



HAL
open science

Contribution to the knowledge of the populations of Nile tilapia (*Oreochromis niloticus*) inhabiting extreme conditions of temperature and alkalinity

Titus Chemandwa Ndiwa

► To cite this version:

Titus Chemandwa Ndiwa. Contribution to the knowledge of the populations of Nile tilapia (*Oreochromis niloticus*) inhabiting extreme conditions of temperature and alkalinity. Biodiversity. Université Montpellier II - Sciences et Techniques du Languedoc, 2014. English. NNT : 2014MON20114 . tel-01542402

HAL Id: tel-01542402

<https://theses.hal.science/tel-01542402v1>

Submitted on 19 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par **UNIVERSITE MONTPELLIER 2**

Préparée au sein de l'école doctorale **SIBAGHE**
Et de l'unité de recherche **ISE-M**

Spécialité : Biologie de l'Evolution et Ecologie

Présentée par Titus Chemandwa Ndiwa

**Contribution à la connaissance des populations
du tilapia du Nil (*Oreochromis niloticus*) vivant
dans des conditions extrêmes de température et
d'alcalinité**

Soutenue le 19/12/2014 devant le jury composé de

Mme Janice Briton Davidian, Directeur de Recherche, CNRS
Mme Dorothy Wanja Nyingi, Researcher, NMK
M. Paul Alibert, MC, Université de Bourgogne
M. Didier Paugy, Directeur de Recherche, IRD
M. Jean-François Agnèse, Directeur de Recherche, IRD

Invitée
Co-encadrant
Rapporteur
Président
Dir. de thèse

Table of contents

Table of contents	1
Acronyms and abbreviations	4
Acknowledgement.....	6
Résumé.....	7
Abstract	8
1. General introduction.....	9
1.1 Context of the research.....	9
1.2 Natural populations of Nile tilapia	9
2. References	16
(I) An important natural genetic resource of <i>Oreochromis niloticus</i> (Linnaeus, 1758) threatened by aquaculture activities in Lobo drainage, Kenya	21
Abstract	21
1. Introduction	22
2. Materials and methods.....	24
2.1 Sampling design	24
2.2 DNA extraction, sequencing and microsatellite analysis	25
2.3 Data analysis.....	26
2.3.1 Sequence analyses	26
2.3.2 Genotype analyses	26
3. Results	27
3.1 mtDNA differentiation	27
3.2 Microsatellite analysis	28
4. Discussion	31
5. Acknowledgements	36
6. References	36
(II) Population differences in the SNPs associated to sex found in the <i>amh</i> gene of Nile tilapia and identification of sex-reversed fish	58
Abstract	58

1. Introduction	59
2. Materials and Methods	62
2.1 Fish sampled and phenotypic sex identification.....	62
2.2 Genomic DNA extraction, PCR amplification and sequencing	63
2.3 Sequence analysis to identify SNP differences	64
2.4 Use of X and Y markers to identify genotypic sex.....	64
3. Results and Discussion.....	65
3.1 SNP identified on the cyp19a1a promoter region	66
3.2 SNPs identified on the wt1b gene between exon 5 and 7.....	67
3.3 High amount of SNPs identified on the amh gene	68
3.4 Finding of SNPs in the amh gene associated with the phenotypic sex.....	68
3.5 FCA analysis of SNPs	70
3.6 Use of X and Y markers to identify sex-reversal individuals.....	70
4. Acknowledgements	73
5. References	74
(III) Morphological variations of wild populations of Nile tilapia living in extreme environmental conditions in the Kenyan Rift-Valley.....	103
Abstract	103
1. Introduction	103
2. Materials and methods.....	107
2.1 Study areas	107
2.2 Collection of samples	107
2.3 Geometric morphometrics.....	107
2.4 Data analysis.....	108
3. Results	109
3.1 Measurement error	109
3.2 Size variation.....	109
3.3 Shape variation (pca).....	109

3.4 Differences between populations (and eventually habitats)	109
4. Discussion	111
5. Acknowledgement.....	114
6. References	115
(IV) General discussion.....	131
1.1 New genetic resources under threat, the case of Loboï Swamp hot spring populations.....	132
1.2 Sex determination in <i>O. niloticus</i> : The usefulness of the Loboï Hot Springs populations to disentangle the complex role of temperature.....	135
1.3 Morphological differentiation in <i>O. niloticus</i> : Genetic versus plasticity, what do we learn from populations living in extreme environments?.....	138
1.4 Future research perspective	139
2. References	141
Résumé détaillé de la thèse	144
Variations morphologiques au sein des populations naturelles de tilapia du Nil vivants dans des conditions environnementales extrêmes dans la vallée du Rift du Kenya.....	147
Des populations naturelles, ressources génétiques importantes d' <i>Oreochromis niloticus</i> (Linnaeus, 1758), mises en danger par des activités aquacoles dans le bassin de la Loboï au Kenya.....	150
Les SNP du gene amh liés aux sexes sont différents suivant les populations de Tilapia du Nil.....	153
References	155

Acronyms and abbreviations

AMH – Anti-Müllerian Hormone

ANOVA – Analysis of Variance

BIC – Bayesian Information Criterion

CeMEB – Centre Méditerranéen de l'Environnement et de la Biodiversité

DM – Distance Method

ESD – Environmental Sex Determination

ESP – Economic Stimulus Programme

FCA – Factorial Correspondence Analysis

GIFT – Genetically Improved Farmed Tilapia

GSD – Genetic Sex Determination

IRD – Institut de Recherche pour le Développement

ISEM – Institut des Sciences de l'Evolution de Montpellier

KENWEB – Kenya Wetlands Biodiversity research group

KWS – Kenya Wildlife Service

LDA – Linear Discriminant Analysis

MANCOVA – Multivariate Analysis of Covariance

MCMC – Markov Chain Monte Carlo

ME – Measurement Error

MEGA – Molecular Evolutionary Genetics Analysis

ML – Maximum Likelihood

MRAC – Musée Royal de l'Afrique Centrale

NMK – National Museums of Kenya

PCA – Principal Component Analysis

SL – Standard Length

SNP – Single Nucleotide Polymorphism

TSD – Temperature Sex Determination

USD – United States Dollar

Acknowledgement

I am greatly indebted to my thesis supervisors: Dr. Jean-Françoise Agnès, Dr. Dorothy Wanja Nyingi, Dr. Helena D’Cotta, Dr. Jean-Françoise Baroiller and Dr. Julian Claude for their guidance and support that saw successful completion of this work. I am also grateful to Erick Desmarais, Joseph Gathua, Frederique Cequeira, Elodie Pepey and Gilbert Kosgei for their assistance in data collection/laboratory work. My sincere thanks to Dr. Didier Paugy, Dr. Paul Alibert and Prof. Janice Briton Davidian for their invaluable comments and insightful suggestions that contributed to improvement of this work.

My deepest appreciation goes to Ministere des Affaires Etrangères (MAE) and Institut de Recherche Pour le Développement (IRD) for financing my studies, Institut des Science de l’Evolution Montpellier (ISEM) and CIRAD for hosting me in their laboratories, and National Museums of Kenya and Kenya Wetlands Biodiversity Research team (KENWEB) for facilitating my field work studies. I also acknowledge the University of Montpellier 2 (UM2) for their support during my three years study in the institution.

Lastly, I thank my wife Tabitha Magero and my son David Ndiwa for their patience and understanding during the entire period of my studies.

Résumé

Les ressources génétiques naturelles du Tilapia du Nil (*Oreochromis niloticus*) se trouvent en Afrique. Ces ressources sont menacées en raison de modifications des habitats naturels des poissons et des introductions incontrôlées d'espèces ou de souches exotiques. Les populations de tilapia du Nil qui vivent dans des habitats extrêmes sont emblématiques de cette situation. Elles représentent des ressources génétiques originales et potentiellement utiles pour l'aquaculture. Cependant, les populations sont menacées soit par des modifications fortes de l'environnement soit par l'introduction d'espèces exotiques dans leur habitat. La caractérisation de ces populations constitue la première étape de leur protection et par conséquent leur utilisation en aquaculture. Dans notre étude, nous nous sommes concentrés sur deux populations différentes de tilapia du Nil. L'une d'entre elles vit dans le lac alcalin 'Crocodile Lake' (10 590 S / cm, pH 10) qui est un lac de cratère situé dans Central Island, au milieu du lac Turkana. La deuxième population (ou groupe de population) vit dans les sources chaudes (Hot Springs) du marais de Loboï (Loboï Swamp) près du lac Bogoria au Kenya. Ces poissons vivent dans une eau caractérisée par des températures élevées, autour de 36 ° c. Toutes les populations ci-dessus (Crocodile Lake et Loboï swamp Hot Springs) ont connu ou connaissent un certain nombre de pressions sélectives de la part de leurs environnements difficiles. Pour le lac Crocodile, les poissons ont certainement du trouver un moyen d'excréter leurs déchets azotés car à pH 10, l'excrétion n'est pas possible par simple diffusion. Pour les populations de sources chaudes, la plupart des individus devraient être des mâles car les hautes températures sont connues pour induire une masculinisation chez *O. niloticus*. Cette population doit avoir accumulé des mutations nécessaires pour lui permettre de surmonter les effets masculinisants des températures élevées. Pour étudier ces populations, nous avons utilisé la morphométrie géométrique et des marqueurs génétiques (microsatellites et 16 ADNmt) et nous les avons comparés à d'autres populations proches de la région. En outre, trois gènes liés au sexe (Cyp19a, Wt1b, AMH) ont été analysés en utilisant des marqueurs SNP dans trois populations de tilapia du Nil habitant les sources chaudes du marais de Loboï, et nous les avons comparés à ceux que l'on observe chez huit autres populations de l'Afrique de l'Est, la région soudano-sahélienne et éthiopiennes. Des différences morphologiques significatives ont été observées entre toutes les populations étudiées, y compris entre les trois populations voisines du marais Loboï, et entre les deux populations génétiquement liées du Crocodile Lake et du lac Turkana. Nous concluons de cette analyse que les différences morphologiques observées peuvent avoir comme origine pour une part des différences génétiques et pour une autre part des facteurs environnementaux. De même, toutes les populations étudiées sont génétiquement différenciées, et nous avons montré que les populations du marais Loboï et celle du lac Baringo ont été introgressées par des gènes de *O. leucostictus*. Les analyses des gènes liés au sexe ont révélé que le gène AMH est un gène candidat pour la détermination du sexe chez le tilapia du Nil, avec 12 SNPs montrant de fortes associations avec le sexe phénotypique des individus. Néanmoins, il n'existe pas de modèle général de la détermination du sexe, il semble plutôt que les mécanismes de détermination du sexe sont différents suivant les populations de cette espèce et qu'il n'existe pas de mécanisme unique pour l'espèce entière.

Abstract

Nile tilapia (*Oreochromis niloticus*) natural genetic resources are found in Africa. These resources are threatened due to modifications of the natural habitats of fishes and uncontrolled introductions of alien species or strains. Nile tilapia populations living in extreme habitats are emblematic of this situation. They represent original and potentially useful genetic resources for Aquaculture. However, the populations are threatened either by strong modifications or introduction of alien species in their habitats. Characterizing these populations constitutes the very first step of their protection and consequently their utilization in aquaculture. In our study we concentrated on two different populations of Nile Tilapia. One living in the alkaline Crocodile Lake (10,590 $\mu\text{S}/\text{cm}$, 10 pH) which is a Crater Lake located in the central Island of Lake Turkana. The second group of population inhabit the hot springs of Loboï Swamp near Lake Bogoria in Kenya. These fish are living in water characterized by high temperatures, around 36°C. All above populations (Crocodile Lake and Loboï Swamp Hot springs) may have experienced some selective pressures to cope with their challenging environments. For Crocodile Lake, fish may have found a way to excrete their nitrogenous wastes because at pH 10, excretion is not possible by simple diffusion. For the hot spring populations, most individuals should have been males as high temperature is known to induce masculinization in *O. niloticus*. This population may have accumulated adequate mutations to enable them overcome masculinizing effects of high water temperature. To study these populations we used geometric morphometrics and genetic markers (16 microsatellites and mtDNA) and compared them with other related populations from the region. In addition, three sex-linked genes (*Cyp19a*, *Wt1b*, *amh*) were analysed using SNP markers in three populations of Nile tilapia inhabiting hot springs of Loboï Swamp, and compared them to eight other populations from East Africa, Sudano-Sahelian and Ethiopian regions. Significant morphological differences were observed in all populations studied, including three closely related populations of Loboï Swamp, and two genetically related populations from Lake Turkana basin. Both genetic differences and environmental factors were responsible for the observed morphological differences. Similarly, all studied populations were genetically differentiated, and we demonstrated that populations from Loboï Swamp and Lake Baringo have been introgressed by *O. leucostictus* genes. Analyses of the sex-linked clustered revealed that *amh* gene is a candidate gene for sex determination in Nile tilapia, with 12 SNPs showing strong associations to phenotypic sex. Nevertheless there is no general pattern of sex determination, rather it seems that sex determination mechanisms are different with respect to populations, but is not characteristic or unique for the entire species.

Key words: Nile tilapia, mtDNA, microsatellites, SNP, natural population, sex determination, morphology, genetic variability

1. General introduction

1.1 Context of the research

Nile tilapia (*Oreochromis niloticus*, Linnaeus 1758) is a member of the family *Cichlidae* (3000 species), which are mainly distributed within tropical and sub-tropical regions of the world. The Nile tilapia has been a major source of proteins in developing and emerging countries, but is now also a prime fish commodity in developed countries (Fitzsimmons, 2001). As a result, the fish has successfully been introduced and cultured in more than 100 countries around the world. From the pioneer work of Trewavas (1983), a number of studies have tried to document the variety of natural populations of Nile tilapia, especially their genetic characterisation (Rognon *et al.*, 1996; Agnèse *et al.*, 1997; Nyingi *et al.*, 2009; Bezault *et al.*, 2011). However, knowledge on their genetic structure still remains insufficient. In addition, some natural populations are poorly known (Lake Tanganyika for example) or unknown as evidenced by the recent discovery of a natural population in Kenya (Nyingi *et al.*, 2009).

Most if not all of the genetic natural resources of Nile tilapia exists in Africa and plays a major role in improvement of its cultured strains in other parts of the world. For example, the Genetically Improved Farmed Tilapia (GIFT) strain was developed from founder populations of Nile tilapia obtained from both wild populations and cultured strains (Eknath *et al.*, 1995; Eknath and Acosta, 1998). They used wild Nile tilapia germplasm acquired from Egypt, Ghana, Kenya and Senegal, and farmed strains of germplasm derived from Israel, Singapore, Taiwan and Thailand (Gupta and Acosta, 2004).

Even though Nile tilapia is an invasive species in many river drainages and has sometimes been accused of being behind the disappearance of some fish species (Ogotu-Ohwayo, 1990), its natural populations are also under threat (Nyingi *et al.*, 2007; Nyingi and Agnèse 2011). The main threat is certainly hybridization with other species or populations. This situation is becoming even more critical, especially after the recent introduction of Genetically Improved Farmed (GIFT) Strains of *O. niloticus* into Africa (Ansah *et al.*, 2014).

1.2 Natural populations of Nile tilapia

Trewavas (1983) was the first to study natural populations of Nile tilapia. Based on the study of their morphological characteristics, she described seven subspecies: *O. n. niloticus* from

West-Africa and River Nile, *O. n. baringoensis* from Lake Baringo, *O. n. sugutae* from River Suguta (Kenya), *O. n. eduardianus* from Lakes Edward, Albert, George and Tanganyika, *O. n. vulcani* from Lake Turkana, *O. n. cancellatus* from River Awash and Lake Tana, and *O. n. filoa* from hot springs in the Awash system.

This work has served as a reference for many years, and even now it is still considered as an important work on Nile and other tilapia species. However, a careful analysis of data provided on differentiation of the subspecies of the Nile tilapia reveals that Trewavas (1983) did not rely on large differentiations and also sometimes is very subjective. For instance, she used genital papilla and proteins of serum as differentiating characteristics between *O. n. filoa* and *O. n. vulcani*, while she does not refer to these traits in the other subspecies. In several instances, she used dorsal fin formula (number of spines and rays of the fin) as a differentiating trait and yet these characteristics overlap largely between the subspecies (see table 35 page 144-145 in Trewavas, 1983). This is probably the reason why the author did not provide an identification key to the subspecies.

More problematic is the absence of any consideration or definition of what is or what should be a subspecies. Are they equivalent to Evolutionary Units (Ryder, 1986; Waples, 1991) or do they simply represent morphologically differentiated populations without any evolutionary significance? It is thus not easy to understand what was the “subspecies concept” adopted by this author. For example, *O. n. cancellatus* and *O. n. filoa* are present in the same drainage, the Awash River. This situation obviously would better fit into a classification of two entirely different species or two different eco-phenotypes of the same species. The same situation is found with *O. n. eduardianus* in Lake Edward, which is in contact with *O. n. niloticus* in the Nile River. Finally it appeared that the author observed some amount of variability between populations and found it “convenient to provide them with names, if only to preserve their identity for any further experimenter or fish manager” (citation from page 141, last sentence).

Hence there has been a need to carry out phenotypic studies to investigate morphological differentiation in Nile tilapia. New and more powerful techniques for morphological studies have been employed in fish studies including landmark-based geometric morphometric methods (Bookstein, 1991). With geometric morphometric, shape variations in different taxa or even populations can efficiently be compared. This technique involves collection of two or three-dimensional coordinates of biologically definable landmarks, elimination of non-shape variations through superimposition and performing statistical analysis of the resulting data on

shape variables (Dean *et al.* 2004). The technique also generates images of centroid shape, which represents the form of the sample, and can be used to infer morphological differences between different groups of organisms being studied.

The technique has been successfully used by Firmat *et al.* (2011, 2012) to differentiate 15 populations of Mozambique tilapia (*Oreochromis mossambicus*) from its native range in Mozambique, and feral populations in New Caledonia and Guadeloupe after their acclimatization to new environments. In Tanganyika, a study to differentiate four wild populations of cichlid *Tropheus moorii* from their pond raised F1 offspring was carried out using geometric morphometrics (Kerschbaumer and Sturmbauer, 2011). The wild populations of *T. moorii* were previously presumed not to differ in their morphology but only in coloration. However, the authors were able to observe clear morphological differentiations between these populations hence demonstrating that geometric morphometric analysis is able to differentiate even closely related populations.

In our study, we decided to employ the use of geometric morphometrics to determine morphological differences between some natural populations of *O. niloticus* in Kenya based on the fact that previous studies by traditional morphometrics had failed to reveal any discrete morphological variations. The main goals were: 1) to demonstrate that this technique can be a useful tool in assessing morphological diversity in this species at a large scale; 2) to investigate the morphological plasticity by comparing two populations of *O. niloticus* with same genetic background but living under different environmental conditions; and 3) to evaluate the impact of temperature on morphology.

More recently, the use of genetic techniques (allozymes, mtDNA sequences, microsatellites) has enabled more detailed and reliable description of different populations of the Nile tilapia. For example, Agnès *et al.* (1997) studied 17 natural populations representing all seven subspecies described by Trewavas (1983), and clustered them into three major groups; West African populations, Ethiopian Rift-Valley populations, and Nile and Kenyan Rift-Valley populations.

Other genetic studies have been conducted on wild populations of Nile tilapia described by Trewavas (1983). Rognon *et al.*, (1996) carried out studies on different populations of Nile tilapia and found clear genetic differentiation between western (Niger, Senegal and Volta populations) and eastern (Manzala and Turkana) populations. To further assess their geographical differentiation, Rognon *et al.* (1997) studied the same populations using

mtDNA. These two studies provided evidence that morphological similarities between the Nile River and West African populations reflects convergence, common ancestral morphology or non-genetic environmental relatedness rather than derived phylogenetic relatedness. This implies that *O. niloticus niloticus* as defined by Trewavas (1983) is not valid. The same authors (2003) studied gene flow between a wild population of *O. niloticus* and *O. aureus*. Other recent studies were carried out by Hassanien and Gilbey (2005), Nyingi and Agnès (2007), Nyingi *et al.* (2009), Angienda *et al.* (2011) and Bezault *et al.* (2011) using mitochondrial and/or microsatellite markers. All these studies, though not exhaustive considering the Rivers and Lakes having natural populations of *O. niloticus*, constitute a valuable documentation of genetic resources of Nile tilapia. Additionally, comparison between these studies and earlier studies by Trewavas's classification have shown discrepancies, hence the need for further phylogenetic relationships studies especially on the East African populations of Nile tilapia. Based on these laid down foundation, this study seeks to characterize some of *O. niloticus* populations found in East Africa using mtDNA and 16 microsatellite markers.

More recently, a new population of *O. niloticus* was discovered in Lake Bogoria Hotel Spring in Kenya (Nyingi *et al.*, 2009) (Fig. 1). This hot spring (~36 °C) flows into Lobo Swamp, which is thought to have formed during a period characterised by high rainfall and high water levels in East African lakes at about 700 years BP (Mohammed *et al.* 1995; Verschuren *et al.*, 2000). At this period, the swamp is thought to have been connected to L. Baringo. However, important climatic shifts that reduced rainfall and increased evaporation (Hay, 1976; Cerling, 1979), together with landform changes in the region led to their isolation (Ashley *et al.* 2004). Currently, the swamp is endoreic, with most of its waters being used to irrigate surrounding agricultural farms. Based on genetic evidences, the population was considered as natural and not introduced. It is characterized by new (unknown) private genetic markers (microsatellites and mtDNA) that make this population close, but differentiated, to the Lake Baringo population (Nyingi *et al.*, 2009). Moreover, Nyingi *et al.*, 2007 demonstrated that *O. niloticus* from Lake Baringo had been recently introgressed by mtDNA of *O. leucostictus*. This latter species is not autochthonous in Kenya (native to Lake Albert in Uganda) but was first introduced in Lake Naivasha (Kenya) from the ponds around Lake Victoria in 1956 (Siddiqui, 1977a, 1979). There are two hypotheses to explain how *O. leucostictus* may have been introduced into Lake Baringo, either to counterbalance the decrease of natural Nile tilapia catch in the lake, or from individuals that may have escaped from local aquaculture farms

within the Lake's catchment. In this first study it appeared that the mtDNA introgression of *O. leucostictus* genes within *O. niloticus* was not accompanied by any nuclear introgression. Two other hot springs exist in this swamp, and host Nile tilapia populations, which have not as yet been investigated.

However, due to increased developments in aquaculture projects within the Lakes Baringo and Bogoria drainages (Mumba, 2008; Nyingi and Agnèse, 2011; The Fish Site, 2011), it is necessary to carry out a more complete survey of the natural populations from the Lake Baringo and from the Loboï swamp hot springs in order to characterize any hybridization events.

Therefore, in order to better genetically characterize the new hot spring populations and to evaluate the introgression by *O. leucostictus* genes, we studied mtDNA sequences (Dloop) and 16 microsatellites in the three Loboï hot spring populations, Lake Baringo population as well as in other natural populations of Nile tilapia in the region.

This study is particularly important because these populations represent important genetic resources for the species. In Kenya, the natural populations of Nile tilapia within the Rift-Valley represent a large portion of the species genetic resources. The most important populations among these genetic resources are the three populations of Loboï Swamp. In order to understand and appreciate the usefulness, it is necessary to consider how the culture of *O. niloticus* is undertaken.

In spite of all the favourable qualities that Nile tilapia offers as a culture fish and the technological advances in its aquaculture, the viability of this commercial production is greatly affected by its prolific breeding which frequently results to overcrowding and stunted growth in the ponds (Hickling, 1960). This led to the introduction of all male mono-sex cultures with the aim of improving productivity, and also to benefit from the faster growth of males as compared to females (Mélard, 1986; Toguyéni *et al.*, 1996; Baras and Mélard, 1997). Additionally, rearing of male mono-sex populations is more advantageous compared to females as the former can allocate more energy towards growth (instead of reproduction) thus increasing productivity (Oldfield, 2005). As a result, several methods of producing monosex populations have been established including: manual sexing, which involves separating males from females; hybridization to produce male monosex populations, hormonal treatment of fry with synthetic male hormone (17 alpha methyltestosterone); genetic manipulation to produce YY 'supermales' or treatment of fry with high temperatures ($\approx 36^{\circ}\text{C}$) (Hicklins, 1960; Mair

et al. 1997; Shelton, 2002; Baroiller *et al.* 1996a; Baras *et al.* 2001). Because hormonal sex reversal is by far the most efficient and the cheapest treatment, it is the main technique used by the farmers to produce male monosex populations. However, there are growing concerns with the elimination of 17alpha-methyltestosterone and its metabolites from the treated fry, their presence in waters (ponds, rivers) or in the soil, their potential effects on wild fish sex differentiation (masculinisation) and/or gametogenesis as well as their effects on other aquatic organisms and finally their putative toxicity. This is why steroid treatments are not allowed in the EU for direct production of monosex populations, and more and more unpopular for consumers in the countries where its use is still allowed. In the framework of a sustainable aquaculture, monosex populations should be produced through environmentally-friendly procedures. In this respect, the most promising approaches are probably the use of YY males or the use of temperature treatments (Baroiller *et al.*, 1995; Baroiller and D’Cotta, 2001; Tessema *et al.*, 2006; Bezault *et al.*, 2007; Wessels and Hörstgen-Schwark, 2007; Baroiller *et al.*, 2009). However, YY males failed to systematically produce all-male populations due to the complex sex determination system of tilapia. Besides, the masculinizing effects of high temperatures on a progeny depend on genetic factors brought by the two parents. This thermosensitivity, as a heritable trait, can be selected in the Nile tilapia (Baroiller and D’Cotta, 2001; Tessema *et al.*, 2006; Bezault *et al.*, 2007; Wessels and Hörstgen-Schwark, 2007; Baroiller *et al.*, 2009). However, because sex determination combines both genetic and environmental factors, it is still difficult to develop these sustainable approaches for sex control.

Most of studies on sex determination in the species have been conducted on domestic stocks used for aquaculture (Bezault *et al.* 2007). Based on observations that sex determination systems can be very diverse even between closely related taxa (Chevassus, 1998), there is a need to carry out sex determination studies on wild populations of Nile tilapia under their natural environments in order to provide a better understanding on temperature-dependent sex determination mechanisms in Nile tilapia.

Bezault *et al.*, (2007) carried out sex determination studies on wild populations of Nile tilapia adapted to three different thermal regimes (extreme stable cold temperatures from Lake Koka, extreme stable warm temperatures from Metahara, and a highly variable environment in Kpandu, Lake Volta). The progenies of these populations were also reared under controlled laboratory conditions (27 °C and 36 °C). Progenies from these three populations exhibited a clear thermosensitivity with large variations in the temperature-induced sex ratios depending

on the parents but also on the populations (Bezault *et al.*, 2007). This study was the first to demonstrate that both genetic and environmental factors are also involved in the sex determination of natural populations of Nile tilapia; it also highlights the importance of field investigations to better understand sex determination in tilapia.

In this regard, the Lobo Hot spring populations become very important. Normally, at temperatures above 32°C (as is the case in hot springs with temperature at 36°C) sex ratios have been observed to be biased towards males. Obviously, this has not been the case in the Lobo Swamp hot springs where the sex ratio is close to 1 male for 1 female. It is then very likely that this population “have found a way” to overcome the temperature problem, a fact that therefore presents a great opportunity to study the mechanisms by which temperature acts on sex differentiation.

Such wild progenies where temperature influences have been probably eliminated (i.e which might be the case for the Kenyan hot springs) will be the best resources for the selection of tilapia strains that can be used for a strict genetic sex control. Conversely, wild populations for which high temperatures can strongly change the sex-ratio towards males are potential resources for selection of tilapias using temperature as a means to produce all-male progenies. In both cases, further study and characterisation of wild populations is therefore essential.

Temperature and genetic influences on sex may be the result of allelic variations on sex-differentiating genes. We therefore studied Single Nucleotide Polymorphism (SNP) in three major genes (anti-Mullerian hormone, *amh*; aromatase, *Cyp19a*; Wilm’s tumor, *Wt1b*) that have previously been linked to sex determination in *O. niloticus* in order to understand the adaptations of these populations against the masculinizing effect of the elevated temperatures.

2. References

- Agnèse, J. F., Adepo-Gourène, B., Abban, E. K., Fermon, E., 1997. Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity*, 79: 88-96.
- Angienda, P. O., Lee, H. J., Elmer, K. R., Abila, R., Waindi, E. N. and Meyer, A., 2010. Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. *Conserv genet*, 12: 243-255.
- Ansah, Y. B., Frimpong, E. A. and Hallerman, E. M., 2014. Genetically-Improved Tilapia Strains in Africa: Potential Benefits and Negative Impacts. *Sustainability*, 6:3697-3721.
- Ashley, G. M., Maitima Mworira, J., Muasya, A. M., Owen, R. B., Driese, S. G., Hover, V. C., Renaut, R. W., Goman, M. F., Mathai, S. and Blatt, S. H., 2004. Sedimentation and recent history of a fresh water wetland in a semi-arid environment: Loboï Swamp, Kenya, East Africa. *Sedimentology*, 51: 1301-1321.
- Baras, E. and Mélard, C., 1997. Individual growth patterns of juvenile Nile tilapia *Oreochromis niloticus* (L.): emergence and dynamics of sexual growth dimorphism. In: Fitzsimmons, K. (Ed.), *Tilapia Aquaculture*. Northeast Regional Agricultural Engineering Service 106, pp. 169–177.
- Baras, E., Jacobs, B. and Mélard, C., 2001. Effects of water temperature on survival, growth and phenotypic sex of mixed (XX-XY) progenies of Nile tilapia *Oreochromis niloticus*. *Aquaculture*, 192: 187-199.
- Baroiller, J. F. and D’Cotta, H., 2001. Environment and sex determination in farmed fish. *Comp. Biochem. Physiol. C: Toxicol. Pharmacol.* 130: 399-409.
- Baroiller, J. F., Clota, F. and Geraz, E., 1995b. Temperature sex determination in two tilapias species, *Oreochromis niloticus* and the red tilapia (Red Florida Strain): effect of high or low temperatures. In: Goetz, F. W., Thomas, P. (Eds), *Proceedings of the 5th International Symposium on the reproduction Physiology of Fish*. The University of Texas, Austin, Texas, USA, pp. 158-160. July 2-8, 1995.
- Baroiller, J. F., D’Cotta, H., Bezault, E., Wessels, S., and Hoerstgen-Schwark, G., 2009. Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology, Part A* 153: 30-38.
- Baroiller, J-F, Jalabert, B., 1989. Contribution of research in reproductive physiology to the culture of tilapias, *Aquat. Living Resour.* 2:105–116.
- Baroiller, J-F., Fostier, A., Cauty, C., Rognon, X. and Jalabert, B., 1996a. Effect of high rearing temperature on the sex ratio of progeny from sex-reversed males of *Oreochromis niloticus*. In: Pullin, R.S.V., Lazard, J., Legendre, M., Amon kothias, J.B., Pauly, D. (Eds.), *Proceedings of the Third Symposium on tilapia in Aquaculture*. ICLARM Conference Proceedings, vol. 41, pp. 246–256.

Beardmore, J. A., Mair, G. C., Lewis, R. I., 2001. Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture*, 197: 283–301.

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

Bezault, E., Clota, F., Derivaz, M., Chevassus, B. and Baroiller, J. F., 2007. Sex determination and temperature induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture*, 272S1: S3-S16.

Bookstein, F. L., 1991. Morphometric tools for landmark data: geometry and biology. New York, Cambridge University Press.

Cerling, T. E., 1979. Palaeochemistry of Plio-Pleistocene Lake Turkana, Kenya. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 27: 247-285.

Chevassus, B., 1998. Modification du phenotype sexuel et du mode de reproduction chez les poissons Salmonidés : inversion sexuelle hormonale, gynogenèse, hybridation interspécifique et polyploidation. Université Paris Sud, Orsay, France, Pp. 162.

Dean, C., Adams, F., Rohlf, J. and Slice, D. E., 2004. Geometric morphometrics: Ten years of progress following the Revolution Italian Journal of Zoology, 71:1, 5-16

Eknath, A. E. and Acosta, B., 1998. Genetic improvement of farmed tilapia project. Final report (1988-1997). ICLARM, Manila.

Eknath, A. E., 1995. Managing aquatic genetic resources. Management example 4. The Nile tilapia. In: Thorpe, J., Gall, G., Lannan, J., Nash, C. (eds) Conservation of fish and shell fish resources: Managing diversity, pp. 176-194. Academic Press, London.

Firmat, C., 2011. Hybridation et goulots d'étranglement induits par l'activité humaine : génétique des populations, morphométrie et parasitologie appliquée au tilapia envahit et envahissant *Oreochromis mossambicus* (Teleostei, Cichlidae). Université de Bourgogne, France. Pp 207.

Firmat, C., Schliewen, U. L., Losseau, M. and Alibert, P., 2012. Body shape differentiation at global and local geographic scale in the invasive cichlid *Oreochromis mossambicus*. *Biological Journal of Linnean Society*, 105: 369-381.

Fitzsimmons, K., 2001. Environmental and conservation issues in tilapia aquaculture, pages 128-131. In: R. Subasinghe and T. Singh (eds.), *Tilapia: Production, Marketing, and Technological Developments*. FAO Infofish, Kuala Lumpur, Malaysia.

Gupta, M. V. and Acosta, B. O., 2004. From drawing board to dining table: The success story of the GIFT project. *NAGA, World Fish Center Quarterly Vol 27 No 3&4*.

- Hassanien, H. A. and Gilbey, J., 2005. Genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*) revealed by DNA microsatellites. *Aquac. Res.* 36: 1450-1457.
- Hay, R. L., 1976. *Geology of Olduvai gorge*. University of California Press, Berkeley. Pp 203.
- Hichling, C. F., 1960. The Malacca Tilapia hybrids: *Journal of Genetics* 57(1): 1-10.
- Hickling, C., 1960. The Malacca tilapia hybrids. *Progressive Fish-Culturalist.* 57: 1-10.
- Kerschbaumer, M. K. and Sturmbauer, C., 2011. The utility of geometric morphometrics to elucidate pathways of cichlids of cichlid evolution. *Int. J. Evol.Biol.* Article ID290245: 8pp.
- Langerhans, R. B., Layman, C. A., Shokrollahi, A. M., DeWitt, T. J., 2004. Predator-driven phenotypic diversification in *Gambusia affinis*. *Evolution*, 58: 2305-18.
- Mair, G. C., Abucay, J. S., Skibinski, D. O. F., Abella, T. A. and Beardmore, J. A., 1997. Genetic manipulation of sex ratio for the large-scale production of all-male tilapia, *Oreochromis niloticus*. *Canadian Journal of Fisheries and Aquatic Science*, 54: 396-404.
- Mélard, C., 1986. Recherche sur la biologie d'*Oreochromis (Tilapia) niloticus* L. (Pisces : Cichlidae) en élevage expérimental : reproduction, croissance, bioénergétique. *Cahiers d'Ethologie Appliquée*, 6 : 1-224.
- Mohammed, M. U., Bonnefille, R. and Johnson, T. C., 1995. Pollen and isotopic records in Late Holocene sediments from Lake Turkana, Kenya. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 119: 371-383.
- Mumba, M., 2008. Adapting to climate change: lessons from Lake Bogoria catchment, Kenya. Programme report No. 3/08.
- Nyingi, D. W. and Agnèse, J-F., 2007. Recent introgressive hybridization revealed by exclusive mtDNA transfer from *Oreochromis leucostictus* (Trewavas, 1933) to *Oreochromis niloticus* (Linnaeus, 1758) in Lake Baringo, Kenya. *J. Fish BioLake* 70 (Supplement A), 148–154.
- Nyingi, D. W., Agnèse, J-F., 2011. Tilapia in eastern Africa – a friend and foe. Species in the Spotlight. Darwall, W.R.T., Smith, K.G., Allen, D.J., Holland, R.A, Harrison, I.J., and Brooks, E.G.E. editors. *The Diversity of Life in African Freshwaters: Under Water, Under Threat. An analysis of the status and distribution of freshwater species throughout mainland Africa*. Cambridge, United Kingdom and Gland, Switzerland: IUCN. Pp. 76-78.
- Nyingi, D., De-Vos, L., Aman, R. and Agnèse, J. F., 2009. Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) (Cichlidae; Teleostei) in the Loboï Swamp (Kenya). *Aquaculture* 297: 57-63.
- Ogutu-Ohwayo, R., 1990. The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of the introduced species, especially the Nile perch, *Lates niloticus* and the Nile tilapia,

- Oldfield, R. G., 2005. Genetic, abiotic and social influence on sex differentiation in cichlid fishes and the evolution of sequential hermaphroditism. *Fish Fisheries*, 6: 93-110.
- Rognon, X., Andriamanga, M., McAndrew, B., and Guyomard, R., 1996. Allozyme variation in the natural and cultured populations in two tilapia species: *Oreochromis niloticus* and *Tilapia zillii*. *Heredity*, 76: 640-650.
- Ryder, O. A., 1986. Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*, 1: 9-10.
- Shelton, W. L., 2002. Tilapia culture in the 21st century. p.1-20. In: Guerrero, R.D. III and M.R. Guerrero-del Castillo (eds.). *Proceedings of the International Forum on Tilapia Farming in the 21st Century (Tilapia Forum 2002)*, 184p. Philippine Fisheries Association Inc. Los Banos, Laguna, Philippines.
- Siddiqui, A. Q., 1977a. Lake Naivasha (East Africa, Kenya) fishery and its management together with a note on the food habits of fishes. *Biol. Conserv.* 12: 217-228.
- Siddiqui, A. Q., 1979. Changes in fish species composition in Lake Naivasha, Kenya. *Hydrobiologia* 64(2): 131-138.
- Tessema, M., Müller-Belecke, A., Hörstgen-Schwark, G., 2006. Effect of rearing temperatures on the sex ratios of *Oreochromis niloticus* populations. *Aquaculture* 258, 270–277.
- The Fish Site website. Developing a Tilapia breeding station in Kenya: Available: <http://www.thefishsite.com/articles/1186/developing-a-tilapia-breeding-station-in-kenya>. Accessed on 2014 August 19.
- Toguyéni, A., Baroiller, J. F., Fostier, A., Le Bail, P. Y., Kuhn, E. R., Mol, K. A. and Fauconneau, B., 1996. Consequences of food restriction on short term growth variations and plasma circulating hormones in *Oreochromis niloticus* in relation to sex. *General and Comparative Endocrinology* 103: 167-175.
- Trewavas, E., 1983. Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*, British Museum (Natural History), London.
- Verschuren, D., Laird, K. R. and Cumming, B. F., 2000. Rainfall and draught in equatorial east Africa during the past 1,100 years. *Nature* 403: 410-414.
- Waples, R. S., 1991. Pacific salmon, *Onchorynchus* spp., and the definition of ‘species’ under the endangered species act. *Marine Fisheries Reviews*, 53: 11-22.
- Wessels, S., Hörstgen-Schwark, G., 2007. Selection experiments to increase the proportion of males in Nile tilapia (*Oreochromis niloticus*) by means of temperature treatment. *Aquaculture* 272, S80–S87.

Figure

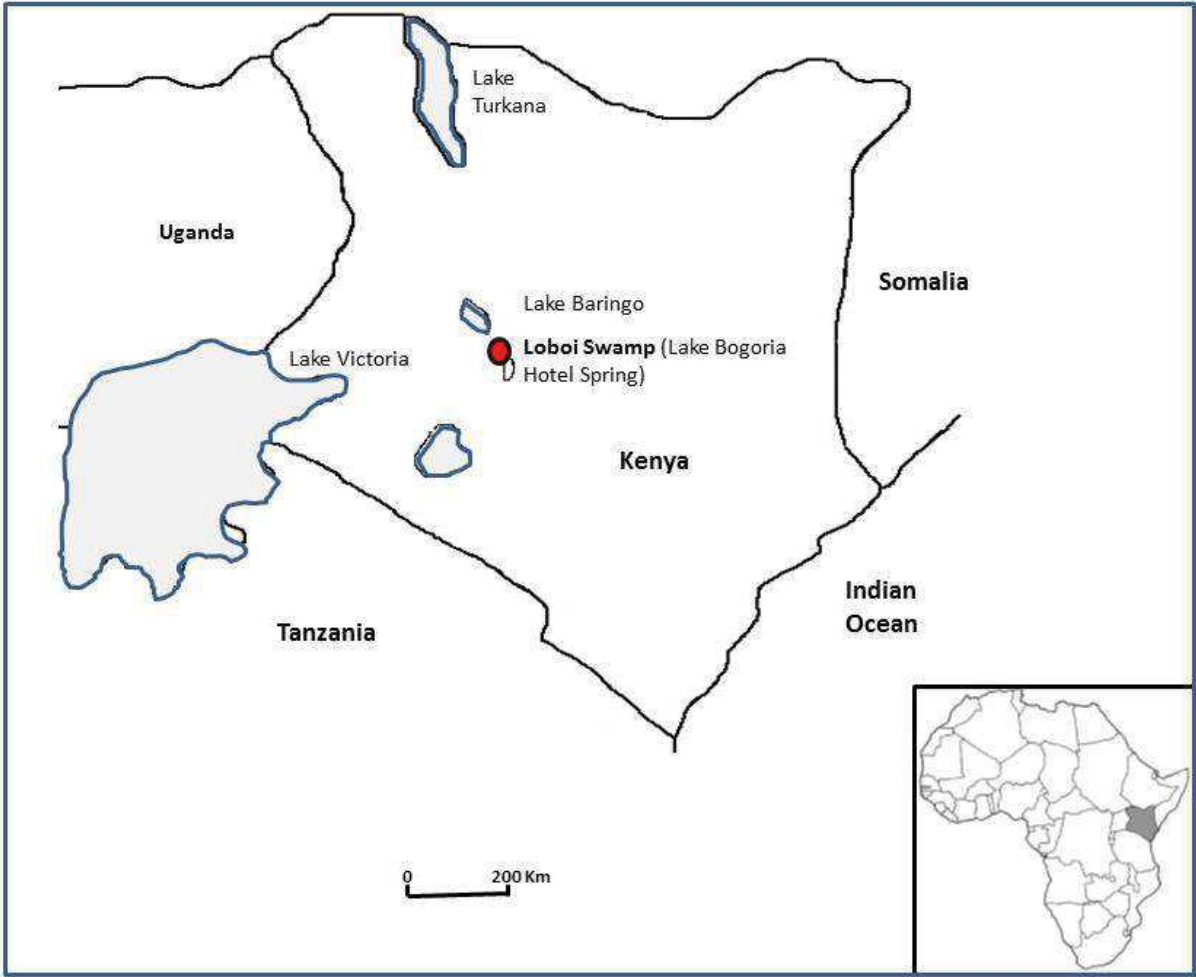


Figure 1: Map showing location of Lake Bogoria Hotel Spring in Lobo Swamp, Kenya

(I) An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Lobo drainage, Kenya

Titus Chemandwa NDIWA^{1,2}, Dorothy Wanja NYINGI^{1*}, Jean-François AGNESE²

¹ Kenya Wetlands Biodiversity Research Group, Ichthyology Section, National Museums of Kenya, Kenya.

² Département Conservation et Domestication, UMR IRD 226 CNRS 5554, Institut des Science de l'Evolution, Université de Montpellier 2, Montpellier, France.

Abstract

The need to improve food security in Africa through culture of tilapias has led to transfer of different species from their natural ranges causing negative impacts on wild fish genetic resources. Lobo swamp in Kenya is fed by three hot springs: Lake Bogoria Hotel, Chelaba and Turtle Springs, hosting natural populations of *Oreochromis niloticus*. The present study aimed at better genetic characterization of these threatened populations. Partial mtDNA sequences of the D-loop region and variations at 16 microsatellite loci were assessed in the three hot spring populations and compared with three other natural populations of *O. niloticus* in the region. Results obtained indicated that the hot spring populations had mitochondrial and nuclear genetic variability similar to or higher than the large closely related populations. This may be attributed to the perennial nature of the hot springs, which do not depend on rainfall but rather receive permanent water supply from deep aquifers. The study also revealed that gene flow between the three different hot spring populations was sufficiently low thus allowing their differentiation. This differentiation was unexpected considering the very close proximity of the springs to each other. It is possible that the swamp creates a barrier to free movement of fish from one spring to the other thereby diminishing gene flow. Finally, the most surprising and worrying results were that the three hot spring populations are introgressed by mtDNA genes of *O. leucostictus*, while microsatellite analysis suggested that some nuclear genes may also have crossed the species barrier. It is very likely that the recent intensification of aquaculture activities in the Lobo drainage may be responsible for these introgressions. Taking into account the importance of these new genetic resources, protection and management actions of the Lobo swamp should be accorded top priority to prevent the loss of these spring populations.

Citation: Ndiwa TC, Nyingi DW, Agnese J-F (2014) An Important Natural Genetic Resource of *Oreochromis niloticus* (Linnaeus, 1758) Threatened by Aquaculture Activities in Lobo Drainage, Kenya. PLoS ONE 9(9): e106972. doi:10.1371/journal.pone.0106972

Editor: Valerio Ketmaier, Institute of Biochemistry and Biology, Germany

Received May 22, 2014; **Accepted** August 4, 2014; **Published** September 15, 2014

Copyright: © 2014 Ndiwa et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its supporting information files. mtDNA sequences are stored at Genbank (accession numbers provided in Table 1).

Funding: TCN received a personal grant from IRD (Institut de Recherche pour le Développement) and MAE (French Ministry of Foreign Affairs). Field work was funded by Kenya Wetlands Research and National Museums of Kenya (NMK). Laboratory work was funded 50% by IRD and 50% by NMK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: dorothyningi@yahoo.com

1. Introduction

Tilapias are some of the most important species for fisheries and aquaculture in Africa. This has contributed to their massive transfer not only within Africa, but to other countries around the world [1,2,3]. These transfers are a major concern because of the invasive nature of tilapia species and their ability to hybridize with local species [3,4,5]. Introgression of alien genes can lead to disruption of specific allele combinations responsible for adaptation of the populations to their environments hence reducing their fitness [6,7].

Within tilapias, *Oreochromis niloticus* (Linnaeus, 1758), the Nile tilapia, is the most economically important species, with a wide natural distribution in Africa. Its natural range covers the entire Nilo-Sudanian province (Senegal to Nile), Ethiopian Rift-Valley province, Kivu province, North Tanganyika Province (Ruzizi) and the Northern part of the East African Rift-Valley [8,9]. Currently, Nile tilapia is cultured in more than 100 countries and its production estimated at 2,790,350 metric tonnes in 2011, and valued at 4.52 billion USD [10]. This species is favoured by fish farmers due to its fast growth rate and its ability to tolerate a wide range of environmental conditions [11]. Resistance to diseases and good consumer acceptance has further promoted its culture worldwide [12].

Seven subspecies of Nile tilapia have been described based on morphological characteristics [8]: *O. n. niloticus* from West-Africa and River Nile, *O. n. baringoensis* from Lake Baringo, *O. n. sugutae* from River Suguta (Kenya), *O. n. eduardianus* from Lakes Edward, Albert,

George and Tanganyika, *O. n. vulcani* from Lake Turkana, *O. n. cancellatus* from River Awash, and Lake Tana and *O. n. filoa* from hot springs in the Awash system. Recently, Nyingi *et al* [5] discovered a natural population from Lake Bogoria Hotel Spring (Kenya). This spring that drains into Loboï swamp is characterized by an elevated water temperature (36°C) and pH ranging from 6.4–6.9 [13]. This population of *O. niloticus* was characterised by 10 private microsatellite alleles and five private mtDNA haplotypes. Both extents of mitochondrial and microsatellite differentiations were in the range of those observed among other naturally occurring discrete populations in the region (Lakes Baringo and Turkana, and River Suguta). These observations indicated that the Lake Bogoria Hotel Spring population was not as earlier hypothesised, introduced from other neighbouring East African natural populations, but represents a previously unknown natural population.

This population offers new opportunities for aquaculture due to various adaptations including its ability to survive in relatively high temperatures (approx. 36°C). The fish may therefore have developed hypoxic resistance mechanisms since dissolved oxygen levels are generally low in warm waters. In addition, special adaptations may also be present that regulate sex determination mechanisms known to involve temperature [14,15]. This population may in this regard potentially offer a model for the study of sex determination in tilapia fishes.

In a first study [5] samples from a single hot spring in the Loboï drainage, the Lake Bogoria Hotel Spring were analysed. At least two other springs with similar ecological conditions exist, each hosting *O. niloticus* population, the Chelaba Spring that is located close to the Lake Bogoria Hotel Spring and the Turtle Spring located at the border of the papyrus marsh of Loboï Swamp. In order to protect these populations from various anthropogenic and environmental threats, there is an urgent need to characterise their natural genetic diversity as a first step of a future action plan for sustainable management. The Loboï Swamp itself has receded in size by over 60% over a short period of 30 years due to expansion of irrigation via a ditch constructed in 1970 [13,16] and conversion of wetlands for agriculture. Another anthropogenic threat is due to the rapid expansion of aquaculture activities in the Rift Valley region enhancing fish transfer from one drainage system to another, and allowing mixing between populations and or species. It has been demonstrated [17], using mtDNA (D-loop) and microsatellite studies, that the *O. niloticus* population from Lake Baringo has been introgressed by *O. leucostictus* from Lake Naivasha. However, even though mtDNA introgression was clearly established, no nuclear introgression was apparent. These introgressions have been attributed to possible introduction of *O. leucostictus* into Lake Baringo to boost tilapia fisheries caused by the decline of *O. niloticus* population by

unsustainable high fishing pressure. Another possible explanation for this may have been the expansion of aquaculture in the region, which has been a continuous threat to wild populations of the Nile tilapia in Kenya. The human population of Lakes Baringo and Bogoria, mainly composed of small-scale crop farmers, have begun to diversify their livelihoods by construction of earthen ponds along the rivers and streams in the region [17,18]. Fish farmers are now breeding different tilapia species with fingerlings from a diversity of sources within East Africa. Most of these ponds are not isolated from streams and wetlands, thus farmed fish can easily escape and hybridize with autochthonous *O. niloticus*.

In order to characterize the three hot spring populations of the Loboï Swamp, we carried out a study of their genetic variation in mtDNA (sequence of partial D-loop region) and microsatellites together with other *O. niloticus* populations from the region: Lake Albert, Lake Turkana, Lake Baringo, River Suguta and *O. leucostictus* from Lake Naivasha.

2. Materials and methods

2.1 Sampling design

Oreochromis niloticus is not an endangered or protected species in Kenya. Seven populations (Figure 1, Table I) were studied and compared, three from the hot springs of the Loboï drainage (Turtle Spring, Chelaba Spring and Lake Bogoria Hotel Spring) (Figures S1-S3), while other populations were from Kenyan Lakes or Rivers (Baringo, Turkana, Suguta). Authorization to collect and sacrifice fish within the protected areas in Lake Turkana was provided by the means of a permit (number KWS/BRM/5001). All fishing and processing of tissue and whole fish specimens was carried out under existing collaborative arrangements with the Kenya Wildlife Service (KWS) and the National Museums of Kenya (NMK) for purposes of biodiversity research. For the other locations outside protected areas, no special authorization was necessary.

As there is no Institutional Animal Care and Use Committee or equivalent animal ethic committee in Kenya, no formal approval was obtained for this work. To minimize suffering of individuals studied, specimens were captured using seine nets and immediately anaesthetized and killed using an overdose of MS-222 (Tricaine-S, Western Chemicals Inc). After death, a fragment of muscle or fin tissue was taken from each specimen and immediately preserved in 95% ethanol for later DNA analyses. Voucher whole fish specimens were fixed in 4% formalin and later preserved in 70% ethanol, and are presently curated at the National

Museums of Kenya (NMK) in Nairobi. Other voucher specimens whose sequences were obtained from Genbank are curated at the Musée Royal de l'Afrique Centrale (MRAC) in Tervuren, Belgium or at the Institut des Sciences de l'Evolution de Montpellier (ISEM), France (Table I).

2.2 DNA extraction, sequencing and microsatellite analysis

Approximately 50 mg of the sample tissue was sheared into fine pieces before being digested at 55 °C overnight using 10 µl proteinase K (10 mM / ml) in 190 µl of an extraction buffer solution (1 M Tris, 0.5 M NaCl₂, 1 % SDS). DNA was then extracted from each sample following the protocol described for genomic DNA extraction for PCR-based techniques [19]. The extracted DNA was suspended in sterile double distilled water and stored in -20 °C until PCR amplification. A 450 bp fragment in the 5' region of the D-loop was amplified in each sample using two primers: 5'-ACCCCTAGCTCCCAAAGCTA-3' (forward) and 5'-CCTGAAGTAGGACCAGATG-3' (reverse). Amplifications were performed in a final volume of 50 µl containing 0.25 mM MgCl₂, 0.2 mM of each dNTP, 1 µM of each primer, 5µl of 10x buffer and 10 units of Taq polymerase (Promega). PCR reaction conditions were: 3 minutes pre-heating at 93 °C followed by 40 cycles of 30 seconds at 93 °C, 30 seconds at 62 °C and 1 minute at 72 °C, with a final elongation phase of 72 °C for five minutes.

A total of 16 microsatellite loci were analysed in this study: PRL1AC and PRL1GT [20], UNH104 (GenBank reference number G12257), UNH115 (G12268), UNH129 (G12282), UNH142 (G12294), UNH146 (G12298), UNH154 (G12306), UNH162 (G12314), UNH189 (G12341), UNH211 (G12362), UNH860 (G68195), UNH874 (G68202), UNH887 (G68210), UNH995 (G68274) and UNH1003 (G68280). Amplifications were carried out using Qiagen Multiplex PCR Kit and following manufacturer's protocol. Reaction conditions were: 5 minutes at 95 °C pre-heating followed by 30 cycles of 30 seconds at 95 °C, 90 seconds at 60 °C (except for UNH115, UNH154, UNH860 and UNH1003 for which the temperature was 50 °C) and 1 minute at 72 °C, with a final elongation phase of 5 minutes at 72 °C. 3 µl of PCR products were diluted in 12 µl of HiDi formamide and 0.2 µl of 500 LIZ size standard added. Electrophoresis was thereafter carried out in an ABI 3730 XL automated sequencer. GeneMapper software was used to analyse the electrophoregrams and allele sizes.

2.3 Data analysis

2.3.1 Sequence analyses

Sequences were aligned manually using BioEdit 5.09 [21]. Additional sequences obtained from Genbank were also included in the analysis (Table I). Genetic diversity was calculated using DnaSP 5 [22]. Aligned sequences were analysed using Maximum Likelihood (ML) and Distance Method (DM) using MEGA (Molecular Evolutionary Genetics Analysis) version 5.1 [23]. Prior to analysis, an evolutionary model for ML was selected by MEGA 5.1 [24] using the Bayesian information criterion (BIC) [25]. Models with the lowest BIC scores are considered to describe the substitution pattern the best. Pair-wise sequence divergences between unique mtDNA haplotypes were calculated using the Kimura two-parameter model [26], followed by Neighbor Joining methods [27] to construct trees. Supports for inferred clades were obtained through the non-parametric bootstrap [28] with 2000 replicates for both methods.

Genetic diversity estimates were computed using DnaSP version 5.10.01 [29] and involved analysis of π , the average number of nucleotide differences per site between two sequences [30], and k , the average number of nucleotide differences between sequences [31].

2.3.2 Genotype analyses

Microsatellite data were checked for scoring errors due to stuttering or large allele dropout and presence of null alleles, using MICRO-CHECKER software [32]. Intra-population genetic variability was measured by estimating observed heterozygosity (H_O), expected heterozygosity (H_E) and the inbreeding coefficient (F_{IS}) using GENETOP version 3.4 [33]. F_{ST} and associated probabilities were estimated using a Markov chain method following [34]. The length of the Markov chain involved a burn in period of 1000 iterations and 100 batches of 1000 iterations thereafter.

A Factorial Correspondence Analysis (FCA) was carried out with GENETIX Version 4.05 [35], in order to investigate the relationships between individuals. This type of analysis can explain a maximal amount of genetic variation using a minimal number of factors and can provide the means for visualizing the genetic relationships between populations or species. We first carried out an analysis of all the specimens studied from both species, *O. niloticus* and *O. leucostictus*, in order to visualize differentiation at the species level. Thereafter, individuals from Lake Baringo and the three hot springs were compared to individuals of *O. leucostictus* in order to observe possible introgression. Finally, genotypes of individuals from

the hot springs and the neighbouring Lake Baringo were compared to establish their relationship.

A Bayesian cluster approach as implemented in STRUCTURE version 2.3.4 [36] was used to assess genetic admixture between *O. leucostictus* and *O. niloticus* at microsatellite loci assuming two populations ($K=2$). The admixture model with correlated allele frequencies was chosen [36,37]. Three different runs to estimate $q(i)$, the fraction of the genome of an individual i of a given species (*O. niloticus* or *O. leucostictus*), inherited from the other species, three different runs were done for 500,000 generations after discarding 200,000 generations as burn-in. The 90% probability intervals around the different $q(i)$ were calculated.

To confirm occurrence of nuclear introgression, we used Bayesian method implemented in NEWHYBRIDS 1.1 [38]. This method is designed to identify new hybrids between populations or species, and unlike STRUCTURE, NEWHYBRIDS model takes into account the predictable patterns of gene inheritance in hybrids [38,39]. An analysis was carried out for 100,000 iterations of Markov Chain Monte Carlo (MCMC) after 100,000 burn-in steps. Affinity of an individual to respective genotype class was assessed by posterior probability values.

3. Results

3.1 mtDNA differentiation

Forty-two new partial D-Loop sequences were obtained during this study (GenBank accession numbers are given in Tab. II). Thirty different haplotypes were identified using these and an additional 58 sequences from GenBank (Tab. II). All populations studied were polymorphic except for River Senegal (*O. niloticus*) and Lake Naivasha (*O. leucostictus*), where a single haplotype was present in 20 and five specimens examined respectively. Sample size (n), number of observed haplotypes (H_{ob}), number of polymorphic sites (p), average number of nucleotide differences (k) and nucleotide diversity (π) are presented in Table II. Out of the 30 haplotypes detected, 22 were unique to specific sample localities: 14 were observed only in Lake Turkana (total number of haplotypes sequenced $n=23$), two in Lake Albert ($n=20$), one in River Suguta ($n=9$), five in Lake Bogoria Hotel Spring ($n=24$). Furthermore, two other haplotypes were only found in hot springs populations.

Surprisingly, the haplotype of *O. leucostictus* from Lake Naivasha, which had been previously

described to occur in *O. niloticus* from Lake Baringo population [17] was also detected in all three hot spring populations: one in Turtle Spring, two in Chelaba Spring and 1 in Lake Bogoria Hotel Spring.

According to the different values of the corrected Akaike Information Criterion (AIC) obtained with MEGA 5.1, the optimal model of sequence evolution was HKY model [40]. This model was used in the ML analysis. Phylogenetic relationships among all haplotypes observed based on ML or DM methods were congruent. Fig. 2 presents a consensus NJ tree where bootstrap values obtained with ML and DM have been indicated. [4,17] reported that *O. niloticus* from West Africa (from River Senegal to the Nile) has been naturally introgressed with mtDNA from *O. aureus*. The haplotype observed in the River Senegal sample corresponds to this alien haplotype. Consequently, the network has been rooted using the single haplotype from River Senegal.

All other haplotypes observed, except one clustered in a single highly supported group (bootstrap value=97% for ML and MD methods). The outlier haplotype present in Lake Baringo, the three hot springs populations and *O. leucostictus* from Lake Naivasha, belongs to *O. leucostictus* as demonstrated earlier [17]. This haplotype was previously observed in the *O. niloticus* population from Lake Baringo, and considered introgressed, but was not observed in the Lake Bogoria Hotel Spring [5], as is the case in the present study. These results gave evidence that introgression by *O. leucostictus* mtDNA has also occurred in all the hot spring populations of the Loboï swamp.

The 28 remaining haplotypes are divided in three groups. The first one (I), supported by high bootstrap values (99% for ML and MD), is the sister group of the two others (II and III) and is composed of two different haplotypes present only in the hot spring populations. The four haplotypes of group II (bootstrap 97% for ML and 96% for DM) are found only in the three hot spring populations and in Lake Baringo. The other 22 haplotypes composed group III, which was supported by important bootstrap values: 72% for ML and 86% for DM. All the populations studied possess at least one haplotype in this group.

3.2 Microsatellite analysis

A total of 197 individuals from seven localities were genotyped at 16 microsatellite loci. All loci were polymorphic in all populations with the number of alleles ranging from nine to 44. Locus UNH154 was the most polymorphic with 44 alleles, while UNH129 was the least polymorphic with nine alleles. Lake Turkana population recorded the highest mean number of

alleles (12.9 ± 1.3 , 217 allele observed; $X \pm SE$ where SE is standard error), while Lake Naivasha population (*O. leucostictus*) had the least (5.0 ± 0.5 , 80 allele observed). The three hot spring populations had mean number of alleles ranging from 8.4 ± 0.6 , (Chelaba Spring), 8.6 ± 0.5 (Turtle Spring) to 8.9 ± 0.9 (Lake Bogoria Hotel Spring), and total number of alleles ranging from 134, 138 to 143 at Chelaba, Turtle and Lake Bogoria Hotel Springs, respectively. Distribution of alleles within the populations is shown in table III.

Using Micro-Checker, the presence of null alleles could not be rejected in the following loci and populations: PRL1AC in all three springs and Lake Baringo populations, UNH154 in Turtle Spring and River Suguta populations, UNH995 in Lake Bogoria Hotel and Chelaba Springs, PRL1GT and UNH115 in Lake Naivasha population, UNH1003 in Lake Bogoria Hotel population, UNH146 in Lake Baringo population, and UNH189 and UNH860 in River Suguta and Lake Turkana populations. No scoring error was detected in the data due to either stuttering or large allele dropout.

All populations had high observed heterozygosities ranging from 0.5828 in Lake Naivasha population to 0.7935 in Lake Baringo. F_{IS} values indicated heterozygote deficiencies in all populations studied. A total of 23 (24.15%) F_{IS} values out of the 105 calculated indicated significant heterozygote deficiencies (Table III).

Pairwise F_{ST} values among the seven populations ranged from 0.0177 to 0.3522 (Table IV) and were all highly significant ($P < 0.001$), indicating that all the samples can be considered genetically differentiated. F_{ST} values between Turtle Spring population and the other two hot spring populations (Lake Bogoria Hotel and Chelaba Springs), were relatively higher (0.0689 and 0.0749, respectively) than the F_{ST} observed between the latter two populations (0.0177). As expected, the highest F_{ST} values were observed for interspecific comparisons between *O. niloticus* and *O. leucostictus* populations, and ranged between 0.2653 at Chelaba Spring to 0.3522 at River Suguta.

All populations had private microsatellite alleles as shown in figure 3. The highest number of private alleles was observed in Lake Turkana population (91). Lake Bogoria Hotel Spring population had the highest number of private alleles (12) within the hot spring populations, while Chelaba Spring and Turtle Spring had nine and three private alleles, respectively.

FCA of all the genotypes observed (Figure 4) highlighted the differences between the two species *O. niloticus* and *O. leucostictus*. Based on the first axis (5.06% of total genetic variations), two main clusters corresponding to the two different species were identified. Within the *O. niloticus* cluster, two distinct groups were observed which correspond on one

hand to populations from Lake Bogoria Hotel Spring, Chelaba Spring, Turtle Spring and Lake Baringo (A), and on the other hand to Lake Turkana and River Suguta populations (B).

Taking into account that some populations have been introgressed by mtDNA from *O. leucostictus*, we carried out another FCA (Figure 5) of all three introgressed populations (the three hot springs and Lake Baringo) and *O. leucostictus* from Lake Naivasha. This enabled a clear separation between populations from *O. niloticus* and the one from *O. leucostictus* based on the first axis (62.24% of the total genetic variation). The fact that mtDNA of the introgressed individuals did not appear related to *O. leucostictus*, and the absence of any intermediate individuals between these two groups suggested that there has been no nuclear introgression between the two species.

Finally, an FCA analysis including the three hot spring populations and the Lake Baringo populations was run (Figure 6). The individuals were separated into four different clusters corresponding to the four populations. The axis 1 (43.34% of the genetic variation) discriminated individuals from Lake Bogoria Hotel and Chelaba Springs from Turtle Spring and Lake Baringo. The second axis (32.05% of the genetic variation) discriminated on one hand Chelaba Spring population from Lake Bogoria Hotel Spring population and on the other hand, Turtle Spring population from Lake Baringo population. It is noticeable that Lake Bogoria Hotel Spring and Chelaba Spring populations which are geographically closer to each other than to Turtle Spring are also genetically close, leading to a genetic isolation by distance pattern.

Further analysis of genotypes from introgressed populations was carried out using STRUCTURE in order to determine the proportions of nuclear genetic admixtures between *O. niloticus* and *O. leucostictus*. No F1 hybrid genotypes were detected from the analysis; this means there were no individuals with Q values around 0.5 (Figure 7). All individuals of *O. leucostictus* had Q values ≤ 0.003 (estimated fraction proportion of their genome inherited from any other species, in this case, *O. niloticus*) and could be considered as pure (not introgressed). All individuals of *O. niloticus* from Lake Turkana and River Suguta, and most of those from the hot springs and Lake Baringo could also be considered as pure (were characterized by Q values ≤ 0.01). Eleven individuals had Q values ≥ 0.01 (five from Chelaba Spring, three from Turtle Spring, two from Lake Bogoria Hotel Spring and one from Lake Baringo) and this number rises to 30 when considering the 90 % probability intervals (13, 6, 6, 5, respectively). Only one individual from Chelaba Spring had a Q value > 0.1 and ten when considering the 90% intervals (4, 3, 2, 1 respectively).

When analysing individual genotypes with NEWHYBRIDS, all specimens of *O. leucostictus*

had a posterior probability of being pure *O. leucostictus* greater or equal to 0.9989. In *O. niloticus* populations, seven specimens had a posterior probability of being pure *O. niloticus* inferior to 0.9574, and could be considered as possible hybrids. For all the concerned specimens, the probability to be F2 hybrid backcrossed with *O. niloticus* was higher than 0.0416 (Table S1).

It is important to take into account that all the specimens identified by NEWHYBRIDS as potential F2 backcrossed hybrids have been considered as hybrids according to their Q values obtained with STRUCTURE (Fig. 7).

Only one specimen from Turtle Spring population was considered as a hybrid by all three methods of analysis used; mtDNA haplotype identification and nuclear genotypes composition analysed by STRUCTURE and NEWHYBRIDS (Table S1).

4. Discussion

In a survey of *O. niloticus* populations in East Africa, the occurrence of a new (previously unknown) native population of *O. niloticus* in Loboï Swamp (Lake Bogoria Hotel Spring), Kenya has been documented [5]. This population was initially assumed to have been introduced from other localities within the region. However, its native status was confirmed due to its significant and unique genetic variability with a large number of private microsatellite alleles and mtDNA haplotypes (partial D-loop). The presence of tilapia in two other hot springs in the Loboï Swamp namely Turtle and Chelaba Springs, prompted us to investigate these fish. The present study clearly establishes that these two hot springs also host native fish closely related to those from Lake Bogoria Hotel Spring and that all three hot spring populations are in addition also closely related to the Lake Baringo population of the Nile tilapia. They share the most common haplotype in group II, and one haplotype in group III (Figure 2). Nevertheless, the three hot spring populations are also characterized by the private haplotypes of group I.

Lake Baringo is considered a modern representation of a much larger middle to upper Pleistocene Lake Kamasia [41], whose shores extended to Kapthurin River to the west and Bogoria River (which Loboï system is part of) headwaters to the south. This may explain the close relationship between the hot springs and the Lake Baringo, as observed in our study.

The analysis of partial mtDNA sequences of the D-loop region showed that the haplotype of *O. leucostictus* from Lake Naivasha previously confirmed to be present in Lake Baringo [17],

is also present in the three *O. niloticus* populations of the springs. The microsatellite study and visual observations of fish specimens during their capture confirmed that the spring populations were composed of a single species, *O. niloticus*. It is therefore evident that some of the fish in each of the hot springs have been introgressed by *O. leucostictus* mitochondrial genome.

The first studies of the Loboï swamp spring population [5,17] detected introgressed specimens in Lake Baringo by *O. leucostictus* but none from the Lake Bogoria Hotel Spring. However, it is possible that the authors failed to detect any introgression of the Lake Bogoria Hotel Spring population due to the low number of hot spring specimens available for their study (16 specimens).

If these introgressions originated from aquaculture practices *i.e.* escapees of *O. leucostictus* specimens or from *O. niloticus* introgressed beforehand in aquaculture farms within the Loboï swamp drainage, as it was hypothesized, then it is possible that these introgressions may have occurred after 2007.

Indeed, fish-farming activities within the region has been enhanced as a result of funding by Kenyan government through the Economic Stimulus Programme (ESP) introduced in 2009. The aim of the ESP project was to construct 200 fishponds worth 10.6 million euros in 140 constituencies by June 2013, with each constituency receiving 70,000 € [42]. Eight fishponds have been constructed between 2012 and 2013, in close proximity to the Loboï Swamp (within a range of 230 – 700 metres), and some stocked with Nile tilapia fingerlings from unspecified sources. Other fishponds in the Baringo-Bogoria catchment are located along the main rivers and streams in the region. Fingerlings used to stock the ponds within the region are produced by Omega Fish Farms located at Ol Kokwe Island on Lake Baringo. The farm was established to produce fingerlings for the local fish farmers [43]. Broodstock for the establishment was obtained from the Lake Baringo, hence may have acted as a source of introgressions observed in the Loboï Swamp fish.

At a first glance, it appears that these introgressions only concern mitochondrial DNA as earlier concluded in the study of Lake Baringo population [17]. Indeed, the results of the Factorial Correspondence Analysis (Figure 5) based on individuals microsatellites genotypes (16 loci) showed a clear separation between the three hot spring populations on one hand and on the other hand, the Lake Baringo, and population of *O. leucostictus* from Lake Naivasha. In case of large amount of introgression, individuals with intermediate genotypes were expected.

Recently STRUCTURE has been used [44,45] to quantify in more details the proportion of individual's genome originating from a different hybridizing species in tilapia from Lake Victoria and in Cichlids from Lake Tanganyika, respectively. Results obtained by these authors concern tilapia species from the Lake Victoria region: *O. niloticus* and *O. esculentus*. These two species are suspected to hybridize even though no mitochondrial introgression has so far been observed in samples (30 individuals of *O. niloticus* from Lakes Victoria, Kanyaboli, Namboyo and Sare, and 30 individuals of *O. leucostictus* from Lakes Kanyaboli and Namboyo). Results of this microsatellite study (eight loci) allowed the separation of both species using a Factorial Correspondence Analysis. However, using STRUCTURE, eight individuals (six *O. esculentus* and two *O. niloticus*) appeared to have 90% probability intervals that extend more than 30% out of a pure species Q value which suggested that these individuals had an important degree of genetic introgression.

In the present study, all *O. leucostictus* had Q values ≤ 0.003 and may be considered as not introgressed by *O. niloticus* genes (Figure 7). At least ten *O. niloticus* specimens (representing 11% of the individuals studied in the three hot springs and Lake Baringo populations) when considering the 90 % probability intervals had 10% of their genes or more that may be considered to have originated from *O. leucostictus* (three in Lake Bogoria Hotel Spring, five in Chelaba Spring, one in Turtle Spring and one in Lake Baringo).

Results obtained with NEWHYBRIDS are congruent with these findings and indicated that seven individuals (as observed using STRUCTURE) may have been introgressed. Taking into account that *O. niloticus* populations are genetically more variable than the *O. leucostictus* population, it is possible that this hybridization pattern was induced by a high genetic variance of the *O. niloticus* samples from the hot springs and Lake Baringo rather than introgression with *O. leucostictus*.

Nevertheless, by assessing the specimens supposed to be hybrids, all seven specimens determined by NEWHYBRIDS were also pointed out by STRUCTURE. It is also worth noting that even if there is a poor correspondence between specimens with mtDNA introgression and specimens designated by the two Bayesian programs as hybrids (only one specimen is concerned), with the FCA analysis, all the mtDNA introgressed individuals appeared within their respective *O. niloticus* cluster, but closer to the *O. leucostictus* cluster (Fig. 5). Lastly, specimens from Lake Turkana population which is the most genetically variable, is clearly composed of non-hybridized individuals (Fig. 7). This suggests that hybrids identified by both STRUCTURE and NEWHYBRIDS are not artefacts due to genetic variance in the *O. niloticus* populations.

All of these observations are congruent and confirm with more detail the previous conclusions [17] pointing out the nuclear introgression of the *O. niloticus* populations from the hot springs of Lobo Swamp and Lake Baringo by nuclear *O. leucostictus* genes.

Even if the extent of the hot spring biotopes are small, and consequently the size of the effective populations, native mtDNA diversity was higher in the hot springs populations than in Lake Baringo and River Suguta populations. Nine different haplotypes were found in the hot springs (36 pure individuals analysed) while three were found in Lake Baringo (eight pure individuals analysed) or River Suguta (nine pure individuals analysed). Differences in samples sizes may partially account for the observed differences but when one considers Chelaba Spring on its own with four haplotypes for six individuals or Turtle Spring with five haplotypes for seven individuals, one can conclude that haplotype diversity in the hot spring is high.

This does not seem to be congruent with the very small size of the biotopes. The hot springs are located within a small swamp of about 1.5 km². The swamp itself is not expected to be a favourable biotope for fish survival as it is completely covered with macrophytes of which the most dominant are *Typha domingensis* and *Cyperus papyrus* [46]. This unexpected high mtDNA diversity can be explained if we consider that the characteristics of the hot springs are not directly related to pluvial water. The hot springs originate from faults and fractures that allow water from deep aquifers to rise to the surface [13,47,48] hence maintaining their perennial existence throughout the year, regardless of climatic changes. The high mtDNA diversity may have been favoured by the stable environmental conditions and consequently decreased levels of environmental stress in the hot springs.

Moreover, gene flow between the three hot spring populations is not enough to prevent their differentiation, as shown by F_{ST} values that were all statistically significant between populations (Table IV), and the FCA analysis (Figure 6). Even if there are only three populations, it seems that the genetic differentiation is congruent with the genetic differentiation through isolation by distance. This was also unexpected considering that the springs are very close to each other. When they join the swamp, Lake Bogoria Hotel and Chelaba Springs are separated only by a few hundred metres, and Turtle Spring is separated from Lake Bogoria Hotel and Chelaba Springs by approximately one kilometre.

To explain this amount of observed genetic differentiation between the springs, the first hypothesis is to consider that the swamp may create a barrier to free movement of fish from one spring to another thereby diminishing gene flow.

Both water temperature and dissolved oxygen concentration within the swamp are lower than within the springs mainly due to the shading effect provided by the dense vegetation and organic decomposition from the decaying matter, respectively. This situation has been described in a study of cichlid fishes from Lake Victoria [49,50]. It has been demonstrated that differences in eco-physiological properties of water can limit gene flow as a result of adaptation or physiological avoidance. Our observations are also congruent with the apparent lack of gene flow between two populations of both *O. esculentus* and *O. niloticus* from different satellite lakes of Lake Victoria separated by a wetland and papyrus swamps [44]. However, if the swamp represents an impenetrable barrier to free movement of fish so that F_{ST} values are statistically significant, it is difficult to explain how *O. leucostictus* or introgressed *O. niloticus* could have crossed the swamp, from fish ponds connected to the swamp drainage, and subsequently hybridized with native fish of the hot springs. It is still possible, though very unlikely, that someone may have directly introduced fish into each hot spring. Whereas this hypothesis cannot be rejected without consideration, there is simply no objective reason to carry out such fish translocation. Even though the exact way by which fishes have been introgressed is still unknown, the reality of the introgression has clearly been demonstrated and that aquaculture activities played a role in this process.

It was also apparent that the three populations of *O. niloticus* in the hot springs are distinct from other related populations by the presences of private alleles and haplotypes. This differentiation could be the consequence of their local adaptations to the extreme thermal conditions of their habitat. Studies indicate that exposure of *O. niloticus* to water temperatures of more than 36°C at an early stage of their development (10 days post fertilization) influences sex ratios towards males [14, 41]. These characteristics are of utmost importance to aquaculture, as they can be utilised to inexpensively produce all-male cultures. Tilapia farming mainly relies on male mono-sex culture in order to avoid overcrowding in the ponds due to breeding, and to benefit from fast growth of males [51,52,53]. Hence the rich genetic diversity and adaptations to masculinizing effect of water temperature makes these populations an important aquaculture resource, and an important natural model for understanding sex determination in Nile tilapias.

These populations and the genetic resources they represent need to be conserved. The conservation and management of Loboï swamp system should thus be accorded top priority in order to safeguard its critical genetic resource that has great potential in development of tilapia aquaculture.

5. Acknowledgements

Molecular genetic data used in this work were produced through molecular genetic analysis technical facilities of “Centre Méditerranéen de l’Environnement et de la Biodiversité” (CeMEB). We are grateful to Frédérique Cerqueira, Erick Desmarais (ISEM) and Joseph Gathua (NMK) and for their technical support in the laboratory and field collections during the study. We are grateful to Dr. Judith Nyunja (KWS) for her support during sample collections in Lake Turkana. Finally, we are grateful to C. Firmat for useful comments on the first draft of this manuscript.

6. References

1. Shelton WL, (2002) Tilapia culture in the 21st century. In: Guerrero III,., Guerrero-del Castillo RD, M.R. (Eds.), *Tilapia Farming in the 21st Century*, Proc. Int. Forum on Tilapia Farming in the 21st Century, February 2002, Los Baños, Philippines, pp. 1–20.
2. Romana-Eguia MRR, Ikeda M, Basiao, ZU, Taniguchi N, (2004) Genetic diversity in farmed Asian Nile and red hybrid tilapia stocks evaluated from microsatellite and mitochondrial DNA analysis. *Aquaculture* 236, 131–150.
3. Eknath AE and Hulata G, (2009) Use and exchange of genetic resources of Nile tilapia (*Oreochromis niloticus*). *Reviews in Aquaculture* 1: 197-213.
4. Rognon X, Guyomard R (2003) Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. *Molecular Ecology* 12: 435–445.
5. Nyingi D, De-Vos L, Aman R, Agnèse J-F (2009) Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) (Cichlidae; Teleostei) in the Loboï Swamp (Kenya). *Aquaculture* 297: 57-63.
6. Ryman N (1991) Conservation genetics considerations in fishery management. *Journal of Fish Biology* 39 (21): 1–224.
7. Lind CE, Brummett, RE, Ponzoni RW (2012) Exploitation and conservation of fish genetic resources in Africa: issues and priorities for aquaculture development and research. *Reviews in Aquaculture* 4: 125-141.
8. Trewavas, E (1983) Tilapiine fishes of the genera *Sarotherodon* *Oreochromis* and *Danakilia*, British Museum (Natural History), London. 583p.
9. Bezault E, Balaesque P, Toguyeni A, Fermon E, Araki H, et al. (2011) Spatial and temporal variation in population structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *Genetics*, 12: 102.

10. FAO website. Fisheries and aquaculture information and statistics service. Available: <http://www.fao.org/figis/servlet/SQServlet?ds=Capture&k1=SPECIES&K1v=1&k1s=3217&outtype=html>. Accessed 2014 August 19.
11. Grammer GL, Slack WT, Peterson MS, Dugo MA (2012) Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) establishment in temperate Mississippi, USA: multi-year survival confirmed by otolith ages. *Aquatic Invasions* 7: 367-376.
12. Chakraborty SB, Banerjee S (2009) Culture of monosex Nile tilapia under different traditional and non-traditional methods in India. *World Journal of Fish and Marine Sciences*, 1 (3): 212-217.
13. Ashley GM, Maitima Mworja J, Muasya AM, Owen RB, Driese SG, et al. (2004) Sedimentation and recent history of a freshwater wetland in a semi-arid environment: Loboï Swamp, Kenya, East Africa. *Sedimentology* 51, 1301–1321.
14. Baroiller JF, D'Cotta H (2001) Environment and sex determination in farmed fish. *Comparative Biochemistry and Physiology Part C* 130: 399-409.
15. Tessema M, Muller-Belecke A, Horstgen-Schwark G (2006) Effects of rearing temperatures on the sex ratios of *Oreochromis niloticus* populations. *Aquaculture*, 258 (1–4): 270–277.
16. Owen RB, Renaut RW, Hover VC, Ashley GM, Muasya AM (2004) Swamps, springs, and diatoms: wetlands regions of the semi-arid Bogoria–Baringo Rift, Kenya. *Hydrobiologia*, 518: 59–78.
17. Nyingi DW, Agnèsè J-F (2007) Recent introgressive hybridization revealed by exclusive mtDNA transfer from *Oreochromis leucostictus* (Trewavas, 1933) to *Oreochromis niloticus* (Linnaeus, 1758) in Lake Baringo, Kenya. *J. Fish BioLake* 70 (Supplement A): 148–154.
18. Mumba M (2008) Adapting to climate change: lessons from Lake Bogoria catchment, Kenya. Programme report No. 3/08.
19. Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* 25(22): 4692-4693.
20. Agnese J-F, Gourene B, Nyingi DW (2009) Functional microsatellite and selective sweep in the black-chinned tilapia *Sarotherodon melanotheron* (Teleostei, Cichlidae) natural populations. *Marine Genomics*, 1: 103-107.
21. Hall TA (1999) BioEdit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series No.* 41: 95-98.

22. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
23. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetic analysis and sequence alignment. *Briefing in Bioinformatics*, 5: 150-163.
24. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5. Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28(10): 2731-2739.
25. Schwarz G (1978) Estimating the dimensions of a model. *The Annals of statistics*, 6(2): 461-464.
26. Kimura M (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111-120.
27. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
28. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4): 783-791.
29. Rozas J, Sanchez-DelBarrio JC, Messeguer X (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497.
30. Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.
31. Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics* 105: 437–460.
32. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P,(2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4: 535-538.
33. Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8:103–106.
34. Weir BS, Cockerham CC (1984) Estimating F-Statistic for the analysis of population structure. *Evolution* 38(6): 1358-1370.
35. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
36. Pritchard JK, Stephens M, Donnelly P, (2000). Inference of population structure using

- multilocus genotype data. *Genetics*, 155(2): 945-959.
37. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164 (4): 1567-1587.
 38. Anderson EC, and Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**: 1217–1229.
 39. Anderson EC (2008) Bayesian inference of species hybrids using multilocus dominant genetic markers. *Phil Trans R Soc* 363: 2841–2850.
 40. Hasegawa M, Kishino H, Yano T (1985) Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22 (2): 160–174.
 41. Fuch VE (1950) Pleistocene events in the Baringo Basin, Kenya Colony. *Geol. Mag.Lond.* 87: 149-174.
 42. Nyingi DW, Agnès J-F (2011) Tilapia in eastern Africa – a friend and foe. Species in the Spotlight. Darwall, W.R.T., Smith, K.G., Allen, D.J., Holland, R.A, Harrison, I.J., and Brooks, E.G.E. editors. *The Diversity of Life in African Freshwaters: Under Water, Under Threat. An analysis of the status and distribution of freshwater species throughout mainland Africa.* Cambridge, United Kingdom and Gland, Switzerland: IUCN. Pp. 76-78.
 43. The Fish Site website. Developing a Tilapia breeding station in Kenya: Available: <http://www.thefishsite.com/articles/1186/developing-a-tilapia-breeding-station-in-kenya>. Accessed on 2014 August 19.
 44. Angienda PO, Lee HJ, Elmer KR, Abila R, Waindi EN, et al. (2010) Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. *Conservation genetics*, 12: 243-255.
 45. Nevado B, Fazalova V, Backeljau T, Hanssens M, Verheyen E (2011) Repeated unidirectional introgression of Nuclear and Mitochondrial DNA between four congeneric Tanganyikan cichlids. *Molecular Biology and Evolution* 28(8): 2253-2267.
 46. Muasya AM, Hover VC, Ashley GM, Owen RB, Goman MF, et al. (2004) Diversity and distribution of macrophytes in a freshwater wetland, Loboï swamp (Rift valley) Kenya. *Journal of East African Natural History*, 93: 39-47.
 47. Deocampo DM (2002) Sedimentary processes and lithofacies in lake-margin groundwater-fed wetlands in East Africa. In: Ashley, G.M., Renaut, R.W. editors. *Sedimentation in Continental Rifts*, vol. 73. Special Publication-SEPM, Tulsa, OK, pp. 295–308.
 48. Ashley GM, Goman M, Hover VC, Owen RB, Renaut RW, et al. (2002) Artesian blister wetlands, a perennial water resource in the semi-arid Rift-Valley of East Africa, *Wetlands* 22(4): 686-695.

49. Chapman LJ, Chapman CA, Nordlie FG, Rosenberger AE (2002) Physiological refugia: swamps, hypoxia tolerance and maintenance of fish diversity in the Lake Victoria region. *Comp Biochem Physiol A* 133:421–437.
50. Crispo E, Chapman LJ, (2008) Population genetic structure across dissolved oxygen regimes in an African cichlid fish. *Molecular Ecology* 17:2134–2148.
51. Baroiller JF, D’Cotta H, Bezault E, Wessels S, Hoerstgen-Schwark G (2009) Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology Part A*, 153: 30-38.
52. Baroiller J-F, Jalabert B (1989) Contribution of research in reproductive physiology to the culture of tilapias. *Aquatic Living Resources* 2: 105-116.
53. Beardmore JA, Mair GC, Lewis RI (2001) Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture* 197: 283-301

Tables

Table I. Original Genbank references and corresponding voucher specimen numbers of the samples studied.

Population	Coordinates	Species	Genbank No.	Voucher Specimens
<i>Loboi swamp</i>				
Bogoria Hotel Spring	0°21'44''N, 36°03'04''E	<i>O. niloticus</i>	FJ440588-440603	(NMK)/1798/1-6
		<i>O. niloticus</i>	KJ746025-746032	(NMK) FW/36549/1-18
<i>Loboi swamp</i>				
Chelaba Spring	0°21'30''N, 36°02'58''E	<i>O. niloticus</i>	KJ746033-746040	(NMK) FW/2997/1-31
<i>Loboi swamp</i>				
Turtle Spring	0°21'44''N, 36°02'41''E	<i>O. niloticus</i>	KJ746051-746058	(NMK) FW/2996/1-25
<i>Lake Baringo</i>				
Kambi samaki	0°36'54''N, 36°01'59''E	<i>O. niloticus</i>	KJ746041-44	(NMK).FW/2660/1-31
Robertson camp	0°36'43''N, 36°01'31''E	<i>O. niloticus</i>	EF016697-016708 FJ440604-440607	MRAC 95-027-P-0074-0084
<i>River Suguta</i>				
Kapedo	1°10'44''N, 36°06'25''E	<i>O. niloticus</i>	EF016709-016714	(ISEM) JFA-5287-5313
<i>Lake Turkana</i>				
North Island	4°05'50''N, 36°02'46''E	<i>O. niloticus</i>	EF016680-016696	(NMK)1639/1-11,1641/1-13
Fergusson Golf	3°30'46''N, 35°54'53''E	<i>O. niloticus</i>	KJ746045-746050	(ISEM) JFA-2010-LT1-30
<i>Lake Albert</i>				
Butiaba	1°48'47''N, 31°16'04''E	<i>O. niloticus</i>	FJ440577-FJ440587	(ISEM) JFA-1922-1932
<i>River Senegal</i>				
Saint Louis	16°03'47''N, 16°28'33''W	<i>O. niloticus</i>	EF016715-016723	(ISEM) JFA-5683-5692
<i>Lake Naivasha</i>				
	0°46'18''N, 36°20'51''E	<i>O. leucostictus</i>	EF016702	(ISEM) JFA-5346-5375

In bold, specimens studied or sequences obtained during the present work. ISEM Institut des Sciences de l'Evolution de Montpellier, France ; MRAC, Musée Royal de l'Afrique Centrale de Tervuren, Belgium; NMK, National Museums of Kenya.

Table II. mtDNA (partial D-loop sequences) variability observed in of 8 populations of *O. niloticus*.

Population	<i>n</i>	<i>Hob</i>	<i>p</i>	<i>k</i>	π
Bogoria Hotel Spring	24 (23)	10 (9)	51 (32)	8.39 (6.41)	0.024 (0.018)
Chelaba Spring	8 (6)	4 (3)	45 (25)	18.57 (9.13)	0.053 (0.026)
Turtle Spring	8 (7)	5 (4)	46 (26)	18.07 (13.33)	0.051 (0.038)
Lake Baringo	15 (8)	4 (3)	37 (12)	18.06 (6.00)	0.051 (0.017)
Lake Turkana	23	15	19	2.53	0.007
River Suguta	9	3	6	1.33	0.004
Lake Albert	20	4	3	0.30	0.001
River Senegal	18	1	0	0.00	0.000

n, number of sequences; *Hob*, number of different haplotypes observed; *P*, number of polymorphic sites; *k*, average number of nucleotide difference; π , nucleotide diversity. In bracket, results without taking into account introgressed specimens.

Table III. Heterozygosity and coefficient of inbreeding in six populations of *O. niloticus* and *O. leucostictus* from Lake Naivasha.

Species		<i>Oreochromis niloticus</i>					<i>O. leucostictus</i>	
Population		Bogoria Spring	Chelaba Spring	Turtle Spring	Lake Baringo	River Suguta	Lake Turkana	Lake Naivasha
Locus	<i>N</i>	31	31	25	30	21	29	20
PRL1AC	H_e	0.6683	0.7472	0.8523	0.8839	0.7971	0.9031	0.5639
	H_o	0.4000	0.5333	0.6818	0.7333	0.8571	0.8571	0.5000
	F_{is}	0.4156	0.3017	0.2222	0.1867	-0.0511	0.0690	0.1300
	P	0.0000	0.0172	0.3320	0.095	0.3632	0.1056	0.2257
	A	10	11	9	18	10	18	4
PRL1GT	H_e	0.5899	0.7733	0.8504	0.8761	0.7088	0.8456	0.6142
	H_o	0.5000	0.6667	0.7600	0.8000	0.6500	0.7600	0.3214
	F_{is}	0.1701	0.1545	0.1264	0.1037	0.1083	0.1214	0.4906
	P	0.0601	0.1910	0.5222	0.0000	0.5784	0.2748	0.0000
	A	10	10	10	13	5	16	4
UNH211	H_e	0.8798	0.8250	0.8620	0.9310	0.7275	0.8644	0.8272
	H_o	0.7419	0.7333	0.9583	0.8966	0.7000	0.7931	0.7857
	F_{is}	0.1727	0.1278	-0.0907	0.0546	0.0634	0.0999	0.0682
	P	0.0046	0.0020	0.2035	0.5198	0.5585	0.0596	0.1343
	A	12	11	12	18	6	12	8
UNH1003	H_e	0.7756	0.7513	0.7456	0.8195	0.7450	0.7111	0.3428
	H_o	0.4333	0.6452	0.6000	0.7153	0.6500	0.5862	0.2667
	F_{is}	0.4548	0.1573	0.2148	0.1463	0.1527	0.1925	0.2381
	P	0.0000	0.0902	0.0785	0.2722	0.3487	0.2712	0.1424
	A	10	10	8	11	7	12	4
UNH104	H_e	0.6239	0.7456	0.7694	0.7564	0.7742	0.8977	0.7487
	H_o	0.600	0.7000	0.8696	0.8571	0.7895	0.9655	0.8571
	F_{is}	0.0552	0.0780	-0.1083	-0.1153	0.0074	-0.0580	-0.1270
	P	0.9434	0.1105	0.9019	0.7534	0.8804	0.6980	0.9064
	A	4	7	8	9	5	14	6
UNH115	H_e	0.7846	0.7456	0.6354	0.7878	0.7506	0.6327	0.3200
	H_o	0.8065	0.8065	0.6667	0.6897	0.6667	0.5714	0.2000
	F_{is}	-0.0115	-0.0653	-0.0279	0.1418	0.1358	0.1333	0.3895
	P	0.9078	0.3620	0.4527	0.0006	0.0250	0.3973	0.0598
	A	8	6	7	7	8	7	2
UNH129	H_e	0.4938	0.5128	0.5408	0.6161	0.5839	0.5344	0.6650
	H_o	0.5161	0.3333	0.5714	0.6000	0.7143	0.5000	0.6667
	F_{is}	-0.0289	0.3647	-0.0323	0.0431	-0.2000	0.0825	0.0144
	P	0.6644	0.0429	0.7965	0.6357	0.5035	0.2638	0.9113
	A	3	3	4	5	3	4	3
UNH142	H_e	0.5994	0.6906	0.7360	0.8033	0.7022	0.8011	0.2494
	H_o	0.5161	0.8000	0.8400	0.8333	0.7778	0.7037	0.2857
	F_{is}	0.1549	-0.1419	-0.1213	-0.0204	-0.0794	0.1401	-0.1279
	P	0.1515	0.5706	0.8908	0.1061	0.2617	0.1590	1.0000
	A	5	6	5	8	5	11	3
UNH146	H_e	0.7594	0.7423	0.7736	0.8733	0.6020	0.5511	0.5594
	H_o	0.7667	0.6071	0.7600	0.6667	0.6190	0.4138	0.5667
	F_{is}	0.0074	0.1997	0.0380	0.2525	-0.0039	0.2656	0.0040

	<i>P</i>	0.7886	0.0496	0.4238	0.0102	0.7029	0.0540	0.3045
	<i>A</i>	10	7	7	10	3	4	4
UNH154	<i>H_e</i>	0.8713	0.8663	0.8624	0.9019	0.8491	0.9477	0.7300
	<i>H_o</i>	0.7692	0.8065	0.6800	0.7931	0.3077	0.8929	0.8000
	<i>F_{is}</i>	0.1364	0.0854	0.2309	0.1379	0.6608	0.0760	-0.0791
	<i>P</i>	0.5460	0.3935	0.0003	0.0568	0.0000	0.0416	0.9411
	<i>A</i>	15	11	10	10	10	20	7
UNH162	<i>H_e</i>	0.7361	0.7283	0.8628	0.8756	0.6429	0.8668	0.7128
	<i>H_o</i>	0.7000	0.6000	0.8095	1.0000	0.6190	0.9655	0.6667
	<i>F_{is}</i>	0.660	0.1926	0.0860	-0.1255	0.0614	-0.0965	0.0816
	<i>P</i>	0.1675	0.1581	0.8164	0.1468	0.8427	0.1697	0.8282
	<i>A</i>	9	7	10	12	5	14	5
UNH189	<i>H_e</i>	0.8538	0.7078	0.8376	0.8389	0.4206	0.8984	0.4923
	<i>H_o</i>	0.9032	0.7000	0.8400	0.7931	0.1905	0.8750	0.5714
	<i>F_{is}</i>	-0.0415	0.0279	0.0175	0.0721	0.5640	0.0473	-0.1429
	<i>P</i>	0.6913	0.3419	0.1503	0.2293	0.0078	0.2097	0.3020
	<i>A</i>	15	9	11	12	3	16	5
UNH860	<i>H_e</i>	0.7877	0.8341	0.7584	0.8029	0.1863	0.9251	0.7483
	<i>H_o</i>	0.6452	0.7586	0.8400	0.8929	0.2000	0.8276	0.7333
	<i>F_{is}</i>	0.1968	0.1079	-0.0874	-0.0940	-0.0483	0.1227	0.0370
	<i>P</i>	0.0201	0.4876	0.8627	0.7953	1.0000	0.1618	0.6471
	<i>A</i>	11	10	10	11	4	19	8
UNH874	<i>H_e</i>	0.7683	0.6906	0.8084	0.8339	0.7687	0.8992	0.6653
	<i>H_o</i>	0.6333	0.7667	0.8095	0.8000	0.9048	0.8571	0.7241
	<i>F_{is}</i>	0.1921	-0.0934	0.0230	0.0576	-0.1533	0.0650	-0.0710
	<i>P</i>	0.0082	0.8139	0.5181	0.0260	0.3119	0.0607	0.1354
	<i>A</i>	10	9	9	9	9	17	4
UNH887	<i>H_e</i>	0.6550	0.7028	0.7760	0.8133	0.6066	0.7478	0.5938
	<i>H_o</i>	0.6129	0.6667	0.8400	0.8333	0.6667	0.7308	0.6786
	<i>F_{is}</i>	0.0807	0.0683	-0.0622	-0.0076	-0.0749	0.0423	-0.1250
	<i>P</i>	0.0201	0.4876	0.8627	0.7953	1.0000	0.1618	0.6471
	<i>A</i>	6	9	8	10	5	8	4
UNH995	<i>H_e</i>	0.6972	0.7494	0.8176	0.7527	0.7637	0.9001	0.7783
	<i>H_o</i>	0.5161	0.5667	0.7600	0.7931	0.7500	0.9310	0.7000
	<i>F_{is}</i>	0.2749	0.2598	0.0907	-0.0362	0.0436	-0.0168	0.1174
	<i>P</i>	0.0006	0.0039	0.0220	0.9307	0.8199	0.3499	0.1225
	<i>A</i>	5	8	10	8	5	14	9
Global mean	<i>H_e</i>	0.7215	0.7383	0.7805	0.8229	0.6643	0.8080	0.6007
	<i>H_o</i>	0.6288	0.6682	0.7679	0.7935	0.6819	0.7607	0.5828
	<i>F_{IS}</i>	0.1444	0.1111	0.0375	0.0536	0.0628	0.0765	0.0481
	<i>A</i>	143	134	180	138	93	217	80

N, sample size; *H_e*, expected heterozygosity; *H_o*, observed heterozygosity; *F_{IS}*, inbreeding coefficient; *P*, associated *F_{IS}* probability and *A*, number of alleles.

Table IV. Pairwise F_{ST} estimates between six *O. niloticus* populations and one *O. leucostictus* population (Lake Naivasha).

Population	Lake Bogoria Hotel Spring	Chelaba Spring	Turtle Spring	Lake Baringo	River Suguta	Lake Turkana
Chelaba Spring	0.0177					
Turtle Spring	0.0689	0.0749				
Lake Baringo	0.0373	0.0408	0.0253			
River Suguta	0.2123	0.2201	0.1822	0.1368		
Lake Turkana	0.1645	0.1642	0.1334	0.1085	0.1366	
Lake Naivasha	0.2893	0.2653	0.2769	0.2591	0.3522	0.2728

Every F_{ST} were statistically highly significant.

Figures

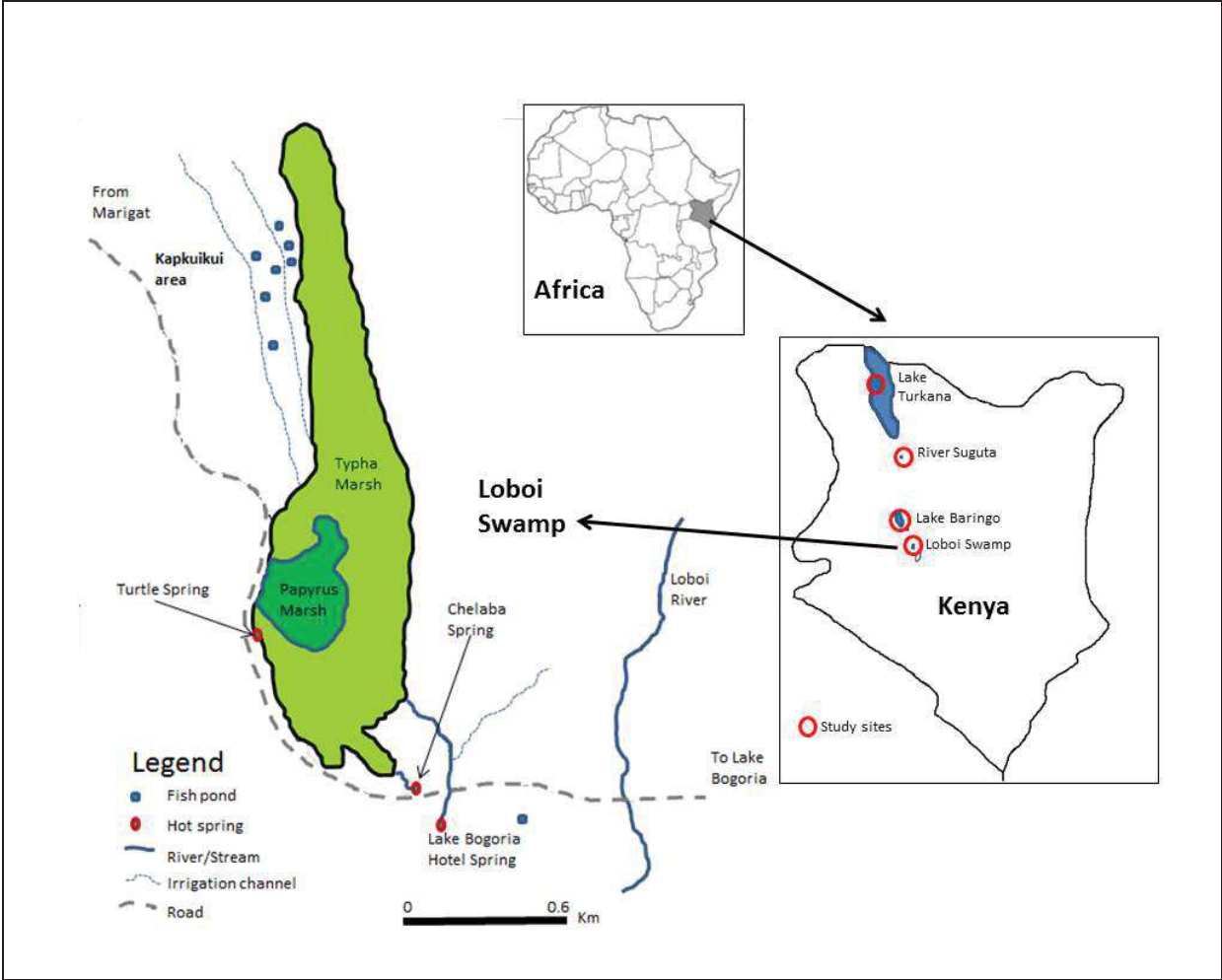


Figure 1: Map of Lobo Swamp showing the location of the three springs and the fish ponds constructed by the end of 2011.

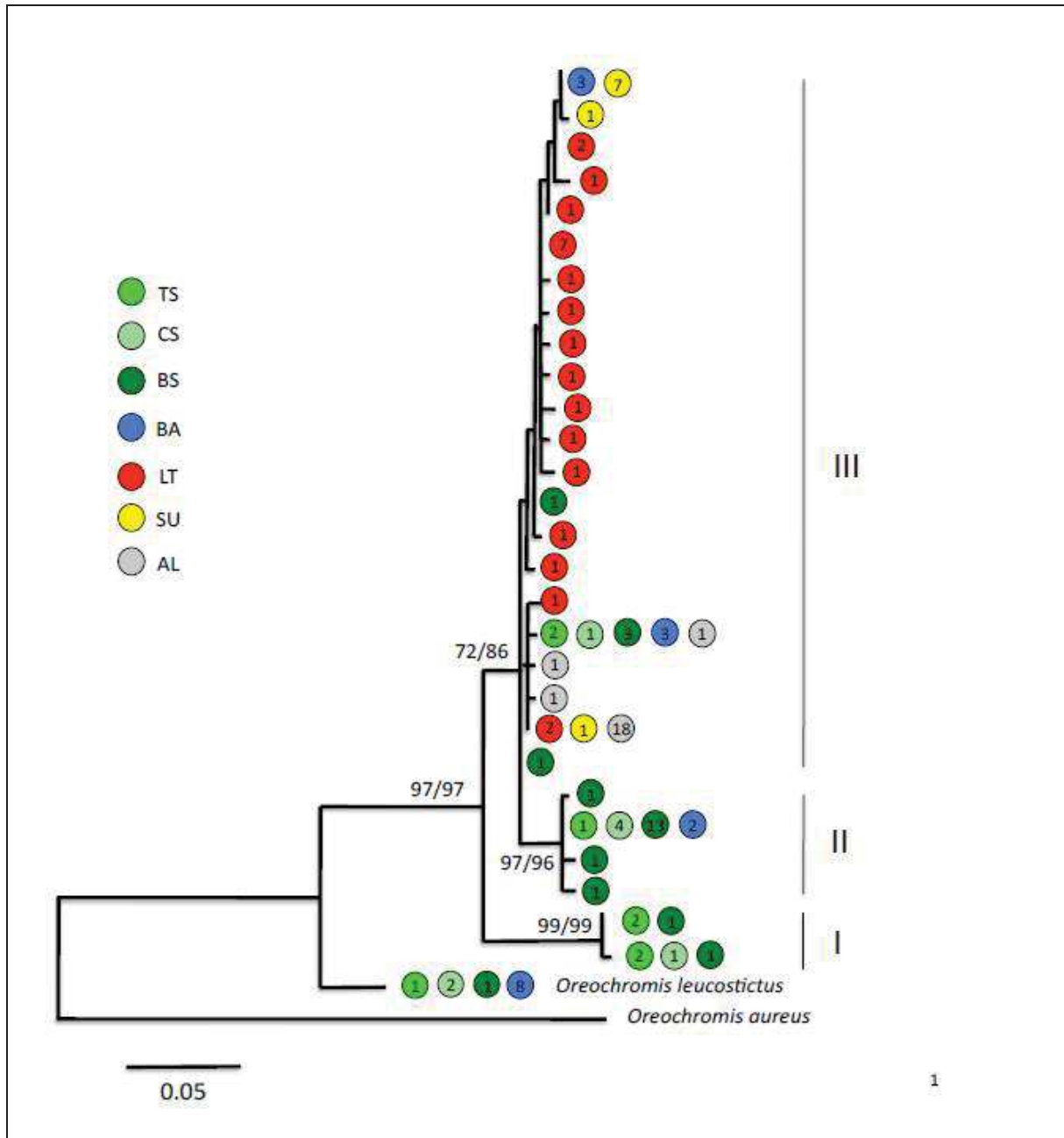
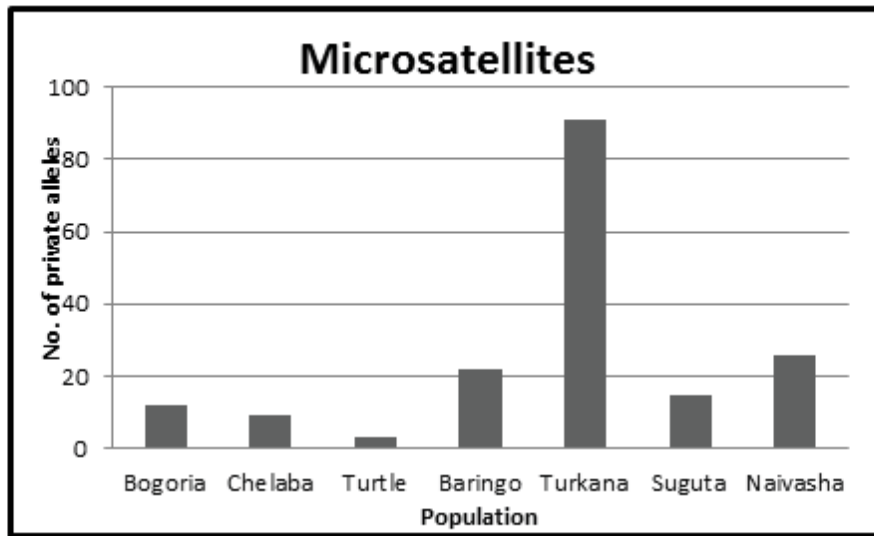


Figure 2: Phylogenetic tree representing the genetic relationships between the 30 haplotypes studied. The network was rooted using haplotype observed in the River Senegal population naturally introgressed by *O. aureus*. *O. niloticus* haplotypes have been clustered in three groups (I, II and III). Numbers above branches indicate bootstrap values (in percentage) based on 2000 replicates for ML and DM methods. The bar represents 0.05 units divergence, TS, Turtle spring, CS, Chelaba Spring, BS, Lake Bogoria Hotel Spring, BA, Lake Baringo, LT, Lake Turkana, SU, Suguta River and AL, Lake Albert. Values inside the circles represent number of haplotype.

(a)



(b)

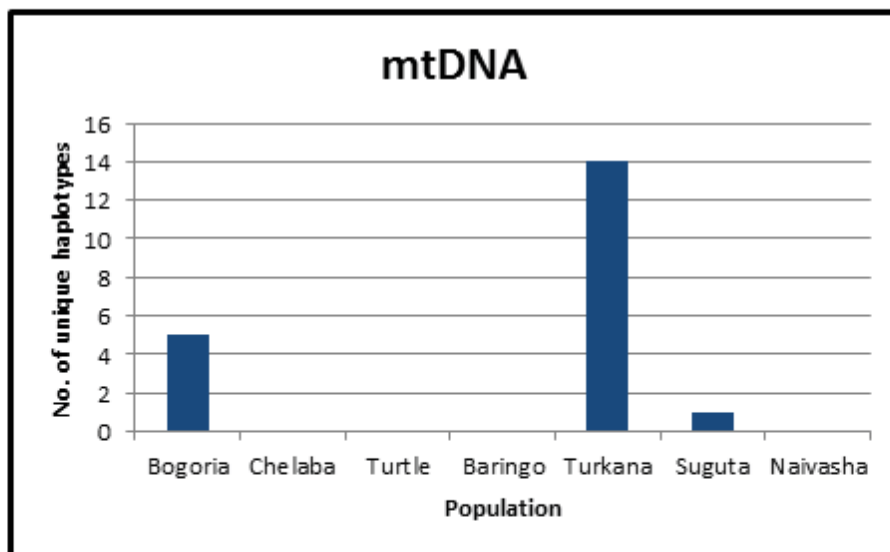


Figure 3: Number of private microsatellite alleles and private mtDNA haplotypes present in six populations of *O. niloticus* and one population of *O. leucostictus* (mtDNA not included for Naivasha population).

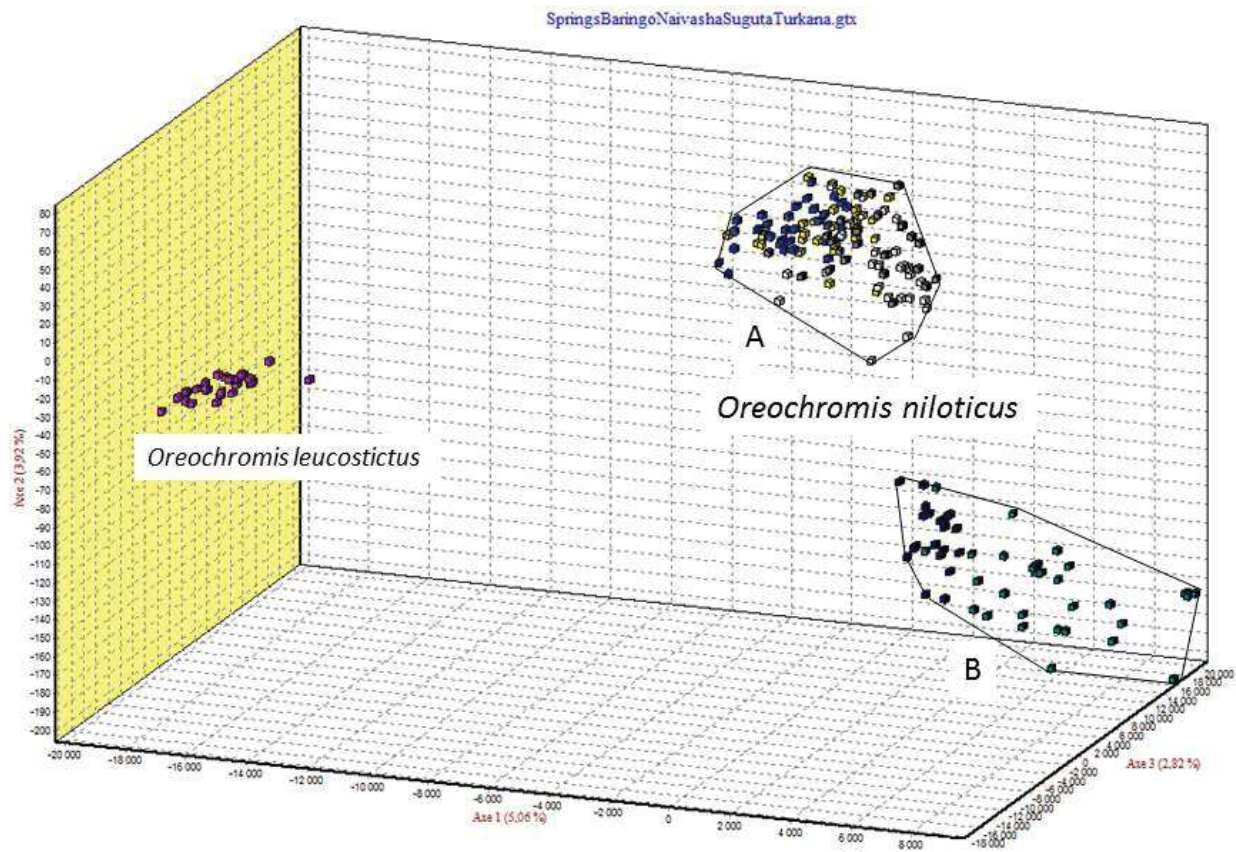


Figure 4: Factorial Correspondence Analysis (GENETIX 4.05) based on individual microsatellite genotypes at 16 loci showing differentiation of the seven populations into two distinct groups corresponding to *Oreochromis leucostictus* and *O. niloticus*. *O. niloticus* populations further clustered into two groups; individuals from Lake Baringo Hotel Spring, Chelaba Spring, Turtle Spring and Lake Baringo (A), and Lake Turkana and River Suguta (B).

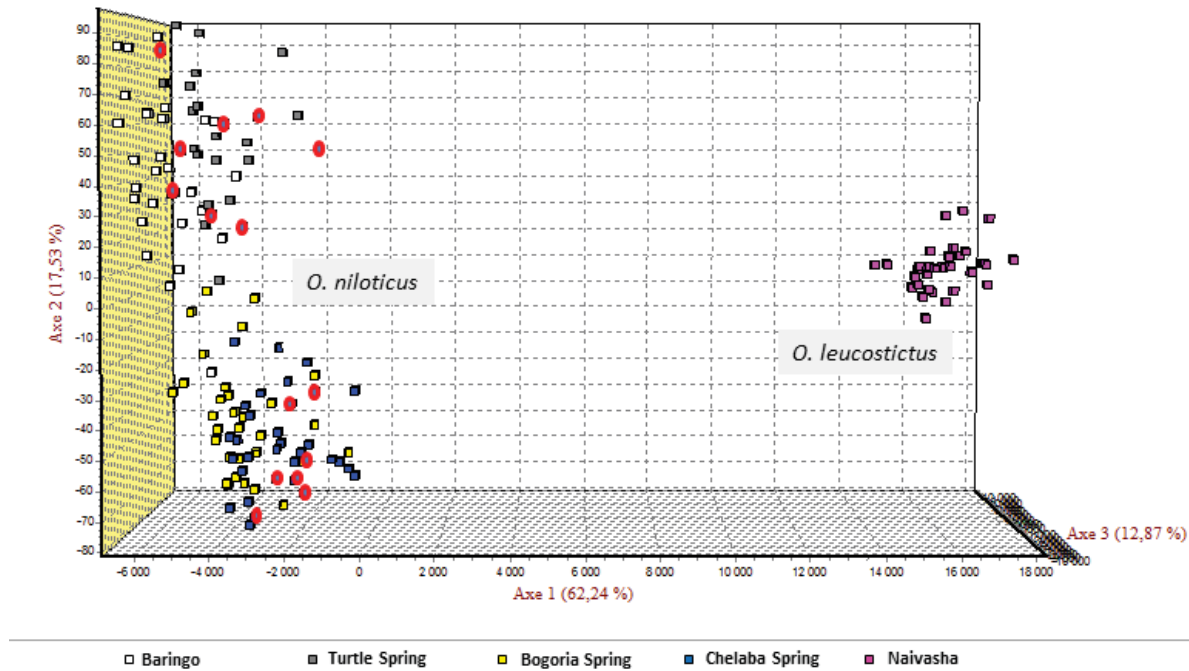


Figure 5: Factorial Correspondence Analysis (GENETIX 4.05) based on individual microsatellite genotypes at 16 loci showing no intermediate genotypes between mtDNA introgressed populations, the three hot springs and Lake Baringo populations, and *O. leucostictus* from Lake Naivasha. Individuals of *O. niloticus* that have *O. leucostictus* mtDNA are highlighted in red. TS, Turtle Spring; BA, Lake Baringo; BS, Bogoria Spring; OL, *Oreochromis leucostictus* from Lake Naivasha. Note: only two introgressed individuals from Lake Baringo are shown instead of eight (Fig. 1) because sequences of the six omitted individuals were obtained from Genbank hence not genotyped.

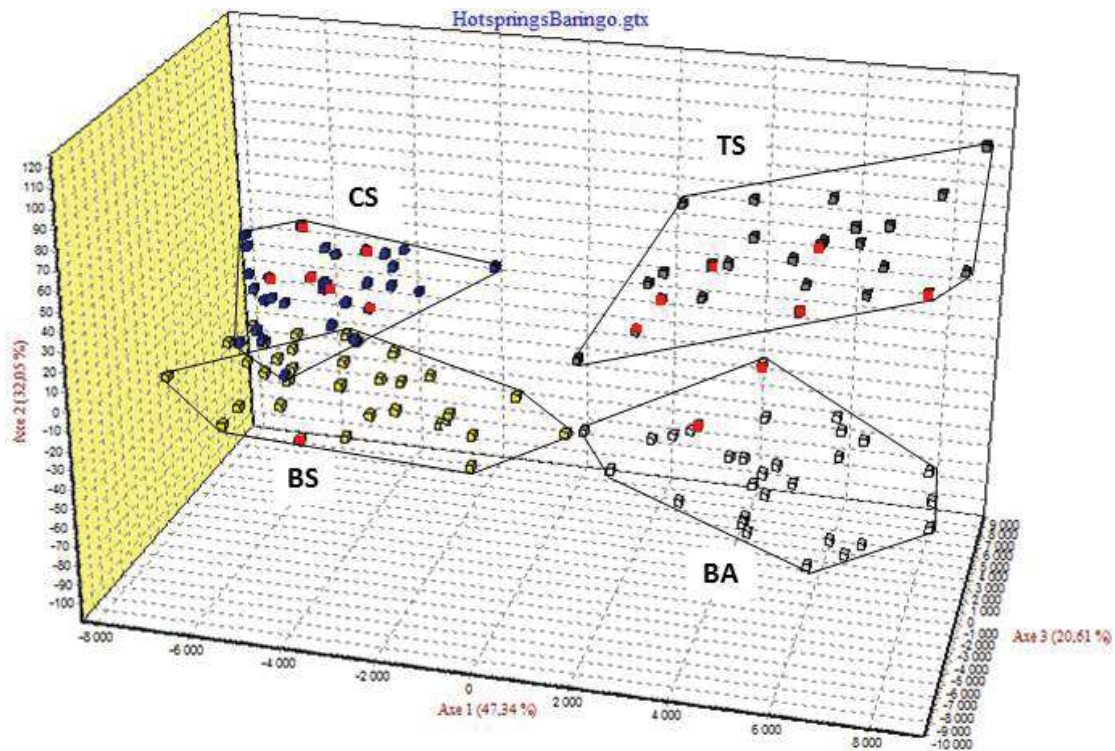


Figure 6: Factorial Correspondence Analysis (GENETIX 4.05) based on individual microsatellite genotypes at 16 loci showing clear separation of the four populations, the three hot spring populations and the lake Baringo population, using the two first axis. CS, Chelaba Spring; BS, Bogoria Hotel Spring; TS, Turtle Spring; BA, Lake Baringo.

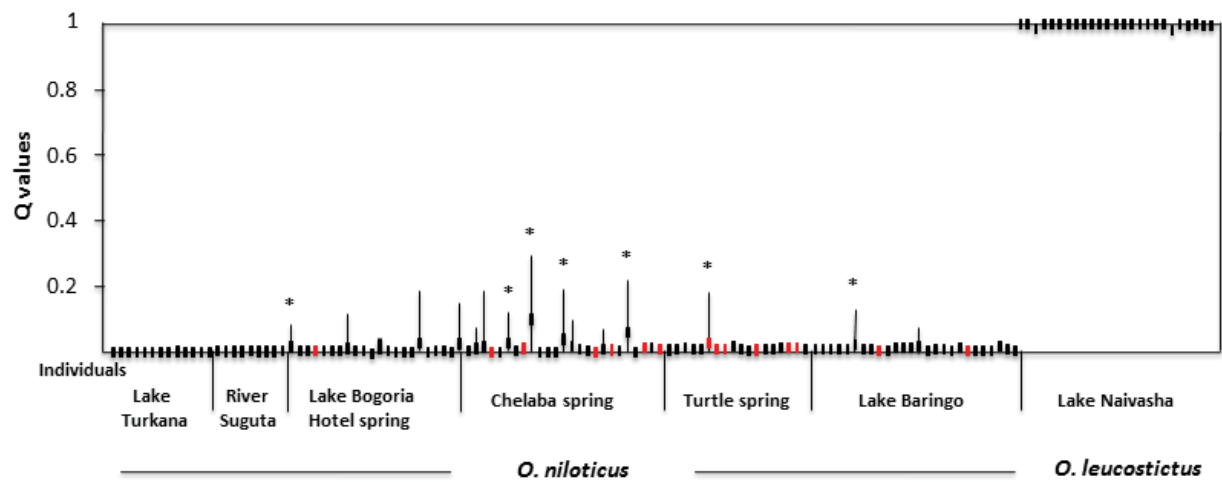


Figure 7: Distribution of specimen membership Coefficient ($Q \pm 90\%$ probability intervals) based on genotypes of *O. niloticus* and *O. leucostictus* in the hot springs and lakes Baringo and Naivasha, identified through microsatellite analysis using STRUCTURE. Each sample along the x axis represents an individual.

Supporting Table and Figures

Table S1: Species specific mtDNA identification, membership probability (STRUCTURE) and Posterior probability (NewHybrids) associated with the introgressed specimens. Post. Prob, posterior probability; *On*, *Oreochromis niloticus*, *Ol*, *Oreochromis leucostictus*, CS, Chelaba Spring; BS, Bogoria Hotel Spring; TS, Turtle Spring; BA, Lake Baringo.

Sample	mtDNA	structure	Structure	NewHybrid	NewHybrid
ID No.	haplotype	membership	pure <i>On</i>	post. Prob.	Post. prob
		Coefficient (<i>On</i>)	%	(<i>On</i>)	backcross (<i>On</i>)
BS0001	<i>On</i>	0.9794	2.06	0.8141	0.1846
BS0004	<i>Ol</i>	0.9975	0.25	0.9983	0.0006
BS10281	<i>On</i>	0.9861	1.40	0.9971	0.0005
BS10292	<i>On</i>	0.9683	3.17	0.9867	0.0120
BS315	<i>On</i>	0.9759	1.57	0.9929	0.0057
CS10223	<i>On</i>	0.9742	2.58	0.9814	0.0171
CS10224	<i>Ol</i>	0.9982	0.18	0.9982	0.0006
CS10226	<i>On</i>	0.9738	0.26	0.8825	0.1161
CS10229	<i>Ol</i>	0.9940	0.06	0.9961	0.0023
CS10230	<i>On</i>	0.8941	10.6	0.0549	0.9400
CS10235	<i>On</i>	0.9638	3.62	0.9574	0.0416
CS10237	<i>On</i>	0.9861	1.39	0.9943	0.0048
CS10240	<i>Ol</i>	0.9978	0.22	0.9982	0.0007
CS10242	<i>Ol</i>	0.9970	0.30	0.9980	0.0010
CS10246	<i>On</i>	0.9322	6.78	0.4192	0.5786
CS10417	<i>Ol</i>	0.9943	0.57	0.9954	0.0033
CS10420	<i>Ol</i>	0.9953	0.47	0.9972	0.0014
BA01887	<i>On</i>	0.9778	2.22	0.9154	0.0833
BA01899	<i>Ol</i>	0.9988	0.12	0.9985	0.0010
BA01921	<i>Ol</i>	0.9921	0.07	0.9987	0.0004

Appendix I continued

TS10255	<i>OI</i>	0.9718	2.82	0.8815	0.1174
TS10256	<i>OI</i>	0.9991	0.09	0.9987	0.0001
TS10257	<i>OI</i>	0.9991	0.09	0.9986	0.0004
TS10263	<i>OI</i>	0.9983	0.17	0.9982	0.0010
TS10267	<i>OI</i>	0.9977	0.23	0.9985	0.0008
TS10268	<i>OI</i>	0.9972	0.28	0.9983	0.0008



Lake Bogoria Hotel Spring

Figure S1: Photograph showing Lake Bogoria Hotel Hot Spring.



Chelaba Spring

Figure S2: Photograph showing Chelaba Hot Spring



Turtle Spring

Figure S3: Photograph showing Turtle Hot Spring.

(II) Population differences in the SNPs associated to sex found in the *amh* gene of Nile tilapia and identification of sex-reversed fish

Titus Chemandwa Ndiwa^{1,2}, Jean-François Agnès², Arnaud Gagnieu³, Etienne Bezault⁴, Dorothy Nyingi¹, Jean-François Baroiller³, Helena D’Cotta^{3*}

1 Kenya Wetlands Research Team, National Museums of Kenya, Kenya

2 Département Conservation et Domestication, UMR IRD 226 CNRS 5554, Institut des Science de l’Evolution, Université de Montpellier 2, Montpellier, France.

3 CIRAD, UMR INTREPID, Dpt. Persyst, Campus International de Baillarguet, F-34398 Montpellier, France

4 UMR BOREA-DYNECAR, Université des Antilles et de la Guyane, Campus de Fouillole B.P. 592 Pointe-à-Pitre, Guadeloupe

*Corresponding author; Helena D’Cotta

Abstract

The Nile tilapia has a complex sex determination system not only dominated by a XX/XY genetic system, but is also influenced by minor genetic factors as well as temperature. Sex-reversed XX males and XY females were found amongst breeders in some wild populations suggesting either influences of temperature or minor genetic factors or both these factors. Temperature and/or genetic influences on sex may be the result of allelic variations on sex-differentiating genes. The aim of the current study was to investigate allelic polymorphisms associated with phenotypic sex in nine wild populations of Nile tilapia living in different thermal regimes, and compare them with two domesticated strains: the Manzala strain for which genotypic sex was known for XX females, XY and YY males and sex-inverted phenotypic males (XX males), and the Bouaké strain. SNP search was performed on three genes (*cyp19a*, *wt1b* and *amh*) implicated in the cascade of sex differentiation in the gonad. These genes are located on two linkage groups LG1, and LG23 which have both been shown to be sex-linked in Nile tilapia. Four SNPs were found in the *cyp19a1a* promoter and the *wt1b*

gene, and correlated with populations but not with sex. SNPs in the *amh* gene are abundant (n=33) and widely distributed along the gene, with some strongly correlated with the phenotypic sex found in exons 2, 6 and 7 and intron 6. Several of these SNPs results in synonymous changes. The sex-linked SNPs appeared population-specific implying that sex determination in Nile tilapia should be analysed at the population rather than at the species level. They were more strongly related to the phenotypic sex in the Kenyan hot spring populations than in the Sudano-Sahelian and Ethiopian populations. Using the recently published *amh* markers, we identified sex-reversed XX males and XY females in several wild populations except in the Turtle hot spring. The *amhy* deleted form of the *amh* is not the universal Y male-marker for Nile tilapia since it is absent in all the males collected from the Ethiopian Rift populations (Lakes Metahara, Hora and Koka). Multivariate analysis conducted across all SNPs show a clustering of the studied populations into four distinct groups suggesting that sex chromosomes in Nile tilapia are still in a continuous state of evolution, with possibly evolution of different sex-linked haplotypes across the entire natural distribution of the species.

1. Introduction

Teleosts have extremely diverse mechanisms by which ovarian or testis differentiation are controlled. Sex determination is the initial decision for this development while sex differentiation is the downstream process generating the testis or ovary phenotypes. Genetic Sex Determination (GSD) triggered by a gene on the sex chromosomes and Environmental Sex Determination (ESD) involving an array of environmental cues including temperature, have evolved numerous times amongst teleost lineages (Baroiller *et al.*, 2009a; Heule *et al.*, 2014) and can even coexist in the same species (Yamamoto *et al.*, 2014). Recently, five male sex-determining genes have been identified on Y chromosomes of different fish species. In four of these cases, the gene has evolved from a duplication of a downstream gene of the sex-differentiating cascade. In the medaka (*Oryzias latipes*), *dmy/dmrt1bY* has been duplicated from *dmrt1*, a conserved testis sex-differentiating gene (Matsuda *et al.*, 2002; Nanda *et al.*, 2002). In a closely related species *O. curvinotus*, the transforming growth factor *gsdf^Y* has been recruited (Myosho *et al.*, 2012). *Amhy* a duplicated gene of the anti-müllerian hormone gene *amh*, has taken the role of master-gene in the Patagonian pejerrey *Odontesthes hatcheri* (Hattori *et al.*, 2012) and even in the temperature-dependent sex determining (TSD) *O.*

bonariensis species (Yamamoto *et al.*, 2014). In three species of *Takifugu*, a single nucleotide polymorphism (SNP) found in the *amh* receptor gene *amhr2*, has been found to be linked to males (Kamiya *et al.*, 2012). In contrast in rainbow trout, the Y testis-triggering gene *sdY* gene has not appeared from a duplicated downstream gene but rather from a duplicated form of the immune gene interferon regulatory factor 9 *Irf9* (Yano *et al.*, 2012).

In the cichlid Nile tilapia (*Oreochromis niloticus*), sex determination and differentiation have been the focus of extended research due to male-growth dimorphism and the need to control sex in this major aquaculture species. Sex determination in this species has proven to be a complex mechanism under the control of various factors, which can broadly be divided into three groups: the sex chromosomes considered as major genetic factors, minor genetic factors and temperature factors (Baroiller *et al.*, 2009b). No sex-determining genes have so far been found in tilapia fish. There is an absence of dimorphic differences between any chromosome pair in male or in female karyotypes (Bezault *et al.*, 2011). Nevertheless, indirect approaches such as progeny testing have demonstrated that *O. niloticus* has a predominant XX/XY sex-determining system while its sister species *O. aureus* has a ZZ/ZW system (Baroiller and Jalabert, 1989). Even so, there is still controversial evidence of which is the sex chromosome pair in tilapia. Foresti *et al.* (1993) and then Carrasco *et al.* (1999), when analysing synaptonemal complex found a terminal non-pairing of the largest pair of bivalents (chromosome 1) only in heterogametic males (XY) and not in the homogametic ones (XX or YY) and suggested that it corresponded to the sex chromosome pair in the Nile tilapia. This large pair of chromosomes presents signs of differentiation into sex chromosomes with accumulation of repetitive elements (Bezault *et al.*, 2011; Martins *et al.*, 2004; Cnaani *et al.*, 2008) as well as sex-segregation differences (Lee *et al.*, 2003; Cnaani *et al.*, 2008). The role of this pair as sex chromosomes has been reinforced by the discovery of an inverted region associated to the genotypic sex (D’Cotta *et al.*, unpublished data). However, the linkage analysis has associated a Y-haplotype dominant locus of males on linkage group 1 (LG1) in the Nile tilapia (Lee *et al.*, 2003; 2004, Cnaani *et al.*, 2008; Lee *et al.*, 2011) located on a small chromosome (Cnaani *et al.*, 2008). Using Restriction Association DNA (RAD) in two families of Manzala, Palaiokostas *et al.* (2013) found two SNPs that explained 96% of the phenotypes, narrowing the SD region to 2cM of LG1. Nevertheless, several variations in segregation patterns have been shown between strains and tilapia species with influences also of LG3 (located on the large chromosome pair) (Lee *et al.*, 2003; Cnaani *et al.*, 2008). In addition, a sex-linked QTL has been found in LG23 and fine-mapping has located the sex-

determining region in an area comprising scaffold 101 assembled from the tilapia genome (Brawand *et al.*, 2014).

Several studies have shown that temperature can override genetic sex determination in the Nile tilapia as has been demonstrated in both domestic and wild populations (Baroiller *et al.*, 1995; Bezault *et al.*, 2007; Baroiller *et al.*, 2009a). In the Bouaké domestic strain, temperature >32°C caused complete masculinisation of some genetic female fries into phenotypic males (Baroiller *et al.*, 1995). The natural distribution of the Nile tilapia represents a very large region across Africa, including Western, Nilo-Sudanian and part of the East African Rift regions, with distinct temperature regimes ranging from Mediterranean, stable tropical climates to cold or extremely hot lakes in Ethiopia and hot springs in Kenya (Baroiller *et al.*, 2009; Bezault *et al.*, 2011; Trewavas, 1983; Nyingi *et al.*, 2009). Detection of masculinisation by high temperature was evidenced using breeders from several wild populations adapted to different thermal regimes but variations amongst populations and families were encountered in the proportion of males achieved by thermal treatment (Baroiller *et al.*, 2009). Progeny testing under controlled conditions provided indirect evidence of the existence of XX males as well as XY females in a wild population of the Volta basin (Bezault *et al.*, 2007).

Family-specific QTLs were associated to temperature-masculinised XX tilapia on LG1, LG3 and LG23 (Wessels *et al.*, 2011). LG23 contains scaffold 101 in which is located the Nile tilapia *amh* gene. Very recently, a deletion of a male-specific duplicated form of the *amh* gene lacking the TGF- β domain has been identified and proposed as the possible sex-determining gene in Nile tilapia referred to as *amhy* (Eshel *et al.*, 2014). *Amh* does indeed have male-specific features. It is expressed in the brain of tilapia in a sex-dimorphic manner at early stages (Poonlaphdecha *et al.*, 2011), followed by a sustained expression in the testes (Ijiri *et al.*, 2008; Poonlaphdecha *et al.*, 2011). Furthermore, it is one of the major genes rapidly up-regulated in temperature-masculinisation (Poonlaphdecha *et al.*, 2013). In addition, allelic differences were found in the *amh* gene with higher effects for the phenotypic males obtained by elevated temperatures (Wessels *et al.*, 2014).

The interaction of genetic and temperature influences on sex determination in the Nile tilapia might stem from SNP variations of sex-differentiating genes such as found for the allelic variant specific for males found in the *amhr2* gene of fugu (Kamiya *et al.*, 2012). In the present study, allelic association to sex was analysed in several wild populations of Nile tilapia living in different thermal regimes including some inhabiting natural hot lakes characterized by high temperatures above 32°C, with newly discovered populations living in

hot springs of the Loboï Swamp in Kenya (Nyingi *et al.*, 2009; Nyingi and Agnèsè, 2012; Ndiwa *et al.*, 2014). Such environment exposes the fish to critical masculinizing temperatures, which would under normal circumstances radically affect the sex ratios of these populations. In spite of this, these populations continue to have normal sex-ratios, which indicate that they have most likely adapted to this situation. To better understand how a thermal regime exerts its effect on wild populations, we performed a SNP screen in specific genes implicated in sex-determination or sex-differentiation, located in two linkage groups associated to sex, LG1 and LG23. One of the genes studied is the gonadal aromatase gene *cyp19a1a* which encodes the enzyme catalysing the conversion of androgens into oestrogens, and which is down-regulated in temperature-masculinisation (D’Cotta *et al.*, 2001). A second gene selected was the Wilms Tumor 1 gene *Wtlb*, which is a fish duplication of the vertebrate *wtl* gene implicated in gonad development (Brennan and Capel, 2004). Both *wtlb* and *cyp19a1a* are located on LG1 and SNP variations were found in the heterogametic sex (Lee and Kocher, 2007). Finally, extensive analysis was also performed on the *amh* gene located on LG23. We have also taken advantage of the *amhy* as a marker to identify putative XX phenotypic males (absence of *amhy*) or XY phenotypic females (presence of *amhy*) amongst the wild populations of Nile tilapia in Africa in order to search for sex-reversed fish. We then compared them to the domesticated strains whose phenotypes and genotypes have been well studied for their sex determination systems.

2. Materials and Methods

2.1 Fish sampled and phenotypic sex identification

Eleven populations of Nile tilapia have been investigated; nine are natural populations while two, the Bouaké (Ivory Coast) and the Manzala (Egypt) are domesticated strains (Table 1). From the Kenyan Rift-Valley (Fig. 1), a total of 5 populations have been studied, three from hot springs (Lake Bogoria Hotel Spring, Chelaba Spring and Turtle Spring) and two from middle/standard temperature habitats (Lakes Baringo and Turkana). Also three populations are from the Ethiopian Rift-Valley region, one from a hot spring (Lake Metahara) and two from cold altitude lakes (Hora and Koka). This has been completed by three populations adapted to standard tropical temperature regime from the Nilo-Sudanian area: one population from Sudano-Sahelian region (Kpandu) and one from the Nile basin (Manzala), while the Bouaké is a synthetic strain composed initially with individuals from both Volta and Nile

rivers populations. These populations have been chosen because they experience different thermal conditions (Bezault *et al.*, 2007; Nyingi *et al.*, 2009 and Ndiwa *et al.*, 2014). The three hot springs populations are exposed to high water temperatures of approximately 36 °C, conditions previously associated with masculinisation of XX females (Baroiller *et al.*, 1995). The population of Lake Metahara connected to the Awash River, is also possibly adapted to high temperature as the lake is supplied by hot water from a hot spring at a temperature of 43 °C (Goërner *et al.*, 2005). In the hot springs, temperatures vary between 32 °C and 39 °C. The Lake Baringo is located between the Loboï hot springs and Lake Turkana and is relatively shallow and has water temperatures around 25 °C. Lake Turkana population lives in medium water temperatures (25 °C) that are slightly saline with pH of 9.3 (Yan *et al.*, 2002). Under experimental conditions both the Bouaké and Manzala domesticated strains resist temperatures of 36 °C with male proportion varying between offspring.

Fish were caught a few meters from the shore of all lakes and hot springs using dip nets. Individual phenotypic sex was determined by macroscopic observation of the gonads. For each population, an average of five individuals characterized as phenotypic males or females were considered for the study, totalling 192 fish analysed. On each fish, a fin clip was collected and preserved in absolute ethanol for DNA analysis.

2.2 Genomic DNA extraction, PCR amplification and sequencing

DNA was isolated from approximately 50mg of fin samples by first digesting at 55°C overnight using 10 µl proteinase K in 180 µl of extraction buffer solution (1M Tris, 0.5M NaCl₂, 1% SDS). The DNA was then extracted using either a phenol/chlorophorm procedure as previously described by Bezault *et al.*(2011) or by the salting-out protocol described by Aljanabi and Martinez (1997). The concentration of the extracted DNA was quantified using NanoDrop spectrophotometer (NanoDrop, Technologies) and concentration of each sample standardized to 30 ng/µl.

PCR reactions were performed using 60 ng of genomic DNA in a total reaction volume of 50 µl containing 5 µl of 10×MP Biomedicals buffer (Tris Cl, KCl, (NH₄)₂SO₄, 15 mM MgCl₂), 2 µl of 10 mM dNTP (0.2 mM of each dNTP), 1 µM of each primer, and 0.5 units of Taq polymerase. PCR T_m amplifications and cycle timing varied depending on the primer pair used (table 2), with primers ordered from Eurogentec (Belgium).

PCR conditions involved an initial denaturing for 5 minutes at 95°C, followed by 40 cycles of amplification, consisting in a denaturing for 1 minute, annealing for 1 minute, and elongation for 3 minutes at 72°C, and ending with 10 minutes elongation at 72°C. After agarose verification of PCR product, they were column purified and then a Sanger sequencing was carried out on an ABI-3130 or privately by Eurofins genomics (Germany), on both the forward and reverse strands using their corresponding specific primers.

2.3 Sequence analysis to identify SNP differences

Chromatogram verification was performed manually where sequencing inconsistencies were found. Sequences were then aligned using both MUSCLE (Edgar, 2004) as well as CLUSTALW with Seaview Version 4.3.5 software (Gouy and Gascuel, 2010). Alignments were corrected manually. SNPs were then manually identified using BioEdit 5.09 software (Hall, 1999). Comparisons and alignments were performed using several reference genes (Figures S1-S3) obtained from: GenBank (table 2), available on Nile tilapia Genome at Ensembl (www.ensembl.org, release 76 last accessed Ago. 2014) as well as from the Bouillabase ([gbrowse:http://cichlid.umd.edu/cgi-bin/gb2/gbrowse/Tilapia_broad_anchored_v1/](http://cichlid.umd.edu/cgi-bin/gb2/gbrowse/Tilapia_broad_anchored_v1/)).

The expected number of male and female phenotypes for each population was calculated based on Hardy-Weinberg Equilibrium (HWE). We then calculated the relationship between sex-linked SNPs and phenotypic sex using Chi-Square test with p value based on Monte Carlo (Hope, 1968). The obtained SNPs data was coded and analysed using Factorial Correspondence Analysis (FCA) on Genetix Version 4.05 (Belkhir *et al.*, 2004) to examine the relationship between the populations with respect to SNP markers. Lastly, linkage disequilibrium between allele pairs was calculated using Genetix version 4.05.2 (Belkhir *et al.* 2004).

2.4 Use of X and Y markers to identify genotypic sex

PCR amplifications were performed on all the individuals using primers Amh-F2443 and Amh-R3490 (table 2) allowing discrimination between XX individuals, and XY/YY individuals (Eshel *et al.*, 2014). PCR reactions were performed using 60 ng of genomic DNA in a total reaction volume of 25 µl containing 2.5 µl of 10×MP Biomedicals buffer (Tris Cl,

KCl, (NH₄)₂SO₄, 15 mM MgCl₂, 2 µl of 10 mM dNTP (0.2 mM of each dNTP), 1 µM of each primer, and 0,75 U of Taq polymerase. PCR conditions involved an initial denaturing for 5 minutes at 95 °C, followed by 38 cycles of amplification, consisting in a denaturing at 95 °C for 30 second, annealing at 58 °C for 30 seconds and elongation for 2 minutes at 72 °C ending with a 10 minute elongation at 72 °C.

An additional discrimination to define the genotypic sex was performed using the X-specific and Y-specific primer couples (table 2) as described by Sun *et al.* (2014). The PCR reaction was performed as above and conditions similar with an annealing at 60°C for 35 cycles. All PCR products were migrated on a 1.5% agarose gel.

3. Results and Discussion

It is well known that in Nile tilapia the genetic sex determination XX/XY can be switched due to high temperature influences (Baroiller *et al.*, 2009a). Sex-reversed XX males and XY females were found amongst breeders in some wild populations (Bezault *et al.*, 2007) suggesting either influences of temperature or minor genetic factors. However, how these influences occur and how frequently the temperature exerts its masculinizing effect in the wild remain unknown. Temperature and genetic influences on sex may be the result of allelic variations on sex-differentiating genes. The current study was carried out in nine wild populations in order to investigate allelic polymorphisms related to the phenotypic sex of Nile tilapia, on three sex-differentiating genes (*cyp19a1a*, *wt1b* and *amh*) previously found to have allelic variants linked to males (Lee & Kocher, 2007; Eshel *et al.*, 2014; Wessels *et al.*, 2014). The *cyp19a1a* and *wt1b* genes are located on linkage groups LG1 (Lee and Kocher, 2007), while *amh* is located on LG23 (Shirak *et al.*, 2006). Studies have shown that these two LGs are associated to sex (Lee *et al.*, 2003; 2004; Cnaani *et al.*, 2008; Eshel *et al.*, 2012).

The nine wild populations cover a large portion of the natural distribution of Nile tilapia in Africa, as well as the entire spectrum of thermal regime encountered by the species in the wild. A selection of 192 individuals were analysed with 99 classed as phenotypic males. Specific sex genotypes produced by chromosome manipulation (Baroiller and Jalabert, 1989; Scott *et al.*, 1989) were also used for the Manzala domesticated-strain, amongst which were included 5 XX males, 5 YY supermales and 1 YY female. Individuals (n=5-7) with poor sequence reads were discarded. We identified a total of 41 SNPs of which 33 allelic variants were found in the *amh* gene, 4 in the aromatase *Cyp19a* promoter and 4 in the *Wt1b* gene. In

addition, our study also found insertions and deletion varying in size for the *cyp19a1a* promoter and *amh* gene, some of which were unique to certain populations.

3.1 SNP identified on the *cyp19a1a* promoter region

Our previous studies performed on the domesticated Bouaké and Manzala strains have shown that the sex-differentiating gene aromatase *cyp19a1a* is implicated in the molecular mechanism of ovarian differentiation (D’Cotta *et al.*, 2001) and temperature masculinisation of XX female fries causes a down-regulation of this gene (Poonlaphdecha *et al.*, 2013). The present study was focused on a 499 bp fragment of the 5’ flanking region of the tilapia *cyp19a1a* gene in which an allelic variant was previously identified g.2124G>A with males being heterozygotes (AG), and females homozygotes (AA) (Lee & Kocher, 2007). This SNP was one of the four polymorphisms found in the *cyp19a1a* promoter region (Table 3) but we detected here no polymorphism associated to the phenotypic sex. It was heterozygous (Cyp_2124GA) with a 38.50 % frequency in all Lake Kpandu individuals and in all fish belonging to the Ethiopian Rift Valley, Lakes of Hora, Metahara and Koka, and in 9 fish out of 13 in the Bouaké strain. All fish from Lake Turkana and the Turtle hot spring were homozygous Cyp_2124AA. Although some individuals had the G allele, the majority of the Chelaba and Bogoria hot springs fish and the domesticated Manzala strain had the A variant.

The three other SNPs (Table 3) showed much higher polymorphism between the hot spring populations of the Loboï swamp and Lake Baringo compared to all the other populations. However, none of these SNPs were correlated with sex. In the hot springs, the Cyp_2016C>T variant predominated together with the Cyp_2175A>G showing a 18.28 % frequency of heterozygous Cyp_2175AG. An exhaustive analysis of the 5’ flanking region of *cyp19a1a* was recently performed in tilapia (Böhne *et al.*, 2013). Based on this analysis, the Cyp_2175 SNP was located in a binding site for SF-1, a transcription factor implicated in the activation or repression of the aromatase *cyp19a1a* gene. We can therefore hypothesize that this SNP might be exerting a repressive action on the *cyp19a1a* gene in some of the hot springs fish.

In addition to SNPs, we detected three nucleotide deletions (table 4) when comparing sequences between our populations with those of Genbank AF472620 and Ensembl ENSONIG00000000155. Two of these deletions (Cyp_1988 and Cyp_1993) involved a single nucleotide each, while one (Cyp_2036-37) involved two nucleotides. The G deletion at Cyp_1988 was observed only in the three Ethiopian populations (Metahara, Hora and Koka)

and might be a specificity of this region since it was observed in both the subspecies *O. n. cancellatus* and *O. n. filoa*. The A deletion located on Cyp_1993 was only seen in the Bogoria and Chelaba hot springs (11/40) with one SNP A>T change seen in 3 individuals. Lastly, deletion on Cyp_2036-37 was observed in every studied population of *O. niloticus* with respect to the references.

Pairwise analysis of the SNPs and deletions identified in the *cyp19a1a* promoter showed linkage disequilibrium (table 5) elevated for the Bogoria and particularly for the Chelaba populations between Cyp_1992, Cyp_2016, Cyp_2124 and Cyp_2308.

The first analysis performed on the *cyp19a1a* promoter identified SNPs and deletions shared between the Manzala strain and the Lake Turkana population indicating a close affinity of Turkana fish to fish of Nilotic origin, with some affinities also visible in fish from the Ethiopian Rift Valley.

3.2 SNPs identified on the *wt1b* gene between exon 5 and 7

A sex-linked SNP was previously detected by Lee and Kocher (2007) on the *wt1b* gene with females being homozygous (g.686CC) and males being heterozygous (g.686CG). Because of this, we restricted the SNP search to a 899 bp fragment comprising exon 5 to exon 7. Four SNPs were found on the *wt1b* gene but the 686CC named wt1b_684 in our study (table 6) was not amongst these. Three SNPs were located on intron 6, while the wt1b_649 SNP was located in the codon between exon 5 and exon 6 causing a non-synonymous substitution. There was a clear population-associated polymorphism found for the Wt1b_398 SNP. Wt1b_398GG was found in 22 out of 23 Kpandu fish, while in contrast all the Ethiopian Rift Valley fish as well as the Turkana fish were found to be homozygous Wt1b_398AA. A higher variability was observed for the Bogoria, Chelaba, Turtle and Baringo fish with Wt1b_398AA, Wt1b_398GG or Wt1b_398AG. The Manzala YY males were mostly Wt1b_398AA whereas the XX males or females were Wt1b_398GG or Wt1b_398AG (also found for the Bouaké fish). Similarly, high polymorphism was observed on SNP Wt1b_464 with specimens from six populations (Bouaké, Kpandu, Manzala, Chelaba, Baringo and Turkana) being homozygous (Wt1b_464AA or Wt1b_464GG), while two hot spring populations (Bogoria and Turtle) were either homozygous (Wt1b_464AA) or heterozygous (Wt1b_464AG). Similarly, the three Ethiopian populations (Metahara, Koka and Hora) were homozygous (Wt1b_464AA). The SNPs Wt1b_466 and Wt1b_649 were less heterozygous

than Wt1b_398 and Wt1b_464. Heterozygous SNPs were observed on only three populations (Bouaké, Kpandu, Manzala) for Wt1b_466, and two (Manzala and Baringo) for the Wt1b_649. No deletions and insertions were observed on this gene. Like *cyp19a1a*, SNPs observed in the wt1b region segregated across populations but was not linked to phenotypic sex.

A linkage disequilibrium was seen between Wt1b_308, Wt1b_464, Wt1b_464 and Wt1b_649 for the Baringo population as well as the Bouaké strain (table 7).

3.3 High amount of SNPs identified on the *amh* gene

We identified an astonishing large amount of SNPs (n=33) in the *amh* gene. These SNPs were found throughout the gene (table 8), but were particularly concentrated in intron 6 (13), exon 6 (4) and exon 7 (8). Out of the 19 SNPs located on the coding region of the *amh* gene (different exons), 6 involved changes on the amino acids while the rest were synonymous (Table 8). The occurrence of deletions and insertions in the *amh* gene was also observed on different populations of *O. niloticus* as shown on Table 9. All observed insertions and deletions occurred on intron 6. Two deletions located from 2677 to 2694 of 17 nt and from 2895 to 2914 of 18 nt were particularly noticeable since they only occurred in the three Kenyan hot spring populations together with Lake Baringo fish. Furthermore, these deletions had strong linkage disequilibrium.

3.4 Finding of SNPs in the *amh* gene associated with the phenotypic sex

In addition to presenting the highest number of SNPs, the *amh* gene also showed SNPs with strong relationship to the phenotypic sex. The initial search for SNPs associated to sex was grouped, but we found large population differences. Even though these associations were observed in most populations studied, stronger associations were observed on the Kenyan populations. A strong correlation was observed for the amh_202 SNP for both the Kpandu population and the Bouaké and Manzala strains with XX fish being homozygous amh_202AA while males were amh_202GG or amh_202GA. This polymorphism was not seen in the other populations. Likewise, females from the Kpandu, Manzala and Metahara populations had a C allele on the amh_754, while the males were mostly heterozygous amh_754TC (including the YY Manzala fish) or homozygous amh_754TT. Within the Kenyan populations (Bogoria,

Chelaba, Turtle, Baringo and Turkana), 12 SNPs were observed on the *amh* gene that had strong association to phenotypic sex (table 10). Analysis of these SNPs indicated significant deviations from the HWE of some populations. For instance, three populations (Bogoria, Chelaba and Turtle) showed significant deviations on *amh_2509*, while four SNPs (*amh_597*, *amh_611*, *amh_2586*, *amh_2807*) deviated significantly in two populations each. Interestingly, the SNP *amh_597* G>T caused a change in amino acid with females having the T variant while males had the G variant thereby, changing from Glutamic acid to Aspartic acid. Likewise, the SNP *amh_611* females had the G variant while males had the A allele changing the amino acid from Serine to Asparagine, respectively. The remaining two SNPs (*amh_2618* and *amh_2655*) recorded significant deviation from HWE each in one population (Table 10).

Very recently, two studies have shown SNPs linked to sex in the *amh* gene of Nile tilapia. Eshel *et al.*, (2014) revealed a 233 bp deleted form of the *amh* gene specific to males when carrying the Y chromosome, on XY and YY individuals, which the authors refer to as *amhy*. In the sequencing of both *amh* and *amhy* they identified a SNP at position 1274 (corresponding to *amh_2992* in the current study) with a T allele present in XX female fish, a C in XY males and both C/T in YY individuals. This SNP was also found in our study but no T and C sex-association was noted in our Manzala strain which is from the same origin as the one used by Eshel *et al.*, (2014). We found XX female fish with the C allele and with both C/T alleles (Fig. 2). Likewise, we found XY males carrying the T allele as well as C/T combination, while YY fish had the T variant. Similar observations were recorded for the Bouaké strain, with females either carrying the C or the C/T alleles. In wild populations such as Lake Turkana, some females had the C allele while others a T allele, as well as both C/T alleles. These combinations were also seen in males. We did not find the A>G substitution at position 1275 found in the *amhy* form since all our fish had only the G allele. We did not find either the SNP at position 1403 (corresponding to *amh-3112*) with a G allele for the *amhy* vs the T for the *amh*.

Another important SNP identified (ss831884014) was associated with temperature masculinisation and autosomal sex-reversal (Wessels *et al.*, 2014). This SNP located at position 2183 in our study was not polymorphic in the individuals studied from different populations or strains.

Considering the different results obtained from SNP analysis, there is existence of significant deviations that varied among populations on many SNPs. These results confirm previous

findings by Baroiller *et al.* (2009a; 2009b) and Cnaani *et al.* (2008), that the associations to phenotypic sex are population-specific. Therefore, determination of sex in *O. niloticus* should be analysed at a population level.

3.5 FCA analysis of SNPs

A Factorial Correspondence Analysis (FCA) including all SNPs observed in the three sex-linked genes clustered *O. niloticus* populations into four distinct groups (Fig. 3). The first group comprised the three Kenyan hot spring populations of Bogoria, Chelaba and Turtle as well as the Lake Baringo population. The second group included exclusively the Lake Turkana population, separating it from all other Kenyan as well as Nilotic populations. The third group comprised two of the three Ethiopian populations (the hot spring of Lake Metahara and cold Lake Koka populations), and the fourth group consisted in all Nilo-Sudanian *sensu lato* populations (the Nilotic Manzala strain, the Sahelo-Sudanian Kpandu population as well as the synthetic Bouaké strain), but also, less expectedly, the Ethiopian population of the cold crater lake of Hora. This observation suggests that sex chromosomes in Nile tilapia are still in a continuous state of evolution, with possibly evolution of different sex-linked haplotypes across the entire natural distribution of the species, hence the observed differences between the studied populations.

3.6 Use of X and Y markers to identify sex-reversal individuals

We found association of the SNPs in the *amh* gene with the phenotypic sex. However, this can be different from the genotypic sex since sex-reversal in tilapias is under the influence of both high temperatures and minor genetic factors (Baroiller *et al.*, 2009a). We therefore, exploited the recent published “amhy” markers as described in Eshel *et al.* (2014) which discriminates two bands, a ~1000 bp size band present in XX females and a second ~800 bp band only present in males carrying the Y chromosome, XY and YY individuals. In the domesticated Manzala strain, the *amh* markers identified correctly all our genetic XX females, XY males, the five YY supermales as well as an YY female (Fig. 4). It also allowed the identification of sex-reversed XX males and an XY female. A XY male that gave biased sex-ratios in progeny testing was in fact identified as an XX male. For the Bouaké strain the 800 bp markers identified a number of individuals which did not correspond to the supposed genotype. When using the markers in the wild populations, they indicated the presence of several sex-reversed XX males although no XY females. A close look at the different

populations evidenced that none of the males from the Ethiopian populations of Lake Metahara, Lake Koka and Lake Hora had the 800 bp band considered to be an *amhy* form (Eshel *et al.*, 2014). These results suggest that the *amhy* markers do not discriminate males in all Nile tilapia populations.

We then analysed the *amhy* markers in the Kenyan wild populations (Fig. 5) and here the majority of males possessed both the XY and YY specific 800 bp band as well as the 1000 bp band. Females in contrast had mostly the unique 1000 bp band, which indicated that the *amhy* most likely discriminated correctly the genotypic sex. Contrary to other populations, the Kenyan populations had five fish out of 97 individuals that were found to be sexually reversed females (phenotypic females that had XY genotypes), while ten phenotypic males had supposedly XX genotypes. Both Chelaba Spring and Lake Baringo populations had three reversed males each, while Lake Bogoria Hotel Spring and Lake Turkana populations had two reversed males each. Considering females only, Lake Baringo had 2 inversions while Lake Bogoria Hotel Spring, Chelaba Spring and Lake Turkana had 1 inversion each. Surprisingly, no sex-reversal was detected in males and females from the Turtle Spring. A possible explanation for this occurrence is that individuals from this population may have been less sensitive to masculinising effect of high temperature compared to individuals from the other two hot springs. This hypothesis seems to agree with the findings of Baroiller and D’Cotta (2001) who observed that some progenies of Nile tilapia displayed higher sensitivity to temperatures resulting in high proportions of males in their sex ratios, while others showed insensitive balanced sex ratios. Lack of inversion may have also resulted from lower water temperature experienced by this population (32-34 °C) compared to the other two hot spring populations (36 °C). Our earlier studies using FISH (unpublished data), which showed presence of reversed individuals within the hot spring (Supporting information Figure S1) disputes this theory. An alternative explanation to this finding is that the markers used in our study failed to identify reversed individuals from Turtle spring population.

We wanted to further validate the presence of these sex-reversed individuals by also using very recently published X and Y specific markers also located on scaffold 101 of LG23, with one of the markers tightly linked to the SD locus, permitting the distinction between normal XY males and YY males (both viable in tilapia) (Sun *et al.*, 2014). However these markers could not be used due to complications in the PCR testing. Nevertheless, these additional markers should be tested in all our wild populations to see whether they discriminate the genotypic sex in a similar manner as the duplicated *amh* form.

The presence of sex-reversed XX males and XY females in several wild populations in this study coincides with our previous studies in which both XX males and an XY female were indirectly identified *via* progeny testing of wild breeders (Bezault *et al.*, 2007). We are however, surprised by the elevated number of sex-reversed individuals encountered. An alternative hypothesis is that the *amhy* is not the sex determining gene but that the *amh* gene is in close proximity to the sex determining locus and in some individuals recombination does not occur. The fact that the *amh* markers did not show a second *amh* form (*amhy*) with a deletion in all the Ethiopian males indicates that this is not an universal male sex-marker (or sex-determinant) in Nile tilapia populations. In addition, other deletions were found amongst the Kenyan populations that were in intron 7 but their presence suggests that a more detailed analysis of all the *amh* gene sequences should be carried out to detect more regions of deletions or insertions between sexes and populations.

Sex reversal of both males and females was observed in two populations of *O. niloticus* from the hot springs (Bogoria and Chelaba Springs). These hot springs are characterized by water temperature of ~36 °C and do not undergo significant changes annually since the springs are served by faults and fractures that allow water from deep aquifers to rise to the surface (Deocampo, 2002; Ashley *et al.*, 2004). The possible presence of XX males amongst our different populations living in varying temperature regimes in the present study and that of Bezault *et al.*, (2007) and Baroiller *et al.*, (2009a) suggests that some wild individuals were sensitive to the effects of high temperatures, causing masculinisation of XX females. Environmental sex reversal is the modification of primary sex during ontogeny into a sex opposite to that coded by the genotype (Bull, 1993; Valenzuela *et al.*, 2003). According to Baras *et al.* (2001) sex reversal is a conservative process that enables the species to not produce sex ratios that can jeopardize its survival. Progeny differences for temperature masculinisation have been detected previously with some progenies or individuals not being sensitive to temperature (Baroiller *et al.*, 1995; Baroiller *et al.*, 2009a). It is therefore, not surprising that for the survival of a population living particularly in hot springs, both genetic and temperature influences coexist so that the sex-ratios are not biased.

The existence of several sex-reversed XY females in wild populations detected in the present study is rather astonishing although it correlates with the finding of an XY female in a previous progeny testing of wild breeders (Bezault *et al.*, 2007). We think that these individuals may result from additional autosomal genetic factors influencing sex which we called minor genetic factors.

In conclusion, our study has detected several SNPs in three sex-differentiating genes *cyp19a1a*, *wt1b* and *amh* amongst wild populations living in different temperature regimes as well as in domesticated strains. Several deletions and insertions were found on *cyp19* and *amh* genes particularly in populations living in hot springs above 34°C constant temperatures, known to cause masculinization and leading to male biased sex ratios. We hypothesized that these deletions and insertions may be responsible for adapting the populations against the effects of high water temperature. Our study shows SNPs in the *amh* gene strongly associated to the phenotypic sex notably in the Kenyan Hot spring populations. The SNPs associated to sex varied depending on the population. In addition, by using *amh* markers we have detected sex reversed XX males as well as XY females in all populations except in the Turtle hot spring. Because none of the males in three Ethiopian populations could be discriminated, the *amh* deleted form is most likely not the tilapia master sex-determinant, but may be in close proximity to the sex-determining locus.

4. Acknowledgements

We express our sincere gratitude to Catherine Ozouf-Costaz (Museum National d'histoire Naturelle, Paris), Elodie Pepey (CIRAD, INTREPID), Joseph Gathua (National Museums of Kenya), and Frederique Cerqueira and Erick Desmarais (Plateforme Génomique Environnementale du Labex Centre Méditerranéen Environnement Biodiversité) for their technical support during the study. We are also grateful to Kenya Wetlands Biodiversity Research Team, CIRAD, IRD and MAE for their financial support. YY and XX males from the Manzala strain were provided by C Méléard (CEFRA-Liege Univ).

5. References

- Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* 25(22): 4692-4693.
- Ashley GM, Maitima Mworira J, Muasya AM, Owen RB, Driese SG, Hover VC, Renaut RW, Goman MF, Mathai S, Blatt SH. 2004. Sedimentation and recent history of a freshwater wetland in a semi-arid environment: Loboï Swamp, Kenya, East Africa. *Sedimentology* 51, 1301–1321.
- Baras E, Jacobs B, Mélard C. 2001. Effect of water temperature on survival, growth and phenotypic sex of mixed (XX-XY) progenies of Nile tilapia *Oreochromis niloticus*. *Aquaculture* 192, 187-199.
- Baroiller JF, Chourrout D, Fostier A, Jalabert B. 1995. Temperature and sex-chromosomes govern sex ratios of the mouthbrooding cichlid fish *Oreochromis niloticus*. *J. Exp Zool* 273: 216-223.
- Baroiller JF, D'Cotta H, Bezault E, Wessels S, Hoerstgen-Schwark G. 2009a. Tilapia sex determination: Where temperature and genetics meet. *Comp Biochem Physiol A Mol Integr Physiol.* 153(1):30-8
- Baroiller JF, D'Cotta H, Saillant E. 2009b. Environmental effects on fish sex determination and differentiation. *Sex Dev.*3(2-3):118-35.
- Baroiller JF, and D'Cotta H. 2001. Environment and sex determination in farmed fish. *Comp. Biochem. Physiol. C: Toxicol. Pharmacol.* 130: 399-409.
- Baroiller JF, Fostier A, Cauty C, Rognon X, Jalabert B. 1996a. Significant effects of high temperatures on sex-ratio of progenies from *Oreochromis niloticus* with sibling sex-reversed males broodstock. In: Pullin, R.S.V., Lazard, J., Legendre, M., Amon Kotias, J.B., Pauly, D. _Eds., Third International Symposium on Tilapia in Aquaculture. ICLARM Conf. Proc., Manila, Philippines, 41 pp. 333–343.
- Baroiller JF, Jalabert B. 1989. Contribution of research in reproductive physiology to the culture of tilapias. *Aquat. Living Resour.* 2: 105–116.
- Baroiller JF, Nakayama I, Foresti F, Chourrout D. 1996b. Sex determination studies in two species of teleost fish, *Oreochromis niloticus* and *Leporinus elongatus*. *Zool. Stud.* 35, 279–285.
- Beardmore JA, Mair GC, Lewis RI. 2001. Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture* 197 : 283–301.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

- Bezault E, Balaesque P, Toguyeni A, Fermon Y, Araki H, Baroiller JF, Rognon X. 2011. Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genet.* 12:102.
- Bezault E, Clota F, Derivaz M, Chevassus B, Baroiller JF. 2007. Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272, S3–S16.
- Bezault E, Ozouf-Costaz C, D'Hont A, Volff JN, Rognon X, Baroiller JF. 2001. Structure and evolution of pure and hybrid genomes of Tilapia. *Chromosom. Res.* 9 (suppl. 1): 30.
- Böhne A, Heule C, Boileau N, Salzburger W. 2013. Expression and sequence evolution of aromatase *cyp19a1* and other sexual development genes in East African cichlid fishes. *Mol Biol Evol.* 30(10):2268-85.
- Brawand D, Wagner C. et al. (68 co-authors). 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 13726 doi :10.1038.
- Brennan J, Capel B. 2004. One tissue, two fates: olecular genetic events that underlie testis versus ovary development. *Nature Genetics*, 5:509-521.
- Bull JJ. 1983. Evolution of sex-determining mechanisms. Benjamin-Cummings, Menlo Park, USA. 316 pp.
- Carrasco LAP, Penman DJ, Bromage N. 1999. Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. *Aquaculture*, 173, 207-218.
- Chang X, Kobayashi T, Senthilkumaran B, Kobayashi-Kajura H, Sudhakumari CC, Nagahama Y. 2005. Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). *Gen Comp Endocrinol.* 141:101–115.
- Cnaani A, Lee BY, Zilberman N, Ozouf-Costaz C, Hulata G, Ron M, D'Hont A, Baroiller JF, D'Cotta H, Penman DJ, Tomasino E, Coutanceau JP, Pepey E, Shirak A, Kocher TD. 2008. Genetics of sex determination in tilapiine species. *Sex Dev.*;2(1):43-54.
- Cnaani A, Tinman S, Avidar Y, Ron M, Hulata G. 2003. Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F2 tilapia hybrid. *Aquaculture*, 223(1):117–128.
- D'Cotta H, Fostier A, Guighen Y, Govoroun M, Baroiller, JF. 2001. Aromatase plays a key role normal and temperature induced sex differentiation of tilapia *Oreochromis niloticus*. *Molecular Reproduction and Development* 59: 265-276.
- D'Cotta H, Baroiller JF, Soler L, Bocs S, Conte M, D'Hont A, Ozouf C, Kocher T. 2013. Genomic characterisation of Chr3-LG3: an ancient sex chromosome in the cichlid tilapia? Conférences Jacques Monod, Roscoff, France. Poster.

- Deocampo DM. 2002. Sedimentary processes and lithofacies in lake-margin groundwater-fed wetlands in East Africa. In: Ashley, G.M., Renaut, R.W. editors. Sedimentation in Continental Rifts, vol. 73. Special Publication-SEPM, Tulsa, OK, pp. 295–308.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792-1797.
- Eshel O, Shirak A, Dor L, Band M, Zak T, Markovich-Gordon M, Chalifa-Caspi V, Feldmesser E, Weller JI, Seroussi E, Hulata G, Ron M. 2014. Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 15: 774. doi: 10.1186/1471-2164-15-774.
- Eshel O, Shirak A, Weller JI, Slossman T, Hulata G, Cnaani A, Ron M. 2010. Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). *Animal Genetic* 42: 222-224.
- Foresti F, Oliveira C, Galetti PM, de Almeida-Toledo LF. 1993. Synaptonemal complex analysis in spermatocytes of tilapia, *Oreochromis niloticus* (Pisces, Cichlidae). *Genome*, 36, 1124-1128.
- Goërner A, Jolie E, Gloaguen R. 2005. Tectonic induced growth of Lake Beseka (Main Ethiopian Rift)? EGU Conference. Geophysical Research Abstracts, vol. 7. Copernic, European Geosciences Union, Vienna, Austria.
- Gouy MG, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27(2): 221-224.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95-98.
- Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, Fernandino JI, Somoza GM, Yokota M, Strüssmann CA. 2012. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc Natl Acad Sci U S A.* 109: 2955–2959.
- Heule C, Salzburger W, Böhne A. 2014. Genetics of sexual development: an evolutionary playground for fish. *Genetics.* 196(3):579-91.
- Hope ACA. 1968. A simplified Monte Carlo significance test procedure. *J. Roy, Statist. Soc. B* 30, 582–598.
- Ijiri S, Kaneko H, Kobayashi T, Wang DS, Sakai F, Paul-Prasanth B, Nakamura M, Nagahama Y. 2008. Sexual dimorphic expression of genes in gonads during early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*. *Biol. Reprod.* 78: 333–341.
- Kamiya T, Kai W, Tasumi S, et al. (16 co-authors). 2012. A trans-species missense SNP in amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet.* 8:e1002798.

- Lee BY, Coutanceau, JP, Ozouf-Costaz C, D’Cotta H, Baroiller JF, and Kocher T.D. 2011. Genetic and Physical Mapping of Sex-Linked AFLP Markers in Nile Tilapia (*Oreochromis niloticus*). *Mar. Biotechnol.* 13: 557-562.
- Lee BY, Hulata G, and . Kocher TD. 2004. Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity* 92: 543–549.
- Lee BY, Kocher TD. 2007. Exclusion of Wilms tumor (WT1) and ovarian cytochrome P450 aromatase (CYP19A1) as candidates for sex determination genes in Nile tilapia (*Oreochromis niloticus*). *Anim. Genet.* 38: 85–86.
- Lee BY, Penman DJ, and Kocher TD. 2003. Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Anim. Genet.* 34: 379–383.
- Mair GC, Scott AG, Penman DJ, Beardmore JA, and Skibinski DOE. 1991. Sex determination in the genus *Oreochromis* 1. Sex reversal, gynogenesis and triploidy in *O. niloticus* (L.) *Theor. Appl. Genet.* 82:144-152.
- Martins C, Oliveira C, Wasko AP, Wright JM. 2004. Physical mapping of the Nile tilapia (*Oreochromis niloticus*) genome by fluorescent in situ hybridization of repetitive DNAs to metaphase chromosomes a review. *Aquaculture* 231, 37–49.
- Matsuda M, Nagahama Y, Shinomiya A, et al. (13 co-authors). 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417:559–563.
- Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, Fujiyama A, Naruse K, Hamaguchi S, Sakaizumi M. 2012. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics* 191:163–170.
- Nanda I, Kondo M, Hornung U, et al. (12 co-authors). 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc Natl Acad Sci U S A.* 99: 11778–11783.
- Ndiwa CT, Nyngi D, Agnès JF. 2014. An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Lobo drainage, Kenya. *PloS ONE* 9(9):e106972.
- Nyngi DW, and Agnès JF. 2012. *Phylogeography of the Nile Tilapia in Africa*. Lambert Academic Publishers.
- Nyngi D, De-Vos L, Aman R, Agnès JF. 2009. Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) (Cichlidae; Teleostei) in the Lobo Swamp (Kenya). *Aquaculture* 297: 57-63.
- Palaiokostas CI, Bekaert M, Khan MG, Taggart JB, Gharbi K, McAndrew BJ, Penman DJ. 2013. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD sequencing. *PLoS One.* 8(7):e68389.

- Poonlaphdecha S, Pepey E, Huang SH, Canonne M, Soler L, Mortaji S, Morand S, Pfennig F, Mélard C, Baroiller JF. 2011. Elevated amh gene expression in the brain of male tilapia (*Oreochromis niloticus*) during testis differentiation. *Sexual Development*, 5, 33-47.
- Poonlaphdecha S, Pepey E, Canonne M, de Verdal H, Baroiller et al., 2013. Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both dmrt1 and amh expressions. *Gen. Comp. Endocrinol.*
- Rousset F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8:103–106.
- Scott AG, Penman DJ, Beardmore JA, Skibinski DOF. 1989. The 'YY' supermale in *Oreochromis niloticus* L. and its potential in aquaculture. *Aquaculture* 78, 237–251.
- Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovskiy R, Zilberman N, Kocher TD, Hulata G, Ron M. 2006. AMH and DMRT2 genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics*, 174(3):1573–1581.
- Sun Y, Jiang D, Zeng S, Hu C, Ye K, Yang C, Yang S, Li M, and Wang D. 2014. Screening and characterization of sex-linked DNA markers and marker-assisted selection in the Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 433: 19-27.
- Trewavas E. 1983. Tilapiine Fishes of the Genera Sarotherodon, Oreochromis and Danakilia, British Museum Natural History, London, UK.
- Valenzuela N, Adams DC, and Janzen FJ. 2003. Pattern does not equal process: Exactly when is sex environmentally determined? *Am. Nat.* 161: 676–683.
- Wessels S, Hoerstgen-Schwark G. 2011. Temperature dependent sex ratios in selected lines and crosses with a YY-male in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 318: 79–84.
- Wessels S, Sharifi RA, Luehmann LM, Rueangsri S, Krause I, Pach S, Hoerstgen-Schwark G, Knorr C. 2014. Allelic variant in the anti-müllerian hormone gene leads to autosomal and temperature-dependent sex reversal in a selected Nile tilapia line. *PLoS One* 9(8):e104795.
- Yamamoto Y, Zhang Y, Sarida M, Hattori RS, Strüssmann CA. 2014. Coexistence of genotypic and temperature-dependent sex determination in pejerrey *Odontesthes bonariensis*. *PLoS One*. 9(7):e102574.
- Yan JP, Hinder M, Einsele G. 2002. Geochemical evolution of the closed basin lakes: general model and application to Lakes Qinghai and Turkana. *Sedimentary Geology* 143: 105-122.
- Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, Cabau C, Bouchez O, Fostier A, Guiguen Y. 2012. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Curr Biol.* 22:1423–1428.

Table 1: Information on the populations and strains sampled of *Oreochromis niloticus*

Population or Strain	Characteristics	Subspecies	Ichthyo. Regions	Basin	Localisation	Country	N° of females	N° of males	Water temperature
Kpandu	Wild	<i>O. n. niloticus</i>	Sudano-Sahelian	Volta	Lake Volta	Ghana	11	10	27 – 32 °C
Bouake	Domesticated Strain	<i>O. n. niloticus</i>	Sudano-Sahelian + Nilotic	Volta + Nile	Bouake Station	Ivory Coast	6	7	27 – 32 °C
Manzala	Domesticated Strain	<i>O. n. niloticus</i>	Nilotic	Nile	Lake Manzala	Egypt	11	17	20.5 - 25.7
Lake Turkana	Wild	<i>O. n. vulcani</i>	Nilotic	Rift-Valley	Lake Turkana	Kenya	10	7	27 – 27 °C
Lake Baringo	Wild	<i>O. n. baringoensis</i>	Nilotic	Rift-Valley	Lake Baringo	Kenya	10	10	25 – 27 °C
Bogoria	Wild	<i>O. n. loboï</i>	Nilotic	Rift-Valley	Lake Bogoria Hotel Spring	Kenya	10	10	36 °C
Chelaba Spring	Wild	<i>O. n. loboï</i>	Nilotic	Rift-Valley	Chelaba Spring	Kenya	10	10	36 °C
Turtle Spring	Wild	<i>O. n. loboï</i>	Nilotic	Rift-Valley	Turtle Spring	Kenya	10	10	36 °C
Lake Hora	Wild	<i>O. n. cancellatus</i>	Ethiopian	Awash	Lake Hora	Ethiopia	5	6	22 - 24.5 °C
Lake Koka	Wild	<i>O. n. cancellatus</i>	Ethiopian	Awash	Lake Koka	Ethiopia	4	4	17 – 26 °C
Metahara	Wild	<i>O. n. filoa</i>	Ethiopian	Awash	Lake Metahara	Ethiopia	4	6	32 – 39 °C

Table 2: Primers used for SNP search for each candidate gene or marker

Linkage Group	Scaffold	Gene/ Marker	Region	Primers Used	Primer Sequence (5'-3')	Tm	Size	GenBank	Alignment		
LG1	scaffold 17	<i>wt1b</i>	Exon 5-Exon 7	Wt1b-F938 Wt1b-R1836	AACCTGTACCAAATGGCATC GATGGGACAGTTTGAAGTAT	57°C	898 bp	EF026080	LOC100701078 ENSONIG00000005545		
		<i>cyp19a1a</i>	Promoter	cyp19a1a-F1872 cyp19a1a-R2371	AAATCTGGGAATCAATGTAATGAAGA GCAGAGATCAGATCCATGAGAA	56°C	499 bp	AF472620	ENSONIG00000000155 Scaffold GL831420.1		
LG23	scaffold 101	<i>amh</i>	Exon 1-Intron 3	Amh-F85 Amh-R1308	CCCGTTTGCTGTCTGATTTT GGATGGCAACCTCGTTTTAC	56°C	1223 bp	EF512167	Scaffold GL831234.1 Mod. EF512167		
			Exon 3-Exon 6	Amh-F885 Amh-R1958	ACAAACCCTGTGCTGCTCTT GTGGGAGCTGTGGAGTTGAT	58°C	1073 bp		ENSONIT0000000602		
			Exon 6-Intron 6	Amh-F1802 Amh-R2883	CTGCAAACCTCCTCCTTTG CGACAAATTCAGCTAAGCA	57°C	1081 bp				
			Intron 6-Exon 7	Amh-F2340 Amh-R3328	TGGCCACGGATTCTCTAAAC GCATCTGGCTTGATGGAGAT	60°C	988 bp				
		<i>amhy</i>	Intron 6-Exon 7	Amh-F2443 Amh-R3490	CGGTCCCAGTGACCTATGAG AAGTACACGTGGTGTATTGTAATTGA	58°C	1047 bp/ 814 bp	HG518784			
		X-specific	scaffold 101			SCAR-5F-X SCAR-5R	CTGGTTTGCAATAGTTAGGGTGCT CAGAAATGTAGACGCCCAGGTATC	60°C	925 bp	KC710223	
						Y-specific	SCAR-5F SCAR-5R-Y	TAAATTAATGACATTTAGTTATG TTACAGCAGCACCCAGAGTCAT	60°C	293 bp	KC710224

Table 3: SNPs identified in the promoter region studied of the aromatase gonadal *cyp19a1a* gene

SNP Name	Gene	Position	Feature	Consequence	SNP type	Frequency	N°Ind.
Cyp_2016	<i>cyp19a1a</i>	2016	Promoter	none	C	70.65	184
					T	28.26	184
					CT	1.08	184
Cyp_2124	<i>cyp19a1a</i>	2124	Promoter	none	G	38.50	187
					A	61.50	187
Cyp_2175	<i>cyp19a1a</i>	2175	Promoter	none	A	77.42	186
					G	4.30	186
					A/G	18.28	186
Cyp_2308	<i>cyp19a1a</i>	2308	Promoter	none	C	28.88	187
					T	68.98	187
					CT	2.14	187

Table 4: Nucleotide deletions found in the *cyp19a1a* promoter region studied vs ref. Genbank AF472620 & ENSONIG00000000155

SNP Name	Gene	Position	Feature	Deletion	Nt change	Population/Strain	Frequency	N°Ind.
Cyp_1988	<i>cyp19a1a</i>	1988	promoter	G		Metahara	100	10/10
						Hora	100	11/11
						Koka	100	8/8
Cyp_1993	<i>cyp19a1a</i>	1992	promoter	A	A to T	Bogoria	25	5/20
						Chelaba	15	6/20
						Chelaba	30	3/20
Cyp_2036-2037	<i>cyp19a1a</i>	2036-2037	promoter	GT		ALL POP.	100	188

Table 5: Pairwise linkage disequilibrium detected for the region of the *cyp19a1a* gene promoter

Allele pair	Population/strain	Linkage disequilibrium (%)	R ²
Cyp_2016-Cyp_2124	Bogoria	22	0.22
Cyp_2016-Cyp_2175	Baringo	40	0.4
	Turtle	12	0.12
Cyp_2016-Cyp_2308	Baringo	2	0.02
Cyp_2016-Cyp_1993	Bogoria	50	0.5
	Chelaba	46	0.46
Cyp_2124-Cyp_2175	Bogoria	14	0.14
	Chelaba	2	0.02
Cyp_2124-Cyp_2308	Bogoria	10	0.1
	Chelaba	6	0.06
Cyp_2124-Cyp_1993	Bogoria	8	0.08
	Chelaba	46	0.46
Cyp_2175-Cyp_2308	Baringo	16	0.16
	Turtle	10	0.1
Cyp_2175-Cyp_1993	Bogoria	32	0.32
	Chelaba	64	0.64
Cyp_2308-Cyp_1993	Bogoria	28	0.28
	Chelaba	56	0.56

Table 6: SNPs identified in the *wt1b* gene

SNP Name	Gene	Position ref.*	Position gene**	Feature	Consequence	Codon	Amino acid	SNP type	Frequency	N°Ind.
wt1b_398	<i>wt1b</i>	398	1277	Intron 6	none			A	53.55	183
								G	28.41	
								AG	18.04	
wt1b_464	<i>wt1b</i>	464	1343	Intron 6	none			A	79.46	185
								G	18.92	
								AG	1.62	
wt1b_466	<i>wt1b</i>	466	1345	Intron 6	none			C	89.19	185
								CT	10.81	
wt1b_649	<i>wt1b</i>	649	1528	Exon5/Exon 6	non-synon.	AGC/AGT	S	C	87.10	186
								T	5.91	
								CT	5.91	
								G	1.08	
wt1b_684	<i>wt1b</i>	684	1563	Exon 6				NON		

* Position with respect to EF026080

** Position with respect to LOC100701078

Table 7: Pairwise linkage disequilibrium detected for the *wt1b* gene gene promoter

Allele pair	Population/strain	Linkage disequilibrium (%)	R ²
Wt1b_398-Wt1b_652	Baringo	8	0.8
Wt1b_398-Wt1b_464	Chelaba	6	0.6
	Baringo	100	1
	Turtle	100	1
	Bouake	100	1
Wt1b_398-Wt1b_466	Bouake	10	0.1
Wt1b_398-Wt1b_649	Baringo	8	0.8
Wt1b_652-Wt1b_464	Baringo	100	1
Wt1b_652-Wt1b_466	Turkana	6	0.6
Wt1b_464-Wt1b_466	Bouake	100	1
	Kpandu	10	0.1
Wt1b_464-Wt1b_649	Baringo	100	1

Table 8: SNPs identified in the *amh* gene

SNP Name	Gene	Position	Feature	Consequence	Codon	Amino acid	SNP type	Frequency	N°Ind.
amh_202	amh	202	Exon 1	upstream			A	74.85	163
							G	15.34	
							AG	9.81	
amh_264	amh	264	Exon 1	upstream			A	87.58	177
							G	6.21	
amh_280	amh	280	Exon 1	upstream			C	87.36	174
							T	4.6	
							CT	8.04	
amh_597	amh	597	Exon 2	synonymous	GAG/GAT	E/D	G	74.29	175
							T	18.86	
							GT	6.86	
amh_611	amh	611	Exon 2	synonymous	AAC/AGC/ ACC/ATC	N/S/ T/I	A	18.97	175
							G	81.03	
							C	51.43	
							T	32.57	
							CT	16.00	
amh_754	amh	754	Intron 2				C	51.43	175
							T	32.57	
							CT	16.00	
amh_875	amh	875	Exon 3	non-synon.	CCA/CCT	P	A	66.67	141
							T	33.33	
amh_916	amh	916	Exon 3	synonymous	AAT/AGT	N/S	A	66.67	132
							G	33.33	
amh_1996	amh	1996	Exon 6	non-synon.	AGC/AGT	S	C	81.32	182
							T	18.68	
amh_2005	amh	2005	Exon 6	non-synon.	TCC/TCT	S	C	86.74	181
							T	6.08	
							CT	7.18	
amh_2127	amh	2127	Exon 6	synonymous	CTG/CCG	L/P	C	74.43	176
							T	6.25	
							CT	19.32	
amh_2128	amh	2128	Exon 6	non-synon.	CCG/CCA	P	A	22.29	175
							G	72.00	
							AG	1.71	
							T	4.00	
amh_2245	amh	2245	Intron 6				A/T	68.13	91
							T	31.87	
amh_2282	amh	2282	Intron 6				G	60.92	174
							T	39.08	
amh_2486	amh	2487	Intron 6				C	78.12	192
							T	12.50	
							CT	9.38	
amh_2509	amh	2509	Intron 6				A	68.62	188
							C	25.53	
							AC	2.85	
amh_2586	amh	2586	Intron 6				C	75.42	179
							G	23.46	
							CG	1.12	
amh_2618	amh	2618	Intron 6				A	14.29	182
							G	81.87	
							AG	3.84	
amh_2643	amh	2643	Intron 6				A	8.99	178
							G	76.40	
							AG	14.61	
amh_2655	amh	2655	Intron 6				A	16.11	180
							T	80.00	
							AT	3.89	

Table 9: Nucleotide Insertions vs ref. Scaffold GL831234.1 and Mod. EF512167

SNP Name	Gene	Position	Feature	Deletion	Insertion	Population/Strain	Frequency	N°Ind.			
amh_2860	amh	2860	Intron 6		G	Bouake	100.0	13/13			
						Kpandu	100.0	21/21			
						Manzala	100.0	28/28			
						Hora	100.0	10/10			
						Koka	100.0	11/11			
amh_2867	amh	2867	Intron 6		A	ALL POP.		8/8			
amh_2871	amh	2871	Intron 6		C	Bouake	100.0	13/13			
						Kpandu	100.0	21/21			
						Manzala	100.0	28/28			
						Metahara	100.0	10/10			
						Hora	100.0	11/11			
						Koka	100.0	8/8			
amh_2898	amh	2898	Intron 6		T	Bouake	100.0	13/13			
						Kpandu	100.0	21/21			
						Manzala	100.0	28/28			
						Metahara	100.0	10/10			
						Hora	100.0	11/11			
						Koka	100.0	8/8			
amh_2909	amh	2909	Intron 6		C	Bouake	100.0	13/13			
						Kpandu	100.0	21/21			
						Manzala	100.0	28/28			
						Metahara	100.0	10/10			
						Hora	100.0	11/11			
						Koka	100.0	8/8			
						Turkana	100.0	17/17			
						Bogoria	60.0	12/20			
						Chelaba	45.0	9/20			
						Turtle	100.0	15/15			
						Baringo	95.0	19/20			
amh_2934	amh	2934	Intron 6		T	ALL POP.	100.0	192			
						2677 to 2694	Intron 6	17 nt	Bogoria	40.0	8/20
									Chelaba	55.0	11/20
									Turtle	25.0	5/20
									Baringo	5.0	1/20
						2895 to 2914	Intron 6	18 nt	Bogoria	40.0	8/20
									Chelaba	55.0	11/20
									Turtle	25.0	5/20
									Baringo	5.0	1/20

Table 10: Chi-square test results for the deviation of sex-linked SNP in females and males from the HWE.

SNP Name	SNP Type	Phenotype	Item	Bogoria Spring	Chelaba Spring	Turtle Spring	Lake Baringo	Lake Turkana
Amh_597	T	Females	χ^2	8.2974	7.4370	10.796	6.8075	3.5451
			P	0.0028	0.0316	0.0001	0.0001	0.1048
	G	Males	χ^2	6.5778	4.3571	9.1000	7.6095	-
			P	0.0546	0.1035	0.0048	0.0001	-
Amh_611	G	Females	χ^2	11.962	9.6103	-	10.936	13.888
			P	0.0008	0.0001	-	0.0001	0.0001
	A	Males	χ^2	9.5093	5.9932	-	2.6037	-
			P	0.0001	0.0034	-	0.5941	-
Amh_754	C	Female	χ^2	2.1173	9.4780	2.5568	4.2743	-
			P	0.2985	0.0018	0.1734	0.1180	-
	T/C	Males	χ^2	3.3812	-	-	6.5714	-
			P	0.1645	-	-	0.0001	-
Amh_2586	C	Female	χ^2	3.2145	7.4370	6.1276	6.5470	7.2193
			P	0.1618	0.0301	0.0214	0.0508	0.0001
	G	Males	χ^2	4.6803	5.9943	8.7575	5.3466	8.6480
			P	0.0703	0.0701	0.0040	0.0800	0.0001
Amh_2618	A	Female	χ^2	3.2399	6.1864	8.6464	6.9242	-
			P	0.2683	0.0636	0.0104	0.0001	-
	G	Males	χ^2	3.1738	-	8.6464	-	-
			P	0.1773	-	0.0096	-	-
Amh_2643	G	Female	χ^2	-	2.3331	2.5076	3.0096	16.009
			P	-	0.3209	0.2716	0.0682	0.0001
	A	Males	χ^2	-	-	5.1361	4.9882	-
			P	-	-	0.0084	0.0001	-
Amh_2655	A	Female	χ^2	3.1644	7.3418	5.5800	6.9327	-
			P	0.1700	0.0234	0.0760	0.0475	-
	T	Males	χ^2	2.6479	4.9155	5.5800	-	-
			P	0.2591	0.0246	0.0740	-	-
Amh_2668	A	Female	χ^2	-	2.7522	7.5030	1.9226	6.0103
			P	-	0.2309	0.0290	0.4274	0.0495
	C	Males	χ^2	1.7287	-	3.9115	3.0720	-
			P	0.5569	-	0.1639	0.1557	-
Amh_2807	T	Female	χ^2	7.8671	6.7877	6.3502	1.1015	-
			P	0.0001	0.0199	0.0350	0.6177	-
	C	Males	χ^2	5.7147	6.2877	10.115	5.9863	-
			P	0.0680	0.0103	0.0001	0.0836	-
Amh_2486	C	Female	χ^2	-	-	-	7.9071	-
			P	-	-	-	0.0001	-
	T	Males	χ^2	5.7193	4.8331	2.7264	2.6300	-
			P	0.1060	0.1142	0.3524	0.6187	-
Amh_2509	A	Female	χ^2	8.8034	7.7724	5.5015	5.1160	-
			P	0.0001	0.0029	0.0232	0.0860	-
	C	Males	χ^2	5.9466	5.5909	6.3127	-	5.4925
			P	0.0275	0.0294	0.0182	-	0.0139

X²= chi-square test value, BOLD indicates population showing significant differences in both males and females.

P= significance difference at 0.05

P values could not be calculated in all the populations due to their sums being zero.

Figures

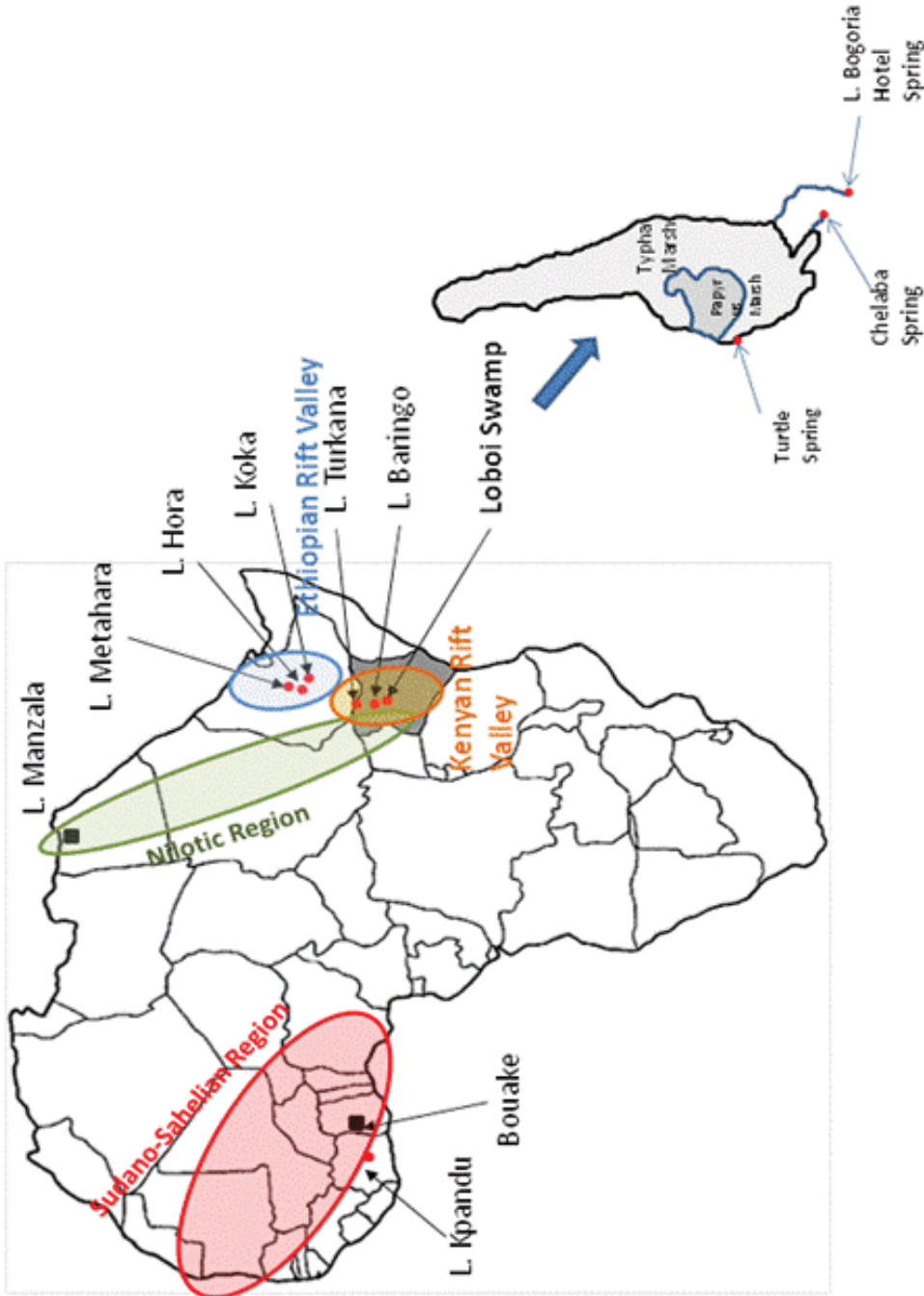


Figure 1: Map showing the locations of studied areas within the Kenyan Rift-Valley, Sudano-Sahelian and Ethiopian regions

AMH_2992

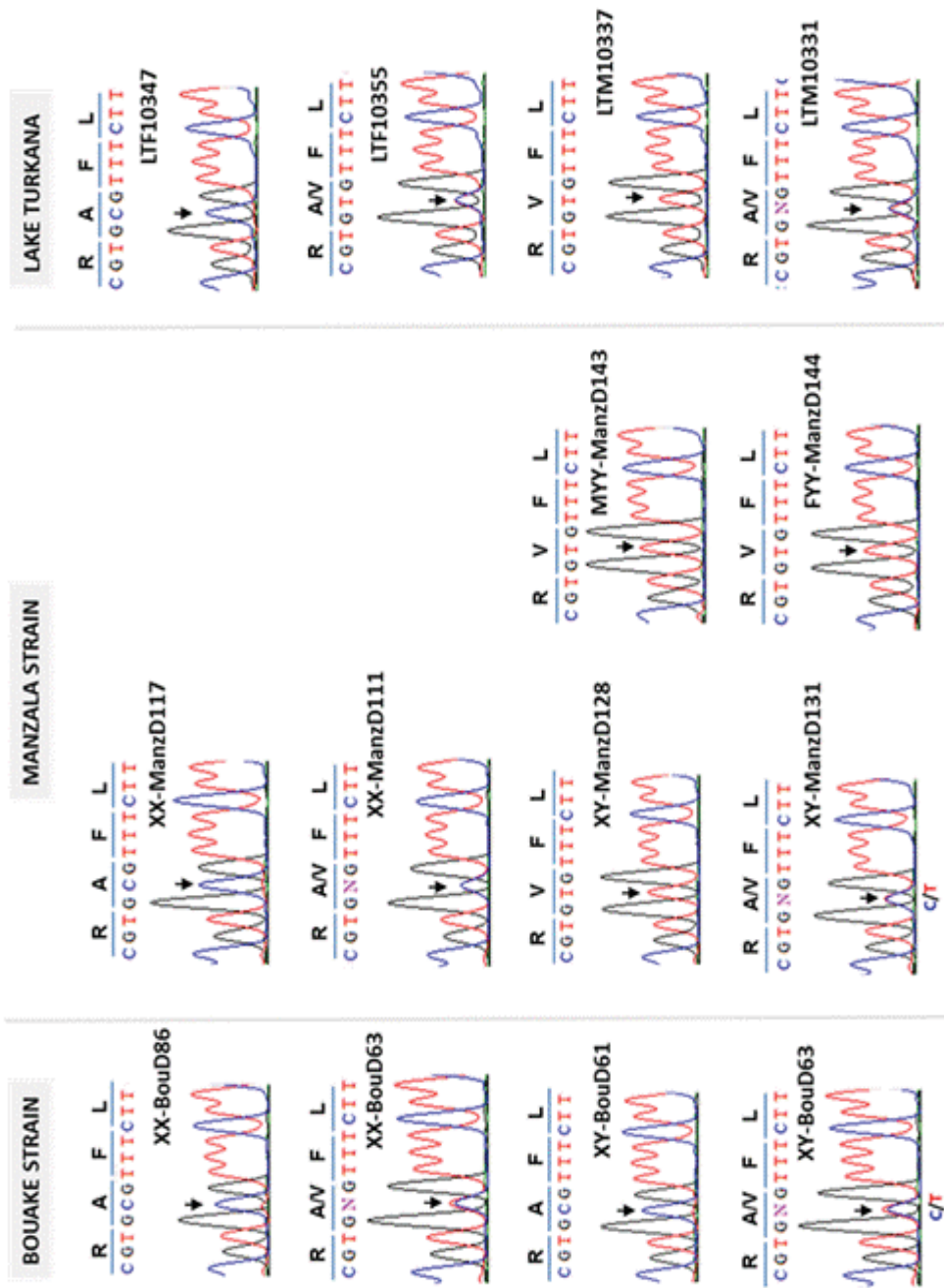


Figure 2: Chromatograms obtained from the domesticated strains and one of the wild population showing that our polymorphism does not coincide with the study of Eshel *et al.* (2014) in which XX females had a T, XY males a C and YY males both

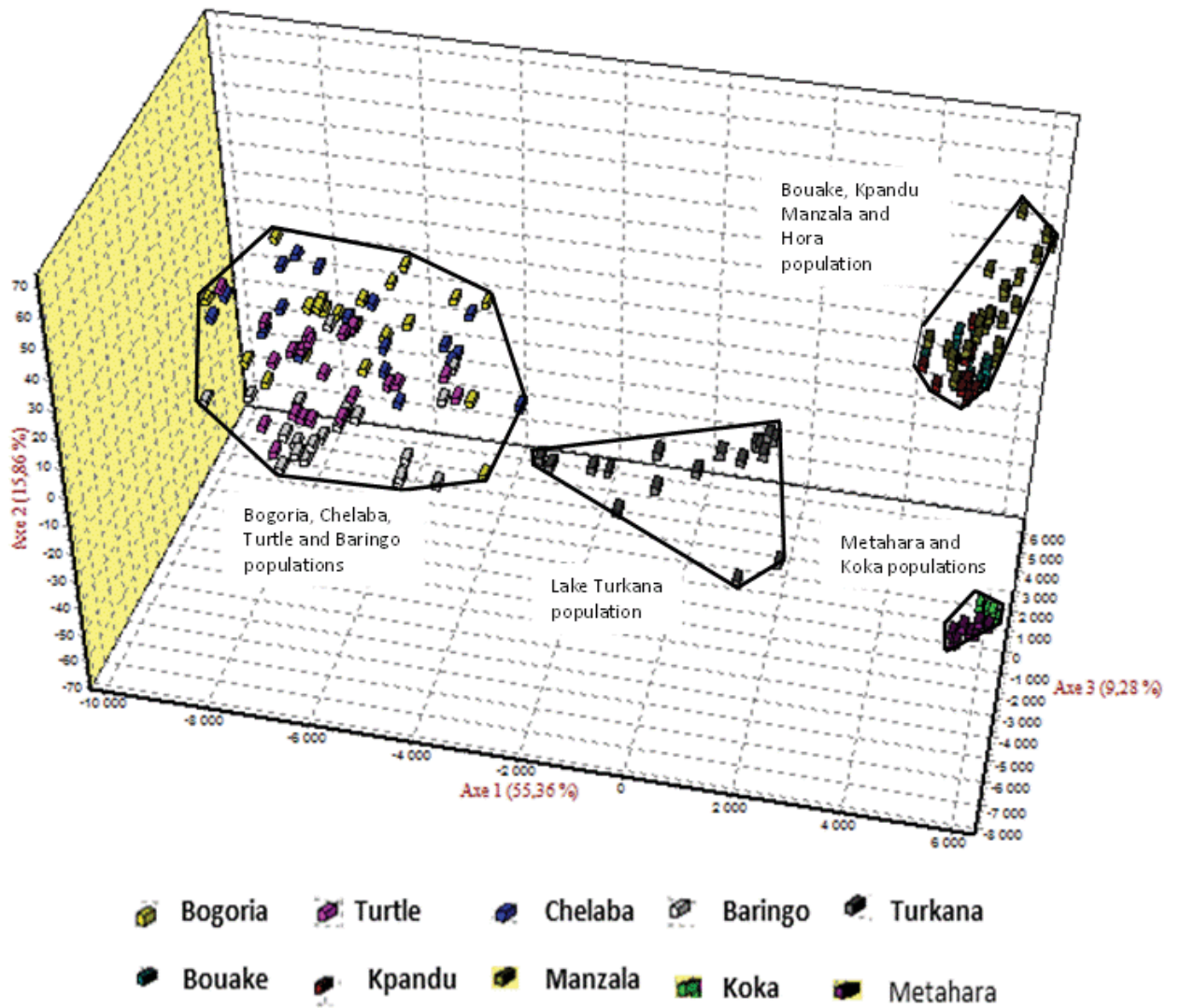


Figure 3: FCA analysis of all the SNPs identified from the *amh*, *cyp19a1a* and the *wt1b* genes



Figure 4: Analysis of the domesticated strain and different wild populations with *amh* makers to discriminate XY and YY individuals with a band of 800 bp. In red are sex reversed individuals whose sexual phenotypes did not coincide with the genotypes .

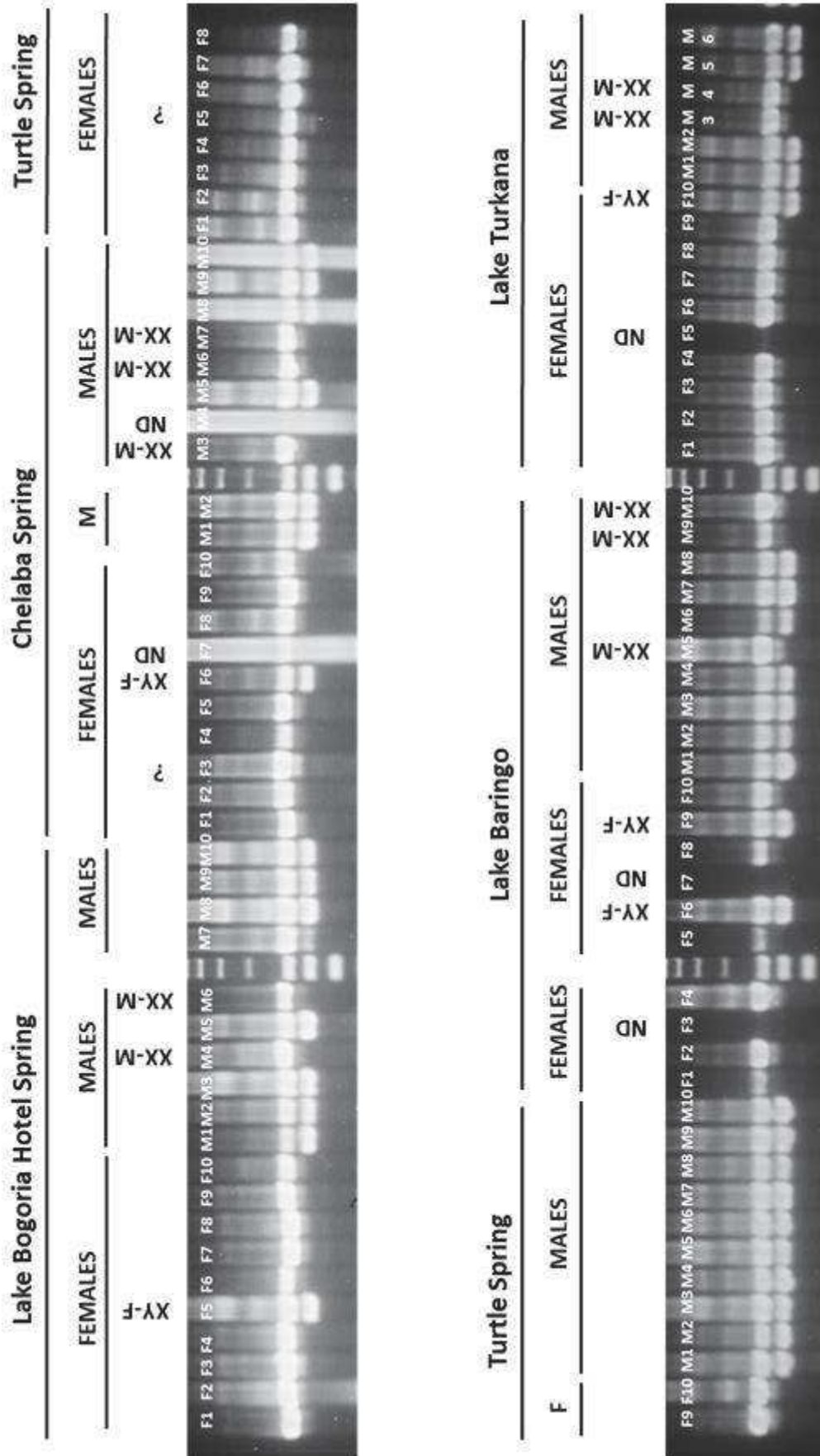
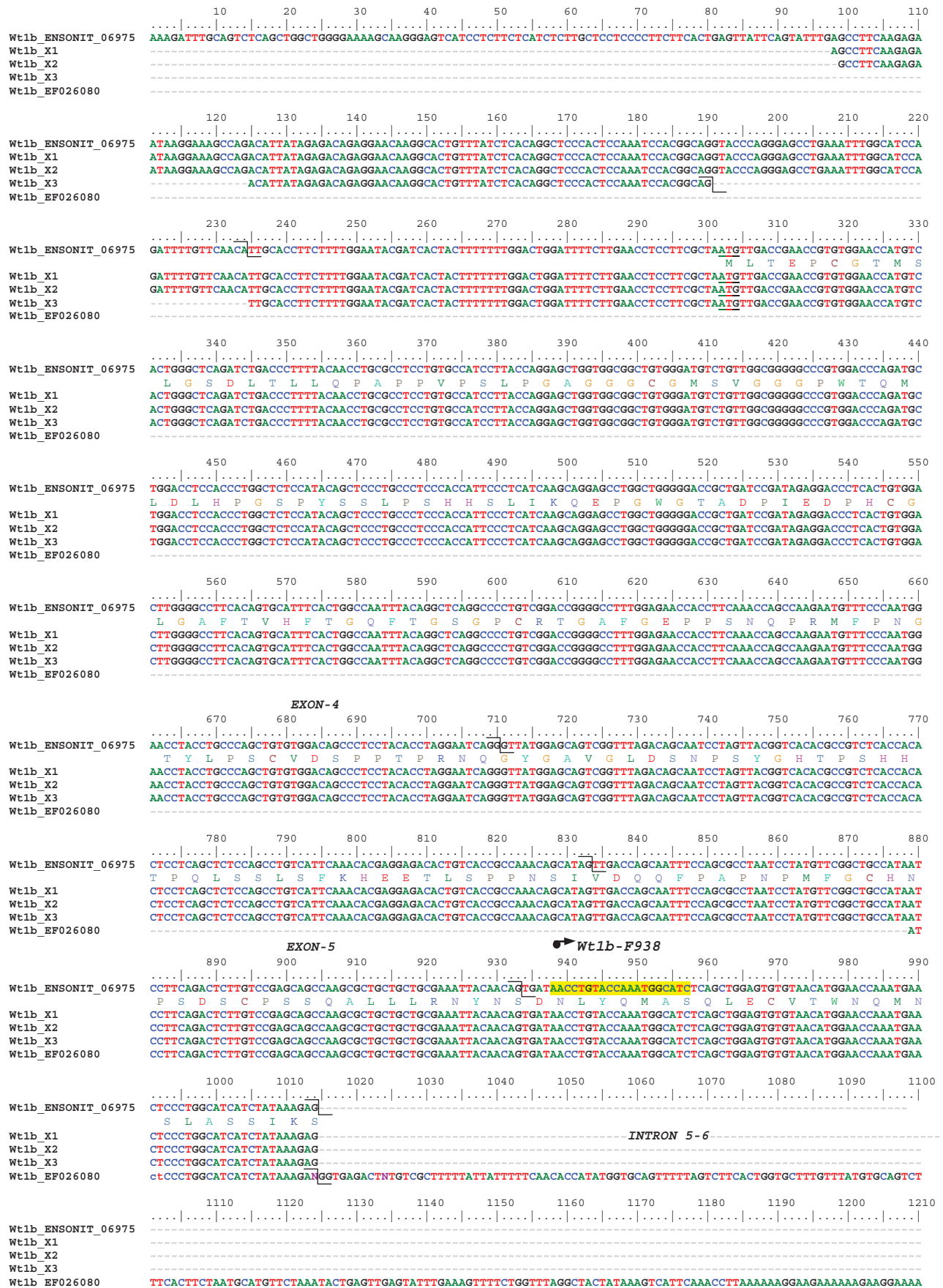


Figure 5: Analysis of the Kenyan wild populations with *amh* markers to discriminate XY and YY individuals with a band of 800 bp. Genotype analysis revealed between the sexual phenotype and genotype suggesting sexual reversal of individuals shown on the gel.

Supporting figure



1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 CAGAAGGAAGGGCCATTGCTAAAGAATGTGAGACGAAGCTTCTCTGCTCAAAAAATTAATAAAAAAGTCTCAGATGAAGTTATGTAGGATCCTGTCACCTAAAT

1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 TACAAGCTGAGTGAGAAAATATATCTTCAGGGCCCTGGATATATTGTTAGAGGCAGGCATGTTTAAAGGGTTGTTTGAAGAGAAGAGGAAAACGGTTTTGAAAATCAAG

1440 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 TTGGAATCCCTCAGGAAACATGGCATTATGTCCTACCTACTCCACTGTTATAATTAACCTGTAATTTTGAAGAAATATTTGTTCTGCTACCAAGCGCCATCCAAC

EXON-6

1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 AGCTATGACAGTGACCCCATACCCCCACCTCCACCCATGCTTGTGAGCAGCCAGTACCACATACACACCCACGGCGTCTTCAGAGGACTTCAG
S Y D S D P I P P P P P M L V S T Q Y H I H T H G V F R G L Q
AGCTATGACAGTGACCCCATACCCCCACCTCCACCCATGCTTGTGAGCAGCCAGTACCACATACACACCCACGGCGTCTTCAGAGGACTTCAG
AGCTATGACAGTGACCCCATACCCCCACCTCCACCCATGCTTGTGAGCAGCCAGTACCACATACACACCCACGGCGTCTTCAGAGGACTTCAG
AGCTATGACAGTGACCCCATACCCCCACCTCCACCCATGCTTGTGAGCAGCCAGTACCACATACACACCCACGGCGTCTTCAGAGGACTTCAG

1660 1670 1680 1690 1700 1710 1720 1730 1740 1750 1760
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 GATGTCAGACGGGTATCTGGTATCACCTTCCCTGTTGCGAAGACT
D V R R V S G I T S P V A K T
INTRON 6-7
GATGTCAGACGGGTATCTGGTATCACCTTCCCTGTTGCGAAGACT
GATGTCAGACGGGTATCTGGTATCACCTTCCCTGTTGCGAAGACT
GATGTCAGACGGGTATCTGGTATCACCTTCCCTGTTGCGAAGACT
ACACATGTAACAAGACAGAAAGTGAATCTTCCAGTGTAAATCCTTCTCCTTCTTCTCCTAAAGCATGTCAGACGGGTATCTGGTATCACCTTCCCTGTTGCGAAGACT

EXON-6-7? **Wt1b-R1836** ←

1770 1780 1790 1800 1810 1820 1830 1840 1850 1860 1870
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 TCAGAGCCAGTGAAGAAGACCCGTTTGTGCGCTTATCCCGGCTGCAGCAAGAGTACTTCAAACCTGCCATCTGCAGATGCATGGCCGCAAAACACACAGAGAGAA
S E A S E K R P F V C A Y P G C S K R Y F K L S H L Q M H G R K H T G E K
TCAGAGCCAGTGAAGAAGACCCGTTTGTGCGCTTATCCCGGCTGCAGCAAGAGTACTTCAAACCTGCCATCTGCAGATGCATGGCCGCAAAACACACAGAGAGAA
TCAGAGCCAGTGAAGAAGACCCGTTTGTGCGCTTATCCCGGCTGCAGCAAGAGTACTTCAAACCTGCCATCTGCAGATGCATGGCCGCAAAACACACAGAGAGAA
TCAGAGCCAGTGAAGAAGACCCGTTTGTGCGCTTATCCCGGCTGCAGCAAGAGTACTTCAAACCTGCCATCTGCAGATGCATGGCCGCAAAACACACAGAGAGAA

1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 ACCATACCAAGTGTGATTTACAGACTGCGGCCGAGATTCTCTCGCTCGGACCAAGTAAAGAGACACCAGCGCAGACACAGAGAGTGAAGCCCTTTCAGTGTGAGACGT
P Y Q C D F T D C G R R S F S R S D Q L K R H Q R R H T G V K P P F Q C E T
ACCATACCAAGTGTGATTTACAGACTGCGGCCGAGATTCTCTCGCTCGGACCAAGTAAAGAGACACCAGCGCAGACACAGAGAGTGAAGCCCTTTCAGTGTGAGACGT
ACCATACCAAGTGTGATTTACAGACTGCGGCCGAGATTCTCTCGCTCGGACCAAGTAAAGAGACACCAGCGCAGACACAGAGAGTGAAGCCCTTTCAGTGTGAGACGT
ACCATACCAAGTGTGATTTACAGACTGCGGCCGAGATTCTCTCGCTCGGACCAAGTAAAGAGACACCAGCGCAGACACAGAGAGTGAAGCCCTTTCAGTGTGAGACGT
ACCATACCAAGTGTGATTTACAGACTGCGGCCGAGATTCTCTCGCTCGGACCAAGTAAAGAGACACCAGCGCAGACACAGAGAGTGAAGCCCTTTCAGTGTGAGACGT

1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090
Wt1b_ENSONIT_06975
Wt1b_X1
Wt1b_X2
Wt1b_X3
Wt1b_EF026080 GTCAGAGAAAGTTTTCACGGTCAGATCATCTTAAGACACACACTCGGACTCATACAGGTAAACAAGTGAAGAAGCCCTTTCATGCCGCTGGTCCAACCTGTCAGAGAAAG
C Q R K F S R S D H L K T H T R T H T G K T S E K P F T C R W S N C Q K K
GTCAGAGAAAGTTTTCACGGTCAGATCATCTTAAGACACACACTCGGACTCATACAGGTAAACAAGTGAAGAAGCCCTTTCATGCCGCTGGTCCAACCTGTCAGAGAAAG
GTCAGAGAAAGTTTTCACGGTCAGATCATCTTAAGACACACACTCGGACTCATACAGGTAAACAAGTGAAGAAGCCCTTTCATGCCGCTGGTCCAACCTGTCAGAGAAAG
GTCAGAGAAAGTTTTCACGGTCAGATCATCTTAAGACACACACTCGGACTCATACAGGTAAACAAGTGAAGAAGCCCTTTCATGCCGCTGGTCCAACCTGTCAGAGAAAG

2100 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
Wt1b_ENSONIT_06975
Wt1b_X1/X2/X3
TTTCCCGCTCTGACAGCTGGTGGCCACCCACAGCATGCACCAGAGAACTGACCAGCTGCAGCTGCCATCTGAGCAGCAAGAGGAGGAGGAGGAGGACAAAGAGG
F A R S D E L V R H H S M H Q R N L T K L Q P A I *
TTTCCCGCTCTGACAGCTGGTGGCCACCCACAGCATGCACCAGAGAACTGACCAGCTGCAGCTGCCATCTGAGCAGCAAGAGGAGGAGGAGGAGGACAAAGAGG

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300 2310
Wt1b_ENSONIT_06975
Wt1b_X1/X2/X3
AATGTGACAGTATGTTGTCAGCTAGTAGAAAAGACGCTGGATCCTGGTCAAGCTTATGGTCTTTTCCAGACTTCCGAAAGTGGCAGCGCAGCTGATGCCTGCG
AATGTGACAGTATGTTGTCAGCTAGTAGAAAAGACGCTGGATCCTGGTCAAGCTTATGGTCTTTTCCAGACTTCCGAAAGTGGCAGCGCAGCTGATGCCTGCG
TCAACGTCAGGACTCACAGATGATCCCGAAGGCTCTCAAGATGCACCTTCAACTACTACTAGCAATAAAAAGTGCATTTTGAAGGGTATAATGCCACAGTAACGCTGAGC
TCAACGTCAGGACTCACAGATGATCCCGAAGGCTCTCAAGATGCACCTTCAACTACTACTAGCAATAAAAAGTGCATTTTGAAGGGTATAATGCCACAGTAACGCTGAGC

2430 2440 2450 2460 2470 2480 2490 2500 2510 2520 2530
Wt1b_ENSONIT_06975
Wt1b_X1/X2/X3
CAAACCTTGTGAGATGAGAAAATGTTGACATACATTTTGAATATTTTACGACACATGCACCTGGCTTCTCTTTATTTTTAAAATGGTCTCACCTCTTGAATATGTTG
CAAACCTTGTGAGATGAGAAAATGTTGACATACATTTTGAATATTTTACGACACATGCACCTGGCTTCTCTTTATTTTTAAAATGGTCTCACCTCTTGAATATGTTG

2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640
Wt1b_ENSONIT_06975
Wt1b_X1/X2/X3
GACATTTGAACTGGAATTTAGAGCAATAAAGAACACTCCCTCAGTGTGTTGCTGCTTAAAGCCGAAACACAGAGAACTTCCGTTGTGGGAATGTCGGTATCTTTATCTTCCTA
GACATTTGAACTGGAATTTAGAGCAATAAAGAACACTCCCTCAGTGTGTTGCTGCTTAAAGCCGAAACACAGAGAACTTCCGTTGTGGGAATGTCGGTATCTTTATCTTCCTA

2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750
Wt1b_ENSONIT_06975
AAGTGTACATTTAGACTGCCAGATGCTTTGTATATCTTCATGTTGTTGCTTACCAATATATGGTCCCTCTATATCTTGTGCAATTTATAAAATCTTGAGTCA

Wt1b_X1/X2/X3 AAGTGTACATTAGACTGCCAGATTGCTTTGTATATACCTTCATGTTGGTGCATTTGCTTACAATATATGGTGCCTTCTATATTCTTGTGCATTTATAAAATCTTGAGTCAC

2760 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860

Wt1b_ENSONIT_06975 ATCAGTCCGCGTTGTGTTGGCGAGAGCAAAGGAGGTGTTGGTGTGTCAGGTTAAGAGGCCACCTATGAGACCTGGTGGGGGTGTGAGGTGGGTGACAGACAGTAAAA

Wt1b_X1/X2/X3 ATCAGTCCGCGTTGTGTTGGCGAGAGCAAAGGAGGTGTTGGTGTGTCAGGTTAAGAGGCCACCTATGAGACCTGGTGGGGGTGTGAGGTGGGTGACAGACAGTAAAA

2870 2880 2890 2900 2910 2920 2930 2940 2950 2960 2970

Wt1b_ENSONIT_06975 CAGCACCCCTCTGTACTGTGACAGGGTGCAGATCTCACCCCTCCTAAATCTCAAGCACACATCTTCACTAACTAGGAACATCTGGAGCCCAATAAATCTCATTAGAAAAA

Wt1b_X1/X2/X3 CAGCACCCCTCTGTACTGTGACAGGGTGCAGATCTCACCCCTCCTAAATCTCAAGCACACATCTTCACTAACTAGGAACATCTGGAGCCCAATAAATCTCATTAGAAAAA

2980 2990 3000 3010 3020 3030 3040 3050 3060 3070 3080

Wt1b_ENSONIT_06975 CACATACCTCAGTGTGACAGGGTGTCTCTCCACTGACTGCCCCCTTACGCGAGGCTGGGCCCAAAATGCACACATGCACATGCTTGAGTCTTGTGTACATAATAGT

Wt1b_X1/X2/X3 CACATACCTCAGTGTGACAGGGTGTCTCTCCACTGACTGCCCCCTTACGCGAGGCTGGGCCCAAAATGCACACATGCACATGCTTGAGTCTTGTGTACATAATAGT

3090 3100 3110 3120 3130 3140 3150 3160 3170 3180 3190

Wt1b_ENSONIT_06975 CTGCAGTTTAAAGCAGCACTTTACACCTTTTATCCACATGATGGCAGTGTCACTCCTTCTCGTCTTGCATAATGGAACTTGTACTCCGAGTTCGACAGTGTGACTGAATTC

Wt1b_X1/X2/X3 CTGCAGTTTAAAGCAGCACTTTACACCTTTTATCCACATGATGGCAGTGTCACTCCTTCTCGTCTTGCATAATGGAACTTGTACTCCGAGTTCGACAGTGTGACTGAATTC

3200 3210 3220 3230 3240 3250 3260 3270 3280 3290 3300

Wt1b_ENSONIT_06975 ATATACCTGTTTATATTTTACCCTCTCAGCCTGAAGATGTCCTCAATTTATCTTGCCTGTCTGTCACGTTGAGTATCAGTACAGACTATATTTGGTCCATGAAATATACATG

Wt1b_X1/X2/X3 ATATACCTGTTTATATTTTACCCTCTCAGCCTGAAGATGTCCTCAATTTATCTTGCCTGTCTGTCACGTTGAGTATCAGTACAGACTATATTTGGTCCATGAAATATACATG

3310 3320 3330 3340 3350 3360 3370 3380 3390 3400 3410

Wt1b_ENSONIT_06975 GTTTCAGTAGAGTGTGTTTTTTCAGCTTATGTCCTCAATCTAGAGGAAAGGCCGATGGGATGATTTCAGCTTCAGCGGGTGTGGAATAAAATATGACTCAGTGTGGGCAGA

Wt1b_X1/X2/X3 GTTTCAGTAGAGTGTGTTTTTTCAGCTTATGTCCTCAATCTAGAGGAAAGGCCGATGGGATGATTTCAGCTTCAGCGGGTGTGGAATAAAATATGACTCAGTGTGGGCAGA

3420 3430 3440 3450 3460 3470 3480 3490 3500 3510 3520

Wt1b_ENSONIT_06975 AAGTGTGTTGTCATGGCTCATAGTATGTTTTTAACTCTTATCAGTGTGCTCGCATATGGAAGCACAGTGTACGTTTAAAGAACATTTATCAACAGAAAGCCCAACAG

Wt1b_X1/X2/X3 AAGTGTGTTGTCATGGCTCATAGTATGTTTTTAACTCTTATCAGTGTGCTCGCATATGGAAGCACAGTGTACGTTTAAAGAACATTTATCAACAGAAAGCCCAACAG

3530 3540 3550 3560 3570 3580 3590 3600 3610 3620 3630

Wt1b_ENSONIT_06975 AGTACAGTATGTACCTGCTCTTTTACCAGCAGCCCTTTACACTTGACGAATTAACAAAGATTAATAATGCTTTCTCCTATTTACAGGCTCACAGCGGCTCATTGAAAT

Wt1b_X1/X2/X3 AGTACAGTATGTACCTGCTCTTTTACCAGCAGCCCTTTACACTTGACGAATTAACAAAGATTAATAATGCTTTCTCCTATTTACAGGCTCACAGCGGCTCATTGAAAT

3640 3650 3660 3670 3680 3690 3700 3710 3720 3730 3740

Wt1b_ENSONIT_06975 GCAGATTCAGGTTGTTTTCCAATCCCAATAAGCAGGGAGGTAGTATCTCTGTGATGTTTTTTCTGCTTCTTAATCCAGTGCACCTTGATTAAAGAGGGCTCGTCTTT

3750 3760 3770 3780 3790 3800 3810 3820 3830 3840 3850

Wt1b_ENSONIT_06975 GTTCTCTCTCAGCGGGGTGTTGGAATTGCTGGAAGATCAGAGGACAGAGTGAAGACAGAGTGAACATTTTACTCCTTCAGTTCAGTCTTACTGGATATTGGCA

3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960

Wt1b_ENSONIT_06975 AGGCTTTGCGGAGCATGAGGTTCTGATGCTTAAATGTCCTCAGTCTTTACTGCCAATGACTATAACACTCACTCAGATCTGTGCCCTATACACTGATCTTCTGCAATTC

3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070

Wt1b_ENSONIT_06975 TTTGCGGGCATGTGATTTACCCCAATAACACCAATAAATGTATGCACCTCAGTCCCATCTGCCAATGTTACCCCGACTCTGAATGCATGAACACACACATAGGCGTACAC

4080 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180

Wt1b_ENSONIT_06975 ACACCTCGCTCGCACACACCGCCACAGATGCCCCCTCAGCTGACTGCTGGCAGTTTGCCTTAAGGGGTGTGCGCATGGGTAATCTCAGGTGACGTACACCTGTTAAGGCAC

Figure S1: *Wt1b* gene sequence

10 20 30 40 50 60 70 80 90 100 110
On_Cyp19a1gene
GGACTTCNATCCACCCCTCCCACTTCTTCCCTCCAGAAATTTAAATTTATAGGACTTACTACAGAAATGAAAGTGAGGGATTACTTGTATTGCAGTAAGTCCCTCACA

120 130 140 150 160 170 180 190 200 210 220
On_Cyp19a1gene
TAGTGTCAACAGAAAAGATAAAGTGGAGTTCCGCGTGTAAATGTCAACACAGTACAGGCTACACTCTCTGTGCACTAGCTTGTAAATGTGCAGTGGCGGAGTGGCTGGGA

230 240 250 260 270 280 290 300 310 320 330
On_Cyp19a1gene
AGCAAGTTCAGCTTACAGCGTTATCATATCTCATGGCCAACTCTCCAGTGTCTTTAAATGTGAATCGTGTATAAATCCCAAGTGTGCTCGGGCAGCATAAAGCT

340 350 360 370 380 390 400 410 420 430 440
On_Cyp19a1gene
GAGCCTCTGCTTTCCTTTGCTGCCGTCAAAAATCATCAGTCTCTCCTCTGTTCTGTTAGTTTGAAGTGAAACGTCACCCCTCTTAACTTCTGAGGTTTTACATCAGGA

450 460 470 480 490 500 510 520 530 540 550
On_Cyp19a1gene
CGTGATTAATAAATAAATCTGATTTTAGCAGGCTATTTAAATACAACCTCAGAAAGGCGGCAACACCTGGCATACTTCCAGCACACTTTTTTTAAAGCAGTGTGAGT

560 570 580 590 600 610 620 630 640 650 660
On_Cyp19a1gene
TCATTGAGATTTGAGGCATTTCTTTCCAGAAAGACTGATTTAGATGTGCCGTTGATCTTTCAGAGAGACCAGCCCTTCCACTGTGGCAACCAACAGTCTATTCTTAA

670 680 690 700 710 720 730 740 750 760 770
On_Cyp19a1gene
TTGACTGCTTGAACCTTGTGATTTAACATGTAGCTGAAGTTCCAGTTTCTTCCAGCACCTGACAGTCTGAATTTGCTGGGACGTCCTACTCCTGCATGACTGTGGA

780 790 800 810 820 830 840 850 860 870 880
On_Cyp19a1gene
TGGCGATGACTGTCAATGAATCACTGTACAGTAGTGATAAATATAGGTTGATAAATTTGATATAACCCGTTGTCAGTATGTAAAGCATTAGAATGAAAGAGGGGTGAC

890 900 910 920 930 940 950 960 970 980 990
On_Cyp19a1gene
TTAATTTTTATGCTTATCTAATATGTGATATAATAATAAATCCTACATAATAACATAAATCTGACATTTAAACCTTTTTTAAACGCTCGTTTCTGAATAACTCCCTA

1000 1010 1020 1030 1040 1050 1060 1070 1080 1090
1100
On_Cyp19a1gene
AACATGCTAGTGTATAAAGATGCGGTTGACCAACAATCAGAGCCAGAGACGAGGTTACGCACACTTCTACTTTTCCAGGAGAAACAAAGGGGGCTGTCGCCCTTAATC

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
1210
On_Cyp19a1gene
ATGCAGAACATCTAGGTCTACGCCATCACACAGGGCATCGTTATCGAGGCAGGACACAGCACAGTCACTTTGAGTGTGCACATCTTGGCTGGAGTGTGGG

1220 1230 1240 1250 1260 1270 1280 1290 1300 1310
1320
On_Cyp19a1gene
CGGTGAGCTTGCAAGGTTCTTTATGCAGAAATGAGCCAGAAACAAAGGCAAGCTCAAGAGCAGACAGCATGGGTTATGTTGAGGATTTTTAGAAATTAATAACTCAGC

1330 1340 1350 1360 1370 1380 1390 1400 1410 1420
1430
On_Cyp19a1gene
AGAGAACCGAAAGTGGCTAAAAGTCAAGATAATGTCAATGTTCACTGAGTTTTTCTGTTCTATTTTTTAGTTTACAGACCAGCTTAACTGATGACCTCTTTTTTT

1440 1450 1460 1470 1480 1490 1500 1510 1520 1530
1540
On_Cyp19a1gene
TCTCCACTTTGTTTTCTGGCAGGAAAATTTCTACACGAGTCTCTCCATTAGCTCCAGCTATCCACTCTTTCTGCAAACTGTCTGGGGCTGGAGTTAACAGATCAGTC

1550 1560 1570 1580 1590 1600 1610 1620 1630 1640
1650
On_Cyp19a1gene
TAGCTTTACTGTAGAACATAGGCATATTAATCTTCAATTTATTTGTCTGATAAACAATAAACAATACAGTATGGTATCTGCATTACCACCTGGTATTATGTAACCTTAT

1660 1670 1680 1690 1700 1710 1720 1730 1740 1750
1760
On_Cyp19a1gene
TCTTGGCTTCACTGGATCTGAGAGCAAGCCCAATAAATAAACCAGAAAGCCAGGTTTAAATGGATAGCGTTTGCCTTTTTTAACTTTAAATTCATTACTGTAAT

1770 1780 1790 1800 1810 1820 1830 1840 1850 1860
1870
On_Cyp19a1gene
ATGATCACATTTGTGAGTATAATGAATACATAGCAGCTCTAATTAACAACAATAGCTTTGATTTTGAATGAAAATGTATCACCTAAAAACCAAGTCAGCTTTTGAACA

1880 1890 1900 1910 1920 1930 1940 1950 1960 1970
1980
On_Cyp19a1gene
TAAATCTGGAATCAATGTAATGAAGATTATAATTCAAATGCATCTACATATGTAATATTAACATTTAATCCATTCCCTTGTCTCGCTTCTTTAATAAACAAGATGAAA

1990 2000 2010 2020 2030 2040 2050 2060 2070 2080
2090
On_Cyp19a1gene
AAATTCAGTTTATGAAAAGCTGATTTTTAACAGGAGCTAGAGCAATAACACTGTGATGTTCTCAAACTCCCAAGGCAAGCAACTCTTTTGAAGTTTAAAGGAGG

2100 2110 2120 2130 2140 2150 2160 2170 2180 2190
2200
On_Cyp19a1gene
TGAAAACCTTATCATCAACAAAATGCCAAATAGGAGCTCAGCCCAACATGCAGCCATTGAAGTGGTCTCAAACTCCCAAGGCAAGGCTGAGCACAACCAATCCCTCAAGG

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
2310
On_Cyp19a1gene
TTGCCACAATATGTTTACCTTTTCTGCGAGTGTGGCTTTTGCATTACCTGACTGGCTGTAACCGACTCAGACCCCTATAAAAAGGCAGATGCACACTTCGGG

```

2420          2320      2330      2340      2350      2360      2370      2380      2390      2400      2410
On_Cyp19a1gene  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
cyp19a1a_ENSONIT0000  ACCCTCGAGTCTGTGCAGGCTTTCTACATCATCACCCCTCTCATGGATCTGATCTCTGCTTGTGAACAGGCGATGAGTCCTGTAGGCTTAGACGCCGTGGTGGCAGAT
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
                                                                M N P V G L D A V V A D

2530          2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
On_Cyp19a1gene  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
cyp19a1a_ENSONIT0000  CTCTGTGACCTCAAATGCCATCCAATCG ATGGGATATCAATGGCAACCAGAACGCTGATACTGCTCGTCTGTCTGCTTTGGTTGCCCTGGAGTCACACGGACAAGAA
CTCTGTGACCTCAAATGCCATCCAATCGCATGGGATATCAATGGCAACCAGAACGCTGATACTGCTCGTCTGTCTGCTTTGGTTGCCCTGGAGTCACACGGACAAGAA
L S V T S N A I Q S H G I S M A T R T L I L L V C L L L V A W S H T D K K

```

Figure S2: *Cyp19&1* gene sequence

EXON 1 -85

10 20 30 40 50 60 70 80 90 100 110

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

120 130 140 150 160 170 180 190 200 210 220

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

230 240 250 260 270 280 290 300 310 320 330

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

340 350 360 370 380 390 400 410 420 430 440

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

INTRON 1

450 460 470 480 490 500 510 520 530 540 550

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

EXON 2

560 570 580 590 600 610 620 630 640 650 660

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

670 680 690 700 710 720 730 740 750 760 770

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

INTRON 2

EXON 3

780 790 800 810 820 830 840 850 860 870 880

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

F-885

890 900 910 920 930 940 950 960 970 980 990

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

1000 1010 1020 1030 1040 1050 1060 1070 1080 1090 1100

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

INTRON 3

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200 1210

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

R-1308

1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

EXON 4

1440 1450 1460 1470 1480 1490 1500 1510 1520 1530 1540

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

1550 1560 1570 1580 1590 1600 1610 1620 1630 1640 1650

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

TTTCCCATCTCTGGTTCATTGTGTACCTACTTACATTTCTCTCATTGTAGAGCAAAACCTAAAAAGCATCTTGATTGGTGA AAAATCAGGAAGTAACATCAGCATGA
TTTCCCATCTCTGGTTCATTGTGTACCTACTTACATTTCTCTCATTGTAGAGCAAAACCTAAAAAGCATCTTGATTGGTGA AAAATCAGGAAGTAACATCAGCATGA
AGCAAAACCTAAAAAGCATCTTGATTGGTGA AAAATCAGGAAGTAACATCAGCATGA

INTRON 4 K Q N L K S I L I G E K S G S N I S M

EXON 5

1660 1670 1680 1690 1700 1710 1720 1730 1740 1750 1760

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

GTCCACTTCTACTTTTCTCCGGCGGAACGGGAACGTATACGAGGTTCAGCCCGGCTTTCTTTCTGCTGATATTCACTGTTCATCAGAGACCGCTCAGCTTTGGTTTTTATT
GTCCACTTCTACTTTTCTCCGGCGGAACGGGAACGTATACGAGGTTCAGCCCGGCTTTCTTTCTGCTGATATTCACTGTTCATCAGAGACCGCTCAGCTTTGGTTTTTATT
GTCCACTTCTACTTTTCTCCGGCGGAACGGGAACGTATACGAGGTTCAGCCCGGCTTTCTTTCTGCTGATATTCACTGTTCATCAGAGACCGCTCAGCTTTGGTTTTTATT

INTRON 5

S P L L L F S G G T G T D T

●→ **F-1802**

1770 1780 1790 1800 1810 1820 1830 1840 1850 1860 1870

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

TTTGGTTTCCAGATGTGCTTCAGGCTCGCCCCGGCATCTCTGCAAACTCCTTCCCTTTGTGAGATGAAACGCTTCTCGGGTCTGTTCCTCCCTCAGGAACACTTCAGC
TTTGGTTTCCAGATGTGCTTCAGGCTCGCCCCGGCATCTCTGCAAACTCCTTCCCTTTGTGAGATGAAACGCTTCTCGGGTCTGTTCCTCCCTCAGGAACACTTCAGC
AGATGTGCTTCAGGCTCGCCCCGGCATCTCTGCAAACTCCTTCCCTTTGTGAGATGAAACGCTTCTCGGGTCTGTTCCTCCCTCAGGAACACTTCAGC

R C A S G S P P A S L Q T S F L C E M K R F L G A V L P Q E H F T

EXON 6 **R-1958** ←

1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

TCCCTCCACTTCTCTGGACTCCTTACAGTCTCTGCCTCCCTCTCCGCTTGGCTTATCCTCCAGCGAGACCTCTGTCGAGTAATGATCAACTCCACAGCTCCACAGT
TCCCTCCACTTCTCTGGACTCCTTACAGTCTCTGCCTCCCTCTCCGCTTGGCTTATCCTCCAGCGAGACCTCTGTCGAGTAATGATCAACTCCACAGCTCCACAGT
TCCCTCCACTTCTCTGGACTCCTTACAGTCTCTGCCTCCCTCTCCGCTTGGCTTATCCTCCAGCGAGACCTCTGTCGAGTAATGATCAACTCCACAGCTCCACAGT

S P L P L D S L Q S L Q S L P P L S L G L S S S E T L L A V M I N S T A P T V

1990 2000 2010 2020 2030 2040 2050 2060 2070 2080 2090

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

CTTTGGCTTCCAGAGTGGGGCTCTGTGTCGGGTGTCACCGGAGAGCTGGCCCTGTCTGCTGCACGTGTTAGAGGAGCTCAGACAGAGACTGGACCAGACTTTGGTGC
CTTTGGCTTCCAGAGTGGGGCTCTGTGTCGGGTGTCACCGGAGAGCTGGCCCTGTCTGCTGCACGTGTTAGAGGAGCTCAGACAGAGACTGGACCAGACTTTGGTGC
CTTTGGCTTCCAGAGTGGGGCTCTGTGTCGGGTGTCACCGGAGAGCTGGCCCTGTCTGCTGCACGTGTTAGAGGAGCTCAGACAGAGACTGGACCAGACTTTGGTGC

F G F T S W G S V L P V C H G E L A L S A A L L E E L R Q R L D Q T L V

2100 2110 2120 2130 2140 2150 2160 2170 2180 2190 2200

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

AAATGACAGAAATAATCAGAGAGGAAGAGGTTTCACTGGGAGCCAAAGGAGAGCTGGGGAGGCTCAAAGAACTGAGTGCCTTACAGGAGAAAGAACATGCCACAGGTGAC
AAATGACAGAAATAATCAGAGAGGAAGAGGTTTCACTGGGAGCCAAAGGAGAGCTGGGGAGGCTCAAAGAACTGAGTGCCTTACAGGAGAAAGAACATGCCACAGGTGAC
AAATGACAGAAATAATCAGAGAGGAAGAGGTTTCACTGGGAGCCAAAGGAGAGCTGGGGAGGCTCAAAGAACTGAGTGCCTTACAGGAGAAAGAACATGCCACAGGTGAC

Q M T E I I R E E V S L G A K E E S L G R L K E L S A L Q E K E H A

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300 2310

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

ATGTGCAAGAACAGTTACAGTTACATAGGAAGCATTTTTTTTTTAAATGAAACAACCCCTGCAGCACGGTGGGGTGCATTAAGTATCCCGCTGCAGCTTTTCAGCTCC
ATGTGCAAGAACAGTTACAGTTACATAGGAAGCATTTTTTTTTTAAATGAAACAACCCCTGCAGCACGGTGGGGTGCATTAAGTATCCCGCTGCAGCTTTTCAGCTCC

●→ **F-2340**

2320 2330 2340 2350 2360 2370 2380 2390 2400 2410 2420

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

CAGCTTTAGGAGATGTGGGTGGAGGAGTGGCCACGGATTTCTTAAACCCCAATTTATAACAGGCCCTTACAGTCAAAGGTTTATGGGCCGAAAACAGTAATCATGCTGCA
CAGCTTTAGGAGATGTGGGTGGAGGAGTGGCCACGGATTTCTTAAACCCCAATTTATAACAGGCCCTTACAGTCAAAGGTTTATGGGCCGAAAACAGTAATCATGCTGCA

INTRON 6

2430 2440 2450 2460 2470 2480 2490 2500 2510 2520 2530

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

GACTGCTGTTTCACTTACAGCGGTCCAGTACCTATGAGTTTCAAGTATTTTAAAGTCTTGTGCAAAACAGATTTCCCAATTTTAAAGTTCTGAAGTCTG
GACTGCTGTTTCACTTACAGCGGTCCAGTACCTATGAGTTTCAAGTATTTTAAAGTCTTGTGCAAAACAGATTTCCCAATTTTAAAGTTCTGAAGTCTG

2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

GCAGATTAATAAATAAATAAATAGATCTCATTATCAAAAACCTAAAAACCTGATGAAATTTAGATAGATATCAGCAGATCTTTTAAAGAAACTAATCATGT
GCAGATTAATAAATAAATAAATAGATCTCATTATCAAAAACCTAAAAACCTGATGAAATTTAGATAGATATCAGCAGATCTTTTAAAGAAACTAATCATGT

2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

GAGTGTGTGAGCTTTGGCTGAGTGTGCAAAAAGGCTTCTGTTCTATTATTATATCATTGCTGTTTTGTCAGTGGTAAAAATTTGGGTCTATATTACCCACTGACCTTTGA
GAGTGTGTGAGCTTTGGCTGAGTGTGCAAAAAGGCTTCTGTTCTATTATTATATCATTGCTGTTTTGTCAGTGGTAAAAATTTGGGTCTATATTACCCACTGACCTTTGA

2760 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

CATATATAGATGAGCAACTATATAAATTTAATTTATTAGATTTTGTATTGTTGCAATTTTCTCTGCTGTGTTTATCAGTTTATCAGCTCCACAGACTCTCAAAACT
CATATATAGATGAGCAACTATATAAATTTAATTTATTAGATTTTGTATTGTTGCAATTTTCTCTGCTGTGTTTATCAGTTTATCAGCTCCACAGACTCTCAAAACT

R-2883 ←

2870 2880 2890 2900 2910 2920 2930 2940 2950 2960 2970

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

GGACTCGTTTCACTTTTGTCTTTTCTTAGCGTGAAATTTGCTGCTCACTCTAAAGGTCACGTATATCCCTTTTAAAGTCTGCTGATGTTTTCTTTCTGCTCCGC
GGACTCGTTTCACTTTTGTCTTTTCTTAGCGTGAAATTTGCTGCTCACTCTAAAGGTCACGTATATCCCTTTTAAAGTCTGCTGATGTTTTCTTTCTGCTCCGC

2980 2990 3000 3010 3020 3030 3040 3050 3060 3070 3080

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

CAGGAGGGAGTCACTTCCGCTGTTTCTTCTGCTGAAGGCTCTGCAGCGTGGCCCAAAGTACGACGCGCAAGAAACTGCGGGCCACCAGAGCAGACCCCAAGTTCCG
CAGGAGGGAGTCACTTCCGCTGTTTCTTCTGCTGAAGGCTCTGCAGCGTGGCCCAAAGTACGACGCGCAAGAAACTGCGGGCCACCAGAGCAGACCCCAAGTTCCG
CAGGAGGGAGTCACTTCCGCTGTTTCTTCTGCTGAAGGCTCTGCAGCGTGGCCCAAAGTACGACGCGCAAGAAACTGCGGGCCACCAGAGCAGACCCCAAGTTCCG

T G G S Q F R V F L L K A L Q T V A Q T Y D A Q R K L R A T R A D P S S

Amh_HG5187878

EXON 7

3090 3100 3110 3120 3130 3140 3150 3160 3170 3180 3190

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005
AmhY_HG5187878

TCAGTGAGGGGCGGCCTCTGTGGGCTGAAGGCTCTCACCGTGTCCCTGACAAAGCTTCTGTCGGCCCAAGCAGCGCAAACATTAACAATTGCCACGGCCCTGCACGTT
TCAGTGAGGGGCGGCCTCTGTGGGCTGAAGGCTCTCACCGTGTCCCTGACAAAGCTTCTGTCGGCCCAAGCAGCGCAAACATTAACAATTGCCACGGCCCTGCACGTT
TCAGTGAGGGGCGGCCTCTGTGGGCTGAAGGCTCTCACCGTGTCCCTGACAAAGCTTCTGTCGGCCCAAGCAGCGCAAACATTAACAATTGCCACGGCCCTGCACGTT
S V R G G V C G L K A L T V S L T K L L V G P S S A N I N N C H G C T F
TCAGTGAGGGGCGGCCTCTGTGGGCTGAAGGCTCTCACCGGTTGTGAGGGAGTGTGATGCCCGTAGAGCTGCTCTTCTGCTACTTTACCCAGCATTTATAACT
S V R G G V C G L K A L T G C E G V W M P L E L L S S A T L P Q H L *

3200 3210 3220 3230 3240 3250 3260 3270 3280 3290 3300

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

CCCTCTGACCAAAGGCAACAACCAACGCCATCCTGCTCAACTCCCACATCGAGACGGCAACCGGATGAGCGTTCGCCCTGCTGTGTGCCCGTGGCATACGAAGCCCTGG
CCCTCTGACCAAAGGCAACAACCAACGCCATCCTGCTCAACTCCCACATCGAGACGGCAACCGGATGAGCGTTCGCCCTGCTGTGTGCCCGTGGCATACGAAGCCCTGG
CCCTCTGACCAAAGGCAACAACCAACGCCATCCTGCTCAACTCCCACATCGAGACGGCAACCGGATGAGCGTTCGCCCTGCTGTGTGCCCGTGGCATACGAAGCCCTGG
P L T N G N N H A I L L N S H I E T G N A D E R S P C C V P V A Y E A L

R-3328 ←

3310 3320 3330 3340 3350 3360 3370 3380 3390 3400 3410

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

AGGTTGTGGAAGGCAACGAGATGGGACCTTCTCTCCATCAAGCCAGATGCGGTTGCGAGGGAGTGTGGATGCCGCTAGAGCTGCT
AGGTTGTGGAAGGCAACGAGATGGGACCTTCTCTCCATCAAGCCAGATGCGGTTGCGAGGGAGTGTGGATGCCGCTAGAGCTGCT
AGGTTGTGGAAGGCAACGAGATGGGACCTTCTCTCCATCAAGCCAGATGCGGTTGCGAGGGAGTGTGGATGCCGCTAGAGCTGCTCTCTTCTGCTACTTTACCCAGC
E V V D W N A D G T F I S I K P D A V A R E C G C R *

3420 3430 3440 3450 3460 3470 3480 3490 3500 3510 3520

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005
AmhY_HG5187878

ATTTATAACTTTACACAACTGATTAGTTATGCGCATGTTTCAGTTATGAAATATCAATTACAATACACCAGTGTACTTCTAATTAATATAACTGCTTTTATGCTTTA

3530 3540 3550 3560 3570 3580 3590 3600 3610 3620 3630

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

TTAAAGATCACAACTCTTAAGATCAAAAAGTTAATGTTAATCTATTTATGTTAAGATATTTTATAAATAGTGTAAATGAATCAGATTTTGTCAATTTGAAAAT

3640 3650 3660 3670 3680 3690 3700 3710 3720 3730 3740

Amh_ON_ModGene
Amh_ON_Gene
gi|542212005

ATTTCTAATTGGAATAATGTTTCTGAGTGTAAAGTTAAACTGTTAAAGGATGAAATATA-----

Figure S3: *Amh* gene sequence of *O. niloticus*

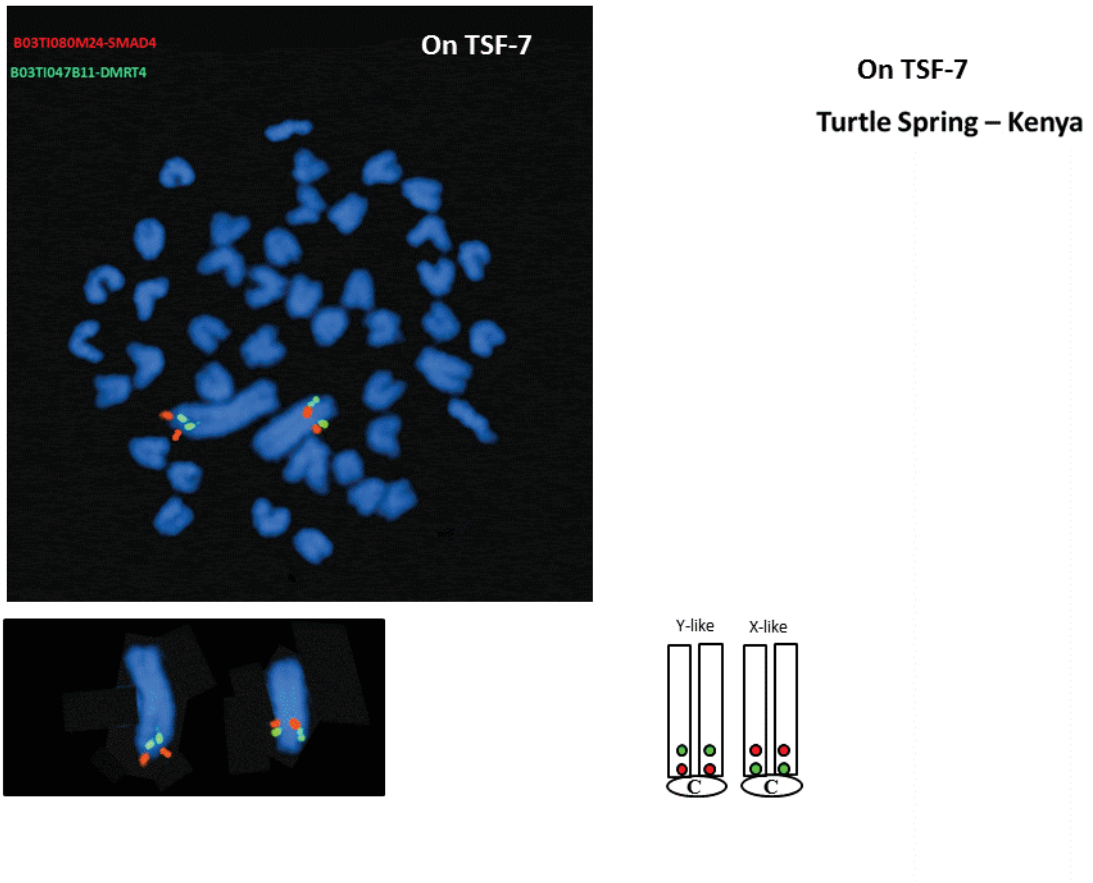


Figure S4: Sex inversion on fish from Turtle Spring based on BAC and FISH

(III) Morphological variations of wild populations of Nile tilapia living in extreme environmental conditions in the Kenyan Rift-Valley

Titus Chemandwa Ndiwa^{1,2}, Julien Claude³, Dorothy Wanja Nyingi^{1,2}, Jean-François Agnès⁴

1 Kenya Wetlands Biodiversity Research Team, National Museums of Kenya, P.O. Box 40658-00100 Nairobi, Kenya

2 Ichthyology Section, National Museums of Kenya, P.O. Box 40658-00100 Nairobi, Kenya

3 Département Forme, UMR IRD 226 CNRS 5554, Institut des Science de l'Evolution, Université de Montpellier 2, Montpellier, France.

4 Département Conservation et Domestication, UMR IRD 226 CNRS 5554, Institut des Science de l'Evolution, Université de Montpellier 2, Montpellier, France.

Abstract

In this study, we use geometric morphometric to establish morphological differences between natural populations of Nile tilapia from two extreme environmental conditions (high temperature and salinity) in Kenya, and compare them to two populations from regions experiencing less extreme conditions. The study observed significant morphological differences between all studied populations, including three closely related hot spring populations (Bogoria, Chelaba and Turtle Springs), and two populations with similar genetic background inhabiting saline environments in Lake Turkana basin (Turkana and Crocodile Lake populations). In addition, allometric growth patterns of the seven populations differed significantly, demonstrating that developmental changes (plastic or adaptive) operated between these populations. All observed differences were attributed either to genetic and/or environmental factors, which seemed to play a major role in influencing morphology of wild Nile tilapia populations. We recommend further studies to be carried out under controlled conditions to confirm the role of temperature, pH and salinity in morphological diversification of Nile tilapia.

1. Introduction

Cichlids are tropical freshwater fishes well known for their high rates of speciation often resulting in rapid radiation. They exhibit remarkably high levels of diversity that affect their morphology, ecology, behaviour and genomes (Nelson, 1994; Barlow, 2000; Chakrabarty, 2005). A large extent of cichlid diversity is observed in three main East African Lakes (Victoria, Tanganyika and Malawi) where 1800 to 2000 species (60% of all cichlid species

worldwide) are present in species flocks (Meyer *et al.*, 1990; Stiassny and Meyer, 1999; Kerschbaumer and Sturmbauer, 2011). These cichlid species flocks are primarily monophyletic groups of closely related species inhabiting the same ecosystem, and evolved from a single ancestral species (Greenwood, 1974). Within these three East African Lakes, some cichlid species went through rapid morphological diversification and speciation, leading to formation of many new species within a few thousand generations (Sage *et al.*, 1984; Meyer *et al.*, 1990; Owen *et al.*, 1990). In Lake Victoria for instance, the radiation of monophyletic haplochromine species occurred within a geologically short time of about 750,000 years and resulted to the emergence of approximately 500 cichlid species flocks (Greenwood, 1984; Abila *et al.*, 2004; Kerschbaumer and Sturmbauer *et al.*, 2011). Additionally, evolutionary older cichlid assemblages of Lake Tanganyika have shown the highest degree of morphological differentiations (Salzburger *et al.*, 2002).

Under the ecological theory of diversification (Schluter, 2000), a strong relationship exists between speciation events and colonization of new environments, or niche diversification, which may in turn induce changes in organism's behaviour, morphology and physiology (Price *et al.*, 2003). Generally, selection pressures that exist in new environments favour divergence of populations from their ancestral forms. A classic example is provided by the case of the three spined sticklebacks (*Gasterosteus aculeatus*) in Canada, where two sympatric forms (limnetic and benthic) emerged as a consequence of ecological divergence (Schluter and McPhail, 1993). This example among others has provided evidence of morphological diversifications as a result of environmental variations. However, Gagnaire *et al.* (2013) proposed that ecological divergence may actually involve multiple facets of species divergence on which divergent selections may operate, hence the need for an integrative approach to study the process of speciation. Some of the factors proposed to influence divergence include variations in quantitative phenotypic traits, regulation in gene expression and differential transmission of particular allelic combinations (Gagnaire *et al.*, 2013).

The Nile tilapia (*Oreochromis niloticus*), the subject of our study, is a teleost fish of the family Cichlidae, has its native range in Africa and is widely known for its importance in aquaculture. This species has been recognized for its ability to tolerate a wide range of salinity, dissolved oxygen and temperature (Avella *et al.*, 1993; Farmer and Beamish, 1969), a high fecundity and a rapid growth rate (Liti *et al.*, 2005) and an omnivorous mode of feeding (El-Sayed, 1999; Peterson *et al.*, 2005). Due to these attributes, the species has the ability to adapt rapidly to new environments, showing a wide range of biological responses to different

environmental conditions both in culture and in nature (Seehausen, 2006; Schofield *et al.*, 2011). As a result of this plasticity (or of the high degree of physiological tolerance found in the Nile Tilapia), the species has been able to successfully establish itself in extreme environmental conditions such as temperate winter conditions (Peterson *et al.*, 2005; Grammer *et al.*, 2012), hot springs (Trewavas, 1983; Nyingi *et al.*, 2009) and saline waters (Schofield *et al.*, 2011). Nile tilapia has in addition been reported to consistently out-compete other species of fish even in areas beyond its natural ranges, especially as colonisers of newly emerging lacustrine ecosystems (Boulenger, 1898). Studies by Collyer *et al.*, (2005) in pupfish (*Cyprinodon tularosa*) noted that the geographical spread of species across a broad range of ecological conditions is accompanied by equally diverse morphological variations, which are strongly correlated to the environmental conditions.

Within the Kenyan Rift Valley, Nile tilapia has colonized numerous newly formed habitats within a system characterised by a number of drainages comprising a series of rivers, springs and lakes. This complex system is a host to four out of the seven described sub-species of *Oreochromis niloticus* by Trewavas (1983): *O. n. baringoensis* from Lake Baringo; *O. n. sugutae* from River Suguta; *O. n. vulcani* from Lake Turkana. Recently Nyingi *et al.* (2009) discovered a natural population from Lake Bogoria Hotel genetically differentiated from but nevertheless close to the Lake Baringo population. The purpose of this current study is to investigate whether populations of the Nile Tilapia have diverged significantly morphologically from each other, and whether these differences can be related to environmental factors (either as the result of phenotypic plasticity or adaptive evolution).

According to Collyer *et al.* (2005, 2007), variations in salinity have the potential of affecting fish body shape. These authors reported a general decrease in body circumference to overall body size at high salinity on pupfish (*Cyprinodon Tularosa*). Similarly, high water temperatures can also influence both the growth rate and development rate of an organism (Atkinson, 1995). In Zebra fish (*Danio rerio*) for instances, raising of juveniles under four different developmental temperatures resulted to highly variable phenotypes (Sfakianakis *et al.*, 2011). A similar study by Georgakopoulou *et al.* (2007) on European seabass (*Dicentrarchus labrax*, L) showed that samples reared under lower temperatures (15 °C) were more slender than those reared under higher temperature (20 °C). Even though no studies have been carried out to establish the effect of water pH on fish shape, extreme changes in water pH have been reported to cause fish deaths by altering other aspects of water chemistry,

for instance increase in pH increases toxicity of ammonia (Schofield, 1976; Wurts and Durborow, 1992).

Taking into consideration the local genetic variations observed in three recently discovered populations of *O. niloticus* within the hot springs (Chelaba, Lake Bogoria Hotel and Turtle) of Loboï Swamp (Nyingi, *et al.*, 2009; Ndiwa *et al.*, 2014), we hypothesized that these fish populations may also differ in shape due to the influence of high salinity and high temperature. In addition to the Loboï Swamp fish, another population of *O. niloticus* found in the northern part of the Kenyan Rift-Valley also experiences extreme environmental conditions. The Crocodile Lake, one of the three Lakes located on the Central Island of Lake Turkana is saline, with electrical conductivity as high as 10,590 $\mu\text{S}/\text{cm}$ and alkaline with a pH of around 10.0 (Avery, 2010). Depending on the level of Lake Turkana, Crocodile Lake can either be submerged or emerged. During high water level the lake has been known to disappear (Hay, 1976; Cerling, 1979) while in low water levels, it remains isolated. The last time the two lakes were reported to have connected was in 1902 (Hopson *et al.*, 1982). So far only three species of fish have been reported to survive in Crocodile Lake namely: *Clarias gariepinus*, *Haplochromis sp.*, and *O. niloticus*.

Both the Loboï and Crater Lake populations have adapted to extreme conditions, high temperature and high salinity levels respectively, while other closely related populations (Lakes Baringo, Turkana and Victoria) do not experience such conditions. These populations thus present an opportunity to observe how extreme conditions affect the morphology of the fish. To achieve this goal, we compared individuals from seven populations, four with extreme conditions (hot springs; Lake Bogoria Hotel Spring, Chelaba Spring, Turtle Spring and saline; Crocodile Lake populations), and three that experience more favourable environmental conditions (Lakes Baringo, Turkana and Victoria).

The study provides an insight into morphological responses of these fish populations to extreme temperature and salinity. Occurrence of phenotypic differences between the populations studied will confirm that the environment or/and the genetic background play an important role in evolution of cichlids, while lack of phenotypic differences will imply that the high degree of tolerance in Nile Tilapia is related to its propensity to invade new environment, and in keeping a very canalized phenotype.

2. Materials and methods

2.1 Study areas

Specimens from three hot spring populations of Loboï swamp drainage system (Chelaba, Lake Bogoria Hotel and Turtle) and other three Lakes (Baringo, Turkana and Crocodile) were studied (Table 1; Figure 1). The three hot springs are characterised by elevated water temperatures (approximately 36°C), low dissolved oxygen concentration (3.5 – 3.9 mg l⁻¹) and slight acidity pH (6.4 – 6.9) (Ashley *et al.*, 2004). The springs drain into Loboï swamp of about 1.5 Km², which is covered by dense vegetation of *Typha domingensis* and *Cyperus papyrus*. As the water exits the swamp, it is poorly aerated (Dissolved Oxygen (DO) = 0.2 - 1.1 mg l⁻¹) and has a higher pH (7.4 – 8.3) (Ashley *et al.*, 2004). Comparatively, Crocodile Lake is a crater formed Island lake within the Lake Turkana. It is characterised by relatively higher water pH (10) and almost three times higher salinity (conductivity, 10,590 µS/cm) than Lake Turkana, which has pH and a conductivity of 9.3 and 3,420 µS/cm respectively (Avery, 2010). Lake Baringo, a fresh water Lake located about 52 Km from the Loboï swamp has D.O. concentration ranging from 5.89-6.7 mg/L, temperature range of 22.1 - 31.8 °C, pH~8.5 and conductivity of about 580 µS/cm (Omondi *et al.*, 2014). Lake Victoria, also fresh water body is characterized by water temperatures ranging from 23.8 to 31.5 °C, pH 6.2 - 9.2, and conductivity of approximately 170 µS/cm (Ochumba, 1990; Crul, 1995).

2.2 Collection of samples

Fish specimens were collected from the hot springs (Chelaba, Lake Bogoria Hotel and Turtle) and Lakes Baringo, Victoria, Turkana and Crocodile using seine nets and scoop nets. Fish were anaesthetised and killed using an overdose of MS-222 before fixing in 4% formalin solution for seven days. The formalin was washed and voucher specimens preserved in 70% ethanol, and are curated as reference material at the National Museums of Kenya. The standard lengths (SL) of the fish were measured using a digital Vernier calliper in order to estimate their sizes.

2.3 Geometric morphometrics

Body shape of the samples was quantified using landmark based geometric morphometric method (Rohlf and Marcus, 1993). Images of preserved samples of fish were obtained by using a flatbed scanner (Herler *et al.*, 2007) and imported into the program tpsUtil version 1.53 (Rohlf, 2012). A total of 14 landmarks (figure 2) were chosen based on their capacity to

capture overall body shape, and their coordinates (x,y) digitized on the left hand side of each sample using the program tpsDig version 2.16 (Rohlf, 2010).

A total of 237 fish specimens from seven wild population of Nile tilapia (Bogoria Hotel, 41; Chelaba, 30; Turtle 24; Baringo, 43; Victoria, 33 Crocodile, 28 and Turkana, 38) were analysed. Their standard sizes ranged from 40mm to 207mm.

The measurement error was estimated by re-scanning and land-marking of 11 randomly chosen fish samples from Lake Turkana. The size and shape measurement Error (ME) was then estimated using ANOVA and Procrustes ANOVA following the procedure of Yezerinac *et al.* (1992), as explained in Claude (2008; 2013). In this approach, the measurement error is compared to among fish variations, and measurement error variation.

2.4 Data analysis

Data analysis was carried out using R software program packages (Claude, 2008; 2013). In order to establish morphological variations between the different populations of fish, we used principal component analysis (PCA) of the superimposed coordinates of the landmarks. Variations were visualised by plotting multivariate ordination of specimens on the first two principal component (PC) axes. Contribution of original variables on corresponding eigenvectors or PC loadings were analysed to establish shape features that had the greatest influence on each PC.

Nested MANCOVA models were conducted on the principal components (PC) of shape data in order to assess the influence of locality, and environmental factors. Habitat and locality were considered as explanatory variables and centroid size as a covariate in the model. Size was added as a covariate to understand the effect of habitat and locality on allometric growth patterns.

$$\text{Response} \sim \text{Size} + \text{Habitat} + \text{Habitat:Size} + \text{Habitat:locality} + \text{Habitat:Locality:Size}$$

To represent and appraise morphological differences between the localities, a linear discriminant analysis was performed on the dataset of the Procrustes coordinates.

A mantel test analysis (Mantel, 1967) was also performed on Fst Values and Mahalanobis distance in order to confirm the influence of genetic factors on the morphology. Lastly, minimum spanning network based on Mahalanobis distance was constructed in order to visualize phenotypic relationship between the seven wild populations of Nile tilapia.

3. Results

3.1 Measurement error

In the current study, we estimated the size and shape measurement Error using ANOVA and Procrustes ANOVA. The percentage measurement error was less than 1 % for the centroid size and 60 % for shape.

3.2 Size variation

Comparison of fish sample sizes between the seven populations of Nile tilapia (Fig. 3) indicated minimal variations between specimens from Lake Baringo, Chelaba Spring, Crocodile Lake and Turtle Spring. However, greater sample size differences were observed between the latter populations and populations from Bogoria Spring, Lake Turkana and Lake Victoria. Size comparison among individuals of same population showed least variations within Crocodile Lake (Fig. 3).

3.3 Shape variation (pca)

The first shape PC accounted for 23% of the observed variations. This PC was mostly related with the variation of the head, caudal peduncle and anal fin (Fig 4). The PC2 on the other hand accounted for 14.5% of variation mostly within the depth of the body and caudal peduncle (Fig. 4).

Even though there were overlaps between most of the populations under study, significant morphological differences were observed between all populations. A strong separation was observed between the hot spring populations and both Crocodile Lake, and Lake Turkana populations based on PC1. However, PC2 showed greater differentiation between Lakes Turkana and Crocodile populations than the latter (Fig 4).

Surprisingly, we observed that populations of Lakes Turkana and Victoria were close morphologically. These results were unexpected since the two systems have different ecological characteristics.

3.4 Differences between populations (and eventually habitats)

Both differences in habitats and localities significantly explained the shape variations between the seven populations of Nile tilapia studied (Table 2).

Linear discriminant analysis based on the sampling locations correctly assigned 84.8 % of the individuals to their respective localities. All samples from Lake Victoria were assigned to their real group (table 3). However, most of individuals from the hot springs populations that were not correctly assigned to their respective localities were either assigned to another hot spring population(s), or to L. Baringo (table. 3). Assignment of individuals from other localities is shown on table 3.

Linear discriminant analysis (LDA) clustered the seven populations of *O. niloticus* into two distinct groups. The first cluster consisted entirely of individuals from Lake Victoria, while the second cluster comprised the other six populations (Fig. 5).

However, partial separations were observed among the hot spring populations and also among Lakes Turkana and Crocodile populations (Fig. 5). Shape variations related with both LDA1 and LDA2 were attributed to differences within the head and the caudal region.

Because we observed a significant interaction of locality with size, we compared small sized individuals to large sized individuals from each locality in order to establish their differences and similarity in growth. Generally, we observed a decrease in head size and length of the mouth with increase in size in all the six populations except Crocodile Lake, where no marked differences were observed between the heads of the small and large sized samples (Fig 6).

The closely related hot spring populations displayed different patterns of growth irrespective of their close spatial relationship. In Bogoria spring, small sized individuals were characterized by more elongated mouths compared to the other two spring populations. On the other hand, the heads of small sized individuals from Chelaba Spring showed slight elongation with greater dorsal displacement of their mouths. We also observed posterior displacement of the anal fin as individuals increase in size for Bogoria and Chelaba populations. Finally, large sized samples from Chelaba and Turtle Springs had elongated caudal peduncle length compared to small sized individuals, while no differences were observed for Lake Bogoria Hotel Spring population (Fig. 6).

Growth patterns of the Lake Turkana system (Turkana and Crocodile) populations differed significantly among themselves, and among other *O. niloticus* populations. For instance, great change in shape was observed between the small sized samples and the large sized samples involving drastic decrease of the depth of the body and caudal peduncle. Dorsal displacement of the mouth and anterior displacement of the eye, dorsal fin and operculum were also noted.

Smaller differences were observed between small and large samples in Lake Turkana; anterior elongation of the mouth for small sized individual, anterior displacement of the lower jaw and pelvic fin of mature individuals, and slight elongation of the caudal peduncle (Fig. 6).

Within the two control populations (Lakes Victoria and Baringo), marked differences were observed between small and large sized samples. Comparison between their growth patterns showed that Lake Victoria samples had deeper caudal peduncles that increased in length with maturity of the fish, while Lake Baringo samples had slender caudal peduncles, which also increased in length with fish growth (Fig. 6).

Additionally, correlation of F_{st} values and Mahalanobis distance based on Mantel test analysis showed existence of significant relationship ($P < 0.05$) (Fig. 7).

Minimum spanning network computed from Mahalanobis distances between localities showed close proximity between the three hot spring populations. Additionally, Lake Turkana population showed close relationship to Lake Victoria population, while Crocodile Lake and Lake Turkana populations were separated (Fig. 8).

4. Discussion

Morphological analysis revealed that specimens studied were first differentiated by their sizes. Fishes from Lakes Turkana and Victoria were the largest with a mean size of 122.8 ± 4.38 mm (mean \pm standard error) and 163.3 ± 4.40 mm, respectively. The smallest specimens were from Turtle Spring and Crocodile Lake, and exhibited mean sizes of 76.6 ± 6.11 mm and 70.1 ± 1.72 mm, respectively. Differences were also observed in terms of variability of sizes; some samples had higher size range (Lake Victoria) than others (Crocodile Lake). These observed differences may partly represent artefacts due to sampling, and may not completely reflect the actual differences of fish sizes in different localities despite the knowledge that Lakes Turkana and Victoria host large specimens of *O. niloticus* (Trewavas, 1983) due to their larger sizes and greater depths, compared to biotopes investigated in the present study.

The shape measurement error reported by our study is relatively higher than expected. This can partly be explained by the fact that it comprises measurement error, digitization error, and variations among the digitizing and scanning devices. Additionally, the study considered a single population in determining shape measurement error implying that the value of among

fish variations (which divides the measurement error variations during calculations of shape measurement error) was low hence contributing to a higher shape measurement error value.

In our study, we observed different allometric patterns of growth in all the seven populations of Nile tilapia. Considering the fact that most of the localities differ greatly in their environmental conditions, the observed allometries may be partly attributed to environmental differences, which might have acted on developmental stages, and shaped the differences observed between *O. niloticus* populations. Whether differences are adaptive or the by-product of phenotypic plasticity is still speculative at that stage but could be tested by doing experimental reciprocal transplantation of fish (see however the end of the discussion).

Shape changes found in our study could be functional and environmentally related. Indeed, our observations are congruent to findings by Ramler *et al.* (2014) who investigated the effect of different temperatures on body size and shape on stickleback (*Gasterosteus aculeatus*). The authors reported significant differences between individuals across different temperature groups. A similar study on zebrafish (*Danio rerio*) juveniles reported occurrence of highly variable phenotypes induced by diverse thermal conditions during the early ontogenic stages (Sfakianakis *et al.*, 2011). However, results obtained by Fujimura and Okada (2008) comparing (during their developmental stages) the lower jaw bones of a cichlid fish (*Haplochromis chilotes*) from Lake Victoria and another Riverine cichlid (*O. niloticus*) contradicted our findings. Irrespective of the habitat differences, the authors noted that most shape changes had a similar pattern during the growth of the two species showing that the development of these fish were relatively robust.

The PC analysis revealed a large amount of shape differences whereby variations of the head, caudal peduncle and anal fin accounted for 23% (Fig. 4). These variations allowed discrimination of the Lake Victoria specimens from those of other Lakes (Baringo, Turkana and the hot springs). Within these latter populations, fishes from Lake Turkana were placed between the two groups of fishes.

Differentiation of the Lake Victoria population from all other population based on linear discriminant analysis was even clearer (Fig. 5). There was also a clear separation between the hot springs specimens on one hand, and Lake Turkana and Crocodile Lake specimens on the other hand. Specimens from Lake Baringo appeared to fall between these two groups of specimens. All observed variations on both LDA1 and LDA2 were related to differences within the head and the caudal region.

These results are congruent with the genetic variation observed by Agnèse *et al.* (1997), who demonstrated that fishes from Lake Baringo and Lake Turkana are genetically close. Later on, Nyingi *et al.* (2009) and Ndiwa *et al.* (2014) observed that fishes from the hot springs are even more closely related to fishes from Lake Baringo. If one assumes that fishes from Crocodile Lake should have a genetic background identical to the fishes of Lake Turkana, and that fishes from Lake Victoria introduced from Lake Albert (Fryer and Iles, 1972, Trewavas, 1983) should be less genetically related to the other fish studied, then morphological differentiation seems to be congruent with genetic differentiation. Indeed, our studies based on mantel test confirmed the role of genetic factors in influencing morphology. In addition, links between morphological and genetical differentiation have been reported by Stewart and Albertson (2010), who integrated both genetic and geometric morphometric approaches in studying scale eating cichlids of Lake Tanganyika, which had been reported to show frequency balanced polymorphism in the left-handedness and right-handedness of the mouth, an adaptation previously reported by Takahashi *et al.* (2007). Their morphometric results strongly supported the genetic basis of jaw laterality.

When looking at closely genetically related populations, one can observe a certain amount of morphological differentiation. The first case concerned specimens from the three hot springs. Specimens from Bogoria and Chelaba or Bogoria and Turtle springs were unevenly distributed along the first and second axis of the LDA analysis respectively (Fig. 5). Differences between Chelaba and Bogoria Springs were mainly due to variations in body length and body depth, while differences between Bogoria/Chelaba and Turtle were as a result of variations in head length and body depth.

If we consider that the genetic background of these three populations is the same (Ndiwa *et al.*, 2014), then these observed morphological differentiations could be attributed to phenotypic plasticity of the species. Very likely, environmental characteristics of these hot springs are slightly different. Bogoria Spring is the biggest of the three hot springs and runs a few hundred meters before entering the swamp. The two others have a smaller flow rate and enter in the swamp after only few meters. Nevertheless, as only a few genes can strongly act on morphological differentiation, one cannot exclude that morphological differences observed between the three hot spring populations was due to genetic differences. Ndiwa *et al.* (2014) observed that gene flow between these populations is limited by the swamp, which is covered by a dense vegetation of *Typha domingensis* and *Cyperus papyrus*. Decomposition of organic matter, and shading provided by these vegetation is responsible for low oxygen concentration

and low temperatures, which act as barriers to free movement of the fish from one place to another. Similarly, Ndiwa *et al.* (2014) also observed that these fish are slightly introgressed by *O. leucostictus* genes from escapees of this species originating from aquacultural ponds within the drainage. The account of possible introgression of alien genes acting on morphological differentiation cannot be excluded. Similar study by Zaccara *et al.* (2013) on three cichlid species (*Alcolapia* spp.) based on 13 populations inhabiting lagoons and hot springs of Lake Natron (located between Kenya and Tanzania), like in the current study also showed significant morphological differences between studied populations.

The second case of genetically related populations in our study concerns fish from Lakes Turkana and Crocodile. There was also a clear morphological differentiation between these two groups of fish as shown in Fig. 5. Specimens from Crocodile Lake were generally smaller than those from Lake Turkana, and differed in body depth, head size and caudal length.

These two populations are genetically very close (unpublished results) since Crocodile Lake population originated from Lake Turkana about one century ago. Most likely, the observed differences can be as a result of phenotypic plasticity. These lakes offer two quite different environments to the fish. Lake Turkana is a large alkaline (pH 9.3) and slightly saline lake (conductivity 3,420 $\mu\text{S}/\text{cm}$) hosting numerous fish species (48 following Hopson *et al.*, 1982). Comparatively, Crocodile Lake is a small size lake (about 0.32 Km^2), with high alkalinity (pH 10) and salinity (conductivity 10,590 $\mu\text{S}/\text{cm}$). Only three species of fish are present, two Cichlids (*Aplochromis* sp, *O. niloticus*) and one Clariid (*Clarias gariepinus*).

In conclusion, our study provided evidence that both genetic and environmental factors play an important role in influencing morphological development of *O. niloticus* populations. Despite the findings, further laboratory studies are necessary in order to confirm the importance of these environmental factors. This can be achieved by breeding genetically identical siblings in different environmental conditions: first under low versus high salinity, and secondly under low versus high temperature.

5. Acknowledgement

We thank Mr. Joseph Gathua (NMK), Mr. Gilbert Kosgei (KENWEB) and Dr. Judith Nyunja (KWS) for their support during collection of samples. We are also grateful to KENWEB, NMK, IRD and French Embassy, Nairobi, Kenya for their financial support during this study.

6. References

- Abila R, Barluenga M, Engelken J, Meyer A, Salzburger W. 2004.** Population-structure and genetic diversity in a haplochromine cichlid fish (corrected) of a satellite lake of Lake Victoria. *Mol. Ecol.* 13(9): 2589-2602.
- Agnèse JF, Adépo-Gouréne B, Abban EK, Fermon Y. 1997.** Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity* 79: 88-96.
- Ashley GM, Maitima MJ, Muasya AM, Owen RB, Driese SG, Hover VC, Renaut RW, Goman MF, Mathai S, Blatt SH. 2004.** Sedimentation and recent history of a freshwater wetland in a semi-arid environment: Loboï Swamp, Kenya, East Africa. *Sedimentology* 51: 1301-1321.
- Atkinson D. 1995.** Effects of temperature on the size of aquatic ectotherms: Exceptions to the general rule. *J. Therm. Biol.* 20: 61-74.
- Avella M, Berhault J, Bornancin M. 1993.** Salinity tolerance of two tropical fishes, *Oreochromis aureus* and *O. niloticus*. I. Biochemical and morphological changes in the gill epithelium. *Journal of Fish Biology* 42: 243–254.
- Avery S. 2010.** *Hydrological impacts of Ethiopias Omo basin on Kenya's Lake Turkana water levels and fisheries.* Final Report prepared for the African Development Bank, Tunis.
- Barlow G. 2000.** *The cichlid fishes: the nature's grand experiment in evolution.* Persuus publishing Cambridge, mass, USA.
- Boulenger GA. 1898.** *Catalogue of the fresh water fishes of Africa in British Museum (Natural History).* London UKL.
- Butzer KW. 1971.** *Recent history of an Ethiopian delta. The Omo River and the levels of Lake Rudolf.* Univ. of Chicago, Dept. of Geography Research Paper 136.
- Cerling TE. 1979.** Palaeochemistry of Plio-Pleistocene Lake Turkana, Kenya. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 27: 247-285.
- Chakrabarty, P., 2005.** Testing conjectures about morphological diversity in cihlids of Lakes Malawi and Tanganyika. *Copeia*, 2: 359-373.
- Claude J. 2008.** *Morphometrics with R.* Springer, New York. Pp 316.
- Claude J. 2013.** Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: three worked examples in R. *Hystrix, It. J. Mamm.* 24(1): 94-102.
- Collyer ML, Novak JM, Stockwell CA. 2005.** Morphological divergence of native and recently established populations of white sand pupfish (*Cyprinodon Tularosa*). *Copeia*, 1-11.

- Collyer ML, Stockwell CA, Adams DC, Reiser MH. 2007.** Phenotypic plasticity and contemporary evolution in introduced populations: evidence from translocated populations of sand pupfish (*Cyprinodon Tularosa*). *Ecological Research* 22: 902-910.
- Crul RCM. 1995.** *Limnology and hydrology of Lake Victoria, Comprehensive and Comparative Study of Great Lakes*. UNESCO/IHP-IV Project M-5.1. UNESCO publishing. Pp 79.
- El-Sayed AFM. 1999.** Alternative dietary protein sources for farmed tilapia, *Oreochromis* spp. *Aquaculture* 179: 149–168.
- Farmer GJ, Beamish FWH. 1969.** Oxygen consumption of *Tilapia nilotica* in relation to swimming speed and salinity. *Z Fish Res Bd Canada* 26(11): 2807–2821.
- Fryer G, Isles TD. 1972.** *The Cichlid Fishes of the Great Lakes of Africa*. Edinburgh: Oliver and Boyd.
- Fujimura K, Okada N. 2008.** Shaping of the lower jaw bone during growth of Nile tilapia *Oreochromis niloticus* and Lake Victoria cichlid *Haplochromis chilotes*: A geometric morphometric approach. *Develop. Growth Differ.* 50: 653-663.
- Gagnaire PA, Normandeau E, Pavey SA, Bernatchez L. 2013.** Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). *Molecular Ecology* 22: 3036-3048.
- Georgakopoulou E, Sfakianakis DG, Kouttouki S, Divanach P, Kentouri M, Koumoundouros G. 2007.** The influence of temperature during early life on phenotypic expression at later ontogenetic stages in sea bass. *J. Fish. Biol.* 70(1): 278 -291.
- Grammer GL, Slack WT, Peterson MS, Dugo MA. 2012.** Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) establishment in temperate Mississippi, USA: multi-year survival confirmed by otolith ages. *Aquatic Invasions* 7(3): 367–376.
- Grant PR, Grant BR, Markert JA, Keller LF, Petren K. 2004.** Convergent evolution of Darwin’s Finches caused by introgressive hybridization and selection. *Evolution* 58 : 1588–1599.
- Greenwood PH. 1974.** The cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock. *Bull. Br. Mus. Nat. Hist. (Zool)* 6: 1-134.
- Greenwood PH. 1984.** *African cichlids and evolutionary theories, in evolution of fish species flocks*, A. A. Echelle and I. Kornfield, Eds. University of Maine at Orono press, Me USA, pp 145-155.
- Hay RL. 1976.** *Geology of Olduvai gorge*. University of California Press, Berkeley. Pp 203.
- Herler J, Lipej L, Makovec T. 2007.** A simple technique for simple imaging of live and preserved small fish specimens. *Cybium* 31: 39-44.
- Hopson AJ. 1982.** *Lake Turkana; A report on the findings of the Lake Turkana Project 1972 1975*, Vol. 16. Overseas Development Administration, London. Pp 1614.

- Johnson TC, Malala JO. 2009.** Lake Turkana and its link to the Nile. In: Dumont, H. J. (Ed.), *The Nile: Origin, environment, limnology and human use*. Springer pp. 287-305.
- Kerschbaumer MK, Sturmbauer C. 2011.** The utility of geometric morphometrics to elucidate pathways of cichlids of cichlid evolution. *Int. J. Evol. Biol.* Article ID290245, pp 8.
- Liti D, Cherop L, Munguti J, Chhorn L. 2005.** Growth and economic performance of Nile tilapia (*Oreochromis niloticus* L.) fed on two formulated diets and two locally available feeds in fertilized ponds. *Aquac. Res.* 36: 746–752.
- Mantel N, 1967.** The detection of disease clustering and a generalized regression approach. *Cancer Research* 27 (2): 209–220.
- Meyer A, Kocher TD, Basasibwaki P, Wilson AC. 1990.** Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, 347: 550-553.
- Ndiwa CT, Nyingi D, Agnèsè JF. 2014.** An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Loboï drainage, Kenya. *PloS ONE* 9(9):e106972.
- Nelson JS. 1994.** *Fishes of the world, 3rd edition*. John Wiley and Sons, New York.
- Nicholson SE. 1998.** Historical fluctuations of Lake Victoria and other Lakes in the Northern Rift-Valley of East Africa. In: Lehman, J. T. (Ed.). *Environmental change and response in East African Lakes*, pp 7-35.
- Nyingi DW, Agnèsè JF. 2012.** *Phylogeography of the Nile Tilapia in Africa*. Lambert Academic Publishers.
- Nyingi D, De-Vos L, Aman R, Agnèsè JF. 2009.** Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) (Cichlidae; Teleostei) in the Loboï Swamp (Kenya). *Aquaculture* 297: 57-63.
- Ochumba PBO. 1990.** Massive fish kills within the Nyanza Gulf of Lake Victoria, Kenya. *Hydrobiol.* 208: 93-99.
- Omondi R, Kembanya E, Nyamweya C, Ouma H, Machua SK, Ogari Z. 2014.** Recent limnological changes and their implication on fisheries in Lake Baringo, Kenya. *Journal of Ecology and the Natural Environment* 6(5): 154-163.
- Owen RB, Crossley R, Johnson TC, Tweddle D, Kornfield I, Davison S, Eccles DH, Engstrom DE. 1990.** Major low levels of Lake Malawi and their implication for speciation rates in cichlid fishes. *Proc. R. Lond. B.* 240:519-553.
- Peterson MS, Slack WT, Woodley CM. 2005.** The occurrence of nonindigenous Nile Tilapia, *Oreochromis niloticus* (Linnaeus) in coastal Mississippi: ties to aquaculture and thermal effluent. *Wetlands* 25: 112-121.

- Price TD, Qvarnstrom A, Irwin DE. 2003.** The role of phenotypic plasticity in driving genetic evolution. *Proc. R. Soc. Lond. B* 270: 1433-1440.
- Ramler D, Mitteroecker P, Shama LNS, Wegner KM, Ahnelt H. 2014.** Nonlinear effects of temperature on body form and developmental canalization in the threespine stickleback. *J. Evol. Biol.* 27: 497-507.
- Rohlf FJ, Marcus LF. 1993.** A revolution in morphometrics. *Trends Ecol. Evol.* 8: 129–132.
- Rohlf FJ. 2010.** *tpsDig, digitize landmarks and outlines, version 2.16.* Department of Ecology and Evolution, State University of New York at Stony Brook.
- Rohlf FJ. 2012.** *tpsUtil, file utility program, version 1.53.* Department of Ecology and Evolution, State University of New York at Stony Brook.
- Sage RD, Loiselle PV, Basasibwaki P, Wilson AC. 1984.** Molecular versus morphological change among cichlid fishes of Lake Victoria. In Echelle, A. A. and Kornfield, I (Eds). *Evolution of fish species flocks.* University of Maine at Orono press, Orono, Maine.
- Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. 2002.** Phylogeny of the Lake Tanganyika Cichlid species flock and its relationship to the central and East African Haplochromine cichlid fish faunas. *Syst. Biol.* 51(1): 113-135.
- Schluter D, McPhail JD. 1993.** Character displacement and replicate adaptive radiation. *Trends Ecol. Evol.* 8: 197-200.
- Schluter D. 2000.** *The ecology of adaptive radiation.* Oxford University Press.
- Schofield CL. 1976.** Acid precipitation: effects on fish. *Ambio.* 5(5-6): 228-230.
- Schofield PJ, Peterson MS, Lowe MR, Brown-Peterson NJ, Slack WT. 2011.** Survival, growth and reproduction of nonindigenous Nile tilapia (*Oreochromis niloticus* (Linnaeus, 1758). I. Physiological capabilities to various temperatures and salinities. *Mar. Freshw. Res.* 62: 439–449.
- Seehausen O. 2006.** African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B* 273: 1987–1998.
- Sfakianakis DG, Leris I, Laggis A, Kentouri M. 2011.** The effect of rearing temperature on body shape and meristic characters in zebra fish (*Danio rerio*) juveniles. *Environ. Biol. Fish.* DOI 10.1007/s10641-011-9833-z.
- Sneath PHA, Sokal RR. 1973.** *Numerical Taxonomy.* W.H. Freeman and Company, San Francisco, pp 230-234.
- Spinage CA. 2012.** *African Ecology-Benchmarks and historical perspectives.* Springer Heidelberg Dordrecht, London, Pp 1541.
- Stiassny MLJ, Meyer A. 1999.** Cichlids of the Rift-Lakes. *Sci. Am.* 280: 44-49.

- Takahashi R, Watanabe K, Nishida M, Hori M, 2007.** Evolution of feeding specialization in Tanganyikan scale-eating cichlids: a molecular phylogenetic approach,” *BMC Evolutionary Biology* 7: 195 doi:10.1186/1471-2148-7-195.
- Trewavas E. 1983.** *Tilapiine fishes of the genera Sarotherodon Oreochromis and Danakilia.* British Museum (Natural History), London. Pp 583.
- Wikelski M, Thom C. 2000.** Marine iguanas shrink to survive El Niño. *Nature* 403, 37–38.
- Wurts WA, Durborow RM. 1992.** *Interaction of pH, carbon dioxide, alkalinity and hardness in fish ponds.* SRAC Publication No. 464.
- Yezerinac SM, Lougheed SC, Handford P. 1992.** Measurement Error and Morphometrics Studies: Statistical power and observer experience. *Systematic Biology* 41(4): 471-482.
- Zaccara, S., Crosa, G., Vanetti, I., Binneli, G., Harper, M. D., Mavuti, K. M., Balarin, J. D. and Britton, J. R. 2013.** Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcilia* flock of East Africa. *Conserv. Genet.* DOI 10.1007/s10592-013-0551-2.

Tables

Table 1: Names of the seven populations of *O. niloticus* studied, their respective sample sizes (N), coordinates of study sites and habitat characteristics

Population	Number	Co-ordinates	Habitat
Bogoria Hotel Spring	41	0°21'44''N, 36°03'04''E	Hotspring
Chelaba Spring	30	0°21'30''N, 36°02'58''E	Hotspring
Turtle Spring	24	0°21'44''N, 36°02'41''E	Hotspring
Crocodile Lake	28	3°29'34''N, 36°01'54''E	Saline
Lake Turkana	38	3°30'46''N, 35°54'53''E	Saline
Lake Baringo	43	0°36'43''N, 36°01'31''E	Normal
Lake Victoria	33	0°38'17''N, 33°58'35''E	Normal

Table 2: Results of Multi-Variate Analysis of Covariance (MANCOVA) performed on body shape considering all the samples

Factor	df	Pillai	approx. F	num. df	den. df	P-value
Size	1	0.8282	40.167	24	200	<0.0001
Habitat	2	1.2667	14.464	48	402	<0.0001
Habitat:Size	2	0.3991	2.088	48	402	<0.0001
Habitat:Locality	4	2.0469	8.864	96	812	<0.0001
Habitat:Locality:Size	4	0.5402	1.321	96	812	0.0268

Table 3: Linear Discriminant Analysis (LDA) assignment of individuals to different localities, percentages of assignment are in brackets.

Origin of the fish	False assignation
Lake Bogoria Hotel Spring (82.5%)	Chelaba Spring (10%)
	Lake Baringo (5%)
	Turtle Spring (2.5%)
Chelaba Spring (78.13%)	Bogoria Spring (12.5%)
	Turtle Spring (9.37%)
Turtle Spring (76.19%)	Chelaba Spring (4.76%)
	Crocodile Lake (4.76%)
	Lake Baringo (14.29%)
Lake Baringo (91.43%)	Bogoria Spring (8.57%)
Lake Turkana (94.59%)	Crocodile Lake (5.41%)
Crocodile Lake (83.87%)	Turtle Spring (3.23%)
	Lake Turkana (3.23%)
	Lake Baringo (9.67%)
Lake Victoria (97.06%)	Lake Baringo (2.94%)

Figures

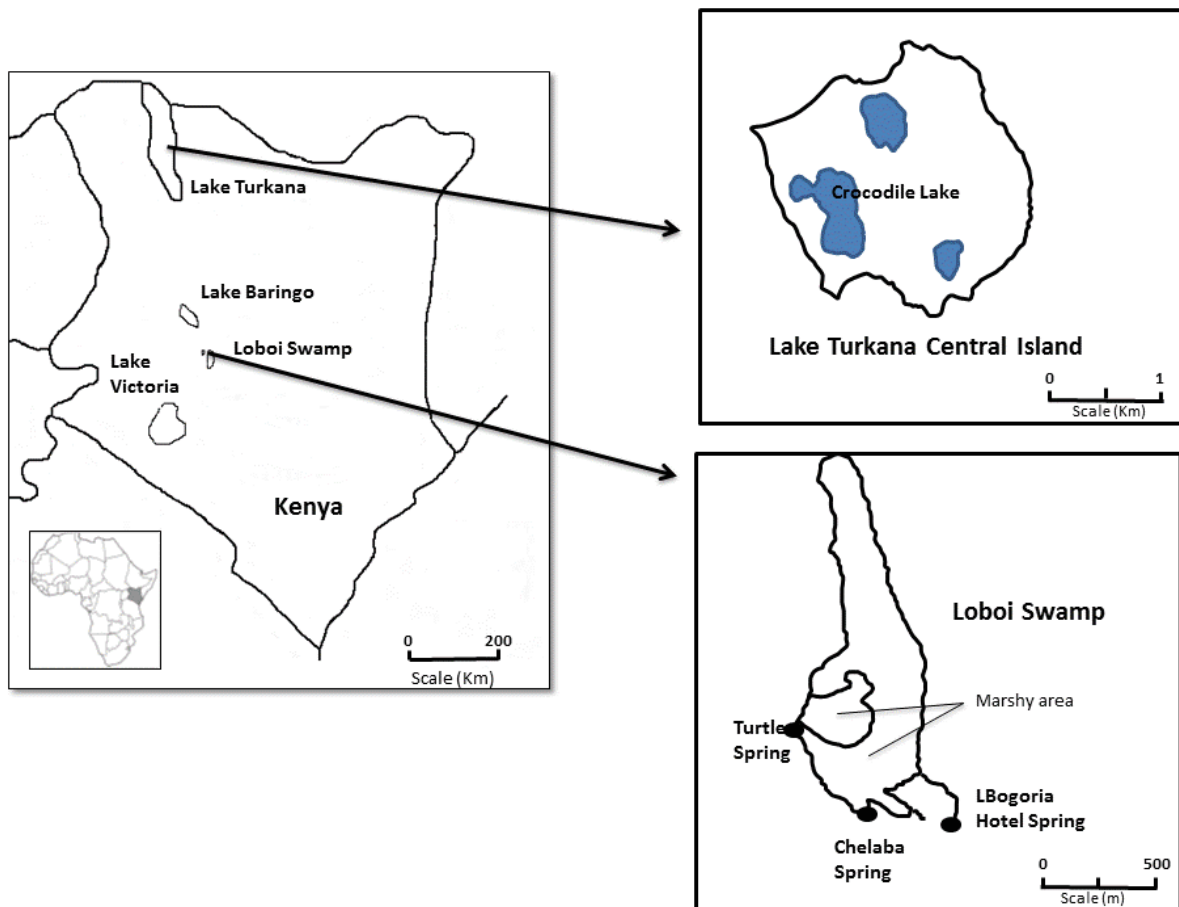


Figure 1: Map showing the location of all the seven natural populations of *O. niloticus*: Lake Turkana at Lake Turkana drainage system, Crocodile Lake in the Central Island of Lake Turkana, Lake Baringo at Lake Baringo drainage system, Turtle Spring, Chelaba Spring and Bogoria Hotel Spring in the Loboï Swamp, and Lake Victoria at L. Victoria drainage system.

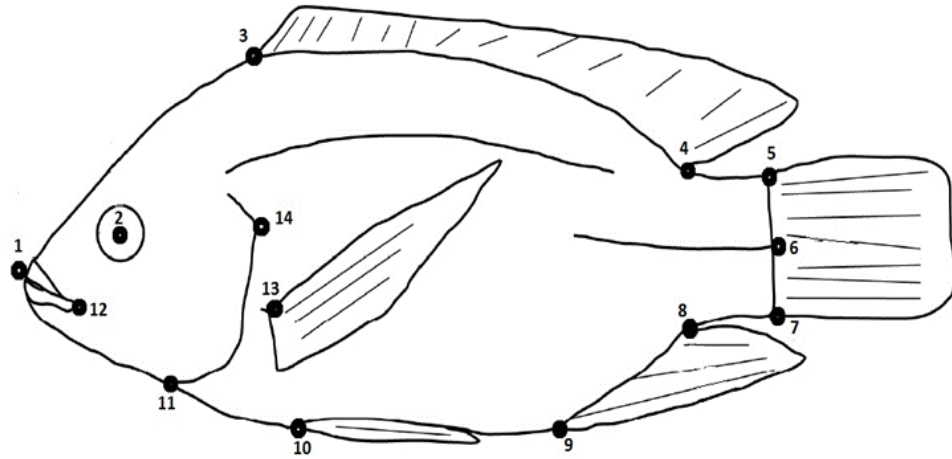


Figure 2: Position of 14 different landmarks used in the study: (1) anterior tip of the snout with mouth closed, (2) centre of the eye, (3 & 4) anterior and posterior insertions of the dorsal fin, (5) dorsal origin of caudal fin, (6) point halfway between dorsal and ventral origin of caudal fin, (7) ventral origin of caudal fin, (8) posterior insertion of anal fin, (9) anterior origin of anal fin, (10) anterior origin of pelvic fin, (11) juncture of the ventral edge of operculum with the ventral outline, (12) most posterior end of the mouth, (13) anterior origin of pectoral fin and (14) most posterior end of operculum.

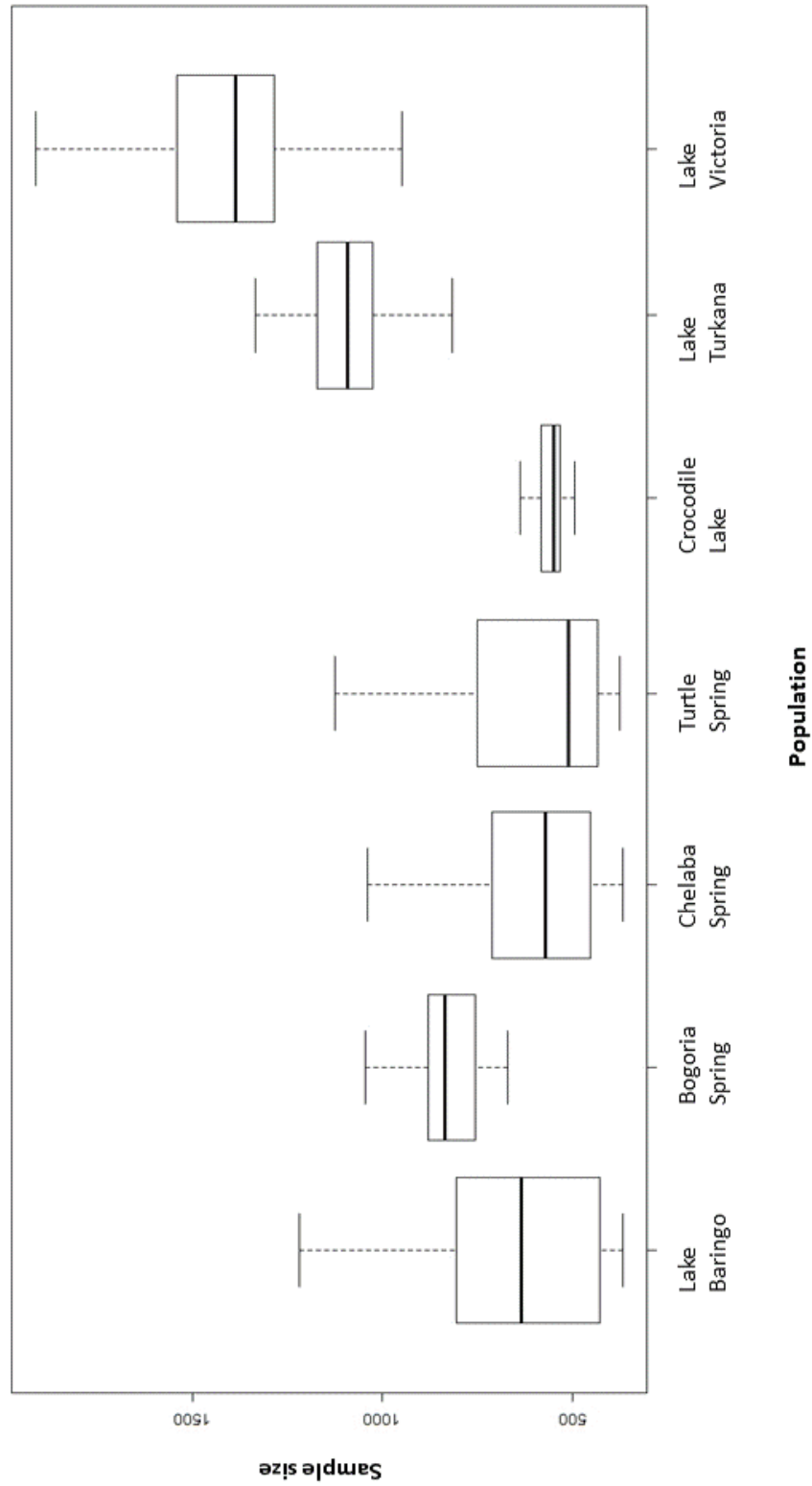


Figure 3: Size distribution of the fish samples collected from the seven localities.

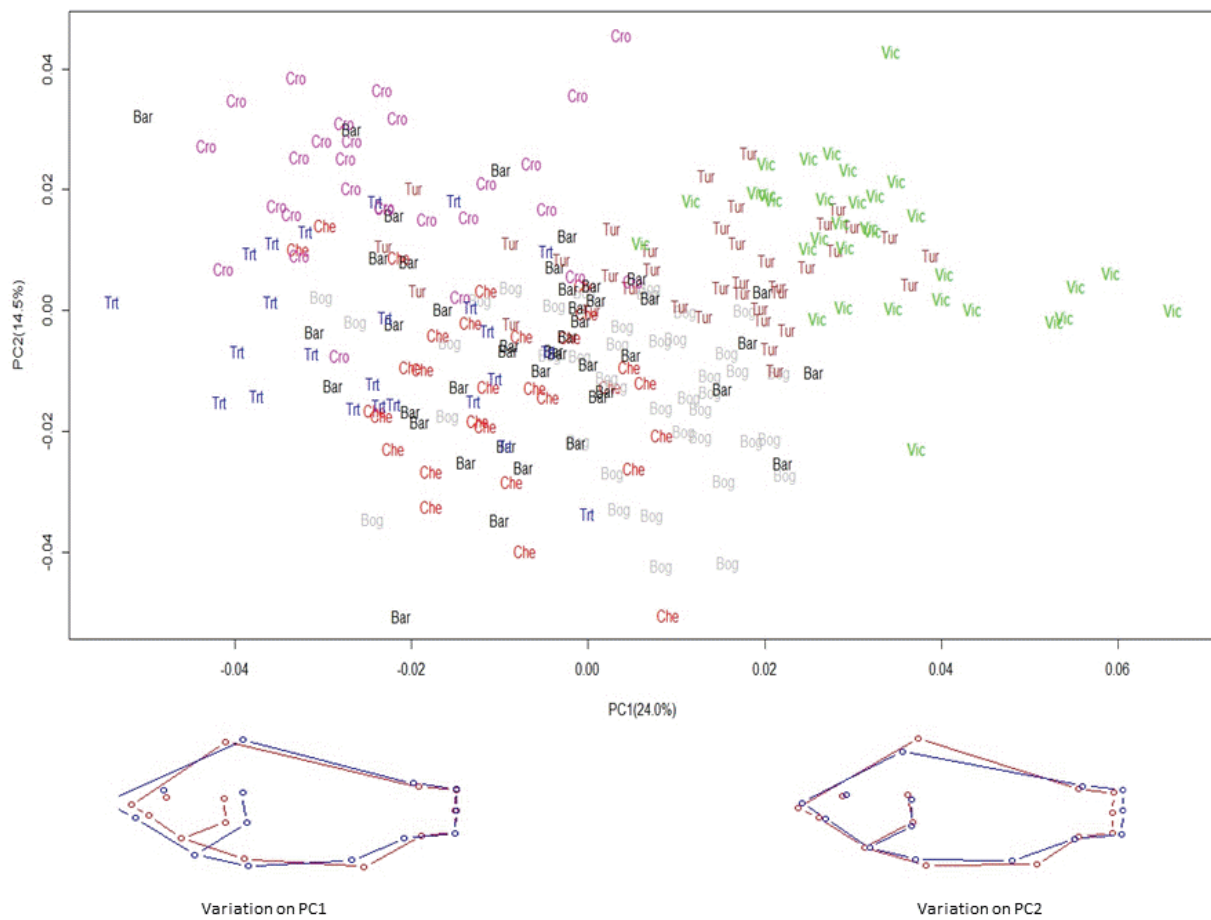


Figure 4: Differentiation of the seven populations of Nile tilapia based on Principal Component Analysis (PCA). On fish shape images (PC 1, 2), red = mature (large samples) and blue small (young) samples. In the scatterplot, cro=crocodile, Tur =Turkana, Che=Chelaba, Bog=Bogoria, Trt =Turtle, Bar=Baringo, Vic=Victoria.

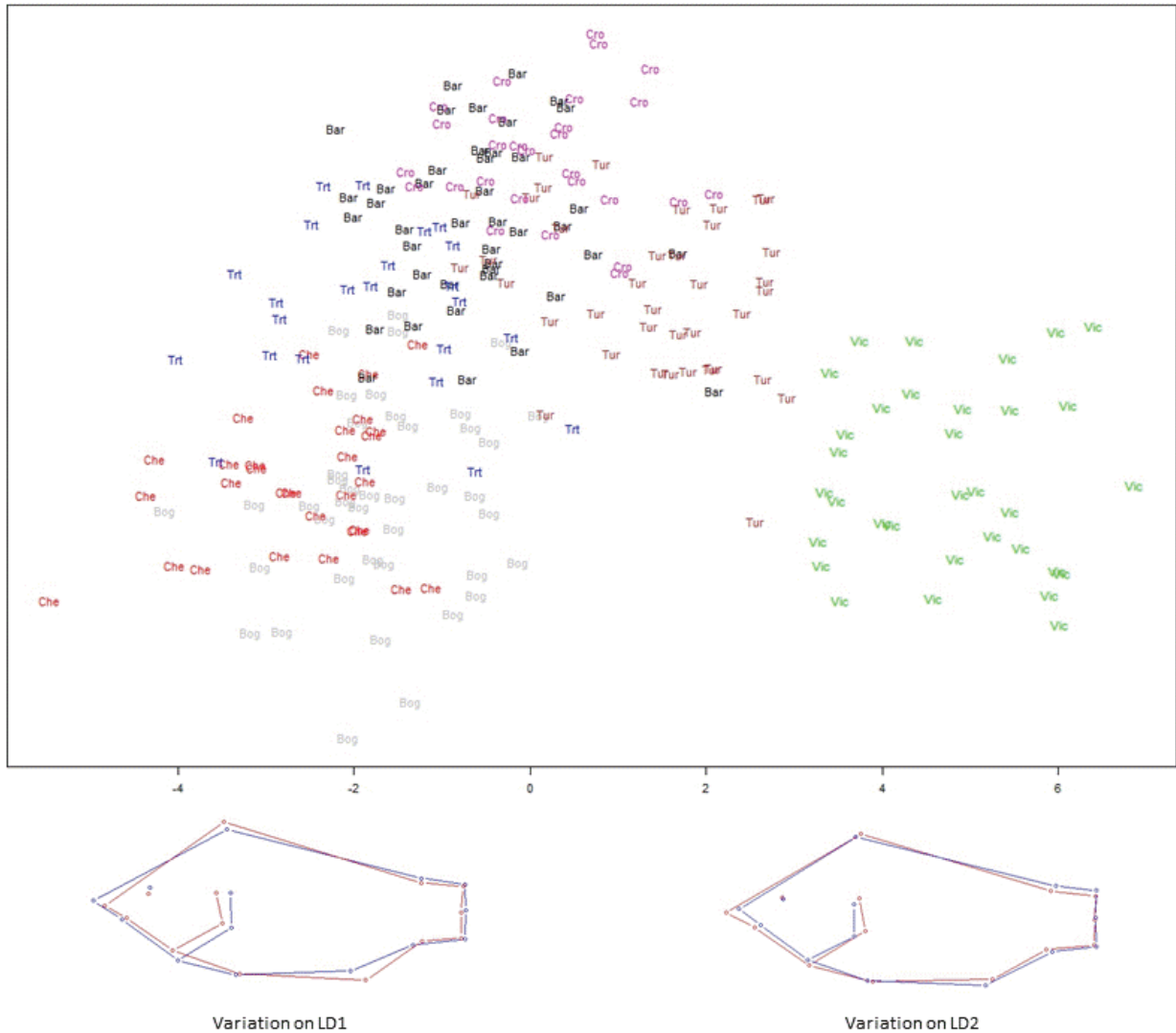


Figure 5: Differentiation of the seven populations of Nile tilapia using Linear Discriminant Analysis (LDA). On fish shape images (LDA 1, 2), red = mature (large samples) and blue small (young) samples. In the scatterplot, cro=crocodile, Tur =Turkana, Che=Chelaba, Bog=Bogoria, Trt =Turtle, Bar=Baringo, Vic=Victoria.

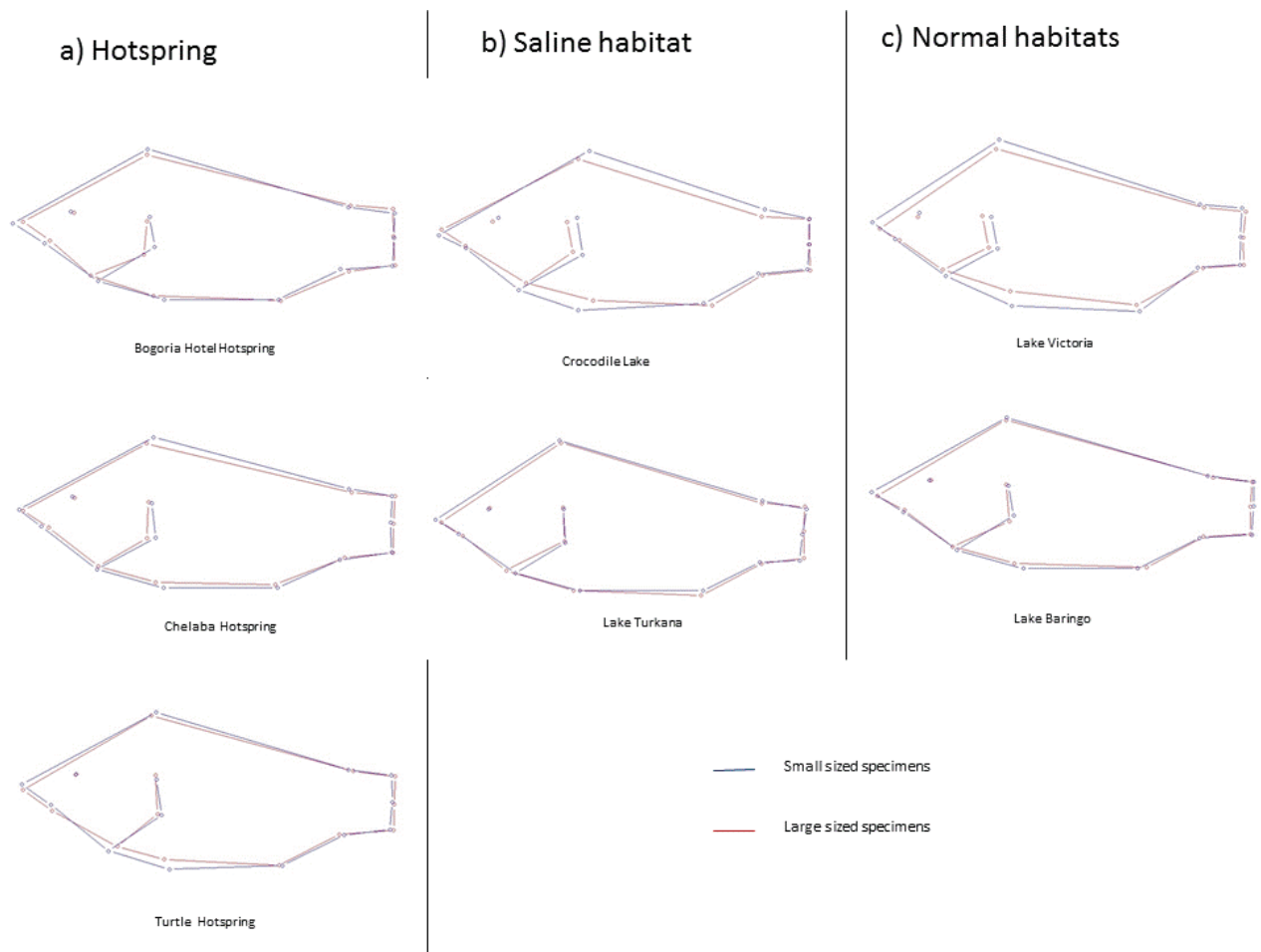


Figure 6: Growth patterns in the seven populations of *Oreochromis niloticus*. Red colour represents the shape of the large samples while the blue colour represents the shape of the small sized individuals.

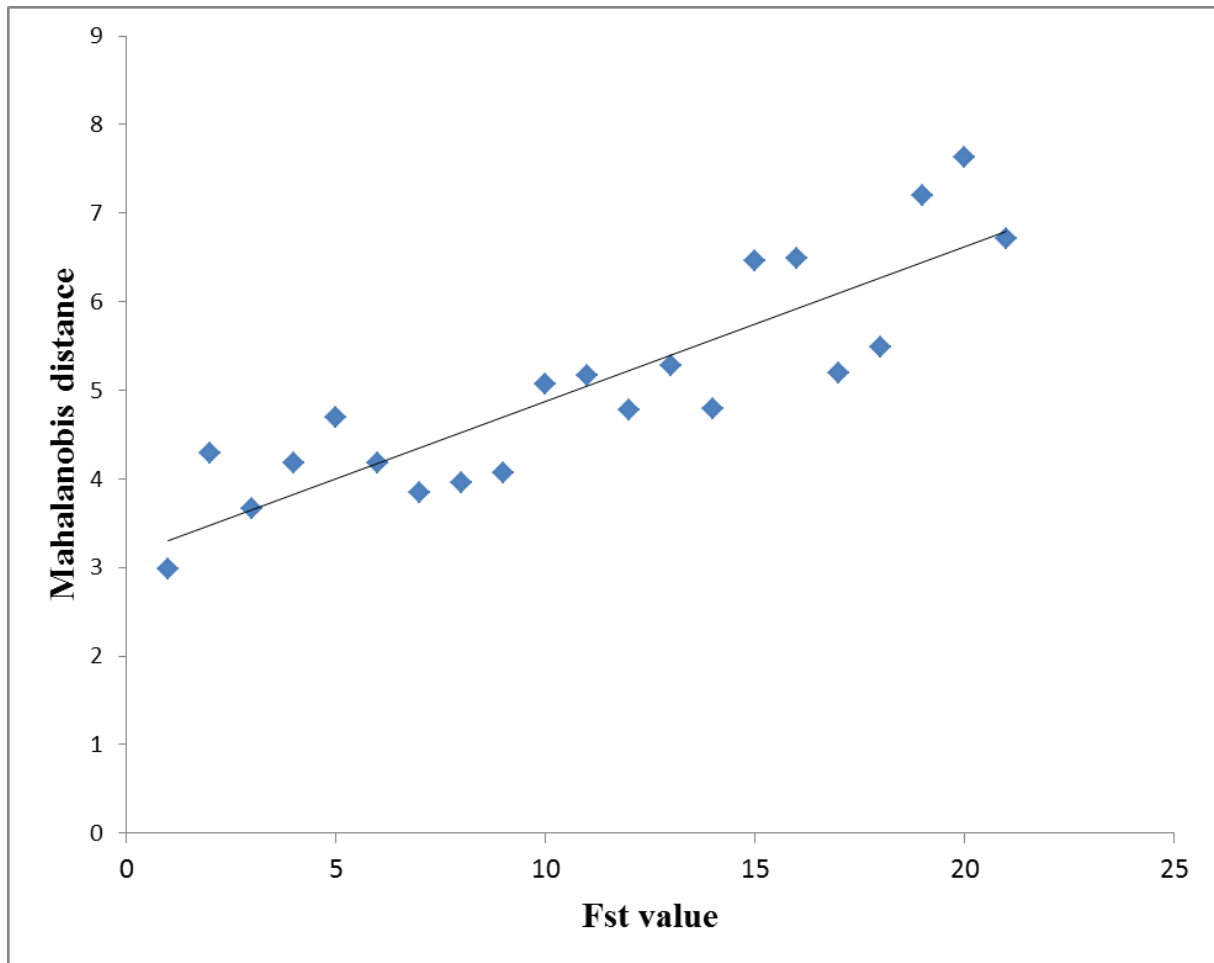


Figure 7: Mantel test of Fst values against Mahalanobis distance showing positive correlation

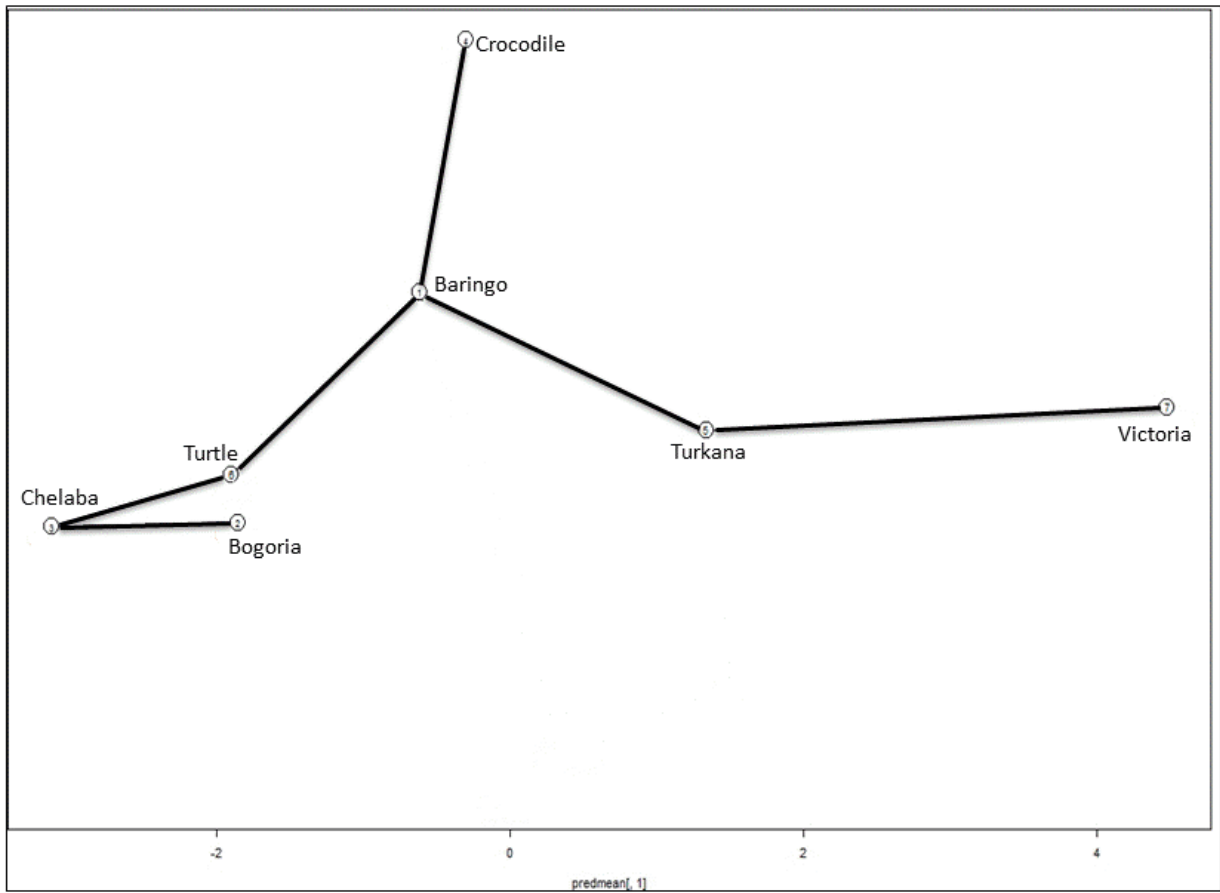


Figure 8: UPGMA dendrogram showing overall similarity in body shape among the *O. niloticus* populations

(IV) General discussion

Nile Tilapia is an emblematic fish from Africa which has been introduced in most if not all the tropical countries. Its ability to adapt to new environments allowed it to colonize most river basins in the countries where it was introduced for farming. One would have thought that culture of this species would have been simple like rearing of either chickens or pigs, both considered as part of the pantheon of domesticated species. However, there are many not well-known characteristics of *Oreochromis niloticus* that still excite scientists and fish farmers.

We know from the pioneer work of Trewavas (1983) that the species is naturally variable, and widely spread in Africa. The author defined seven groups based on morphological differences (meristic and morphometric characters) and called them subspecies. However, we must recognize that this work had a lot of discrepancies, and for a long time needed to be revised, not only with new morphometric methods (as geomorphometrics), but also with genetic tools (sequences of mtDNA, microsatellites) and with evolutionary considerations more in line with the modern view of evolution.

Oreochromis niloticus is an economically important fish species whose worldwide popularity has risen rapidly over the recent past. Its development in aquaculture is partially hindered by the fact that it is difficult to produce male mono-sex fingerlings. Indeed, there are differences in growth rates of males and females, and males tend to grow faster than females on one hand. On the other hand, use of mixed sex cultures sharply reduces growth of fish as a result of reproduction activities that occur within the grow-out ponds. Despite sex determination having been studied extensively in the Nile tilapia, genetic control to produce males is still unreliable. Therefore, the production of male monosex population mainly relies upon sex reversal treatment using 17alpha-methyltestosterone with important concerns about environment and food safety. Sex in this species (as in many other fish species) is determined by the interactions between major (sex chromosomes) and minor (brought by both parents) genetic factors as well as environmental factors. The environmental factor influencing sex in the Nile tilapia is temperature. High temperatures (>32°C) have a strong masculinizing effect on XX female tilapia (Baroiller *et al.*, 1995; Baroiller *et al.*, 2009). This temperature sensitivity of sex differentiation is under genetic control, carried by both parents in an additive way (Baroiller and D'Cotta, 2001; Baroiller *et al.*, 2009; Wessels *et al.*, 2011). Thermosensitivity can be rapidly selected (in three generations) under laboratory conditions

giving breeders which produce ~93% males in their offspring (Wessels *et al.*, 2011). Hence, high temperatures could be exploited in a larger scale for tilapia farming as a way to produce all male progenies. Therefore, further research on sex determination and particularly the effects of temperature in *O. niloticus* is of utmost importance in the context of aquaculture.

One of the most popular strains of *O. niloticus*, especially in Asia, is the GIFT strain. The strain was developed using germplasm obtained from both domestic strains, and wild populations obtained from Egypt, Ghana, Kenya and Senegal (Eknath *et al.*, 1995; Eknath and Acosta, 1998). The success of this strain is certainly due to the selection program employed in its production. However, this success would not have been achieved if the genetic basis of the original strains would have been poor. In other words, it is because the original strain was made from a mixture of different natural genetic resources of *O. niloticus* that the selection was so successful. This provides evidence of the importance of these natural genetic resources for aquaculture.

With the use of genetic tools, these genetic resources are now more studied than before (Rognon *et al.*, 1996; Agnès *et al.*, 1997; Nyingi and Agnès, 2007; Nyingi *et al.*, 2009, 2012; Bezault *et al.*, 2011). However, it is clear that some of *O. niloticus* natural resources are on one hand not fully studied and some are even not yet identified. For example fishes from Lake Tanganyika are yet to be studied and some natural populations may still be waiting to be discovered (like the recent population of Loboï Swamp). On the other hand, these natural genetic resources are threatened by fish farming activities through hybridization with other tilapias species (Ndiwa *et al.*, 2014) or strains (Ansah *et al.* 2014).

In this context, we take the opportunity that Kenya is a hot spot of *O. niloticus* biodiversity in order to study some of these natural populations. From all the possible populations, we focused on those that live in extreme environmental conditions since they could represent cases of specific genetic adaptation, possible morphological differentiation and are under strong anthropogenic threats.

1.1 New genetic resources under threat, the case of Loboï Swamp hot spring populations

In a survey of *O. niloticus* populations in East Africa, the occurrence of a new (previously unknown) native population of *O. niloticus* in Loboï Swamp (Lake Bogoria Hotel Spring), Kenya has been documented (Nyingi *et al.* 2009, 2012). This population was initially

assumed to have been introduced from other localities within the region. However, its native status was confirmed due to its significant and unique genetic variability with a large number of private microsatellite alleles and mtDNA haplotypes (partial D-loop). The presence of tilapia in two other hot springs in the Loboï Swamp namely Turtle and Chelaba Springs, prompted us to investigate these fish. The present study clearly establishes that these two hot springs also host native fish closely related to those from Lake Bogoria Hotel Spring.

Moreover, partial mtDNA sequences of the D-loop region analysis showed that the haplotype of *O. leucostictus* from Lake Naivasha previously confirmed to be present in Lake Baringo, is also present in the three *O. niloticus* populations of the springs thus providing evidence that some of the fish in each of the hot springs have been introgressed by *O. leucostictus*. The previous study of Lake Bogoria Hotel Spring specimens (Nyingi *et al.*, 2007, 2012) did not detect any introgression. This can be due to insufficient number of specimens studied (16) or as a result of very recent introgression events. Indeed, if these introgressions originated from aquaculture practices *i.e.* escapees of *O. leucostictus* specimens or from *O. niloticus* introgressed beforehand in fish farms within the Loboï swamp drainage, as it was hypothesized, then it is possible that these introgressions may have occurred after 2007. The second hypothesis is more likely than the first.

Indeed, fish farming activities within the region have been enhanced as a result of funding by Kenyan government through the Economic Stimulus Programme (ESP) introduced in 2009. The aim of the ESP project was to construct 200 fishponds worth 10.6 million euros in 140 constituencies by June 2013, with each constituency receiving 70,000 € (Nyingi *et al.* 2011). Eight fishponds have been constructed between 2012 and 2013, in close proximity to the Loboï Swamp (within a range of 230–700 metres), and some stocked with Nile tilapia fingerlings from unspecified sources. Other fishponds in the Baringo-Bogoria catchment are located along the main rivers and streams in the region. Fingerlings used to stock the ponds within the region are produced by Omega Fish Farms located at Ol Kokwe Island on Lake Baringo. The farm was established to produce fingerlings for the local fish farmers. Broodstock for the establishment was obtained from the Lake Baringo, hence may have acted as a source of introgressions observed in the Loboï Swamp fish.

Our results based on the 16 microsatellites loci studied clearly indicated that alien nuclear genes are present in *O. niloticus* in the hot springs and Lake Baringo populations. Within these populations, most of the individuals exhibited no or a very small amount of nuclear introgression, and a few specimens appeared to have 90% probability intervals that extended

more than 30% out of a pure species Q value which suggested that these individuals had an important degree of genetic introgression.

Taking into account that the hot spring biotopes are small, and consequently the size of the effective populations, native mtDNA diversity was higher in the hot springs populations (four haplotypes for six individuals in Chelaba Spring; five haplotypes for seven individuals in Turtle Spring) than in Lake Baringo and River Suguta populations. This high diversity may have been favoured by the stable environmental conditions and consequently decreased levels of environmental stress in the hot springs (Ashley *et al.*, 2002, 2004; Deocampo, 2007).

The present ecological situation is now breaking this millenary equilibrium in two ways. We first discussed the threats of aquaculture activities, but there is another important menace: water catchment. The Loboï Swamp receded in size by about 60% over a short period of 30 years due to expansion of irrigation via a ditch constructed in 1970 (Ashley *et al.*, 2004; Owen *et al.*, 2004). In addition, periodic avulsion events have caused changes to river courses in this region. The most recent was during the El-Nino that induced heavy rains of 1997 that caused changes to the course of Loboï River. The river that used to feed Lake Baringo partially changed its course and now also flows to Lake Bogoria. Changes in flow were also influenced by intensive encroachment by local farmers leading to weakening of stream banks (Harper *et al.*, 2003, Owen *et al.*, 2004).

If it was apparent that the three populations of *O. niloticus* in the hot springs are distinct from other related populations by the presences of private alleles and haplotypes, at least partly, this differentiation could be the consequence of their local adaptations to the extreme thermal conditions of their habitat. Studies indicate that exposure of *O. niloticus* to water temperatures of 36 °C or more at an early stage of their development (10 days post fertilization) can influence sex ratios towards males (Baroiller and D'Cotta, 2001; Baroiller *et al.* 2009b). These characteristics are of utmost importance to aquaculture, as they can be utilised to inexpensively produce all-male cultures. Tilapia farming mainly relies on male monosex culture in order to avoid overcrowding in the ponds due to breeding, and to benefit from fast growth of males (Baroiller *et al.* 2009b; Baroiller and Jalabert, 1989; Beardmore *et al.* 2001). Hence the rich genetic diversity and adaptations to masculinizing effect of water temperature makes these populations 1) an important aquaculture resource, and 2) an important natural model for understanding sex determination in Nile tilapias.

1.2 Sex determination in *O. niloticus*: The usefulness of the Lobo Hot Springs populations to disentangle the complex role of temperature.

Temperature and genetic influences on sex may be the result of allelic variations on sex-differentiating genes. Moreover, how these influences occur and how frequently does temperature exert an effect in the wild remain unknown. Our study was carried out in order to investigate allelic polymorphisms related to the phenotypic sex of Nile tilapia in nine wild populations and compare these to two domesticated strains (the Bouaké strain produced in Ivory Coast and the Manzala strain from Egypt). The nine wild populations cover a large portion of the natural distribution of Nile tilapia in Africa, as well as the entire spectrum of thermal regime encountered by the species in the wild. A selection of 192 individuals were analysed with 99 classed as phenotypic males. SNPs were searched on three sex-differentiating genes (*cyp19a1a*, *wt1b* and *amh*) previously found to have allelic variants linked to males (Lee & Kocher, 2007; Eshel *et al.*, 2014; Wessels *et al.*, 2014). The *cyp19a1a* and *wt1b* genes are located on linkage groups LG1 (Lee and Kocher, 2006) while *amh* is located on LG23 (Shirak *et al.*, 2006) for which the linkage analysis of these LGs showed segregation with sex (Lee *et al.*, 2003; 2004; Cnaani *et al.*, 2008; Eshel *et al.*, 2012).

Our study revealed a total of 41 SNPs of which 33 allelic variants were found in the *amh* gene, 4 in the aromatase *Cyp19a* promoter and 4 in the *Wt1b* gene region studied. In addition we also found insertions and deletion varying in size for the *cyp19a1a* promoter and *amh* gene, some of which were unique to certain populations. One of the SNPs was Cyp_2124 which had previously been linked to sex with females being homozygotes (AA) and males heterozygotes (AG) (Lee and Kocher, 2007). In our study we found no polymorphism associated to sex but rather segregation according to populations. No association to sex was found for the other three SNPs found in the *cyp19a1a* promoter. These four SNPs were strongly polymorphic in Lake Baringo and the Lobo fish particularly in the Chelaba hot spring showing linkage disequilibrium between the SNPs. Interestingly, one of the SNPs (Cyp_2175) was found in the binding site for SF1, an important transcription factor responsible for either the activation or repression of the *cyp19a1a* gene. An hypothesis could be that this SNP may be one of the factors causing a repressive action on the *cyp19a1a* gene (reducing the amount of estrogens and causing masculinisation) in temperature sensitive fish or perhaps the reverse, activating the gene and thus increasing the amount of estradiol causing ovarian differentiation (in fish where temperature causes no masculinisation). Likewise, no

polymorphism was seen with sex for the four SNPs identified in the region of the *wt1b* gene studied, but all segregated across populations.

The most interesting result is that in addition to presenting the highest number of SNPs, the *amh* gene also exhibits SNPs with strong relationship to the phenotypic sex. Even though these associations were observed in most populations studied, stronger associations were observed on the Kenyan populations. Within these populations (i.e. Bogoria, Chelaba, Turtle, Baringo and Turkana), 11 SNPs were observed on the *amh* gene that had strong association to phenotypic sex. Analysis of these SNPs indicated significant deviations from the HWE of some populations. For instance, three populations (i.e. Bogoria, Chelaba and Turtle) showed significant deviations on *amh_2509*, while four SNPs (*amh_597*, *amh_611*, *amh_2586*, *amh_2807*) deviated significantly in two populations each. Interestingly, the SNP *amh_597* G>T caused a change in amino acid with females having the T variant while males had the G variant thereby changing from Glutamic acid to Aspartic acid. Likewise, the SNP *amh_611* females had the G variant while males had the A allele changing the amino acid from Serine to Asparagine respectively. SNPs *amh_2618* and *amh_2655* recorded significant deviation from HWE each in one population. Of the 11 SNPs linked to sex, two imply nucleotide changes causing amino acid changes (are synonymous) which might significantly modify the function/activity of the hormone. The other 9 SNPs are all located in intron 6, and why and how they affect the gene is not clear.

Other precedent studies have shown SNPs linked to sex in the *amh* gene in domestic strains of Nile tilapia (Eshel *et al.*, 2014) with some also found in intron 6 (Wessels *et al.*, 2014). Association to sex of these SNPs was not found in our study despite having also used the Manzala strain which might be due to laboratory specific selections. Our results show that SNPs linked to sex are population-specific reinforcing previous findings by Baroiller *et al.*, (2009a; 2009b) and Cnaani *et al.*, (2008), that the LGs associations to phenotypic sex are population-specific. Therefore, determination of sex in *O. niloticus* should be analysed from a population point of view.

Another interesting result from this thesis was the clustering of all the SNPs identified in the sex-linked genes of the *O. niloticus* populations in four distinct groups: (1) the three Kenyan hot spring populations of Bogoria, Chelaba and Turtle as well as the Lake Baringo population, (2) Lake Turkana population, (3) two of the three Ethiopian populations (i.e. the hot spring of Lake Metahara and cold Lake Koka populations), (4) and the fourth group consisted in all Nilo-Sudanian *sensu lato* populations (i.e. the Nilotic Manzala strain, the Sahelo-Sudanian

Kpandu population as well as the synthetic Bouaké strain), but also, less expectedly, the Ethiopian population of the cold crater lake of Hora. This observation suggests that sex chromosomes in Nile tilapia are still in a continuous state of evolution, with possible evolution of different sex-linked haplotypes across the entire natural distribution of the species, hence the observed differences between the studied populations.

We also wanted to address the question of whether the phenotypic sex of the fish studied was the same as the genotypic sex. In other words how many sex-reversed fish could be found amongst our wild populations since tilapia sex is under the influence of both high temperatures and minor genetic factors (Baroiller *et al.*, 2009a). We exploited the recent published “amhy” markers as described in Eshel *et al.*, (2014) which discriminates two bands, a ~1000 bp size band present in XX females and a second ~800 bp band only present in males which carry the Y chromosome, XY and YY individuals. We observed the presence of sex-inverted XX males as well as XY females in several wild populations. A finding which correlates with previous studies in which both XX males and an XY female were indirectly identified *via* progeny testings of wild breeders (Bezault *et al.*, 2007). Nevertheless, as observed in the SNPs, it clearly appeared that the “amhy” markers do not discriminate correctly genetic males and females in all populations. The *amhy* form which has a 2XX deletion was not found in any of the male fish from Lake Metahara, Lake Koka and Lake Hora, meaning that *amhy* does not exist in fish from the Ethiopian Rift Valley. This indicates that the *amhy* is not the universal male-determinant of tilapia.

It would be very informative to use, in all the populations studied, other recently identified X and Y markers (Sun *et al.*, 2014) to see whether the genotypes identified coincide. Unlike *amhy*, these markers can differentiate XY from YY fish. We do not know the origin of the Chinese strain used for the analysis but considering our studies it would not be surprising to find that these markers do not discriminate males from the Ethiopian populations.

Another way to identify sex-reversed fish from the Kenyan populations would be the use of Bacterial Artificial Chromosomes (BAC) markers. An inverted region of more than 1,5 Mb has recently been identified in a region near the centromere in linkage group 3 which corresponds to the large bivalent pair. Using two BAC probes, Fluorescent *In Situ* Hybridizations (FISH) were compared in Nile tilapias between XX females, XY males and YY supermales, showing an inverted region in one of the large chromosomes in XY males with both chromosomes being inverted in YY males (Ozouf *et al.* 2008; D’Cotta *et al.*, 2014). In order to identify the genotypic sex of the Kenyan wild populations and see whether

temperature may be causing a masculinization, we attempted to perform a FISH analysis using the two BACs on metaphase spreads of males and females. The FISH signals seem to be inverted in some individuals such as a TSF7 (female from Turtle Spring) shown in supporting Fig. 1, suggesting the existence of sex-reversed tilapias. However, the FISH analysis was not conclusive because the quality and number of metaphases were low. It would therefore be interesting to continue this study further increasing the number of individuals and the quality of metaphases.

1.3 Morphological differentiation in *O. niloticus*: Genetic versus plasticity, what do we learn from populations living in extreme environments?

East Africa exhibits the largest extent of cichlid diversity with more than 2000 species (60% of all cichlid species worldwide) present, mainly in species flocks (Meyer *et al.*, 1990; Stiassny and Meyer, 1999; Kerschbaumer and Sturmbauer, 2011). Within the three East African Lakes some cichlid species went through rapid morphological diversification and speciation, leading to formation of many new species within a few thousand generations (Sage *et al.*, 1984; Meyer *et al.*, 1990; Owen *et al.*, 1990). Additionally, evolutionary older cichlid assemblages of Lake Tanganyika have shown the highest degree of morphological differentiations (Salzburger *et al.*, 2002).

A strong relationship exists between speciation events and colonization of new environments, or niche diversification, which may in turn induce changes in organism's behaviour, morphology and physiology (Price *et al.*, 2003). Indeed, there are many evidences of morphological diversifications as a result of environmental variations.

Oreochromis niloticus can be a model to studies aiming at determining the role of genetic *versus* environment in morphological differentiation of fish. On one hand natural populations are morphologically differentiated as demonstrated by the pioneer work of Trewavas (1983). On the other hand these populations are genetically differentiated as many authors observed (Rognon *et al.*, 1996; Agnès *et al.*, 1997; Nyingi *et al.*, 2009; Bezault *et al.*, 2011). Moreover, *O. niloticus* was able to colonize extreme environments like hyper-saline lakes (Crocodile Lake) or hot springs (Loboi swamp).

Our study on such remarkable populations revealed that morphological differentiation in *O. niloticus* is both determined by genetic and environmental factors. Genetically close

populations (Lake Turkana and Crocodile Lake on one hand, and Hot Springs and Baringo on the other hand) tend to be morphologically closely related when considering the overall populations. Closer examination revealed that within these groups of populations, a certain amount of morphological differentiation exists. Very likely, the origin of morphological differentiations between Crocodile and Turkana populations could be due to phenotypic plasticity rather than genetic differentiation. The two populations are assumed to share the same genetic background, but live under different environmental conditions (varying in pH and salinity). On the contrary, slight morphological differentiation between Chelaba, Bogoria and Turtle Hot spring populations could be due to genetic factors. Even though the three hot spring populations had similar genetic background, their subsequent separation and low gene flow seems to have played a major role in their differentiation since they live in habitats with very small environmental differences.

Despite the confirmation that the two factors (genetics and environment) play a major role in morphological differentiation, some laboratory experiments are necessary to further confirm these results. Genetically identical fish (siblings) could be bred in different environments (low *versus* high saline concentrations, low *versus* high temperatures) in order to determine the importance of pH, salinity and temperature.

1.4 Future research perspective

The current study clearly established occurrence of genetic introgression between *O. leucostictus* and *O. niloticus* populations of the three hot springs of Loboï Swamp. It will be interesting to monitor and precisely elucidate the origin of this introgression. In addition, follow up studies are necessary in order to establish how the introgressed populations will evolve over time.

Despite their close proximity and connection via the swamp, low gene flow was observed between the three hot spring populations. It is therefore necessary to conduct further studies using more powerful approaches involving high number of loci i.e SNP. This will enable accurate evaluation of gene flow between the three hot spring populations.

Our morphological study clearly established that both genetic and environmental factors play a major role in influencing morphology. It will be interesting to carry out further morphological studies on fish raised under controlled conditions in order to establish the role

of various environmental factors such as temperature, pH and salinity in influencing morphology.

Understanding of sex determination in the hot spring populations can be achieved by pursuing the SNP analysis studies in the inverted region of LG3 notably on the *dmrt4* gene implicated in fish development, which is found in this LG3 region. A first SNP analysis was performed in this thesis in scaffold 88 corresponding to part of the inverted region of LG3. Although some SNPs had allelic variations, it is still early to know if they are associated to the populations and to the phenotypic sex because not all the wild populations could be analyzed, and the number of individuals was reduced in some populations.

Finally, if the sex-reversed individuals result from a temperature effect on sex differentiation, it will then suggest that 2 mechanisms of sex determination are still present in populations living under high thermal regimes. The persistence of these two mechanisms could be associated to an adaptive role of thermosensitivity in these populations. In order to answer this question, it will be necessary to estimate the frequency of sex reversed individuals in a population, and then, to compare the fitness of sex reversed *versus* classic individuals. In addition, Bacterial Artificial Chromosomes (BAC) markers can be used to further confirm our sex-reversed results in Kenyan populations.

2. References

- Agnèse, J. F., Adepo-Gourène, B., Abban, E. K., Fermon, E., 1997. Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity*, 79: 88-96.
- Ansah, Y. B., Frimpong, E. A. and Hallerman, E. M., 2014. Genetically-Improved Tilapia Strains in Africa: Potential Benefits and Negative Impacts. *Sustainability*, 6:3697-3721.
- Ashley, G. M., Goman, M., Hover, V. C., Owen, R. B., Renaut, R. W., and Muasya, A. M. 2002. Artesian blister wetlands, a perennial water resource in the semi-arid Rift-Valley of East Africa, *Wetlands* 22(4): 686-695.
- Ashley, G. M., Maitima-Mworia, J., Muasya, A. M., Owen, R. B., Driese, S. G., Hover, V. C., Renaut, R. W., Goman, M. F., Mathai, S., and Blatt, S. H., 2004. Sedimentation and recent history of a freshwater wetland in a semi-arid environment: Lobo Swamp, Kenya, East Africa. *Sedimentology* 51, 1301–1321.
- Baroiller, J. F., D'Cotta, H., and Saillant, E., 2009b. Environmental effects on fish sex determination and differentiation. *Sex Dev.* 3(2-3):118-35.
- Baroiller, J. F., and D'Cotta, H., 2001. Environment and sex determination in farmed fish. *Comp. Biochem. Physiol. C: Toxicol. Pharmacol.* 130: 399-409.
- Baroiller, J. F., Jalabert, B., 1989. Contribution of research in reproductive physiology to the culture of tilapias. *Aquat. Living Resour.* 2: 105–116.
- Bezault, E., Balaesque, P., Toguyeni, A., Fermon, E., Araki, H., Baroiller, J. F. and Rognon, X. 2011. Spatial and temporal variation in population structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *Genetics*, 12: 102.
- D'Cotta H, Coutanceau JP, Pepey E, Baaziz SB, Coriton O, Bonillo C, Rodier-Goud M, Soler L, D'Hont A, Kocher T, Ozouf-Costaz C, Baroiller JF, 2014. Fishing for sex genes by chromosome microdissection in two tilapia species with different sex determining systems. Meeting Abstract. *Chromosome Research*, 22(3): 433
- Deocampo, D. M., 2002. Sedimentary processes and lithofacies in lake-margin groundwater-fed wetlands in East Africa. In: Ashley, G.M., Renaut, R.W. editors. *Sedimentation in Continental Rifts*, vol. 73. Special Publication-SEPM, Tulsa, OK, pp. 295–308.
- Eknath, A. E., and Acosta, B., 1998. Genetic improvement of farmed tilapia project. Final report (1988-1997). ICLARM, Manila.
- Eknath, A. E., 1995. Managing aquatic genetic resources. Management example 4. The Nile tilapia. In: Thorpe, J., Gall, G., Lannan, J., Nash, C. (eds) *Conservation of fish and shell fish resources: Managing diversity*, Academic Press, London. pp. 176-194.
- Harper, D.M., Brooks Childress, M., Harper, M.M., Boar, R.R., Hickley, P., Mills, S.C., Otieno, N., Drane, T., Vareschi, E., Nasirwa, O., Mwatha, W.E., Darlington, J.P.E.C., Escuté-

Gasulla, X., 2003. Aquatic biodiversity and saline lakes: Lake Bogoria National Reserve, Kenya. *Hydrobiologia* 500, 259–276.

Kerschbaumer, M. K., and Sturmbauer, C., 2011. The utility of geometric morphometrics to elucidate pathways of cichlids of cichlid evolution. *Int. J. Evol. Biol.* Article ID290245, pp 8.
Meyer, A., Kocher, T. D., Basasibwaki, P. and Wilson, A. C., 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, 347: 550-553.

Ndiwa, C. T., Nyngi, D., and Agnèse, J. F., 2014. An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Lobo drainage, Kenya. *PloS ONE* 9(9):e106972.

Nyngi D. W. and Agneses J. F. 2012. *Phylogeography of the Nile Tilapia in Africa*. Lambert Academic Publishers.

Nyngi, D. W. and Agnèse, J-F. 2007. Recent introgressive hybridization revealed by exclusive mtDNA transfer from *Oreochromis leucostictus* (Trewavas, 1933) to *Oreochromis niloticus* (Linnaeus, 1758) in Lake Baringo, Kenya. *J. Fish BioLake* 70 (Supplement A), 148–154.

Nyngi, D. W., Agnèse, J-F., 2011. Tilapia in eastern Africa – a friend and foe. Species in the Spotlight. Darwall, W.R.T., Smith, K.G., Allen, D.J., Holland, R.A, Harrison, I.J., and Brooks, E.G.E. editors. *The Diversity of Life in African Freshwaters: Under Water, Under Threat*. An analysis of the status and distribution of freshwater species throughout mainland Africa. Cambridge, United Kingdom and Gland, Switzerland: IUCN. Pp. 76-78.

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

Owen, R. B., Crossley, R., Johnson, T. C., Tweddle, D., Kornfield, I., Davison, S. Eccles, D. H. and Engstrom, D. E. 1990. Major low levels of Lake Malawi and their implication for speciation rates in cichlid fishes. *Proc. R. Lond. B.* 240:519-553.

Owen, R. B., Renaut, R. W., Hover, V. C., Ashley, G. M., Muasya, A. M., 2004. Swamps, springs, and diatoms: wetlands regions of the semi-arid Bogoria–Baringo Rift, Kenya. *Hydrobiologia* 518, 59–78.

Ozouf-Costaz C., Coriton O., D'Cotta H., Cnaani A., Coutanceau J.P., Bonillo C., Peppey E., Kocher T., Baroiller J. F. 2008. Chromosomal localisation of sex-specific BACs and of selected cDNAs identified by microdissection in two tilapia species, *Oreochromis niloticus* and *Oreochromis aureus*. *Chromosome Research*, 16 (7): 1060-1060.

Price, T. D., Qvarnstrom, A., and Irwin, D. E., 2003. The role of phenotypic plasticity in driving genetic evolution. *Proc. R. Soc. Lond. B* 270: 1433-1440.

Rogon, X., Andriamanga, M., McAndrew, B., and Guyomard, R. 1996. Allozyme variation in the natural and cultured populations in two tilapia species: *Oreochromis niloticus* and *Tilapia zillii*. *Heredity*, 76: 640-650.

Sage, R. D., Loiselle, P. V., Basasibwaki, P., and Wilson, A. C., 1984. Molecular versus morphological change among cichlid fishes of Lake Victoria. In Echelle, A. A. and Kornfield, I (Eds). Evolution of fish species flocks. University of Maine at Orono press, Orono, Maine.

Salzburger, W., Meyer, A., Baric, S., Verheyen, E., and Sturmbauer, C., 2002. Phylogeny of the Lake Tanganyika Cichlid species flock and its relationship to the central and East African Haplochromine cichlid fish faunas. Syst. Biol. 51(1): 113-135.

Stiassny, M. L. J., and Meyer, A., 1999. Cichlids of the Rift-Lakes. Sci. Am. 280: 44-49.
Trewavas, E (1983) Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*, British Museum (Natural History), London. pp 583.

Résumé détaillé de la thèse

Le tilapia du Nil (*Oreochromis niloticus*, Linné, 1758) fait partie de la famille des Cichlidae qui regroupe environ 3000 espèces principalement distribuées dans les régions tropicales et sub-tropicales du monde. Cette espèce, originaire d'Afrique, a été introduite dans plus de 100 pays à travers le monde pour y être élevée. Depuis le premier travail de Trewavas (1983), un certain nombre d'études ont tenté de documenter la diversité des populations naturelles d'*O. niloticus*, en particulier leur caractérisation génétique (Rognon *et al.*, 1996; Agnèse *et al.*, 1997; Nyingi *et al.* 2009; Bezault *et al.*, 2011). Cependant, la connaissance de leur diversité génétique reste encore insuffisante. En outre, certaines populations naturelles sont mal connues comme celle du lac Tanganyika par exemple, ou peut être même encore inconnues comme en témoigne la récente découverte d'une population naturelle au Kenya (Nyingi *et al.*, 2009).

La plupart, sinon toutes, des ressources naturelles génétiques du tilapia du Nil se trouvent en Afrique et jouent un rôle majeur dans l'amélioration des souches cultivées dans d'autres parties du monde. Par exemple, le tilapia GIFT (Genetically Improved Farmed Tilapia) souche génétiquement améliorée a été développée à partir de populations fondatrices du tilapia du Nil provenant de populations sauvages et de souches cultivées (Eknath, 1995; Eknath et Acosta, 1998). Notamment, des poissons en provenance d'Egypte, du Ghana, du Kenya et du Sénégal ont été utilisés, ainsi que des souches d'élevage provenant d'Israël, de Singapour, de Taiwan et de Thaïlande (Gupta et Acosta, 2004).

Même si le tilapia du Nil est une espèce envahissante dans de nombreux bassins fluviaux et a parfois été accusé d'être à l'origine de la disparition de certaines espèces de poissons (Ogutu-Ohwayo, 1990), ses populations naturelles sont aussi menacées (Nyingi *et al.*, 2007; Nyingi et Agnèse 2011). La principale menace est certainement l'hybridation avec d'autres espèces ou populations. Cette situation devient de plus en plus critique, surtout après la récente introduction de souches synthétiques comme celle du tilapia GIFT en Afrique (Ansah *et al.*, 2014).

Trewavas (1983) fut la première à étudier les populations naturelles du tilapia du Nil. A partir de l'étude de leurs caractéristiques morphologiques, elle a décrit sept sous-espèces: *O. n. niloticus* d'Afrique de l'Ouest et du Nil, *O. n. baringoensis* du lac Baringo, *O. n. sugutae* de la rivière Suguta (Kenya), *O. n. eduardianus* des lacs Edouard, Albert, George et Tanganyika,

O. n. vulcani du lac Turkana, *O. n. cancellatus* de la rivière Awash et du lac Tana, et *O. n. filoa* de sources d'eau chaude dans le système de la rivière Awash.

Ce travail a servi et sert encore de nos jours de référence sur *O. niloticus* et les autres espèces de tilapia. Cependant, une analyse minutieuse des données fournies sur la différenciation des sous-espèces du tilapia du Nil révèle que Trewavas (1983) a utilisé à plusieurs reprises des arguments bien faibles et très subjectifs. Parfois elle a utilisé des caractères comme le nombre de rayons des nageoires comme caractère de différenciation alors que l'éventail des valeurs observées pour ce caractère entraîne des chevauchements entre les sous-espèces (voir par exemple le tableau 35 pages 144-145). Ceci explique sans doute pourquoi l'auteur n'a pas fourni une clé d'identification des sous-espèces.

Plus problématique est l'absence de toute considération ou de définition de ce qui est ou ce qui devrait être une sous-espèce. Sont-elles équivalentes à des unités évolutives (Ryder, 1986; Waples, 1991) ou représentent-elles simplement des populations morphologiquement différenciées sans signification évolutive? Il est difficile de comprendre quel était le concept de "sous-espèce" adopté par cet auteur. Par exemple, *O. n. cancellatus* et *O. n. filoa* sont présents (paraâtrique) dans le même bassin, celui de la rivière Awash. Cette situation laisse plutôt penser que l'on a affaire soit à deux espèces différentes soit à deux éco-phénotypes différents de la même espèce. La même situation se retrouve avec *O. n. eduardianus* dans le lac Edouard, qui est en contact avec *O. n. niloticus* dans le Nil. A la lecture de son ouvrage, il apparaît que son auteur a observé une certaine quantité de variabilité entre les populations et a trouvé "pratique de leur fournir des noms, pour préserver leur identité pour sévir aux futurs expérimentateurs ou gestionnaires des pêches" (citation de la page 141, dernière phrase).

Depuis Trewavas (1983) peut d'avancées on été faites sur la différenciation phénotypiques des populations sauvages du tilapia du Nil, ce qui nous a amené à réaliser des études sur la différenciation morphologique des populations à l'aide de techniques à la fois nouvelles et plus puissantes comme les analyses géomorphométriques (Bookstein, 1991). Avec ces analyses les variations de formes entre les différents taxons ou même les différentes populations peuvent efficacement être appréhendées.

La technique a été utilisée avec succès par exemple par Firmat *et al.* (2011, 2012) pour différencier 15 populations de tilapia du Mozambique (*Oreochromis mossambicus*) de son aire de répartition naturelle au Mozambique, des populations sauvages de Nouvelle-Calédonie et de la Guadeloupe après leur acclimatation dans ces nouveaux environnements. L'analyse

géomorphométrique est également capable de différencier les populations même étroitement liés comme (Kerschbaumer et Sturmbauer, 2011) l'ont montré en étudiant des populations sauvages de *Tropheus moorii*.

Bien après Trewavas (1983), l'utilisation de techniques génétiques (étude des allozymes, des séquences d'ADN mitochondrial, des microsatellites) a permis une description plus détaillée des différentes populations de tilapia du Nil. Par exemple, Agnèse *et al.* (1997) ont étudié 17 populations naturelles représentant les sept sous-espèces décrites par Trewavas (1983), et les ont regroupées en trois grands groupes : 1) populations d'Afrique occidentale, 2) les populations éthiopiennes de la vallée du Rift, et 3) les populations du Nil et de la vallée du Rift du Kenya.

D'autres études génétiques ont été menées comme celles de Rognon *et al.* (1996, 2003) Hassanien et Gilbey (2005), Nyingi et Agnèse (2007), Nyingi *et al.* (2009), Angienda *et al.* (2011) et Bezault *et al.* (2011). Toutes ces études constituent une précieuse documentation des ressources génétiques du tilapia du Nil mais restent malgré tout encore incomplètes.

Récemment par exemple, une nouvelle population naturelle d'*O. niloticus* a été découverte dans une source chaude du marais de Loboï (Nyingi *et al.*, 2009).

Cette population méritait amplement qu'on s'y intéresse par le double fait qu'elle représente une ressource génétique originale et hautement intéressante pour l'élevage d'*O. niloticus* et d'autre part parce qu'elle est en danger.

La productivité des élevages d'*O. niloticus* repose en grande partie sur la possibilité de pouvoir réaliser des cultures mono-sexe mâles. Plusieurs méthodes existent pour y parvenir : sexage manuel, hybridation, traitement hormonal, manipulations génétique. Aucune ne donne pleinement satisfaction. Par ailleurs On connaît les effets masculinisants des températures élevées (Baroiller *et al.*, 2009a) et le fait que les populations varient de ce point de vue, par des thermosensibilité différentes et donc génétiquement déterminées (Bezault *et al.*, 2007). La population des sources chaudes du marais de Loboï est très vraisemblablement adaptée aux hautes températures et par conséquent offre l'opportunité d'étudier ce phénomène de thermosensibilisation sur une population naturelle.

Par ailleurs, cette population est en danger. Comme la population du lac Baringo (toute proche), elle peut craindre l'introduction d'espèces étrangères au sein de son bassin hydrographique en raison des activités aquacoles dans la région. En outre, la taille du marais

s'est considérablement réduite ces dernières années (Fuch, 19950 ; Nyingi *et al.*, 2009) mettant en danger l'existence même de la rivière.

Pour ces raisons, nous avons mis en place deux types d'études, d'une part l'étude de la caractérisation génétique de cette population du marais Loboï, et d'autre part l'étude de la variabilité génétique au niveau des SNP situés dans des régions génomiques connues pour avoir des liens avec la différenciation du sexe chez *O. niloticus* (Baroiller *et al.*, 2009b).

Notre travail se compose donc de trois parties, d'une part l'analyse géomorphologique de certaines populations naturelles de *O. niloticus* afin de mettre en évidence leur diversité morphologique, leur plasticité phénotypique et l'effet environnemental sur les phénotypes, d'autre part, nous avons caractérisé génétiquement ces populations afin d'évaluer leur originalité et leur intégrité génétique. Enfin nous avons cherché à mettre en évidence des phénomènes adaptatifs qui concerne le rôle de la température dans la détermination du sexe.

Variations morphologiques au sein des populations naturelles de tilapia du Nil vivants dans des conditions environnementales extrêmes dans la vallée du Rift du Kenya.

Les échantillons provenant de trois populations de sources chaudes du marais de Loboï, Chelaba, Lake Bogoria et Turtle Springs, de trois autres localités, les lacs Baringo, Turkana et Crocodile, et du lac Victoria, ont été étudiés.

Chaque spécimen a été scanné à l'aide d'un scanner à plat (Herler *et al.*, 2007) les images ont été ensuite importées dans le programme tpsUtil la version 1.53 (Rohlf, 2012). La forme du corps des poisson a été appréhendée à l'aide de la méthode des repères morpho-géométrique (Rohlf et Marcus, 1993). Un total de 14 points (repères) ont été choisis en fonction de leur capacité à capter forme générale du corps, et leurs coordonnées (x, y) numérisées sur le côté gauche de chaque poisson en utilisant le programme tpsDig la version 2.16 (Rohlf, 2010).

Un total de 237 spécimens provenant de sept populations naturelles du tilapia du Nil (Bogoria hôtel spring, 41 individus; Chelaba spring, 30 ; Turtle spring, 24; lac Baringo, 43; Lac Victoria, 33 ; Lac Crocodile, 28 et lac Turkana, 38) ont été analysés. Leurs dimensions standard ont varié de 40 mm à 207 mm.

L'analyse des données a été réalisée à l'aide de logiciels de logiciels R (Claude, 2008; 2013). Afin d'établir les variations morphologiques entre les différentes populations de poissons, nous avons utilisé l'analyse en composantes principales (ACP) des coordonnées superposées des points de repère. Les variations ont été visualisées en traçant les représentation des coordonnées multivariées des échantillons sur les deux premières composantes principales (PC) axes.

Sans surprise, l'analyse morphologique a révélé que les échantillons étudiés ont été tout d'abord différenciés par leurs tailles. Les poissons des lacs Turkana et Victoria ont les tailles moyennes les plus importantes $122,8 \pm 4,38$ mm (moyenne \pm écart-type) et $163,3 \pm 4,40$ mm, respectivement. Les plus petits spécimens provenaient de Turtle Spring et du lac Crocodile ($76,6 \pm 6,11$ mm et $70,1 \pm 1,72$ mm, respectivement). Les différences ont également été observées en termes de variabilité des tailles. Certains échantillons avaient une plus grande gamme de taille (lac Victoria) que d'autres (lac Crocodile). Ces différences observées peuvent représenter des artefacts en partie dus à l'échantillonnage et peuvent ne pas refléter complètement les différences réelles de la taille des poissons dans les différentes localités en dépit du fait que l'on sait que les lacs Turkana et Victoria accueillent de grands spécimens de *O. niloticus* (Trewavas, 1983) en raison de leurs grandes tailles et de leurs plus grandes profondeurs, par rapport aux autres biotopes étudiés au cours de ce travail.

Dans notre étude, nous avons observé différents modèles allométriques de croissance dans chacune des sept populations étudiées. Compte tenu du fait que la plupart des localités diffèrent grandement de par leurs conditions environnementales, les allométries observées peuvent être attribuées en partie aux différences environnementales (température et salinité), qui auraient agi sur certains stades de développement. Il est encore trop tôt à ce stade pour dire si les différences observées sont d'ordre adaptatives ou simplement dues à la plasticité, mais ces deux hypothèses pourraient être testées en faisant par exemple des transplantations réciproques expérimentales de poissons.

Les analyses ont montré un grand nombre de différences de forme parmi lesquelles les variations de la tête, du pédoncule caudal et de la nageoire anale représentaient 23%. Ces variations ont permis de discriminer les spécimens du lac Victoria de ceux des autres populations (lacs Baringo, Turkana et les sources d'eau chaude). Au sein de ces autres populations, les poissons du lac Turkana ont été clairement différenciés et occupent une place intermédiaire entre la population du lac Victoria et les autres. Ces résultats sont en accord

avec les relations génétiques observées par Agnèse *et al.* (1997). En effet, ces auteurs ont montrés que les poissons du lac lac Baringo et ceux du lac Turkana sont génétiquement proches. Par la suite, Nyingi *et al.* (2009) et Ndiwa *et al.* (2014) ont observé que les poissons des sources chaudes sont encore plus étroitement liés aux poissons du lac Baringo. Si l'on suppose que les poissons du lac Crocodile doivent avoir une base génétique identique aux poissons du lac Turkana, et que les poissons du lac Victoria qui ont été introduites à partir de poissons originaires du lac Albert (Fryer et Iles, 1972, Trewavas, 1983) devraient être moins liés génétiquement à tous les autres poissons étudiés, la différenciation morphologique des populations est congruente avec leur différenciation génétique

Lorsque l'on regarde de plus près les populations génétiquement liées, on peut tout de même observer une certaine quantité de différenciation morphologique. Le premier cas concerne des spécimens des trois sources d'eau chaude. Les échantillons de Bogoria et Chelaba ou Bogoria et Turtle spring sont morphologiquement différenciées. Par exemple, les individus de Chelaba et Bogoria Springs varient en ce qui concerne les proportions longueur du corps et profondeur du corps. Les individus des populations Bogoria / Chelaba et Turtle spring se différencie au niveau de la longueur de la tête et la hauteur du corps.

Si l'on considère que la base génétique de ces trois populations est la même (Ndiwa *et al.*, 2014), ces différenciations morphologiques observées pourraient être attribuées à la plasticité phénotypique de l'espèce. Les caractéristiques environnementales très probables de ces sources chaudes sont légèrement différentes. Bogoria spring est la plus grande des trois sources d'eau chaude et coule pendant quelques centaines de mètres avant d'entrer dans le marais. Les deux autres ont un débit plus faible et entrent dans le marais après seulement quelques mètres. Néanmoins, seuls quelques gènes peuvent agir fortement sur la différenciation morphologique, on ne peut exclure que les différences morphologiques observées entre les trois populations thermales soient tout de même dues à des différences génétiques.

Ndiwa *et al.* (2014) ont observé que les flux de gènes entre ces populations est limité par le marais, qui est couverte d'une végétation dense de *Typha domingensis* et *Cyperus papyrus*. De plus, Ndiwa *et al.* (2014) ont également observé que ces poissons sont légèrement introgressés par des gènes de *O. leucostictus* échappés de fermes aquacoles toutes proches. L'action de gènes étrangers issus de ces introgressions agissant sur la différenciation morphologique ne peut être exclue.

Le deuxième cas de populations génétiquement liés dans notre étude porte sur les poissons des lacs Turkana et Crocodile. Il y avait aussi une différenciation morphologique claire entre ces deux groupes de poissons qui diffèrent par la hauteur du corps, la proportion de la tête et la longueur de la caudale.

Ces deux populations sont génétiquement très proches (résultats non publiés) puisque la population du lac Crocodile tire son origine, il ya environ un siècle, de celle du lac Turkana. Très probablement, les différences observées peuvent être attribuées à de la plasticité phénotypique. Ces lacs offrent deux environnements très différents pour les poissons. Le lac Turkana est un grand lac alcalin (pH 9,3) et légèrement salé (conductivité 3420 $\mu\text{S} / \text{cm}$) qui hébergement de nombreuses espèces de poissons (48 d'après Hopson *et al.*, 1982). En comparaison, le lac Crocodile est un petit lac (environ 0,32 km²), avec une alcalinité (pH 10) et une salinité (conductivité 10 590 uS / cm) plus élevées. Seules trois espèces de poissons sont présents, deux cichlidés (*Aplochromis sp* et *O. niloticus*) et un Clariidae (*Clarias gariepinus*). La réduction observée de la forme du corps peut être une réponse adaptative aux niveaux élevés de salinité dans l'habitat. Cette hypothèse soutient les conclusions de Collyer *et al.* (2005; 2007) qui ont noté une réduction de taille chez les « pupfish » (*Cyprinodon tularosa*) en fonction de l'augmentation du niveau de salinité.

En conclusion, notre étude tend à montrer qu'à la fois des facteurs génétiques et environnementaux jouent un rôle important en influençant le développement morphologique de *O. niloticus*. D'autres études, en laboratoire notamment, seront nécessaires pour confirmer l'importance de ces facteurs environnementaux. Ceci pourra être obtenu par reproduction de fratries génétiquement identiques dans des conditions environnementales différentes, d'abord dans des conditions de salinités différentes, puis dans des conditions de températures différentes.

Des populations naturelles, ressources génétiques importantes d'*Oreochromis niloticus* (Linnaeus, 1758), mises en danger par des activités aquacoles dans le bassin de la Loboï au Kenya

Sept populations ont été étudiées et comparées : trois populations des sources d'eau chaude du bassin de la rivière Loboï (Turtle, Chelaba et lac Bogoria Hôtel springs), et celles des lacs Baringo, Turkana et de la rivière Suguta auxquelles nous avons ajouté une population de *O. leucostictus* du lac Naivasha.

Les travaux ont porté sur les séquences d'une portion de l'ADN mitochondrial (Dloop) et l'analyse du polymorphisme à 16 microsatellites.

En premier lieu, l'analyse de l'ADN mitochondrial a confirmé ce que Nyingi *et al.* (2007) avaient montré, la population d'*O. niloticus* du marais de Loboï est bien une population naturelle de Tilapia du Nil, génétiquement proche de celle du lac Baringo. L'étude des deux nouvelles populations des sources chaudes de Turtle et Chelaba spring, confirme donc ces résultats.

En revanche, il apparaît, contrairement aux résultats obtenus en 2007 et 2009 (Nyingi *et al.*, 2009) que les populations du marais de Loboï sont, comme l'était déjà la population du lac Baringo, introgressées par de l'ADN mitochondrial de *O. leucostictus*.

Il est probable que cette introgression soit très récente et date de l'entrée en vigueur d'un plan de développement de l'aquaculture dans le bassin de la Loboï en 2009.

L'analyse des 16 locus microsatellites a elle aussi dans un premier temps confirmé les résultats de Nyingi *et al.* 2007. Les populations du marais de la Loboï sont bien génétiquement originales et proches de celle du lac Baringo. Une analyse sommaire par Analyse Factorielle des Correspondances sur la base des génotypes observés, comme celle qui avait été faite par Nyingi *et al.* en 2007, pourrait laisser penser que l'introgression mitochondriale ne s'est pas accompagnée d'une introgression de gènes nucléaires. Une analyse plus attentive, utilisant le logiciel Structure, révèle sans ambiguïté la présence d'allèles provenant d'*O. leucostictus* dans la population du lac Baringo et dans les trois sources chaudes du marais de Loboï.

Par contraste, et comme indice supplémentaire de la réalité de cette introgression, aucun échange de gène n'est détecté avec les populations de la rivière Suguta et du lac Turkana. Une analyse avec le programme NEWHYBRIDS suggère que les 7 individus introgressés sont des F2 issus de backcross avec *O. niloticus*.

Même si l'étendue des biotopes de sources chaudes sont de petites tailles, et par conséquent aussi la taille de leurs populations efficaces, la diversité de l'ADNmt est élevée dans les sources chaudes avec par exemple 9 haplotypes différents chez 36 individus (en dehors des individus introgressés).

Cette observation semble paradoxale si l'on considère que les sources chaudes se jettent dans un petit marais d'environ 1,5 km². Le marais lui-même ne constitue pas un biotope favorable à la survie des poissons car il est entièrement couvert de macrophytes dont les plus dominantes sont *Typha domingensis* et *Cyperus papyrus*. Pour expliquer cette diversité mitochondriale, il faut considérer que ces sources chaudes sont approvisionnées par des failles et fractures qui permettent à l'eau des aquifères profonds de remonter à la surface. Ces sources sont donc insensibles aux changements climatiques et sans doute stables sur de très longues périodes. Par ailleurs, il est remarquable d'observer que, même si ces populations se trouvent proches les unes des autres de quelques dizaines à quelques centaines de mètres, il existe un certain degré de différenciation génétique qui est proportionnel à la distance géographique. Pour expliquer cette structuration génétique, il faut admettre que le marais crée un obstacle à la libre circulation des poissons d'une source à l'autre. En effet, la concentration en oxygène dissous au niveau du marais est plus faible que dans les sources, principalement en raison de la décomposition de la matière organique. Cette situation a été déjà décrite au niveau du lac Victoria (Deocampo, 2002 ; Ashley *et al.* 2002). Nos propres observations sont congruentes avec l'isolement observé entre d'une part des populations lacustres de *O. esculentus* vivant dans les lacs satellites du lac Victoria, et d'autre part des populations d'*O. niloticus* séparés par des zones humides constituées par des marais à papyrus (Nyingi *et al.* 2011). Toutefois, si le marais représente une barrière infranchissable pour les poissons il devient alors difficile d'expliquer comment *O. leucostictus* a pu introgresser *O. niloticus* en traversant le marais à partir des étangs d'élevage en connexion avec la rivière Loboï. Il est possible, bien que peu probable, que quelqu'un ait directement introduit des poissons dans chaque source chaude. Même si cette hypothèse ne peut être rejetée simplement, il faut noter qu'il n'y a tout simplement aucune raison objective pour réaliser de telles translocations de poisson. Même si les modalités précises par lesquelles les poissons ont été introgressés ne sont pas encore connues, il n'en reste pas moins que la réalité de l'introgression a clairement été démontrée et que les activités aquacoles jouent très certainement un rôle dans ce processus.

Ces populations et les ressources génétiques qu'elles représentent doivent être préservées. La conservation et la gestion des Marais de la Loboï devraient être prioritaires dans un programme de gestion des ressources génétiques de *O. niloticus*.

Les SNP du gene *amh* lies aux sexes sont différents suivant les populations de Tilapia du Nil.

La détermination du sexe est la prédisposition initiale de son développement tandis que la différenciation sexuelle est le processus en aval qui va générer les testicules ou les ovaires. La détermination Génétique du sexe (DGS) qui est induite par un ou des gènes sur les chromosomes sexuels et la détermination environnementale du sexe (DES) impliquant une série de signaux environnementaux comme la température, ont évolué de nombreuses fois entre lignées de téléostéens (Baroiller *et al*, 2009a; Heule *et al*, 2014).

Chez le tilapia du Nil, la détermination du sexe et la différenciation ont fait l'objet de longues recherches en raison de dimorphisme de croissance et la nécessité de contrôler le sexe dans ce principales espèces aquacoles. La détermination du sexe chez cette espèce est un mécanisme complexe sous le contrôle de divers facteurs, qui peuvent être divisés en trois groupes: Les facteurs génétiques majeurs portés par des chromosomes sexuels, les facteurs génétiques mineurs et les facteurs environnementaux (Baroiller *et al.*, 2009b). Bien qu'il n'existe pas de chromosomes sexuels différenciés, les approches indirectes, telles que les tests de descendance ont démontré que *O. niloticus* dispose d'un système prédominant de détermination du sexe XX / XY alors que par exemple, son espèce sœur *O. aureus* a un système ZZ / ZW (Baroiller et Jalabert, 1989).

En utilisant la technique de RAD sequencing dans deux familles de la souche Manzalla, Palaioikostas *et al.* (2013) ont trouvé deux SNP qui expliquent 96% des phénotypes dans le groupe de liaison LG1. Cependant il a été montré dans d'autres souches des influences du groupe de liaison LG3 et des analyses QTL ont également indiquées des liaisons avec le sexe dans le groupe LG 23.

Plusieurs études ont montré que la température peut être un facteur dominant par rapport à la détermination génétique du sexe (Baroiller *et al.*, 1995, 2009a ; Bezault *et al.*, 2007). Dans la souche domestique Bouaké, une température > 32 °C provoque la masculinisation complète des femelles génétiques en mâles. La répartition naturelle du tilapia du Nil amène cette espèce à se trouver dans des eaux dont la température peut largement dépasser les 35°C

Au cours de notre étude, nous avons cherché à savoir dans quelle mesure, des populations sauvages de tilapia du Nil vivant dans différents régimes thermiques, y compris certaines sources chaudes, présentent des différences significatives au niveau des SNP situés dans des zones connues pour être liées à la différenciation sexuelle.

Pour cela, nous avons étudié des SNP qui se trouvent au niveau des gènes *wt1b* et *cyp19a1a* qui sont situés sur LG1 et au niveau du gène de l'AMH situé sur LG23.

Onze populations d'*O. niloticus* ont été étudiés, neuf sont des populations naturelles, tandis que deux, Bouaké (Côte-d'Ivoire) et Manzala (Egypte) sont des souches domestiques. Les populations naturelles du Kenya étaient au nombre de cinq, trois des sources chaudes (Lake Bogoria Hôtel, Chelaba et Turtle spring) et deux de milieu / habitats de température standard (lacs Baringo et Turkana). Nous avons également étudié trois populations de la Vallée du Rift éthiopien, un d'une source d'eau chaude (lac Metahara) et deux de lacs d'altitude froids (Hora et Koka).

Parallèlement aux SNP, nous avons amplifiés par PCR une portion du génome qui permet de discriminer les individus XX, XY et individus / YY.

Le résultat le plus intéressant est que, en plus de présenter le plus grand nombre de SNP, le gène *amh* présente également SNP avec une forte relation avec le sexe phénotypique. Même si ces associations ont été observées dans la plupart des populations étudiées, des associations plus fortes ont été observées sur les populations kenyanes. Au sein de ces populations (Bogoria, Chelaba, Turtle, Baringo et Turkana), 11 SNP ont été observés sur le gène de l'*amh* qui avaient une forte association au sexe phénotypique. L'analyse de ces SNP indique des écarts importants aux fréquences attendues sous l'hypothèse de Hardy-Weinberg dans certaines populations. Par exemple, trois populations (Bogoria, Chelaba et Turtle) ont montré des écarts importants sur *amh_2509*, tandis que quatre SNP (*amh_597*, *amh_611*, *amh_2586*, *amh_2807*) ont dévié de manière significative dans deux populations chacun. Fait intéressant, le SNP *amh_597* G> T provoqué un changement de l'acide aminé avec des femelles ayant la variante T tandis que les mâles ont la variante G de ce fait, entraînant le changement de l'acide glutamique par l'acide aspartique. Parmi les 11 SNP liés au sexe, deux impliquent des changements nucléotidiques entraînant des changements d'acides aminés qui pourraient modifier de manière significative la fonction ou l'activité de l'hormone *amh*. Les 9 autres SNP sont tous situés dans l'intron 6 leurs rôles et leurs importances restent largement à déterminer.

Nos résultats montrent que les SNP liés au sexe sont spécifiques à certaines populations et cette observation vient confirmer les précédents résultats (Baroiller *et al.*, 2009a; 2009b). Il en résulte que la détermination du sexe chez *O. niloticus* devrait être analysée en priorité au niveau populationnel.

References

- Agnèse, J. F., Adepo-Gourène, B., Abban, E. K., Fermon, E., 1997. Genetic differentiation among natural populations of the Nile tilapia *Oreochromis niloticus* (Teleostei, Cichlidae). *Heredity*, 79: 88-96.
- Angienda, P. O., Lee, H. J., Elmer, K. R., Abila, R., Waindi, E. N. and Meyer, A., 2010. Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. *Conserv genet*, 12: 243-255.
- Ansah, Y. B., Frimpong, E. A. and Hallerman, E. M., 2014. Genetically-Improved Tilapia Strains in Africa: Potential Benefits and Negative Impacts. *Sustainability*, 6:3697-3721.
- Ashley GM, Goman M, Hover VC, Owen RB, Renaut RW, et al., 2002. Artesian blister wetlands, a perennial water resource in the semi-arid Rift-Valley of East Africa, *Wetlands* 22(4): 686-695.
- Baroiller, J-F, Jalabert, B., 1989. Contribution of research in reproductive physiology to the culture of tilapias, *Aquat. Living Resour.* 2:105–116.
- Baroiller, J. F., Clota, F. and Geraz, E., 1995. Temperature sex determination in two tilapias species, *Oreochromis niloticus* and the red tilapia (Red Florida Strain): effect of high or low temperatures. In: Goetz, F. W., Thomas, P. (Eds), *Proceedings of the 5th International Symposium on the reproduction Physiology of Fish*. The University of Texas, Austin, Texas, USA, pp. 158-160. July 2-8, 1995.
- Baroiller, J. F., D'Cotta, H., and Saillant, E., 2009a. Environmental effects on fish sex determination and differentiation. *Sex Dev.* 3(2-3):118-35.
- Baroiller, J. F., D'Cotta, H., Bezault, E., Wessels, S., and Hoerstgen-Schwark, G. 2009b. Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology, Part A* 153: 30-38.
- Bezault, E., Clota, F., Derivaz, M., Chevassus, B. and Baroiller, J. F. 2007. Sex determination and temperature induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture*, 272S1: S3-S16.
- Bezault, E., Balaesque, P., Toguyeni, A., Fermon, E., Araki, H., Baroiller, J. F. and Rognon, X. 2011. Spatial and temporal variation in population structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *Genetics*, 12: 102.
- Bookstein, F. L. 1991. *Morphometric tools for landmark data: geometry and biology*. New York, Cambridge University Press.
- Claude J. 2008. *Morphometrics with R*. Springer, New York. Pp 316.
- Claude J. 2013. Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: three worked examples in R. *Hystrix, It. J. Mamm.* 24(1): 94-102.

- Collyer ML, Novak JM, Stockwell CA. 2005. Morphological divergence of native and recently established populations of white sand pupfish (*Cyprinodon Tularosa*). *Copeia*, 1-11.
- Collyer ML, Stockwell CA, Adams DC, Reiser MH. 2007. Phenotypic plasticity and contemporary evolution in introduced populations: evidence from translocated populations of sand pupfish (*Cyprinodon Tularosa*). *Ecological Research* 22: 902-910.
- Deocampo DM, 2002. Sedimentary processes and lithofacies in lake-margin groundwater-fed wetlands in East Africa. In: Ashley, G.M., Renaut, R.W. editors. *Sedimentation in Continental Rifts*, vol. 73. Special Publication-SEPM, Tulsa, OK, pp. 295–308
- Eknath, A. E., and Acosta, B., 1998. Genetic improvement of farmed tilapia project. Final report (1988-1997). ICLARM, Manila.
- Eknath, A. E., 1995. Managing aquatic genetic resources. Management example 4. The Nile tilapia. In: Thorpe, J., Gall, G., Lannan, J., Nash, C. (eds) *Conservation of fish and shell fish resources: Managing diversity*, Academic Press, London. pp. 176-194.
- Firmat, C., 2011. Hybridation et goulots d'étranglement induits par l'activité humaine : génétique des populations, morphométrie et parasitologie appliquée au tilapia envahit et envahissant *Oreochromis mossambicus* (Teleostei, Cichlidae). Université de Bourgogne, France. Pp 207.
- Firmat, C., Schliewen, U. L., Losseau, M. and Alibert, P. 2012. Body shape differentiation at global and local geographic scale in the invasive cichlid *Oreochromis mossambicus*. *Biological Journal of Linnean Society*, 105: 369-381.
- Fryer G, Isles TD. 1972. *The Cichlid Fishes of the Great Lakes of Africa*. Edinburgh: Oliver and Boyd.
- Fuch VE, 1950. Pleistocene events in the Baringo Basin, Kenya Colony. *Geol. Mag.Lond.* 87: 149-174
- Gupta, M. V. and Acosta, B. O., 2004. From drawing board to dining table: The success story of the GIFT project. *NAGA, World Fish Center Quarterly* Vol 27 No 3&4.
- Hassanien, H. A. and Gilbey, J., 2005. Genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*) revealed by DNA microsatellites. *Aquac. Res.* 36: 1450-1457.
- Herler J, Lipej L, Makovec T. 2007. A simple technique for simple imaging of live and preserved small fish specimens. *Cybium* 31: 39-44.
- Heule C, Salzburger W, Böhne A. 2014. Genetics of sexual development: an evolutionary playground for fish. *Genetics.* 196(3):579-91.
- Hopson AJ. 1982. *Lake Turkana; A report on the findings of the Lake Turkana Project 1972 1975*, Vol. 16. Overseas Development Administration, London. Pp 1614.
- Kerschbaumer, M. K. and Sturmbauer, C. 2011. The utility of geometric morphometrics to elucidate pathways of cichlids of cichlid evolution. *Int. J. Evol.Biol.* Article ID290245: 8pp.

- Ndiwa, C. T., Nyngi, D., and Agnèse, J. F., 2014. An important natural genetic resource of *Oreochromis niloticus* (Linnaeus, 1758) threatened by aquaculture activities in Lobo drainage, Kenya. *PloS ONE* 9(9):e106972.
- Nyingi, D., De-Vos, L., Aman, R. and Agnèse, J. F., 2009. Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) (Cichlidae; Teleostei) in the Lobo Swamp (Kenya). *Aquaculture* 297: 57-63.
- Ogutu-Ohwayo, R., 1990. The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of the introduced species, especially the Nile perch, *Lates niloticus* and the Nile tilapia,
- Oldfield, R. G. 2005. Genetic, abiotic and social influence on sex differentiation in cichlid fishes and the evolution of sequential hermaphroditism. *Fish Fisheries*, 6: 93-110.
- Palaiokostas CI, Bekaert M, Khan MG, Taggart JB, Gharbi K, McAndrew BJ, Penman DJ. 2013. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD sequencing. *PLoS One*. 8(7):e68389.
- Rognon, X., Andriamanga, M., McAndrew, B., and Guyomard, R. 1996. Allozyme variation in the natural and cultured populations in two tilapia species: *Oreochromis niloticus* and *Tilapia zillii*. *Heredity*, 76: 640-650.
- Rognon X, Guyomard R (2003) Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. *Molecular Ecology* 12: 435–445.
- Rohlf FJ, Marcus LF. 1993. A revolution in morphometrics. *Trends Ecol. Evol.* 8: 129–132.
- Rohlf FJ. 2010. *tpsDig, digitize landmarks and outlines, version 2.16*. Department of Ecology and Evolution, State University of New York at Stony Brook.
- Rohlf FJ. 2012. *tpsUtil, file utility program, version 1.53*. Department of Ecology and Evolution, State University of New York at Stony Brook.
- Ryder, O. A., 1986. Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*, 1: 9-10.
- Trewavas, E., 1983. Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*, British Museum (Natural History), London.
- Waples, R. S., 1991. Pacific salmon, *Onchorynchus* spp., and the definition of ‘species’ under the endangered species act. *Marine Fisheries Reviews*, 53: 11-22.