAAGA	60	55	60
Ln76	cons428_2_A_0	TG	13	Dinucleotide	141	AATGCGTTCAGGCAACT	60	TCAGTTCCTGTGAGCTCT	59	45	55
Ln77	cons83_3_A_0	GT	13	Dinucleotide	143	TTTGGTGGGAATGTTTAA	60	TACTGTGAAAGGGGACG	60	35	55
Ln78	IJMZ0LK03F0QU2_0	CA	13	Dinucleotide	137	GCAGAAACACAGTGCGA	60	TCCCTCTAGGCTAATGCT	59	55	52
Ln79	IJMZ0LK03F1MEA_0	GT	13	Dinucleotide	195	GTGTCATTGTTCGTGGCC	62	GCCGAGACAGTCAAGGG	60	58	60
Ln80	IJMZ0LK03F58OQ_0	CA	13	Dinucleotide	100	GGTGATTCAAATTAAGA	59	CCTAGTGCGTTGTCAGA	61	39	60
Ln81	IJMZ0LK03FLBXN_0	GT	13	Dinucleotide	175	TGGGGCAGCTAATCTAA	60	CCCTGCAGGCTCAGATA	60	50	55
Ln82	IJMZ0LK03FRVHD_0	GATA	13	Tetranucleotide	239	GTTCTTGGGAGCATGTA	60	TGCTATTAAGGAAATTA	57	55	32
Ln83	IJMZ0LK03FT13W_0	AC	13	Dinucleotide	110	TCTTTCACTCTTGGGACT	59	TTAACAGGGGTCGAGCA	60	48	55
Ln84	IJMZ0LK03FVUYZ_0	TG	13	Dinucleotide	138	GTTTCCTGACCCAGCTCA	60	TCAGCCTCACTTCACAG	58	55	48
Ln85	IJMZ0LK03FW36T_0	CA	13	Dinucleotide	105	TTCATTTATCAGGCCCCA	60	ACCCCTCCAATTCTGAG	60	45	55
Ln86	IJMZ0LK03G635Q_0	CA	13	Dinucleotide	113	GCACCAAAACCCAGAAA	59	TAGGAGTGAGGTTTGGC	60	43	55
Ln87	IJMZ0LK03G6S44_0	AC	13	Dinucleotide	102	GACACCTGTTCTTGTTCT	59	CAGCAGCAACTTTTGTT	61	50	45
Ln88	IJMZ0LK03G70QF_0	CA	13	Dinucleotide	215	TGGTCTTTGGAGTGTCAG	60	TAACCGCTCTCCTCCTGA	60	52	55
Ln89	IJMZ0LK03G7NF6_0	AC	13	Dinucleotide	175	CCTTTCAAATGTGGTGA	58	ATCTGATCTGGCAACCC	60	39	50
Ln90	IJMZ0LK03GCBKP_0	AC	13	Dinucleotide	186	TGACCTTCTTCGCTCCAC	61	GAGGTGGTAACGACATC	59	55	52
Ln91	IJMZ0LK03GDFZ7_0	TA	13	Dinucleotide	140	TTTCAATATTTGCCATGT	58	CGAGCAGTGAAACCAAA	59	32	41
Ln92	IJMZ0LK03GFN5V_0	AC	13	Dinucleotide	104	GCCTGCCATCACACTTA	60	TTCGTCCTCGTTATTGTG	61	48	50
Ln93	IJMZ0LK03GGOGW_0	CA	13	Dinucleotide	166	TCACTGACCTTTTGTCGC	60	CGCATTTGTTATCCATAA	59	50	43
Ln94	IJMZ0LK03GH8Q8_0	CA	13	Dinucleotide	146	TGGGAACCAACTGGTGA	60	ACATCCTGAGCGTAAAA	60	52	50
Ln95	IJMZ0LK03GIDES_0	AC	13	Dinucleotide	111	AATCCCTCCCATCTGAAC	60	ATCAGTGACACCAGCTT	59	53	50
Ln96	IJMZ0LK03HCO3U_0	GT	13	Dinucleotide	141	GGCAGTGGAACTGCTCT	60	CAGCATGGTCTGTCTTC	59	55	48
Ln97	concat9_A_0	AC	12	Dinucleotide	205	AACCTGTTACCCGAAGG	61	TCAGAGTGCAGATACGT	60	55	55
Ln98	cons197_2_A_0	TG	12	Dinucleotide	100	TTCTTCTTTTGTTCCCAC	60	GCGGTCATGCACACTAG	59	45	61
Ln99	cons214_2_A_0	CA	12	Dinucleotide	143	TGCCTTTCCCACAGTTAA	60	GCTGCTTTTCTTTCCTAA	59	50	38
Ln100	cons333_3_A_0	TG	12	Dinucleotide	110	CATGTATAGTTTATCAGT	60	CCTTGATATCTTCAACA	58	41	39
Ln101	cons564_2_A_0	AT	12	Dinucleotide	134	TGAAAAGTGCCCTGAGA	60	TGGCCTCTGTTATAGTA	60	48	48
Ln102	cons89_4_A_0	AC	12	Dinucleotide	148	CCACAGAGCAGTTCTGA	61	TATTGTCCACGCAAGCC	59	60	50
Ln103	IJMZ0LK03F0JLP_0	GT	12	Dinucleotide	101	TCCTTGGTATCGGTAGGC	60	GATTCATGTTGAGCCCA	60	55	53
Ln104	IJMZ0LK03F8F08_0	AC	12	Dinucleotide	106	TTTCTCTGCATTTCACCA	60	TCAGTAAAGTCAAAGTA	58	45	43
Ln105	IJMZ0LK03FNAT2_0	AT	12	Dinucleotide	146	AAAGCTTTTGTGCAAGTT	57	TGCAGATTTCTAAGCAC	60	33	45
Ln106	IJMZ0LK03FR44J_0	GT	12	Dinucleotide	128	AGAGCTCTGGCGTGATA	59	ACCAAGAGCCTGTCTCC	59	52	55
Ln107	IJMZ0LK03FRCFN_0	AC	12	Dinucleotide	161	CTGACTCGACCCGTTAG	60	CAGATATGGAAGGTGGT	60	60	55
Ln108	IJMZ0LK03FVKXV_0	AG	12	Dinucleotide	100	CCTTTCTCAGGGTGTGAG	60	AGTACTGATCTGCTGCT	60	55	55
Ln109	IJMZ0LK03FVQN8_0	CA	12	Dinucleotide	229	TGCATCCCATAATTTCAC	61	CAGTGAAACAGCTTTGC	60	45	50
Ln110	IJMZ0LK03FXARW_0	TG	12	Dinucleotide	176	GGTTGCTGCTGTTGTGTA	60	TGAAAGGAGGTCATGTT	61	55	50
Ln111	IJMZ0LK03FXXWF_0	CA	12	Dinucleotide	114	TAGATGAGCTGCAGGTG	60	CCCTCGTGTCTTCTCCTC	60	55	60
Ln112	IJMZ0LK03G1W5O_0	AC	12	Dinucleotide	117	CCAAGGAGGTCAGGTAG	59	ATAGCGTCCCTCGACTC	59	60	55
Ln113	IJMZ0LK03GH148_0	CA	12	Dinucleotide	250	CTCATTGTCACAGACGCT	60	GAGAAAGAATGGGAGG	60	55	55
Ln114	IJMZ0LK03GKL8S_0	CA	12	Dinucleotide	159	TGAGCATGTCCTCAGCA	60	TGTGGTGACTGTCTTCA	60	50	52
Ln115	IJMZ0LK03GO2SE_0	AC	12	Dinucleotide	136	ATGTGAACCTACGGTCT	60	CAGCAAGGTAAATGGTG	61	55	42
Ln116	IJMZ0LK03GON8X_0	GT	12	Dinucleotide	115	TTCATCCAGAGCAGAGA	59	CTGGCTGCAGACAAACA	62	50	55
Ln117	IJMZ0LK03GPI56_0	TG	12	Dinucleotide	136	TTATTGTTCGCTGTCGGT	60	TGAGTGAGAGCAAAGAA	59	43	50

Ln118	IJMZ0LK03GQJHO_0	AC	12	Dinucleotide	103	GGCCACTCATTCATCATT	60	TGGATCTGTGGACTGGT	60	50	50
Ln119	IJMZ0LK03HB1S4_0	GT	12	Dinucleotide	105	TTTGCACCAGAAACAAA	60	AGGTACCTGATTGGACG	60	40	55
Ln120	cons21_2_A_0	AC	11	Dinucleotide	133	TCATGGGTTTTGCTTTCT	60	GCTTGTCATCAGCAGAC	60	45	55
Ln121	cons266_2_A_0	AC	11	Dinucleotide	123	GGTCCCAAGCTGCATAG	60	CCATTGTCCAGTCATTTG	60	55	45
Ln122	cons313_2_A_0	CA	11	Dinucleotide	102	AAGCCTCTTGAGGGAAG	60	TGAGAGTGTTGGTCTGC	60	48	55
Ln123	cons395_2_A_0	CA	11	Dinucleotide	164	CAAGGGAAACAACAGTA	57	AGTCTGTCTCCCAAACA	60	43	55
Ln124	cons695_2_A_0	GT	11	Dinucleotide	145	ACAGAGACTCCCCAGAC	59	CTGGCATCAGAACACCA	60	60	55
Ln125	cons82_3_A_0	TGC	11	Dinucleotide	142	AGCTGTTGTTGCTGCTGT	60	CGTTCCAGTTCAAATGA	60	50	50
Ln126	IJMZ0LK03F03LL_0	GT	11	Dinucleotide	152	CAAAGACAGGGGAGGA	61	ACCTCACTACTGTCCCGT	59	60	60
Ln127	IJMZ0LK03F1DA1_0	TG	11	Dinucleotide	121	TTGTGTATCTCTTTCTCA	59	GATAGTGAGCGCACAAC	59	35	55
Ln128	IJMZ0LK03F24U3_0	CA	11	Dinucleotide	121	AATGATGTAATGTTAAC	58	CAAAACTGTTTGCACAG	59	32	41
Ln129	IJMZ0LK03F6TDJ_0	TG	11	Dinucleotide	115	CGTCTGCCCTGTGTAGTG	59	GAAACATCAGTGCAAGG	60	60	48
Ln130	IJMZ0LK03F6YCN_0	TG	11	Dinucleotide	231	AAATGCATGTGTTGCAG	60	TAGCTTCTTCTCTGGCTG	60	45	55
Ln131	IJMZ0LK03F7LJB_0	AC	11	Dinucleotide	126	ATACACAAGCACAGGCC	60	TTCTGTACTTCAGCTCCT	59	55	52
Ln132	IJMZ0LK03FGIEX_0	TG	11	Dinucleotide	106	ACGAGAAAGGCAATCCA	60	TGTTGTAACAGTGAGGA	59	50	35
Ln133	IJMZ0LK03FH6CK_0	CA	11	Dinucleotide	128	GACGCGTGGATTTATCT	59	GTTAAAATAGTCAGCGG	60	50	50
Ln134	IJMZ0LK03FRJBT_0	AC	11	Dinucleotide	127	ATTAGCGATGACCTCGTT	60	GGCTGTGTTATATCAGA	57	50	43
Ln135	IJMZ0LK03FRTSH_0	TG	11	Dinucleotide	183	TCTGCGTGCTGACAGAG	60	AAGGACTTTCAGGTTTG	59	55	48
Ln136	IJMZ0LK03FT0MM_0	GT	11	Dinucleotide	140	AGCGTTGCTATGGTGTTG	60	TTCTTTCTCATACGAACA	57	53	30
Ln137	IJMZ0LK03FVAMD_0	TA	11	Dinucleotide	201	TCACCAACATGTAATGC	59	ATGTCTGCAGGCTGTCT	59	41	55
Ln138	IJMZ0LK03FX1E6_0	TG	11	Dinucleotide	131	CACAGGTGAGTTATTTA	59	ACCACCTGTGTCAAACT	60	44	55
Ln139	IJMZ0LK03FZGPU_0	GT	11	Dinucleotide	101	CACAGGAAGCAGCGATG	62	AGGAGTCTGGTCTGACG	60	58	60
Ln140	IJMZ0LK03G1R05_0	AC	11	Dinucleotide	130	CCAAGGCCCCTAGTTAA	59	TGGGTTTAGGTGCAAGT	59	52	48
Ln141	IJMZ0LK03G28GD_0	CT	11	Dinucleotide	127	TGCCTTGATTACTTTCCA	60	GAGAAGGCGTCGAGGA	60	43	63
Ln142	IJMZ0LK03G5B4H_0	CA	11	Dinucleotide	172	AAGCAAACCAAAACCAC	60	AGGAAACTGGTGGCTGA	59	45	55
Ln143	IJMZ0LK03G82C7_0	CA	11	Dinucleotide	186	ACCATCCCACAGCTGTTC	60	CCGCGTCCTTATTCATTT	60	55	45
Ln144	IJMZ0LK03GA1UF_0	AC	11	Dinucleotide	200	TCCTAATGACCTCTGGTG	60	CAGCTATCCCAAATCTT	60	55	50
Ln145	IJMZ0LK03GA3J0_0	AC	11	Dinucleotide	118	TCAATTGCGCGTTAAGT	60	CGCATTGTGTCAATGAG	60	45	48
Ln146	IJMZ0LK03GCNI1_0	GT	11	Dinucleotide	174	CGCTTTGGAGGTTAAAG	60	AACCTAACGAAGACCCG	60	50	55
Ln147	IJMZ0LK03GH8S8_0	AC	11	Dinucleotide	102	AGGCTTTTGAAAGGGTC	61	GCATGTGTTGCTACTGG	61	50	45
Ln148	IJMZ0LK03GJ2GV_0	TG	11	Dinucleotide	104	CAATAGGTATCATCCCTC	61	TAGAGCATCGCACTCAT	60	52	50
Ln149	IJMZ0LK03GR9GK_0	GT	11	Dinucleotide	112	CTTGGTATTCTCCCACCA	59	CCAGAGCCCTGACGTAA	60	55	55
Ln150	IJMZ0LK03GU7JI_0	AC	11	Dinucleotide	103	ACTGAGCCAAAACCCAA	60	TTGTCTTACTGGAAGCT	58	50	48
Ln151	IJMZ0LK03GZ2S6_0	GT	11	Dinucleotide	118	TGGAAAGCTGTACATCA	60	GATGAACTTTTCCTCAG	60	50	50
Ln152	cons10_2_A_0	GT	10	Dinucleotide	133	TGCTTCAGTCAGCACCA	60	GCTCAACCCACTGTCTTT	59	50	52
Ln153	cons126_2_A_0	AC	10	Dinucleotide	128	CGCAAAGCTACCCGTAA	60	TGACTCGCTGCTCTGTA	59	50	55
Ln154	cons149_2_A_0	AC	10	Dinucleotide	200	TGTCTGAATGGAGGCAG	60	TGATTTACTTACTCCCCA	58	50	38
Ln155	cons17_2_A_0	GT	10	Dinucleotide	102	GAACTTAAATTGCCCTCT	60	GGAAAGTAGGTGCGGAA	59	45	50
Ln156	cons207_2_A_0	CA	10	Dinucleotide	109	TTGCATGCTCTTTGACTT	60	AGGACCTTGACAGATGA	59	45	52
Ln157	cons219_3_A_0	AC	10	Dinucleotide	176	ACACGCTTCAAAATTCA	60	GAGACTGTCAAACATAC	57	45	35
Ln158	cons268_3_A_0	TG	10	Dinucleotide	162	TGAATCACTAATGGGCC	60	TGAGGAAAAAGAAGCA	58	50	45
Ln159	cons295_2_A_0	TG	10	Dinucleotide	121	AAGCTCCAAACACTCCA	58	TGATGATGGCTGTTGGC	60	43	50
Ln160	cons37_2_A_0	AC	10	Dinucleotide	123	TGATGACGGGAACAAAG	59	GGAACTGAACGGCAGTC	60	45	55

Ln161	cons474_2_A_0	ACA	10	Dinucleotide	292	AGCGTCTGCAGTGTATC	59	ACGGACTGTGCTGTGAT	60	55	55
Ln162	cons5_2_A_0	AC	10	Dinucleotide	108	GTGCAAGGAGTTTGAAA	60	CAACAGGTCACAACCCA	60	50	55
Ln163	cons588_2_A_0	AC	10	Dinucleotide	191	CATCCACACTCACATAC	59	GTGCAATGTTATCGAAA	59	48	45
Ln164	cons632_2_A_0	CA	10	Dinucleotide	198	CATTCAACCAAAATCCT	60	TATCACCACCAAACACT	59	45	50
Ln165	IJMZ0LK03F269E_0	AC	10	Dinucleotide	114	CTGGAACTGCTTACCTCA	60	CAATGCTTATAAATCCA	58	52	39
Ln166	IJMZ0LK03F7BC2_0	CA	10	Dinucleotide	100	AGGTTTTCAGGTCAGGA	59	ATACTACATTTCCGCAC	60	50	50
Ln167	IJMZ0LK03F7CHQ_0	TG	10	Dinucleotide	193	TTGTCTGCCAAACAGTC	60	GCCGTCTAATCTCTGCC	60	50	55
Ln168	IJMZ0LK03F9VAS_0	TG	10	Dinucleotide	103	ATCTTTTGGATGATTGCG	58	CACTCACGCTCACTTCA	59	40	55
Ln169	IJMZ0LK03FGPAW_0	AC	10	Dinucleotide	239	TCTGAAGTATGGATCAG	58	CAAAAGGTCAGGAGGTC	60	43	55
Ln170	IJMZ0LK03FJ2PX_0	TG	10	Dinucleotide	186	TGTCCTCTTCCCACTTCT	59	TGTCCTCCAGGATTAGG	60	55	55
Ln171	IJMZ0LK03FJ729_0	AC	10	Dinucleotide	104	GCAACAAACCTGAGGGT	60	ACCAAGATTACGGCCAT	60	55	45
Ln172	IJMZ0LK03FJJJN_0	CA	10	Dinucleotide	234	TATTGGTGCTGCTGTGAA	60	GGCTGCTCTTTTCACCTG	60	50	55
Ln173	IJMZ0LK03FK6UW_0	AC	10	Dinucleotide	128	ACCCTTCGAGGAGAGGT	61	TTGCAGGATTTCTAATC	59	60	33
Ln174	IJMZ0LK03FM0P9_0	TG	10	Dinucleotide	101	CCCTGTCCTGCTTTGATC	60	CTTTGACAGACACACTC	60	55	48
Ln175	IJMZ0LK03FM1K9_0	GT	10	Dinucleotide	100	CACATGCGCAGATAACT	60	AACACGCTACTCTTCTTC	59	43	48
Ln176	IJMZ0LK03FOY8T_0	TG	10	Dinucleotide	177	CTCGTCCCTCGATCTGTC	60	CTGTCTGTCTGACTGCCT	60	60	60
Ln177	IJMZ0LK03FRECC_0	TC	10	Dinucleotide	171	TGCTGCTATAAATAACCT	61	GCGAGCTCATCAAACTT	60	38	50
Ln178	IJMZ0LK03FRURN_0	GT	10	Dinucleotide	100	TCGGTCCCCACTACCATT	62	TGAGAGGAAGTTGTCGG	60	55	55
Ln179	IJMZ0LK03FUHJZ_0	CT	10	Dinucleotide	169	TAATTCCCAGATTTCACC	60	AAACCATCGTGGCGTTT	60	45	45
Ln180	IJMZ0LK03FW16A_0	CA	10	Dinucleotide	167	GATCTATCACTCTCCGAG	60	AAGCATTTTGGCTAACC	60	57	47
Ln181	IJMZ0LK03FYXZL_0	AC	10	Dinucleotide	111	AGAGATGGCAAGATGGG	60	CCTGCTTTTCCTCTTTTT	59	50	43
Ln182	IJMZ0LK03G3Y26_0	AC	10	Dinucleotide	100	TTAAAACGGAAATCCCC	61	CAGAAGAGTGGCTGCTG	61	45	60
Ln183	IJMZ0LK03G5EBQ_0	TG	10	Dinucleotide	102	GTTGCACTTGAAATTGA	58	TTCTTAGTACATAATGTG	58	45	38
Ln184	IJMZ0LK03G79VT_0	AT	10	Dinucleotide	107	AGGGTGTTGAAAGCAAG	59	GAACCATATTCCACATA	59	48	38
Ln185	IJMZ0LK03GB1J3_0	GT	10	Dinucleotide	116	TGTATTGTCATGTTGACT	58	CTTGGGAGACCTCACAA	60	41	55
Ln186	IJMZ0LK03GBWST_0	CA	10	Dinucleotide	145	ACAACACACAGACACAG	59	TTACTCATGCCTAATCG	59	48	48
Ln187	IJMZ0LK03GCPXM_0	AC	10	Dinucleotide	135	CGCTGGAAAAGTTGTCA	59	TTTAATGAGGAGACGCA	59	45	41
Ln188	IJMZ0LK03GDUVB_0	AC	10	Dinucleotide	107	CTCTGCATGCAAGTTCCC	61	CATCAGCTGTACGTCAA	58	55	50
Ln189	IJMZ0LK03GG4HV_0	TG	10	Dinucleotide	124	GGTTAAGGTTAGATTAA	59	CAATGTAATTAGTCCCA	60	38	43
Ln190	IJMZ0LK03GG6VU_0	GT	10	Dinucleotide	136	GATGCTCTGGTGCAGTTT	59	AGCATCACTTTGTGCTCC	59	50	50
Ln191	IJMZ0LK03GQWSX_0	CA	10	Dinucleotide	100	TTACGCACATGCTGAAC	60	CATCCATCTCCCAGCAC	60	48	58
Ln192	IJMZ0LK03GZ9JX_0	GA	10	Dinucleotide	108	TGTAGGTTGCATAGAGG	60	TGGACTGCTACGTTCAC	60	46	55
Ln193	cons634_2_A_0	TG	9	Dinucleotide	135	CTGCACTCCTCAGTTCAG	61	CTGCTGTGCTCACACAG	60	60	57
Ln194	IJMZ0LK03F9XCD_0	GT	9	Dinucleotide	133	TCGACAGCAAGACACAG	60	TTCTCAAACCCCTACCTC	60	55	55
Ln195	IJMZ0LK03FR8N6_0	AC	9	Dinucleotide	138	AGCTATCTGCAGAGCAC	58	CTGAGGTCTGGGTCGAA	61	55	60
Ln196	IJMZ0LK03G8MH3_0	GT	9	Dinucleotide	138	GATGGTATCTCATCCCCG	59	TCCAGATGTTGCGAGAA	60	55	50
Ln197	IJMZ0LK03G94VJ_0	AC	9	Dinucleotide	107	TCCACAAAAGAGAAACC	58	TCTGAGCTATTCCCTGTG	60	45	55
Ln198	IJMZ0LK03GU2RW_0	TG	9	Dinucleotide	147	TGTGGTCAGAGCTGTCCT	60	AGCCAGCTGTGTTGTTTT	59	55	45
Ln199	IJMZ0LK03GXPKO_0	CA	9	Dinucleotide	136	GAGCTCCCAACTAACCT	60	GCTGCTGTTTCAGCATGT	60	55	50
Ln200	IJMZ0LK03GYZRF_0	GA	9	Dinucleotide	250	AGGCCCTCACTGAACAA	59	CAGTACCTGGAGCACAT	59	55	55
Ln201	cons444_2_A_0	AC	8	Dinucleotide	104	AAAATGCAAAGCCTCAT	59	TCCTCAACTGTAGGAAG	59	38	52
Ln202	IJMZ0LK03F2DUJ_0	GT	8	Dinucleotide	113	CGCTTTTGCTTTACATTA	58	ATTGGGCACTTGTTTGC	60	35	47
Ln203	IJMZ0LK03FPUOX_0	TG	8	Dinucleotide	115	TCATCAGTTTATGTGTCT	59	GCCTGCTCCATTTCCAGC	62	33	61

Ln204	IJMZ0LK03FSRSF_0	GA	8	Dinucleotide	130	CGTGTGACTCTGAAACA	59	TGACGTCTTCATCGCACT	60	55	50
Ln205	IJMZ0LK03GHYI3_0	CA	8	Dinucleotide	154	ATTCATTGTGGTTGCTGT	59	ACGACAATGGACCTGAG	60	45	55
Ln206	IJMZ0LK03GQTP9_0	GT	8	Dinucleotide	128	TCACACCATGCTAACAG	60	TGTTGTGTAGTCGGTGG	60	48	55
Ln207	IJMZ0LK03HBEKM_0	AC	8	Dinucleotide	140	GTCATCAGGAAGAGGGA	60	GCATCACAGGAATCATG	60	55	50
Ln208	cons335_2_A_0	AC	7	Dinucleotide	139	AGTGACCACTCCTCCCA	60	TTCTCACTGCTTTCGTCT	59	60	50
Ln209	IJMZ0LK03F5YPR_0	GT	7	Dinucleotide	100	AGCCAATTTCACTGAGC	60	TTGACAGGAACTGCTTC	59	50	48
Ln210	IJMZ0LK03G3R1B_0	TC	7	Dinucleotide	167	ATGTGAGCGCTGTCAGA	60	ATCTTGCCATTCAATTTG	60	58	40
Ln211	IJMZ0LK03GBFTT_0	TC	7	Dinucleotide	116	CCGTATGTCACTCTCACC	60	TACCAGAGCTCTCCGAC	61	55	60
Ln212	IJMZ0LK03GKMV1_0	AC	7	Dinucleotide	134	TAGTGTCGCGTGTTCTCA	60	TCAACAGGTGTGAGGGA	60	55	58
Ln213	IJMZ0LK03GYWJ9_0	AC	7	Dinucleotide	107	CACAGCTGAGGCAATAA	60	TTGATTTACATGAGATC	59	52	38
Ln214	IJMZ0LK03HEXWW_0	AGG	7	Trinucleotide	126	ATTTAAAGTGGCCTCCTC	60	TATCGACAGTCACACCT	60	50	55
Ln215	cons125_2_A_0	CA	6	Dinucleotide	102	TTTAGATGCCCTGAAAT	60	ACGCTTAGCTCTTTCACC	61	45	55
Ln216	cons200_3_A_1	CT	6	Dinucleotide	113	CATACAGAGGCCCACTG	62	TTCTTTCCAGGTGATCGT	60	63	50
Ln217	cons54_2_A_0	TTGT	6	Tetranucleotide	111	CACAGGTACGCGAATCA	59	AGCTCTGCTGTTCGCTCT	60	50	58
Ln218	IJMZ0LK03F4H3R_0	TG	6	Dinucleotide	110	GCAGGAAACAGAGGAA	60	AACGAGCCGGACTAGTT	60	50	55
Ln219	IJMZ0LK03FK0TH_0	GT	6	Dinucleotide	106	GTATTGTGCATCGGCCA	60	GGAGGGCCTCTAGATGT	61	53	60
Ln220	IJMZ0LK03FLO3G_0	GT	6	Dinucleotide	117	GGTCTGCAGGGTGTACA	59	TCTGTCTGTCAGGCATC	60	60	52
Ln221	IJMZ0LK03FP3XH_0	TG	6	Dinucleotide	100	GACTATTCAGGCATAAT	58	TGTTGCCATCTCAGCAA	60	45	50
Ln222	IJMZ0LK03FR8N6_1	TG	6	Dinucleotide	143	ATCAGCCTGTCCTTCTCA	60	TAACACATACACCCTCC	61	55	55
Ln223	IJMZ0LK03FTJ2D_0	GT	6	Dinucleotide	100	TAGAAAAGGGAACAGTG	58	ATCCAACCTGGTTTCAC	60	45	53
Ln224	IJMZ0LK03FU45C_0	GCA	6	Trinucleotide	128	GAGACTGTGCTGTGAGA	60	TCACCTGCAGCTTCACT	60	57	55
Ln225	IJMZ0LK03G0T2W_0	AC	6	Dinucleotide	227	ACAAGAAAGCACTGGCT	59	TGGGGACCAAACTGCTA	60	53	55
Ln226	IJMZ0LK03G1RBW_0	AC	6	Dinucleotide	130	AGGGATGCTGTTACTGA	58	GGTTAAAGTAAGGAGCT	59	48	52
Ln227	IJMZ0LK03G55VJ_0	CA	6	Dinucleotide	171	AAACACGGTCTGGGAAA	59	TGCCCTCGAAAAGACAA	60	50	45
Ln228	IJMZ0LK03G8E13_0	AC	6	Dinucleotide	137	TTGTCTGTCTACCATTAT	59	TAAATGGGCCTCAGTCA	60	42	50
Ln229	IJMZ0LK03GAN0Q_0	AC	6	Dinucleotide	114	ACGACCGGGCTATACAA	60	TGTAGCCAACAACAGCT	58	55	50
Ln230	IJMZ0LK03GEG27_0	GT	6	Dinucleotide	168	AGCCTGAGGCGTAAACA	60	TTCTCACTGATACTACCT	60	55	46
Ln231	IJMZ0LK03GHJKA_0	CA	6	Dinucleotide	116	TGGCTAATTTACTGCAGC	60	TCTGCGTTCAAGCAAGT	60	50	50
Ln232	IJMZ0LK03GL1MW_0	TG	6	Dinucleotide	332	GTGCTCTGAGGACTTCTG	60	CTGCAGGGTGTTGCTCT	59	60	61
Ln233	IJMZ0LK03GLU1Q_0	CT	6	Dinucleotide	103	AAGGAGAGCAAAAGGA	60	CCTTTAAAGAGCCCTCC	60	50	55
Ln234	IJMZ0LK03GRMM0_0	AT	6	Dinucleotide	189	TTTGGCACAAATGAATG	59	AGAGGCAAAAGCTCACC	60	38	50
Ln235	IJMZ0LK03GS4QA_0	TC	6	Dinucleotide	110	TTCTCCCTCTGTCCAGTT	60	GGGCAATATCAGATTGG	59	55	50
Ln236	IJMZ0LK03HAQ5A_0	GT	6	Dinucleotide	185	CACTGCTACAGCCCCTTC	60	GAGAACAAACATGCCCA	59	60	50
Ln237	concat7_A_0	TGTT	5	Tetranucleotide	141	TGGGGACTATTTTCTGTG	60	TGACCTGGTACAGAGAT	60	50	55
Ln238	cons200_3_A_0	CA	5	Dinucleotide	103	AAAGTTTCACACTCCAC	60	CACCAGACCAGTGGTGG	61	50	67
Ln239	cons228_3_A_0	AG	5	Dinucleotide	108	CATGTGGGAGTTCAGAC	59	CTGATCCACCTCCTCTCT	60	55	60
Ln240	cons478_2_A_0	CT	5	Dinucleotide	121	CGCACACTACTCTGTCTC	59	AACGAGACGCGGTTGCT	63	55	61
Ln241	cons62_2_A_0	AC	5	Dinucleotide	113	TCCCACAAAATCCTTCCA	60	CCAAAATGCTGCACTGT	60	45	50
Ln242	cons628_2_A_0	CA	5	Dinucleotide	146	TCCATTATCTTTTCTCTG	59	TGATTGAGTGCCAGCAA	61	41	45
Ln243	cons685_2_A_0	TG	5	Dinucleotide	121	GGAAAGTGAGGACATTC	60	ACATCTCTACACATGGC	61	50	55
Ln244	cons75_2_A_0	TG	5	Dinucleotide	117	GGAAGTAAATGGTGGTA	59	CCTATGGCAGATCACTG	59	46	55
Ln245	cons80_2_A_0	GAA	5	Trinucleotide	178	AAGCACCATGGGTGTAA	60	TTGTGATTGGTCAGAGC	60	50	50
Ln246	IJMZ0LK03F4P2R_0	TG	5	Dinucleotide	100	TTTACACAAACAACAGA	59	TCTGAGTTCATCAACAC	60	39	37

Ln247	IJMZ0LK03F5FYH_0	CA	5	Dinucleotide	119	AACTGAAGTAAAGGAAG	57	CTGAAACAGGCAGCTCA	60	35	55
Ln248	IJMZ0LK03F5K4C_0	GA	5	Dinucleotide	102	GCAGGAAAGGACCGAAA	60	CAGCTGCTTCTCTGCCA	60	53	58
Ln249	IJMZ0LK03F5TW7_0	TG	5	Dinucleotide	105	GTAACAGGGCTATTCAC	60	ATCACCACTCTGAAGCC	60	55	55
Ln250	IJMZ0LK03F94OF_0	TG	5	Dinucleotide	166	CAGTTGGTACCCAGTGG	61	TATGCCACATGTGTTGG	60	60	50
Ln251	IJMZ0LK03FIBBS_0	GT	5	Dinucleotide	201	TTACTTGCCCTCCAGCAA	60	TGCAGGTAAGACACATA	60	50	44
Ln252	IJMZ0LK03FLY16_0	TG	5	Dinucleotide	103	TCACAGTTAAACATTCA	59	TCCCCAGGAAACTTTGA	60	32	50
Ln253	IJMZ0LK03FM0P9_1	TG	5	Dinucleotide	125	GCATCAGAGTGTGTCTG	60	GCACCCATGCATGCAGA	62	48	55
Ln254	IJMZ0LK03FOUNE_0	TA	5	Dinucleotide	102	ATCGTATGGGGATGCTC	60	CAAGAGAATAAGAACA	59	50	43
Ln255	IJMZ0LK03FR5K2_0	AC	5	Dinucleotide	105	GCATGTAGGTTAATTAG	60	ACATTGCGGCAAAGATG	60	42	47
Ln256	IJMZ0LK03FSQ1Q_0	AAT	5	Trinucleotide	141	GTTCCAGAGTCGGAGGT	60	CCATATTTGTAAGTTTCC	60	55	38
Ln257	IJMZ0LK03FTB4R_0	GT	5	Dinucleotide	186	GTCCATATCCAAGCTGA	60	GGTGAAAGTTCCCTTCC	61	55	55
Ln258	IJMZ0LK03FUL9P_0	AC	5	Dinucleotide	100	CATGCATGTATTCAAGG	60	CCCCTTCGTGTATGTGTC	59	42	55
Ln259	IJMZ0LK03FUZ4N_0	CA	5	Dinucleotide	140	CCTCAAATGCATTAGTAT	60	TGCTGTGTCGACGTAAG	60	38	55
Ln260	IJMZ0LK03FVO4J_0	CCT	5	Trinucleotide	104	CACCTGCTTAGAAACAG	60	GACACGCTCACTGAAGT	57	46	50
Ln261	IJMZ0LK03FXHR8_0	TC	5	Dinucleotide	117	GCCAATCTCGAGAGCAA	61	TAAACAAGGGCATAGCG	60	50	50
Ln262	IJMZ0LK03FYI1K_0	AG	5	Dinucleotide	113	GGATGAATGGTGCCAGA	60	ATGAATGCCCACACGCA	62	55	56
Ln263	IJMZ0LK03G0HAF_0	TG	5	Dinucleotide	119	CAAGTCAAGCAGTTTAT	59	CATCTGCCTTGCTGTGA	60	42	55
Ln264	IJMZ0LK03G46TD_0	CA	5	Dinucleotide	121	GGCAGGTTCACTTGCCTT	60	TGTACATGGTCTATTTGA	59	55	42
Ln265	IJMZ0LK03G77YK_0	CA	5	Dinucleotide	110	GAGCCATCAACAGCAAA	59	TGTTTACATGGTGCAAC	60	50	41
Ln266	IJMZ0LK03G813H_0	GA	5	Dinucleotide	112	AGAGTCTGTACAGCCCA	60	TGACTCATCTCCACACCT	60	60	55
Ln267	IJMZ0LK03G8CI2_0	TAA	5	Trinucleotide	109	TGGTTATAGAGAAGAAG	58	TCCTCCGTCACAGAGTTT	60	40	55
Ln268	IJMZ0LK03GESU3_0	TG	5	Dinucleotide	101	CAGCCACATGATGAGAA	61	GAGGCTGCAGCTGACTG	61	48	67
Ln269	IJMZ0LK03GGCFT_0	CT	5	Dinucleotide	108	CAGAGCCATGTGACCAA	59	ACCGACTGAGGTTTGAG	60	50	55
Ln270	IJMZ0LK03GJAXJ_0	AC	5	Dinucleotide	211	ATCGTTCTGCTACCTGCC	60	TTAGGTCGTGTTTGCGA	60	55	50
Ln271	IJMZ0LK03GJE9M_0	AC	5	Dinucleotide	100	TGTCTTCTCTGTCAAGGT	59	TATCGCTGGGGACCTAA	59	50	50
Ln272	IJMZ0LK03GJYMG_0	TG	5	Dinucleotide	113	TGGAGCAGAGGCTAATC	60	TGTGAGCTGCTCTTATCC	61	55	52
Ln273	IJMZ0LK03GKXXJ_0	TG	5	Dinucleotide	133	TCTAACTGGCATTGGCTC	59	TCCTGCAGTGATTCAAG	60	50	50
Ln274	IJMZ0LK03GMJZ2_0	TG	5	Dinucleotide	132	CTCAGAGGCACTCAGGC	60	TATCATGTGACGACGAT	61	63	50
Ln275	IJMZ0LK03GMSH2_0	GT	5	Dinucleotide	210	CCATTTTAACAATGCCG	60	TCTCGGTAACAGCAGGG	60	45	55
Ln276	IJMZ0LK03GQ0YH_0	AC	5	Dinucleotide	106	TGTATGACTTAAACAAA	58	CAGTGACGAGCTCCTCT	59	39	60
Ln277	IJMZ0LK03GQR8Q_0	AT	5	Dinucleotide	108	TTGTGGGAAAACTTGCT	60	AAAGATTTTCCATATCA	57	45	30
Ln278	IJMZ0LK03GW1IM_0	AC	5	Dinucleotide	114	GCAGGTTTTAGGTCCACT	60	CGCTTTTACCAGCCACA	62	50	50
Ln279	IJMZ0LK03GWBU7_0	CA	5	Dinucleotide	155	ATGTTCAGGTGGTTTCGA	60	CAACCACTTGTTTCTGGC	59	50	50
Ln280	IJMZ0LK03GXCH6_0	AG	5	Dinucleotide	124	AATGTACTCTGGCCTGTC	59	TTAACTCGGCCTCGTTGT	60	55	50
Ln281	IJMZ0LK03HBX63_0	TG	5	Dinucleotide	107	TGCAGAAAATGAAATGA	59	TGCGCACTCACATCTTC	60	40	48
Ln282	IJMZ0LK03HDIKA_0	TC	5	Dinucleotide	100	TTTCACTGACGAAATTG	60	GGAGAAGGAGAGGATC	60	41	60
Ln283	IJMZ0LK03HDIKA_1	CTC	5	Trinucleotide	124	CCCTGATCCTCTCCTTCT	60	AAACATTGATTGCCTCA	59	60	43
Ln284	IJMZ0LK03HF5QN_0	AC	5	Dinucleotide	116	AAAACAAAGAGAAGAA	57	TCCTTACACAGATATAG	60	33	42
Ln285	IJMZ0LK03HFVFN_0	AC	5	Dinucleotide	115	CCAGCCAACACTCATAT	59	GCATGCATCTCTGGATTT	61	50	43

Appendix 3.1: *Bagrus docmak* haplotypes and corresponding sample source details

			Haplotype frequency					
Haplotype name	Sequence name	Sample location	Lake Victoria	Victoria Nile River	Lake Albert	Lake Edward	Kazinga Channel	Total Number Sequences(n)
HAP1	ALBBDMURC_195	Lake Albert	0	0	1	0	0	1
HAP23	ALB_NTK27	Lake Albert	0	0	1	0	0	1
HAP2	ALB_WSK3	Lake Albert	0	1	5	13	14	33
HAP3	ALB_WSK4	Lake Albert	0	1	2	0	0	3
HAP4	ALB_WSK5	Lake Albert	1	2	1	0	0	4
HAP17	ALB_WSK9	Lake Albert	0	0	2	0	0	2
HAP21	EDWM4	Lake Edward	0	0	0	1	0	1
HAP18	EDWM5	Lake Edward	0	0	0	1	0	1
HAP22	EDWM7	Lake Edward	0	0	0	1	0	1
HAP25	EDWV19	Lake Edward	0	0	0	1	0	1
HAP5	EDWV6	Lake Edward	0	0	0	1	1	2
HAP24	EDWV8	Lake Edward	0	0	0	1	0	1
HAP6	KAZ_F1	Kazinga Channe	0	0	0	1	2	3
HAP9	KAZ_F12	Kazinga Channel	0	5	10	0	1	16
HAP10	KAZ_F15	Kazinga Channel	0	0	0	0	1	1
HAP7	KAZ_F2	Kazinga Channel	0	0	0	0	3	3
HAP11	KAZ_F21	Kazinga Channel	0	0	0	0	1	1
HAP12	KAZ_F25	Kazinga Channel	0	0	3	0	1	4
HAP8	KAZ_F3	Kazinga Channel	0	0	0	0	1	1
HAP19	KAZF27	Kazinga Channel	0	0	0	0	1	1
HAP20	KAZF29	Kazinga Channel	0	0	0	0	1	1
HAP13	NILE_RNK29	Victoria Nile	0	1	0	0	0	1
HAP14	NILE_RNK39	Victoria Nile	0	1	0	0	0	1
HAP15	NILE_RNK47	Victoria Nile	6	2	0	0	0	8
HAP16	VIC_RNK7	Lake Victoria	1	0	0	0	0	1
Total								93

Appendix 3.2: List of Microsatellite Primers for *Bagrus docmak*

Locus	Primer sequence	T_M (°C)	Motif	Size range (bp)
†Bd04	F: TGTGGACCAAGAGACAGGTG R: AATGAACAAGGCAGGTGATG	59	$(AGAT)^{18}$	200-208
†Bd18	F: ATGGGGAGGAAAAGTGGAG R: CCTGAGTGCATTGCTCATGG	61	$(AC)^{15}$	100-102
†Bd01	F:TTGCCAATCCTGATGACACTC R:TAAAGCTGGGCAACTGATCC	60	$(TTCT)^{15}$	203-219
†Bd02	F:TGTGCTCTGACCCCTACCTC R:GGGTATCGCATCCCAGATAG	60	$(AGAT)^{17}$	110-130
†Bd12	F: CCGACCATCTCAAATACAAGTC R: CTCTTCCCCAAGGCTATTCC	60	$(AAT)^{18}$	237-258
†Bd09	F: ACTGTTCCCATGAAGTTGGG R: TGGTCAACTTTAGATGTGCAGC	60	$(ATT)^{19}$	223-238
†Bd06	F: TTCTGAAGCCCAAAGTAGACG R: GCCCACACTATTGACACAGG	59	$(GATA)^{16}$	171-199
†Bd20	F: TCCTGGAGACCAAGACCAAG R: TGCAGGTTAAGAATGGAGGC	60	$(CA)^{11}$	156-168
†Bd05	F: GCTGGCAACATGCAGTAATC R: CAGCATTTCATTGCTATGTGC	59	$(ATAC)^{15}$	136-172
BD07	F: GAGCACACGAAACATTGCAG R: TTGTAGATTCCCTTTGGGATG	60	$(GATA)^{15}$	125-157
Bd16	F: GCAATCGCACTCTTGTTATCG R: TAGTAGCGCACCCAGGAAAC	61	$(ATT)^{13}$	83
BD08	F: TTACCTCACACTCTGGGGTTG R: GGTAAAGGTTTACACTGTGGGG	60	$(ATCT)^{16}$	179-187
†BD03	F:CCTGCAGGAGTTTGTTTGTG R: CGTGCCATAGGCATTTATCC	60	$(TAGA)^{15}$	159-191
†BD14	F: CTTTAATGACACTGCGCTGC R: CTCAAAGCGCTTGAAGTGG	60	$(TAT)^{17}$	218-233
†Bd10	F: GTCCCACGGACTGAAAAGTG R:TCAACTTCTTAGCACAAAATCAGAC	60	$(TTA)^{15}$	266-287

T_M, optimal primer melting temperature; †Polymorphic loci used for analyses in the current study

Source: Extracted from Basiita et al (2015)

Appendix 4. 1: Hardy Weinberg equilibrium multi-population global test results considering all Nile perch populations sampled across loci

Locus	P-value Het deficit	sig BF corrected (p<0.001)	P-value Het excess	sig BF corrected (p<0.001)
*Ln*31	0.0067	ns	1	ns
*Ln*29	0.0001	ns	0.8623	ns
*Ln*23	0.1377	ns	0.9957	ns
*Ln*19	0.0043	ns	0.5273	ns
*Ln*17	0.4727	ns	0.9634	ns
*Ln*15	0.0366	ns	0.8873	ns
*Ln*11	0.1127	ns	0.9677	ns
*Ln*10	0.0323	ns	0.9677	ns
*Ln*09	0.0136	ns	0.9864	ns
*Ln*02	0.1886	ns	0.8114	ns
*Lca*98	0.3861	ns	0.6139	ns
*Lca*74	0.5017	ns	0.4983	ns
*Lca*70	0.0323	ns	0.9677	ns
*Lca*69	0.0825	ns	0.9175	ns
*Lca*64	0.0308	ns	0.9692	ns
*Lca*58	0.8265	ns	0.1735	ns
*Lca*21	0.0201	ns	0.9799	ns
*Lca*20	0.726	ns	0.2741	ns
*Lca*08	0.1559	ns	0.8441	ns

Appendix 4. 2: Significant deviations following a sequential Bonferroni (at P<0.001) correction are indicated by sig and non-significant by ns.

Locus Pair	P- value	Significance after BF correction (P<0.001)
Lca08 - Lca21	0.473845	ns
Lca08 - Lca69	0.134888	ns
Lca08 - Lca70	0.856118	ns
Lca08 - Lca98	0.631919	ns
Lca08 - Ln09	0.751508	ns
Lca08 - Ln10	0.754913	ns
Lca08 - Ln11	0.871627	ns
Lca08 - Ln31	0.889322	ns
Lca20 - Lca08	0.53766	ns
Lca20 - Lca21	0.679519	ns
Lca20 - Lca70	0.488531	ns
Lca20 - Lca98	0.867584	ns
Lca20 - Ln09	0.593766	ns
Lca20 - Ln10	0.484361	ns
Lca20 - Ln11	0.139973	ns
Lca20 - Ln31	0.281037	ns
Lca21 - Lca70	0.083587	ns
Lca21 - Lca98	0.09418	ns
Lca21 - Ln09	0.909244	ns
Lca21 - Ln10	0.823577	ns
Lca21 - Ln11	0.725201	ns
Lca21 - Ln31	0.596987	ns
Lca58 - Lca08	0.995697	ns
Lca58 - Lca20	0.5959	ns
Lca58 - Lca21	0.737602	ns
Lca58 - Lca69	0.28715	ns
Lca58 - Lca70	0.874892	ns
Lca58 - Lca98	0.862074	ns
Lca58 - Ln09	0.211763	ns
Lca58 - Ln10	0.620908	ns
Lca58 - Ln11	0.999849	ns
Lca58 - Ln31	0.432461	ns
Lca64 - Lca08	0.045761	ns
Lca64 - Lca20	0.919774	ns
Lca64 - Lca21	0.093184	ns
Lca64 - Lca58	0.033216	ns
Lca64 - Lca69	0.925175	ns
Lca64 - Lca70	0.444409	ns
Lca64 - Lca98	0.086117	ns
Lca64 - Ln09	0.912414	ns
Lca64 - Ln10	0.994245	ns
Lca64 - Ln11	0.933869	ns
Lca64 - Ln31	0.784254	ns

Lca69 - Lca21	0.680395	ns
Lca69 - Lca70	0.650975	ns
Lca69 - Lca98	0.202709	ns
Lca69 - Ln09	0.627302	ns
Lca69 - Ln10	0.377689	ns
Lca69 - Ln11	0.704024	ns
Lca69 - Ln31	0.095364	ns
Lca70 - Lca98	0.209319	ns
Lca70 - Ln09	0.590476	ns
Lca70 - Ln10	0.253418	ns
Lca70 - Ln11	0.623025	ns
Lca70 - Ln31	0.943863	ns
Lca74 - Lca08	0.936267	ns
Lca74 - Lca20	0.192856	ns
Lca74 - Lca21	0.261683	ns
Lca74 - Lca58	0.538154	ns
Lca74 - Lca64	0.554683	ns
Lca74 - Lca69	0.71156	ns
Lca74 - Lca70	0.804785	ns
Lca74 - Lca98	0.491614	ns
Lca74 - Ln09	0.291263	ns
Lca74 - Ln10	0.119031	ns
Lca74 - Ln31	0.185223	ns
Lca98 - Ln09	0.966977	ns
Lca98 - Ln10	0.290625	ns
Lca98 - Ln11	0.727798	ns
Lca98 - Ln31	0.499494	ns
Ln02 - Lca08	0.56349	ns
Ln02 - Lca20	0.295426	ns
Ln02 - Lca21	0.781242	ns
Ln02 - Lca58	0.387967	ns
Ln02 - Lca64	0.155556	ns
Ln02 - Lca69	0.390843	ns
Ln02 - Lca70	0.357571	ns
Ln02 - Lca74	0.013067	ns
Ln02 - Lca98	0.548307	ns
Ln02 - Ln09	0.710578	ns
Ln02 - Ln10	0.808529	ns
Ln02 - Ln17	0.136822	ns
Ln02 - Ln23	0.239636	ns
Ln09 - Ln10	0.374678	ns
Ln09 - Ln10	0.374678	ns
Ln11 - Ln09	0.774707	ns
Ln11 - Ln10	0.094837	ns
Ln15 - Lca08	0.468595	ns
Ln15 - Lca20	0.509858	ns
Ln15 - Lca21	0.218403	ns
Ln15 - Lca58	0.615616	ns

Ln15 - Lca64	0.271526	ns
Ln15 - Lca69	0.323635	ns
Ln15 - Lca70	0.661926	ns
Ln15 - Lca74	0.85427	ns
Ln15 - Lca98	0.273414	ns
Ln15 - Ln02	0.266995	ns
Ln15 - Ln09	0.996638	ns
Ln15 - Ln10	0.993954	ns
Ln15 - Ln11	0.011413	ns
Ln15 - Ln23	0.053381	ns
Ln15 - Ln29	0.91718	ns
Ln15 - Ln31	0.341141	ns
Ln17 - Lca08	0.61627	ns
Ln17 - Lca20	0.743607	ns
Ln17 - Lca21	0.872606	ns
Ln17 - Lca64	0.645709	ns
Ln17 - Lca69	0.576378	ns
Ln17 - Lca70	0.247501	ns
Ln17 - Lca74	0.417542	ns
Ln17 - Lca98	0.865505	ns
Ln17 - Ln09	0.594995	ns
Ln17 - Ln10	0.578207	ns
Ln17 - Ln11	0.042318	ns
Ln17 - Ln23	0.636463	ns
Ln19 - Lca08	0.997054	ns
Ln19 - Lca20	0.300788	ns
Ln19 - Lca21	0.211957	ns
Ln19 - Lca58	0.722913	ns
Ln19 - Lca64	0.962531	ns
Ln19 - Lca69	0.624461	ns
Ln19 - Lca70	0.865987	ns
Ln19 - Lca74	0.006648	ns
Ln19 - Lca98	0.787921	ns
Ln19 - Ln02	0.472136	ns
Ln19 - Ln09	0.785886	ns
Ln19 - Ln10	0.262912	ns
Ln19 - Ln11	0.300643	ns
Ln19 - Ln15	0.019652	ns
Ln19 - Ln17	0.726416	ns
Ln19 - Ln23	0.583389	ns
Ln19 - Ln29	0.656511	ns
Ln19 - Ln31	0.621049	ns
Ln23 - Lca08	0.141831	ns
Ln23 - Lca20	0.282408	ns
Ln23 - Lca21	0.269076	ns
Ln23 - Lca58	0.298156	ns
Ln23 - Lca64	0.189877	ns
Ln23 - Lca69	0.650719	ns

Ln23 - Lca70	0.527844	ns
Ln23 - Lca74	0.778772	ns
Ln23 - Lca98	0.41623	ns
Ln23 - Ln09	0.457027	ns
Ln23 - Ln10	0.635062	ns
Ln23 - Ln11	0.20239	ns
Ln23 - Ln31	0.517196	ns
Ln29 - Lca08	0.389832	ns
Ln29 - Lca20	0.035329	ns
Ln29 - Lca21	0.390994	ns
Ln29 - Lca58	0.947353	ns
Ln29 - Lca64	0.797242	ns
Ln29 - Lca69	0.126584	ns
Ln29 - Lca70	0.71274	ns
Ln29 - Lca74	0.707676	ns
Ln29 - Lca98	0.551725	ns
Ln29 - Ln02	0.02523	ns
Ln29 - Ln09	0.23468	ns
Ln29 - Ln10	0.905091	ns
Ln29 - Ln11	0.193666	ns
Ln29 - Ln17	0.996391	ns
Ln29 - Ln31	0.158705	ns
Ln31 - Ln09	0.240827	ns
Ln31 - Ln10	0.457154	ns
Ln31 - Ln11	0.427965	ns
Lca20 - Lca69	Highly sig.	sig
Lca74 - Ln11	Highly sig	sig
Ln02 - Ln11	Highly sig	sig
Ln02 - Ln31	Highly sig	sig
Ln15 - Ln17	Highly sig	sig
Ln17 - Lca58	Highly sig	sig
Ln17 - Ln31	Highly sig	sig
Ln29 - Ln23	Highly sig	sig

ns- not signifcant and sig for significant deviations from HW even after sequential Bonferroni (BF) correction with $P<0.001$.

Appendix 4.3: Pairwise F_{ST} estimates for *Lates niloticus*, Nile perch, for all loci as estimated in Genepop on the web 1.2 (Rousset, 2008).

Population	SEN	NIG	ALB	KYO	VIC	TUR
SEN						
NIG	0.12					
ALB	0.36	0.31				
KYO	0.39	0.33	0.11			
VIC	0.40	0.37	0.07	0.15		
TUR	0.40	0.37	0.13	0.19	0.17	

SEN- Senegal River, NIG- Lake Kainji, ALB- Lake Albert, KYO- Lake Kyoga, VIC – Lake Victoria and TUR- Lake Turkana. All pairwise comparison were significant at $P<0.05$, as shown in the table below

P-value for each population pair across all loci (Fisher's method) implemented in Genepop v 4.2

Population pair		Chi2	df	P-Value	
SEN	NIG	Infinity	38	Highly	sign.
SEN	ALB	Infinity	38	Highly	sign.
NIG	ALB	Infinity	38	Highly	sign.
SEN	KYO	Infinity	38	Highly	sign.
NIG	KYO	Infinity	38	Highly	sign.
ALB	KYO	Infinity	38	Highly	sign.
SEN	VIC	Infinity	38	Highly	sign.
NIG	VIC	Infinity	38	Highly	sign.
ALB	VIC	Infinity	38	Highly	sign.
KYO	VIC	Infinity	38	Highly	sign.
SEN	TUR	Infinity	38	Highly	sign.
NIG	TUR	Infinity	38	Highly	sign.
ALB	TUR	Infinity	38	Highly	sign.
KYO	TUR	Infinity	38	Highly	sign.
VIC	TUR	Infinity	38	Highly	sign.

Appendix 4.4: Genetic diversity tests for Nile perch, *Lates niloticus*

a) Allelic Richness per locus and population (FSTAT vs 2.9.3.2)

	SEN	NIG	ALB	KYO	VIC	TUR	Overall
*Ln*19	2.00	2.88	2.99	3.00	2.89	3.94	6.33
*Ln*15	7.01	8.98	6.40	5.14	4.95	7.44	8.20
*Ln*29	3.64	4.91	4.38	3.36	2.66	2.61	5.15
*Ln*02	7.54	7.29	5.34	4.94	4.58	4.46	8.32
*Ln*17	6.48	7.62	9.15	7.34	7.79	8.31	9.54
*Ln*23	2.61	2.96	2.64	1.88	2.41	1.00	3.64
*Lca*74	5.51	5.33	7.00	5.68	6.37	6.19	8.31
*Lca*64	2.73	2.35	2.06	1.29	2.02	3.47	2.44
*Lca*58	8.83	10.65	4.52	3.97	4.27	6.88	7.47
*Lca*20	2.00	4.15	3.49	3.00	3.29	2.98	4.02
*Lca*08	2.67	1.39	2.90	2.95	2.00	3.75	3.49
*Lca*69	2.00	2.87	4.26	3.00	3.30	3.00	4.66
*Lca*21	3.28	1.00	1.92	2.76	1.30	3.37	2.37
*Lca*70	3.00	3.66	2.85	2.95	2.00	3.60	4.65
*Lca*98	7.31	8.31	5.07	6.09	4.10	6.16	8.28
*Ln*31	2.64	5.24	3.76	2.81	3.60	7.42	5.98
*Ln*11	3.61	6.29	3.98	2.59	3.09	4.14	5.81
*Ln*09	4.28	3.76	2.42	1.30	2.19	1.00	4.79
*Ln*10	4.71	3.03	4.55	4.55	4.13	4.46	4.74
Mean	4.31	4.88	4.19	3.61	3.52	4.43	5.69
SE	0.47	0.58	0.40	0.35	0.35	0.46	0.47

b) Number of alleles sampled (FSTAT vs 2.9.3.2)

Locus	SEN	NIG	ALB	KYO	VIC	TUR	Total
*Ln*19	2	3	3	3	3	5	11
*Ln*15	8	12	10	6	7	9	22
*Ln*29	4	6	5	4	3	3	8
*Ln*02	8	9	6	5	5	5	14
*Ln*17	7	9	12	9	10	9	17
*Ln*23	3	3	4	2	4	1	5
*Lca*74	6	7	10	8	7	7	16
*Lca*64	3	3	4	2	3	4	6
*Lca*58	9	11	9	5	7	8	24
*Lca*20	2	5	5	3	4	3	5
*Lca*08	3	2	4	3	2	5	8
*Lca*69	2	3	5	3	4	3	5
*Lca*21	4	1	4	3	2	4	7
*Lca*70	3	4	4	3	2	4	8
*Lca*98	8	14	6	7	6	8	24
*Ln*31	3	7	6	5	5	8	16
*Ln*11	4	8	6	4	4	5	12
*Ln*09	5	4	5	2	3	1	9
*Ln*10	5	4	5	7	5	5	8
Total	89	115	113	84	86	97	225

c) Nei's estimation of heterozygosity (FSTAT vs 2.9.3.2)

Locus	Ho	Hs	Ht	Dst	Dst'	Ht'	Gst	Gst'	Gis
Ln19	0.398	0.471	0.791	0.321	0.385	0.856	0.406	0.45	0.154
Ln15	0.704	0.751	0.846	0.095	0.114	0.865	0.112	0.131	0.062
Ln29	0.362	0.548	0.711	0.163	0.195	0.743	0.229	0.263	0.34
Ln02	0.775	0.787	0.868	0.081	0.097	0.884	0.093	0.109	0.016
Ln17	0.861	0.85	0.891	0.041	0.05	0.9	0.046	0.055	-0.014
Ln23	0.231	0.262	0.58	0.318	0.381	0.643	0.548	0.592	0.117
Lca74	0.806	0.761	0.87	0.109	0.13	0.892	0.125	0.146	-0.058
Lca64	0.145	0.231	0.248	0.017	0.021	0.252	0.069	0.082	0.374
Lca58	0.692	0.649	0.801	0.152	0.182	0.831	0.19	0.22	-0.068
Lca20	0.662	0.596	0.71	0.113	0.136	0.732	0.16	0.186	-0.11
Lca08	0.352	0.339	0.676	0.337	0.405	0.743	0.499	0.544	-0.039
Lca69	0.608	0.614	0.748	0.133	0.16	0.774	0.178	0.206	0.011
Lca21	0.151	0.182	0.194	0.011	0.014	0.196	0.059	0.07	0.172
Lca70	0.452	0.496	0.737	0.241	0.289	0.785	0.327	0.369	0.089
Lca98	0.745	0.735	0.868	0.132	0.159	0.894	0.153	0.178	-0.013
Ln31	0.47	0.475	0.724	0.248	0.298	0.774	0.343	0.385	0.012
Ln11	0.46	0.489	0.664	0.175	0.21	0.699	0.264	0.301	0.06
Ln09	0.267	0.268	0.696	0.428	0.514	0.782	0.615	0.657	0.003
Ln10	0.619	0.653	0.738	0.085	0.102	0.754	0.115	0.135	0.052
Overall	0.514	0.535	0.703	0.168	0.202	0.737	0.24	0.274	0.039

d) F_{IS} Per population at each of the 19 loci examined: (FSTAT)

Locus	SEN	NIG	ALB	KYO	VIC	TUR
Ln19	0.6	0.521	0.067	0.147	-0.001	-0.215
Ln15	0.244	-0.054	-0.028	0.343	-0.118	-0.019
Ln29	0.435	0.302	0.419	0.036	0.102	0.745
Ln02	-0.091	-0.015	0.136	0.002	0.195	-0.125
Ln17	-0.004	-0.001	0.145	-0.04	-0.021	-0.167
Ln23	-0.178	0.043	0.46	-0.072	0.09	NA
Lca74	-0.223	0.008	0.174	-0.022	-0.013	-0.277
Lca64	0.533	0.444	0.178	0.000	0.231	0.308
Lca58	-0.017	-0.008	-0.149	0.051	0.003	-0.236
Lca20	-0.467	0.032	-0.002	0.112	-0.354	-0.048
Lca08	-0.05	0.000	-0.058	-0.221	0.18	-0.053
Lca69	-0.385	0.066	0.215	0.009	-0.083	0.122
Lca21	0.211	NA	-0.02	0.143	0.000	0.255
Lca70	-0.045	0.483	0.195	-0.149	-0.068	0.279
Lca98	-0.054	0.274	0.061	-0.123	-0.114	-0.151
Ln31	0.342	-0.137	0.109	-0.011	0.372	-0.186
Ln11	0.042	0.155	0.171	0.008	-0.136	-0.085
Ln09	0.017	-0.181	0.405	0.000	0.379	NA
Ln10	-0.156	-0.02	0.33	0.027	0.034	0.042
Mean	0.009	0.078	0.148	0.019	0.001	-0.029

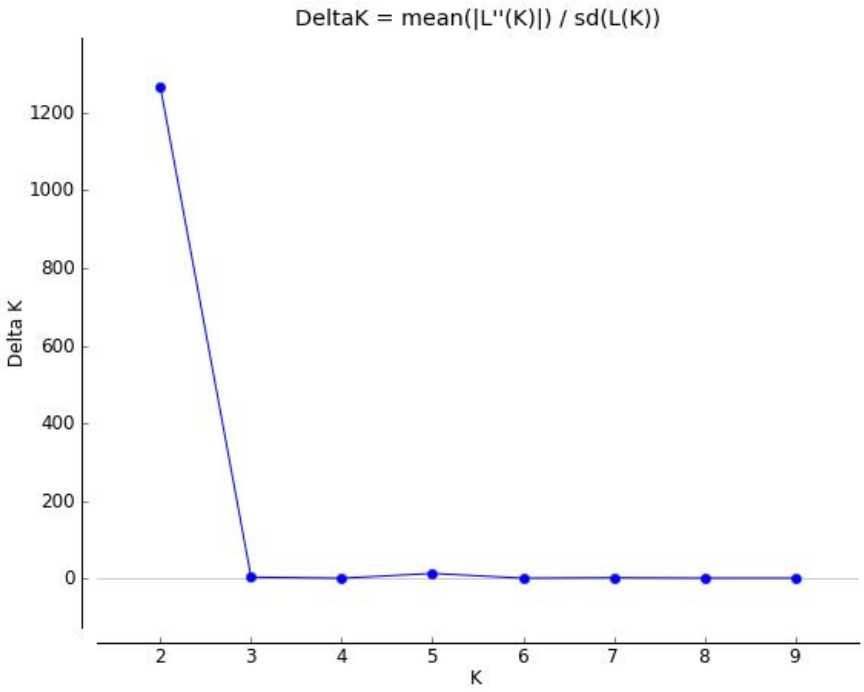
e) Private allelic richness calculated in HP-Rare vs 1.1

Locus	SEN	NIG	ALB	KYO	VIC	TUR
Ln19	0.03	0.91	0.00	0.00	2.89	2.29
Ln15	2.67	1.88	1.51	0.10	0.63	1.28
Ln29	0.00	1.46	0.11	0.08	0.00	0.48
Ln02	2.60	1.02	0.04	0.09	0.03	0.08
Ln17	0.72	0.07	1.41	0.08	0.69	0.45
Ln23	0.02	0.00	0.41	0.00	0.00	0.00
Lca74	1.24	0.54	0.53	0.11	0.06	1.62
Lca64	0.04	0.38	0.27	0.01	0.02	0.88
Lca58	3.01	3.28	0.58	0.12	0.93	2.02
Lca20	0.00	0.33	0.07	0.00	0.09	0.00
Lca08	1.30	0.03	0.09	0.00	0.00	1.87
Lca69	0.00	0.00	0.00	0.00	0.00	0.00
Lca21	0.85	0.00	0.27	0.76	0.00	0.85
Lca70	1.17	0.91	0.43	0.19	0.00	0.02
Lca98	3.31	4.42	0.04	0.64	0.60	0.73
Ln31	0.17	0.97	0.14	0.57	0.69	4.45
Ln11	0.06	1.72	0.59	0.00	0.03	1.47
Ln09	1.80	0.28	0.29	0.23	0.39	0.00
Ln10	1.08	0.00	0.11	0.32	0.03	0.12
Mean	1.06	0.96	0.36	0.17	0.37	0.98
SE	0.26	0.27	0.10	0.05	0.16	0.26

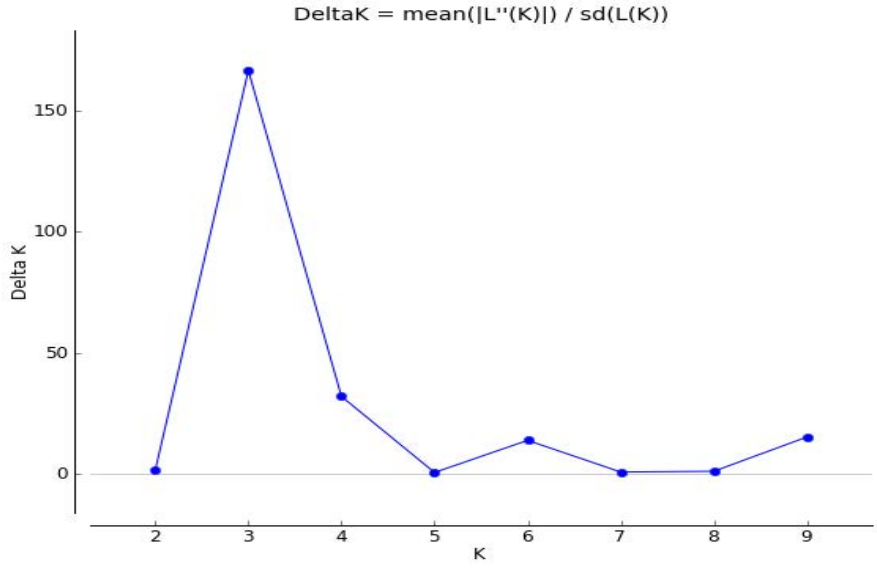
f) Rare allele calculated for both translocated and native Nile perch populations in East Africa

Locus	Population				Total
	ALB	KYO	VIC	TUR	
Ln19	0	0	0	3	3
Ln15	5	1	4	3	13
Ln29	1	1	1	1	4
Ln17	6	2	3	2	13
Ln23	2	0	2	0	4
Ln11	5	3	1	2	11
Lca74	1	4	1	2	8
Ln10	3	1	2	1	7
Ln09	4	1	1	0	6
Lca58	5	1	4	2	12
Lca20	2	0	1	0	3
Lca8	2	0	0	4	6
Lca69	1	0	1	0	2
Lca21	3	0	1	2	6
Lca70	2	0	0	1	3
Lca98	2	1	2	4	9
Lca31	3	3	2	2	10
Lca11	3	2	1	3	9
Mean	2.78	1.11	1.50	1.78	7.17
SE	0.38	0.28	0.28	0.30	0.83

Appendix 4.5: Delta K value for Nile perch, following both posterior probability and delta log likelihood methods as implemented in STRUCTURE HARVESTER 0.6



Plot for detecting the number of K groups of Nile perch, *Lates niloticus* that best fit the data; in the current study K = 2 (West and East African genetic groupings)



Graph showing the number of clusters estimated based on delta K = 3, for Nile perch, *Lates niloticus* in the East African region.

Appendix 5.1: Mutation Model based of the sequence data as calculated in Mega (7)

Model	No. of Parameters	BIC	AICc	lnL	Invariant	Gamma
T92+G	166	2810.52	1538.01	-601.27	n/a	0.46
HKY+G	168	2812.33	1524.53	-592.48	n/a	0.42
T92+I	166	2814.35	1541.84	-603.18	0.46	n/a
T92	165	2815.10	1550.23	-608.40	n/a	n/a
HKY+I	168	2816.52	1528.71	-594.58	0.48	n/a
HKY	167	2818.79	1538.63	-600.55	n/a	n/a
T92+G+I	167	2820.20	1540.04	-601.26	0.05	0.50
HKY+G+I	169	2822.01	1526.56	-592.48	0.06	0.47
TN93+G	169	2822.01	1526.57	-592.48	n/a	0.42
TN93+I	169	2826.12	1530.68	-594.54	0.49	n/a
TN93	168	2828.36	1540.56	-600.50	n/a	n/a
TN93+G+I	170	2831.68	1528.59	-592.47	0.06	0.47
K2+G	165	2848.06	1583.20	-624.88	n/a	0.43
GTR+G	172	2849.27	1530.90	-591.58	n/a	0.44
GTR+I	172	2853.20	1534.82	-593.54	0.48	n/a
GTR	171	2854.66	1543.93	-599.12	n/a	n/a
K2	164	2856.88	1599.66	-634.13	n/a	n/a
K2+G+I	166	2857.66	1585.15	-624.83	0.21	0.67
K2+I	165	2858.33	1593.46	-630.01	0.35	n/a
GTR+G+I	173	2858.95	1532.93	-591.58	0.05	0.49
JC+G	164	2869.14	1611.92	-640.26	n/a	0.60
JC	163	2869.82	1620.24	-645.44	n/a	n/a
JC+I	164	2871.62	1614.40	-641.50	0.35	n/a
JC+G+I	165	2881.07	1616.20	-641.38	0.17	0.62

Appendix 5.2: Details of diversity indices per population of *Bagrus bayad* at each microsatellite loci

Population	Locus	*N*	*Na*	*Ne*	*I*	H_o	H_e	*UHe*	*F*
ALB	*Bd04*	26	19.000	9.455	2.568	0.808	0.894	0.912	0.097
	Bd18	23	9.000	3.245	1.623	0.652	0.692	0.707	0.057
	Bd01	23	10.000	5.688	1.965	0.826	0.824	0.843	-0.002
	Bd02	24	6.000	3.056	1.364	0.542	0.673	0.687	0.195
	Bd12	27	4.000	2.695	1.160	0.519	0.629	0.641	0.176
	Bd09	28	5.000	2.533	1.182	0.607	0.605	0.616	-0.003
	Bd05	27	12.000	3.488	1.773	0.444	0.713	0.727	0.377
	Bd08	26	5.000	4.265	1.519	0.654	0.766	0.781	0.146
	Bd14	28	7.000	2.835	1.353	0.571	0.647	0.659	0.117
NIG	*Bd04*	25	8.000	2.815	1.378	0.720	0.645	0.658	-0.117
	Bd18	22	8.000	4.084	1.633	0.636	0.755	0.773	0.157
	Bd01	19	4.000	2.854	1.142	0.737	0.650	0.667	-0.134
	Bd02	19	7.000	5.870	1.844	0.842	0.830	0.852	-0.015
	Bd12	22	6.000	2.839	1.352	0.545	0.648	0.663	0.158
	Bd09	23	4.000	2.513	1.108	0.652	0.602	0.615	-0.083
	Bd05	25	13.000	6.410	2.143	0.400	0.844	0.861	0.526
	Bd08	20	6.000	3.524	1.468	0.850	0.716	0.735	-0.187
	Bd14	17	8.000	2.766	1.402	0.647	0.638	0.658	-0.014
SEN	*Bd04*	42	6.000	3.741	1.481	0.690	0.733	0.742	0.058
	Bd18	36	8.000	2.112	1.115	0.500	0.527	0.534	0.051
	Bd01	37	6.000	1.182	0.408	0.108	0.154	0.156	0.299
	Bd02	23	11.000	3.712	1.774	0.652	0.731	0.747	0.107
	Bd12	40	2.000	1.025	0.067	0.025	0.025	0.025	-0.013
	Bd09	40	3.000	1.134	0.265	0.125	0.118	0.120	-0.055
	Bd05	38	13.000	8.805	2.377	0.500	0.886	0.898	0.436
	Bd08	35	5.000	3.490	1.350	0.371	0.713	0.724	0.479
	Bd14	41	8.000	5.303	1.801	0.805	0.811	0.821	0.008

Populations ALB = Lake Albert, NIG = Lake Kainji on Niger River and SEN- Senegal River. Ho = observed heterozygosity, He = expected heterozygosity

Appendix 5.3: Total number of alleles for *Bagrus bayad* as calculated in HP-Rare (Kalinowski, 2005)

Locus	Overall	ALB	NIG	SEN
Bd04	23	19	8	6
Bd18	17	9	8	8
Bd01	14	10	4	6
Bd02	12	6	7	11
Bd12	8	4	6	2
Bd09	10	5	4	3
Bd05	26	12	13	13
Bd08	6	5	6	5
Bd14	10	7	8	8
Total	126	77	64	62

Appendix 5.4: Allelic richness for 95 individuals of *Bagrus bayad* calculated in FSTAT

Locus	Population			Overall
	ALB	NIG	SEN	
Bd04	15.2	6.7	5.4	13.9
Bd18	8.3	7.3	5.7	10.9
Bd01	9.3	3.9	3.5	8.0
Bd02	5.6	7.0	9.6	8.7
Bd12	4.0	5.9	1.4	6.2
Bd09	4.8	4.0	2.3	7.7
Bd05	10.0	11.1	11.9	16.6
Bd08	5.0	5.8	4.7	5.3
Bd14	6.2	8.0	7.1	8.7

ALB, Lake Albert, NIG, Lake Kainji, SEN, Senegal River