Conserv Genet (2007) 8:475–488
DOI 10.1007/s10592-006-9186-x

ORIGINAL PAPER

Hybridization and phylogeography of the Mozambique tilapia *Oreochromis mossambicus* in southern Africa evidenced by mitochondrial and microsatellite DNA genotyping

María Eugenia D'Amato · Maria M. Esterhuyse · Ben C. W. van der Waal · D. Brink · Filip A. M. Volckaert

Received: 20 October 2005 / Accepted: 4 July 2006 / Published online: 14 December 2006

Abstract Many *Oreochromis* species utilized in aquaculture were extensively introduced outside their native range in Africa. Given their recent evolutionary radiation, these species hybridize easily, posing a threat to the integrity of local adaptation. The objective of this work was to study the genetic diversity of the Mozambique tilapia (*Oreochromis mossambicus*) in its native range, southern Africa, and provide a method for identifying hybrids with genetic markers. We genotyped the mitochondrial DNA (mtDNA) control region (385 bp) of wild and farmed *O. mossambicus*, wild and farmed *O. niloticus* and morphologic wild hybrids. These data were complemented with published sequences of parapatric and sympatric *Oreochromis* taxa. Phylogeographic analysis showed the presence of two *O. mossambicus* lineages, the southernmost representing a recent Holocene radiation. Hybridization of *O. mossambicus* was indicated by the presence of *O. niloticus* and *O. mortimeri–andersonii* mtDNA specimens in the Limpopo basin and of *O. karongae* mtDNA in specimens from Malawi. We also genotyped seven suspected hybrid individuals from the Limpopo River, and 137 wild and farmed Mozambique and Nile tilapia samples with five microsatellite markers. Factorial Component Analysis, Bayesian clustering and assignment analyses consistently delineated an *O. mossambicus* and an *O. niloticus* group, with the putative hybrids positioned in between. Different levels of hybridization were detected by the Bayesian assignment. The complex nature of hybridization and introgression between cichlid species raises major concerns for the long-term integrity of Mozambique tilapia.

Keywords Cichlidae · Mitochondrial control region · Exotics · Fish · Microsatellite DNA

M. E. D'Amato (✉) · M. M. Esterhuyse · D. Brink
Department of Genetics, University of Stellenbosch, Private Bag X1, 7602 Stellenbosch, South Africa
e-mail: medamato@uwc.ac.za

B. C. W. van der Waal
Department of Biological Sciences, University of Venda, PO Bag X 5050, Thohoyandou 0950, South Africa

F. A. M. Volckaert
Laboratory of Aquatic Ecology, Katholieke Universiteit Leuven, Ch. Deberiotstraat 32, B-3000 Leuven, Belgium

Present Address:
M. E. D'Amato
Department of Biotechnology, University of the Western Cape, Private Bag X17, 7535 Bellville, Cape Town, South Africa

Introduction

African cichlids are represented by two major groups: Haplochromines and Tilapiines. Tilapiine fishes are grouped into 10 genera by breeding habit (Trewavas 1983). The genera *Tilapia*, *Sarotherodon* and *Oreochromis* are generically known as tilapias. The mouthbreeding genus *Oreochromis* (31 species) occurs throughout the African continent and inhabits riverine and lacustrine habitats (Trewavas 1983). The recent evolutionary radiation of Haplochromines in the East African Great Lakes (Kocher 2004 and references therein) and palaeolake Makgadikgadi (Joyce et al. 2005) have been extensively studied; studies of phylogenetic relationships or phylogeographic patterns of Tilapiines are not as well represented (e.g., Nagl et al.

2000, 2001; Klett and Meyer 2002; Falk et al. 2003). As with other Cichlidae, Tilapiines underwent a recent evolutionary radiation. Recent or incomplete speciation processes allow them to hybridize readily, a quality that has been exploited for aquaculture purposes (Wohlfarth and Hulata 1981; Trewavas 1983). Well-established tilapias strains were produced by hybridizing *O. niloticus*, *O. aureus* and *O. mossambicus* (Romana-Eguia et al. 2004) among other species.

The adaptive value of hybridization varies between taxa (Barton 2001). When natural hybrids are less fit or viable than the parental species, secondary contact results in reinforcement of the reproductive isolation and represents a phylogenetic dead end. However, natural hybridization can also play an important role in facilitating adaptive radiation and evolution (see Seehausen 2004) by originating hybrid taxa, introgression, or establishment of a hybrid zone (Allendorf et al. 2001 and references therein). Gene flow between recently separated species is probably widespread in cichlids, and their mode of evolution may be better represented by a reticulate model rather than divergent dichotomies (Hey et al. 2004). Moreover, the hybrid origin of Lake Victoria cichlids was detected by the use of nuclear markers (Seehausen et al. 2003), challenging the original view of monophyletic origin detected by mtDNA (e.g., Meyer et al. 1990). Speciation by introgressive hybridization also seems to have contributed to rapid adaptive radiation in cichlids (Salzburger et al. 2002). Hybridization in the natural environment is poorly documented; it was only recorded for the sympatric *O. niloticus* and *O. aureus* in Western Africa (Rognon and Guyomard 2003). Escapees from aquaculture farms are a main concern for the conservation of *Oreochromis*. At least one case of extinction was recorded: in Lake Victoria where *O. variabilis* was out-competed and disappeared after introduction of Nile and Mozambique tilapia (Welcomme 1967).

O. niloticus, which is probably the most widely utilized cichlid for aquaculture, was extensively introduced throughout Africa, including SW South Africa and Natal (Skelton 1993), and Zambia and Zimbabwe (Schwank 1995). Hybridization with the introduced Nile Tilapia in South Africa has been documented for the Limpopo River (Van der Waal and Bills 2000; Moralee et al. 2000). *O. mossambicus* × *O. niloticus* hybrids were first described by Trewavas (1983); they are difficult to identify morphologically, as backcrosses resemble the parental species. Natural hybridization of *O. mossambicus* with other parapatric *Oreochromis* species is also expected from the ease with which Tilapiine fish seem to hybridize. Support for this hypothesis is provided by (1) the observation that hybridization between tilapia species generally produces viable progeny (Wohlfarth and Hulata 1981), (2) the sympatric hybridization of *Oreochromis* in Western Africa mentioned above, and (3) hybridization in Zimbabwe between *O. mossambicus* or *O. mortimeri* and the allopatric species *O. macrochir* in the Upper Zambezi (Gregg et al. 1998).

Our first aim was the identification of hybrids in the Limpopo River where anthropogenic hybridization has been taking place. Wild and farmed fish from the expected parental species *O. mossambicus* and *O. niloticus* were also genotyped to facilitate the examination of the extent of hybridization or introgression in individual fish. Our second aim was the delineation of Evolutionary Significant Units (ESUs) for *O. mossambicus*. Its geographic range extends from the lower Zambezi (Mozambique) to the Boesmans River (South Africa). *O. mossambicus* appears to have phylogenetic affinities to other *Oreochromis* species from central African lakes Tanganyika and Malawi (Nagl et al. 2001). Therefore we hypothesize that the most ancestral lineages in the *O. mossambicus* phylogeographic reconstruction may be represented in the northernmost part of the distribution range.

Methods

Samples

Samples were collected in southern Africa, northern Nile River, the Central Luzon State University research station (Nueva Ecija, the Philippines), and a fish farm in Thailand with a breeding program founded on fish from Lake Manzala, Egypt (Table 1 and Fig. 1). Initial species identification was based on morphological and morphometric traits observed on site (following Trewavas 1983; Skelton 1993, 2001; and observations of Moralee et al. 2000). Non-destructive sampling was performed by preserving fin clips in 99% ethanol. A total of 144 phenotypically identified individuals were genotyped with microsatellite markers. MtDNA control region sequence information was obtained for 41 of the above mentioned individuals (Table 1). Given the apparently multiple introductions of tilapias from unknown origin into Southern Africa and Malawi, mtDNA genotyping was extended to another 17 *O. mossambicus* phenotypic fish from farms and the wild. Captive fish were collected in four South African farms (Valley Farm, Amatikulu, Makathini, and Nick James). Wild fish were collected in Mozambique, the northernmost distribution range of

Table 1 Sample sites of *Oreochromis mossambicus, O. niloticus* and their hybrids

Phenotypic identification	Sample	*N*	Location	Site, country	Wild/aquaculture
O. mossambicus	MILE	7 (1)	34° S; 18°45′ E	Le Pommier Farm, Western Cape, SA	Farm, introduced in 1940's
O. mossambicus	MWBO	10 (2)	33°30′ S; 26°30′ E	Boesmans River, SA	Wild
O. mossambicus, red coloured variety	MFNJ	(2)	33° S; 27° E	Nick James Farm, SA	Farm
O. mossambicus	MFAM	(2)	29° S; 31°30′ E	Amatikulu, Natal, SA	Farm
O. mossambicus	MFMA	(2)	27° S; 31°30′ E	Makathini research station, Pongola River, Natal, SA	Farm
O. mossambicus	MWND	9 (2)	26°45′ S; 32° E	Pongola/Usutu River at Ndumu, SA	Wild
O. mossambicus, red coloured variety	MFVA	(2)	25° S; 31° E	Valley Farm, SA	Farm
O. mossambicus	MWOL	(6)	24° S; 31°45′ E	Olifants River, Limpopo basin, SA	Wild
O. mossambicus	MWLI	21 (10)	22°16′ S; 29°16′ E	Dam at Limpopo River, South Africa	Wild
Hybrids					
O. mossambicus × *O. niloticus*	HWLI	7 (8)	22°16′ S; 29°16′ E	Dam at Limpopo River, SA	Wild
O. mossambicus	MWSU	(7)	17° S; 35°30′ E	Sucoma, lower Shire river system, MA	Wild
O. mossambicus,	MWKA	14 (6)	17° S; 35°30′ E	Research station on the Shire River system at Kasinthula, MA	Wild
O. niloticus	NWLI	(4)	22°10′52′ ′ S; 29°14′4′ ′ E	Dam at Limpopo River, SA	Wild, introduced.
O. niloticus	NWE1	20 (5)	Northern Nile River	Nile River, EG	Wild
O. niloticus	NWE2	16 (4)	Northern Nile River	Nile River, EG	Wild
O. niloticus	NFTH	20	Thailand	Thailand	Farm
O. niloticus	NFPH	20 (5)	Philippines	Central Luzon State University, PH	Farm

N, number of individuals that were genotyped with microsatellites; (*N*), number of individuals that were genotyped at the mitochondrial control region. SA, South Africa; MA, Malawi; EG, Egypt; PH, Philippines

O. mossambicus (at Sucoma and Kasinthula), where it occurs in sympatry with *O. karongae*, and in the Limpopo basin at the Olifants River. Additionally four phenotypic *O. niloticus* specimens were collected from close to the site of hybrid collection (further details in Table 1). Nile tilapias from the native range and farms were included to cover as much *O. niloticus* diversity as possible, as the origin of the introduced fish is unknown. Overall, 68 sequences were obtained.

DNA was extracted following an adapted salting out method of Nucleon 1 (SCOTLAB, UK).

Genotyping

MtDNA

A fragment of 385 bp at the 5′ end of the mtDNA control region was amplified from individuals morphologically identified as *Oreochromis mossambicus*, *O. niloticus* and their hybrids, collected from the wild and from fish farms. We used the primers H16498 (Meyer et al. 1990) and L19 (Bernatchez et al. 1992), 25 ng DNA, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.025 U μl^{-1} *Taq* DNA polymerase (Promega, USA), 0.15 μM of each primer, and 1 × Promega buffer. Cycling conditions were 3 min at 94°C, followed by 35 cycles of denaturation at 94°C for 30 s, 40 s at 51°C and extension at 72°C for 45 s, finalized by a final extension step of 72°C for 5 min.

Microsatellites

Microsatellite loci have been characterized previously for *Oreochromis niloticus* (Lee and Kocher 1996). Polymerase Chain Reactions (PCR) were performed in a final volume of 10 μl, containing 1.5 mM $MgCl_2$ (2 mM $MgCl_2$for UNH 129), 0.2 mM dNTPs, 0.025 U μl^{-1} *Taq* DNA polymerase (Promega, USA), 0.1 μM of each primer, 1 × Promega buffer, and 25 ng genomic DNA. For each primer pair, one of the primers was labelled with FAM, NED or HEX dyes (Applied Biosystems, USA).

PCR amplifications were conducted in a GeneAmp 2700 thermal cycler (Applied Biosystems, USA). Cycling conditions comprised an initial denaturation step of 5 min at 94°C, followed by 35 cycles of denaturation at 94°C for 30 s, 30 s at the optimal annealing temperature and extension at 72°C for 45 s finalized by a final extension step of 72°C for 7 min. Annealing temperature was 55°C for locus UNH 104, 56°C for locus UNH 111, 58°C for locus UNH 123, and 50°C for loci UNH 129 and UNH 146.

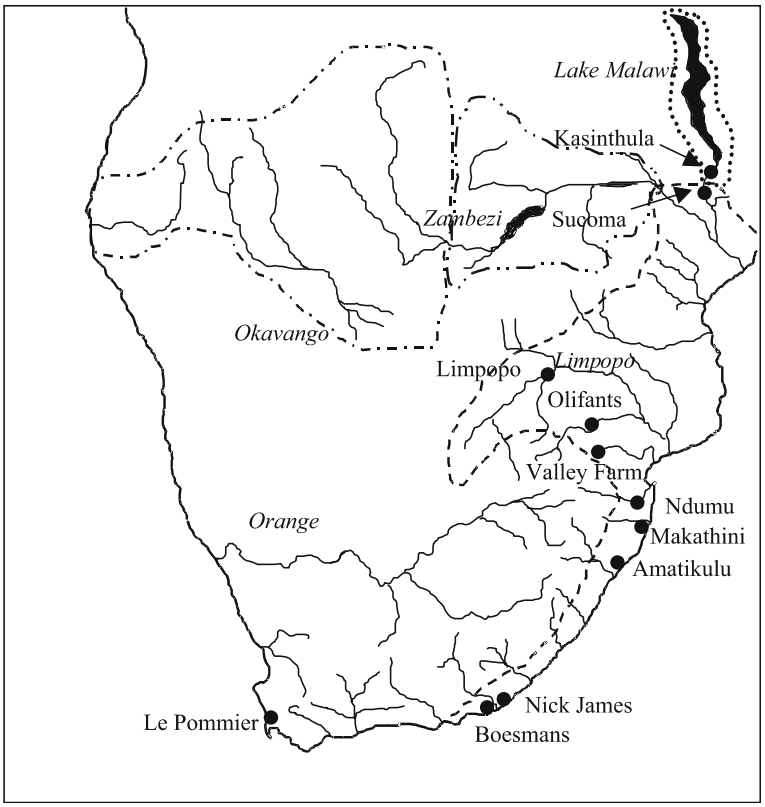

Fig. 1 Native ranges of *O. mossambicus* (— —), *O. karongae* (.....), *O. andersonii* (– · – ·) and *O. mortimeri* (– · · -) and sampling sites for *O. mossambicus*

Sequencing and PCR products were run on an ABI 3100 automated capillary sequencer. Genotypes were determined using the Genotyper™ v.2.1 software (Applied Biosystems, USA).

Data analysis

mtDNA sequence data

The 68 sequences generated during this study were aligned with published GenBank sequences of *O. niloticus, O. mossambicus*, other parapatric (*O. mortimeri* and *O. karongae*) and an allopatric species (*O. andersonii*) found in the Zambezi basin (Fig. 1), other extensively used aquaculture species (*O. aureus*), and sequences that showed a high similarity to our sequences during BLASTN searches (*O. urolepis* and *O. malagarasi*) (Table 2 and Appendix 1). MtDNA control region sequences were aligned with the software BioEdit v.7.0 (Hall 1999) and checked manually. Nucleotide composition was determined using the software MEGA v.2.1 (Kumar et al. 2001). The substitution model that best fitted the data was determined with Modeltest v.3.0 (Posada and Crandall 1998). The best likelihood scores were obtained for the TrN 93 (Tamura and Nei 1993) with gamma shape parameter model.

A phylogenetic parsimony network was reconstructed using the median-joining algorithm (Bandelt et al. 1999) as implemented in NETWORK v.4.1.1.2. (http://www.fluxus-engineering.com/sharenet.htm). The algorithm reconstructs median vectors, which are parsimony consensus sequences that can be interpreted either as ancestral or unsampled. All sites were given equal weights, parameter ε was set to 0, and the final network was drawn using the union of all shortest minimum-spanning trees.

A Maximum Likelihood (ML) tree was constructed with Tree Puzzle v.5.0 using the quartet–puzzling algorithm, the Tamura and Nei (1993) substitution model and an eight-category γ-distribution rate of the substitution rates across variable sites (Yang 1994) of $\alpha = 0.21$, estimated from the ML approach in Tree-Puzzle (Schmidt et al. 2002). The support of internal branches was tested with 10,000 puzzling steps (Schmidt et al. 2002). Additionally, a more detailed phylogenetic analysis was restricted to only *O. mossambicus* sequences and one outgroup using Maximum Parsimony (MP) with close-neighbour interchange with a search level of two and maximum likelihood as above. Divergence time between groups was inferred from the mutational distance and the average substitution rate for haplochromines of 5.6% Myr^{-1} (Nagl et al. 2000) using NETWORK.

Table 2 Frequency listing of all 38 tilapiine haplotypes forming the six haplotype clusters (see Fig. 2)

	Accession number	Haplotypes	Sampling sites (site code)												
			LePommier (MILE) (F)	Boesmans (MWBO)	Nick James (MFNJ) (F)	Amatikulu (MFAM) (F)	Makathini (MFMA) (F)	Ndumu (MFNA) (F)	Valley (MFMA) (F)	Olifants (MWOL)	Limpopo (MWLI–MHLI–NWLI)	Sucoma (MWSU)	Kasinthula (MWKA)	Nile River (NWE1–NWE2)	Philippines (NFPH)
O. mossambicus	AF328851	C1-1-*Oaureus**													
	AY833435 AF296494*	C1-2-*Ourolepis*	1												
	AY833436–40	C2a-1		1	2			1		1					
	AY833441–42	C2a-2								2					
	AY833443	C2a-3		1											
	AY833444	C2a-4						1							
	AY833445–6	C2a-5					2								
	AY833447	C2a-6											1		
	AF328843*	C2b-													
	AY833448–9	C2c-1										1	1		
	AY833450	C2c-2											1		
	AY833451	C2c-3											1		
	AY833452	C2c-4											1		
	AY833453	C2c-5											1		
	AY833454–58 AF296466*	C2c-6				2			2	1					
	AY833459	C2c-7								1					
	AF015006	C3-*Omalagarasi**													
	AF328844	C4a-*Okarongae**													
	AY833460-65	C4b										6			
O. niloticus	ONI237397	C5-1-*Onbaringoensis**													
	AY833466–72	C5–2									4			3	
	AF328849	C5-3*													
	AY833473	C5-4												1	
	AF296468	C5-5 *Onvulcani**													
	AY833474–81	C5-6									8				
	AY833482	C5-7												1	
	AY833483	C5-8												1	
	AY833484–85	C5-9												2	
	AY833486	C5-10												1	
	AY833487–91	C5-11													5
O. andersonii	AF296488	C6a-*Oandersoni**													
	AY833492	C6b-1								1					
	AY833493	C6b-2									1				
	AY833494	6c-1									1				
	AY833495–97	6c-2									3				
	AY833498–99	6c-3									2				
	AY833500–501	6c-4									2				
	AY833502	6c-5									1				

Haplotype labelling: C1–C6 refers to the cluster number, the characters a–c (if present) refer to the subcluster, and the number is the haplotype number within the cluster; *: DNA sequences from GenBank; (F): farm samples

Microsatellite data

Allele frequencies for each locus were calculated with the software GENEPOP v. 3.1 (Raymond and Rousset 1995). Observed and expected unbiased heterozygosity (Nei 1987) were calculated using the software FSTAT v.2.9.1 (Goudet 2001). Tests for departures from Hardy–Weinberg equilibrium were performed over all loci per species and per sample using a Markov chain method (Raymond and Rousset 1995), to estimate without bias the exact *P*-value (Guo and Thompson 1992). Even though each microsatellite locus was mapped to a different linkage group (Lee et al. 2005), we performed a genotypic linkage disequilibrium test with GENEPOP. Exact *P*-values were estimated with a Markov chain method. Heterogeneity in allele frequency distribution between species was tested with the analysis of variance based methods assuming the Infinite Allele Model (IAM) (Weir and Cockerham 1984) and the Stepwise Mutation Model (SMM; Goodman 1997). F_{is} was estimated as in Weir and Cockerham (1984). The significance of these estimators was tested using resampling procedures provided by the package FSTAT.

The extent of population differentiation and relationship among samples were assessed with Factorial Correspondence Analysis (FCA) using GENETIX v.4.05.2 (Belkhir 1999). In addition, we applied the model-based program STRUCTURE v.2.1 (Pritchard et al. 2000), which uses a Bayesian clustering approach on individuals' multilocus data to infer the most likely number of populations (K) and the proportion of each K within individual genomes. We used a model with correlated gene frequencies allowing for admixture, with $1 < K < 10$. We performed five runs for each K with 600,000 MCMC steps after a burn-in of 50,000. Then, we estimated the true number of K following the method of Evanno et al. (2005).

Results

Phylogenetic reconstruction

After alignment of all 73 sequences (68 newly obtained and five downloaded from GenBank), 38 haplotypes were recognized (Appendix 1 and Table 2). Our dataset contained 95 variable sites and 83 parsimony informative sites. Mean G + C content was 32.9%, and the observed transition:transversion parameter was 7.75 ± 2.29 (expected ts:tv ratio = 6.66).

Phylogenetic reconstruction using a median-joining parsimony network (Fig. 2) showed six main clusters: (1) an *Urolepis–Aureus* group, with *O. urolepis* (NCBI Accession No. AF296494), *O. aureus* (AF328851) and one individual identified as *O. mossambicus* from the Le Pommier farm; (2) a *Mossambicus* group, containing individuals identified as *O. mossambicus* collected in the Limpopo, Olifants and Boesmans Rivers, Shire River at Sucoma and Kasinthula, and also the Amatikulu, Makathini, Valley and Nick James fish farms; (3) one distinct *O. malagarasi* haplotype; (4) a *Karongae* group, containing *O. karongae* (AF328844) and six fish from Sucoma, Malawi, morphologically identified as *O. mossambicus*; (5) a *Niloticus* cluster, containing *O. niloticus* from the Nile River, a Philippino fish farm, the Limpopo River, the seven fish morphologically identified as *O. mossambicus* × *O. niloticus* hybrids

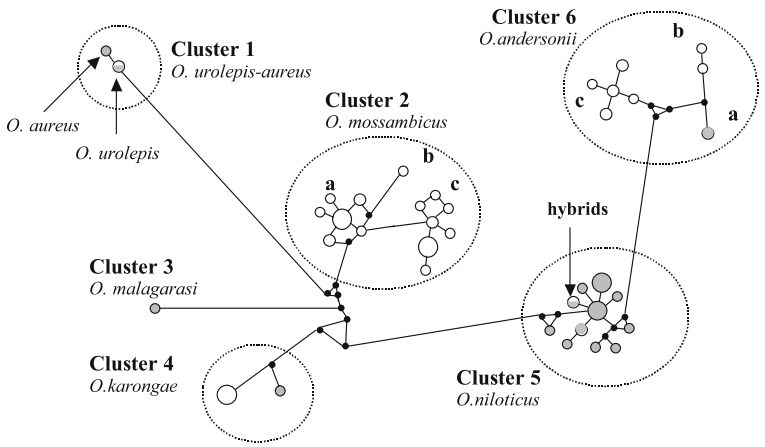

Fig. 2 Median-Joining network of the 38 unique haplotypes based on 385 bp of the mitochondrial control region. The area of the circles is proportional to the frequency of the haplotypes. The six distinctive clusters corresponding to the mtDNA of the species *O. aureus–urolepis, O. mossambicus, O. malagarasi, O. karongae, O. niloticus* and *O. andersonii* are numbered from 1 to 6. Fish with a "mossambicus" phenotype are indicated in white; all other phenotypes are indicated in grey

from the Limpopo River, and one fish morphologically identified as *O. mossambicus*, also from the Limpopo River; and lastly (6) a *Mortimeri–Andersonii* group, containing *O. mortimeri*, *O. andersonii* and sequences of fish identified as *O. mossambicus* from the Limpopo and Olifants Rivers. The *Niloticus* cluster falls in an intermediate position between the *Andersonii* and the *Karongae* + *Mossambicus* groups, while the *Mortimeri–Andersonii* group is more closely related to *Niloticus* than to any other group. The distribution of all these 38 haplotypes among sampling sites is shown in Table 2.

The most significant observation on this phylogenetic reconstruction is that fish identified as *O. mossambicus* were found to carry mtDNA of many different *Oreochromis* species: some fish from Sucoma carried *O. karongae* mtDNA, fish from the Limpopo River carried *O. niloticus* and *O. andersonii–mortimeri* mtDNA. This *O. niloticus* haplotype was identical to that of the hybrids, providing additional evidence of the interspecific hybridization in this drainage. The control regions of *O. andersonii* (AF296488) and *O. mortimeri* (AF328845) are identical.

The network (Fig. 2) showed a higher resolution than the ML tree (not shown). It rooted with *O. aureus*, and showed three main clusters: *Niloticus*, *Mossambicus* (that includes *O. malagarasi* and *O. karongae*) and *Mortimeri–Andersonii* with high levels of bootstrap support (89%, 82% and 75%, respectively). TrN93 genetic distances between clusters are shown in Appendix 2.

Phylogeography of *Oreochromis mossambicus*

The *Mossambicus* group (Fig. 2, Cluster 2) is composed of three clusters. Cluster 2a occurs mainly at the southern locations and cluster 2c at the northernmost sampling sites. Both lineages co-exist in the Olifants River and Kasinthula, Malawi (Table 2). The network shows *O. mossambicus* lineages rooted on a missing intermediate. The rooting of the trees at a different position would have a major influence on the phylogeographic inference and interpretation of the ancestral haplotypes and recent—ancient radiations. Hence, a more detailed analysis was restricted to the 15 haplotypes of *O. mossambicus* and *O. karongae*. The median-joining network roots *O. mossambicus* to a missing intermediate haplotype within cluster 2a (as in Fig. 2), which might be an artifact (Cassens et al. 2003). The total number of mutations in this tree is 52, and ε was kept at a value of zero to reduce the level of homoplasies.

The MP tree (not shown), placed the haplotype C2b basally to the two other clusters (bootstrap support is 77% in the MP tree with all haplotypes). This tree instead showed a total length of 39 mutations, therefore rendering a more reliable tree under the parsimony criteria than the network. The NJ tree (not shown) showed a topology identical to that of the MP tree, whereas the ML tree showed haplotype 2b unresolved. All these trees also show a longer internode for cluster 2c, indicating the more recent radiation of cluster 2a. Applying a rate of 5.6% Myr^{-1}, the age in mutations (rho) of the three *O. mossambicus* 2a, 2b, and 2c lineages indicates 98,000 (±46,400); 417,600 (±122,700) and 464,400 (±139,200) years, respectively. The within-group genetic distance of cluster 2a indicates a radiation 17,800 years before present. The age of the *O. mossambicus* lineage divergence from its node in common with *O. malagarasi* renders approximately 464,400 (±46,400) years before the radiation of the lineages 2a–c.

Genetic diversity and genetic differentiation of microsatellite genotypes

The number of alleles detected in *O. niloticus* is slightly higher than in *O. mossambicus* (Table 3). Heterozygosity levels are similar, and the smallest differences between H_E and H_O were detected in the farmed *O. mossambicus* (MILE, Le Pommier farm), *O. niloticus* (NWE2, Nile River at unknown location in Egypt) and the hybrids (Table 3). Only locus UNH146 was in Hardy–Weinberg equilibrium over all samples. However, due to the small sample sizes these estimations should be considered with caution.

We tested the differentiation between the two species by pooling all *O. mossambicus* (excluding the farm MILE and hybrids from Limpopo, MWLI, HWLI). Both F_{st} and R_{st} indicated these two species to be highly differentiated ($F_{st} = 0.3$; $R_{st} = 0.43$, $P < 0.01$). Between populations analysis of F_{st} indicated strong differentiation between all wild samples except for the Nile populations (Appendix 3).

FCA of the microsatellite genotypes by population separates populations mainly by the *x*-axis. Two main groups are formed with the parental species samples; *O. niloticus* fits to the right of the centre and *O. mossambicus* to the left, and the hybrid sample takes an intermediate position (Fig. 3a). Along the *y*-axis, the most important separation occurs between the Philippino farm sample and the other *O. niloticus* samples. Among the *O. mossambicus* samples, Le Pommier (MILE) farm is more closely related to the wild samples from Malawi (Kasinthula) and Ndumu (eastern South Africa), corresponding with their documented Malawi origin (E. Hall, pers. comm.).

Table 3 Genetic variability at five microsatellite loci for *O. mossambicus*, *O. niloticus* and wild hybrid individuals

Locus	N	*O. mossabicus*						Hybrids	*O. niloticus*					Overall
		MWLI 21	MWBO 10	MWND 9	MILE 7	MWKA 14	Total 61	HWLI 7	NFTH 20	NFPH 20	NWE1 20	NWE2 16	Total 76	144
UNH102	H_E	0.56	0.69	0.36	0.82	0.8		0.43	0.11	0	0.46	0		
	H_O	0.48	0.7	0.43	0.25	0.62		0	0.12	0	0.38	0		
	N_A	4	4	2	4	6	6	2	2	1	4	1	4	7
	F_{is}	0.147	– 0.016	0.2	0.727	0.238		1	– 0.32	–	0.161	–		
	P_{HWE}	NS	*	NS	NS	NS		NS	NS	NS	NS	–		*
UNH124	H_E	0.63	0.66	0.65	0.51	0.61		0.86	0.72	0.48	0.87	0.79		
	H_O	0.44	0.2	1	0.2	0.5		0.75	0.67	0.18	0.83	0.56		
	N_A	4	4	3	3	4	8	5	5	2	7	7	11	11
	F_{is}	0.304	0.707	– 0.6	0.636	0.18		0.143	0.071	0.636	0.039	0.297		
	P_{HWE}	NS	*	NS	NS	NS		NS	NS	NS	NS	NS		*
UNH129	H_E	0.54	0.53	0.33	0.30	0.47		0.2	0.06	0.37	0.46	0.62		
	H_O	0.29	0.75	0.13	0.33	0.42		0.2	0.06	0.47	0.56	0.31		
	N_A	3	2	2	2	3	3	2	2	2	4	5	6	6
	F_{is}	0.476	– 0.448	0.632	– 0.111	0.113		–	–	– 0.286	– 0.239	0.503		
	P_{HWE}	*	NS	NS	NS	NS		–	–	NS	NS	*		*
UNH146	H_E	0.61	0.48	0.39	0.62	0.56		0.53	0.41	0.30	0.63	0.68		
	H_O	0.76	0.7	0.44	0.86	0.83		0.4	0.47	0.35	0.67	0.38		
	N_A	3	2	3	3	3	3	2	4	2	5	4	7	7
	F_{is}	– 0.267	– 0.5	– 0.164	– 0.44	– 0.517		0.273	0.164	– 0.187	– 0.061	0.456		
	P_{HWE}	*	NS	NS	NS	NS		NS	NS	NS	NS	NS		NS
UNH192	H_E	0.43	0.42	0	0.67	0.62		0.71	0.78	0.53	0.61	0.61		
	H_O	0.2	0.1	0	0.14	0.38		0.57	0.85	0.95	0.47	0.63		
	N_A	5	3	1	4	3	8	4	5	3	5	5	8	9
	F_{is}	0.541	0.769	–	0.8	0.385		0.213	– 0.088	– 0.814	0.213	– 0.017		
	P_{HWE}	**	NS	–	*	NS		NS	NS	*	NS	NS		**
Total	H_E	0.55	0.56	0.34	0.58	0.61		0.55	0.42	0.34	0.60	0.54		
	H_O	0.43	0.49	0.40	0.36	0.55		0.38	0.43	0.39	0.58	0.38		
	N_a	21	15	11	16	19		15	18	10	25	28		
	P_{HWE}	**	**	NS	*	NS		NS	NS	*	NS	**		**

H_E, expected heterozygosity without bias (Nei 1987); H_O, the observed heterozygosity; N_A, number of alleles; P_{HWE}, probability of Hardy–Weinberg equilibrium; *$P < 0.05$; **$P < 0.01$

The genetic composition of each population and the strong differentiation among populations is more evident from the results of the assignment analyses. Several runs for each K produced consistent results and the highest likelihood for $K = 7$, $[P\ (D/K) = 1)$ with 600,000 iterations after a burn-in of 50,000. The differential distribution of these components among samples also indicates strong population structure (not shown). The number of K was subsequently reduced to two after analysing the rate of variation in L (K/D) between successive K (ΔK statistic of Evanno et al. 2005). These two clusters represent different genomic contribution and are unequally distributed between species (Fig. 3b): one cluster represents 97–99% of the *O. mossambicus* wild sample genomes, and the second cluster represents 89–98% of the *O. niloticus* genomes. These ''mossambicus'' and ''niloticus'' components are equally represented in the hybrids (49.5% and 50.5%, respectively). The Limpopo sample MWLI shows 77% of the *O. mossambicus* genome. However, these fish also carry *O. mortimeri–andersonii* and *O. niloticus* DNA. We prefer more cautiously to refer to ''mossambicus'' and ''non-mossambicus'' components in this case, as we cannot infer the level of introgression from the present data.

Individual identification of hybrids

FCA analysis of the individual microsatellite genotypes by their corresponding mtDNA haplotype shows that *O. niloticus* and *O. mossambicus* individuals slightly overlap along the first axis (data not shown). Phenotypic *O. mossambicus* from the Limpopo River (MWLI) plotted mostly in the negative quadrant; the haplotypes were typical *O. niloticus* or *O. andersonii* mtDNA. The distribution of the proportion of the *O. mossambicus* genome q appears bimodal among MWLI fish: lower than 40% and higher than 75%. By contrast, in fish with a hybrid phenotype, q ranged from 86% to 18% (Fig. 4), which follows approximately the distribution along the x-axis from left to right in the FCA plot of individuals (not shown).

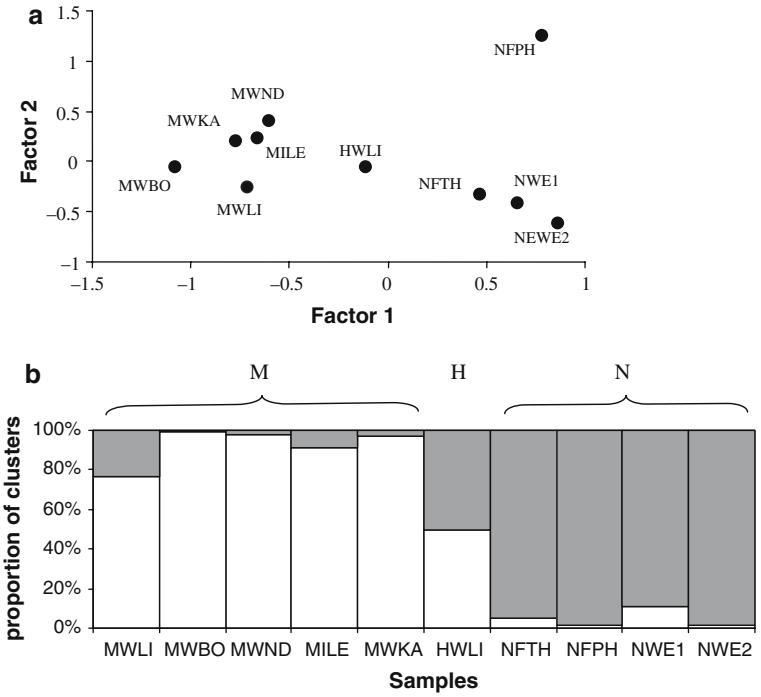

Fig. 3 (**a**) Factorial Correspondence Analysis of *Oreochromis mossambicus* and *O. niloticus* using sample microsatellite allele frequencies. For sample abbreviations see Table 1. Factor 1 explains 36.03% of inertia while Factor 2 explains 25.41%. (**b**) Proportion of the model-based clusters ($K = 2$) in the ancestry of each sample; in white: ''mossambicus'' cluster; in grey: ''non-mossambicus'' cluster. Sample labels correspond to Table 1. The phenotypic identification of the fish are indicated on top: M, *O. mossambicus*; H, hybrid; N, *O. niloticus*

Discussion

Interspecific hybridization

The most significant finding from the phylogenetic reconstruction is the clustering of *Oreochromis mossambicus* phenotypes with other recognized species: *O. mortimeri–O. andersonii* in the Limpopo River; *O. karongae* in Malawi, and *O. urolepis–O. aureus* in a farm in the Western Cape, South Africa (Le Pommier).

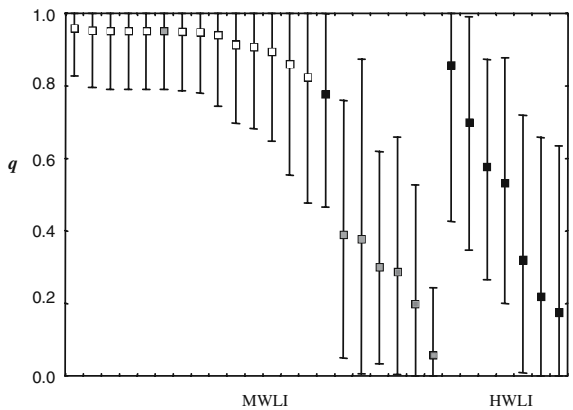

Fig. 4 Distribution of q (proportion of *O. mossambicus* genome) among fish from the Limpopo River with *O. mossambicus* (MWLI) and hybrid phenotype (HWLI). Dots in black: fish carrying *O. niloticus* mtDNA; dots in grey: fish carrying *O. andersonii–O. mortimeri* mtDNA. The values of q have been ranked for each sample, and the ranks are plotted against q. The lines indicate the 95% posterior probability intervals

O. andersonii and *O. mortimeri* occupy in parapatry of the upper and middle Zambezi basins, respectively. The latter is separated from the upper and lower basins by the Victoria Falls and Kafue Gorge, respectively. The *O. andersonii* range also extends to the Okavango and Cunene (Skelton 2001). Of special consideration is the taxonomic status of both species. Given that the sequences of their mitochondrial control region are identical, they may represent a single species. Additional sampling along the Zambezi, Kafue, Cunene and Limpopo basins combined with morphometric analysis await confirmation of this hypothesis. We therefore predict that hybridization might extend to the lower Zambezi, if the physical barrier separating the lower from the middle Zambezi is crossed by these species.

The Zambezi basin is of mixed palaeofluvial origin (reviewed in Dollar 1998). While the upper Zambezi was part of the Limpopo system, and linked to the Okavango by palaeolakes, the Middle Zambezi was part of the Shire system. Different hypotheses that explain the river capture, place this event in the Holocene (Dollar 1998). The biogeographic affinities between the Cunene–Okavango–Upper Zambezi are well known (see e.g., Skelton 1993; Stewart 2001). It is therefore not surprising that the upper Zambezi *O. mortimeri–O. andersonii* are phylogenetically not the most closely species to *O. mossambicus*. We estimated the divergence between *O. andersonii* and *O. mossambicus* from their common ancestor at 1.55 ± 0.5 Myr ago.

O. karongae approximate divergence time is 556,800 ± 65,600 years. It is endemic to Lake Malawi (Trewavas 1983) where it is fished and used in aquaculture (Turner 2003). The clustering of *O. malagarasi*, a lacustrine species, along with *O. karongae* and *O. mossambicus* might indicate a lacustrine common ancestor. Likely, the *O. mossambicus* tolerance to brackish water could be associated with the Plio-Pleistocene drought that produced pronounced water level fluctuations affecting the central African lakes (reviewed in Sturmbauer et al. 2001). Alternatively, it would be derived as an adaptation to estuaries, river mouths and marine coastal waters, which might have favoured its dispersal southwards. We estimated that the divergence of *O. malagarasi* occurred approximately 649,300 (± 173,500) years ago. This implies that divergence might have taken place no later than the lowest water level for Lake Tanganyika, dated at 0.4 Myr (Sturmbauer et al. 2001).

This phylogenetic reconstruction allows for the inference of anthropogenic translocations, examples being the presence of *O. andersonii–mortimeri* in the Limpopo basin and *O. karongae* in Sucoma. Gillmore (1978) reported earlier translocations of various Okavango Delta fish species, including *O. andersonii and O. macrochir* to the newly built Shashe Dam in the Limpopo system (Botswana). Unlike *O. andersonii* the latter has since that time not been collected in the Limpopo River (Kleynhans and Hoffman 1993), possibly because it cannot be easily distinguished from *O. mossambicus* in the field.

The presence of *O. karongae* mtDNA in Sucoma may likely result from hybridization with escapees from fish farms. However, the lineage identified in this paper differs in five transitions from specimens collected in Lake Malawi, and could represent an *O. karongae* lineage adapted to riverine systems. This hypothesis should be tested by sampling rivers without aquaculture interference flowing into and out of Lake Malawi, or even remote locations of the Shire River system. *O. karongae* (synonymous to *Sarotherodon squamipinnis*, D. Tweddle and G. Turner, pers. comm.) was occasionally reported in the Lower Shire (Tweddle and Willoughby 1979; Tweddle and Lewis 1979).

The Le Pommier farm in the Western Cape, an area outside the natural range of *Oreochromis* spp., is apparently breeding hybrids between *O. mossambicus* and *O. aureus*. Given the capacity of *Oreochromis* species to hybridize, it is possible that the *O. urolepis* taken from the Wami River in Tanzania (Nagl et al. 2001) resulted from hybridization with *O. aureus*. The latter is extensively used in aquaculture and its native range is the lower Nile, Lake Chad, Niger and Senegal, while *O. urolepis* is naturally restricted to the Wami River, Tanzania (Trewavas 1983). The *O. urolepis* and *O. aureus* mitochondrial control region sequences are separated by only two transitions.

Phylogeography of *Oreochromis mossambicus*

The most important implication for the phylogeography of *O. mossambicus* is the presence of two or three lineages, defined mostly by differences in transitions. The topology of the trees shows that the most basal *O. mossambicus* lineage inhabits the Zambezi basin (C2b). Cluster 2c, occurring predominantly in Malawi, may represent an older lineage, and the shallower node for cluster 2a argues for a recent southern radiation or colonization. This species has been observed in marine waters in northern KwaZulu-Natal (van der Waal, pers. comm.), suggesting that dispersal occurs along the coast. The hypothetical old group must have been linked to rivers of Lake Malawi and nearby lakes, and colonized the Zambezi basin via the Shire system. Anthropogenic dispersal can be hypothesized from the phylogeographic reconstruction. The presence of the Malawi lineage in the Valley and Amatikulu fish farms (both in South Africa) and the Olifants River suggests translocation of animals out of their native range. A research station on the Olifants River bank used to breed and distribute *Oreochromis* fish (van der Waal, pers. comm.), possibly being responsible for multiple translocations and escapes into the Limpopo basin.

Population structure and phylogenetic relationship among populations

The identification of discrete populations and various levels of gene flow have been of primary interest for the conservation of genetic variation. The potential to discriminate between discrete population units under different scenarios and levels of gene flow, marker type, sample sizes and methods of data analysis have been reviewed recently (Waples and Gaggiotti 2006). Despite the small size of some of our samples, which decreases the power of discrimination, our data display two of the most important factors that favour the detection of differentiation: markers of a high mutation rate and a reduced gene flow ($Nm < 1$) (Appendix 3). The performance of STRUCTURE in this scenario was found to be reliable (Waples and Gaggiotti 2006). Therefore, the differentiation between *mossambicus* and *non-mossambicus* fish, along with the detection of mixed ancestry should be considered seriously. The differentiation among wild *O. mossambicus* samples is

seemingly apparent from the STRUCTURE and F_{st} analyses. However, missing intermediate populations might bias the estimates of θ (Beerli 2004; Slatkin 2005); hence isolation by distance cannot be discarded with the present data. If so, it is imperative to apply methods of detection of mixed ancestry when studying intervening populations due to the apparent extended undocumented anthropogenic introductions and translocations.

Molecular identification of hybrids and the implications for conservation genetics

The misleading nature of morphological identification is apparent from the extensive hybridization found across samples identified as *O. mossambicus*. Morphological identification represents then a very preliminary approximation to the identification of hybrids. The mtDNA genotyping represented an invaluable tool to detect hybrids. This method could be particularly useful in areas where the secondary contacts were once-off events, and introgression occurred mainly through one species' females. This is likely the case of the presence of *O. andersonii–O. mortimeri* in the Limpopo basin. The FCA corresponded well with the Bayesian assignment. For this, we recommend that the FCA method be applied along with the Bayesian analysis of genomic composition combined with mtDNA sequence information (e.g., van Houdt et al. 2005) to infer hybrid status and monitoring of tilapias across basins in Southern Africa. More accuracy in the assignment would be gained by increasing the number of loci (Pritchard et al. 2000; Waples and Gaggiotti 2006), although deep introgression might remain undetected by any method but mtDNA genotyping.

The extensive hybridization that is apparent in Southern Africa is mostly the consequence of secondary contact with introduced species. Understandably, the hybridization also occurs between species that overlap in geographical range, like *O. mossambicus* and *O. andersonii* (already detected by Gregg et al. 1998). Of no less importance is the observation of several species outside their originally described geographic distribution, demonstrating a need for revision. The natural hybridization between native tilapias does not pose a major concern for the conservation of these species, as it represents a natural process. The gene pool of parental species is potentially intact in other southern locations (e.g., *O. mossambicus* at Boesmans and Ndumu).

The conservation status of the Limpopo, where unique (*O. andersonii*) and episodic (*O. niloticus* escapees from surrounding farms) introductions are apparent is a critical case for conservation. Hybridization with introduced species seems to have swamped and replaced the original natural diversity by a hybrid swarm. Following Allendorf et al. (2001), this hybrid population is of no conservation value. Luckily, if adaptation to brackish water were lost in the hybrids, these new genetic combinations would remain secluded to this basin, as *O. mossambicus* dispersal seems to occur through marine waters. The impact of the hybridization and introgression is further heightened by the temporal character of the Limpopo River during droughts, when fish populations are reduced to a few permanent pools and populations experience bottlenecks.

Following Crandall et al. (2000) conservation should be aiming at preserving the evolutionary potential of species, for which it is imperative to test the ecological and genetic exchangeability between populations. The data presented are preliminary to delineate ESUs, but a set of working hypotheses and recommendations can de drawn. Genetic data suggest historical and recent restricted exchangeability with mtDNA and microsatellites, respectively. However, we cannot ascertain ecological exchangeability between lineages or basins. A group of individuals with restricted gene flow accompanied by ecological exchangeability should be managed as a single population (Crandall et al. 2000). *O. mossambicus*, however, may not be represented by a network of relatively discrete units restricted per basin, but by a patched pattern of intact basins and rivers with anthropogenic disturbance. Thus, several population units of conservation value might be recognized within lineages. The populations of the Boesmans River and Usuthu/Pongola (Ndumu sample) can be considered representative of two ESUs within the southern lineage, but their geographic extension must be verified. It is imperative, then, to extend this study to geographically close rivers of different basins and assess the levels of gene flow.

Finally, the hybridization of *O. mossambicus* in its natural range with the introduced *O. niloticus* indeed raises concerns among conservationists and aquaculturists in southern Africa (van der Waal and Bills 2000). Genetic adaptations of *O. mossambicus* might be disrupted by hybridization (Moralee et al. 2000), e.g., its resistance to drought, capability of surviving and reproducing in seawater and resistance to temperatures as low as 11°C. The probability of finding pure Mozambique tilapia is challenging. The rivers from SE South Africa where no aquaculture activities or introductions have occurred represent prime sampling targets. Further north the chances of encountering tilapia conspecifics decreases, because

of human introductions and overlaps with the geographical ranges of other *Oreochromis* species. In addition, stretches of lowland rivers easily accessible for farm activities are more likely to have suffered from introgression. Upper basins are more likely to harbour the reservoirs of native genetic diversity (Van Houdt et al. 2005). Dam or reservoir walls forming insurmountable fish migration barriers in rivers in southern Africa may after all have been a blessing in disguise, preserving a species with a man-made worldwide distribution but no defence to maintain its genetic integrity in its native environment.

Acknowledgements This project was funded by the Department of International Development (DFID, UK) and benefited from a Bilateral Scientific Cooperation Grant between the Republic of South Africa—Flanders (Belgium) (Project BIL/00/30). H. Heyns of the farm Samaria is sincerely thanked for permission to collect samples, and G. Maes, D. Tweddle, J. Snoeks, two anonymous reviewers and Editor E. Nielsen for providing useful comments on the manuscript. We also thank N. Leat for critically reading the manuscript.

Appendices

Appendix 1 Variable sites of the 385 bp fragment of the mitochondrial control region among the 38 recognized *Oreochromis* haplotypes. The haplotype tags correspond to Table 3 and Fig. 3a

```
                                                  11111111 1111111111 1111111111 1111111111 1111222222 2222222222 2233333333 33333
                                1111112223 5568889999 9900001111 1122333334 4455566777 7788888888 9999002223 5566677777 7711224555 77777
                                0456894580 5917891234 7816790123 4589125890 1201238145 6712356789 0378061799 6935923456 8959051046 01238
#C2a1 O.mossambicus              GTCAGCTAAG TACACTTTTC AAACCCTGCT CCCCACCTTA CAGAAGGATC TTAAACATTT CAACACCTCT CCCT-TTCCA ATTTATTATT TAATC
#C2a2                            .......... .......... .......... .......... .......... .......... .......... ....-..... .......... .....
#C2a3                            .......... .......... .......... ....G..... .......... .......... .......... ....-..... .......... .....
#C2a4                            .......... .......... .......... .......... .......... .......... .......... ....-..... .......... .....
#C2a5                            .......... .......... .......... .......... .......... .......... .......... ....-..... .......... .....
#C2a6                            .......... .......... .......... .......... ......A... .......... .......... ....-..... .......... .....
#C2b                             .......... .....C...T .......... .......... ..A...A... C......... .....T.... ....-..T.G .......... .....
#C2c1                            .......... .........T ..G....... .......... ......AGCT .......... .......... ....-....G .......G.C ....T
#C2c2                            C......... .........T ..G....... .......... ......AGCT .......... .......... ....-....G .......G.C ....T
#C2c3                            .......... .........T ..G....... .......... ......AGCT ..G.G..... .......... ....-....G .......G.C ....T
#C2c4                            .......... .........T ..G....... .......... ......AGCT ..G.G..... .......... ....-....G .......G.C ....T
#C2c5                            .......... .........T ..G....... .......... ......AGCT .......... .......... ....-....G .......G.C ....T
#C2c6                            .......... .........T ..G....... .......... ......AGCT .......... .......... ....-....G ...C...G.C ....T
#C2c7                            ........-. .........T ..G....... .......... ......AGCT .......... .......... ....-....G ...C...G.C ....T
#C11 O.aureus                    ..TGA..G.. C..T.A..CT .T.AG....C ..A.T...C. ..C.TAC.CT .CGT.AT..G A...G..CTC TATGCACT.- .A.C.....C ...A.
#C12 O.urolepis                  ..TGA..G.. C..T.A..CT .T.AG....C ..A.T...C. ..C.TAC.CT .CGT.AT..G A...G..C.C TATGCACT.- .A.......C ...A.
#C3a O.malagarasi                ..T..T.G.. ..T..CC.CT ......CA.. .T.T.G.... ..A..AA... .......... ..C.GT..T. .A..-C.T.. .........C .....
#C4a O.karongae                  .C........ .....CCCCT .....T..T. T..T...... ..A..AA..T .......... T..TG..... .A..-..... ..C..C.... C..C.
#C4b                             .C........ .....CCCCT .....T.... T..T...... ..A..AAGCT .......... T..TG....C .A..-....G ..C..C.... C....
#C51 O.niloticus                 .........A .T...AC.CT ...A....T. TTTTT..C.. T.AG.AA..A ...T..GA.. AT..G..... .A..-...T. ..CGT..... ....T
#C52                             .........A .T...AC.CT ...A....T. TTT.T..C.. T.AG.AA..A ...T..GA.. AT..GT.... .A..-...T. ..CAT.G... ....T
#C53                             .........A .T...AC.CT ...A....T. TTT.T..C.. T.AG.AA..A ...C...A.. AT..GT.... .A..-...T. ..CAT..... ....T
#C54                             ......C.GA .T...AC.CT ...A....T. TTT.T..C.. T.AG.AA..A ...T...A.. AT..GT.... .A..-...T. ..CAT..... ....T
#C55                             ........GA .T...AC.CT ...A....T. .TT.T..C.. T.AG.AA..A ...T...A.. AT..GT.... .A..-...T. ..CAT..... ....T
#C56                             .........A .T...AC.CT ...A....T. TTT.T.TC.. T.AG.AA..A ...T..GA.. AT..GT.... .A..-...T. ..CAT..... ....T
#C57                             .........A .T...AC.CT ...A....T. TTT.T..C.. T.AG.AA..A C..T..GA.. AT..GT.... .A..-...T. ..CAT..... ....T
#C58                             .........A .T...AC.CT ...A....T. TTT.T..C.. T.AG.AA..A ...T..GA.. AT..GT.... .A..-...T. ..AAT..... ....T
#C59                             .........A .T...AC.CT ...A....T. TTT.T..C.. T.AG.AA..A ...T.TGA.. AT..GT.... .A..-...T. ..CAT..... ....T
#C510                            .........A .T...AC.CT ...A....T. TTT.T..C.. T.A..AA..A ...T.TGA.. AT..GT.... .A..-...T. ..CAT..... ....T
#C511                            .........A .T...AC.CT ...A....T. .TT.T..C.. T.AG.AA..A ...T..GA.. AT..GT.... .A..-...T. ..CAT..... ....T
#C6a O.andersonii                ..T..T.G.A ....TAC..T ...A.T..T. TTTTT..... ..A..AA... ........C. ....GT...C .AT.A.CTTG G....C.GC. .TC.T
#C6b1                            ..T..T.G.A ....TAC..T ...A.T..T. TTTTT....G ..A..AA... .......... ....GT.... .AT.A.CTTG G.C..C.GA. ..C.T
#C6b2                            ..T..T.G.A ....TAC..T ...A.T..T. TTTTT....G ..A..AA... .......C.. ....GT.... .AT.A.CTTG G.C..C.GA. ..C.T
#C6c1                            ..-..T.G.A .....AC..T ...A.T..T. TTT.T..... ..A..AA... ........C. ....GT.... .AT.A.CT.G ..C..C.GC. ..C.T
#C6c2                            ..-..T.G.A .....AC..T G..A.T..T. TTT.T..... ..A..AA... ........C. ....GT.... .AT.A.CT.G ..C..C.GC. ..C.T
#C6c3                          ..-..T.G.A .....AC..T G..A.T..T. TT..T..... ..A..AA... ........C. ....GT.... .AT.A.CT.G ..C..C.GC. ..C.T
#C6c4                          ..-..T.G.A .....AC..T G..A.T..T. TTT.T..... .GA..AA... .......... ....GT.... .AT.A.CT.G ..C..C.GC. ..C.T
#C6c5                          ..-..T.G.A .G...AC..T G..A.T..T. TTT.T..... ..A..AA... ........C. ....GT.... .AT.A.CT.G ..C..C.GC. ..C.T
```

Appendix 2 Average TrN93 + Γ genetic distance between clusters (below diagonal) and SE (above diagonal). On the diagonal: within group mean distance with SE between brackets

	Clusters									
	C1	C2a	C2b	C2c	C3	C4	C5	C6a	C6b	C6c
C1	0.005 (0.004)	0.034	0.034	0.032	0.038	0.049	0.051	0.052	0.054	0.047
C2a	0.162	0.002 (0.001)	0.009	0.01	0.021	0.024	0.024	0.034	0.031	0.026
C2b	0.162	0.021	0	0.013	0.018	0.023	0.025	0.032	0.029	0.024
C2c	0.16	0.031	0.041	0.005 (0.002)	0.024	0.023	0.026	0.036	0.034	0.029
C3	0.179	0.081	0.068	0.099	0	0.029	0.029	0.03	0.026	0.026
C4a	0.2	0.083	0.082	0.091	0.104	0.017 (0.007)	0.028	0.034	0.029	0.028
C5	0.215	0.114	0.116	0.128	0.127	0.115	0.009 (0.002)	0.027	0.022	0.022
C6a	0.226	0.135	0.127	0.148	0.117	0.131	0.111	0	0.008	0.009
C6c	0.21	0.111	0.103	0.122	0.105	0.111	0.093	0.026	0.026	0.005 (0.002)

Appendix 3 F_{st} values (Weir and Cockerham 1984) and *Nm* between samples

	MWLI	MWBO	MWND	MILE	MWKA	HWLI	NFTH	NFPH	NWE1	NWE2
MWLI		*2.1*	*0.41*	*2.54*	*1.11*	*1.12*	*0.51*	*0.32*	*0.64*	*0.52*
MWBO	0.106*		*0.49*	*1.5*	*1.25*	*0.69*	*0.3*	*0.24*	*0.48*	*0.37*
MWND	0.378*	0.336*		*1.43*	*1.63*	*0.43*	*0.26*	*0.21*	*0.4*	*0.29*
MILE	0.089*	0.143*	0.148*		*0*	*1.94*	*0.56*	*0.35*	*0.77*	*0.55*
MWKA	0.184*	0.167*	0.133*	– 0.022		*1.01*	*0.48*	*0.36*	*0.66*	*0.51*
HWLI	0.183*	0.266*	0.368*	0.114	0.199*		*1.11*	*0.53*	*1.69*	*1.09*
NFTH	0.328*	0.455*	0.493*	0.310*	0.340*	0.184*		*0.62*	*2.28*	*1.38*
NFPH	0.436*	0.511*	0.538*	0.417*	0.411*	0.319*	0.286*		*0.81*	*0.62*
NWE1	0.281*	0.344*	0.387*	0.245*	0.274*	0.129*	0.099*	0.236*		*10.47*
NWE2	0.324*	0.401*	0.462*	0.311*	0.327*	0.187*	0.153*	0.286*	0.023	

*Indicates the significant values at the $P < 0.05$ level after 1,000 replications

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. TREE 16:613–622

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Barton NH (2001) The role of hybridization in evolution. Mol Ecol 10:551–568

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. Mol Ecol 13:827–836

Belkhir K (1999) GENETIX, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome et Populations, Université de Montpellier II. http://www.univ-montp2.fr/~genetix

Bernatchez L, Guyomard R, Bonhomme F (1992) DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. Mol Ecol 1:161–173

Cassens I, Van Waerebeek K, Best PB, Crespo EA, Reyes J, Milinkovitch MC (2003) The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of network methods and rooting procedures. Mol Ecol 12:1781–1792

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. TREE 15:290–295

Dollar ESJ (1998) Palaeofluvial geomorphology in southern Africa: a review. Prog Phys Geog 22:325–349

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Falk TM, Teugels GG, Abban EK, Villwock W, Renwrantz L (2003) Phylogeographic patterns in populations of the black-chinned tilapia complex (Teleostei, Cichlidae) from coastal areas in West Africa: support for the refuge zone theory. Mol Phylogenet Evol 27:81–92

Gillmore KS (1978) A fisheries survey of Shashe Dam, Botswana. Unpublished report of the Department of Agriculture and Fisheries, Ministry of Agriculture, Gaborone Botswana, p 26

Goodman SJ (1997) Rst Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and a determining their significance. Mol Ecol 6:881–885

Gregg RE, Howard JH, Shonhiwa F (1998) Introgressive hybridization of tilapias in Zimbabwe. J Fish Biol 52:1–10

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Update from Goudet 1995. Available from http://www.unil.ch/izea/softwares/fstat.html.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportions for multiple alleles. Biometrics 48:361–372

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Hey J, Won Y-J, Sivasundar A, Nielsen R, Market JA (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. Mol Ecol 13:909–919

Joyce D, Lunt DH, Bills R, Turner GF, Katongo C, Duftner N, Sturmbauer C, Seehausen O (2005) An extant fish radiation emerged in an extinct Pleistocene Lake. Nature 435:90–94

Klett V, Meyer A (2002) What, if anything, is a Tilapia? Mitochondrial ND2 phylogeny of Tilapiines and the evolution of parental care system in the African cichlid fishes. Mol Biol Evol 19:865–883

Kleynhans CJ, Hoffman A (1993) First record of *Oreochromis macrochir* (Boulenger, 1912) (Pisces: Cichlidae) from the Limpopo River in South Africa. SAJ Aquat Sci 18:104–107

Kocher TD (2004) Adaptive evolution as explosive radiation: the cichlid fish model. Nature Rev Genet 5:288–298

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17:1244–1245

Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD (2005) A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). Genetics 170:237–244

Lee WJ, Kocher TD (1996) Microsatellite DNA markers for genetic mapping in the tilapia, *Oreochromis niloticus*. J Fish Biol 49:169–171

Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of lake Victoria cichlid fishes suggested by mitochondrial-DNA sequences. Nature 347:550–553

Moralee RD, Van der Bank FH, Van der Waal BCW (2000) Biochemical genetic markers to identify hybrids between the endemic *Oreochromis mossambicus* and the alien species, *O. niloticus* (Pisces: Cichlidae). Water SA 26:263–268

Nagl S, Tichy H, Mayer WE, Samonte IE, McAndrew BJ, Klein J (2001) Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. Mol Phylogenet Evol 20:361–374

Nagl S, Tichy H, Mayer WE, Takezaki N, Takahata N, Klein J (2000) The origin and age of haplochromine fishes in Lake Victoria, East Africa. Proc R Soc Lond B Biol Sci 267:1049–1061

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Raymond M, Rousset F (1995) GENEPOP (version 1.2): a population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rognon X, Guyomard R (2003) Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O-niloticus* in West Africa. Mol Ecol 12:435–445

Romana-Eguia MRR, Ikeda M, Basiao ZU, Taniguchi N (2004) Genetic diversity in farmed Asian Nile and red hybrid tilapia stocks evaluated from microsatellite and mitochondrial DNA analysis. Aquaculture 236:131–150

Salzburger C, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? Mol Ecol 11:619–625

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Schwank EJ (1995) The introduced *Oreochromis niloticus* is spreading on the Kafue floodplain, Zambia. Hydrobiologia 315:143–147

Seehausen O (2004) Hybridization and adaptive radiation. TREE 19:198–207

Seehausen O, Koetsier E, Schneider MV, Chapman LJ, Chapman CA, Knight ME, Turner GF, van Alphen JJM, Bills R (2003) Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. Proc R Soc Lond B 270:129–137

Skelton P (1993) A complete guide to the freshwater fishes of Southern Africa. Southern Book Publishers, Halfway House, Harare

Skelton P (2001) A complete guide to the freshwater fishes of Southern Africa. Struik, Cape Town

Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Mol Ecol 14:67–73

Stewart KM (2001) The freshwater fish of Neogene Africa (Miocene–Pleistocene): systematics and biogeography. Fish Fisheries 2:177–230

Sturmbauer C, Baric S, Salzburger W, Rüber L, Verheyen E (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African Lakes. Mol Biol Evol 18:144–154

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Trewavas E (1983) Tilapiine fishes of the genera *Sarotherodon, Oreochromis* and *Danakilia*. Cornell University Press, Ithaca New York, p 583

Turner GF (2003) *Oreochromis (Nyasalapia) karongae* (Trewavas). http://malawicichlids.com/mw10001a.htm. In: ''The Cichlid Fishes of Lake Malawi, Africa'' (M. K. Oliver, webmaster), http://malawicichlids.com. Accessed September 2005

Tweddle D, Willoughby NG (1979) Part 2: An annotated checklist of the fish fauna of the river Shire south of Kapachira Falls, Malawi. Ichthyol Bull JLB Smith Inst Ichthyol 39:10–12

Tweddle D, Lewis DSC (1979) Part1: the nature of the barrier separating the Lake Malawi and Zambezi fish faunas. Ichthyol Bull JLB Smith Instit Ichthyol 39:1–9

Van Houdt JKJ, Pinceel J, Flamand M-C, Briquet M, Dupont E, Volkaert FAM, Baret PV (2005) Migration barriers protect indigenous brown trout (*Salmo trutta*) population from introgressions with stocked hatchery fish. Conserv Genet 6:175–191

Van der Waal BCW, Bills R (2000) *Oreochromis niloticus* (Teleostei: Cichlidae) now in the Limpopo River System. S Afr J Sci 96:47–48

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15:1419–1439

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Welcomme R (1967) Observations on the biology of the introduced species of *Tilapia* in Lake Victoria. Revue Zool Bot Afr 76:249–279

Wohlfarth GW, Hulata GI (1981) Applied genetics of tilapias. ICLARM Stud Rev 6:1–26

Yang ZH (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306–314