

# Eco-morphological differentiation in Lake Magadi tilapia, an extremophile cichlid fish living in hot, alkaline and hypersaline lakes in East Africa

GERALDINE D. KAVEMBE,\*†<sup>1</sup> ANDREAS F. KAUTT,\*‡<sup>1</sup> GONZALO MACHADO-SCHIAFFINO\* and AXEL MEYER\*‡

\*Chair in Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany, †Department of Biology, School of Pure and Applied Sciences, South Eastern Kenya University, P. O. Box 170, 90200 Kitui, Kenya, ‡International Max Planck Research School for Organismal Biology, University of Konstanz, Universitätsstrasse 10, 78457 Konstanz, Germany

## Abstract

Ecological diversification through divergent selection is thought to be a major force during the process of adaptive radiations. However, the large sizes and complexity of most radiations such as those of the cichlids in the African Great Lakes make it impossible to infer the exact evolutionary history of any population divergence event. The genus *Alcolapia*, a small cichlid lineage endemic to Lakes Magadi and Natron in East Africa, exhibits phenotypes similar to some of those found in cichlids of the radiations of the African Great Lakes. The simplicity within *Alcolapia* makes it an excellent model system to investigate ecological diversification and speciation. We used an integrated approach including population genomics based on RAD-seq data, geometric morphometrics and stable isotope analyses to investigate the eco-morphological diversification of tilapia in Lake Magadi and its satellite lake Little Magadi. Additionally, we reconstructed the demographic history of the species using coalescent simulations based on the joint site frequency spectrum. The population in Little Magadi has a characteristically upturned mouth—possibly an adaptation to feeding on prey from the water surface. Eco-morphological differences between populations within Lake Magadi are more subtle, but are consistent with known ecological differences between its lagoons such as high concentrations of nitrogen attributable to extensive guano deposits in Rest of Magadi relative to Fish Springs Lagoon. All populations diverged simultaneously only about 1100 generations ago. Differences in levels of gene flow between populations and the effective population sizes have likely resulted in the inferred heterogeneous patterns of genome-wide differentiation.

**Keywords:** niche width, RAD-seq, site frequency spectrum, soda lakes, stable isotopes

Received 7 July 2015; revision received 1 November 2015; accepted 3 November 2015

## Introduction

Adaptation to ecologically different environments has often been regarded as one of the initial steps leading to population divergence (Nosil 2012). Alongside com-

petition for resources, this process forms the foundation of the ecological theory of speciation (Schluter 2000; Rundle & Nosil 2005). The significance of these two processes in driving divergence and generating biodiversity are best understood in the context of adaptive radiations (Schluter 2000; Losos 2010) for which cichlid fishes are a prime example. Cichlids (mostly the *Haplochromines*) are among the most species-rich vertebrate lineages, with much of their diversity attributable to trophic diversification and adaptive radiations

Correspondence: Gonzalo Machado-Schiaffino and Axel Meyer, Fax: +49 7531883018; E-mails: gonzalo.machado-schiaffino@uni-konstanz.de and axel.meyer@uni-konstanz.de

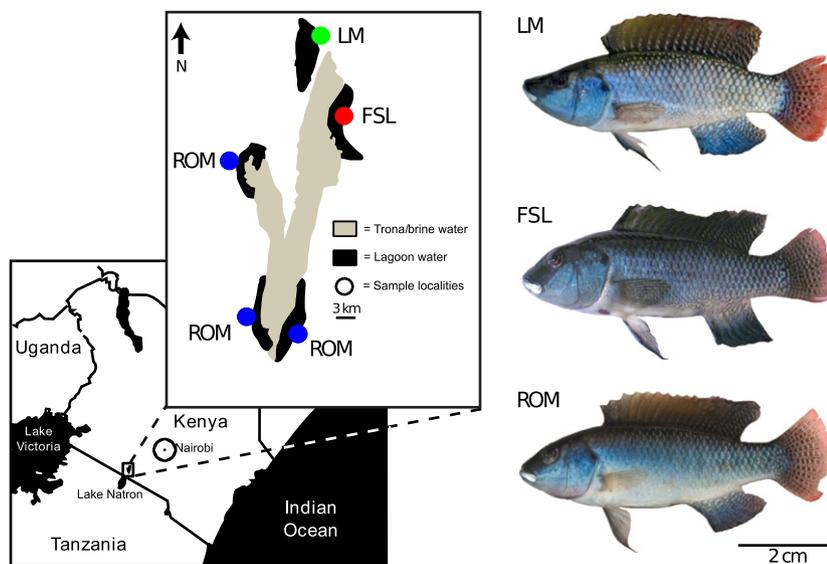
<sup>1</sup>These authors contributed equally to this work.

(Meyer 1993; Rüber *et al.* 1999; Verheyen *et al.* 2003; Kocher 2004; Salzburger *et al.* 2005; McMahan *et al.* 2013; Henning & Meyer 2014). Indeed, the astonishing ability of cichlids to rapidly adapt to different trophic niches has particularly intrigued evolutionary biologists (e.g. reviewed in Takahashi & Koblmüller 2011). Morphological studies mainly conducted on the Tanganyikan cichlids have revealed multiple morphological adaptations in trophic features, including position and orientation of the mouth (Clabaut *et al.* 2007; Takeuchi & Hori 2008; Stewart & Albertson 2010), number, size and shape of teeth (McKaye & Marsh 1983; Yamaoka 1983; Meyer 1990; Trapani 2004), intestine length (Yamaoka 1985; Wagner *et al.* 2009) and jaws (Huysseune 1995; Parsons *et al.* 2011).

While the radiations of cichlids in the African great lakes such as Victoria, Malawi and Tanganyika are intriguing for their unparalleled diversity, the sheer size of the lakes and complexity of their species flocks makes it difficult to disentangle individual evolutionary processes that have shaped their diversity. For example, both allopatric isolation and hybridization have been implicated to play a role in driving speciation in these radiations (Rüber *et al.* 2001; Salzburger *et al.* 2002; Schelly *et al.* 2006; Koblmüller *et al.* 2010; Kirchberger *et al.* 2012). But the relative effects of these processes is difficult to test, as detailed demographic investigations about population splits and gene flow encompassing all potentially interacting populations are virtually impossible for these species flocks composed of hundreds of species. Thus, the relative importance of the processes that generate and maintain the phenotypic diversity of the African cichlids remain debated (reviewed in Stiasny & Meyer 1999; Kocher 2004; Seehausen *et al.* 2008;

Henning & Meyer 2014; Brawand *et al.* 2014; Seehausen 2015). Focusing instead on very young and less species-rich lineages in a confined area (e.g. islands or lakes) seems a promising research approach (Losos & Ricklefs 2009; Elmer *et al.* 2010), because the effects of different evolutionary processes can be discerned and the demographic history can be reconstructed more easily and reliably in these simpler systems.

Soda lake tilapia (Natron and Magadi tilapias) of the genus *Alcolapia* (Seegers *et al.* 1999) form a monophyletic group of four morphologically distinct cichlid species endemic to the Magadi–Natron ecosystem in East Africa. These lakes are arguably some of the world's most extreme aquatic environments: salinity 60% seawater, pH ~10, and temperatures often exceeding 40 °C. Phylogenetically, the Soda tilapias are derived from a freshwater ancestor that inhabited Palaeolake Orolonga (Seegers *et al.* 1999). This palaeolake's water level dropped drastically following abrupt climatic changes in East Africa after the Pleistocene (~13 000 years) leading to a split into two distinct water bodies—Lake Natron and Lake Magadi (Butzer *et al.* 1972; Hillaire-Marcel *et al.* 1986). A further drop in water level led to the split of Lake Magadi and its satellite Lake Little Magadi (LM), although details on the time of formation of LM are scant. Presently, these two lakes are separated by a land barrier of <1 km (Fig. 1). Extant populations of Magadi tilapia, the focal species of this study, are restricted to small isolated pools (lagoons) along the shores of the Lake Magadi basin and its satellite lake LM (Fig. 1). Movement of fish between lakes is restricted by a land barrier, whereas trona (deposits of crystalline sodium carbonate and sodium bicarbon-



**Fig. 1** Map of the sampling locations in Lake Magadi basin, Kenya. Representative fish from each of the populations are shown on the right. Note the upturned mouth in Little Magadi fish (LM) relative to a terminal mouth in the populations from Lake Magadi (Fish Springs Lagoon and Rest of Lake Magadi).

ate salts) and inhabitable water prevent fish movement between lagoons (Coe 1966, 1967; Wilson *et al.* 2004).

Traditionally, soda lake tilapia in Lake Magadi and LM have been assigned to a single species, *Alcolapia grahami*, which has been regarded as morphologically less variable compared to the Lake Natron clade. In comparison, three Natron tilapia species (*Alcolapia alcalicus*, *A. ndalalani* and *A. latilabris*), distinguishable by head shape, mouth position as well as number, shape and position of teeth have been described (Seegers & Tichy 1999; Tichy & Seegers 1999; Seegers *et al.* 2001). However, recent studies have revealed several morphological and anatomical differences between some populations of Magadi tilapia (Wilson *et al.* 2004). Perhaps the most apparent difference is the upturned mouth found among all individual fish in LM compared to the terminal mouth found in Fish Springs Lagoon (FSL) and other populations in the Rest of Lake Magadi (ROM) lagoons (Wilson *et al.* 2004). Interpopulation variation in gut lengths has also been reported in Magadi tilapia. The ROM population which exploits guano-rich substrates displays much shorter guts compared to the FSL population that predominantly browse on algae (Wilson *et al.* 2004; Pörtner *et al.* 2010). Moreover, chemical characterization of the Lake Magadi lagoons has revealed varying water parameters (e.g. temperature and salinity) (Wilson *et al.* 2004). Consequently, the observed morphological differences may represent signs of local adaptations and ecological diversification. This is corroborated by a recent study based on microsatellite markers that revealed three genetically distinct population clusters of Magadi tilapia designated as FSL, ROM and LM (Kavembe *et al.* 2014).

Despite the evidence that Magadi tilapia is not as morphologically invariant as previously thought, no detailed studies of the eco-morphology and demographic history of the species have been conducted. It is currently not known whether there is a correlation between phenotype (e.g. mouth shape) and environment, a key feature of adaptive radiations (Schluter 2000). Furthermore, little is known about their population history, especially the timing and order of population splits, levels of gene flow and, consequently, the speed of population divergence.

Given the known population structure (Kavembe *et al.* 2014) and geographical distribution of the species (Fig. 1), three scenarios seem plausible: (i) all three populations could have split simultaneously, (ii) the LM population could have been isolated from the FSL and ROM populations (inhabiting the main Lake Magadi) before they were themselves separated by trona deposits, and (iii) the barrier imposed by trona

between FSL and ROM, may have preceded the geological separation of LM and the Lake Magadi populations, leading to the split of ROM from a combined population of LM and FSL before they in turn got separated. Earlier studies in the Magadi system have suggested that migration between the lagoons of Lake Magadi is rare leading to genetic differentiation among the populations (Wilson *et al.* 2000, 2004; Kavembe *et al.* 2014). However, a recent study found significant genetic exchange between populations of congeneric species inhabiting various lagoons of Lake Natron (Zaccara *et al.* 2014; Ford *et al.* 2015). Knowledge about the amount of gene flow is essential for the interpretation of population divergence and to draw informed inferences of how this will be reflected at the genomic level (Feder *et al.* 2013; Martin *et al.* 2015).

Here, we use a complementary approach of population genomics, geometric morphometrics, stable isotope (SI) analyses and coalescent simulations to investigate the phenotype–environment correlation in populations of Magadi tilapia and reconstruct the evolutionary history of the diversification process. We quantified morphological differences between the three previously identified populations and tested whether significant differences in body shape exist. Based on the results of the morphometric analyses we investigated whether there is a correlation between morphology and resource utilization, using carbon and nitrogen SI signatures. Using coalescent simulations, we tested different demographic models to establish the model that best fits our genomic data (summarized in the site frequency spectrum) and estimated the associated parameters. In this way, we inferred how and when the three populations were separated, estimated the amount of gene flow between them and calculated the past and contemporary effective population sizes. Finally, we scanned the genomes of the three populations for regions potentially under selection.

## Materials and methods

### *Sampled populations*

Sampling was based on the three previously identified genetic clusters corresponding to the populations of LM, and within Lake Magadi, FSL and ROM (Kavembe *et al.* 2014) (Fig. 1). Information on the samples used is provided in Table S1 (Supporting information). Details of sampling sites and methods are provided in (Kavembe *et al.* 2014). Sampling was carried out with permission (Number: NCST/RR1/12/1/MAS/99/4) of the Kenya National Commission for Science, Technology and Innovation (NACOSTI).

### Body shape analyses

Morphological variation among the Magadi tilapia populations was determined using landmark-based geometric morphometric shape analyses. Standardized images of the left side of individual fishes ( $n$ : FSL = 18, ROM = 24, LM = 18) were taken using a 10 $\times$  megapixel Canon PowerShot D10 digital camera (Canon USA, Inc.) mounted on a tripod stand. Twelve landmarks were digitized on each image in the same order after setting the scale factor in TPSDIG v2 software (Rohlf 1999); for landmarks used, see Fig. S1a (Supporting information). Two-dimensional data extracted from the digitized images were imported into the program MORPHOJ v1.02e (Klingenberg 2011) for shape analyses. Prior to shape analyses, the landmark coordinates were superimposed as shape variables by performing a generalized Procrustes analysis which minimizes the sum of squared distances between homologous landmarks by translating, rotating and scaling them to unit (centroid) size (Rohlf 1999). This step removes variation associated with differences in the scale, position and orientation of specimens to ensure the differences observed between landmarks are only due to shape (Rohlf 1999; Klingenberg 2002). Following superimposition, the procrustes distances become relative measures of shape differences between biological groups (Webster & Sheets 2010). Thus, to determine if populations of Magadi tilapia were statistically distinct in morphospace, procrustes distances among populations were calculated and the significance of each of these distances were assessed using a permutation test with 10 000 iterations.

To assess the effect of size on shape variation among samples (Sidlauskas *et al.* 2011), a multivariate allometric regression was performed in MORPHOJ using procrustes coordinates and standard length as dependent and independent variables, respectively. Due to a significant effect of allometry (see results of shape analysis), downstream analyses were performed using the regression residuals as the new shape variables, thus allowing for meaningful comparisons of the specimens. We used the thin plate spline grid technique (Bookstein 1989; Dryden & Mardia 1998) and the wire frame outline (Klingenberg 2013) to visualize shape changes. The thin plate technique allows the differences between two shapes to be compared as a grid of the first shape (e.g. the average body shape of the individuals from LM population) being deformed into the second (a consensus body shape of all individuals included in the shape analyses) (Dryden & Mardia 1998). The body regions with the largest differences between the groups and the average shape are depicted by the largest distortion on the grid. On the other hand, the wireframe

option compares wireframe graphs that connect landmarks with straight lines for the starting (a consensus body shape of all fish included in the analysis) and the target shapes (population of interest) (Klingenberg 2013).

Interpopulation shape variation between Magadi tilapia samples was determined using a canonical variate analysis (CVA) and discriminant function analyses (DFAs) using predefined population IDs as the classifying variable. CVA allows for identification of shape features that best distinguish multiple groups of specimens (Mardia *et al.* 1979; Albrecht 1980), while DFA performs pairwise comparison of the groups. For the DFA, population predictions were based on a leave-one-out cross-validation (Lachenbruch 1967). Additionally, we performed principal component analyses (PCAs) on our data set (Fig. S1b, Supporting information).

### SI analysis

Muscle was excised from the dorsal left side of individual Magadi tilapia fish ( $n$ : FSL = 11, ROM = 10, LM = 11) for SI analysis of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ . These two isotopes are used for estimating trophic positions and carbon flow to consumers in food webs, respectively (Fry 2006). A main advantage of SI approaches over other conventional dietary methods is that isotope ratios in tissues yield time-integrated dietary information that reflects what has been assimilated and not just ingested materials (Vander Zanden & Rasmussen 1999). White muscle in particular is considered most suitable for teleost SI analysis owing to a long integration time (4–6 months) and low variability in  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  (Pinnegar & Polunin 1999). To reduce isotopic variability commonly associated with differences in muscle lipid content (Wessels & Hahn 2010; Ryan *et al.* 2012), individual samples were treated using a modification of the lipid-extraction method from Bligh & Dyer (1959). We used the same methods for the SI work as in previous studies on cichlids and other fish radiations (Bootsma *et al.* 1996; Elmer *et al.* 2010, 2014; Rutschmann *et al.* 2011).

All statistical analyses were conducted using JMP software v11.2.1 (SAS Institute Inc., Cary, NC, USA, 2010) and considered significant at  $P < 0.05$ . We performed correlation analysis (Pearson's correlation;  $P = 0.05$ ) to test for allometric effects of size on  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values. The Shapiro–Wilk test (Shapiro & Wilk 1965) for normal distribution and Levene's test (Levene 1960) for equal variance were performed on all data sets prior to analysis. All data follow a normal distribution according to the Shapiro–Wilk test ( $P < 0.05$ ), but did not pass the homogeneity test (see Results), thus nonparametric statistics were used to compare the interpopulation

differences in  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ . Specifically, we used the Kruskal–Wallis nonparametric test (Kruskal & Wallis 1952) that allows the comparison of medians among multiple independent groups followed by an all-pair-wise comparison according to Dunn's method (Dunn 1964).

#### *RAD library preparation and sequencing*

Total genomic DNA was extracted from muscle or fin tissues (n: FSL = 16, ROM = 20, LM = 21) using a standard salt extraction method following proteinase K digestion with RNase treatment and a final purification step with QIAmp DNA Mini Kit (Qiagen). For each individual 900  $\mu\text{g}$  of purified genomic DNA was double digested using the rare-cutting enzyme *Pst*I-HF (20 U/reaction) and the frequent cutter *Msp*I-HF (20 U/reaction) in a final volume of 100  $\mu\text{L}$ . For details on the ddRADseq methods as implemented in the Meyer Lab, see our previous publications (Recknagel *et al.* 2013; Henning *et al.* 2014). The genomic libraries were each single-end-sequenced (101 cycles) on one lane of an Illumina HiSeq 2000 platform.

#### *Mapping and genotyping*

Sequence quality was visually inspected with FASTQC and no quality drop-off over the length of the reads was observed. Individually barcoded full-length reads were demultiplexed using the *process\_radtags* script included in the STACKS v.1.20 beta2 software pipeline (Catchen *et al.* 2011, 2013). Additionally, reads containing uncalled bases and/or showing an average quality score of <20 in a sliding window of 10% of the total read length were discarded in this step. The remaining 96 bp reads were then mapped to the anchored *Oreochromis niloticus* genome assembly version ORENIL1.1 (NCBI Assembly GCA\_000188235.2, Brawand *et al.* 2014) with BOWTIE v.1.0.1 (Langmead *et al.* 2009) allowing for a maximum of three mismatches in the seed (-n 3). Only reads mapping to a unique position were retained (-m 1). Loci construction and genotyping was conducted with STACKS. Loci within individuals were formed using a minimum stack depth of five reads. Subsequently population-specific catalogues were built and the *rxstacks* correction module was used to filter out loci with a log-likelihood of <-10, being confounded in more than 25% of the individuals, or showing excessive haplotypes. Additionally, this module uses population-level information to correct individual SNP calls (i.e. the assignment of a homozygote or heterozygote state at a certain position). SNP calling was performed with an upper bound of 0.05 for the error rate and using a 5% significance level cut-off.

On average, data for  $30\,832 \pm 12\,110$  (sd) loci per individual were obtained with a mean coverage of  $25.8 \pm 8.9$  (sd) reads per locus and individual (see Table S1 for more details).

The distribution in the number of SNPs over the lengths of loci showed an increase in the last four base pairs and, strangely, at position 47. Thus, SNPs at these sites were removed from the internal STACKS files using custom bash scripts. Note that this resulted not from a drop of sequence quality, as the pattern was the same even after trimming these bases before mapping and catalogue construction. Tests using different read mapper software indicated that this is a spurious result depending on the mapping procedure (data not shown). To minimize problems due to hidden paralogy, loci deviating from HWE at 5% significance level within populations were excluded from further analyses using the blacklist option in the *populations* module of STACKS. HWE exact tests (Wigginton *et al.* 2005) were performed in PLINK v.1.07 (Purcell *et al.* 2007). Note that using a higher *P*-value is more conservative when testing for a deviation of HWE; that is, more loci potentially under HWE are excluded. Furthermore, overlapping loci and loci containing more than three SNPs within a population, as well as loci that were successfully genotyped in <10 individuals per population were blacklisted and excluded from all subsequent analyses.

#### *Population structure*

Population structure was investigated with ADMIXTURE (Alexander *et al.* 2009) and by means of PCAs with EIGENSOFT (Patterson *et al.* 2006). The support for the number of clusters found by ADMIXTURE was evaluated by 10 iterations of the implemented cross-validation procedure. For PCAs, statistical significance of eigenvectors (principal components) was determined based on the build-in Tracy–Widom statistics. Missing data were accounted for by solving least square equations as implemented in EIGENSOFT (*lsproject* option). Only one SNP per RAD-tag locus was used for these analyses to reduce the effect of nonindependence among markers due to linkage. PCAs were visualized in R v.3.1.2 (R Core Team 2014).

#### *Demographic inference*

Demographic estimates were obtained by simulation and fitting of different demographic models to the information contained in the multidimensional joint minor site frequency spectrum (MSFS) as implemented in FASTSIMCOAL2 (Excoffier *et al.* 2013). To reduce the potentially biasing effect of selection, loci presumably located in coding regions, identified by a *blastx* search against the protein-coding database of *Oreochromis*

*niloticus*, were excluded (cut-off: e-value  $<1e^{-10}$ ). Again, only one SNP per locus was used for these analyses to reduce the effect of linkage. The MSFS was created as follows: data were parsed from variant call format (VCF) files using a custom python script and transformed into the MSFS using *dadi* (Gutenkunst *et al.* 2009). To account for missing data, the MSFS was projected down to a minimum number of 20 alleles per population (as specified in *STACKS* to create the VCF files). The number of monomorphic sites (i.e. invariable among all populations and individuals) was manually added to the MSFS and theoretically equals the respective number of loci times the 86 potentially variable sites (obtained by subtracting the 5 bp of restriction site and the five manually edited sites from the 96 bp reads) minus the number of segregating sites. As using only one SNP per locus decreases the ratio of polymorphisms and thus biases the estimates, we corrected for this by first calculating the ratio of monomorphic to polymorphic sites using all SNPs. The resulting number of monomorphic sites is then the number of SNPs (using one per locus) multiplied by this ratio culminating in an effective sequence length of 1.054 Mb with 5433 segregating sites. To convert the inferred parameters into demographic units, the default substitution rate of  $2.5 \times 10^{-8}$  per site and generation was assumed (Excoffier *et al.* 2013).

To infer the demographic scenario leading to the divergence of the post-Orolonga Magadi tilapia into the three genetic clusters found in (Kavembe *et al.* 2014), we tested three different main plausible hypotheses: (i) a simultaneous split of the three populations; (ii) LM split first from a common ancestor, which later diverged into FSL and ROM populations; and (iii) isolation of FSL and ROM by trona preceded the allopatric isolation of LM and Lake Magadi. We denote the three scenarios 'sim-split', 'LM-first' and 'ROM-first', respectively. For each of these three main scenarios several different demographic models were tested. Migration between demes can be symmetrical or asymmetrical. Population size changes can be positive or negative and can occur in the ancestral Magadi tilapia population or each of the three populations separately, as well as coincide with population splits. In more detail, allowing for a change in the ancestral population was denoted as 'ANC-change', and changes in the three populations accordingly as 'FSL-change', 'ROM-change' and 'LM-change'. A population size change during a population split was denoted as 'split-change'. In the case of the 'sim-split' scenario there is only one population split, whereas in the 'LM-first' and 'ROM-first' scenarios there are two splits. In this case, a change during the 'older/first' split (forward in time) was denoted as 'split1-change', and during the 'younger/second' split as 'split2-change' (Fig. S2, Supporting information).

Migration parameters were denoted by the number of different migration rates that were included in a model, that is, the term 'single-mig' denotes a model with only a single migration rate. However, this does not mean that there is only migration between two of the populations. Instead, migration occurs between all the three demes, but its rate is symmetrical and identical in all instances. Accordingly, the term 'six-mig' denotes a model that includes six different migration rates. In the case of the 'LM-first' and 'ROM-first' scenarios, two additional migration parameters between the combined ancestral population of two of the populations and the population that split 'first' are possible. These models are denoted as 'eight-mig'. After running the 'six-mig' and 'eight-mig' models, it became clear that migration rates between FSL and LM, and ROM and LM were almost identical. Thus, in addition, a 'two-mig' setting with a symmetrical migration rate between FSL and ROM, and another symmetrical migration rate between FSL and LM, as well as between ROM and LM was implemented. Models without migration are denoted as 'no-mig'. The basic demographic events and parameters are visualized in Fig. S2 (Supporting information). Naturally, several population size changes can occur in any one model and in any combination with migration rates.

Starting with simple models and adding migration rates and population size changes, in total, 19, 22, and 22 models for the 'sim-split', 'LM-first' and 'ROM-first' scenarios were tested, respectively. The support for different models to fit our data was evaluated based on the Akaike Information Criterion (AIC) after converting the log-likelihood values to ln-likelihoods. For each model, 25 independent runs with 40 cycles each of the implemented ECM optimization algorithm were performed. The first 20 cycles always consisted of 100 000 coalescent simulations each, following which step size was increased by 5000 simulations per cycle up to the specified maximum number of cycles. The single best models for the three main scenarios were analysed in a total of 100 independent runs with 50 cycles each, to enhance accuracy of parameter estimates. Uncertainty in parameter estimates was gauged from 100 parametric bootstrap replicates and is reported as 95% confidence intervals. Upper and lower bounds were calculated as  $\theta \pm 1.96\sigma$ , where  $\theta$  denotes the mean and  $\sigma$  the standard deviation of the bootstrap results. Briefly, bootstrap replicates were obtained by simulating 100 minor site frequency spectra, based on the same overall corrected sequence length as the empirical data (in unlinked regions of 86 bp) and according to the highest likelihood parameter point estimates, and then re-estimating the parameters each time. For each bootstrap replicate, we performed 25 independent runs with 30 cycles each.

### Genome scans

Genome-wide differentiation was analysed in terms of AMOVA  $F_{ST}$  values (Weir & Cockerham 1984) as implemented in the populations script of STACKS. The relatively low frequency of polymorphic sites precluded a window-based approach. Instead, the position-independent approach implemented in BAYESCAN v2.1 (Foll & Gaggiotti 2008; Fischer *et al.* 2011) was used to scan for signatures of selection between the different Magadi tilapia populations. All three populations were used together in a single analysis with default settings. Outlier status was evaluated based on a 5% false discovery rate (FDR) threshold.

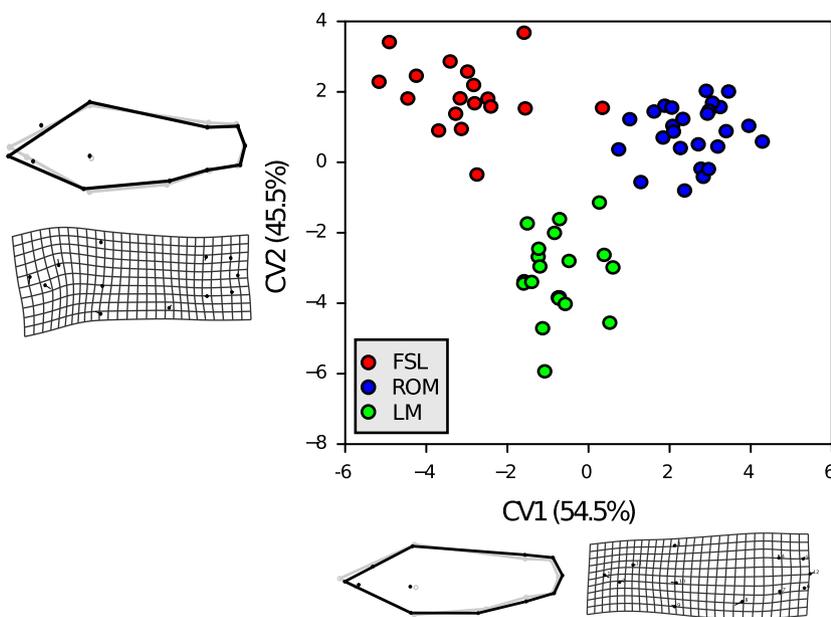
## Results

### Body shape analyses

Regression of the superimposed shape data on size (standard length) indicated that 7.9% of shape variation in Magadi tilapia covaries with size. Importantly, the relationship between the procrustes residuals and size is statistically significant ( $P < 0.0001$  at 10 000 permutations) for the entire data set. The permutation test for procrustes distance between groups found statistically significant distances for all pairs of populations: a similar distance value of 0.03 was found between the comparisons of LM vs. FSL, and LM vs. ROM. The procrustes distance between FSL and ROM was slightly higher with 0.0327 ( $P < 0.0001$  for 10 000 permutation tests for all the three group comparisons).

Canonical variate analysis of the residuals of the multivariate regression resulted in a clear separation of the three a priori defined populations of Magadi tilapia in morphospace. Based on the CVA, there is no overlap of specimens from different populations (Fig. 2). Canonical covariate 2, which explains 45.5% of the total variance, clearly separates LM from the populations of FSL and ROM driven mainly by a change in the position of the mouth (Landmark 1) (Fig. 2). This pattern is consistent with the two distinct head shapes previously observed in Magadi tilapia (an upturned mouth in Lake LM vs. a terminal mouth among the populations of FSL and ROM) (Fig. 1; Wilson *et al.* 2004; Kavembe *et al.* 2014). The first canonical axis (CV1), which explains 55.5%, suggested a difference in shape due to elongation as well as on the caudal peduncle (Fig. 2). The distinctiveness of the populations was however not particularly obvious in the PCA (Fig. S1b, Supporting information). As PCA employs a hypothesis-free approach and assumes no a priori groupings it is prone to biases especially when dealing with data with unequal variances. For instance, populations with the greatest amount of variance may bias interpretations of the common morphospace, rendering PCA results of little significance in capturing the biologically important shape differences (Parsons *et al.* 2009).

Further differences in morphospace between populations were highlighted using DFA (Fig. S3, Supporting information). Between FSL and ROM, relative to LM, most differences were associated with the position of the mouth (Landmark 1). FSL and ROM differed by an elongation of the caudal peduncle (landmarks 4, 5, 6, 7). Complementarily, based on the results of the discrimi-



**Fig. 2** Canonical variance analysis (CVA) based on geometric morphometrics. Wireframe plots beside the axis show the main shape changes.

nant function analysis (DFA), most of the specimens could be assigned to their source populations. The highest similarity according to Mahalanobis distances was between FSL and ROM, and the lowest similarity was found between FSL and LM (Mahalanobis distances per pairwise comparisons: FSL-LM 6.787, FSL-ROM 5.615 and LM-ROM 6.608 ( $P < 0.0001$  for 10 000 permutation tests for all the three group comparisons).

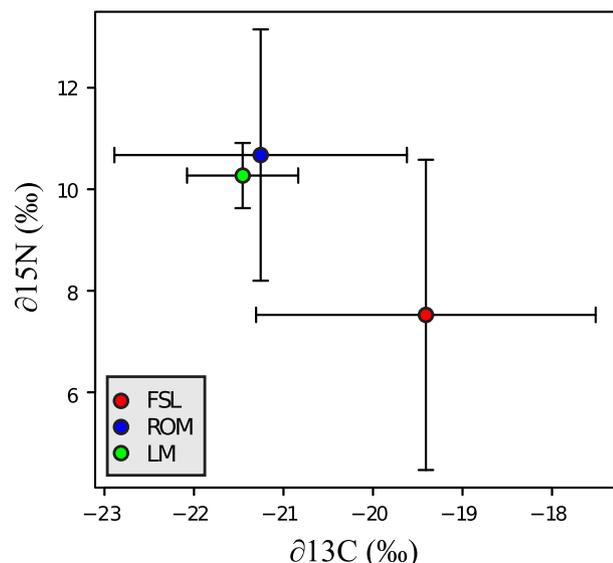
### Ecological divergence

In isotopic niche space, FSL and ROM populations were indistinguishable, but the two were clearly distinct from LM. The mean isotopic values were  $-19.41\text{‰}$ ,  $-21.26\text{‰}$ ,  $-21.46\text{‰}$  for  $\delta^{13}\text{C}$  and  $7.52\text{‰}$ ,  $10.67\text{‰}$ ,  $10.27\text{‰}$  for  $\delta^{15}\text{N}$  in FSL, ROM and LM, respectively. Fish size (standard length) was not a significant covariate and was thus omitted from the final model (Pearson's coefficient,  $P = 0.05$ ). All our data sets were normally distributed (Shapiro–Wilk test,  $P < 0.05$ ). Levene's test of homogeneity of variance returned statistically significant differences ( $\delta^{13}\text{C}$   $F_{2,29} = 3.941$ ,  $P = 0.031$ ;  $\delta^{15}\text{N}$   $F_{2,29} = 11.634$ ,  $P = 0.000$ ) suggesting unequal population variances for both isotopes. The smallest variance was found for LM ( $\delta^{15}\text{N} = 0.423$ ;  $\delta^{13}\text{C} = 0.387$ ) and the largest for FSL ( $\delta^{15}\text{N} = 9.363$ ;  $\delta^{13}\text{C} = 3.603$ ) followed by ROM ( $\delta^{15}\text{N} = 6.1$ ;  $\delta^{13}\text{C} = 2.667$ ) (Fig. 3). ROM, on average, had the highest values of  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$ , suggest-

ing its fish could be exploiting high trophic levels. Overall we found a high intraspecific range of isotopic signatures in Magadi tilapia ( $\delta^{15}\text{N} = 9.16\text{‰}$ ;  $\delta^{13}\text{C} = 5.64\text{‰}$ ). Among populations, the isotopic ranges were highest in FSL ( $\delta^{15}\text{N} = 8.27\text{‰}$ ;  $\delta^{13}\text{C} = 5.64\text{‰}$ ), followed by ROM ( $\delta^{15}\text{N} = 6.82\text{‰}$ ;  $\delta^{13}\text{C} = 4.35\text{‰}$ ) and the least in LM ( $\delta^{15}\text{N} = 2.34\text{‰}$ ;  $\delta^{13}\text{C} = 2.363\text{‰}$ ). There was a strong population effect on the isotopic signatures of both  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  (Welch's test,  $\delta^{13}\text{C}$  F-ratio = 5.524,  $P = 0.016$ ;  $\delta^{15}\text{N}$  F-ratio = 4.257,  $P = 0.036$ ). The mean C:N ratios (% of weight) were 3.11, 3.17 and 3.22 in FSL, ROM and LM, respectively. All these values were within the expected range for biological proteins (DeNiro 1985). Comparison of all the populations combined revealed significant variation of both isotopes (Kruskal–Wallis,  $H = 6.609$ ,  $P = 0.0261$  for  $\delta^{15}\text{N}$ ;  $H = 7.291$ ,  $P = 0.036$  for  $\delta^{13}\text{C}$ ). Pairwise comparisons showed a significant difference between ROM and FSL for  $\delta^{15}\text{N}$  (Dunn's test  $Z = 2.525$ ,  $P = 0.035$ ), while the rest of the population comparisons tests were nonsignificant ( $P$ -values  $> 0.05$ ). We found a marginal level of significance between ROM and FSL using  $\delta^{13}\text{C}$  data (Dunn's test  $Z = -2.285$ ,  $P = 0.066$ ) and no differences among the rest of the pairwise population comparisons of the  $\delta^{13}\text{C}$  data (data not shown).

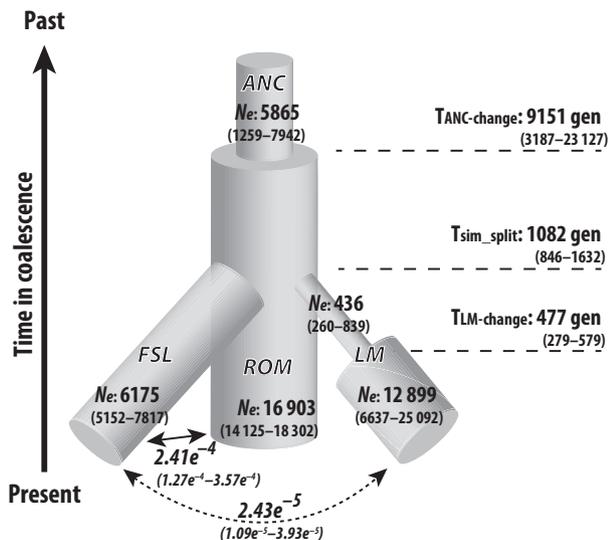
### Demographic inference supports a recent simultaneous split and different levels of gene flow among the Magadi tilapia populations

As all subsequent genetic analyses hinge on the correct assignment of individuals, we first sought to test whether our samples would be assigned into the three previously defined genetic clusters (Kavembe *et al.* 2014). To do this, we investigated the population structure of Magadi tilapia with 7171 SNPs using the software ADMIXTURE v 1.23 and with a model-free principal component analysis (PCA). Assuming the same number of clusters as populations ( $K = 3$ ), ADMIXTURE groups all samples into three distinct genetic clusters that match perfectly with the three previously assigned populations (Fig. S4, Supporting information). However, with cross-validation tests, two rather than three genetic clusters are most supported corresponding to the two lakes (Fig. S4, Supporting information). The model-free PCA revealed three distinct clusters. Only the first two eigen-vectors (PCs) are significant with  $P$ -values of  $3.87 \times 10^{-33}$  and  $1.86 \times 10^{-2}$  for PC1 and PC2, respectively. PC1 explains 9.34% of the variation and separates LM from the other two populations, whereas PC2 explains 3.82% of the variation and separates FSL from ROM (Fig. S5, Supporting information). Neither of the two analyses showed evidence for admixed individuals.



**Fig. 3** In comparison with Fish Springs Lagoon (red) and Rest of Magadi (blue), the population in Little Magadi (green) has a narrower ecological niche width as shown here by the mean  $\pm$  standard deviation (SD) of  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$  stable isotope signatures.

To reconstruct the demographic history of the three populations we used coalescent simulations. For each of the three main scenarios ('sim-split', 'LM-first', 'ROM-first') we explored the fit of different migration rates as well as population size changes in the ancestral population, in the extant populations, and during population splits/mergers to our data. Overall, the best model for our data set, namely the simultaneous split of the three populations, has a 43% probability of being correct. The model contains 10 parameters, and is denoted as 'sim-split, ANC-change, LM-change, two-mig, no-MIG ROM-LM'. It can be verbally described as follows: the ancestral population ( $N_{ANC}$ ) of 5865 (1259–7942; 95% CI) individuals undergoes an about threefold population size expansion to 16 903 (14 125–18 302) individuals about 9151 (3187–23 127) generations ago ( $T_{ANC-change}$ ); this is also equivalent to the current population size of ROM (see below and Fig. 4). The three populations then split simultaneously 1082 (846–1632) generations ago ( $T_{sim-split}$ ). The effective population sizes of FSL and ROM remains stable with 6175 (5152–7817) and 16 903 (14 125–18 302) individuals, respectively, whereas LM undergoes a substantial increase from 436 (260–839) ( $N_{LMpast}$ ) to 12 899 (6637–25 092) individuals



**Fig. 4** Schematic representation of the most supported demographic model. Looking from the past to the present (top to bottom), first, the ancestral population (ANC) undergoes an increase in size at  $T_{ANC-change}$ . The three populations then split simultaneously at  $T_{sim-split}$ . The population sizes of Fish Springs Lagoon and Rest of Lake Magadi (ROM) remain stable whereas Little Magadi (LM) undergoes a substantial increase in size at  $T_{LM-change}$ . Effective population sizes ( $N_e$ ) are given in number of diploid individuals and times are given in number of generations. Arrows indicate migration rates, which are symmetrical and denote the migration probability of an allele per generation. Note that there is no migration between ROM and LM.

( $N_{LMpresent}$ ) about 477 (279–579) generations ago ( $T_{LM-change}$ ).

Migration rates between both FSL and ROM ( $MIG_{FSL-ROM}$ ), as well as FSL and LM ( $MIG_{FSL-LM}$ ) are symmetrical, yet they differ in an order of magnitude with  $2.41 \times 10^{-4}$  ( $1.27-3.57 \times 10^{-4}$ ) and  $2.43 \times 10^{-5}$  ( $1.09-3.93 \times 10^{-5}$ ), respectively. There is no migration between ROM and LM. Migration rates give the probability for an allele to migrate from one deme to another per generation. The actual number of effective immigrants (forward in time) is a product of the migration rate and the effective population size of the population of interest. In our case, this translates to 1.5 migrants per generation from ROM to FSL ( $2.41e^{-4} \times 6175$ ) and 4.1 migrants per generation from FSL to ROM. In the case of migration from FSL into LM the population size change has to be taken into account. Accordingly, we inferred 0.3 and 0.01 (on average 0.14) FSL migrants per generation for the last 477 and 605 generations until the population split, respectively. The number of migrants from LM into FSL is 0.15 per generation.

Importantly, the next best models in our data set (Table S2, Supporting information) are qualitatively and quantitatively similar to the model described above. The five best models capture 92% of the overall probability and differ from each other only in the presence of one vs. two population splits and the presence or absence of migration between ROM and LM (Table S2, Supporting information).

#### Genome-wide differentiation and signatures of selection

Overall genetic differentiation is low with average pairwise  $F_{ST}$  values of 0.027 (FSL vs. ROM), 0.090 (FSL vs. LM) and 0.071 (ROM vs. LM). The profiles of genetic differentiation across the genome are qualitatively similar between FSL vs. LM and ROM vs. LM, whereas FSL vs. ROM exhibits less differentiation (Fig. S6, Supporting information). The difference in heterogeneity is also reflected in the standard deviations of  $F_{ST}$  values of 0.133, 0.116 and 0.033 of the above comparisons, respectively. In each of the two comparisons including LM, five SNPs are fixed (two are shared), whereas the highest  $F_{ST}$  value between FSL and ROM is only 0.516. As the frequency of polymorphic sites was too low for a window-based outlier detection approach, we decided to use a position-independent method (see methods). Only one SNP was found to be potentially under divergent selection with this method—this SNP mapped to position 15 159 712 bp in linkage group six (LG6) of the Nile tilapia genome. The SNP was one of the two sites alternately fixed in LM as compared to both FSL and ROM. A close examination of the corresponding

genomic region within the vicinity of the SNP (10 kb upstream and downstream) in Nile tilapia did not reveal any gene that we could associate to the eco-morphological divergence of LM. Further examination of the other SNPs that were fixed between LM and the other Magadi tilapia populations revealed functions related to physiology (Table S3, Supporting information).

## Discussion

The cichlid fishes (genus *Alcolapia*) of the Magadi–Natron basin are a more amenable model system to study the processes involved in ecological diversification and adaptive radiations compared to the exceedingly species-rich adaptive radiations of cichlids in the African Great Lakes. Here, we perform for the first time a detailed and integrative examination of the eco-morphology and demographic history of three populations of Magadi tilapia. Our data suggest that the three Magadi tilapia populations separated simultaneously only about 1100 generations ago (Fig. 4). There is virtually no gene flow between LM and the other two populations of FSL and ROM, whereas gene flow between the latter two is in the order of a few migrants per generation (Fig. 4). Furthermore, fish in LM are not only genetically most distinct (Figs S4 and S5, Supporting information), but also exhibit a characteristically upturned mouth and a strikingly narrow ecological niche width (Fig. 3). We find subtle differences in body shape between FSL and ROM, consistent with the low genetic differentiation between them. We interpret these results as signs of early stages of diversification, with fish in LM probably being specialized for catching prey items that fall on the water surface.

### *Eco-morphological divergence*

Our data show that the Magadi tilapia in LM not only have a characteristic upturned mouth shape, but also a strikingly narrow ecological niche width compared to the other two populations. Other fish species with upturned mouths often feed on insects from the surface of the water column (Keast & Webb 1966; Langerhans *et al.* 2004) [e.g. *Thoracocharax stellatus* (Netto-Ferreira *et al.* 2007)]. Indeed, the surface film in the soda lakes made up of bacteria, insects and other organic matter blown on the water by wind has been suggested as a source of food for soda tilapia (Seegers & Tichy 1999). Additionally, during sampling, numerous flying insects were seen on the shores of LM (G.D.K., *personal observation*). Thus, we propose that the tilapia in LM might have evolved an upturned mouth as an adaptation to exploit this resource. Further, we hypothesize that the

occurrence of a high number of insects in LM compared to the lagoons within Lake Magadi could be attributed to differences in water conditions; as the former has more extreme conditions compared to the latter (Wilson *et al.* 2004). Differences in niche space among the Magadi tilapia populations is further supported by a narrow niche width in LM compared to FSL and ROM consistent with patterns observed in specialized and generalist consumers, respectively (Bearhop *et al.* 2004; Bolnick *et al.* 2007; but see Flaherty & Ben-David (2010)). While we also find significant differences in body shape between FSL and ROM, no obvious shape–environment link could be inferred for these populations.

Trophic traits such as jaws and craniofacial morphology have a strong genetic basis in cichlids (Albertson *et al.* 2005; Albertson & Kocher 2006; Roberts *et al.* 2011). In concordance, comparisons of teeth in wild and aquarium-raised *Oreochromis ndalalani* (sixth generation) of Lake Natron found no differences in dentition despite their difference in diet (Seegers *et al.* 2001). Nonetheless, one potential caveat of our study is that we cannot rule out the role of phenotypic plasticity in contributing to eco-morphological differentiation among the Magadi tilapia. Yet, while plasticity may play a role it seems unlikely to be the sole explanation for the pronounced differences we find in regard to the position of the mouth in LM compared to the other Magadi tilapia populations.

An intriguing finding was the exceptionally broad ranges of  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$  values in ROM and FSL, relative to reported intrapopulation ranges of 3–5‰ for N (Minagawa & Wada 1984) and 1–3‰ for C (DeNiro & Epstein 1978). In particular, the broad range and the high  $\delta^{15}\text{N}$ -values in more than half of the individuals in ROM are puzzling. However, SI analyses of the alkaline Crater Lake Empakai inhabited by flamingos in Northern Tanzania have revealed high levels of  $\delta^{15}\text{N}$  (Muzuka 2009) and there are suggestions that nitrogen-rich flamingo guano might constitute a major diet for some populations of Magadi tilapia (Wilson *et al.* 2004; Pörtner *et al.* 2010). Thus, the high enrichment in  $\delta^{15}\text{N}$  may be associated with exploitation of flamingo guano.

### *Pronounced genetic differentiation despite microscale isolation*

Consistent with our previous study based on microsatellite loci (Kavembe *et al.* 2014), using a panel of 7171 SNPs we find evidence for three genetic clusters: FSL, ROM and LM. The highest differentiation is clearly between LM and the other two populations, as reflected by the consistency in the results of pairwise

$F_{ST}$  comparisons, PCA, and ADMIXTURE analyses. Genetic differentiation between FSL and ROM is very low. Indeed, the support for only two genetic clusters ( $K = 2$ )-LM being one cluster and the other two being combined-in our ADMIXTURE analysis is slightly higher than for three clusters ( $K = 3$ ). However, in light of the perfect correspondence of the three populations and genetic clusters assuming  $K = 3$  and their clear separation in the PCA, we suggest that the rather subtle differentiation between FSL and ROM does not invalidate the biological reality of their distinctiveness; thus the Magadi tilapia populations may represent a case of incipient speciation. As our demographic analyses suggest a simultaneous split of the Magadi tilapia populations, the most parsimonious explanation for the high genetic differentiation in LM relative to the other populations is a combination of differences in gene flow and its past small population size. In other words, the effect of genetic drift would be much stronger in LM and not be counteracted by gene flow leading to a pronounced differentiation of LM (e.g. Whiteley *et al.* 2010; Méndez *et al.* 2011). In contrast, gene flow between FSL and ROM coupled with their relatively large and stable population sizes would maintain the allele frequencies in both populations at relatively similar levels.

Although the possibility of gene flow between FSL and ROM populations has been questioned (Wilson *et al.* 2004; Kavembe *et al.* 2014), migration is strongly supported in our demographic model; our otherwise best model without migration is only  $2.3 \times 10^{-5}$  times as likely as the one with migration (Table S2, Supporting information). Furthermore, the inferred number of migrants between FSL and ROM populations (1.4 and 4.1 per generation), is quite similar to the 0.5–2.3 migrants per generation reported between isolated populations in Lake Natron using microsatellite data (Zaccara *et al.* 2014) and recently using ddRADseq (Ford *et al.* 2015). In these latter studies, the authors suggested that intense rains or movement of fish by large piscivorous birds may facilitate migration of fish within Lake Natron. As a land barrier or trona proposed in Kavembe *et al.* (2014) would not hinder birds' movement, yet we find almost no migration between LM and the Lake Magadi (FSL and ROM combined) populations, but significant migration among the latter two, our results suggest migration in Magadi tilapia may occur primarily through intermittent water connections following heavy rains. Indeed, Lake LM is at about 10 m higher elevation than Lake Magadi which probably hinders intermittent water connections of the lake to the Lake Magadi lagoons even during the rare flood events suggested in this region. Moreover, unlike Lake Magadi, which has no single outlet or inlet, Lake Natron is connected to several rivers and streams that

drain water into the lake. This means that during periods of heavy rains and occasional floods, Lake Natron may experience abrupt mixing of water compared to Lake Magadi and as a consequence have frequent exchange of fish between lagoons leading to the high levels of gene flow among its species/populations.

While the small land barrier (of less than a kilometre at the closest range) between Lake Magadi and the LM may play a role in restricting gene flow between their populations, other possible mechanisms may also play a role. For instance, actual migration (not gene flow) into LM might occur to a larger extent, but appear to be much more reduced due to selection against immigrants (Nosil *et al.* 2005; Plath *et al.* 2013) as a result of ecological differences between the habitats. Investigating whether this process contributes to reduced gene flow in this system seems an interesting research objective, but is beyond the scope of this study.

#### *The Magadi tilapia demographic history*

The simplicity of the Magadi tilapia system allowed us to infer the demographic history of all its populations in a single framework. This is important, as excluding populations from an analysis can bias demographic inferences of a species' evolutionary history (Beerli 2004; Slatkin 2005). For instance, gene flow between LM and ROM is strongly supported in a pairwise comparison, but not in a three-population comparison (data not shown). This might suggest that some level of gene flow occurs between LM and ROM, but indirectly through FSL [stepping stone model (Kimura 1953)]. Another advantage of this system is the availability of geological data (Seegers & Tichy 1999), which allows a correlation of our inferred demographic events to the geological history of the region. For instance, the separation of the Palaeolake Orolonga into Lake Natron and the Lake Magadi basins is estimated to have occurred about 9000 years ago following periods of extreme drought towards the end of the Pleistocene (Butzer *et al.* 1972; Tichy & Seegers 1999). Correspondingly, we find support for an expansion of the ancestral Magadi tilapia population about 9000 generations ago (Fig. 4). Assuming one to maximally two generations per year places this expansion to sometime after the drought, consistent with recovery patterns observed in populations that have undergone periods of extreme environmental stress (Genner *et al.* 2010; Koblmüller *et al.* 2011). Similarly, our estimate of a simultaneous split only about 1100 generations is congruent with a suggested major flood in the Magadi region about 1450 years ago (White 1953). Although it is possible that the populations may have been isolated prior to the proposed flood, any signals of prior isolation might have been overridden by

subsequent genetic homogenization facilitated by the drought.

With regard to the translation of our demographic estimates to absolute times (in years), we caution that the accuracy of the time estimates hinges on the availability of a good estimate of the generation time in this species. Although generation times of 6 months have been proposed for most tilapia species (Philippart & Ruwet 1982), much shorter times of about 42–60 days have been suggested for Magadi tilapia (Coe 1966; Wilson *et al.* 2004). For instance, applying the 42–60 days generation time estimates to our results would translate into much more recent times for the inferred demographic events in Magadi tilapia (e.g. 180 years for the populations split with six generation per year). Considering the known geological history and the biology of reproduction of this and other cichlid species, these values appear to be unrealistically low. Yet, this discrepancy might be explained by the fact that, like in any other population genetic study, demographic parameters scale by the neutral substitution rate, which is often associated with high uncertainty. Although the absolute values might change depending on the assumed substitution rate, this does not change their relative values nor does it affect the model likelihoods and thus our main conclusions.

Nonetheless, our estimates of population sizes are close to the estimated census sizes of >10 000 individuals for FSL and *c.* 32 000 individuals for the Lake Magadi (FSL and ROM) populations combined (Wilson *et al.* 2004). Unfortunately, no census estimate is available for LM. However, using mtDNA sequence data, Wilson *et al.* (2000) estimated an (long-term) effective population size of only a few hundreds and found evidence for recent population expansion for the population, which is congruent with the expansion inferred in our study. Whether the dramatic population expansion in LM is related to its adaptation to a new niche (floating surface biota) is an interesting hypothesis that needs further investigation.

#### *Genome-wide patterns of population differentiation and outlier loci*

Not surprisingly, the patterns of differentiation across the genome are consistent with the inferred demographic history. The occurrence of gene flow between FSL and ROM and their relatively large stable population sizes is paralleled by low and rather invariable genomic differentiation between them. In LM, genetic drift has not only resulted in a high average genetic differentiation, but also a more heterogeneous pattern. The single locus that was found to be potentially under selection resides in a region in the Tilapia genome that does not seem to contain a clear candidate gene with a function related to the eco-morphological divergence we found. Interestingly,

however, the four other fixed SNPs fall within genes that have putative functions related to physiology such as signalling (e.g. Wong *et al.* 2001; McElhinny *et al.* 2008) and neural transduction (e.g. Nguyen *et al.* 2012) or in chemical detoxification (e.g. Lopez-Huertas *et al.* 2000) (Table S3, Supporting information). While these genes might be interesting candidates for local adaptation in Magadi tilapia, further investigations are required. Nonetheless, Magadi tilapia have evolved eco-morphological adaptations within a very short time of 1,100 generations only and here we provide an account of the demographic details behind this process. Thus, future genome scans with a higher resolution of markers and taking our results into account may be a promising research avenue.

#### *Magadi tilapia as a system for studying the early stages of ecological diversification*

Our data provide evidence that the ancestral Magadi tilapia population split simultaneously into three populations only about 1100 generations ago. Following the split, the LM population has evolved a distinct upturned mouth, which we suggest might be an adaptation to feed on prey items on the water surface. Fish in FSL and ROM seem to be more generalist feeders, probably exploiting a wider range of food items. Eco-morphological differences exist and are consistent with knowledge about their ecology, but are more subtle. Gene flow between the lagoons is happening and maintains the genetic similarity of the populations of FSL and the ROM, whereas only 1 km of land are sufficient to block almost all gene flow with LM. The Magadi tilapia provide a remarkable example of how fast trophic diversity (especially in mouth shapes) in cichlids can evolve to exploit different ecological niches in probably <1100 generations—and lead in <1000 years to three genetically and morphologically distinguishable cichlid lineages that might be considered an example of incipient speciation.

#### **Acknowledgements**

We thank Tata Chemicals Magadi, National Commission of Science and Technology and the Department of Fisheries (Kenya) for support during sample acquisition. We are grateful to the Limnological Institute at the University of Konstanz, and especially Elizabeth Yohannes, for help with the SI analyses. GDK is supported by the MOHEST (Kenya)—DAAD (Germany) collaborative programme. AFK is supported by the Landesgraduiertenförderung (LGFG) of the state of Baden-Württemberg and the International Max Planck Research School (IMPRS) for Organismal Biology. GMS is supported by an Alexander von Humboldt Foundation postdoctoral fellowship and a grant from the Deutsche Forschungsgemeinschaft (MA 6144/1-1). We thank Romulus Abila, Ben Murage, Ann

Mokoro, Achiya J. Onyango, the Magadi team and Dishon Muthee for important support during sampling. This work was partly performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and Arts and the Universities of the State of Baden-Württemberg, Germany, within the framework programme bwHPC, and partly on the High Performance Computing (HPC) cluster of the University of Konstanz. The study was funded partly by support of the University of Konstanz, grants of the Deutsche Forschungsgemeinschaft to AM and by a NSERC Canada Discovery grant to Chris Wood.

## References

- Albertson RC, Kocher TD (2006) Genetic and developmental basis of cichlid trophic diversity. *Heredity*, **97**, 211–221.
- Albertson RC, Streebman JT, Kocher TD, Yelick PC (2005) Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 16287–16292.
- Albrecht GH (1980) Multivariate analysis and the study of form, with special reference to canonical variate analysis. *American Zoologist*, **20**, 679–693.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.
- Bearhop S, Adams CE, Waldron S, Fuller RA, Macleod H (2004) Determining trophic niche width: a novel approach using stable isotope analysis. *Journal of Animal Ecology*, **73**, 1007–1012.
- Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, **13**, 827–836.
- Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, **37**, 911–917.
- Bolnick DI, Svanbäck R, Araújo MS, Persson L (2007) Comparative support for the niche variation hypothesis that more generalized populations also are more heterogeneous. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10075–10079.
- Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 567–585.
- Bootsma HA, Hecky RE, Hesslein RH, Turner GF (1996) Food partitioning among Lake Malawi nearshore fishes as revealed by stable isotope analyses. *Ecology*, **77**, 1286–1290.
- Brawand D, Wagner CE, Li YI *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375–381.
- Butzer KW, Isaac GL, Richardson JL, Washbourn-kamau C (1972) Radiocarbon dating of East African Lake levels. *Science*, **175**, 1069–1076.
- Catchen JM, Amores A, Hohenlohe P *et al.* (2011) Stacks: building and genotyping loci de novo from short-read sequences. *Genes, Genomes, Genetics*, **1**, 171–182.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
- Clabaut C, Bunje PME, Salzburger W, Meyer A (2007) Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. *Evolution*, **61**, 560–578.
- Coe MJ (1966) The biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya. *Acta Tropica*, **XXIII**, 146–175.
- Coe M (1967) Local migration of *Tilapia grahami* Boulenger in Lake Magadi, Kenya in response to diurnal temperature changes in shallow water. *East African Wildlife Journal*, **5**, 171–174.
- DeNiro MJ (1985) Postmortem preservation and alteration of in vivo bone collagen isotope ratios in relation to palaeodietary reconstruction. *Nature*, **317**, 806–809.
- DeNiro M, Epstein S (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta*, **42**, 495–506.
- Dryden IL, Mardia KV (1998) *Statistical Shape Analysis*. John Wiley and Sons, Chichester.
- Dunn OJ (1964) Multiple comparisons using rank sums. *Technometrics*, **6**, 241–252.
- Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A (2010) Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology*, **8**, 60.
- Elmer KR, Fan S, Kusche H *et al.* (2014) Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature Communications*, **5**, 5168.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.
- Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P (2013) Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 73–97.
- Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*, **20**, 1450–1462.
- Flaherty EA, Ben-David M (2010) Overlap and partitioning of the ecological and isotopic niches. *Oikos*, **119**, 1409–1416.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.
- Ford AGP, Dasmahapatra KK, Rüber L *et al.* (2015) High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*, **24**, 3421–3440.
- Fry B (2006) *Stable Isotope Ecology*. Springer Science + Business Media, LLC, New York.
- Genner MJ, Knight ME, Haesler MP, Turner GF (2010) Establishment and expansion of Lake Malawi rock fish populations after a dramatic Late Pleistocene lake level rise. *Molecular Ecology*, **19**, 170–182.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.
- Henning F, Meyer A (2014) The evolutionary genomics of cichlid fishes: explosive speciation and adaptation in the postgenomic era. *Annual Review of Genomics and Human Genetics*, **15**, 417–441.

- Henning F, Lee HJ, Franchini P, Meyer A (2014) Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD markers for dense linkage mapping. *Molecular Ecology*, **23**, 5224–5240.
- Hillaire-Marcel C, Carro O, Casanova J (1986) <sup>14</sup>C and dating of Pleistocene and Holocene stromatolites from East African paleolakes. *Quaternary Research*, **25**, 312–329.
- Huysseune A (1995) Phenotypic plasticity in the lower pharyngeal jaw dentition of *Astatoreochromis alluaudi* (Teleostei: Cichlidae). *Archives of Oral Biology*, **40**, 1005–1014.
- Kavembe GD, Machado-Schiaffino G, Meyer A (2014) Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. *Hydrobiologia*, **739**, 55–71.
- Keast A, Webb D (1966) Mouth and body form relative to feeding ecology in the fish fauna of a small lake, Lake Opinicon, Ontario. *Journal of the Fisheries Research Board of Canada*, **23**, 1845–1874.
- Kimura M (1953) Stepping-stone model of population. *Annual Report of the National Institute of Genetics*, **3**, 62–63.
- Kirchberger PC, Sefc KM, Sturmbauer C, Koblmüller S (2012) Evolutionary history of Lake Tanganyika's predatory deep-water cichlids. *International Journal of Evolutionary Biology*, Article ID 716209.
- Klingenberg CP (2002) Morphometrics and the role of the phenotype in studies of the evolution of developmental mechanisms. *Gene*, **287**, 3–10.
- Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, **11**, 353–357.
- Klingenberg CP (2013) Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. *Hystrix, the Italian Journal of Mammalogy*, **24**, 15–24.
- Koblmüller S, Egger B, Sturmbauer C, Sefc KM (2010) Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. *Molecular Phylogenetics and Evolution*, **55**, 318–334.
- Koblmüller S, Salzburger W, Obermüller B *et al.* (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. *Molecular Ecology*, **20**, 2272–2290.
- Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews. Genetics*, **5**, 288–298.
- Kruskal WH, Wallis AW (1952) Use of ranks in one-criterion variance analysis. *American Statistical Association*, **47**, 583–621.
- Lachenbruch P (1967) An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, **23**, 639–645.
- Langerhans BR, Layman CA, Shokrollahi MA, Dewitt TJ (2004) Predator-driven phenotypic diversification in *Gambusia affinis*. *Evolution*, **58**, 2305–2318.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.
- Levene H (1960) Robust tests for equality of variances. In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (eds Olkin I, Ghurye SG, Hoeffding W, Madow WG, Mann HB), pp. 278–292. Stanford University Press, Stanford.
- Lopez-Huertas E, Charlton WL, Johnson B, Graham IA, Baker A (2000) Stress induces peroxisome biogenesis genes. *EMBO Journal*, **19**, 6770–6777.
- Losos JB (2010) Adaptive radiation, ecological opportunity and evolutionary determinism. *The American Naturalist*, **175**, 623–639.
- Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature*, **457**, 830–836.
- Mardia K, Kent J, Bibby J (1979) *Multivariate Analysis*. Academic Press, London.
- Martin CH, Cutler JS, Friel JP *et al.* (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, **69**, 1406–1422.
- McElhinny AS, Li JL, Wu L (2008) Mastermind-like transcriptional co-activators: emerging roles in regulating cross talk among multiple signaling pathways. *Oncogene*, **27**, 5138–5147.
- McKaye KR, Marsh A (1983) Food switching by two specialized algae-scraping cichlid fishes in Lake Malawi, Africa. *Oecologia*, **56**, 245–248.
- McMahan CD, Chakrabarty P, Sparks JS, Smith WL, Davis MP (2013) Temporal patterns of diversification across global cichlid biodiversity (Acanthomorpha: Cichlidae). *PLoS One*, **8**, e71162.
- Méndez M, Tella JL, Godoy JA (2011) Restricted gene flow and genetic drift in recently fragmented populations of an endangered steppe bird. *Biological Conservation*, **144**, 2615–2622.
- Meyer A (1990) Morphometrics and allometry in the trophically polymorphic cichlid fish, *Cichlasoma citrinellum* alternative adaptations and ontogenetic changes in shape. *Journal of Zoology London*, **221**, 237–260.
- Meyer A (1993) Trophic polymorphisms in cichlid fish: do they represent intermediate steps during sympatric speciation and explain their rapid adaptive radiation? In: *Trends in Ichthyology* (eds Schroeder JH, Bauer J, Schartl M), pp. 257–266. GSF-Bericht and Blackwell, London.
- Minagawa M, Wada E (1984) Stepwise enrichment of <sup>15</sup>N along food chains: further evidence and the relation between  $\delta^{15}\text{N}$  and animal age. *Geochimica et Cosmochimica Acta*, **48**, 1135–1140.
- Muzuka A (2009) Elemental and isotopic compositions of organic carbon and nitrogen of recently deposited organic matter in Empakai crater and its implication for climatic changes in northern Tanzania. *Tanzania Journal of Science*, **30**, 87–96.
- Netto-Ferreira AL, Albrecht MP, Nessimian JL, Caramaschi ÉP (2007) Feeding habits of *Thoracocharax stellatus* (Characiformes: Gasteropelecidae) in the upper rio Tocantins, Brazil. *Neotropical Ichthyology*, **5**, 69–74.
- Nguyen HM, Miyazaki H, Hoshi N *et al.* (2012) Modulation of voltage-gated K<sup>+</sup> channels by the sodium channel 1 subunit. *Proceedings of the National Academy of Sciences*, **109**, 18577–18582.
- Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford.
- Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705–719.

- Parsons KJ, Cooper WJ, Albertson RC (2009) Limits of principal components analysis for producing a common trait space: implications for inferring selection, contingency, and chance in evolution. *PLoS One*, **4**, e7957.
- Parsons KJ, Cooper JW, Albertson RC (2011) Modularity of the oral jaws is linked to repeated changes in the craniofacial shape of African cichlids. *International Journal of Evolutionary Biology*, Article ID 641501.
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, 2074–2093.
- Philippart J-C, Ruwet J-C (1982) Ecology and distribution of tilapias. In: *The Biology and Culture of Tilapias* (eds Pullin R, Lowe-McConnell R), pp. 15–60. Manila, ICLARM.
- Pinnegar JK, Polunin NVC (1999) Differential fractionation of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  among fish tissues : implications for the study of trophic interactions. *Functional Ecology*, **13**, 225–231.
- Plath M, Pfenninger M, Lerp H *et al.* (2013) Genetic differentiation and selection against migrants in evolutionarily replicated extreme environments. *Evolution*, **67**, 2647–2661.
- Pörtner HO, Schulte PM, Wood CM, Schiemer F (2010) Niche dimensions in fishes: an integrative view. *Physiological and Biochemical Zoology*, **83**, 808–826.
- Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.
- R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>.
- Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *Genes, Genomes, Genetics*, **3**, 65–74.
- Roberts RB, Hu Y, Albertson RC, Kocher TD (2011) Craniofacial divergence and ongoing adaptation via the hedgehog pathway. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 13194–13199.
- Rohlf JF (1999) Shape statistics: procrustes superimpositions and tangent spaces. *Journal of Classification*, **16**, 197–223.
- Rüber L, Verheyen E, Meyer A (1999) Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 10230–10235.
- Rüber L, Meyer A, Sturmbauer C, Verheyen E (2001) Population structure in two sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for introgression. *Molecular Ecology*, **10**, 1207–1225.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.
- Rutschmann S, Matschiner M, Damerau M *et al.* (2011) Parallel ecological diversification in Antarctic notothenioid fishes as evidence for adaptive radiation. *Molecular Ecology*, **20**, 4707–4721.
- Ryan C, McHugh B, Trueman CN *et al.* (2012) Accounting for the effects of lipids in stable isotope ( $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values) analysis of skin and blubber of balaenopterid whales. *Rapid Communications in Mass Spectrometry*, **26**, 2745–2754.
- Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, **11**, 619–625.
- Salzburger W, Mack T, Verheyen E, Meyer A (2005) Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology*, **5**, 17.
- Schelly R, Salzburger W, Koblmüller S, Duftner N, Sturmbauer C (2006) Phylogenetic relationships of the lamprologine cichlid genus *Lepidolamprologus* (Teleostei: Perciformes) based on mitochondrial and nuclear sequences, suggesting introgressive hybridization. *Molecular Phylogenetics and Evolution*, **38**, 426–438.
- Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.
- Seegers L, Tichy H (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya, with descriptions of two new species. *Ichthyological Exploration of Freshwaters*, **10**, 97–146.
- Seegers L, Sonnenberg R, Yamamoto R (1999) Molecular analysis of the Alcolapia flock from lakes Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their systematics and evolution. *Ichthyological Exploration of Freshwaters*, **10**, 175–199.
- Seegers L, Sonnenberg R, Tichy H (2001) The Alcolapia group, a remarkable species flock from Lakes Natron, Tanzania, and Magadi, Kenya: a further piece of puzzle of cichlid evolution? *Journal of Aquariculture and Aquatic Science*, **IX**, 335–364.
- Seehausen O (2015) Process and pattern in cichlid radiations – inferences for understanding unusually high rates of evolutionary diversification. *New Phytologist*, **207**, 304–312.
- Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.
- Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.
- Sidlauskas BL, Mol JH, Vari RP (2011) Dealing with allometry in linear and geometric morphometrics: a taxonomic case study in the *Leporinus cylindriciformis* group (Characiformes: Anostomidae) with description of a new species from Suriname. *Zoological Journal of the Linnean Society*, **162**, 103–130.
- Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology*, **14**, 67–73.
- Stewart TA, Albertson RC (2010) Evolution of a unique predatory feeding apparatus: functional anatomy, development and a genetic locus for jaw laterality in Lake Tanganyika scale-eating cichlids. *BMC Biology*, **8**, 8.
- Stiassny MLJ, Meyer A (1999) Cichlids of the Rift Lakes. *Scientific American*, **280**, 64–69.
- Takahashi T, Koblmüller S (2011) The adaptive radiation of cichlid fish in Lake Tanganyika: a morphological perspective. *International Journal of Evolutionary Biology*, Article ID 620754.
- Takeuchi Y, Hori M (2008) Behavioural laterality in the shrimp-eating cichlid fish *Neolamprologus fasciatus* in Lake Tanganyika. *Animal Behaviour*, **75**, 1359–1366.
- Tichy H, Seegers L (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from lakes Natron and Magadi, Tanzania

- and Kenya : a model for the evolution of “new” species flocks in historical times? *Ichthyological Exploration of Freshwaters*, **10**, 147–174.
- Trapani J (2004) A morphometric analysis of polymorphism in the pharyngeal dentition of *Cichlasoma minckleyi* (Teleostei: Cichlidae). *Archives of Oral Biology*, **49**, 825–835.
- Vander Zanden MJ, Rasmussen JB (1999) Primary consumer  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  and the trophic position of aquatic consumers. *Ecology*, **80**, 1395–1404.
- Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science*, **300**, 325–329.
- Wagner CE, McIntyre PB, Buels KS, Gilbert DM, Michel E (2009) Diet predicts intestine length in Lake Tanganyika’s cichlid fishes. *Functional Ecology*, **23**, 1122–1131.
- Webster M, Sheets HD (2010) A practical introduction to landmark-based geometric morphometrics. *The Palaeontological Society Papers*, **16**, 163–188.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wessels FJ, Hahn DA (2010) Carbon 13 discrimination during lipid biosynthesis varies with dietary concentration of stable isotopes: implications for stable isotope analyses. *Functional Ecology*, **24**, 1017–1022.
- White TH (1953) Some speculations on the sudden occurrence of floods in the history of Lake Magadi. *East Africa Natural History Society*, **22**, 69–71.
- Whiteley AR, Hastings K, Wenburg JK *et al.* (2010) Genetic variation and effective population size in isolated populations of coastal cutthroat trout. *Conservation Genetics*, **11**, 1929–1943.
- Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**, 887–893.
- Wilson PJ, Wood CM, Maina JN, White BN (2000) Genetic structure of Lake Magadi tilapia populations. *Journal of Fish Biology*, **56**, 590–603.
- Wilson PJ, Wood CM, Walsh PJ *et al.* (2004) Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological and Biochemical Zoology*, **77**, 537–555.
- Wong K, Ren X, Huang Y *et al.* (2001) Signal transduction in neuronal migration: roles of GTPase activating proteins and the small GTPase Cdc42 in the Slit-Robo pathway. *Cell*, **107**, 209–221.
- Yamaoka K (1983) Feeding behaviour and dental morphology of algae scraping cichlids (Pisces: Teleostei) in Lake Tanganyika. *African Study Monographs*, **4**, 77–89.
- Yamaoka K (1985) Intestinal coiling pattern in the epilithic algal-feeding cichlids (Pisces, Teleostei) of Lake Tanganyika, and its phylogenetic significance. *Zoological Journal of the Linnean Society*, **84**, 235–261.
- Zaccara S, Crosa G, Vanetti I *et al.* (2014) Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcolapia* flock of East Africa. *Conservation Genetics*, **15**, 429–440.
- G.D.K., G.M.S. and A.M. conceived of the study. All authors were involved in the study design. G.D.K. participated in sampling. A.F.K. analysed the genomic data. G.D.K. analysed the morphological and isotope data. A.F.K. and G.D.K. wrote the manuscript with input from G.M.S. and A.M.

## Data accessibility

Demultiplexed Illumina sequence reads and individual mapped files used for the population genomic analyses have been deposited in the European Nucleotide Archive under the study accession number PRJEB11365. Stable isotope data and MORPHOJ input files, as well as STACKS catalogue and variant files, the VCF file and BAYESCAN results, have been deposited in Dryad doi: 10.5061/dryad.h13jv.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** (a) Positions of the digitized landmarks and semi-landmarks used for body shape analyses. (b) The first two axes of a principal component analysis (PCA) of overall body shape variation.

**Fig. S2** Schematic representation of the three hypothesized scenarios for the split of the Magadi tilapia populations (top row), as well as population size changes (middle rows) and migration rates (bottom row).

**Fig. S3** Average shape differences between pairs of populations in discriminant function analyses (DFAs).

**Fig. S4** Admixture plots based on 7171 SNP markers showing the assignment of individual samples to a predefined number of two (top;  $K = 2$ ) or three (middle;  $K = 3$ ) genetic clusters.

**Fig. S5** Principal component analysis (PCA) based on 7171 SNP markers.

**Fig. S6**  $F_{ST}$  plots showing the genome-wide patterns of differentiation between the Magadi tilapia populations.

**Table S1** Sample information.

**Table S2** Tested Models and their support.

**Table S3** Genetic markers fixed in Little Magadi, their position, and candidate genes in their vicinity.

**Appendix S1** Data accessibility.