The genetic population structure of Lake Tanganyika's *Lates* species flock, an endemic radiation of pelagic top predators

Jessica A. Rick*[1], Julian Junker*[2,3], Ismael A. Kimirei[4], Emmanuel A. Sweke[4,5], Julieth B. Mosille[4], Christian Dinkel[2], Salome Mwaiko[2,3], Ole Seehausen[2,3], Catherine E. Wagner[1]

* denotes equal contribution

[1] Department of Botany and Program in Ecology, University of Wyoming, 1000 E University Dr., Laramie, Wyoming 82072 USA

[2] EAWAG Swiss Federal Institute of Aquatic Science and Technology, CH-6047 Kastanienbaum, Switzerland

[3] Division of Aquatic Ecology & Evolution, Institute of Ecology & Evolution, University of Bern, CH-3012 Bern, Switzerland

[4] Tanzania Fisheries Research Institute (TAFIRI), Dar es Salaam, Tanzania

[5] Deep Sea Fishing Authority (DSFA), Zanzibar, Tanzania

Corresponding authors: Jessica Rick (jrick@uwyo.edu), Julian Junker (Julian.junker@eawag.ch) and Catherine E. Wagner (Catherine.Wagner@uwyo.edu)

Running title: Population genetics of *Lates* fishes

Abstract

Life history traits are important factors in shaping gene flow within species, and these traits can thus be determinants of whether a species exhibits genetic homogeneity across its range, or considerable population structure. Furthermore, understanding genetic connectivity plays crucially into species conservation decisions, and genetic connectivity is an important component of modern fisheries management in fishes exploited for human consumption. In this study, we investigated the four endemic *Lates* species of Lake Tanganyika (*Lates stappersii*, *L. microlepis*, *L. mariae* and *L. angustifrons*), sampled along the Tanzanian shoreline, using reduced-representation genomic sequencing methods. Based on previous studies, we predicted little genetic population structure in the entirely pelagic *L. stappersii* and the predominantly pelagic *L. microlepis*. In contrast, we expected the highest genetic differentiation in populations of *L. mariae*, which is the most resident and for which females aggregate over spawning grounds. We predicted that *L. angustifrons* would show intermediate structure. We indeed find the most strongly differentiated genetic clusters in *L. mariae.* However, contrary to our predictions, we find evidence for genetically distinct (albeit weakly differentiated) groups within the pelagic *L. stappersii* and *L. microlepis*, and no genetic structure in *L. angustifrons*. We call for management approaches accounting for genetically differentiated populations of *L. stappersii*, *L. microlepis*, and *L. mariae*, which are commercially important species in the region.

***Keywords*:** endemic radiation, Lake Tanganyika, population structure, gene flow

Introduction

Intraspecific gene flow is shaped by an organism's life history traits (Ellegren & Galtier 2016). Life history traits, including reproduction, growth and dispersal, strongly influence the genetic structure of populations (Manier & Arnold 2006; Stearns 1992). Differences in life history traits also explain why ecologically similar taxa inhabiting the same environment may differ in their genetic population structure (Peterman *et al.* 2015; Young *et al.* 2015). Conversely, landscape features may lead to similar population structures even in species with different life history traits (e.g. Petren *et al.* 2005).

Aquatic ecosystems are often inhabited by fish species with major differences in life history traits sharing the same environment. In teleost fishes, life history traits—such as spawning location, philopatry, timing of spawning and duration of larval stage—have been shown to shape population structure (Olsen *et al.* 2011; Pettersson *et al.* 2019; Young *et al.* 2015). Anadromous salmonids are particularly notorious for displaying homing behaviour that leads to genetic distinction among spawning grounds, dividing populations into many genetic units despite the populations spending most of their lives together in the ocean (Wenburg *et al.* 1998). Divergence between populations caused by differences in the timing of spawning also leads to genetic structure in salmonids (Brannon *et al.* 2004) and has also been observed in Atlantic herrings (Pettersson *et al.* 2019). In the oceans, coral reef fishes with reproduction tied to the substrate, exhibit parental care strategies, and lack planktonic larvae, generally have significantly greater population structure than pelagic spawners or substrate spawners with planktonic larvae (Riginos *et al.* 2014). Additionally, a shorter larval phase among pelagic spawners can lead to a more structured population, because this leads to a shorter larval dispersal distance (Selkoe & Toonen 2011). Understanding how life history traits like dispersal, reproduction, and growth, shape gene flow in fish species is crucial for the goal of improving our understanding of local adaptation, speciation, persistence of populations, migration, and the spatial scale of effective management (Sunday *et al.* 2014).

The use of next generation sequencing methods, which can resolve even fine-scale genetic structure through sampling a large proportion of the genome, is a useful tool to describe population structure, particularly in species with low genetic differentiation. The advent of these methods, and their application to understanding genetic structure, have made important contributions to understanding management units in exploited populations. Examples include the classification of spring and fall spawners in Atlantic herring (Martinez Barrio *et al.* 2016; Pettersson *et al.* 2019) and the differentiation between migrating and resident stocks of Atlantic cod (Kirubakaran *et al.* 2016).

The pelagic fish community and the fishery of Lake Tanganyika

The pelagic fish community of Lake Tanganyika is relatively low in species diversity compared to that of other large African lakes: it consists predominantly of six endemic fish species belonging to two families (Coulter 1991). These include a monophyletic pair of mainly planktivorous clupeids (*Stolothrissa tanganicae*, *Limnothrissa miodon*; Junker *et al.* 2020; Wilson *et al.* 2008) and four mainly piscivorous latids of the genus *Lates* (*L. stappersii*, *L. mariae*, *L. microlepis*, *L. angustifrons*). Cichlid fish, so diverse and abundant in the pelagic communities of other African Great Lakes, are rare in the pelagic zone of Lake Tanganyika. Of the four *Lates* species, only *L. stappersii* is sufficiently abundant to be critical for the current fishery of Lake Tanganyika, the second largest inland fishery on the continent of Africa (Coulter 1976, 1991; Kimirei *et al.* 2008; Mannini 1998a, b; Mölsä *et al.* 1999; Mölsä *et al.* 2002; Munyandero 2002; Sarvala *et al.* 2002; van der Knaap 2013; Van der Knaap *et al.* 2014; van Zwieten *et al.* 2002). Populations of *L. mariae*, *L. microlepis* and *L. angustifrons* suffered major population declines during the 1960s when an industrial purse seine fishery supplemented the traditional fishery (Coulter 1976, 1991; Mölsä *et al.* 2002). The industrial fishery was outcompeted by the artisanal fishery as fishermen transitioned to using powerful outboard engines on their traditional wooden fishing boats, allowing boats to fish further offshore (Kimirei *et al.* 2008). The last industrial fishing vessel stopped operating in 2011 (LTA Secretariat 2011; Paffen *et al.* 1997; Van der Knaap *et al.* 2014), but the populations of *L. microlepis*, *L. mariae* and *L. angustifrons* have remained low ever since. The very large size and slow time to maturation of these species makes them particularly vulnerable to fishing pressures and all three are listed by the IUCN as vulnerable (*L. mariae*) or threatened (*L. microlepis* and *L. angustifrons*) (Ntakimazi 2006a, b, c).

During the last decades, a general downward trend in fish catches was observed for Lake Tanganyika's fishery (Kimirei *et al.* 2008; van der Knaap 2013; Van der Knaap *et al.* 2014; van Zwieten *et al.* 2002). This trend results from a combination of an increase in the number of fishermen and vessels on Lake Tanganyika, changes in fishing practices (e.g. the use of beach seining, which targets fish in their in-shore nursery habitats) (Kimirei *et al.* 2008; van der Knaap 2013; Van der Knaap *et al.* 2014; van Zwieten *et al.* 2002), and warming lake surface temperatures (Cohen *et al.* 2016; O'Reilly *et al.* 2003). Although the practice of beach seining is illegal according to the fishing regulations of the Lake Tanganyika Authority (LTA Secretariat 2011), it still occurs due to difficulties associated with enforcing the regulation across the four riparian countries. Decreases in fish abundance are likely also linked to reduced productivity in the lake caused by stronger water column stratification due to climate change (O'Reilly *et al.* 2003; Verburg *et al.* 2003). Consequently, there is increasing recognition of the need to develop sustainable management strategies for the lake's pelagic fish stocks (Kimirei *et al.* 2008; Mölsä *et al.* 1999; Mölsä *et al.* 2002; van der

Knaap 2013; Van der Knaap *et al.* 2014; van Zwieten *et al.* 2002). Sustainable management of the fishery and safeguarding the survival of the more vulnerable species (i.e. the three large *Lates* species, *L. mariae*, *L. angustifrons* and *L. microlepis*) requires complete information on the factors influencing the health of these stocks. This includes an understanding of the stock complexity and the impacts of differences in life history and environment on population structure of these species. This study focuses on adding this additional important information to our understanding of the four *Lates* species in Lake Tanganyika.

The fish populations of Lake Tanganyika are influenced by a complex interplay of bottom-up control via changes in the supply of nutrient-rich deep waters, driven by different mixing regimes and upwelling rates (Bergamino *et al.* 2010; O'Reilly *et al.* 2003; Stenuite *et al.* 2007; Verburg & Hecky 2009; Verburg *et al.* 2003), and a top-down control via predators and fishing (Kimirei *et al.* 2008; Mannini 1998a; Mannini *et al.* 1996; Munyandero 2002; van der Knaap 2013; Van der Knaap *et al.* 2014; van Zwieten *et al.* 2002). If environmental differences, such as mixing regimes and nutrient supply, are consistent over generations and coupled with limited gene flow between spatially segregated populations, divergent selection on life history traits (e.g. spawning phenology, developmental timing and recruitment success) of species in different parts of the lake could develop and persist. The spatial environmental variation of Lake Tanganyika, in combination with the large distances between populations, might therefore generate intraspecific genetic differentiation among populations of pelagic fish, as has been shown for pelagic cichlids in Lake Tanganyika and Lake Malawi (Genner *et al.* 2010; Koblmüller *et al.* 2019). Here, we test these predictions using genomic data collected from each of the four *Lates* species from across their distribution within Lake Tanganyika.

The biology of the four Lates species

Relatively little is known about the location of spawning grounds of all four *Lates* species. Coulter (1976) mentions that ripe-running females of *L. angustifrons* and *L. mariae* are occasionally caught in benthic gill nets and the *L. mariae* females were aggregated, suggesting that spawning females concentrate over benthic spawning grounds. Eggs and freshly hatched larvae of all species are likely planktonic given other data from this genus (van Zwieten *et al.* 2016). Juveniles of all four species up to a size of 2.5cm are caught in plankton nets (Coulter 1976). The smallest species, *L. stappersii*, is the only species of the four endemic *Lates* that spends its entire life in the pelagic, whereas juveniles (>3cm) and adolescents of the three larger species are found in shared inshore nursery habitats, predominantly in the macrophyte *Ceratophyllum* (Coulter 1976; Kondo and Abe 1995).

Pelagic nursery habitats for *L. stappersii* have been recorded near Kigoma in the northern basin (Mannini et al. 1996), and in the central basin (Chapman and van Well 1978). Juvenile *L. stappersii* feed on zooplankton and switch to a more piscivorous diet when they reach a size of more than 130mm (Ellis 1978). The juveniles of *L. microlepis*, *L. mariae,* and *L. angustifrons* in the weed beds feed on zooplankton, prawns, cichlid fishes, insects, and insect larvae, and no difference between the three species' diets have been discerned (Coulter 1976). Once these species leave the inshore environment as they mature, they begin to differentiate in habitat and diet affinities. From the weed patches, *L. mariae* move to deeper water with increasing size, adopting a benthic lifestyle, with benthic cichlids as their dominant prey (Coulter 1991). *L. microlepis* is the top predator in pelagic surface waters, feeding mainly on clupeids and on *L. stappersii* up to lengths of 40% of the predator's body length (Coulter 1976, 1991). Large *L. angustifrons* can be found throughout the water column. Based on diet composition, adult *L. angustifrons* predominantly hunt in inshore rocky habitats, feeding mainly on littoral cichlids (Coulter 1976, 1991). If clupeids are abundant, mature individuals of all three large species, *L. mariae*, *L. microlepis* and *L. angustifrons*, are found preying upon them during the night (Coulter 1976). Spawning individuals of all three of the large *Lates* species (*L. microlepis*, *L. mariae* and *L. angustifrons*) exist throughout the year but spawning peaks were reported for the southern end of the lake from August to December (Coulter 1976).

Evidence for intraspecific differences in ecology and life history traits are also found in *L. stappersii* and *L. mariae*. Mannini *et al.* (1999) found that *L. stappersii* in the northern basin have a diverse diet, dominated by clupeids, while *L. stappersii* in the southern basin mainly fed on atyid and palaemonid shrimps. Furthermore, the recruitment periods differ between regions of the lake and fish mature at larger sizes in the north than the south (Mannini et al. 1996). Population sizes of *L. stappersii* are also heterogeneous along the length of Lake Tanganyika, with fish occurring in greater abundance in the north of the lake than the south. In contrast, catch rates of *L. mariae* and *L. microlepis* increase from north to south, with the most abundant populations being found in the southern portion of the lake.

The need for a population genetic study

For conservation and fisheries management, which seeks to avoid depletion of fish stocks and to ensure sustainable harvest (Reiss *et al.* 2009), understanding the genetic population structure of species is essential. Disregarding population structure and managing a population complex as one population can lead to overfishing or even extinction of the more vulnerable populations (Reiss *et al.* 2009; Sterner 2007), and to loss of genetic diversity in the species as a whole. Reduced genetic variation that comes with decreased abundance

decreases the ability of a species to adapt and evolve in response to changes in its environment (Mace & Purvis 2008). Defining Management Units (MUs)—demographically independent and genetically distinct populations—in pelagic mixed stocks is therefore key to understanding populations' resilience to fishing pressure (Belgrano & Fowler 2011; Berg *et al.* 2017; Botsford *et al.* 1997; Hutchinson 2008; Kirubakaran *et al.* 2016).

The pelagic zone of the two large lakes in the western arm of the East African rift valley comprises such mixed stocks. In Lake Malawi the pelagic is inhabited by 21 of the 800 different cichlid species present in this lake, nearly all of them endemic (Malinsky *et al.* 2018; Snoeks & Carvalho 2004). Among them is the benthopelagic taxon *Diplotaxodon*, which comprises eight species (Snoeks & Carvalho 2004). Genner *et al.* (2010) sampled individuals on spawning grounds and found genetic structure among spawning grounds within four of the species. They suggest that strong spawning site fidelity explains this genetic structure. Besides the most abundant species (clupeids and *Lates*) in Lake Tanganyika's pelagic, there is the endemic cichlid tribe Bathybatini, including nine species (seven *Bathybates* and two *Hemibates*), which all inhabit the eu- or benthopelagic habitat (Coulter 1991; Koblmüller *et al.* 2019). All nine species are mouthbrooders, all species exceed 30cm (except for *B. minor*), and all are predatory. In a study where four of these species were tested for intraspecific genetic structure, Koblmüller *et al.* (2019) found no genetic structure in the two species inhabiting the eupelagic, *Bathybates fasciatus* and *B. leo*, but clear differences in the two benthopelagic species, *B. graueri* and *Hemibates stenosoma.* These two species showed genetic distinction between populations in the north and south of the lake. Koblmüller *et al.* (2019) hypothesize that the specialization on different types of prey might explain these differences in population structure. While *B. fasciatus* and *B. leo* mainly hunt pelagic clupeids (and therefore perform long distance movements in the open water), *B. graueri* and *H. stenosoma* hunt benthic and benthopelagic cichlids and therefore do not need to move long distances through open water to find their prey. Koblmüller *et al.* (2019) hypothesize that such foraging behavior might translate into different rates of dispersal.

Compared to the *Lates* of Lake Tanganyika, these cichlids have dramatically different life histories such as mouth brooding, much larger larvae that are fully developed miniature versions of their parents when released from their mother's mouth, and no pelagic juvenile phase. However, similar to the cichlid genus *Bathybates*, the *Lates* species flock is comprised of two species with closer ties to the littoral (*L. angustifrons*) and the bentho-pelagic (*L. mariae*) and two species more associated with the eupelagic (*L. stappersii* and *L. microlepis*). Given these differences in life history among the four *Lates* species, it is possible that the genetic consequences of their shared spatial environment differ between species. For example, due to *L. stappersii* having an entirely pelagic lifestyle, including spawning, we hypothesize that its population is more panmictic than those of the other

species, which have aggregated spawning grounds (*L. mariae*) and inshore nursery habitats (*L. angustifrons, L. microlepis, L. mariae*).

Analyzing and understanding the possibly differential responses of genetically differentiated stocks to exploitation and environmental variation is of great importance for developing successful conservation and fisheries management strategies (Hutchinson 2008; Therkildsen *et al.* 2013). Stock diversity can contribute to the resilience of a fishery to environmental fluctuations (Schindler *et al.* 2015) and the loss of differentiated populations can compound to major losses in the consistency of fisheries yields (Hutchinson 2008; Schindler *et al.* 2010). It is therefore crucial to understand the extent and nature of population genetic structure and differential adaptation in fishes that are important food and economic resources. The loss of population diversity may have drastic effects on ecosystem services and species persistence, and understanding the extent to which population structure exists is crucial for predicting these consequences (Schindler *et al.* 2015; Schindler *et al.* 2010).

Here, we used reduced-representation genomic sequencing datasets from four endemic *Lates* species, *L. stappersii, L. mariae, L. microlepis,* and *L. angustifrons*, sampled throughout the Tanazanian waters of Lake Tanganyika to examine their genetic connectivity. Since this is the first population genetic work including all four species (see Koblmüller *et al.* 2021 for a phylogenetic assessment of African Lates perches), we first assessed whether phenotypic identification coincides with genetically based species boundaries. We then examined the extent to which each of the four species is genetically structured, as well as how the genetic structure observed in each species corresponds to its known life history and environment. Based on the literature, we hypothesized to find little genetic population structure in the entirely pelagic *L. stappersii* and mainly pelagic *L. microlepis* (Coulter 1976). In contrast, we expected the highest genetic differentiation in populations of *L. mariae*. Finally, we hypothesized that *L. angustifrons* would show intermediate amounts of structure since there is no evidence for large scale movements of this species (Coulter 1976). Despite absence of isolation-by-distance in any of the four species, we find evidence for strongly differentiated geographically linked genetic clusters in *L. mariae,* as well as less strongly differentiated genetic groups within *L. stappersii* and *L. microlepis,* while we find no distinct genetic clusters in *L. angustifrons*.

Material and Methods

Study system and sampling

Between 2001 and 2019, we collected tissue samples (fin clips) from *L. stappersii, L. mariae, L. microlepis,* and *L. angustifrons* at 9 general locations along the Tanzanian shore of Lake Tanganyika, which together span the ~490 km shoreline from the northern to southern border

(Fig 1). Most fish were obtained opportunistically from fishermen, and therefore were already dead when we received them. All fish collections were made in partnership with researchers of the Tanzanian Fisheries Research Institute (TAFIRI) in Kigoma, Tanzania. Fish were identified in the field by sampling personnel following United Nations Food and Agricultural Organization (FAO) identification guidelines (Eccles 1992) and guidance from Tanzanian researchers and local fishermen, and genetic methods were later used to confirm species identities for each fish. For fish collected from 2016-2019, we recorded length and weight, and took standardized pictures of each fish. For the few live fish that were caught as part of other research, we took cuvette photographs of the live fish and subsequently euthanized the fish with an overdose of MS222. We then took fin clips for genetic analysis and muscle tissue samples for stable isotope analysis from all fish. Specimens smaller than 600mm were preserved in formaldehyde and archived in the collections at TAFIRI (Kigoma, Tanzania), EAWAG (Kastanienbaum, Switzerland), the University of Wyoming Museum of Vertebrates (Laramie, WY, USA), or the Cornell University Museum of Vertebrates (Ithaca, NY, USA).

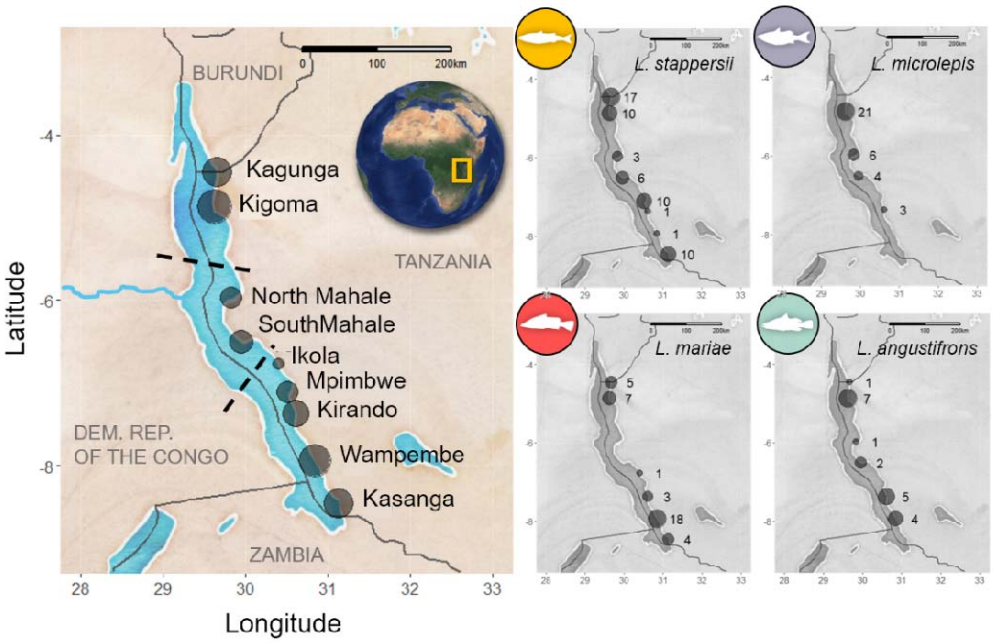

Figure 1. Sampling site locations (left) and number of individuals retained in the final filtered datasets at each site for each of the four Lake Tanganyika-endemic *Lates* species (right). In all maps, circle sizes are proportional to the number of samples in the final dataset.

Genomic sequencing

We extracted DNA from fin clips using DNeasy Blood and Tissue Kits (Qiagen, Inc.) following the standard protocol, with the addition of an RNAse A preparation step. We then prepared

genomic libraries for a combination of genotyping-by-sequencing (GBS, for *L. stappersii, L. mariae, L. microlepis, L. angustifrons*) and restriction site associated DNA sequencing (RAD, for *L. stappersii*).

We prepared GBS libraries following protocols outlined in Parchman *et al.* (2012), using MseI and EcoRI restriction enzymes. Following fragmentation via restriction enzyme digestion, fragmented DNA was barcoded by ligating short, unique individual-identifying DNA fragments (barcodes) to each individual's fragmented DNA. These barcoded fragments were then amplified by PCR and pooled for sequencing. We ran two replicate PCRs for each individual and pooled the final PCR products into two libraries. The prepared libraries were size-selected for 200-350bp fragments using Blue Pippin (Sage Science, MA). One library was sequenced on one lane of Illumina HiSeq4000 (150bp single-end) at the University of Texas at Austin's Genome Sequencing and Analysis Facility (UT GSAF; Austin, TX) and the other library was sequenced on one lane of Illumina HiSeq4000 (150bp, single-end) at the University of Oregon's Genomics and Cell Characterization Core Facility (Eugene, OR). Each library contained 192 individuals, and select individuals were duplicated across libraries to check for library compatibility.

The RAD libraries containing the *L. stappersii* were prepared for sequencing following the protocol by Baird *et al.* (2008) with the following modifications: we used between 400ng and 1000ng genomic DNA per sample and digested with *SbfI* overnight. We tagged each individual using P1 adapters (synthesized by Microsynth) with custom six to eight base pair barcodes and multiplexed 67 barcoded individuals per library. The libraries were sheared using an S220 series Adaptive Focused Acoustic (AFA) ultra-sonicator (Covaris) with the manufacturer's settings for a 400 bp mean fragment size. We size-selected for fragments between 300 and 700bp using a sageElf (Sage Scientific Electrophoretic Lateral Fractionator; Sage Science, Beverly, MA). The enrichment step was done in 6 aliquots with a total volume of 200 µl. Volumes were combined prior to the final size selection (again 300-700bp) step using the sageELF. Sequencing was done by the Lausanne Genomic Technologies sequencing facilities (University of Lausanne, Switzerland). All three libraries were single-end sequenced on one lane each of the Illumina HiSeq2000 (100bp SE), along with spiked-in bacteriophage PhiX genomic DNA (7-8%).

Sequence data preparation

We filtered raw sequencing reads from each library by first removing the reads of the PhiX genome and other common contaminants (e.g. excess barcodes, primers, and adapters)

using bowtie2 (v2.0.0, Langmead & Salzberg 2012). We filtered reads for an intact enzyme restriction site (*SbfI* for RAD libraries, MseI and EcoRI for GBS libraries), de-multiplexed the fastq files, and matched sequence barcodes to individual fish using custom perl and bash scripts.

For the GBS libraries, all data were assembled to the chromosome-level *Lates calcarifer* reference genome (v3, NCBI Genome Assembly GCA_001640805.1; Vij *et al.* 2016) using bwa mem (v0.7.17; Li & Durbin 2009) with default settings. Following alignment, we excluded any individual with < 50,000 reads or less than 60% of raw reads assembled to the reference genome. We identified variable sites (i.e. single nucleotide polymorphisms; SNPs) in the assembly using SAMtools mpileup (v1.8; Li *et al.* 2009) and bcftools (v1.8; Li *et al.* 2009). In calling variable sites, we omitted indels and kept only high-quality biallelic variant sites (QUAL < 20 and GQ > 9). We then filtered SNPs by minor allele frequency and amount of missing data using vcftools (Danecek *et al.* 2011), additionally only calling genotypes with a minimum read depth of 2. We created two separate SNP data sets, using a minor allele frequency threshold of 0.01 for both and keeping only sites that were present in 50% (for the first dataset, used for population genetic analyses) or 90% of individuals (for the second dataset, used in phylogenetic analyses). Finally, we thinned each data set to retain one SNP per locus (-thin 90).

For the RAD libraries, we barcode-trimmed reads down to 84 nucleotides using process_radtags from Stacks (Catchen *et al.* 2013). The FASTX-toolkit v.0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) was used for quality filtering. In a first step, we kept only reads with all base quality scores greater than 10; in a second step, we removed all reads with more than 5% of the bases with quality score below 30. The RAD library was then assembled to the *Lates calcarifer* reference genome (Vij et al. 2016) and filtered, following the same steps as for the GBS libraries above. Individuals included in the RAD libraries were distinct from those included in the GBS libraries.

After filtering variants, we calculated the distribution of minor allele reads at heterozygous sites for each individual in both RAD and GBS libraries, using a custom bash script (available at http://github.com/jessicarick/lates-popgen). If there is an excess of sequencing errors or contamination for an individual, then we would expect the ratio of minor allele to major allele reads to be significantly less than 1:1. Thus, we plotted the distribution of minor allele reads for each individual and removed individuals with a deficiency in minor allele reads across the majority of heterozygous sites from further analyses.

Species assignment

Not all of our samples could be identified to species in the field. Some fish were caught deep and brought quickly to the surface by the fishermen. These fish were often difficult to identify, since their morphology was impaired due to the rapid change in pressure. Furthermore, juveniles of the large species *L. microlepis*, *L. mariae* and *L. angustifrons* can be difficult to distinguish. Because this is the first genetic study of the four Lake Tanganyika *Lates* species, we used three different lines of evidence to confirm species identification and ensure that morphological identification matched genetic assignment: (1) phenotypic identification based on FAO species descriptions; (2) individual ancestry assignment using model-based assignment; and (3) inference of monophyletic groups in phylogenetic analysis. The details for each of these steps are described below.

With the GBS data, we moved forward with the SNP dataset filtered for < 50% missing data and minor allele frequency above 0.01. We first visualized clusters in our data using principal component analysis (prcomp in R) on genotype covariance matrices calculated from the SNP data. We then inferred species groups using the model-based genetic clustering program entropy (Gompert *et al.* 2014). Entropy estimates individual ancestry from multilocus SNP data, using Bayesian estimation from genotype likelihoods. In taking uncertainty about individual genotypes into account via genotype likelihoods, the model integrates outcomes over genotype uncertainty. We converted our VCF containing sites with <50% missing data to the mpgl genotype likelihood format using the custom vcf2mpgl.pl script. We then ran entropy for K = 4 clusters, running three independent MCMC chains of 80,000 total steps, discarding the first 10,000 steps as burn-in, and retaining every 10th value (thin=10), resulting in 7000 samples from the posterior distribution of each chain. We checked MCMC chains for mixing and convergence of parameter estimates by plotting a trace of the MCMC steps. We then visualized assignments in R and assigned individuals to species based on group assignment probabilities (q), using a threshold of 0.6 for assignment.

We further confirmed species identities using maximum likelihood phylogenetic inference methods in RAxML (v8.1.17; Stamatakis 2014). We concatenated all SNPs within our <10% missing data dataset, removed invariant sites using the raxml_ascbias python script (v1.0, from https://github.com/btmartin721/raxml_ascbias), and used the Lewis correction for invariant sites with the ASC_GTRGAMMA model of molecular evolution (Lewis 2001) within RAxML to construct the maximum likelihood phylogeny, including *L. calcarifer* as the outgroup. We used 100 rapid bootstraps to estimate confidence in our maximum likelihood tree. We visually identified monophyletic groups on our maximum likelihood tree to delineate species and used these groupings to further verify species identities from entropy.

Genetic diversity

We calculated genetic diversity for each species using the aligned BAM files in ANGSD (v0.931, Korneliussen *et al.* 2014). From BAM alignment files, we first calculated the site allele frequency likelihoods based on individual genotype likelihoods (option -doSaf 1) using the samtools model (option -GL 1), with major and minor alleles inferred from genotype likelihoods (option -doMajorMinor 1) and allele frequencies estimated according to the major allele (option -doMaf 2). We filtered sites for a minimum read depth of 2 and a maximum depth of 100, minimum mapping quality of 20 and minimum quality (*q*-score) of 20. From the site allele frequency spectrum, we then calculated the maximum-likelihood estimate of the folded site frequency spectrum (SFS) using the ANGSD realSFS program (with option -fold 1). The folded SFS was used to calculate genome-wide genetic diversity (Watterson's Θ), using the ANGSD thetaStat program (Korneliussen *et al.* 2013).

Population genetic structure and isolation-by-distance

For each species, we then re-called variants to create a species-specific set of SNPs from the GBS data. As with the species combined GBS data set, we identified variable sites in each species' assembly using SAMtools mpileup and bcftools, omitting indels and keeping only high-quality biallelic variant sites (QUAL < 20 and GQ > 9). We then filtered SNPs by minor allele frequency (threshold 0.01) and amount of missing data (50% threshold) using vcftools (Danecek et al. 2011), additionally only calling genotypes with a minimum read depth of 2. Finally, we thinned to one SNP per locus (-thin 90).

For each species, we then performed principal component analysis (PCA) using EMU (Meisner *et al.* 2021) to visualize the genetic affinities of individuals within each species. EMU uses an iterative method to infer population structure in the presence of missing data and is therefore able to infer population structure even for datasets of ultra-low coverage sequencing data with very high missingness rates, in addition to being able to handle non-random missingness patterns, which tend to produce bias in other PCA methods (Meisner *et al.* 2021). We also performed PCA in EMU on the RAD dataset for *L. stappersii,* as a comparison for our GBS dataset.

We then used the Reich-Patterson F_{ST} estimator (Reich *et al.* 2009), which is consistent even with small sample sizes, to calculate genetic divergence between sampling sites within each species, using a custom function in R (available at http://github.com/jessicarick/reich-fst). We

then tested for isolation-by-distance (IBD) using a Mantel test (mantel.randtest from ade4 in R; v1.7-16, Dray & Dufour 2007) to calculate the correlation between the natural logarithm of Euclidean geographic distance and genetic distance ($1/1-F_{ST}$) for all pairwise combinations of sampling sites within each species. We chose to pool these tests by sampling site, as we do not have more specific locations for where most of the fish were caught.

To test for intraspecific population subdivision more formally, we again used entropy. We conducted clustering analyses using the dataset with < 50% missing data per site and minor allele frequency > 0.01. We ran entropy for K=1 to K=8 for each species to infer the most likely number of ancestral groups within each species, and to assign individuals to these clusters. We ran three independent MCMC chains of 80,000 total steps, discarding the first 10,000 steps as burn-in, and retaining every 10th value (thin=10), resulting in 7000 samples from the posterior distribution of each chain. We checked MCMC chains for mixing and convergence of parameter estimates by plotting a trace of the MCMC steps. We then calculated the deviance information criterion (DIC) for each value of K and used the model with the lowest DIC as our best for modeling the variation observed in our data. From our entropy results, we assigned individuals to intraspecific groups using a cutoff of q > 0.6 at each value of K. We then calculated the mean of Reich-Patterson F_{ST} estimates between these K groups in R to understand how genetically distinct these groups are from one another.

Results

The GBS libraries yielded an average of 245 million reads across 200 individuals. The RAD libraries containing the *L. stappersii* individuals yielded an average of 271 million reads across 58 individuals, including 2.5% - 8.6% bacteriophage PhiX genomic DNA. On average, the mapping rate for *L. stappersii* individuals' GBS reads to the *L. calcalifer* reference genome was 89.8% (RAD, 89.6%), whereas it was 91.6% for *L. microlepis*, 91.3% for *L. mariae* and 88.7% for *L. angustifrons* individuals.

These mapping rates resulted in 1,463,702 unfiltered variable sites in *L. stappersii*, 1,286,810 in *L. microlepis*, 961,153 in *L. mariae* and 764,456 in *L. angustifrons*. After filtering for < 50% missing data, minor allele frequency (MAF > 0.01), and keeping only one SNP per locus (-thin 90), our species-specific GBS data sets contained 54,193 SNPs from 58 *L. stappersii* individuals (RAD, 34,044 SNPs from 58 different individuals), 40,887 SNPs from 34 *L. microlepis* samples, 43,143 SNPs from 38 *L. mariae* samples and 25,322 SNPs from

20 *L. angustifrons* samples. The data set with all four species contained 4,999,192 unfiltered and 61,094 filtered SNPs.

Species assignment

The species-combined SNP dataset for GBS data contained 200 individuals, of which we removed 50 individuals due to bad read quality or for having highly unbalanced reads at heterozygous sites, potentially indicative of PCR duplicates and contamination. The remaining individuals had > 50,000 reads that aligned to the *L. calcarifer* reference genome and > 80% of genotypes called. The SNP alignment for RAxML contained 38,737 sites after filtering for < 10% missing data, a minimum read depth of 5, and removing invariant sites. The rooted maximum likelihood phylogeny (rooted on *L. calcarifer*; Fig. 2A) shows a sister taxon relationship between *L. mariae* and *L. angustifrons,* and a second sister taxon relationship between *L. stappersii* and *L. microlepis*; however, short internode distances between these splits suggests that all three speciation events may have occurred within a short period, and more investigation of the evolutionary history of the clade is warranted. The PCA based on the GBS data was consistent with the species affinities from monophyletic groups in RAxML (Fig. 2B). The first genetic PC axis (79.9%) separates *L. stappersii* from the remaining three species. The second PC axis (14.8%) then separates *L. mariae* and *L. microlepis* from one another, and the third axis (4.8%) separates *L. angustifrons* from the other three species. Using entropy at K=4, we found each of the four species to make a distinct genetic cluster (Fig. 2C). Based on these combined analyses, 58 individuals were assigned to *L. stappersii*, 34 to *L. microlepis*, 38 to *L. mariae*, and 20 to *L. angustifrons.* While the genetic approaches revealed consistent species identifications, we had phenotypically misidentified one *L. microlepis*, one *L. mariae* and three *L. angustifrons* (Table S1).

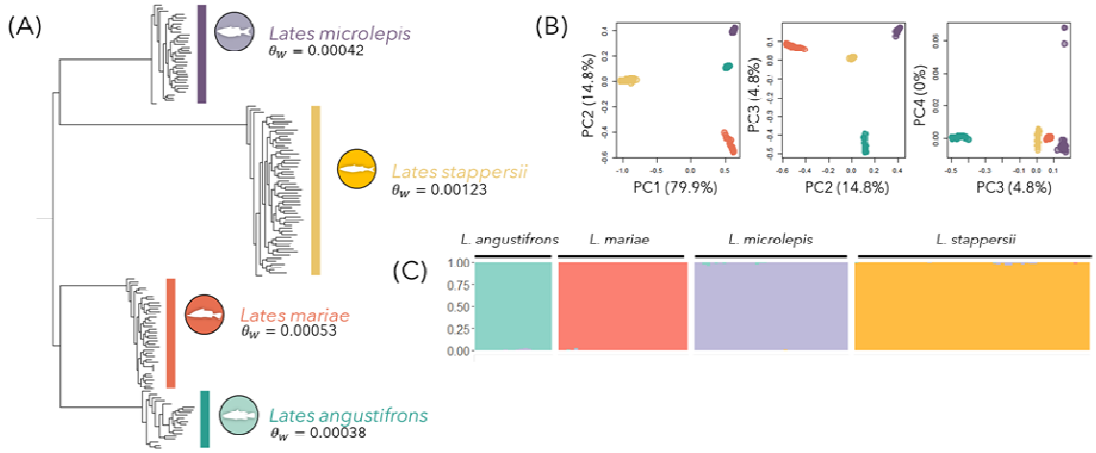

Figure 2. Relationships among the four endemic *Lates* species in Lake Tanganyika based on the GBS data set, as demonstrated by their (a) phylogenetic relationships, (b) relationships using principal component analysis, (c) genetic clustering using entropy at K=4. The phylogeny has been rooted using *L. calcarifer* as an outgroup. In (a), genetic diversity of the given species, as measured using Watterson's Θ in ANGSD, is indicated. Percentages in (b) indicate that amount of variance explained by the given principal component axis (axes 1-4 shown). In (c) each vertical bar represents an individual, and the colors represent the proportion of ancestry assigned to each of the four species.

Population structure

All four species had low but similar levels of genetic diversity (mean Θ_W = 0.00043), and genetic diversity was slightly elevated in *L. stappersii* compared to the three other species (Fig 2A). Reflecting the distinctiveness seen in the PCA, F_{ST} values are high between species (mean Reich-Patterson F_{ST} = 0.607; Table S2). None of the four species shows a significant relationship between genetic and geographic distance, suggesting an overall lack of isolation by distance (Fig S1 all Mantel test p-values > 0.2).

The PCA for *L. stappersii* based on GBS data (Fig 3A) did not reveal any clear genetic clusters, but most individuals from the sampling sites Kigoma, North Mahale and South Mahale do appear genetically distinct. We did not see any clustering in the RAD data for *L. stappersii* (Fig S2); however sampling from Kigoma was minimal (n=2) and North and South Mahale were not sampled at all in this dataset. The lack of structure among the locations that were sampled was, hence, consistent with the GBS data. In entropy, K = 1 was the best fit according to DIC (Fig 4A); however, at K = 2, individuals from Kigoma, North Mahale and South Mahale form a distinct genetic cluster (except for one fish from South Mahale). This weak but distinct genetic structure (between-cluster Reich-Patterson F_{ST} = 0.00168, 95% bootstrap CI: 0.00118-0.00216, Fig 4A and Table S3) was also partly reflected in F_{ST} values

between sampling sites, where F_{ST} values of comparisons between individuals of South Mahale or North Mahale with samples from other sites were elevated, albeit small (<0.007, Fig S3A and S4A).

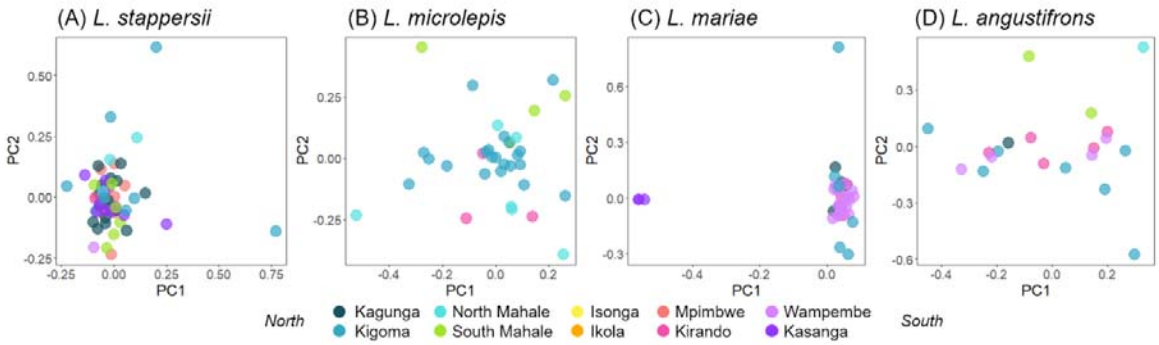

Figure 3. Principal component analysis of individual species based on the GBS data set, with colors indicating sampling locations. For each species, the plot shows the location of each individual on the two most dominant PCA axes (with percentages indicating the proportion of variance explained by the given axis).

A similar pattern to that observed in *L. stappersii* was apparent in *L. microlepis*. The PCA based on GBS data did not reveal clear genetic clustering but it did highlight several individuals from the sampling sites North Mahale and South Mahale that were differentiated from samples of Kigoma and Kirando (Fig 3B). In entropy, K = 1 was again found to be the best fit using DIC (Fig 4B) but at K = 2 all six samples from North Mahale and one individual out of 4 from South Mahale form a distinct genetic cluster (Reich-Patterson F_{ST} between entropy-identified groups = 0.00257, 95% CI 0.00151-0.00393, Fig 4B and Table S3). The F_{ST} estimates between samples from different sites are again very small (Fig S4B), but pairwise comparisons involving North Mahale samples differ significantly from zero (Fig S3B).

The PCA based on GBS data for *L. mariae* separates three individuals from Kasanga on the first PC axis, while the second PC axis separates out some individuals from Kigoma (Fig 3C). In entropy, K = 2 appears to be the best fit using DIC, where the three individuals from Kasanga form a distinct cluster (Reich-Patterson FST between entropy-identified groups = 0.132, 95% bootstrap CI: 0.128-0.138; Fig 4C and Table S3). At K = 3, five of our seven samples from Kigoma form the third, less distinct genetic cluster (mean Reich-Patterson F_{ST} between entropy-identified groups = 0.0892, Fig 4C). The F_{ST} estimates between Kasanga (fish from both clusters pooled) and other sites are all significantly different from zero (Fig S3C). Additionally, the samples from Kigoma are significantly different from Wampembe although this F_{ST} is an order of magnitude smaller (Fig S3C).

For *L. angustifrons*, the PCA based on GBS data did not reveal any genetic structure (Fig 3D). In entropy, K = 1 appears to be the best fit using DIC and the entropy plot does not show any clear clustering for higher values of K (Fig 4D). In addition, no pairwise F_{ST} estimates between sampling sites were significantly different from zero (Fig S3D).

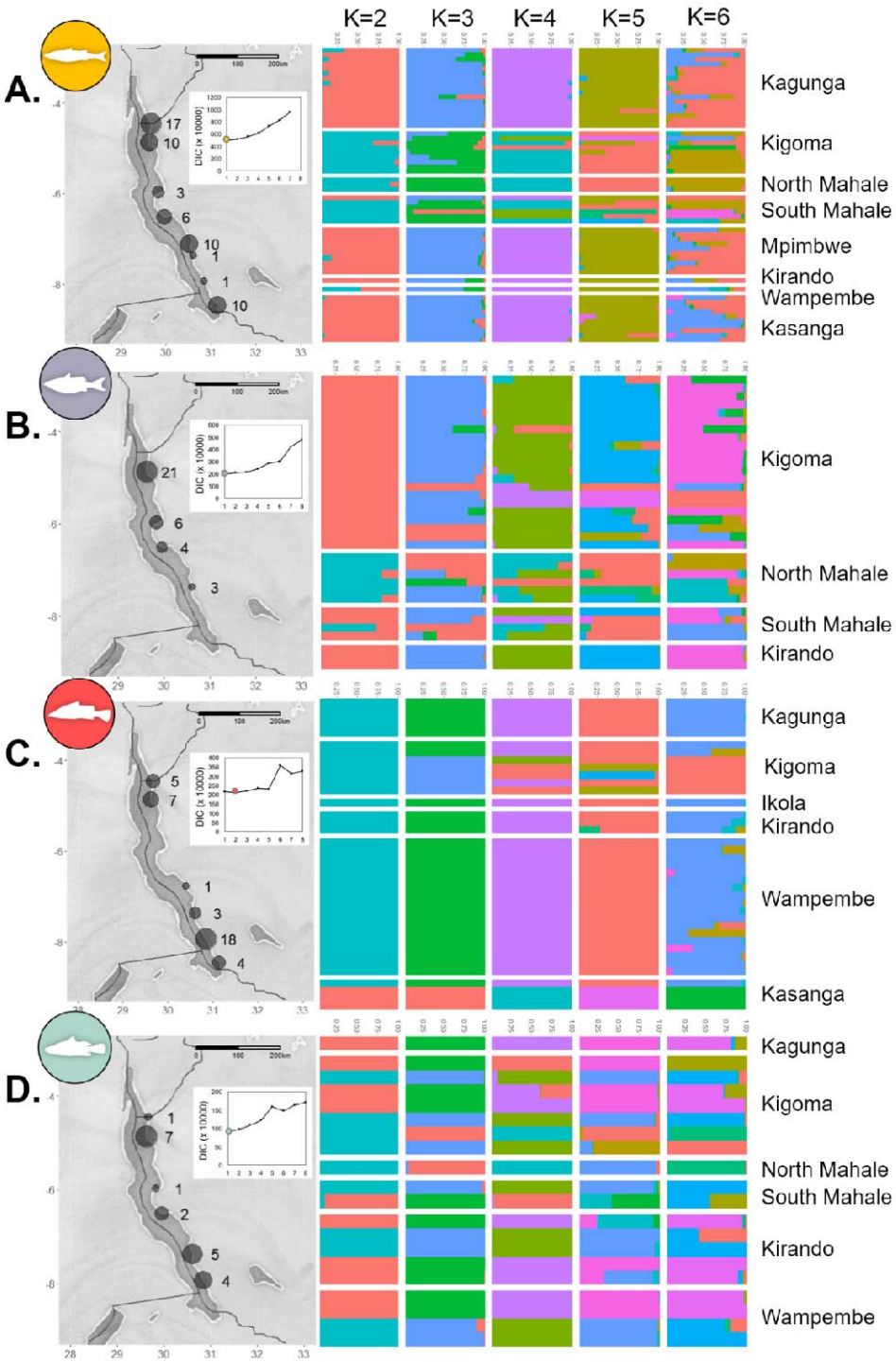

Figure 4. Intraspecific population structure inferred from the GBS data set and entropy analyses at K=2 to K=6 for (A) *L. stappersii*, (B) *L. microlepis*, (C) *L. mariae*, and (D) *L. angustifrons*. Maps show

distribution of individuals and inset plots show discriminant information content (DIC) values for each value of K, with the lowest DIC for each species indicated with a colored dot.

Discussion

Differences in life history characteristics can have significant influence on differences in population genetic structure among taxa, even in closely related species. In this study, we tested for genetic population structure within the four pelagic top predators of Lake Tanganyika, using reduced-representation genomic sequencing data sets (GBS and RAD). We find weak but distinct genetic structure in *L. stappersii* and *L. microlepis*, apparent in genetic PCA analyses, in pairwise F_{ST}s between samples of the intraspecific groups, and with the model-based genetic clustering program entropy (Gompert *et al.* 2014). In *L. mariae* we find three out of four samples from the most southern sampling site, Kasanga (and to a lesser extent five out of seven samples from Kigoma) to be genetically distinct from all other samples. This result was again apparent in genetic PCA analysis, pairwise F_{ST}s between the intraspecific groups, and with clustering analysis in entropy. Finally, we found no pattern of isolation-by-distance in any of the species.

Genetic patterns in the four Lates species of Lake Tanganyika

We found at least two genetic groups in each of *L. stappersii*, *L. microlepis,* and *L. mariae*. These clusters are weak in *L. stappersii* and *L. microlepis* and more pronounced in *L. mariae*. Although the samples were all collected along the 450km Tanzanian shoreline of Lake Tanganyika, none of the genetic differences between populations seems to be related to geographic distance. Instead, the clusters were associated with individuals from the central and southern section of the north basin in *L. stappersii* and from the central basin in *L. microlepis,* and by southern samples and to a lesser degree by samples from the southern north basin in *L. mariae*. Although these results suggest some genetic differentiation in these groups, they are not strictly nor ubiquitously tied to particular sampling sites.

In *L. mariae*, the majority of samples from the most southern sampling site, Kasanga, are distinct from all other samples. The magnitude of F_{ST} estimates between these distinct individuals and other *L. mariae* individuals (F_{ST} ~0.13) in the absence of geographical isolation could suggest not only intraspecific structure, but potentially even two distinct species in *L. mariae*. Anecdotally, catches of both *L. mariae* and *L. microlepis* are also highest in the southern part of the lake, suggesting a high abundance of these two species. It is noteworthy that the southern part of the lake experiences intense mixing of cold, nutrient rich waters from the deep with oxygenated water from the surface (Verburg *et al.*

2011) and consequently has the highest nutrient concentrations and primary productivity rates, which in turn fuels zooplankton growth (Plisnier *et al.* 2009; Loiselle *et al.* 2014). Furthermore, the oxygenated layer in the south extends down to more than 120m, and water temperatures are reduced (Bergamino *et al.* 2010; Verburg *et al.* 2011). As a result, there is high environmental variability over the course of the year in the south. Such a variable environment might lead to local differential selection, and this selection combined with restricted movement of individuals between this and other populations of *L. mariae* could lead to the genetic differentiation in most individuals from Kasanga. A variable environment may also lead to within-site differentiation among individuals caught in Kasanga. While all individuals sampled from Kasanga were mature and *L. mariae* is thought to be the most resident species of all four *Lates* species (Coulter 1976), it is not clear how resident these individuals really are; it is possible that some individuals remain resident while others migrate, such as is the case in Atlantic cod (Kirubakaran *et al.* 2016; Berg et al. 2017). Future studies should focus on expanding sampling here, including sampling various life stages of this and other *Lates.* In addition, several *L. mariae* from Kigoma formed a genetically distinct cluster from other samples caught in the region, suggesting differentiation in multiple regions of the lake in this species not explained by geography. If there is site fidelity to regionally differentiated spawning grounds but between-region movement outside of spawning, this could create such patterns of differentiation. However, we do not have data on the spawning status of the fish collected in this study, which would be required to investigate this hypothesis.

Despite the differences in life history between pelagic cichlids such as *Diplotaxodon* and *Bathybatini* and the *Lates* species, the genetic patterns between the benthopelagic cichlids and the benthopelagic *L. mariae* resemble each other in part. For the four benthopelagic *Diplotaxodon* species of Lake Malawi with genetically structured populations, Genner *et al.* (2010) implicated spawning site fidelity in driving the structure. Koblmüller *et al.* (2019) hypothesize that preying upon benthic cichlids and therefore reduced movements can explain the genetic structure in the *Bathybatini* of Lake Tanganyika. Similar to the *Bathybatini, L. mariae* hunts benthic cichlids and is described as relatively resident (Coulter 1976). Our results did not recover any evidence for significant isolation-by-distance in *L. mariae*, suggesting that the genetic structure that we observe is not the result of dispersal limitation. However, females of *L. mariae* aggregate over spawning grounds, similar to *Diplotaxon*. We obtained all samples from fishermen, and thus a range of ages and reproductive maturities of fish were included in our study. If samples of *L. mariae* were taken over spawning grounds, it is possible that results would reveal genetically clearly distinct populations as would be expected from strong spawning site fidelity and differentiation between these stocks. However, the location of spawning grounds is unclear for all four

Lates species; furthermore, it is unclear how the pelagic phase of eggs, larvae and small juveniles (<2.5cm) of *L. mariae* influences dispersal in the species. Future studies should take care to sample fish at spawning, and to work to understand the location of spawning grounds, to further assess whether the intraspecific divergences that we see could be linked to spawning site differentiation and whether some of this divergence may actually reflect multiple species within *L. mariae*.

Koblmüller *et al.* (2019) found no evidence of genetic population structure in two eupelagic cichlid species of Lake Tanganyika, whereas we do find distinct genetic groups even in the entirely pelagic species *L. stappersii*, and the mainly pelagic *L. microlepis*. Given the differences between mouthbrooding cichlids with large actively swimming offspring and free-spawning *Lates* with tiny pelagic larvae, this contrast is surprising. Although we do not find evidence for spatial genetic structure in *L. stappersii*, we note that individuals of this species caught during the dry season in the south of the lake and samples caught in the north show distinct carbon stable isotope signals (Junker, unpublished data). These differences in carbon isotope ratios correlate with different rates of primary productivity in the northern and the southern basin during the dry season (Loiselle *et al.* 2014), or with differences in prey composition in different areas of the lake (Plisner *et al.* 2009; Kurki *et al.* 1999). This is evidence for at least partly resident populations of *L. stappersii*, which must stay in the northern and southern basin during the dry season for long enough to incorporate the different isotopic signals of these basins. Such movement patterns may reduce gene flow between *L. stappersii* populations from the basins. However, the distinct genetic groups in *L. stappersii* are not associated with the basins: the same genetic group is widespread in the south and the north of the lake but is rare in the middle section. This is very similar to the patterns which we also find in *L. microlepis* and *L. mariae*. Although the geographically widespread groups agree with observations of long-distance movements in *L. stappersii* and *L. microlepis*, it conflicts with the evidence of a rather resident lifestyle in *L. mariae* (Chapman 1976; Coulter 1976, 1991; Muller *et al.* 2001).

We find groups with restricted gene flow in three of the four *Lates* species. One potential explanation for this consistent structure could lie in the long history of Lake Tanganyika. With an estimated age of 9 -12 million years, Lake Tanganyika is one of the oldest lakes in the world and has been influenced by extensive lake level fluctuations during its history (Cohen *et al.* 1993). Based on the bathymetry of the lake, it is thought that Lake Tanganyika was divided into three paleo-lakes during periods of extremely low water levels (Cohen *et al.* 1997). Environmental fluctuations such as these may have led to oscillations between sympatry or parapatry and allopatry for fish populations within the lake. Such fluctuations could lead to persistent genetic structure within species, for example if spawning site fidelity within separate basins is maintained after the basins reconnect. Such processes may also

have ties to the origin of these species through allopatric speciation, a phenomenon referred to as a “species pump” (Greenwood 1981; Sedano *et al.* 2010). In Lake Tanganyika, these water level fluctuations have been implicated in speciation in several different cichlid genera (Janzen & Etienne 2017; Sturmbauer *et al.* 2016). However, in contrast to the *Lates*, all of these cichlid genera are strictly littoral and highly philopatric. In addition, it has recently been suggested that the origin of the Lake Tanganyika *Lates* radiation is much younger than other endemic fish radiations in the lake (Koblmüller *et al.* 2021). Moreover, it remains unclear whether the timing of these events could explain the current weak differences in the *L. stappersii* and *L. microlepis*. For this mechanism to maintain intraspecific genetic structure, the genetic groups would have remained reproductively isolated even after the re-establishment of sympatry with rising lake levels.

In *L. angustifrons*, we found no genetic structure, no systematic differentiation in PCA space, and no evidence for genetic differentiation in pairwise F_{ST} estimates. This makes the possibility of localized populations or site fidelity to aggregated spawning grounds of mature individuals for this species unlikely, despite its benthic habitat affiliations. However, because of the challenge in collecting samples of this largest and rarest *Lates* species, our sample sizes of *L. angustifrons* are smaller than in our datasets for the three other species, and we lack samples from the southernmost sites. Additionally, since all the *Lates* species have a pelagic larval phase before settling into the inshore nursery habitat, it is possible that larval dispersal in this species prevents the emergence of any geographic structuring despite limited movement of adult fish.

Recommendations for conservation and sustainable management of Lake Tanganika’s Lates

We find evidence for weak yet genetically distinct groups within *L. stappersii* and *L. microlepis*. We furthermore find evidence for genetically distinct groups within *L. mariae*, in agreement with previous studies suggesting that this species is the most resident of all four *Lates* species (Coulter 1976, 1991). Intriguingly, this structure is not linked to isolation-by-distance, and instead seems to involve strong differentiation between some individuals at multiple sites, suggesting the existence of a mechanism for maintaining differentiation in sympatry, such as spawning site fidelity, differences in spawning time, or behavioral assortative mating. In contrast, the low F_{ST} estimates between samples belonging to the same genetic cluster from distant sites suggest that these species do disperse over large distances.

To maintain intraspecific genetic variation, viable population sizes, and possible adaptive differences between genetic groups within the *Lates* species, management strategies should

seek to restrict catches, especially in the spawning areas. These practices have been implemented in regions around Lake Tanganyika where community conservation areas are active (Kimirei & Sweke 2018), but are not in place throughout the lake. Such management practices are performed successfully in the cod fishery of Denmark (albeit after collapse of this fishery and extinction of many of the original stocks; Dahle *et al.* 2018), and in the salmon fishery of Bristol Bay (Schindler *et al.* 2015; Schindler *et al.* 2010). However, implementation of such a strategy requires updated scientific evidence on the spawning grounds and seasons, and the ability for fishermen to distinguish between the genetically distinct stocks. Therefore, studies are needed to shed more light on distinct genetic groups within each species, both phenotypically and ecologically, and to assess the spawning grounds for the four species. In addition, studies are needed to examine whether fish from different spawning grounds form the distinct genetic clusters that we identified in this study. Furthermore, a sustainable management approach is crucial to protect the inshore nursery habitats and ensure recruitment of the species in Lake Tanganyika.

Funding

This work was supported by the Swiss National Science Foundation (grant CR23I2-166589 to OS), the University of Wyoming (start-up funding to CEW), the National Science Foundation (grant DEB-1556963 to CEW), the National Institute of General Medical Sciences of the National Institutes of Health (Institutional Development Award grant P20GM103432), and the American Genetics Association (Ecological, Evolutionary, and Conservation Genomics Research Award to JAR).

Acknowledgements

Computing was accomplished with an allocation from the University of Wyoming's Advanced Research Computing Center, on its Teton Intel x86_64 cluster (https://doi.org/10.15786/M2FY47) and the Genetic Diversity Center (GDC) of ETH Zürich. Special thanks go to Mupape Mukuli for facilitating logistics during fieldwork and to the crew of the MV Maman Benita for fieldwork assistance. We thank the whole team at the Tanzanian Fisheries Research Institute for their support. A special thanks goes to Mary Kishe for her support during fieldwork permission processes and to the Tanzanian Commission for Science and Technology (COSTECH) for the permits to do this research. Live fish samples were collected under the approved University of Wyoming IACUC protocol #20160602CW00241-01. We appreciate The Nature Conservancy's Tuungane Project for providing funding and logistical support, with special thanks going to Colin Apse and Peter

Limbu. Thanks to the Wagner lab at the University of Wyoming and to the FishEc group at EAWAG for discussions about the analyses and manuscript. Thanks to C. Alex Buerkle for discussion of entropy and clustering results.

Data Availability

We have deposited the primary data underlying these analyses as follows:

- Sampling locations, morphological data, filtered VCF files: Dryad XXXXX
- Fastq files for each individual: NCBI SRA XXXXX
- Scripts for data analysis: Github (https://github.com/jessicarick/lates-popgen)

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**.

Belgrano A, Fowler CW (2011) *Ecosystem-based management for Marine Fisheries: An evolving perspective* Cambridge University Press, Cambridge.

Berg PR, Star B, Pampoulie C*, et al.* (2017) Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity* **119**, 418-428.

Bergamino N, Horion S, Stenuite S*, et al.* (2010) Spatio-temporal dynamics of phytoplankton and primary production in Lake Tanganyika using a MODIS based bio-optical time series. *Remote Sensing of Environment* **114**, 772-780.

Botsford LW, Castilla JC, Peterson CH (1997) The Management of Fisheries and Marine Ecosystems. *Science* **277**, 509-515.

Brannon EL, Powell MS, Quinn TP, Talbot A (2004) Population Structure of Columbia River Basin Chinook Salmon and Steelhead Trout. *Reviews in Fisheries Science* **12**, 99-232.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**, 3124-3140.

Chapman DW (1976) Acoustic Estimates of Pelagic Ichthyomass in Lake Tanganyika with an Inexpensive Echo Sounder. *Transactions of the American Fisheries Society* **105**, 581-587.

Cohen AS, Gergurich EL, Kraemer BM*, et al.* (2016) Climate warming reduces fish production and benthic habitat in Lake Tanganyika, one of the most biodiverse freshwater ecosystems. *Proc Natl Acad Sci U S A* **113**, 9563-9568.

Cohen AS, Lezzar K-E, Tiercelin J-J, Soreghan M (1997) New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. *Basin Research* **9**, 107-132.

Cohen AS, Soreghan MJ, Scholz CA (1993) Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. *Geology* **21**, 511-514.

Coulter GW (1976) The biology of Lates species (Nile perch) in Lake Tanganyika, and the status of the pelagic fishery for Lates species and Luciolates stappersii (Blgr.). *Journal of Fish Biology* **9**, 235-259.

Coulter GW (1991) *Lake Tanganyika and its Life* British Museum (Natural History) Cromwell Road, London SW7 5BD & Oxford University Press, Walton Street, Oxford OX2 6DP.

Dahle G, Johansen T, Westgaard J-I, Aglen A, Glover KA (2018) Genetic management of mixed-stock fisheries “real-time”: The case of the largest remaining cod fishery operating in the Atlantic in 2007–2017. *Fisheries Research* **205**, 77-85.

Danecek P, Auton A, Abecasis G*, et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *2007* **22**, 20.

Eccles DH (1992) Field guide to the freshwater fishes of Tanzania. In: *species identification sheets for fishery purposes* (ed. FAO), p. 145. FAO, Rome.

Ellegren H, Galtier N (2016) Determinants of genetic diversity. *Nature Reviews Genetics* **17**, 422-433.

Ellis CMA (1978) Biology of Luciolates-Stappersi in Lake Tanganyika (Burundi). *Transactions of the American Fisheries Society* **107**, 557-566.

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.

Genner MJ, Nichols P, Shaw PW*, et al.* (2010) Population structure on breeding grounds of Lake Malawi's 'twilight zone' cichlid fishes. *Journal of Biogeography* **37**, 258-269.

Gompert Z, Lucas LK, Buerkle CA*, et al.* (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Mol Ecol* **23**, 4555-4573.

Greenwood, PH (1981) Species-flocks and explosive evolution. *Chance, change and challenge- the evolving biosphere*, 61-74.

Gruber B, Unmack PJ, Berry OF, Georges A (2018) dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources* **18**, 691-699.

Hutchinson WF (2008) The dangers of ignoring stock complexity in fishery management: the case of the North Sea cod. *Biol Lett* **4**, 693-695.

Janzen T, Etienne RS (2017) Inferring the role of habitat dynamics in driving diversification: evidence for a species pump in Lake Tanganyika cichlids.

Junker J, Rick JA, McIntyre PB*, et al.* (2020) Structural genomic variation leads to genetic differentiation in Lake Tanganyika's sardines. *Mol Ecol* **29**, 3277-3298.

Kimirei IA, Mgaya YD, Chande AI (2008) Changes in species composition and abundance of commercially important pelagic fish species in Kigoma area, Lake Tanganyika, Tanzania. *Aquatic Ecosystem Health and Management* **11**, 29-35.

Kimirei IA, Sweke EA (2018) Community-based conservation areas as a fisheries management tool in the East African Great Lakes: an example from Tanzania. In: 34th SIL Congress, 19-24 August 2018. Nanjing, China. pp. 230 Book of abstracts.

Kirubakaran TG, Grove H, Kent MP*, et al.* (2016) Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol Ecol* **25**, 2130-2143.

Koblmüller S, Zangl L, Börger C*, et al.* (2019) Only true pelagics mix: comparative phylogeography of deepwater bathybatine cichlids from Lake Tanganyika. *Hydrobiologia* **832**, 93-103.

Koblmüller S, Schöggl CA, Lorber CJ, *et al.* (2021) African lates perches (Teleostei, Latidae, *Lates*): Paraphyly of Nile perch and recent colonization of Lake Tanganyika. *Molecular Phylogenetics and Evolution* **160**: 107141.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* **15**, 1179-1191.

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* **14**, 289.

Kurki H, Mannini P, Vuorinen I, *et al*. (1999) Macrozooplankton communities in Lake Tanganyika indicate food chain differences between the northern part and the main basins. *Hydrobiologia* **407**, 123-129.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359.

Lewis PO (2001) A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology* **50**, 913-925.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754-1760.

Li H, Handsaker B, Wysoker A*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Loiselle S, Cózar A, Adgo E, *et al.* (2014) Decadal Trends and Common Dynamics of the Bio-Optical and Thermal Characteristics of the African Great Lakes. *PLoS One* **9**, e93656.

LTA Secretariat (2011) Report on Regional Lakewide Fisheries Frame Survey on Lake Tanganyika 2011. LTA/TECH DOC/2012/01. Bujumbura, Burundi.

Mace GM, Purvis A (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Mol Ecol* **17**, 9-19.

Malinsky M, Svardal H, Tyers AM*, et al.* (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution* **2**, 1940-1955.

Manier MK, Arnold SJ (2006) Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. *Proceedings of the Royal Society B: Biological Sciences* **273**, 3001-3009.

Mannini P (1998a) *Ecology of the Pelagic Fish Resources of Lake Tanganyika*, University of Hull.

Mannini P (1998b) Geographical distribution patterns of pelagic fish and macrozooplankton in Lake Tanganyika. FAO/FINNIDA Research for the Management of the Fisheries on Lake Tanganyika. . In: *GCP/RAF/271/FIN-TD/83 (En)*, p. 103 p.

Mannini P, Aro E, Katonda I*, et al.* (1996) Pelagic fish stocks of Lake Tanganyika: biology and exploitation. FAO/FINNIDA Research for the Management of the Fisheries of Lake Tanganyika. In: *GCP/RAF/271/FIN—TD/53 (En)*, p. 60p.

Mannini P, Katonda I, Kissaka B, Verburg P (1999) Feeding ecology of Lates stappersii in Lake Tanganyika. *Hydrobiologia* **407**, 131-139.

Martinez Barrio A, Lamichhaney S, Fan G*, et al.* (2016) The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *Elife* **5**.

Meisner J, Liu S, Huang M, Albrechtsen A (2021) Large-scale inference of population structure in presence of missingness using PCA. *Bioinformatics*, btab027.

Mölsä H, Reynolds JE, Coenen EJ, Lindqvist OV (1999) Fisheries research towards resource management on Lake Tanganyika. *Hydrobiologia* **407**, 1-24.

Mölsä H, Sarvala J, Badende S*, et al.* (2002) Ecosystem monitoring in the development of sustainable fisheries in Lake Tanganyika. *Aquatic Ecosystem Health and Management* **5**, 267-281.

Muller P, Li XP, Niyogi KK (2001) Non-photochemical quenching. A response to excess light energy. *Plant Physiol* **125**, 1558-1566.

Munyandero J (2002) The lake Tanganyika clupeid and latid fishery system: Indicators and problems inherent in assessments and management. *African Study Monographs* **23**, 117-145.

Ntakimazi G (2006a) Lates angustifrons. The IUCN Red List of Threatened Species 2006: e.T60833A12416551. http://dx.doi.org/10.2305/IUCN.UK.2006.RLTS.T60833A12416551.en . Downloaded on 30 September 2015.

Ntakimazi G (2006b) Lates mariae. The IUCN Red List of Threatened Species 2006: e.T60836A12416789. http://dx.doi.org/10.2305/IUCN.UK.2006.RLTS.T60836A12416789.en . Downloaded on 30 September 2015.

Ntakimazi G (2006c) Lates microlepis. The IUCN Red List of Threatened Species 2006: e.T60837A12416878. http://dx.doi.org/10.2305/IUCN.UK.2006.RLTS.T60837A12416878.en . Downloaded on 30 September 2015.

O'Reilly CM, Alin SR, Piisnier PD, Cohen AS, McKee BA (2003) Climate change decreases aquatic ecosystem productivity of Lake Tanganyika, Africa. *Nature* **424**, 766-768.

Olsen JB, Crane PA, Flannery BG*, et al.* (2011) Comparative landscape genetic analysis of three Pacific salmon species from subarctic North America. *Conservation Genetics* **12**, 223-241.

Paffen P, Coenen E, Bambara S*, et al.* (1997) Synthesis of the 1995 simultaneous frame survey of Lake Tanganyika Fisheries. FAO/FINNIDA Research for the Management of the Fisheries on Lake Tanganyika. In: *GCP/RAF/271/FIN—TD/60 (En)*, p. 22p.

Parchman TL, Gompert Z, Mudge J*, et al.* (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol Ecol* **21**, 2991-3005.

Peterman WE, Anderson TL, Ousterhout BH*, et al.* (2015) Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conservation Genetics* **16**, 59-69.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. *Mol Ecol* **14**, 2943-2957.

Pettersson ME, Rochus CM, Han F*, et al.* (2019) A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. *Genome Research* **29**, 1919-1928.

Plisnier PD, Mgana H, Kimirei I, *et al.* (2009) Limnological variability and pelagic fish abundance (Stolothrissa tanganicae and Lates stappersii) in Lake Tanganyika. *Hydrobiologia* **625**,117-134.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* **461**, 489-494.

Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ (2009) Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries* **10**, 361-395.

Riginos C, Buckley YM, Blomberg SP*, et al.* (2014) Dispersal Capacity Predicts Both Population Genetic Structure and Species Richness in Reef Fishes. *The American Naturalist* **184**, 52-64.

Sarvala J, Tarvainen M, Salonen K, Mölsä H (2002) Pelagic food web as the basis of fisheries in Lake Tanganyika: A bioenergetic modeling analysis. *Aquatic Ecosystem Health and Management* **5**, 283-292.

Schindler DE, Armstrong JB, Reed TE (2015) The portfolio concept in ecology and evolution. *Frontiers in Ecology and the Environment* **13**, 257-263.

Schindler DE, Hilborn R, Chasco B*, et al.* (2010) Population diversity and the portfolio effect in an exploited species. *Nature* **465**, 609-612.

Sedano RE, Burns KJ, Riddle B (2010) Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). *Journal of Biogeography* **37**, 325-343.

Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* **436**, 291-305.

Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693-702.

Snoeks J, Carvalho GR (2004) *The Cichlid Diversity of Lake Malawi/Nyasa/Niassa: Identification, Distribution and Taxonomy* Cichlid Press.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.

Stearns SC (1992) *The evolution of life histories* Oxford Univ. Press, New York.

Stenuite S, Pirlot S, Hardy MA*, et al.* (2007) Phytoplankton production and growth rate in Lake Tanganyika: evidence of a decline in primary productivity in recent decades. *Freshwater Biology* **52**, 2226-2239.

Sterner T (2007) Unobserved diversity, depletion and irreversibility The importance of subpopulations for management of cod stocks. *Ecological Economics* **61**, 566-574.

Sturmbauer C, Börger C, Van Steenberge M, Koblmüller S (2016) A separate lowstand lake at the northern edge of Lake Tanganyika? Evidence from phylogeographic patterns in the cichlid genus Tropheus. *Hydrobiologia* **791**, 51-68.

Sunday JM, Popovic I, Palen WJ, Foreman MG, Hart MW (2014) Ocean circulation model predicts high genetic structure observed in a long-lived pelagic developer. *Mol Ecol* **23**, 5036-5047.

[dataset] Temasek Life Sciences Laboratory (2016). Lates calcarifer isolate:ASB-BC8 Genome sequencing and assembly. National Center for Biotechnology Information Sequence Read Archive, Assembly GCA_001640805.1.

Therkildsen NO, Hemmer-Hansen J, Als TD*, et al.* (2013) Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. *Mol Ecol* **22**, 2424-2440.

van der Knaap M (2013) Comparative analysis of fisheries restoration and public participation in Lake Victoria and Lake Tanganyika. *Aquatic Ecosystem Health and Management* **16**, 279-287.

Van der Knaap M, Katonda KI, De Graaf GJ (2014) Lake Tanganyika fisheries frame survey analysis: Assessment of the options for management of the fisheries of Lake Tanganyika. *Aquatic Ecosystem Health and Management* **17**, 4-13.

van Zwieten PAM, Roest FC, Machiels MAM, Van Densen WLT (2002) Effects of inter-annual variability, seasonality and persistence on the perception of long-term trends in catch rates of the industrial pelagic purse-seine fishery of northern Lake Tanganyika (Burundi). *Fisheries Research* **54**, 329-348.

van Zwieten PAM, Kolding J, Plank MJ*, et al.* (2016) The Nile perch invasion in Lake Victoria: cause or consequence of the haplochromine decline? *Canadian Journal of Fisheries and Aquatic Sciences* **73**, 622-643.

Verburg P, Antenucci JP, Hecky RE (2011) Differential cooling drives large-scale convective circulation in Lake Tanganyika. *Limnology and Oceanography* **56**, 910-926.

Verburg P, Hecky RE (2009) The physics of the warming of Lake Tanganyika by climate change. *Limnology and Oceanography* **54**, 2418-2430.

Verburg P, Hecky RE, Kling H (2003) Ecological consequences of a century of warming in Lake Tanganyika. *Science* **301**, 505-507.

Vij S, Kuhl H, Kuznetsova IS*, et al.* (2016) Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-layered Scaffolding. *PLOS Genetics* **12**, e1005954-e1005954.

Wenburg JK, Bentzen P, Foote CJ (1998) Microsatellite analysis of genetic population structure in an endangered salmonid: the coastal cutthroat trout (Oncorhynchus clarki clarki). *Mol Ecol* **7**, 733-749.

Wilson AB, Teugels GG, Meyer A (2008) Marine incursion: the freshwater herring of Lake Tanganyika are the product of a marine invasion into west Africa. *PLoS One* **3**, e1979.

Young EF, Belchier M, Hauser L*, et al.* (2015) Oceanography and life history predict contrasting genetic population structure in two Antarctic fish species. *Evolutionary Applications* **8**, 486-509.