

Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya

Geraldine D. Kavembe · Gonzalo Machado-Schiaffino · Axel Meyer

Abstract Lake Magadi, an alkaline hypersaline lake in Kenya, is one of the most extreme water bodies known. Although its water temperatures often exceed 40°C, a particular lineage of ‘dwarf’ tilapia, *Alcolapia grahami*, has evolved remarkable adaptations to survive in this hostile environment. Magadi tilapia exists in small fragmented populations in isolated lagoons within Lake Magadi and its satellite Lake, Little Magadi. In spite of the potential this tilapia holds for understanding evolutionary processes in stressful environments, few genetic studies have focused on this species. We examined the genetic diversity and spatial genetic relationships of Magadi tilapia populations using microsatellite and mitochondrial markers. High levels of genetic variation were found to be supporting the hypothesis that *A. grahami* populations represent remnants of a much larger fish population that inhabited paleo-lake Orolonga. In contrast to previous studies, we found a well-supported genetic structure of *A. grahami* consisting of three differentiated genetic clusters (a) Little Magadi, (b) Fish Spring Lagoon and (c) Rest of Magadi. Given the

importance of this species to the Magadi ecosystem and its potential evolutionary significance, the three genetic clusters should be considered as separate gene pools and conservation strategies aimed at protecting the species based on these clusters are recommended.

Keywords Paleo-lake Orolonga · Lake Natron · Trona · *Alcolapia grahami* · Evolution · Microsatellites · mtDNA

Introduction

Saline lakes are estimated to constitute 85,000 km³ of global water compared to 105,000 km³ of freshwater lakes (Shiklomanov, 1990). Varying in their size, salinity, ionic and biotic compositions, saline lakes have significant economic, ecological, biodiversity and cultural value (Williams, 2002). Ecologically, salt lake ecosystems are considered as keystone habitats that offer refuge to various endemic and unique species adapted to the harsh hypersaline water conditions (Jellison et al., 2004). Despite their ecological significance and their size that almost equals the freshwater resources, saline aquatic systems have received limited attention and data on the composition and distribution of their biota are often scarce.

Lake Magadi in Kenya is a shallow hypersaline basin located to the East of the African Great Rift Valley. Together with Lake Natron another hypersaline lake in Tanzania, the Magadi lake basin is considered a remnant

Guest editors: T. von Rintelen, R. M. Marwoto, G. D. Haffner & F. Herder / Speciation in Ancient Lakes – Classic Concepts and New Approaches

G. D. Kavembe · G. Machado-Schiaffino · A. Meyer (✉)
Lehrstuhl für Zoologie und Evolutionsbiologie,
Department of Biology, University of Konstanz,
Universitätstrasse 10, 78457 Constance, Germany
e-mail: axel.meyer@uni-konstanz.de

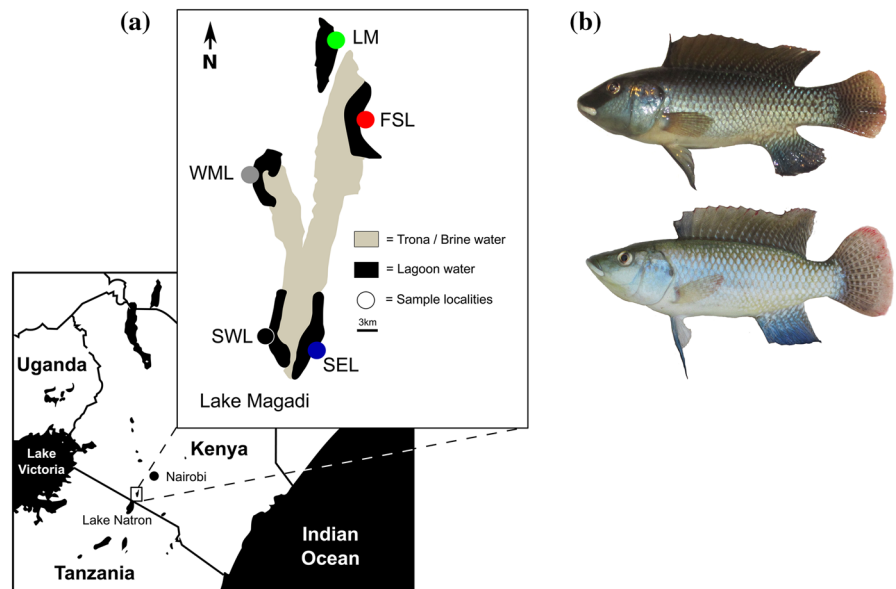
of an old, less alkaline Pleistocene Lake Orolonga that divided and formed the two current lake basins (Fig. 1a) $\sim 13,000$ yr BP (Behr, 2002; c.f. Seegers & Tichy, 1999). It is thought that Lake Magadi further split forming the satellite lake, Little Magadi north of it following adverse climatic conditions in the early Holocene $\sim 7,000$ yr BP (Seegers & Tichy, 1999). Presently, the Magadi lake basin is a closed system with no obvious inflow and outflow and is mainly recharged by occasional torrential rains and geothermal springs found along the margins of the lakes (Bützer et al., 1972). Fast and high rates of water evaporation within the main Lake Magadi have led to the formation of huge deposits of trona (hydrated sodium carbonate salt, $\text{Na}_3(\text{CO}_3)(\text{HCO}_3)\cdot 2\text{H}_2\text{O}$) as well as some pools of anoxic water (brine) that altogether cover about 80% of the lake surface. By comparison, Little Magadi's surface area is less broken up by accumulations of trona.

In spite of its extremely hostile conditions, a unique fauna is found in the Magadi lake basin; notably, a single endemic Magadi tilapia (*Alcolapia grahami*, Boulenger, 1912) thrives in several isolated alkaline lagoons along the shores of the lake. Although there is only one described species, two different mouth forms; a terminal mouth form and a slanting mouth form have been described in the main Lake Magadi and Little Magadi, respectively (Fig. 1b) (Wilson et al., 2004). In contrast, Lake Natron, located a few kilometres south of Lake

Magadi, harbours three closely related but morphologically distinguishable tilapia species: *A. alcalicus*, *A. ndalalani* and *A. latilabris* (Seegers & Tichy, 1999). To survive in the physicochemical extreme parameters associated with their habitat, Magadi tilapia has developed several unique mechanisms that ensure a balance between growth and reproduction (Pörtner et al., 2010). Among the most unusual characteristics of this species is its ability to excrete all nitrogen waste in form of urea (not ammonia, as in most fish) thus enabling the fish to survive in the highly alkaline ($\sim \text{pH}10$) and buffered ($\text{CO}_2 \sim 180 \text{ mm}^{-1}$) Magadi water (Pörtner et al., 2010). The species is listed as vulnerable by the World Conservation Union's (IUCN) Red List of threatened species (Bayona & Akinyi, 2006); yet very little is known about its population dynamics, genetic composition or phenotypic aspects of the diversity within the species (Wilson et al., 2004).

Owing to the patchy distribution of habitable lagoons within the Magadi Lake basin and the associated fluctuations of the physicochemical parameters of the Magadi water (Wilson et al., 2004), Magadi tilapia offers an exceptional opportunity to study the effect of geographical isolation on population genetic structures. One might predict that habitat loss and fragmentation, as was the case with the ancestral Magadi tilapia population, potentially affected the population size and geographical genetic structure and may have shaped the

Fig. 1 a Map of Magadi lake basin showing the lagoons (highlighted in black). Exact sampling locations are indicated by small circles (see Table 1 for geographical coordinates and the corresponding population colour codes). A map of Kenya showing the position of Lake Magadi, Little Magadi and Lake Natron. b The two morphs of Magadi tilapia (Lake Magadi morph at the top and Little Magadi at the bottom).



genetic composition of the resultant populations. Adverse effects of fragmentation may as well lead to loss of genetic diversity due to small effective population sizes, loss of allelic variation due to inbreeding, patterns of gene flow, all of which might lead to a decline in genetic variation and possibly fitness and could potentially even lead to local extinction (Sterling et al., 2012). Such effects are, of course, predicted to have adverse effects on small isolated populations living under stressful conditions where local selective pressures might differ and be particularly strong (Fraser et al., 2011).

Mitochondrial (mtDNA) haplotype data collected so far does not provide evidence for strong genetic differentiation between the populations of Magadi tilapia. Wilson et al. (2004), using 350 bp of the mtDNA control region on samples drawn from the two lakes within the Magadi lake basin and a single location in Lake Natron, found a very weak genetic structure among the populations studied. Haplotypes were shared among lagoons and lakes. The individual haplotypes could all be traced to a single ancestral haplotype that was probably present in the populations that originally inhabited the paleo-lake Orolonga (Wilson et al., 2004). Their findings are consistent with findings of a previous study conducted mainly based on Lake Natron species and a small subset of Lake Magadi tilapia (Seegers et al., 1999). This latter study reported substantial sharing of haplotypes, for both mtDNA control region and cytochrome b, among the Natron and Magadi tilapia, supporting the hypothesis of a common ancestry from the paleo-lake Orolonga (Seegers et al., 1999; Wilson et al., 2004). However, an earlier study by Wilson et al. (2000) based on minisatellites using samples from only two lagoons of the main Lake Magadi and a subset of samples from a single location in Lake Natron reported substantial genetic differentiation between the three populations studied. The difference in the studies by Wilson et al. (2000, 2004) could be attributed to the low resolution power of the markers and the limited number of samples used (Wan et al., 2004).

Given the contrasting results generated by previous studies on the genetic structure of *Alcolapia grahami*, fine-scale genetic studies are crucial to reveal the actual geographical genetic structure of the species. In the present study, we used a combination of markers with varying mutation rates including microsatellites,

mitochondrial ND2 and the complete control region to investigate the genetic consequences of habitat fragmentation in populations of *Alcolapia grahami*. More specifically, we sought to quantify the genetic diversity and to determine the genetic structure of the extant populations of Magadi tilapia. The relevance of trona and land barriers in restricting gene flow among Magadi tilapia populations was also assessed.

Materials and methods

Sampling and DNA extraction

Magadi tilapia fish were collected ($n = 259$) using seine nets from four different lagoons in the main Lake Magadi and one location in Little Magadi (Table 1, Fig. 1a). Fin clips and whenever possible muscle samples were collected and stored in 95% ethanol for DNA analysis. Genomic DNA was extracted by sodium chloride/ethanol precipitation following proteinase K digestion (Bruford et al., 1998).

Haplotype sequencing and microsatellite genotyping

Approximately 889-bp-long fragment of the mitochondrial control region was amplified in 91 samples (Table 1) using primers LproF (5'-aacctcacccttagctcccaaag-3') (Meyer et al., 1994) and 12S5R (5'-ggcggatactgcatgt-3') (Hrbek & Farias, 2008). Additionally, a 1,042 bp fragment of the mtDNA NADH-dehydrogenase subunit 2 (ND2) was amplified in a similar number of samples (Table 1). The ND2 marker was amplified in two fragments using primer pairs ND2Met (5'-catacccaaacatgttgg-3') and ND2B (5'-tggttaatccgctca-3') for the initial fragment and primer pair ND2.2A (5'-ctgacaaaactgccccctt-3') and ND2Trp (5'-gagatttctactccgctta-3') (Kocher et al., 1995) for the rest of the fragment. PCR amplifications and sequencing reactions were carried out following methods similar to those previously described (Klett & Meyer, 2002; Angienda et al., 2011). The purified sequencing reactions were run on a 3130xl DNA Analyzer (Applied Biosystems) and analysed with ABI PRISM DNA Sequencing Analysis Software v5.3.1. Both forward and reverse strands of each individual sample were sequenced to crosscheck for sequencing errors.

Table 1 Sampling localities of Magadi tilapia and their individual geographical coordinates

Locality	Abbreviation	Population colour code	Geographical location		<i>N</i> (Microsatellites)	<i>N</i> (Mitochondrial—control/ND2)
			Latitude	Longitude		
Little Magadi	LM	Green	S 01° 43' 39.3''	E 036° 16' 50.7''	62	17
Fish spring lagoon	FSL	Red	S 01° 53' 52.1''	E 036° 18' 15.7''	51	24
South East lagoon	SEL	Blue	S 01° 59' 38.0''	E 036° 13' 34.7''	50	18
South West lagoon	SWL	Black	S 02° 0' 03.9''	E 036° 13' 55.2''	47	16
West Magadi Lagoon	WML	Grey	S 01° 52' 20.2''	E036° 14' 39.7''	49	16
Total					259	91

The abbreviations of the populations derived from the corresponding localities (see Fig. 1a) as well as the total number of individuals (*N*) analysed per marker per population are indicated

Ten polymorphic microsatellite loci (that were developed and optimised in cichlids) were amplified using standard PCR conditions: Abur04, Abur30, Abur110, Tmom27, UNH002 (Sanetra et al., 2009) and UNH843, UNH874, UNH891, UNH915 and UNH989 (Carleton et al., 2002). These loci were used to genotype 259 individuals (Table 1). The forward primer of each microsatellite loci pair was labelled with a FAM or HEX fluorescent dye (Applied Biosystems). Up to four different loci were pooled and fragmented using a 96-capillary automatic sequencer ABI3130XL (Applied Biosystems, CA, USA) with an internal Rox-size standard. Microsatellite alleles were scored using GENEMAPPER v 4.0 (Applied Biosystems).

mtDNA sequence analysis

The sequences were assembled into contigs using SeqMan v7.2 (DNASTAR, Inc., 2007). Individual sequences were then edited and aligned using ClustalW (Larkin et al., 2007) as implemented in Bioedit v7.1.3 (Hall, 1999). As a final control, the sequences were manually inspected by eye. All sequences have been deposited in Genbank (Accession numbers KC572497-KC572527 and KC572533-KC572546). Unique haplotypes were identified using DnaSP v5.10 (Librado & Rozas, 2009). Median joining (Bandelt et al., 1999) haplotype networks for both mtDNA markers were generated using NETWORK v4.6.1.1 (Fluxus Technology Ltd, 1999–2012). The levels of contemporary genetic diversity in Magadi tilapia populations were determined by calculating the number of polymorphic sites; and number of haplotypes (*Ha*), haplotype (*H*) and

nucleotide (π) diversities of individual populations using ARLEQUIN v3.11 (Excoffier et al., 2005). Sequence partitioning of the ND2 sequences was performed using PartitionFinder v1.0.1 (Lanfear et al., 2012) to determine the best partition scheme under the Akaike Information Criterion (AICc). Based on codon positions, three partition schemes were assumed namely: a) each position separately (1) (2) (3), b) the first and second positions grouped, (12) (3) and c) all the three positions together (i.e., (123)). Additionally, concatenated sequences ($n = 91$) consisting of both the ND2 and the control region combined into a single alignment of 1,931 bp were analysed in PartitionFinder v1.0.1 (Lanfear et al., 2012). For this concatenated alignment, the partition schemes tested included: (a) each marker separately, (b) coding (partitioned ND2) and noncoding (control region) markers and (c) no partition. The programme also served to test the best-fit model of nucleotide substitution for the individual datasets. Separately, the best-fit models of nucleotide substitution for the partitioned datasets and the control region were tested under the Akaike Information Criteria (AIC) as implemented in MetaPiga v2.0 (Helaers & Milinkovitch, 2010) and jmodeltest v2.0 (Darriba et al., 2012), respectively. Phylogenies were reconstructed using maximum-likelihood (ML) algorithms as follows: control region using MEGA v5.1 (Tamura et al., 2011) and the partitioned ND2 and the concatenated sequences using MetaPiga v2 (Helaers & Milinkovitch, 2010). For the MetaPiga analysis, starting trees were constructed using loose neighbour joining and consensus pruning. The trees were rooted using *Sarotherodon melanotheron*.

To infer the position of *Alcolapia grahami* in relation to other tilapia lineages, phylogenetic analysis using maximum likelihood approach was performed as implemented in MetaPiga v2 (Helaers & Milinkovitch, 2010). In previous studies, *Alcolapia alcalicus* (a sister species found in Lake Natron) has been shown to group closely with members of the genus *Oreochromis* followed by those of genus *Sarotherodon* (Nagy et al., 2001; Schwarzer et al., 2009; Dunz & Schliewen, 2013). Thus, based on these findings, publicly available ND2 sequences of the genus *Oreochromis* (accession: AF317230-AF317242 and AF317246 (Klett & Meyer, 2002)) and representative sequences of the genus *Sarotherodon* (accession AF317243-AF317251 (Klett & Meyer, 2002) and GQ16711 (Schwarzer et al., 2009)) were used to reconstruct the phylogenetic position of *Alcolapia grahami*. Four of the most common ND2 haplotypes from our study and the only published ND2 sequence of *A. alcalicus* of L. Natron (accession GQ1678781 (Schwarzer et al., 2009)) were also included. For this analysis, a sequence alignment of 979 bp ($n = 23$) was used. *Astolotilapia burtoni*, a distantly related species was used to root the tree. The robustness of all the phylogenetic trees was tested with 1,000 bootstrap replicates.

Microsatellites data analysis

MICRO-CHECKER v2.2.3 (van Oosterhout et al., 2004) was used to test for the presence of null alleles and/or scoring errors and to evaluate their possible impacts on the estimated values of genetic differentiation. To assess the microsatellite genetic diversity, the number of alleles per locus (N_a), average allelic richness (A_r), observed (H_o) and expected heterozygosities (H_e) and Wright's F_{IS} parameter (Wright, 1978) were calculated using FSTAT v2.9.3.2 (Goudet, 1995). Departures from Hardy-Weinberg equilibrium (HWE) at each locus and linkage disequilibrium between loci were tested using GENEPOP v4.2 (Raymond & Rousset, 1995). Significance levels were estimated using Markov chain permutations with 10,000 steps and 1,000 dememorization steps. Sequential Bonferroni corrections were applied in all multiple comparisons to correct for the statistical significance level (Rice, 1989).

Population genetic structure

To assess the genetic admixture within the Magadi tilapia, three different approaches were used. First,

indices of population differentiation F_{ST} (Wright, 1965) and R_{ST} (Slatkin, 1995) were estimated using FSTAT (Goudet, 1995) with a significance of 10,000 permutations. Second, a Bayesian model-based clustering algorithm implemented in STRUCTURE v2.3.4 (Pritchard et al., 2000) was used to infer the number of genetic clusters (K) in the dataset for all Magadi tilapia samples. The software was run under the admixture model, allowing for some mixed ancestry within individuals. For each K value ($k = 1$ to $k = 7$), 20 independent runs were performed, with a burn-in period of 50,000 iterations and 500,000 replications. We employed the correction method described by Evanno et al. (2005) to determine the most likely number of clusters. Third, in order to assess how the differentiation among the a priori defined populations (based on the sampling locations) and those based on the Bayesian clusters (generated with structure) contributed to the global genetic variation, a hierarchical AMOVA analysis (Excoffier et al., 1992) was performed using both mtDNA control region and microsatellite markers. The significance of the AMOVA test was assessed through 999 permutations. Finally, a Mantel test (Mantel, 1967) was performed to assess the correlation between spatial (geographical) and genetic distances, as implemented in Isolation by Distance (IBD) v3.23 (Jensen et al., 2005). Geographical distances among populations were calculated following the nearest waterway and estimated in Google earth v5.1 (Google Corporation, 2007).

Inferring the historical demography of *Alcolapia grahami*

Signatures of population demographical changes (bottlenecks or expansions) in *Alcolapia grahami* populations were examined first by Tajima's D (Tajima, 1980) and Fu's F (Fu, 1997) statistics using DnaSP v5.0 (Librado & Rozas, 2009) based on 1,000 coalescent simulations and Garza & Williamsons (2001) M statistic as implemented in ARLEQUIN v3.11 (Excoffier et al., 2005). In accordance with Garza & Williamson (2001), a critical M value of 0.68 was used to identify signatures of genetic bottleneck.

Demographical changes in Magadi tilapia were also examined by estimating the sum of squared deviations (SSD) and the raggedness index (Harpending, 1994) as implemented in ARLEQUIN 3.11 (Excoffier et al.,

2005). These statistics are used to test the goodness-of-fit of observed mismatch distributions to the theoretical distribution under a model of exponential population growth (Rogers & Harpending, 1992). Populations that have undergone demographical expansions are expected to have low values of squared differences (SSD) and smaller raggedness index in their mismatch distributions than non-expanded populations. Higher values of SSD and raggedness index suggest static or bottlenecked populations (Harpending, 1994).

Estimates of recent migration rates and levels of gene flow between populations were computed based on the genetic clusters identified by STRUCTURE analysis using a coalescent-based approach as implemented in MIGRATE v3.2.6 (Beerli & Felsenstein, 2001). As programme settings, we employed a stepwise mutation model (Brownian motion approximation) and default settings for other parameters. For each run, starting estimates for Theta (Θ) were based on F_{ST} values, with burn-in 10,000 trees, 100 short chains with a total of 10,000 genealogies sampled, and three long chains with one million genealogies sampled, for each locus. Adaptive Chain heating, with four different temperatures, was employed to get an efficient exploration of the data.

Results

Using the jModeltest v2 (Darriba et al., 2012) the model HKY85 (Hasegawa et al., 1985) with rate heterogeneity and a proportion of invariant sites (HKY + G + I) was selected as the best-fit model for the mtDNA control region. Using PartitionFinder v1.0.1 (Lanfear et al., 2012), the scheme that assumed each ND2 codon position separately was selected as the best. Thus for this partitioned dataset, the model HKY + G was selected for the first and second codon position datasets and HKY was selected for the 3rd codon dataset, respectively. Similarly the scheme that assumed each marker separately was selected for the concatenated DNA sequence alignment and the model HKY + G + I was identified as the best-fit model for both markers. The best-fit models selected for each dataset were then used for the subsequent analyses.

Phylogenetic analyses

Phylogenetic reconstruction to infer the position of *A. grahami* using already published ND2 sequences of

Oreochromis and *Sarotherodon* species, identified *Oreochromis variabilis* (Boulenger, 1906) as the most closely related extant relative to the genus *Alcolapia* with a strong bootstrap support of 96% (Fig. 2). *O. variabilis* is a tilapia species endemic to Lake Victoria and presently listed in the IUCN Red List of threatened species as critically endangered (Maithya et al., 2012).

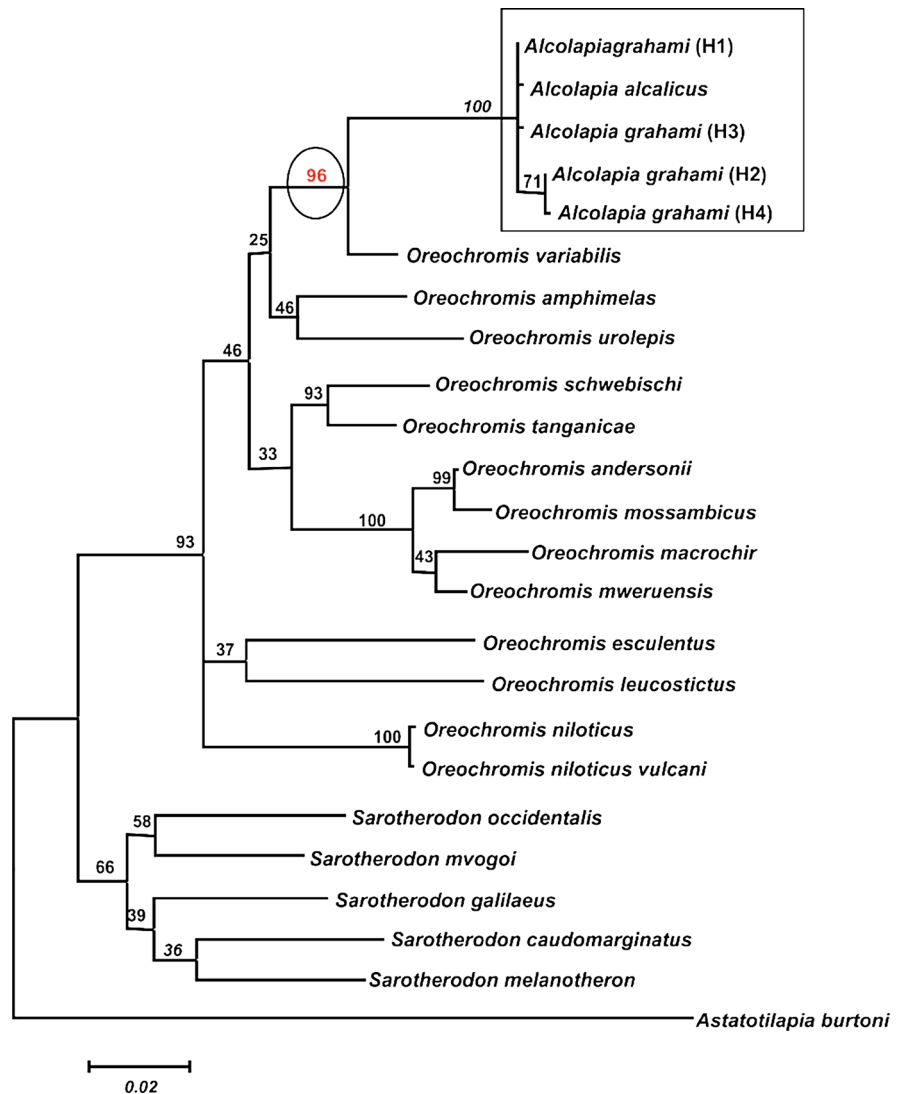
Separately, phylogenetic analysis performed using *Alcolapia grahami* ND2 sequences generated in the current study detected subdivision of the species into two main clades; one for each of the Magadi lakes. Using this marker, the single haplotype of Little Magadi population formed a basal clade (bootstrap support 88%). The second clade consisted of haplotypes exclusive to populations of the main Lake Magadi (Fig. 3a). However, using the fast-evolving control region sequences, a slightly different tree topology was obtained. Two clades were observed; one clade consisted of haplotypes exclusive to the tilapia of the main Lake Magadi (bootstrap value 84%) whereas the other clade consisted of haplotypes drawn from both lakes with a relatively low bootstrap support (68%) (Fig. 3b). Consistent with the control region results, the phylogenetic tree inferred from the concatenated sequence data (1,931 bp) identified two clades; one clade with haplotypes exclusive to the main Lake Magadi and the other clade consisted of haplotypes from both lakes. The clade that had haplotypes drawn from both lakes was poorly (bootstrap 19%) supported whereas the clade that had haplotypes exclusive to the main Lake Magadi was highly (bootstrap 90%) supported (tree not shown). The differences in the observed topology of the mtDNA phylogenies could be explained by the presence of specific Little Magadi haplotypes and the sharing of haplotypes among the other populations of the main Lake Magadi (Fig. 3a, b).

Genetic diversity

mtDNA

High levels of genetic variation were found using both the mitochondrial markers. A total of 31 mtDNA control region haplotypes differing in 25 positions were identified in the 91 samples analysed (Table 2, Fig. 4a). Overall, high levels of haplotype diversity (0.929 ± 0.014) and low levels of nucleotide diversity (0.006 ± 0.000) were detected. Consistent with its

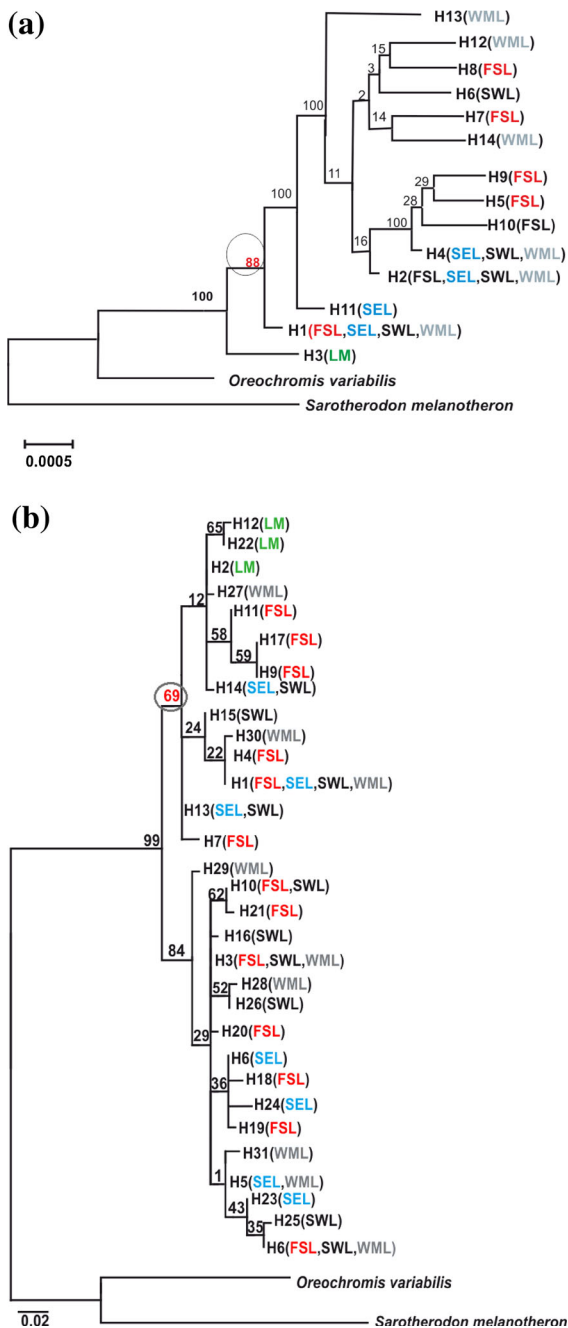
Fig. 2 Maximum likelihood phylogenetic tree reconstructed to infer the position of the genus *Alcolapia* relative to other closely related genera. Numbers on the branches are percentage bootstrap values (1,000 replicates). The tree was generated with MetaPiga v2 (Helaers & Milinkovitch, 2010). *Oreochromis variabilis* grouped closely (bootstrap support highlighted by a circle) with the *Alcolapia* species (highlighted in a box)



young age (7,000 years), Little Magadi tilapia population (Table 2) had the lowest values of haplotype and nucleotide diversities. In contrast, tilapia populations of the main Lake Magadi had relatively high estimated values that were almost similar among the populations (Table 2). Haplotype one (H1) was the most common and widespread haplotype; representing 17.6% of all the samples analysed but was notably absent in Little Magadi (Fig. 4a). Haplotype two (H2) was found in 82.4% of the samples from Little Magadi and it represented 15.4% of all the samples analysed. The other two samples of Little Magadi had unique haplotypes (H12 and H22) that were separated from the common Little Magadi haplotype by single

substitutions (Fig. 4a). Several lagoon-specific mtDNA control region haplotypes were found in other populations of the main Lake Magadi (Fig. 4a). The absence of haplotype two (H2) from all other populations distinguishes the Little Magadi population from all the other populations of Magadi tilapia (Table 4).

Using the ND2 marker, 14 haplotypes were found for an alignment of 1,042 bp in the 91 samples analysed (Table 2). Overall, haplotype (h) and nucleotide diversities (π) for all sequences were 0.782 (± 0.025) and 0.001 (± 0.008), respectively. Haplotype one (H1) was the most common representing 36% of all samples while haplotype three (H3) was fixed for



◀ **Fig. 3** Phylogenetic tree reconstruction of Magadi tilapia using maximum likelihood: **a** ND2 and **b** control region. Numbers on the branches are percentage bootstrap values (1,000 replicates). Each haplotype has been tagged with the name (s) of representative populations for easy reference of the haplotype distribution among the populations. The population names *colour codes* correspond to the colours used in haplotype network construction. The trees were generated using MetaPiga v2 (Helaers & Milinkovitch, 2010) and MEGA v5.1 (Tamura et al., 2011), respectively. Bootstrap support for the Little Magadi haplotypes is *circled*

Microsatellites

MICRO-CHECKER analysis gave no evidence of allelic dropout, null alleles or scoring errors due to stuttering. All the 10 microsatellite loci were highly polymorphic. The average number of alleles (N_a) ranged from three (locus Abur30 in FSL) to 35 (locus UNH989 in WML). The allelic richness (A_r) ranged from 2.27 (locus Abur30 in FSL) to 29.84 (locus TmoM27 in SWL). Again, the Little Magadi population was the least variable ($N_a = 10$; $A_r = 13$) while South West Lagoon population was the most variable ($N_a = 16.5$; $A_r = 24$). Most of the loci were under Hardy–Weinberg equilibrium; no excess or deficiency of heterozygotes was detected at all loci and populations. F_{IS} values were also not significant for all loci and populations ($P > 0.05$) (Table 3).

Population structure of *Alcolapia grahmi*

In general, F_{ST} values for all populations indicated moderate to high levels of inter-population structuring (Tables 4, 5). All pairwise comparisons involving Little Magadi population showed the highest F_{ST} values, ranging from 0.118 to 0.170 for microsatellites, 0.200 to 0.563 for control region and 0.546 to 0.752 for ND2. Significant differences were also found when Fish Spring Lagoon (FSL) was compared with any other population of the main Lake Magadi (Tables 4 and 5). Based on mtDNA data, we found no genetic differentiation between populations from South East Lagoon (SEL), South West Lagoon (SWL) and West Magadi Lagoon (WML) (Table 4). Although significant, only slight differences were found within these three populations of the main Lake Magadi based on microsatellite data (Table 5).

The model-based clustering method implemented in STRUCTURE v2.3.4 (Pritchard et al., 2000) found

Little Magadi (Fig. 4b). As was the case with the mtDNA control region, lagoon-specific haplotypes were discovered using the ND2 marker as well. Due to the highly conserved nature of the ND2 gene and the limited resolution obtained from this marker (Table 2), further inference about mtDNA data will be mainly based on the control region data.

Table 2 mtDNA summary statistics of Magadi tilapia populations

Pop	<i>N</i>	HA	NP	<i>K</i>	Hd	π	Tajima's D (<i>P</i> -value)	Fu's F (<i>P</i> -value)	SSD value (<i>P</i> -value)	Raggedness index (<i>P</i> -value)
Control										
LM	17	3	2	0.529	0.323	0.001	-0.262 (0.3397)	-0.227 (0.311)	0.211 (0.110)	0.362 (0.05)
FSL	24	13	17	5.489	0.902	0.006	0.737 (0.820)	-2.25 (0.156)	0.041 (0.26)	0.043 (0.480)
SEL	18	7	13	4.922	0.83	0.005	1.127 (0.907)	1.27 (0.717)	0.065 (0.130)	0.173 (0.030)
SWL	16	10	14	5.2	0.942	0.006	0.904 (0.862)	-0.1665 (0.179)	0.072 (0.070)	0.081 (0.320)
WML	16	9	15	5.108	0.9	0.006	0.509 (0.730)	-0.781 (0.365)	0.053 (0.160)	0.067 (0.440)
ND2										
LM	17	1	0	0	0	0	0.000 (1.000)	0.000 (0.000)	0.000 (0.00)	0.000 (0.000)
FSL	24	7	10	1.985	0.677	0.002	-0.864 (0.198)	-0.863 (0.353)	0.065 (0.13)	0.250 (0.08)
SEL	18	4	3	1.137	0.628	0.001	0.842 (0.788)	0.843 (0.533)	0.067 (0.130)	0.249 (0.110)
SWL	16	4	4	1.533	0.767	0.005	0.851 (0.809)	0.851 (0.686)	0.120 (0.137)	0.070 (0.590)
WML	16	6	66	1.566	0.717	0.001	-0.457 (0.361)	-0.457 (0.157)	0.035 (0.26)	0.129 (0.430)

Pop population, *N* number of sequences, *HA* number of haplotypes, *NP* number of polymorphic sites, *K* average number of nucleotide differences, *Hd* haplotype diversity, π nucleotide diversity; *SSD* sum of squared deviations; *P* < 0.05

that the most probable number of genetic clusters of *A. grahami* is $k = 3$ (Fig. 5). Little Magadi (LM) population forms the first distinct cluster, Fish Spring Lagoon (FSL) population forms another and the last cluster consisted of the rest of the populations of the main Lake Magadi (RLM). These results were consistent for each of the 20 independent runs performed. With regard to the Little Magadi population, the results are in accordance with the F_{ST} results. If the number of expected genetic clusters (k) is allowed to vary, it is apparent that individuals from the allopatric Little Magadi will form a most distinct genetic cluster separate from those of the main Lake Magadi. However, assuming three different genetic clusters, FSL is clearly separate from the RLM populations raising a number of questions concerning the extent of population sub-structuring of *A. grahami* within the main Lake Magadi. Furthermore, the clustering of three populations within the main Lake Magadi into a single genetic cluster could point to current or recent gene flow between the respective lagoon populations.

Hierarchical AMOVA analysis examining the partitioning of genetic variance among the different hypothesized groupings of Magadi tilapia found that most of the genetic variation arose from differences within populations followed by variance among groups and the least amount of variation arose from differences among populations (Table 6). None of the computations yielded statistically significant values

for the support of the assumed population groupings. As expected, the support for the variation resulting from among population differences considering three groupings based on mitochondrial data was not significant due to the sharing of haplotypes among tilapia populations within the main Lake Magadi.

Finally, even though strong genetic differentiation was suggested from the pairwise F_{ST} comparisons, no evidence of isolation by distance was detected using the Mantel test (Fig. 6). These results are consistent with patterns observed in species that exhibit patchy distribution.

Historical demography of *Alcolapia grahami*

Several tests performed to detect genetic signatures of demographical changes such as Tajima's D and Fu's F tests, mismatch analysis and raggedness indices yielded non-significant results for most populations indicating no signal of past or recent population size changes in Magadi tilapia (Table 2). Moreover, tests for recent bottlenecks using Garza and William's statistic (*M*) yielded values that were higher than the critical value of 0.68 consistent with natural populations that have not undergone recent reduction in population size (Table 3) (Garza & Williamsons, 2001).

Generally, low levels of gene flow were predicted by MIGRATE (Beerli & Felsenstein, 2001). Approximately one to three migrants per population as well as

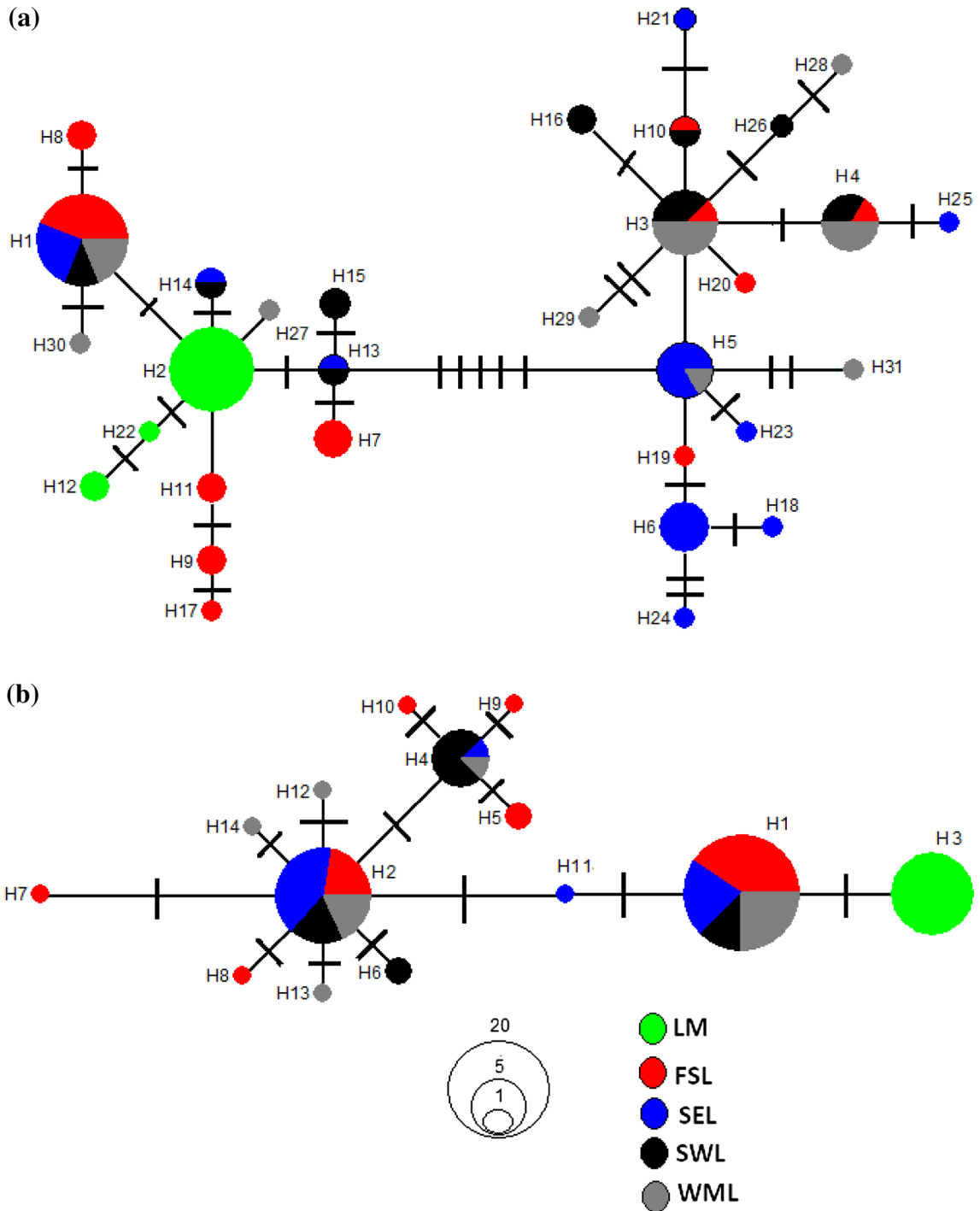


Fig. 4 Median joining haplotype networks of *Alcolapia* mtDNA. **a** Control region and **b** ND2 sequences. Haplotypes are represented as *circles*, with the size of each circle

proportional to the haplotype's overall frequency. *Short line strokes* represent the number of nucleotide sequence changes between the individual haplotypes

Table 3 Summary statistics for 10 microsatellite loci amplified in Magadi tilapia populations

Pop	<i>N</i>	MNA	(<i>Av_{AR}</i>)	Hobs	Hexp	(<i>Av_{GD}</i>)	<i>F_{IS}</i> (<i>P</i> -value)	GW (<i>M</i>)	HWE
LM	51	10.20	13	0.607	0.629	0.595	0.008 (0.4159)	0.794	ns
FSL	62	15.80	21.7	0.773	0.806	0.733	−0.001 (0.520)	0.7414	ns
SEL	50	15.50	23.4	0.752	0.761	0.663	−0.065 (0.998)	0.690	ns
SWL	47	16.50	24	0.759	0.808	0.691	−0.015 (0.797)	0.739	ns
WML	49	16.20	24.2	0.765	0.761	0.649	−0.091 (0.999)	0.680	ns

Pop population, *N* number of samples per population, *MNA* mean number of alleles, *Av_{AR}* mean allelic range, *H_{obs}* observed heterozygosity, *H_{exp}* expected heterozygosity, *Av_{GD}* average gene diversity, *F_{IS}* Inbreeding coefficient, *GW (M)* Garza and Williams' Statistic, *HWE* Hardy–Weinberg equilibrium

Table 4 *F_{ST}* for mtDNA ND2 (above diagonal) and control region (below diagonal) between Magadi tilapia populations

Population	LM	FSL	SEL	SWL	WML
LM	–	0.546**	0.738**	0.751**	0.659**
FSL	0.200**	–	−0.010 ^{NS}	0.078 ^{NS}	−0.029 ^{NS}
SEL	0.549**	0.153*	–	0.068 ^{NS}	−0.040 ^{NS}
SWL	0.552**	0.146 ^{NS}	−0.077 ^{NS}	–	0.089 ^{NS}
WML	0.563**	0.153 ^{NS}	−0.016 ^{NS}	−0.049 ^{NS}	–

(**P* < 0.05, ***P* < 0.001) Significant after sequential Bonferroni correction, *NS* non-significant

Table 5 Microsatellite genetic differentiation among Magadi tilapia populations: *R_{ST}* (above diagonal) and *F_{ST}* (below diagonal)

	LM	FSL	SEL	SWL	WML
LM	–	0.195**	0.231**	0.187**	0.197**
FSL	0.136**	–	0.048*	0.044**	0.094**
SEL	0.141**	0.037**	–	0.021 ^{NS}	0.026*
SWL	0.118**	0.020**	0.016*	–	−0.002 ^{NS}
WML	0.141**	0.050*	0.037*	0.029*	–

overlapping values of effective migrants at 5% and 95% confidence intervals were detected among all population pairs considered (Table 7). The highest proportion of migrants (3.44) was found for computations involving migrations from the Rest of Lake Magadi (RLM) (source population) to Fish Spring Lagoon (FSL) (recipient population) while the lowest proportion of migrants (1.88) was found between Little Magadi (LM) (source population) and Rest of Lake Magadi (RLM) (recipient population). Consistent with the genetic structure and the *F_{ST}* analysis, MIGRATE (Beerli & Felsenstein, 2001) estimated very low proportions of migrants moving from LM to any other lagoon within the main Lake Magadi (RLM). FSL also had much lower estimates of migrants to other

lagoons compared to RLM. Restricted movement of fish probably facilitated by trona or unfavourable water conditions could explain the low levels of gene flow from FSL population to other lagoon populations within the main Lake Magadi (RLM) while the presence of land barrier between Little Magadi and the main Lake Magadi populations provides a strong barrier to gene flow.

Discussion

This study provides information from sensitive genetic markers on the genetic composition and connectivity of *Alcolapia grahami* populations throughout its distribution range. By employing several markers with different mutation rates, this study allowed for investigation of various genetic parameters to infer the fine-scale genetic structure of Magadi tilapia. Lack of monophyletic association of *A. grahami* populations especially with reference to Little Magadi tilapia is reported. In contrast to a weak genetic structure proposed in Magadi tilapia by previous studies that examined the genetic aspects of the species, the current study revealed a well-supported genetic structure consisting of three strongly differentiated genetic

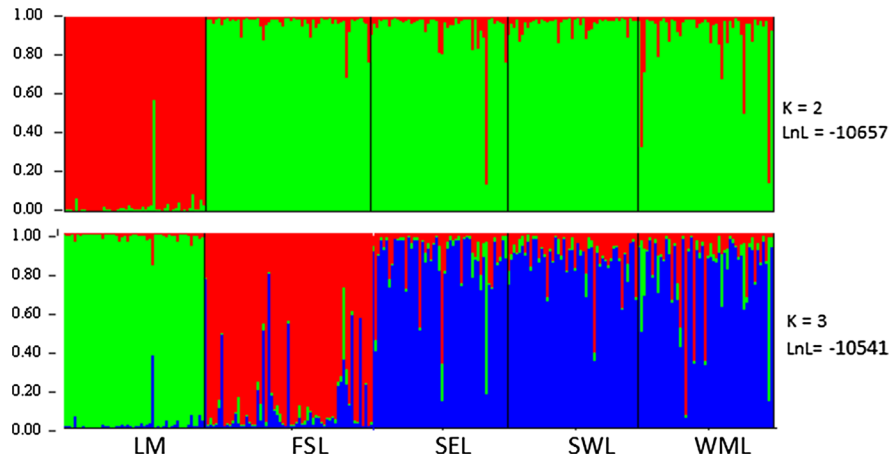


Fig. 5 Results of the analysis of the genetic clustering of Magadi tilapia using STRUCTURE. Abbreviations of sampling localities are indicated in Table 1 and Fig. 1. Colours correspond to the different genetic clusters estimated by the analysis: each individual has a probability of being assigned to a given cluster that is proportional to the height of that coloured bar; the

clusters; Little Magadi (LM), Fish Spring Lagoon (FSL) and the Rest of Lake Magadi (RLM).

Phylogenetic relationship of Magadi tilapia to other tilapiines

The phylogenetic placement of *Alcolapia* has been elusive (Nagy et al., 2001). Based on the first fossils dated to 700,000 years and other recent fossils, it is generally agreed that a cichlid fish that inhabited the paleo-lake Orolonga formed the ancestral stock of the *Alcolapia* species flock (Isaac, 1967; Coe, 1969). Several recent studies further assigned *Alcolapia* to the genus *Oreochromis* (Trewavas, 1983) and later to the genus *Alcolapia* (Seegers & Tichy, 1999), but the phylogenetic position of the *Alcolapia* species remained unresolved as reflected in their repeatedly revised taxonomy (Seegers et al., 1999). Due to the proximity of the Magadi-Natron basin to Lake Manyara, another saline lake in Tanzania, it has been proposed that its endemic species, *Oreochromis amphimelas*, that is also morphologically similar to the Lake Natron species (*A. alcalicus*), is closely related to the *Alcolapia* species. However, two studies (Seegers et al., 1999; Nagy et al., 2001) using mitochondrial control region do not support such relationship with several *Oreochromis* lineages appearing between the two species in the phylogenetic tree. Nagy et al. (2001)

more uniform the colour of the bar, the more probable it is that the individual is entirely composed of genetic material composed of the given cluster. The log likelihood of each assumed number of populations ($k = 2$ and $k = 3$) is shown. Population abbreviations are shown in Table 1

proposed that the sharing of similar phenotypes and behaviour between *Oreochromis amphimelas* and the *Alcolapia* species could be a case of convergent evolution driven by the need to survive in similar environments. Our finding of *O. variabilis*, an endangered tilapia species endemic to Lake Victoria, as the closest relative to *A. grahami* complicates the puzzle given the fact that the two lakes are separated by many kilometres with no obvious water connection pathways. However, evidence from past geological and hydrological patterns indicates the presence of several rivers and intermediate lakes during the periods when East Africa experienced frequent and heavy rains (Becht et al., 2006). These rivers and intermediate lakes could have provided the opportunity for fish from Lake Victoria to move far south of the Great Rift to colonise lakes including the paleo-lake Orolonga.

Genetic diversity of Magadi tilapia

Consistent with past records, that Magadi tilapia represent remnants of an old fish population of paleo-lake Orolonga, high levels of genetic diversity were observed with all markers and in all populations studied. The high number of haplotypes, high mtDNA haplotypic diversity, low mtDNA nucleotide diversity and high variation in microsatellite markers support the association of the present-day tilapia to an old large fish population.

Table 6 AMOVA analysis examining the partitioning of genetic variance at three hierarchical levels

	df	Variation (%)	F statistic	P-value
Microsatellites				
Two groups (LM vs. Rest of Lake Magadi)				
Among groups	1	8.50	0.116	0.193
Among populations	3	3.10	0.033	0.000**
Within populations	513	88.4	0.116	0.000**
Three groups (LM vs. FSL vs. Rest of Lake Magadi)				
Among groups	2	5.99	0.059	0.101
Among populations	2	2.31	0.024	0.000**
Within populations	513	91.70	0.083	0.000**
mtDNA control				
Two groups (LM vs. Rest of Lake Magadi)				
Among groups	1	28.83	0.288	0.191
Among populations	3	8.10	0.114	0.015*
Within populations	86	63.07	0.369	0.000**
Three groups (LM vs. FSL vs. Rest of Lake Magadi)				
Among groups	2	30.28	0.302	0.093
Among populations	2	-1.14	-0.016	0.636
Within populations	86	70.86	0.291	0.000**

Significant genetic variation is indicated by asterisks on the P-values (* $P < 0.05$, ** $P < 0.01$)

Polyphyly of Little Magadi as inferred from mtDNA control region

The mtDNA haplotype network clearly shows that the *Alcolapia* population of Little Magadi does not form a monophyletic group (Fig. 4a). Several scenarios may lead to such a pattern; incomplete lineage sorting resulting in shared polymorphisms, several repeated colonisation events, existence of hybridization and subsequent introgressive gene flow among groups or inadequate taxonomic data which underestimates the extant species diversity within a set of individuals (Sunnucks & Hales, 1996; Maddison & Knowles, 2006). Magadi tilapia is considered a relic of a fish that inhabited paleo-lake Orolonga which split to form the present-day Lake Magadi and Lake Natron over

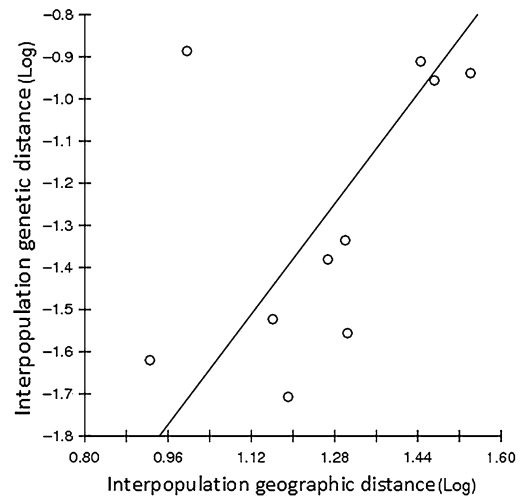


Fig. 6 Pairwise genetic differentiation by geographical distances (log transformed) indicate no pattern of isolation-by-distance among populations of Magadi tilapia ($r^2 = 0.229$, $P = 0.342$)

13,000 years ago. Subsequently, the last major geological processes that led to the current geography of the Magadi basin (i.e., the complete separation of Lake Magadi and Little Magadi) are dated to over 7,000 years ago (Bützer et al., 1972). This time may have been sufficient to have allowed populations in individual lagoons to differentiate and diverge into unique populations. Similar time period has been found in Crater Lake cichlids in Nicaragua, to be sufficient even for the origin of new species (Barluenga et al., 2006; Kautt et al., 2012).

Although the route and timing of the colonisation history of the Magadi-Natron lake basin remains unclear, several periods of high water levels have been reported within the entire soda lake region and could have provided opportunities for exchange of fish between the three lakes. Soda lakes are still poorly explored systems in East Africa and may contain cryptic species since many of their species taxonomic delineation is based on possibly unrepresentative samples partly due to difficulties in sampling and inaccessibility of most of the populations' habitat (Jones et al., 1998). Accurate species delimitation, however, requires integration of information from various datasets including morphological, behavioural, genetic and thorough population sampling (Funk & Santiago, 2012) which was not within the scope of our study.

Table 7 Number of migrants per generation among the three genetic clusters of Magadi tilapia identified using Migrate software

Donor pop.	Recipient pop.		
	LM	FSL	RM
LM	–	1.914 (1.168–1.919)	1.885 (1.313–2.450)
FSL	2.044 (1.471–2.616)	–	2.364 (1.801–2.930)
Rest of Lake Magadi (RLM)	2.830 (2.230–3.416)	3.438 (2.479–4.366)	–

The values were calculated using average theta (θ) and M values from three independent runs. The values at 2.5 and 97.25% confidence interval are indicated in brackets

Population structure

Both the microsatellite and mtDNA analysis revealed high levels of genetic differentiation between the Magadi tilapia populations, particularly between LM and all other populations and between FSL compared to RLM. While the differentiation of the LM population could have been expected owing to its missing physical connection with the main Lake Magadi; the differentiation of FSL from RLM and the clustering of RLM tilapia populations was surprising. Trona has been proposed to provide sufficient barriers to gene flow but its relative contribution in preventing gene flow is still a matter of debate (Coe, 1969; Wilson et al., 2004). Fish Spring Lagoon is the largest lagoon within the main Lake Magadi and probably represents one of the oldest and most stable lagoon in terms of water levels. The lagoon is separated from the other lagoons by several kilometres of trona stretching along its shore on one side and is bordered by barren land on the other side. Increased anthropogenic activities mainly as a result of mining activities along the shores of Fish Spring Lagoon could also have remarkable effects on the lagoon's fish population. For instance, in mid-1900s following increased demand for water for industrial and human use, a retention wall was built along the shore of Fish Spring Lagoon to hold water. This could in effect prevent gene flow between Fish Spring Lagoon population and any other Magadi tilapia population. Although the effects of such anthropogenic activities on *A. grahami* have not been assessed, reduction in numbers of the Lesser Flamingo, a common bird that breeds on the Natron-Magadi basin, has been reported since the mid-1900s. This time period coincides with increased anthropogenic activities in the Lake Magadi basin (Mlingwa & Baker, 2006).

Lack of genetic differentiation among some tilapia populations within the main Lake Magadi could be

explained by lack of physical barriers coupled by the relatively short stretches of trona separating the lagoons. Infrequent floods on the Magadi basin could thus provide opportunities for gene exchange between lagoon populations that are separated by short stretches of trona (White, 1953). The presence of fish-eating birds and flamingos could also facilitate movement of fish between the lagoons and hence gene flow albeit in low levels. Nevertheless, very low levels of gene flow were found between populations further supporting the strong differentiation detected between the genetic clusters of Magadi tilapia.

Although an earlier study (Wilson et al., 2004) proposed two ecomorphs of Magadi tilapia based on their mouth position, lack of morphometric data in the present study could not allow for conclusive inferences to be made on the relationship of the ecomorphs and the genetic structure. Future studies should aim at profiling the molecular, morphological, ecological and behavioural characteristics of the Magadi tilapia populations to better understand the forces shaping their evolution.

Conclusion

In contrast to past studies that found only subtle genetic differentiation among *A. grahami* populations, our study revealed a pronounced genetic sub-structuring consisting of three genetic clusters; Little Magadi (LM), Fish Spring Lagoon (FSL) and the rest of Lake Magadi (RLM) seem to be clearly differentiated due to land barrier to gene flow in the case of the allopatric Little Magadi and trona and anthropogenic activities in the case of Fish Spring Lagoon. Given the strong differentiation among the Magadi tilapia populations and the associated low levels of gene flow, further studies aimed at investigating the evolutionary patterns

at the genome level can help to shed light on the evolution processes shaping the populations. Finally, based on the extent of genetic structuring and the associated genetic clusters identified in this study, management strategies aimed at protecting these unique gene pools (genetic clusters) of *A. grahami* are recommended.

Acknowledgments We are grateful to the National Council of Science and Technology (NCST) of Kenya for permission to conduct our work in Lake Magadi. We sincerely thank Chris M. Wood (NSERC Canada Discovery Grant), John Maina (National Research Foundation (NRF) of South Africa grant) and Adalberto Bianchini (International Canada Research Chair Program from the International Development Research Centre—IDRC grant) for important financial support to G. D. K. during sample acquisition. G.D.K is supported by a DAAD-NCST collaborative grant. Further support came from Alexander von Humboldt fellowship to G.M.S and support from the University of Konstanz to A.M. We thank Andreas Kautt and two anonymous reviewers for useful comments that greatly improved this manuscript. The Magadi team (H. Bergman, O. Ojoo, O. E. Ora Johannsson, M. Papah, P. Laurent, L. Bianchini and C. Chevalier) provided important advice during the initial stages of this study. Mr. Ndonga of the Magadi Company, D. Muthee and G. Muthee provided important logistical support.

References

- Angienda, P. O., H. J. Lee, K. Elmer, R. Abila, E. N. Waindi & A. Meyer, 2011. Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*) compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. *Conservation Genetics* 12: 243–255.
- Bandelt, H. J., P. Forster & A. Rohlf, 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.
- Barluenga, M., K. N. Stolting, W. Salzburger, M. Muschick & A. Meyer, 2006. Sympatric speciation in Nicaraguan Crater Lake cichlid fish. *Nature* 439: 719–723.
- Bayona, J. & E. Akinyi, 2006. *Alcolapia grahami*: IUCN 2011. IUCN Red list of Threatened Species.
- Becht, R., F. Mwango & F. A. Muno, 2006. Groundwater links between Kenyan rift valley lakes. In Odawa, E. O. et al. (eds), Proceedings of the 11th world lakes conference, 31 October–4th November 2005, Nairobi, Kenya. Ministry of Water and irrigation, Nairobi; International Lake Environment Committee (ILEC): 7–14.
- Berli, P. & J. Felsenstein, 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* 98: 4563–4568.
- Behr, H. J., 2002. Magadiite and Magadi chert: a critical analysis of the silica sediments in the Lake Magadi Basin, Kenya. *Society of Sedimentary Geology* 73: 257–273.
- Bruford, M. W., O. Hanotte, J. F. Y. Brookfield & T. Burke, 1998. Multi and single locus DNA fingerprinting. In Hoelzel, A. R. (ed.), *Molecular Analysis of Populations: A Practical Approach*, 2nd ed. IRL Press, Oxford: 287–336.
- Bützer, K. W., G. L. Isaac, J. L. Richardson & C. Washbourn-Kamau, 1972. Radiocarbon dating of East African lake levels. *Science* 175: 1069–1076.
- Carleton, K. L., J. T. Streebman, B. Y. Lee, N. Garnhart, M. Kidd & T. D. Kocher, 2002. Rapid isolation of CA microsatellites from the tilapia genome. *Animal Genetics* 33: 140–144.
- Coe, M. J., 1969. Observations on *Tilapia alcalica* in Lake Natron. *Revue de zoologie et de botanique africaines* 80: 1–14.
- Darriba, D., G. L. Taboada, R. Doallo & D. Posada, 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.
- DNASTAR, Inc. 2007. Lasergene SeqMan Pro (v7.2) Software suite for sequence analysis.
- Dunz, R. A. & U. K. Schlieffen, 2013. Molecular phylogeny and revised classification of the haplotilapine cichlid fishes formerly referred to as ‘‘Tilapia’’. *Molecular Phylogenetics and Evolution* 68: 64–80.
- Evanno, G., S. Regnaut & J. Goudet, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.
- Excoffier, L., P. E. Smouse & J. M. Quattro, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
- Excoffier, L., G. Laval & S. Schneider, 2005. ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.
- Fluxus Technology Ltd 1999–2012. Network v.4.6.1.1. <http://www.fluxus-engineering.com>.
- Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen & E. B. Taylor, 2011. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity* 106: 404–420.
- Fu, Y. X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
- Funk, M. C. & R. R. Santiago, 2012. High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of Royal Society B* 279: 1806–1814.
- Garza, J. C. & E. G. Williamson, 2001. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 10: 305–318.
- Google Corporation, 2007. Google Earth Version 5.1.
- Goudet, J., 1995. FSTAT (Version 1.2): a computer program to calculate F-statistics. *Heredity* 86: 485–486.
- Hall, T. A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
- Harpending, H. C., 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66: 591–600.
- Hasegawa, M., H. Kishino & T. Yano, 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *Molecular Evolution* 22: 160–174.

- Helaers, R. & M. C. Milinkovitch, 2010. MetaPIGA v2.0: maximum likelihood large phylogeny estimation using the metapopulation genetic algorithm and other stochastic heuristics. *BioMed Central. Bioinformatics* 11: 379.
- Hrbek, T. & I. P. Farias, 2008. The complete mitochondrial genome of the pirarucu (*Arapaima gigas*, Arapaimidae, Osteoglossiformes). *Genetics and Molecular Biology* 31: 293–302.
- Isaac, G. L., 1967. The stratigraphy of the Peninj Group—early middle Pleistocene formations west of Lake Natron, Tanzania. In Bishop, W. W. & J. D. Clark (eds), *Background to Evolution in Africa*. University of Chicago Press, Chicago: 229–257.
- Jellison, R., Y. S. Zadereev, P. A., DasSarma, J. M. Melack, M. R. Rosen, A. G. Degermendzhy, S. DasSarma & G. Zambana, 2004. Conservation and management challenges of saline lakes: a review of five experience briefs. *Lake Basin Management Initiative. Thematic Paper* 1–28.
- Jensen, J. L., A. J. Bohonak & S. T. Kelley, 2005. Isolation by distance v. 3.23, web service. *BioMed Central. Genetics* 6: 13.
- Jones, B. E., W. D. Grant, A. W. Duckworth & G. G. Owenson, 1998. Microbial diversity of soda lakes. *Extremophiles* 2: 191–200.
- Kautt, A. F., K. R. Elmer & A. Meyer, 2012. Genomic signatures of divergent selection and speciation patterns in a 'natural experiment', the young parallel radiations of Nicaraguan Crater Lake cichlid fishes. *Molecular Ecology* 21: 4770–4786.
- Klett, V. & A. Meyer, 2002. What, if anything, is a Tilapia? Mitochondrial ND2 phylogeny of Tilapiines and the evolution of parental care systems in the African cichlid fishes. *Molecular Biology and Evolution* 19: 865–883.
- Kocher, T. D., J. A. Conroy, K. R. McKaye, J. R. Stauffer & S. F. Lockwood, 1995. Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. *Molecular Phylogenetics and Evolution* 4: 420–432.
- Lanfear, R., B. Calcott, S. Y. Ho & S. Guindon, 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.
- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & D. G. Higgins, 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
- Librado, P. & J. Rozas, 2009. DnaSP v5: software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451.
- Maddison, W. P. & L. L. Knowles, 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.
- Maithya, J., M. Njiru, J. B. Okeyo-Owuor & J. Gichuki, 2012. Some aspects of the biology and life-history strategies of *Oreochromis variabilis* (Boulenger 1906) in the Lake Victoria Basin. *Lakes and Reservoirs: Research and Management* 17: 65–72.
- Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209–220.
- Meyer, A., J. M. Morrissey & M. Schartl, 1994. Recurrent origin of a sexually selected trait in Xiphophorus fishes inferred from a molecular phylogeny. *Nature* 368: 539–542.
- Mlingwa, C. & N. Baker, 2006. Lesser Flamingo *Phoenicopterus minor* counts in Tanzanian soda lakes: implications for conservation. In Boere, G. C., C. A. Galbraith & D. A. Stroud (eds), *Waterbirds Around the World*. The Stationery Office, Edinburgh, UK: 230–233.
- Nagy, S., H. Tichy, W. E. Mayer, I. E. Samonte, B. J. McAndrew & J. Klein, 2001. Classification and phylogenetic relationships of African Tilapiine fishes inferred from mitochondrial DNA sequences. *Phylogenetics and Evolution* 20: 361–374.
- Pörtner, H. O., P. M. Schulte, C. M. Wood & F. Schiemer, 2010. Niche dimensions in fishes: an integrative view. *Physiological Biochemistry and Zoology* 83: 808–826.
- Pritchard, J. K., M. Stephens & P. Donnelly, 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Raymond, M. & F. Rousset, 1995. Genepop: population genetics software or exact tests and ecumenism. *Heredity* 86: 248–249.
- Rice, W. R., 1989. Analyzing tables of statistical tests. *Evolution* 43: 223–225.
- Rogers, A. R. & H. Harpending, 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552–569.
- Sanetra, M., F. Henning, S. Fukamach & A. Meyer, 2009. A microsatellite-based genetic linkage map of the cichlid fish, *Astatotilapia burtoni* (Teleostei): a comparison of genomic architectures among rapidly speciating cichlids. *Genetics* 182: 387–397.
- Schwarzer, J., B. Misof, T. Diethard & U. K. Schlieven, 2009. The root of the East African cichlid radiations. *BioMed Central Evolutionary Biology* 9: 186.
- Seegers, L. & H. Tichy, 1999. The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with descriptions of two new species. *Ichthyological Exploration of Freshwaters* 10: 97–146.
- Seegers, L., R. Sonnenberg & R. Yamamoto, 1999. Molecular analysis of the *Alcolapia* flock from lakes Natron and Magadi, Tanzania and Kenya, with descriptions of two new species. *Ichthyological Exploration of Freshwaters* 10: 175–199.
- Shiklomanov, I. A., 1990. Global water resources. In: *Nature and Resources* (UNESCO), Vol. 26: 34–43.
- Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457–462.
- Sterling, K. A., D. H. Reed, B. P. Noonan & M. L. Warren, 2012. Genetic effects of habitat fragmentation and population isolation on *Etheostoma raneyi* (Percidae). *Conservation Genetics* 13: 859–872.
- Sunnucks, P. & D. Hales, 1996. Numerous transposed sequences of mitochondrial cytochrome-oxidase I-II in aphids of the genus *Sitobion* (Hemiptera, Aphididae). *Molecular Biology and Evolution* 13: 510–524.
- Tajima, F., 1980. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei & S. Kumar, 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.

- Trewavas, E., 1983. "Tilapiine Fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*," Bulletin of the British Museum (Natural History). London and Cornell & University Press, Ithaca, New York: 583.
- van Oosterhout, C., W. F. Hutchison, D. P. M. Wills & P. Shipley, 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.
- Wan, Q. H., H. Wu, T. Fujihara & S. G. Fang, 2004. Which genetic marker for which conservation genetics issue? *Electrophoresis* 25: 2165–2176.
- White, T. H., 1953. Some speculations on the sudden occurrence of floods in the history of Lake Magadi. *East Africa Natural History Society* 22: 69–71.
- Williams, W. D., 2002. Environmental threats to salt lakes and the likely status of inland saline ecosystems in 2025. *Environmental Conservation* 29: 154–167.
- Wilson, P. J., C. M. Wood, J. N. Maina & B. N. White, 2000. Genetic structure of Lake Magadi tilapia populations. *Fish Biology* 56: 590–603.
- Wilson, P. J., C. M. Wood, P. J. Walsh, A. N. Bergman, H. L. Bergman, P. Laurent & B. N. White, 2004. Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological Biochemistry and Zoology* 77: 537–555.
- Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19: 395–420.
- Wright, S., 1978. *Evolution and the Genetics of Populations*. Vol. 4. Variability within and among Natural Populations. University of Chicago Press, Chicago, IL, 580 pp.