ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/320085609

Genetic diversity and population structure of the African catfish, Clarias gariepinus (Burchell, 1822) in Kenya: Implication for conservation and aquaculture

Article · September 2017
DOI: 10.26496/bjz.2017.9

CITATION
1

READS
262

11 authors, including:

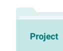
Robert Skilton
icipe – International Centre of Insect Physiology and Ecology
199 PUBLICATIONS **1,362** CITATIONS
SEE PROFILE

Emily Chemoiwa
University of Eldoret
8 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Otto George Dangasuk
University of Eldoret
20 PUBLICATIONS **200** CITATIONS
SEE PROFILE

Boaz Kaunda-Arara
University of Eldoret
39 PUBLICATIONS **703** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:


Northern Kenya Grevy's Zebra Project View project

Reinforcing academic capacities to Monkeypox outbreaks reponse : Diagnostic and origin of eruptive fevers in the Democratic Republic of the Congo (DRC) View project

All content following this page was uploaded by Erik Verheyen on 12 October 2017.

The user has requested enhancement of the downloaded file.

Belgian Journal of Zoology
www.belgianjournalzoology.be

This work is licensed under a Creative Commons Attribution 3.0 License.

ISSN 2295-0451

Research article

https://doi.org/10.26496/bjz.2017.9

Genetic diversity and population structure of the African catfish, *Clarias gariepinus* (Burchell, 1822) in Kenya: implication for conservation and aquaculture

James E. Barasa [1,*], Sinebongo Mdyogolo [2], Romulus Abila [3], Johannes Paul Grobler [2], Robert A. Skilton [4,†], Hesmari Bindeman [2], Moses Ndotono Njahira [4,†], Emily J. Chemoiwa [5], Otto G. Dangasuk [5], Boaz Kaunda-Arara [1] & Erik Verheyen [6,7]

[1] Department of Fisheries and Aquatic Sciences, University of Eldoret, P.O. Box 1125-30100, Eldoret, Kenya.
[2] Department of Genetics, University of the Free State, P.O. Box 399, Bloemfontein, 9300, South Africa.
[3] Department of Environmental Studies, Forestry and Agriculture, Maasai Mara University, P.O. Box 861-20500, Narok, Kenya.
[4] Biosciences eastern and central Africa-International Livestock Research Institute Hub (BecA-ILRI Hub), P.O. Box 30079, Nairobi 00100, Kenya.
[5] Department of Biological Sciences, University of Eldoret, P.O. Box 1125-30100, Eldoret, Kenya.
[6] Royal Belgian Institute of Natural Sciences, OD Taxonomy and Phylogeny-Vertebrates, Vautierstraat 29, 1000, Brussels, Belgium.
[7] University of Antwerp, Biology Department, Evolutionary Ecology, Campus Drie Eiken, 2610 Antwerpen, Belgium.
[†] Current address: International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772-00100, Nairobi, Kenya.

[*] Corresponding author: barkanoti@gmail.com, telephone: +254-725-842098

ABSTRACT. African catfish, *Clarias gariepinus*, is an important species in aquaculture and fisheries in Kenya. Mitochondrial D-loop control region was used to determine genetic variation and population structure in samples of *C. gariepinus* from 10 sites including five natural populations (Lakes Victoria (LVG), Kanyaboli (LKG), Turkana (LTA), Baringo (LBA) and Jipe (LJP), and five farms (Sangoro Aquaculture Center (SAN), Sagana Aquaculture Centre (SAG), University of Eldoret Fish Farm (UoE), Kibos Fish Farm (KIB), and Wakhungu Fish Farm (WKU)) in Kenya. Similarly, samples from eight localities (four natural populations: LVG/LKG, LTA, LBA, and four farmed: SAN, SAG, KIB, UoE) were genotyped using six microsatellite DNA loci. For the D-loop control region, samples from natural sites exhibited higher numbers of haplotypes and haplotype diversities compared to farmed samples, and 88.2% of haplotypes were private. All except LJP and LTA shared haplotypes, and the highest number of shared haplotypes (8) was detected in KIB. The 68 haplotypes we found in 268 individuals grouped into five phylogenetic clades: LVG/LKG, LTA, LBA, LJP and SAG. Haplotypes of farmed *C. gariepinus* mostly have haplotypes typical of LVG/LKG, and some shared haplotypes of the LBA population. Microsatellite analysis showed farmed samples have higher numbers of alleles than natural samples, but higher observed and expected heterozygosity levels were found in samples of natural

populations. Fifteen pair-wise comparisons had significantly different FST values. All samples were in Hardy-Weinberg equilibrium. Samples from the eight localities grouped into four genetic clusters (LVG/LKG, LTA, LBA and SAG), indicating genetically distinct populations, which should be considered for aquaculture and conservation.

KEYWORDS. Mitochondrial DNA, aquaculture, African catfish, microsatellites, population structure, conservation.

Barasa J.E., Mdyogolo S., Abila R., Grobler J.P., Skilton R.A., Bindeman H., Ndotono N.M., Chemoiwa E.J., Dangasuk O.G., Kaunda-Arara B. & Verheyen E. 2017. Genetic diversity and population structure of the African catfish, *Clarias gariepinus* (Burchell, 1822) in Kenya: implication for conservation and aquaculture. *Belgian Journal of Zoology* 147 (2): 105–127. https://doi.org/ 10.24696/bjz.2017.9

Introduction

The African catfish, *Clarias gariepinus* (Burchell, 1822), is an important species for aquaculture in the tropics. In Kenya, the species is second only to Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758), as a fin fish aquaculture species. There have been recent efforts to increase *C. gariepinus* aquaculture for higher food production (Rasowo *et al.* 2008; Chepkirui-Boit *et al.* 2011; Musa *et al.* 2012), and to provide bait for use in the Lake Victoria Nile perch, *Lates niloticus* (Linnaeus, 1758), long line fishery activities (Macharia *et al.* 2005; Barasa *et al.* 2014). In natural ecosystems, the species is an important predator in food webs (Goudswaard & Witte 1997). The species has high fecundity, grows fast, even at high stocking densities, utilizes a wide range of food items, and tolerates a wide range of environmental conditions (Bruton 1988). These traits have facilitated the translocation of *C. gariepinus* to many countries outside its natural range, especially to Thailand, Malaysia and the Netherlands, where they have been introduced for aquaculture (Nazia *et al.*, 2010), and new strains such as the Dutch strain developed (Cambray & van der Waal 2006; Anene & Tianxiang 2007).

A total of 3 million bait samples are required daily by fishermen on the Kenyan side of Lake Victoria, for use in the *L. niloticus* long line fishery activities (Fisheries Frame Survey 2006). Of the total daily supply of live bait to fishermen, however, only 41% are catfish fry (Mkumbo & Mlaponi 2007), with the rest being haplochromines, *Rastrineobola argentae* (Pellegrin, 1904), *Schilbe* (Linnaeus, 1758) and *Labeo* (Boulenger, 1901) species (Chitamwebwa *et al.* 2009). Of the five species exploited as bait for *L. niloticus*, artificial propagation techniques are in place only for *C. gariepinus*, with the rest being harvested exclusively from natural populations. However, of the total supply of *C. gariepinus* fry for use as bait, only 2% comes from fish farms and hatcheries (Chitamwebwa *et al.* 2009), with the rest being sourced from natural populations. Exploitation of natural populations of fish species in the Lake Victoria basin for bait, in addition to being fished for food by local communities, exacerbates pressure on fish stocks whose populations dramatically declined due to *L. niloticus* predation (Witte *et al.* 1992), and hampers the recovery of indigenous fish stocks in the lake (Mkumbo & Mlaponi 2007). The need to promote artificial propagation of *C. gariepinus* in hatcheries for bait as a strategy to conserve the indigenous fish species of the Lake Victoria basin, as well as increase income for farmers, has long been recognized (Kaufman & Ochumba 1993; Mkumbo & Mlaponi 2007; Chitamwebwa *et al.* 2009; Barasa *et al.* 2014).

Low survival of fry of *Clarias* is a major handicap in the expansion of aquaculture of the species in Africa (Sulem *et al.* 2006), because it causes a shortage of seeding stock for ponds and also for use as bait. Poor diets for larval catfish (Sulem *et al.* 2006; Nyina-Wamwiza *et al.* 2010; Chepkirui-Boit *et al.* 2011; Musa *et al.* 2012), cannibalism among batches of catfish siblings (Sulem *et al.* 2006), presence of predators in catfish fry nursery units (Sulem *et al.* 2006; Nyina-Wamwiza *et al.* 2010), and poor quality brood stock of mixed ancestry (Barasa *et al.* 2014) are reported to potentially reduce the survival of fry of *C. gariepinus*.

In artificial propagation of *C. gariepinus*, a male brooder is sacrificed for milt to fertilize the eggs (Hogendoorn & Vismans 1980). Although some hatcheries use a commercial synthetic ovaprim hormone to stimulate ovulation, most prefer using fish pituitary to obtain a natural hormone (De Graaf *et al.* 1995), because of the possible impact of ovaprim on the environment. Therefore, hatcheries frequently collect male brooders from natural populations to provide milt and the pituitary hormone (Barasa *et al.* 2014), and when brood stock size declines, additional brooders are collected from natural populations or other hatcheries. These sources of brood stock may be unrelated to the original source, and the aquatic habitats may be in different drainage basins. Similarly, female brooders are re-used for propagation, which potentially causes inbreeding.

Mixed ancestry of source populations poses risks to animal populations, including fish (McCleland & Naish 2007), via outbreeding depression, where offspring show reduced fitness. Out-breeding depression by inter-population hybridization introduces nonlocal alleles, creates phenotypes intermediate between the parents (Hatfield & Schluter 1999), and reduced fitness is observed in the F1 or F2 generation (Huff *et al.* 2011). These genetic factors, together with poor husbandry practices at hatcheries, may play a significant role in reducing the survival of fry and juveniles of *C. gariepinus*, especially in countries where strict regulations on inter-basin translocation of fish are absent or not enforced. Lower survival, growth, or fecundity may therefore be expressed in juveniles from brooders of *C. gariepinus* of mixed genetic origin, or those maintained under a poor feeding regime (Izquierdo *et al.* 2001).

On the other hand, out-breeding increases population genetic variation, which may be an important factor influencing fitness traits in some fish species (Knaepkens *et al.* 2002), and persistence of a fish species in the habitat (Policansky & Magnuson 1998). Research on genetic diversity of fish populations is applied in genetic improvement programmes, to develop a suitable base population (Ponzoni *et al.* 2009), a need already reported for *C. gariepinus* (Ponzoni & Nguyen 2008). Similarly, research on genetic diversity is used to monitor and manage stocks, and identify quantitative trait loci for commercially important traits in fish breeding (Kocher *et al.* 1998). Maintenance of genetic variation within and among populations is an important goal in management and conservation of biodiversity (Allendorf & Waples 1996).

Natural fish populations often outperform domesticated populations (Eknath *et al.* 1993) under culture, possibly due to inbreeding depression in domesticated populations (Kocher *et al.* 1998; Teichert-Coddington & Smitherman 1988), related to lower genetic variation (Popoola *et al.* 2014). Since out-breeding can also negatively affect population fitness, the use of catfish populations of higher genetic variation and purity as brood stock may help increase fry production at hatcheries through higher survival, and contribute to higher income and food availability among farmers, and also reduce exploitation pressure on natural populations of catfish in Kenya. To assess genetic variation and structure in natural and farmed populations of *C. gariepinus*, mitochondrial D-loop region and nuclear microsatellite DNA markers were used to determine genetic variation and population structure in both natural and farmed populations of *C. gariepinus* in Kenya. The control region has a high rate of base substitution and changes in the genome are accumulated here faster, making the region suitable for addressing questions of population genetic variation (Meyer 1994). Similarly, microsatellite markers are highly polymorphic, abundant, short sized and repetitive arrays and are conserved between species and families, and hence offer a higher resolution in inferring genetic variation among populations (Waldbieser *et al.* 2001; Cunningham & Meghen 2001).

Material and methods

Study sites and sample collection

We collected samples of *C. gariepinus* from 10 different sites in Kenya. The 10 sites comprised five lakes or natural habitats (Lakes Victoria (LVG), Kanyaboli (LKG), Turkana (LTA), Baringo (LBA) and Jipe (LJP)) and five farms (Sangoro Aquaculture Center (SAN), Sagana Aquaculture Center (SAG),

University of Eldoret Fish Farm (UoE), Wakhungu Fish Farm (WKU), and Kibos Fish Farm (KIB)) as shown in Table 1 and Fig.1. On the other hand, the nuclear microsatellite DNA marker analysis involved samples of *C. gariepinus* from four natural populations (LVG, LKG, LTA and LBA) and four farms (SAN, SAG, UoE and KIB). The sample sizes and sequence accession numbers (publicly available in GenBank) of the fish samples are shown in Table 1. Fish samples were collected from lakes by gill

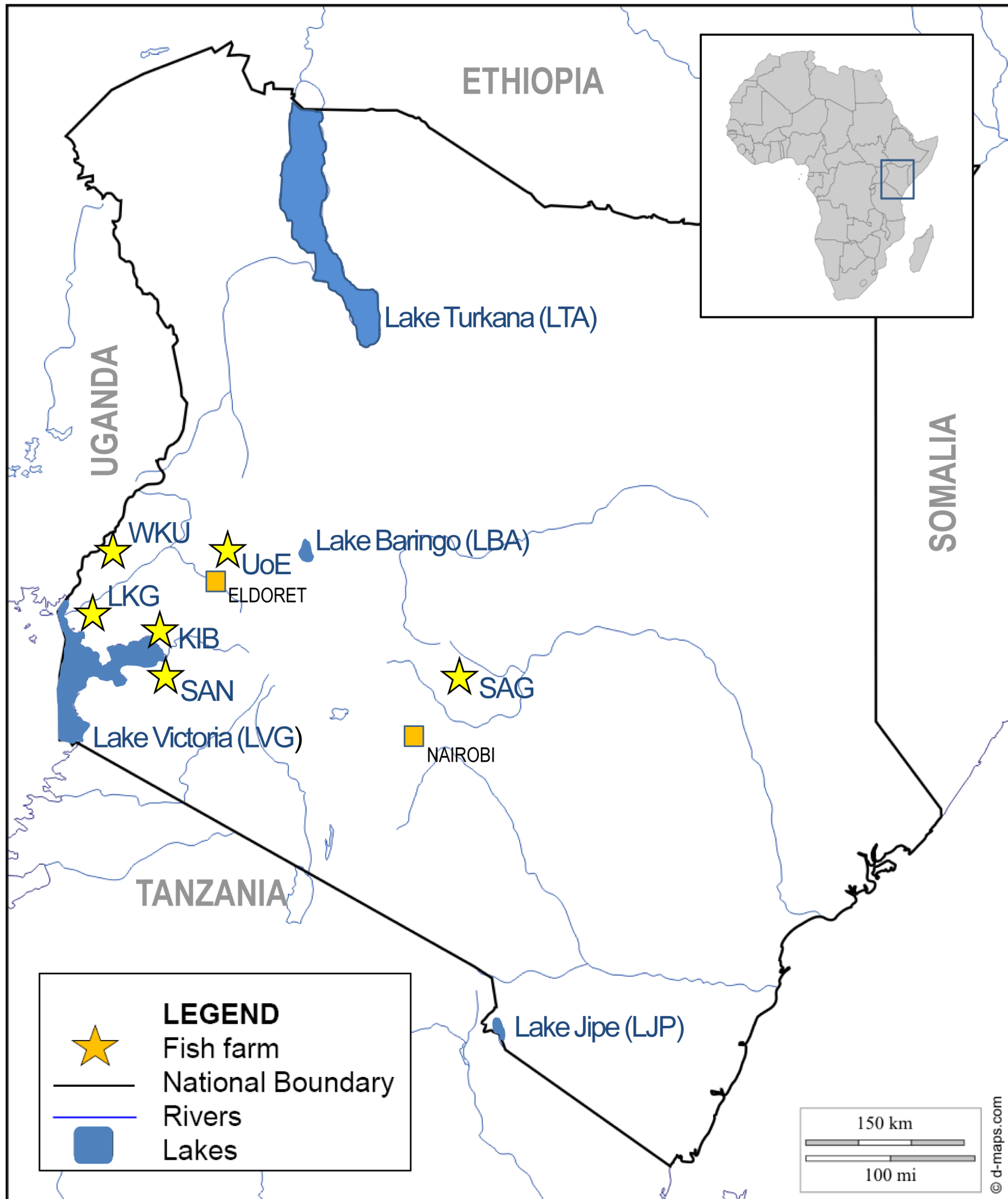


Figure 1 – Map of Kenya showing sampling sites for 10 populations of *C. gariepinus* (Burchell, 1822). Natural populations were collected from Lakes Victoria (LVG), Kanyaboli (LKG), Turkana (LTA), Baringo (LBA) and Jipe (LJP), represented on the map in light blue colour. Farmed populations were collected from fish farms SAN, SAG, UoE, KIB and WKU, represented by a star in deep blue colour.

TABLE 1

Sampling sites, coordinates of sampling sites, sample sizes and GenBank accession numbers of sequences for 268 samples of *C. gariepinus* (Burchell, 1822) collected from 10 different localities (five natural and five farms) in Kenya.

Site	Population Code	Coordinates	Sample size	GenBank sequence accession numbers
Lake Victoria	LVG	34°38′ E, 0°21′ S	24	KC594181-KC594205
Lake Kanyaboli	LKG	00°04′30″ N, 34°09′36″ E	27	KC594206-KC594232
Lake Turkana	LTA	3°37′ N, 36°0′ E	28	KJ814254-KJ814281
Lake Baringo	LBA	0°38′ N, 36°05′ E	24	KJ814282-KJ814305
Lake Jipe	LJP	3°35′ S, 37°45′ E	32	KJ814306-KJ8143037
Sangoro Aquaculture Centre	SAN	0°30′ N, 0°45′ E	30	KJ814338-KJ814367
Sagana Aquaculture Centre	SAG	0°39′ S, 37°12′ E	23	KJ814368-KJ814390
University of Eldoret Fish Farm	UoE	0°57′ N, 35°30′ E	29	KJ814391-KJ814419
Kibos Fish Farm	KIB	0°04′ S, 4°48′ E	26	KJ722140-KJ722165
Wakhungu Fish Farm	WKU	0° 30′ N, 0° 00′ E	25	KJ814420-KJ814444

netting, hook and line, and traps, while seine netting was used at fish farms to catch brooders kept in earthen ponds. Species were identified using identification keys for field studies (WITTE & VAN DENSEN 1995). Fin clips were collected as the source of DNA, and immediately preserved in 95% ethanol in cryovials until laboratory genetic studies were done. All the farms sampled sourced their brood stock in the early 1990s, with no records of replacement of original stocks. Actual numbers of brood stock at each is unknown, and some of the farms retain a portion of their offspring from propagation to augment brood stock sizes, and none practices selective breeding.

DNA extraction

Genomic DNA was extracted from approximately 25 mg of fin clip tissue using the Invitrogen PureLink genomic DNA mini kit (cat no. K1820-02), according to the manufacturer's instructions. The purity and concentration of eluted DNA was determined by spectrophotometry using a Nanodrop 2000-Spectrophotometer (Thermo Scientific). Extracted DNA was stored at -20°C until required for further analysis.

Mitochondrial DNA analysis

DNA samples from 268 samples of *C. gariepinus* from 10 sites were PCR amplified in a thermal cycler (ABI 9700) using a pair of mitochondrial D-loop primers (forwards primer: L16473 5′-CTAAAAGCATCGGTCTTGTAATCC-3′; reverse primer: H355 5′-CCTGAAATGAGGAGGAACCAGATG- 3′ (NAZIA *et al*., 2010). PCR reactions were done using AccuPower PCR PreMix (Bioneer South Korea), in a 20 µl volume containing 10 µM each of forward and reverse primers, and 50ng DNA. PCR conditions were as described by NAZIA *et al.* (2010). The success of the PCR reaction was confirmed by 2.0% agarose gel electrophoresis of PCR products. PCR products were purified by ethanol precipitation (UTHICE & BENZIE 2003; BARASA *et al.* 2014). Precipitated DNA was washed once with 70% ethanol, air dried for 20 minutes, re-suspended in 20 µl distilled water and stored at -20°C. The reverse primer H355 was used to sequence the PCR product of the D-loop region. The BigDye terminator premix sequencing kit (cat no. 4336911 from ABI Life Technologies) was used for sequencing reactions, following the manufacturer's protocol. Products of the sequence reaction

were cleaned by precipitation in absolute alcohol, re-suspended in Hi DiTM Formamide, and BigDye terminator premix, and run on the ABI 3730xl Genetic Analyzer with a capillary length of 50 cm.

Microsatellite DNA analysis

Genotyping at six loci was done on amplified products from 160 samples of *C. gariepinus* from eight sampling sites (LVG, LKG, LTA, LBA, SAN, SAG, UoE and KIB; LJP and WKU were not included) in Kenya. DNA was extracted as described above in the mtDNA study, its purity and concentration quantified on Nanodrop 2000 spectrophotometer (Thermo Scientific), and PCR amplified. Of the six microsatellite markers (Table 4), four (Cga1, Cga3, Cga9, Cga10) were developed by Galbusera *et al.* (1996), while two (Cba2, Cba19) were developed by Yue *et al.* (2003). For the four markers (Cga1, Cga9, Cga10 and Cba2), PCR reaction was carried out in 12.5 µl reaction volume, comprising of 0.625 µl each of forward and reverse primers (reconstituted to 10 pmoles/µl each), 4 µl of distilled water, 6.25 µl of 2× Kapa2G™ Robust HotStart Ready mix, and 1 µl of template DNA sample.

The PCR reaction mixes for Cga3 and Cba19 were similar as for above, except 3 µl water and 1 µl of 25 mM magnesium chloride and 3.5 µl water and 1 µl of 25 mM magnesium chloride were used for Cga3 and Cba19 respectively. For Cga3 and Cga10, samples were amplified at an annealing temperature of 60°C while for Cga1 and Cga9, annealing temperature was 56°C and 65°C respectively. Initial denaturation of samples was done for 3 minutes at 95°C, followed by 35 cycles of amplification each at 95°C for 15 seconds, 15 seconds of annealing, 15 seconds at 72°C, with a final primer extension step of 10 minutes at 72°C. For Cba2 and 19, the cycling conditions were an initial denaturation at 94°C for 3 minutes, followed by 35 cycles of amplification at 94°C for 30 seconds, primer annealing at 56°C for 30 seconds, and primer extension at 72°C for 1 minute, with final primer extension step of 5 minutes at 72°C. 3 µl of each PCR product was electrophoresed on a 2% agarose gel, to confirm the success of PCR amplification.

PCR products were prepared for analysis as follows: each reaction consisted of 9.25 µl of Hi-Di Formamide, 0.25 µl of Genescan-500 LIZ size standard and 0.5 µl PCR sample. The mix was heated at 95°C for 3 minutes, then chilled immediately on ice and analyzed on the ABI 3730xl Genetic Analyzer (Applied Biosystems).

Mitochondrial DNA data analysis

DNA sequences were trimmed, assembled and aligned in CLC main Work bench (version 6.7.2, Inqaba Biotech). Duplicate haplotypes were identified using DnaSP (version 5) (Librado & Rozas 2009). Genetic diversity within populations was determined as number of distinct haplotypes, haplotype frequencies and nucleotide diversities, using DnaSP and ARLEQUIN (version 3.5) (Excoffier *et al.* 2005). ARLEQUIN (version 3.5) was also used to determine genetic differentiation between groups, expressed as FST (Wright 1965). A maximum likelihood tree (with *C. liocephalus* as an outgroup) was drawn using MEGA (Tamura *et al.* 2007), with 1,000 bootstrap repeats. Modeltest 3.7 (Posada & Crandall 1998) was used to choose the most likely model of evolution for the *Clarias* mtDNA data set. A Minimum Spanning Network showing the relationship between haplotypes was drawn using Network 4.56, with a median joining approach (Bandelt *et al.* 1999).

Microsatellite DNA data analysis

Fragments were scored using GENE MARKER version 2.6.3 for size and number of alleles at loci. Mean observed heterozygosity and expected heterozygosity were determined in Microsatellite Toolkit version 3.1.1 (Park 2001). Mean number of private alleles and effective alleles were computed in ARLEQUIN version 3.5 (Excoffier *et al.* 2005). Deviations from Hardy Weinberg expectations were tested using GenAlEx (version 6.502) (Peakall & Smouse 2012). The number of clusters in the samples was determined by the Bayesian clustering algorithm implemented in the program STRUCTURE version

2.3.4 (Pritchard *et al.* 2000), under an admixture model with correlated allele frequencies. The program was run with a burn-in period of 10,000 and 100,000 Monte Carlo simulations. The number of K was tested for K= 2 to 7 (i.e., the presumed number of populations or clusters in the samples), and for each of these K values, 10 runs were made. STRUCTURE HARVESTER (web version 0.6.94) (Earl & Von Holdt 2012) was used to compute the number of genetic clusters in the samples. This was inferred from the distribution of delta K (ΔK), an *ad hoc* statistic based on the rate of change in the log probability of data between successive K values (Evanno *et al.* 2005). The actual K value most suitable for the data set is the modal value of the ΔK distribution (Evanno *et al.* 2005). Identity of particular populations making up the genetic clusters revealed by STRUCTURE HARVESTER was determined by a series of eight (representing the eight sampling sites) runs in STRUCTURE and STRUCTURE HARVESTER, excluding samples from one site for each run. The membership coefficient of each of the eight pre-defined populations in each of the genetic clusters was determined as an average of outputs of ancestry proportions generated by STRUCTURE.

Results

Catfish genetic diversity inferred from mitochondrial D-loop DNA sequences

Nucleotide diversity ranged from 0.006 to 0.067, with LKG, LBA and WKU having the lowest value of 0.006, while SAG had the highest value of 0.067 (Table 2). Therefore among the five natural populations of catfish, nucleotide diversity was lowest in LBA, and highest in LJP, but highest in SAG for the farmed populations. The number of haplotypes ranged from seven to 14 for the 10 populations, with LVG and SAG having the highest among natural and farmed populations respectively. LBA and LTA had the lowest number of haplotypes with eight each among the samples of natural populations, while WKU and UoE had the lowest number with seven each among samples from the farms (Table 2). Of the total number of haplotypes, all populations except LTA and LJP shared haplotypes. A total of eight haplotypes were shared among populations. This therefore left each of the populations with a number of singletons (private haplotypes), totalling 88.2% of the haplotypes, with LJP having the highest number at 12 (haplotypes H-18 to 29), while KIB and WKU had a single private haplotype each (H-7 and H-68 respectively). Therefore, with the exception of SAG, all the farmed populations had a lower number of singletons than natural populations (Table 2). Haplotype 2 was the most frequent, appearing in 76 individuals of the LVG group.

The diversity of haplotypes (*h*) was consistent with the number of haplotypes, where populations with a higher number of haplotypes also had a higher diversity of haplotypes. However, LJP which had the third highest number of haplotypes had the lowest haplotype diversity. Haplotype diversity values ranged from 0.679 to 0.941, with SAG reporting the highest haplotype diversity (Table 2). The D-loop region in the populations segregated at a total of 386 polymorphic sites, with LJP having the highest with 161sites, while for the rest of the populations, the sites ranged from 11 to 112 (Table 2).

Pairwise comparisons for FST values among populations revealed significant differentiation in 38 out of 45 comparisons (Table 3). Among the natural populations, LVG, LTA, LBA and LJP were all differentiated ($p \leq 0.001$). However, LVG and LKG were not significantly differentiated ($p \geq 0.01$), but LKG was differentiated from LTA, LBA and LJP ($p \leq 0.001$). Among the farmed populations, SAG-SAN, SAG-WKU, SAG-UoE and KIB-SAG pairwise comparisons were significantly different ($p \leq 0.001$), (Table 3), showing the possibility of SAG having been sourced from a site different from that of other farms, or a combination of sites. Similarly, WKU-SAN and WKU-UoE were significantly different at $p \leq 0.01$, while KIB-UoE and SAN-UoE were significantly different at $p \leq 0.05$. Pair-wise FST comparisons between natural and farmed *C. gariepinus* showed all farmed populations were significantly differentiated ($p \leq 0.01$) from LTA, LBA and LJP. This differentiation in populations, where LVG, LTA, LBA, LJP and SAG show significantly different values of FST when compared with each other, or with samples from fish farms, is also clearly reflected in phylogenetic relationships (Figs 2–3). In both the

TABLE 2

Genetic diversity values for the 268 samples of *C. gariepinus* (Burchell, 1822) collected from 10 different sampling sites in Kenya (five natural and five farms), as inferred from 427 base pairs of mtDNA D-loop control region sequences. Π is the nucleotide diversity, while *h* is the haplotype diversity.

Population	**LVG**	**LKG**	**LTA**	**LBA**	**LJP**	**SAN**	**SAG**	**WKU**	**UoE**	**KIB**
Sample	24	27	28	24	32	30	23	25	29	26
Π	0.009	0.006	0.009	0.006	0.037	0.006	0.067	.006	.009	0.009
No of haplotypes	13	10	8	8	12	9	14	7	7	9
h	0.813	0.745	0.791	0.794	0.679	0.766	0.941	0.72	0.732	0.812
Shared haplotypes	6	5	0	1	0	5	4	6	4	8
Singletons	7	5	8	7	12	4	10	1	3	1
Polymorphic sites	21	18	12	16	161	12	112	11	12	11

TABLE 3

F_{ST} pair wise comparisons of 268 samples of *C. gariepinus* (Burchell, 1822) from 10 different sampling localities in Kenya, based on 427 bp of mitochondrial D-loop control region sequences. Values in bold are significantly different. A total of 38 of 45 pair wise comparisons are significantly different.

	LVG	**LKG**	**LTA**	**LBA**	**LJP**	**SAN**	**SAG**	**WKU**	**UoE**	**KIB**
LVG	0.000									
LKG	0.014	0.000								
LTA	**0.853*****	**0.877*****	0.000							
LBA	**0.534*****	**0.652*****	**0.870*****	0.000						
LJP	**0.957*****	**0.959*****	**0.959*****	**0.958*****	0.000					
SAN	0.011	0.000	**0.883*****	**0.660*****	**0.962*****	0.000				
SAG	**0.253*****	**0.281*****	**0.393*****	**0.337*****	**0.921*****	**0.290*****	0.000			
WKU	**0.058***	**0.155****	**0.870*****	**0.562*****	**0.958*****	**0.163****	**0.275*****	0.000		
UoE	0.019	**0.092****	**0.847*****	**0.483*****	**0.959*****	**0.087***	**0.253*****	**0.058****	0.000	
KIB	0.040	**0.141****	**0.851*****	**0.410*****	**0.957*****	0.153	**0.268*****	0.043	**0.069***	0.0

*p≤0.05, **p≤0.01, ***p≤0.001

Maximum likelihood tree (Fig. 2) and the minimum spanning networks for haplotypes (Fig. 3), the samples clustered into five clades: LVG/LKG, LTA, LBA, LJP and SAG.

Catfish genetic diversity inferred from Microsatellite DNA markers

The mean number of alleles per locus (Na) was generally moderate in *C. gariepinus* populations, ranging from 3.80±0.84 in LBA to 10.83±3.66 in SAN. The mean observed heterozygosity (HO) ranged from 0.465±0.054 in LBA to 0.795±0.037 in SAN (Table 5). Similarly, values for expected heterozygosity (HE), ranged from 0.582±0.045 in LBA to 0.839±0.049 in LVG, and were slightly higher than HO values. However, AMOVA analysis showed that variation in the populations was mainly attributed to individuals, with 96.72% of the variation in samples of *C. gariepinus*. Variation among individuals within populations was 7.49%, while variation among populations was minimal.

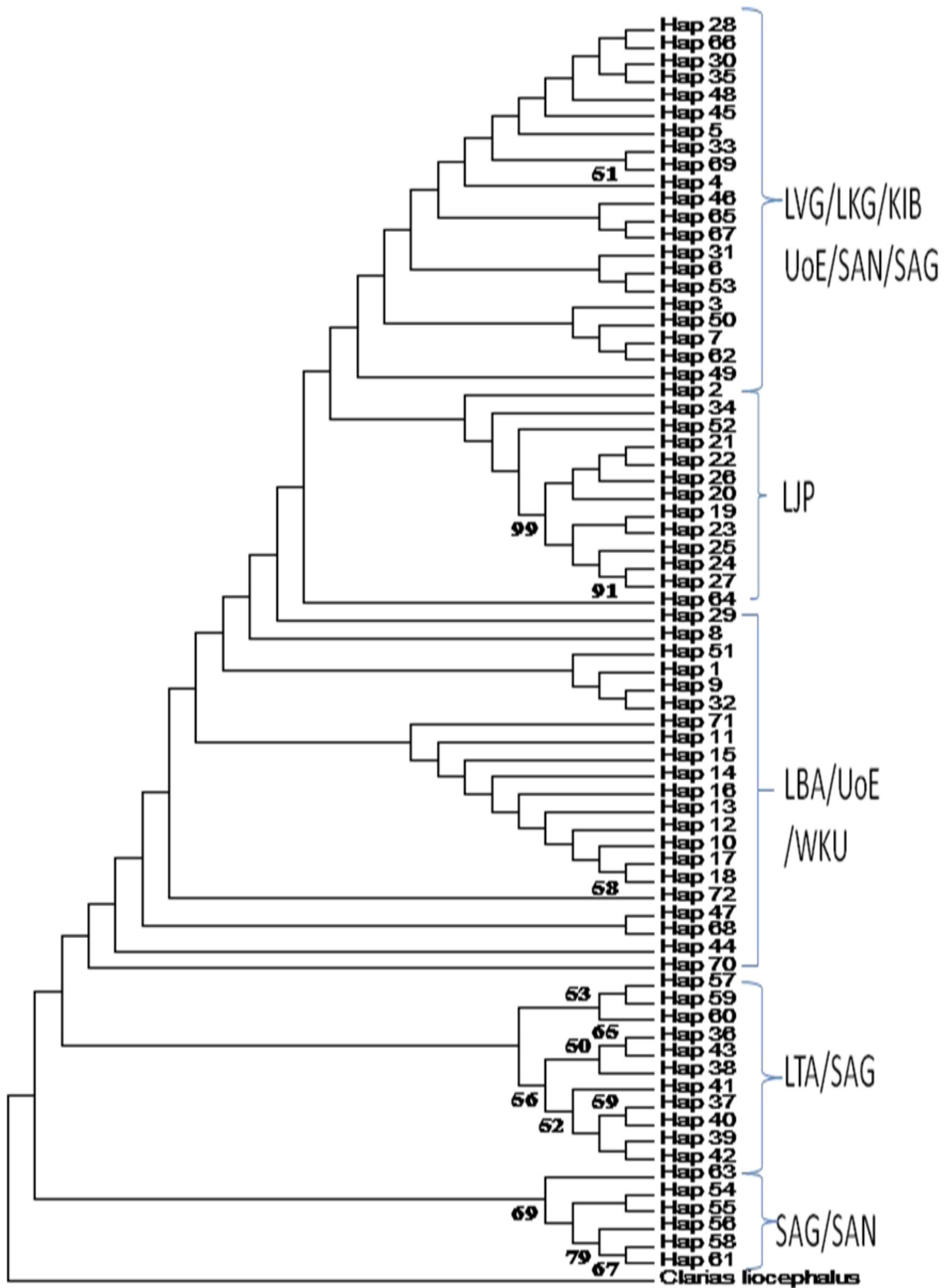


Figure 2 – Maximum likelihood tree for 68 haplotypes of 268 samples of *C. gariepinus* (Burchell, 1822) from 10 different sampling sites in Kenya. Five clusters are discerned, consisting of LVG/LKG, LJP, LBA, LTA and SAG. The label is borrowed from the median network (Fig. 3). Numbers at nodes indicate confidence levels, based on 1,000 bootstrap iterations. Nodes without numbers have confidence levels less than 50%. The LVG/LKG cluster groups LVG and LKG, and farmed samples UoE, KIB, SAN and SAG. LTA cluster has samples from SAG. The LBA cluster comprises samples from UoE and WKU. *C. liocephalus* is the outgroup.

TABLE 4

Microsatellite primer sequences with nucleotides, range of allele size, dye colour and reference for the six loci used to genotype 160 samples of *C. gariepinus* (Burchell, 1822) from eight different sampling localities (four natural and four farms) in Kenya.

Primer	Nucleotides	Size range	Dye colour	Reference
CGA01	5′ GGC TAA AAG AAC CCT GTC TG 3′ 3′ TAC AGC GTC GAT AAG CCA GG 5′	92–104	Green	Galbusera *et al.* 1996
CGA03	5′ CAC TTC TTA CAT TTG TGC CC 3′ 3′ ACC TGT ATT GAT TTC TTG CC 5′	142–168	Blue	Galbusera *et al.* 1996
CGA09	5′ CGT CCA CTT CCC CTA GAG CG 3′ 3′ CCA GCT GCA TTA CCA TAC ATG 5′	180–196	Green	Galbusera *et al.* 1996
CGA10	5′ GCT GTA GCA AAA ATG CAG ATG C 3′ 3′ TCT CCA GAG ATC TAG GCT GTC C 5′	102–138	Green	Galbusera *et al.* 1996
CBA02	5′ GCC CTG CGA ACA TCT CCA 3′ 3′ TGG CTC CAG CAC TCA CAA 5′	176–190	Yellow	Yue *et al.* 2003
CBA19	5′ CAG GGC TAA ATT ACC CAT AAT CA 3′ 3′ GGC ATG TGT TAT AAC ATG TGA GG 5′	215–255	Green	Yue *et al.* 2003

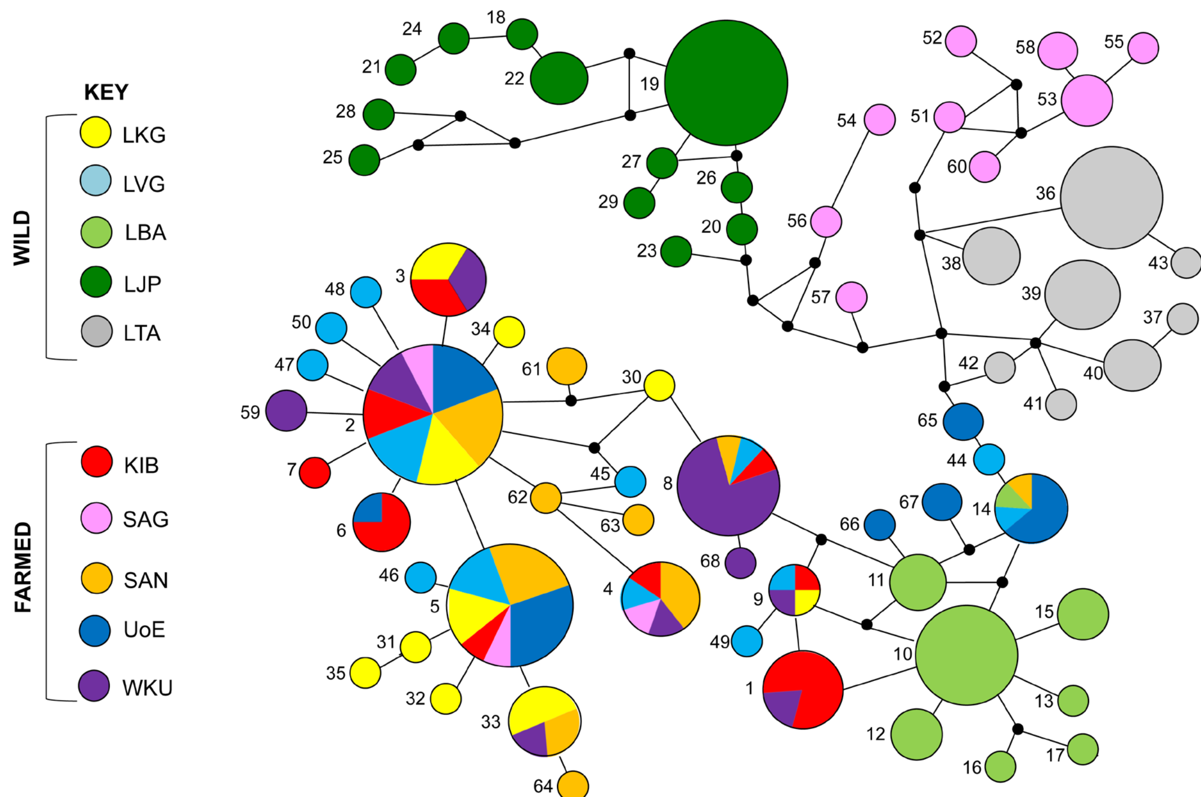


Figure 3 – Minimum spanning network showing mutational relationships between 68 haplotypes from 268 individuals of *C. gariepinus* (Burchell, 1822) from 10 different sampling sites in Kenya. Size of circle is proportional to the frequency of haplotypes. The small black dots denote mutational steps, with a single dot denoting one mutational step. Five clusters are discerned, consisting of four natural (LVG/LKG, LTA, LBA and LJP) and one farmed population (SAG). Samples from farms group in to LVG and LBA clusters.

TABLE 5

Level of genetic diversity (Na is the number of alleles; H*o* is the observed heterozygosity; and H_E is the expected heterozygosity) in *C. gariepinus* (Burchell, 1822) sampled from eight different localities in Kenya and genotyped at six microsatellite loci. F_{IS} values in bold are significantly different. Data is based on 160 fish samples from eight sampling sites (four natural and four farmed).

Population	N	Na	H_o	H_E	No. of private alleles	No. of effective alleles	Coefficient of inbreeding (F_{IS})
LVG	23	8.0±3.5	0.79±0.05	0.839±0.05	5.86±0.79	6.66± 0.32	-0.614
LKG	20	8.17±3.31	0.72±0.05	0.827±0.05	5.47±0.87	6.35± 0.32	**0.112***
LTA	20	8.00±2.68	0.74±0.05	0.824±0.04	5.53±1.11	6.66± 0.40	0.266
LBA	18	3.80±0.84	0.465±0.05	0.582±0.05	5.75±0.43	6.35± 0.61	0.198
SAN	20	10.83±3.66	0.795±0.04	0.835±0.04	5.50±0.78	5.79± 0.84	0.016
SAG	20	7.67±2.73	0.55±0.05	0.758±0.04	5.55±0.62	6.66± 0.32	**0.250****
UoE	20	8.83±2.56	0.737±0.04	0.820±0.04	5.61±0.56	6.66± 0.61	**0.095***
KIB	20	9.67±2.88	0.699±0.05	0.846±0.04	5.91±0.67	6.66± 0.32	**0.069***

*p≤0.01; **p≤0.05

All the populations had private alleles, i.e., alleles restricted or specific to a particular population. The number of private alleles ranged from 5.47±0.87 in LKG to 5.91±0.67 in KIB. The coefficient of inbreeding (FIS) was significantly different (p≤0.05) in LKG and three farmed populations UoE, KIB and SAG, but when adjusted for Bonferroni correction, only LKG and two farmed populations (UoE and KIB) were significantly different (p≤0.01) (Table 5).

None of the fish populations showed significant deviations of genotypic frequencies from those expected under HWE at all six loci. However, deviation was observed only in LBA at locus Cga9, and LVG at locus Cga3, with a significantly different p-value (p≤ 0.05) for Chi-square. After Bonferroni correction, both loci were significantly different (p≤0.01), but the difference was not significant at the lower confidence level (p≥0.001). Locus Cba2 for LBA was left out of the analysis for this population because it was monomorphic (Table 6).

FST values were significantly different (p ≤ 0.01) in 15 out of 28 pairwise comparisons (Table 7). Of these, LVG-LTA and LVG-LBA comparisons were highly significant (p≤ 0.001). Most of the significantly different pairwise comparisons were between samples of natural populations from different sites, and between natural and farmed samples.

From the STRUCTURE HARVESTER, samples grouped into four genetic clusters (Figs 4–5), with the first three clusters being LVG/LKG, LTA, and LBA. The fourth cluster was deduced as SAG. Samples from the LVG pre-defined population constituted the highest proportion of the LVG/LKG cluster at 0.4483 (Table 8). Similarly, samples from the LBA gene pool constituted the highest proportion of the LBA cluster, with a membership coefficient of 0.9748, while SAG samples dominated in the SAG cluster with a membership coefficient of 0.6219. Likewise, the LTA gene pool constituted the highest proportion of the LTA cluster, with a membership coefficient of 0.8241. SAN, UoE and KIB shared the highest proportion of membership with the LVG cluster (0.5385 to 0.7092) (Table 8).

TABLE 6

P-values for HWE for 160 samples of *C. gariepinus* (Burchell, 1822) from eight sampling sites (four fish farms and four natural habitats) in Kenya, genotyped at six loci. Values in bold are significantly different at p ≤0.01 and 0.05.

Population	Locus					
	Cga1	Cga3	Cga9	Cga10	Cba2	Cba19
LVG	0.366	**0.003**	0.411	0.366	0.451	0.015
LKG	0.064	0.116	0.014	0.423	0.317	0.711
LTA	0.058	0.127	0.042	0.029	0.258	0.637
LBA	0.013	0.563	**0.010**	0.348	–	0.651
SAN	0.579	0.246	0.015	0.023	0.821	0.033
SAG	0.026	0.270	0.014	0.015	0.057	0.140
UoE	0.015	0.763	0.016	0.883	0.546	0.474
KIB	0.077	0.954	0.020	0.170	0.166	0.135

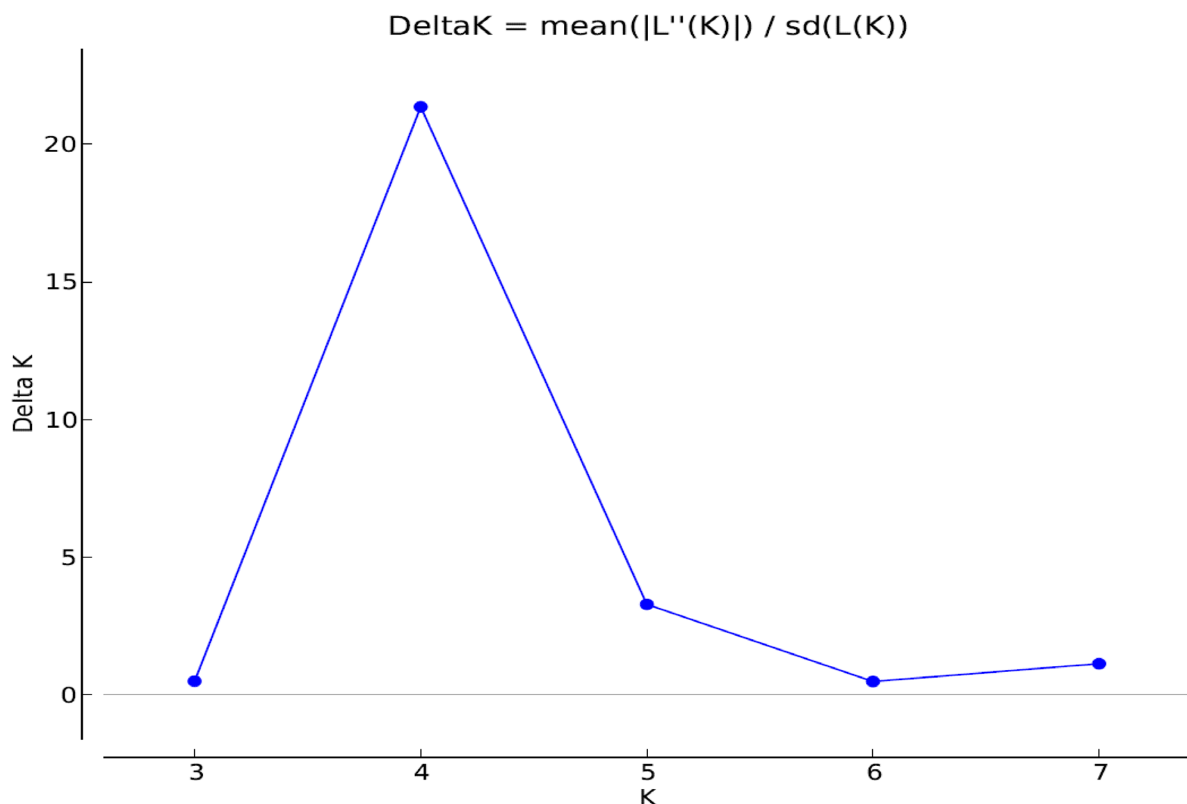


Figure 4 – Graphical representation of results from STRUCTURE, using STRUCTURE HARVESTER. K is the most likely number of genetic clusters into which the samples group, representing the genetic population structure. The actual K for the data set is inferred as the modal value of the delta K distribution. Delta K is an *ad hoc* statistic that shows the rate of change in the log probability of data between successive K values being tested. The 160 samples of *C. gariepinus* (Burchell, 1822) from eight different sites in Kenya most likely represent four genetic clusters. Samples of LJP and WKU were not included in the genotyping.

TABLE 7

Pair wise F_{ST} values for 160 samples of *C. gariepinus* (Burchell, 1822) from eight different sampling localities (four natural and four farmed) in Kenya, genotyped at six loci. Values in bold are significantly different. A total of 15 out of 28 pair wise comparisons are significantly different.

	LVG	LKG	LTA	LBA	SAN	SAG	UoE	KIB
LVG	0.000							
LKG	0.247	0.000						
LTA	**0.071****	0.172	0.000					
LBA	**0.000****	**0.018***	**0.243***	0.000				
SAN	0.187	0.089	0.189	**0.253***	0.000			
SAG	0.227	0.159	**0.163***	**0.290***	**0.077***	0.000		
UoE	0.206	0.112	**0.152***	**0.267***	**0.033***	**0.092***	0.000	
KIB	0.115	0.083	**0.095***	**0.210***	0.029	0.088	**0.040***	0.000

*p≤0.01; **p≤0.001

Discussion

Mitochondrial DNA

The nucleotide diversity of the catfish populations was low, while haplotype diversity was high, but comparable to values reported by other studies (ROODT-WILDING *et al.* 2010, NAZIA *et al.* 2010). This shows that genetic variation in the populations could be accounted for by within population variation, which was comparatively higher than values reported for populations of *C. macrocephalus* from Malaysia (NAZIA *et al.* 2010), whose nucleotide and haplotype diversities were 0.003 in all the three populations and 0.657 to 0.765 respectively. However, haplotype and nucleotide diversities in the current

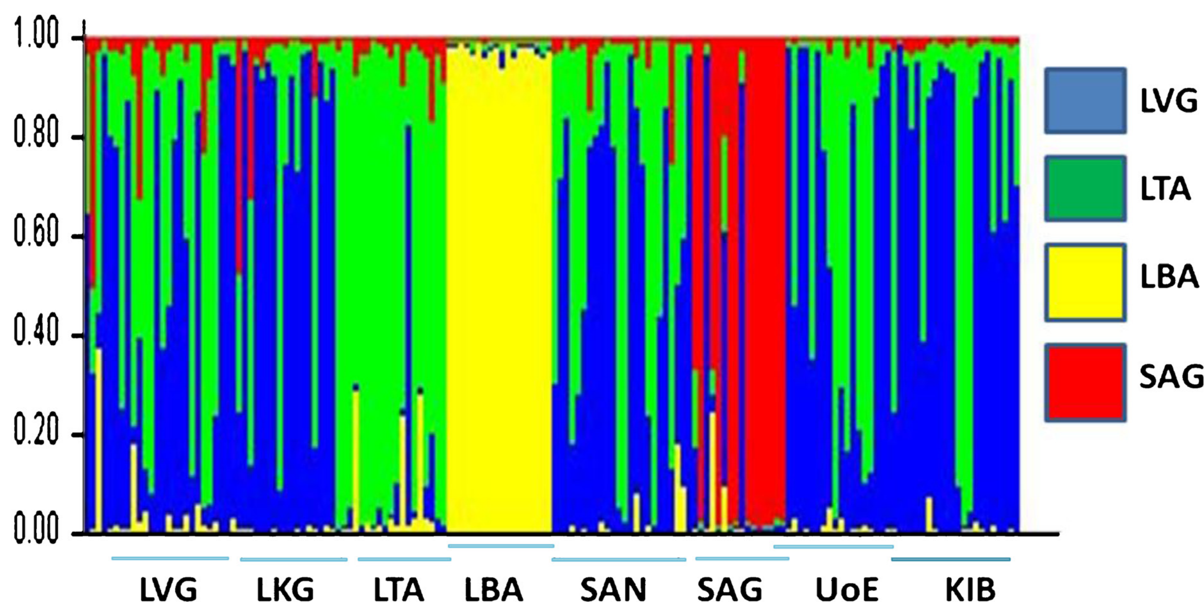


Figure 5 – STRUCTURE bar plot of 160 individuals of *C. gariepinus* (Burchell, 1822) from eight different sampling sites in Kenya, genotyped at six microsatellite loci. Each vertical bar represents an individual, and each colour represents a probability that the individual is assigned to each genetic cluster identified (K=4), using the admixture proportion, qi, from the output of the STRUCTURE software. LVG is the Lake Victoria, LTA the Lake Turkana, LBA is the Lake Baringo, while SAG is the Sagana Fish Farm populations of *C. gariepinus*.

TABLE 8

Proportion of membership of each pre-defined population in each of the four genetic clusters (K=4). STRUCTURE was run with 10,000 Burn-in period and 100,000 Monte Carlo simulations.

Pre-defined population	Inferred genetic cluster				Sample size
	LVG/LKG	LBA	SAG	LTA	
LVG	0,4482	0,04	0,069	0,4429	23
LKG	0,7592	0,0094	0,0538	0,1776	20
LTA	0,0733	0,0561	0,0466	0,8241	19
LBA	0,0099	0,9748	0,0052	0,0101	18
SAN	0,5385	0,0119	0,0239	0,4257	20
SAG	0,2339	0,038	0,6219	0,1062	20
UoE	0,5873	0,0131	0,0131	0,3865	20
KIB	0,7092	0,011	0,0138	0,266	20

study were similar to values reported for populations of *C. gariepinus* from South Africa, which ranged from 0.838±0.030 to 0.904±0.019 and 0.006±0.003 to 0.008±0.040 respectively (Roodt-Wilding *et al.* 2010). The presence of shared haplotypes is attributed to gene flow among the populations, and this was restricted to farmed populations (SAN, UoE, SAG, KIB and WKU), which were mainly sourced from Lake Victoria, or from other fish farms that initially collected their brood stock from Lake Victoria. Human translocation of fish populations across drainage basins (Nazia *et al.* 2010) for aquaculture is common, especially in *Clarias* aquaculture that frequently depends on collection of males from natural populations to provide the pituitary hormone and milt during artificial propagation at hatcheries (Barasa *et al.* 2014), and when fish escape, this could lead to gene flow and homogeneity of geographically isolated populations of the species. However, SAG formed a distinct cluster of its own, unlike samples from the rest of the fish farms, which grouped into LVG and LBA. This finding, together with SAG reporting the highest number of haplotypes, and haplotype and nucleotide diversities among the 10 sampling sites seem to suggest that farmed catfish at Sagana Aquaculture Centre were sourced from multiple sites, including LVG. However, most of these sites seem not to have been sampled by this study, and could include rivers and farms around the Mount Kenya region. and Sagana Aquaculture Centre. A combination of genetically distinct stocks is reported to increase genetic variation in farmed *C. gariepinus* (Van Der Bank *et al.* 1992; Grobler *et al.* 1997).

However, samples of SAG appeared in the LVG/LKG and LBA cluster. This could be attributed to human-assisted translocation of the population for aquaculture, especially in the late 1990s and early 2000s, when the farm was the main source of fish seeds (*O. niloticus* and *C. gariepinus*) in Kenya after a rehabilitation and refurbishment of the hatchery in the mid 1990s. At the same time the Aquaculture/Collaborative Research Support Programme project was initiated at the SAG Centre and on-farm trials of fish growth in western Kenya were undertaken, using seeds from the Centre (Ngugi *et al.* 2003). Similarly, seeds of *C. gariepinus* from the Centre were transported and stocked in farmers' ponds in Uasin Gishu county (BARASA, personal observation), before the UoE fish farm in the county was established. From these ponds, the fish were distributed widely in the region, and could be the source of SAG in the LVG and LBA clusters.

In addition, farmed populations, except SAG, had lower haplotype diversities than natural populations. This could be attributed to genetic drift, which caused a reduction in gene frequencies in the samples. Generally, the number of brooders maintained at most small scale farms of *C. gariepinus* is low, which declines further because of cannibalism among the fish (Sulem *et al.* 2006) and the sacrifice of male

brooders for milt and the pituitary hormone (De Graaf *et al.* 1995). These factors also cause an imbalanced sex ratio among brood stock and increase genetic drift in populations (Allendorf 1986; Vrijenhoek 1998).

Of the five natural populations sampled, LKG and LJP had the lowest haplotype diversities. Both lakes are small and isolated (L. Kanyaboli is 10.5 km^2, L. Jipe 30 km^2, L. Baringo 130 km^2, L. Turkana 7,000 km^2, while L. Victoria is 69,000 km^2 in surface area), and higher fishing pressure, especially in the LKG population (Aloo 2003), could have reduced its genetic diversity, through fishing mortality (Van Der Walt *et al.* 1993). Small and isolated populations also suffer lower genetic diversities due to genetic drift that results from founder effects and lower effective population sizes (Franklin & Frankharm 1998). Although the haplotype diversities in LVG, LTA, LBA and LJP were comparable, ranging from 0.679 to 0.813, LVG had a slightly higher diversity, which was expected, because of its larger population size.

On the other hand, the high number of private haplotypes in the populations could be attributed to the rapid rate of evolution of the mitochondrial D-loop region as a marker for population genetic structure (Meyer 1994). Changes in the genome are captured by the mitochondrial region because its effective population size is four-fold lower than the nuclear genome (Triantafyllidis *et al.* 1999). This leads to private haplotypes in populations, even in populations that are under genetic drift, as reported in *C. gariepinus* of the satellite lakes of Yala swamp, Kenya (Barasa *et al.* 2016) and siluroid catfishes of Europe (Triantafyllidis *et al.* 1999).

High differentiation was noted among populations, which could be due to geographic isolation of the main populations. According to Beadle (1974) and Giddelo *et al.* (2002), rifting caused the elevation of Lake Victoria on to an uplifted plateau (1,000 m), separating it from the connectivity with the western rift. Although both lakes Baringo and Turkana are in the old eastern or Gregory rift, Lake Turkana and the Omo River were separated from the Nile River system less than 10,000 years ago (Dumont 1986), when the Kanguen river became affected by tectonic uplifting (Beadle 1974). This disrupted the connectivity of Lakes Turkana and Victoria, and Turkana and Baringo. The geographical separation of the water masses could have led to high differentiation of populations of *C. gariepinus*, which was also reported by Roodt-Wilding *et al.* (2010) and Giddelo *et al.* (2002).

These natural populations of *C. gariepinus* therefore represent important genetic resources that could be harnessed for intensive aquaculture programmes for higher food security and incomes among fish farmers. In addition, these genetically differentiated populations represent important biodiversity resources or genetic diversity that should be conserved.

Microsatellite DNA

The populations were generally in HWE, since no population showed a deviation from HW at all loci. A departure from HW equilibria was observed only at locus Cba9 for LBA and LVG at locus Cga3, showing a heterozygote deficit, attributed to small sample sizes. Significantly higher values of FIS ($p \leq 0.01$) observed in LKG, SAG, UoE and KIB could be attributed to loss of heterozygosity due to small population size. Although the exact number of brooders used at the hatcheries is unknown due to poor record keeping, maintaining lower numbers of brood stock is common at hatcheries for *C. gariepinus* (such as SAG, SAN, UoE and KIB), where certain females are selected and re-used in artificial propagation. Although the size of a fish population is not entirely a function of the habitat size, numbers of *C. gariepinus* inhabiting Lake Kanyaboli may be comparatively smaller than the populations in other lakes. The smaller size of LKG therefore, coupled with overfishing (Aloo 2003) that leads to fishing mortality, may cause loss of heterozygosity in LKG. Insignificant values of FIS in LVG, LTA and LBA show that inbreeding was absent in these populations, which was expected because of the larger size of the water masses, and therefore a comparatively higher population size than LKG and farmed populations. Total landings of *C. gariepinus* in 2013 were 2368, 61, 26 and 38 metric tons

for Lakes Victoria, Turkana, Baringo and Kanyaboli respectively (State Department of Fisheries 2013), although the dynamics influencing fish production at each of the lakes differ. A larger watermass provides more spawning grounds, reducing inbreeding in fish (Castric *et al.* 2002).

Generally, higher genetic variation in natural than in farmed fish populations is common, as reported in the Indian major carp *Catla catla* (Hansen *et al.* 2006; Alam & Islam 2005), and the Atlantic salmon, *Salmo salar* (Norris *et al.* 1999). These studies demonstrated the potential impacts of domestication on genetic variability in fish, which occurs by inbreeding that decreases the fitness of populations (Hansen *et al.* 2006), fixation of deleterious alleles or due to absence of variation at loci showing over-dominance (Hedrick & Kalinowski 2000).

In this study however, all farmed populations except SAG had higher mean number of alleles than natural populations. Artificial propagation of *C. gariepinus* at hatcheries involves collection of males from wild habitats to provide milt for egg fertilization, and the pituitary gland for induced breeding. This introduces novel alleles, increasing the mean number of alleles without increasing heterozygosity in the fish stocks at the hatcheries, a practice widely used to increase genetic variation after several generations of repeated breeding of the same brooders (Van Der Bank *et al.* 1992; Hedrick & Kalinowski 2000; Wachirachaikarn *et al.* 2009; Sukmanomon *et al.* 2012).

However, among the natural populations, the lower number of alleles of LBA could be attributed to a recent colonization of the lake by *C. gariepinus*, after the lake witnessed frequent drying during the Holocene (Verschuren *et al.* 2000), especially given its shallow depth (maximum depth 2.1 m) and a small surface area of 130 km^2 (Beadle 1974; Bessem *et al.* 2008). Empirical evidence shows that Lake Baringo was completely dry about 200 years ago (Bessem *et al.* 2008). This implies that the evolutionary age of a fish lineage influences genetic variation. Lower genetic variation is reported in fish with recent colonization history than in older and more stable lineages (Barluenga & Meyer 2010). Although both LTA and LBA are located in the eastern Rift, LTA showed a higher mean number of alleles and therefore higher genetic variation than LBA, because of a historical connectivity with the Western Nile (Dumont 1986). Rifting cut off this connectivity, but LTA population retained the higher genetic diversity extant in *C. gariepinus* of the Nile system, but LBA was seeded by the Kanguen River after re-filling in the late Holocene (Dumont 1986; Bessem *et al.* 2008).

On the other hand, LVG, LKG and LTA showed a uniform number of alleles and HE, which were much higher than for LBA. This could be a result of historical connectivity of L. Victoria to the western arm of the rift, sharing ichthyofauna with Lakes Kyoga, Albert and George before it was uplifted to 1,000 m by tectonic uplifting (Beadle 1974). This uplifting occurred in the Miocene (6 million years before present), and disconnected Lake Victoria from this historical connectivity to the western rift (Partridge *et al.* 1995), but the population of *C. gariepinus* retained the higher genetic variability of the Nile system. This diversity in genetic variability could have been maintained by the large size of the Lake Victoria water mass (and therefore the large size of the LVG population) (Barasa *et al.* 2014), despite predation of LVG population of *C. gariepinus* by the exotic *L. niloticus* (Goudswaard & Witte 1997). The historical connectivity of LVG and LTA to the western arm of the rift could also account for comparable genetic diversities between the two populations of *C. gariepinus*.

The number of alleles and HE among natural catfish populations showed minimal variation (apart from LBA), while both the number of alleles (Na) and HE varied strongly among farmed populations. This is consistent with the results of Hansen *et al.* (2006), where strong variation in both Na and HE among farmed fish samples occurs due to genetic drift in samples, related to breeding practices, such as sacrificing male brooders for milt and pituitary, and re-use of selected females in artificial propagation of *C. gariepinus* at hatcheries.

A total of four genetic clusters were inferred from the data, comprising the LVG/LKG, LTA, LBA and SAG populations. Of the four populations, SAG was the only farmed population forming a distinct

genetic cluster, which is also evident in the maximum likelihood tree (Fig. 2) and median spanning network (Fig. 3). This reflects a high divergence of the SAG population, which could be attributed to the possibility of multiple sources of the *C. gariepinus* in use at the farm. Identity of the populations making up the four genetic clusters was supported by membership coefficients, with only LVG, LTA, LBA and SAG comprising high memberships of respective samples from pre-defined populations (0.4483 to 0.9748). Overall, both markers have shown a clear distinction between the natural populations. The mitochondrial D-loop control region revealed a higher among-population genetic variation than did microsatellite markers, which were more informative for variation within populations.

Conclusions

In conclusion, a combination of the high resolution microsatellite markers and the rapidly evolving D-loop control region in this study has established high extant genetic variation in natural populations of *C. gariepinus* compared to farmed populations in Kenya. LVG/LKG, LTA, LBA, LJP and SAG are significantly different, at both markers, and therefore constitute important catfish genetic resources that could be exploited for higher aquaculture production in Kenya. For instance, the populations could be used in selective breeding or genetic improvement programmes to develop a base population with several gene pools. These populations should therefore be managed separately, and inter basin translocation of the populations should be avoided in order to maintain the genetic distinctness of each. A high number of haplotypes in the study samples were private, indicating uniqueness, which could be conserved by avoiding translocation of natural populations of *C. gariepinus* across drainage basins. However, Kapkuikui area of Lake Baringo drainage with intensive aquaculture activities (Ndiwa *et al.* 2016), has been recently stocked with *C. gariepinus* from Dominion Farms Limited (Barasa *et al.* 2016), a multinational company reclaiming the Yala swamp, which hosts Lake Kanyaboli, for commercial food production. The translocation could erode the genetic distinctness of *C. gariepinus* of Lake Baringo.

Maintaining a high number of brooders at hatcheries, and sound brood stock husbandry through better feeding regimes may prevent loss of genetic variation in farmed *C. gariepinus*. Reducing exploitation pressure on natural populations will help to maintain within-population genetic variation in populations of lower genetic variation such as LBA. Habitat destruction should also be avoided, to prevent negative impacts on populations. Unfortunately, Lake Baringo has recently been infested with the exotic water hyacinth, *Eichorrnia crassipes* (BARASA, personal observation), thought to have been translocated by fishermen relocating with their fishing nets from Lake Victoria, in search better fishing opportunities. Development of aquaculture in Africa has resulted from identification and improvement of natural genetic resources through marker-assisted selection (Lind *et al.* 2012), and our findings provide support for the need to protect Kenya's natural fish genetic resources and reservoirs of genetic diversity.

Acknowledgements

Mitochondrial DNA analysis was done at the Biosciences eastern and central Africa-International Livestock Research Institute Hub (BecA-ILRI Hub) laboratories in Nairobi, Kenya. We acknowledge the excellent technical support by the Segolip, the BecA-ILRI Hub sequencing and genotyping facility. Financial assistance to BecA-ILRI Hub by Syngenta Foundation for sustainable Agriculture (SFSA), the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership, the UK Department for International Development (DFID), the Bill & Melinda Gates Foundation (BMGF) and the Swedish International Development Agency (SIDA) is appreciated. Microsatellite DNA analysis was carried out at the Department of Genetics, University of the Free State, South Africa. This study was supported through the Kenya-South Africa bilateral Scientific and Research co-operation (SAK/2nd call) (NACOSTI-Kenya grant to R.A. and an NRF- South Africa grant to J.P.G). J.E.B held a PhD Science, Technology and Innovation research grant (NACOSTI/5/003/3rd call) from NACOSTI-Kenya. The manuscript was developed during a research fellowship of J.E.B at the Royal Belgian Institute

of Natural Sciences and the University of Antwerp supported by VLIR-UOS, Belgium. We thank the management of the Fish farms for allowing the sampling of their catfish brood stock for this study. Comments by four anonymous reviewers greatly improved the original manuscript.

References

Alam M.S. & Islam M.S. (2005). Population genetic structure of the major Indian carp *Catla catla* (Hamilton) revealed by microsatellite DNA markers. *Aquaculture* 246: 151–160. https://doi.org/10.1016/j.aquaculture.2005.02.012

Allendorf R.W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* 5: 181–190. https://doi.org/10.1002/zoo.1430050212

Allendorf F.W. & Waples R.S. (1996). Conservation and genetics of salmonid fishes. *In*: Avise J.C. & Hamrick J.L. (eds) *Conservation Genetics: Case Histories from Nature*: 238–501. New York, Chapman and Hall.

Aloo P.A. (2003). Biological diversity of Yala swamp lakes, with special emphasis on fish species composition, in relation to changes in the Lake Victoria basin Kenya: threats and conservation measures. *Biodiversity Conservation* 12: 905–920. https://doi.org/10.1023/A:1022869624524

Anene S.N. & Tianxiang G. (2007). Is the Dutch domesticated strain of *Clarias gariepinus* (Burchell, 1822) a hybrid? *African Journal of Biotechnology* 6 (8): 1072–1076.

Bandelt H.J., Forster P. & Röhl A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Barasa J.E., Abila R., Grobler J.P., Dangasuk O.G., Njahira M.N. & Kaunda-Arara B. (2014). Genetic diversity and gene flow in *Clarias gariepinus* from Lakes Victoria and Kanyaboli, Kenya. *African Journal of Aquatic Science* 39 (3): 287–293. https://doi.org/10.2989/16085914.2014.933734

Barasa J.E., Abila R., Grobler J.P., Agaba M., Chemoiwa E.J. & Kaunda-Arara B. (2016). High genetic diversity and population differentiation in *Clarias gariepinus* (Burchell, 1822) of Yala swamp: Evidence from mitochondrial DNA sequences. *Journal of Fish Biology* 89 (6): 2557–2570. https://doi.org/10.2989/16085914.2014.933734

Barluenga M. & Meyer A. (2010). Phylogeography, colonization and population history of the Midas species complex (*Amphilophus* spp.) in the Nicaraguan crater lakes. *BMC Evolutionary Biology* 10: 326–346. https://doi.org/10.1186/1471-2148-10-326

Beadle L.C. (1974). *The Inland Waters of Tropical Africa: An Introduction to Tropical Limnology*. Longman, London.

Bessem I., Verschuren D., Russell J.M., Hus J., Mees F. & Cumming B.F. (2008). Palaeolimnological evidence for widespread late 18th century drought across equatorial East Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology* 259: 107–120. https://doi.org/10.1016/j.palaeo.2007.10.002

Bruton M.N. (1988). Systematics and biology of clariid catfish. *In*: Hecht T., Uys W. & Brit P.J. (eds) *The culture of sharptooth catfish in southern Africa*. South African National Scientific Programmes Report no. 153. J.L.B. Smith Institute of Ichthyology.

Cambray J.A. & Van Der Waal B.C.W. (2006). Dutch domesticated *Clarias* with mixed genetic background now used in aquaculture in South Africa, with unpredictable consequences for biodiversity. *African Journal of Aquatic Science* 31 (1): 151–153.

CASTRIC V., BERNATCHEZ L., BELKHIR L. & BONHOMME F. (2002). Heterozygote deficiencies in small lacustrine populations of brook charr *Salvelinus fontinalis* Mitchill (Pisces, Salmonidae): a test of alternative hypotheses. *Heredity* 89: 27–35.

CHEPKIRUI-BOIT V., NGUGI C.C., BOWMAN J., OYOO E.O., RASOWO J., BUNDI J.M. & CHEROP L. (2011). Growth performance, survival, feed utilization and nutrient utilization of the African catfish, *Clarias gariepinus* co-fed Artemia and a micro-diet containing atyid shrimp (*Caridina nilotica*) during weaning. *Aquaculture Nutrition* 17: e82–e89. https://doi.org/10.1111/j.1365-2095.2009.00737.x

CHITAMWEBWA D., KAMANYI J., KAYUNGI J., NABBONGO H., OGOLLA A. & OJUOK J. (2009). The present status of the hook fishery and its impact on fish stocks of Lake Victoria. *African Journal of Tropical Hydrobiology and Fisheries* 12: 78–82.

CUNNINGHAM E.P. & MEGHEN C.M. (2001). Biological identification systems: genetic markers. *Scientific and Technical Review of the Office International des Epizooties* 20 (2): 491–499.

DE GRAAF G.J., GALEMONI F. & BANZOUSSI B. (1995). Artificial reproduction and fingerling production of the African catfish, *Clarias gariepinus* in protected and unprotected ponds. *Aquaculture Research* 26: 233–242. https://doi.org/10.1111/j.1365-2109.1995.tb00908.x

DUMONT H.J. (1986). The Nile River system. *In*: DAVIS B.R. & WALKERS K.F. (eds) *The Ecology of River Systems*: 61–74. W. Junk Publishers, Dordrecht.

EARL D.A. & VONHOLDT B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361. https://doi.org/10.1007/s12686-011-9548-7

EKNATH A.E., TAYAMEN M.M., PALADA DE-VERA M.S., DANTING J.C., REYES R.A., DIONISIO E.E., CAPILI J.B., BOLIVAR H.L., ABELLA T.A., CIRCA A.V., BENTSEN H.B., GJERDE B., GJEDREM T. & PULLIN R.S.V. (1993). Genetic improvement of farmed tilapias: the growth performance of 8 strains of *Oreochromis niloticus* tested in different farm environments. *Aquaculture* 111: 171–188. https://doi.org/10.1016/0044-8486(93)90035-W

EVANNO G., REGNAUT S. & GODET J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

EXCOFFIER L., LAVAL G. & SCHNEIDER S. (2005). *Arlequin: an integrated software package for population genetics data analysis. Version 3.0. Computational and Molecular Population Genetics Laboratory (CMPGL)*. Institute of Zoology, University of Berne, Berne (Switzerland).

FISHERIES FRAME SURVEY (2006). Implementation of a Fisheries Management Plan. *In*: *National Report on the Lake Victoria (Kenya) EDF Project no. 8 ACP ROR 029*: 8–24. Ministry of Livestock and Fisheries Development, Nairobi.

FRANKLIN I.R. & FRANKHARM R. (1998). How large must populations be to retain evolutionary potential? *Animal conservation* 1: 69–73. https://doi.org/10.1111/j.1469-1795.1998.tb00228.x

GALBUSERA P., VOLCKAERT F.A.M., HELLEMANS B. & OLLEVIER F. (1996). Isolation and characterization of microsatellite markers in the African catfish, *Clarias gariepinus* (Burchell 1822). *Molecular Ecology* 5: 703–705. https://doi.org/10.1111/j.1365-294X.1996.tb00366.x

GIDDELO C.S., ARNDT A.D. & VOLCKAERT F.A.M. (2002). Impact of rifting and hydrography on the genetic structure of *Clarias gariepinus* in eastern Africa. *Journal of Fish Biology* 60: 1252–1266. https://doi.org/10.1111/j.1095-8649.2002.tb01718.x

GOUDSWAARD K.P.C. & WITTE F. (1997). The catfish fauna of Lake Victoria after the Nile perch upsurge. *Environmental Biology of Fishes* 49: 21–43. https://doi.org/10.1023/A:1007311708377

Grobler J.P., Hoffman L.C. & Prinsloo J.F. (1997). A comparison of allozyme heterozygosity and life history variables in four strains of African catfish *Clarias gariepinus*. *South African Journal of Aquatic Science* 23 (1): 31–41. https://doi.org/10.1080/10183469.1997.9631386

Hansen M.M., Simonsen V., Mensberg K.L.D., Sarder M.D.R.I. & Ala M.D.S. (2006). Loss of genetic variation in hatchery-reared Indian major carp. *Journal of Fish Biology* 69 (Supplement B): 229–241. https://doi.org/10.1111/j.1095-8649.2006.01285.x

Hatfield T. & Schluter D. (1999). Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* 53: 866–873.

Hedrick P.W. & Kalinowski S.T. (2000). Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics* 31: 139–162.

Hogendoorn H. & Vismans M.M. (1980). Controlled propagation of the African catfish, *Clarias lazera*. II. Artificial reproduction. *Aquaculture* 21: 39–53. https://doi.org/10.1016/0044-8486(80)90124-6

Huff D.D., Miller M.L., Chizinski J.C. & Bruce V. (2011). Mixed-source reintroductions lead to out-breeding depression in second-generation descendants of a native North-American fish. *Molecular Ecology* 20: 4246–4258. https://doi.org/10.1111/j.1365-294X.2011.05271.x

Izquierdo M.S., Fernandez-Palacios H. & Tacon A.G.J. (2001). Effect of brood stock nutrition on reproductive performance of fish. *Aquaculture* 197: 25–42. https://doi.org/10.1016/S0044-8486(01)00581-6

Kaufman L. & Ochumba P. (1993). Evolutionary and conservation biology of cichlid fishes as revealed by faunal remnants in the Northern Lake Victoria. *Conservation Biology* 7 (3): 719–730. https://doi.org/10.1046/j.1523-1739.1993.07030719.x

Knaepkens G., Knapen D., Bervoets L., Hanfling B., Verheyen E. & Eens M. (2002). Genetic diversity and condition factor: a significant relationship in Flemish but not in German populations of the European bullhead (*Cottus gobio*). *Heredity* 89: 280–287.

Kocher T.D., Lee W.J., Soboleska H., Penman D. & Mcandrew B. (1998). A genetic linkage map of a cichlid fish, the Tilapia (*Oreochromis niloticus*). *Genetics* 148: 1225–1232.

Librado P. & Rozas J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lind C.E., Brummett R.E. & Ponzoni R.W. (2012). Exploitation and conservation of fish genetic resources in Africa: issues and priorities for aquaculture development and research. *Reviews in Aquaculture* 4: 125–141. https://doi.org/10.1111/j.1753-5131.2012.01068.x

Macharia S.K., Ngugi C.C. & Rasowo J. (2005). Comparative study of hatching rates of African catfish (*Clarias gariepinus* Burchell 1822) eggs on different substrates. *NAGA, World Fish Center Quarterly* 28 (3–4): 23–27.

Mcclelland E.K. & Naish K.A. (2007). What is the fitness outcome of crossing unrelated fish populations? A meta-analysis and an evaluation of future research directions. *Conservation Genetics* 8: 397–416. https://doi.org/10.1007/s10592-006-9178-x

Meyer A. (1994). DNA technology and phylogeny of fish. *In*: Beaumont A.R. (ed.) *Molecular Systematics and Evolution of Marine Organisms*: 219–248. Chapman and Hall, London.

Mkumbo O.C. & Mlaponi E. (2007). Impact of the baited hook fishery on the recovering endemic fish species in Lake Victoria. *Aquatic Ecosystem Health and Management* 10 (4): 458–466. https://doi.org/10.1080/14634980701704197

MUSA S.M., AURA C.M., NGUGI C.C. & KUNDU R. (2012). The effect of three different feed types on growth performance and survival of *Clarias gariepinus* fry reared in a hatchery. *ISRN Zoology* 2012: 1–6. https://doi.org/10.5402/2012/861364

NAZIA A.K., SUZANA M., AZHAR H., NGUYEN THUY T.T. & SITI AZIZAH M.N. (2010). No genetic differentiation between geographically isolated populations of *Clarias macrocephalus* in Malaysia revealed by sequences of mtDNA cytochrome b and D-loop gene regions. *Journal of Applied Ichthyology* 26: 568–570. https://doi.org/10.1111/j.1439-0426.2010.01469.x

NDIWA T.C., NYINGI D.W., CLAUDE J. & AGNESE F. (2016). Morphological variations of wild populations of Nile tilapia (*Oreochromis niloticus*) living in extreme environmental conditions in the Kenyan Rift-Valley. *Environmental Biology of Fishes* 99: 4773–485. https://doi.org/10.1007/s10641-016-0492-y

NGUGI C.C., AMADIVA J., VEVERICA K.L., BOWMAN J., IMENDE S., NYANDAT B. & MATOLLA G. (2003). *On-farm trials in Kenya change attitudes of fish farmers and extensionists*. Samaki News I.

NORRIS A.T., BRADLEY D.G. & CUNNINGHAM E.P. (1999). Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture* 180: 247–264. https://doi.org/10.1016/S0044-8486(99)00212-4

NYINA-WAMWIZA L., WATHELET B., RICHIR J., ROLLIN X. & KESTEMONT P. (2010). Partial or total replacement of fish meal by local agricultural by-products in diets of juvenile African catfish: growth performance, feed efficiency and digestibility. *Aquaculture Nutrition* 16: 237–247. https://doi.org/10.1111/j.1365-2095.2009.00658.x

PARK S.D.E. (2001). *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection*. PhD Thesis, University of Dublin, Dublin.

PARTRIDGE T.C., WOOD B.A. & DEMENOCAL P.B. (1995). The influence of global climate change and regional uplift on large-mammalian evolution in East and southern Africa. *In*: VRBA E.S., DENTON G.H., PARTRIDGE G.C. & BURCKLE L.H. (eds) *Paleoclimate and Evolution, with Emphasis on Human Origins*: 331–355. Yale University Press, London and New Haven.

PEAKALL R. & SMOUSE P.E. (2012). GenALEx 6.5. Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

POLICANSKY D. & MAGNUSON J.J. (1998). Genetics, metapopulations and ecosystem management of fisheries. *Ecological Applications* 8 (supplement 1): S119–S123. https://doi.org/10.1890/1051-0761(1998)8[S119:GMAEMO]2.0.CO;2

PONZONI R.W. & NGUYEN N.H. (2008). Foreword. *In*: PONZONI R.W. & NGUYEN N.H. (eds) *Proceedings of a workshop on the development of a Genetic improvement programme for the African catfish,* Clarias gariepinus*, Accra, Ghana, 5–9th November, 2007*: vi. The WorldFish Center, Penang (Malaysia).

PONZONI R.W., NGUYENI N.H. & KHAW H.I. (2009). Genetic improvement programmes for aquaculture species in the developing countries: Prospects and challenges. *In*: *Proceedings of the 18th Conference of the Association for Advancement of Animal breeding and Genetics, September 2011, Perth, Australia*: 342–349. AAABG, Roseworthy, South Australia. Available from http://www.aaabg.org/proceedings18/files/ponzoni342.pdf [accessed 8 August 2017].

POPOOLA M.O., FASAKIN E.A. & AWOPETU J.I. (2014). Genetic variability in wild and cultured populations of *Clarias gariepinus* using Random Amplified Polymorphic DNA (RAPD) markers. *Croatian Journal of Fisheries* 72 (1): 5–11. https://doi.org/10.14798/72.1.712

POSADA D.A. & CRANDALL K.A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

Pritchard J.K., Stephens M. & Donnelly P. (2000). Inference of population structure using multi-locus genotype data. *Genetics* 155: 945–959.

Rasowo J., Auma E., Ssanyu G. & Ndunguru M. (2008). Does African catfish (*Clarias gariepinus*) affect rice in integrated rice-fish culture in Lake Victoria Basin, Kisumu? *African Journal of Environmental Science and Technology* 2 (10): 336–341.

Roodt-Wilding R., Swart B.L. & Impson N.D. (2010). Genetically distinct Dutch domesticated *Clarias gariepinus* used in aquaculture in southern Africa. *African Journal of Aquatic Science* 35: 241–249. https://doi.org/10.2989/16085914.2010.538507

State Department of Fisheries (2013). Fisheries statistical Bulletin, 2013. Ministry of Agriculture, Livestock and Fisheries, Nairobi.

Sukmanomon S., Senanan W., Kapuscinski A.R. & Na-Nakorn U. (2012). Genetic diversity of feral populations of Nile tilapia (*Oreochromis niloticus*) in Thailand and evidence of genetic introgression. *Kasetsart Journal of Natural Sciences* 46: 200–216.

Sulem Y.S., Tomedi E.T., Mounchili S., Tekeng S. & Brummet R.S. (2006). Survival of *Clarias gariepinus* fry in earthen ponds: effects of composts and leaks. *Aquaculture* 260 (1–4): 139–144. https://doi.org/10.1016/j.aquaculture.2006.06.008

Tamura K., Dudley J., Nei M. & Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596–1599. https://doi.org/10.1093/molbev/msm092

Teichert-Coddington D.R. & Smitherman D.R. (1988). Lack of response by *Tilapia nilotica* to mass selection for rapid early growth. *Transactions of the American Fisheries Society* 117: 227–300. https://doi.org/10.1577/1548-8659(1988)117<0297:LORBNT>2.3.CO;2

Triantafyllidis A., Abatzopoulos T.J., & Economidis P.S. (1999). Genetic differentiation and phylogenetic relationships among Greek *Silurus glanis* and *Sirulus aristotelis* populations, assessed by PCR-RFLP analysis of mitochondrial DNA segments. *Heredity* 82: 503–509.

Uthice S. & Benzie A.H. (2003). Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology* 12: 2635–2648. https://doi.org/10.1046/j.1365-294X.2003.01954.x

Van Der Bank F.H., Grobler J.P. & Du Preez H.H. (1992). A comparative biochemical genetic study of three populations of domesticated and wild African Catfish, *Clarias gariepinus*. *Comparative Biochemistry and Physiology* 101B: 387–390. https://doi.org/10.1016/0305-0491(92)90017-L

Van Der Walt L.D., Van Der Bank F.H. & Steyn G.J. (1993). The suitability of using cryopreservation of spermatozoa for the conservation of genetic diversity in African catfish, *Clarias gariepinus*. *Comparative Biochemistry and Physiology* 106 (2): 313–318. https://doi.org/10.1016/0300-9629(93)90519-A

Verschuren D., Laird K.R. & Cumming B.F. (2000). Rainfall and drought in equatorial East Africa during the past 1,100 years. *Nature* 403: 410–413. https://doi.org/10.1038/35000179

Vrijenhoek R.C. (1998). Conservation genetics of freshwater fish. *Journal of Fish Biology* 53 (Supplement A): 394–412. https://doi.org/10.1111/j.1095-8649.1998.tb01039.x

Wachirachaikarn A., Rungsin W., Srisapoome P. & Na-Nakorn U. (2009). Crossing of African Catfish, *Clarias gariepinus* (Burchell, 1822), strains based on strain selection using genetic diversity data. *Aquaculture* 290: 53–60. https://doi.org/10.1016/j.aquaculture.2009.01.036

Waldbieser G.C., Bosworth B.G., Nonneman D.J. & Wolters W.R. (2001). A microsatellite based genetic linkage map for channel catfish, *Ictalurus punctatus*. *Genetics* 158: 727–734.

WITTE F. & VAN DENSEN W.L.T. (1995). *Fish Stocks and Fisheries of Lake Victoria: a Handbook for Field Observations*. Samara Publishing Limited, Dyfed, UK.

WITTE F., GOLDSCHMIDT T., LIGTVOET W., OIJEN M.J.P. & WANINK J.H. (1992). Species extinction and concomitant ecological changes in Lake Victoria. *Netherlands Journal of Zoology* 42: 214–232.

WRIGHT S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19: 395–420. https://doi.org/10.2307/2406450

YUE G.H., KOVACS B. & ORBAN L. (2003). Microsatellites from *Clarias batrachus* and their polymorphism in seven additional catfish species. *Molecular Ecology Notes* 3: 465–468. https://doi.org/10.1046/j.1471-8286.2003.00486.x

Manuscript received: 28 December 2015
Manuscript accepted: 2 June 2017
Published on: 28 September 2017
Branch editor: Peter Galbusera

View publication stats