Genetic Structure of Pelagic and Littoral Cichlid Fishes from Lake Victoria

武田, 深幸

https://doi.org/10.15017/1441023

出版情報：九州大学, 2013, 博士（理学）, 課程博士
バージョン：
権利関係：全文ファイル公表済

Genetic Structure of Pelagic and Littoral Cichlid Fishes from Lake Victoria

Miyuki Takeda

Department of Biology, Faculty of Sciences,

Kyushu University, Fukuoka, 812-8581, Japan

1. Abstract

The approximately 700 species of cichlids found in Lake Victoria in East Africa are thought to have evolved over a short period of time, and they represent one of the largest known examples of adaptive radiation. To understand the processes that are driving this spectacular radiation, we must determine the present genetic structure of these species and elucidate how this structure relates to the ecological conditions that caused their adaptation. I analyzed the genetic structure of two pelagic and seven littoral species sampled from the southeast area of Lake Victoria using sequences from the mtDNA control region and 12 microsatellite loci as markers. Using a Bayesian model-based clustering method to analyze the microsatellite data, I separated these nine species into four groups: one group composed of pelagic species and another three groups composed mainly of rocky-shore species. Furthermore, I found significant levels of genetic variation between species within each group at both marker loci using analysis of molecular variance (AMOVA), although the nine species often shared mtDNA haplotypes. I also found significant levels of genetic variation between populations within species. These results suggest that initial groupings, some of which appear to have been related to habitat differences, as well as divergence between species within groups took place among the cichlid species of Lake Victoria.

2. Introduction

Cichlid fishes are tropical freshwater fish, and form one of the most species rich families of vertebrates (Barlow 2002). A large number of cichlid fish species have been identified in each of the three great lakes of East Africa, Lakes Tanganyika, Malawi, and Victoria. Turner *et al*. (2001) have estimated that there are approximately 250, 700 and 700 cichlid species endemic to Lakes Tanganyika, Malawi, and Victoria, respectively. Moreover, these species possess a wide variety of adaptations to specific environments, which seem to have developed over fairly short periods of time (Seehausen 2006). These adaptations have often involved changes to morphology and sensory organ structure, and they appear to have occurred independently within each lake (Kocher 2004). Previous geological and molecular genetic studies have revealed the age of each cichlid fauna (Salzburger & Meyer 2004, Johnson *et al*. 1996, Turner *et al*. 2001, Verheyen *et al*. 2003, Genner *et al*. 2007, Coyne & Orr 2004). According to these studies, the species flock in Lake Tanganyika is the oldest and followed by the species flock in Lake Malawi. Lake Victoria harbors the youngest species flock. In Lake Victoria, for example, the species flock of endemic cichlids is thought to be either monophyletic (Verheyen *et al*. 2003) or of hybrid origin from colonizing lineages (Seehausen *et al*. 2003) and includes morphologically and ecologically diverse species (Greenwood 1974). Furthermore, clear examples of ecological speciation are known to have occurred for the cichlids in this lake (e.g., Terai *et al*. 2006, Seehausen *et al*. 2008). Therefore, these species provide us with an excellent opportunity to study adaptive radiation.

An important question with respect to adaptive radiation is how its initial stages are affected by different habitats. Using ultrametric trees of the Lake Victoria radiation,

Seehausen *et al*. (2003) inferred that the radiation can be thought of as a starburst pattern with either very short or no branches separating any two speciation events. In this case, species groupings based on habitat would not be apparent. On the other hand, Danley & Kocher (2001) proposed that adaptive radiation first occurred by adaptation to different habitats – rocky and sandy habitats, in the case of Lake Malawi – followed by diversification with respect to trophic morphology and male nuptial color within each habitat. In this scenario, species groupings based on habitat would be apparent even at early stages and could be identified as a hierarchical genetic structure related to habitat. Danley & Kocher (2001) proposed this scenario of evolutionary radiation based on the phylogenetic relationships between cichlid species in Lake Malawi, and they also cited examples from other species groups to suggest that this may represent a general mode of adaptive diversification. However, because many species may have gone extinct following the initial burst of adaptive radiation (Seehausen 2006), it would be difficult to reconstruct the initial stages of diversification by examining only surviving species generated by older adaptive radiations.

The species flock of cichlids found in Lake Victoria is an excellent group of organisms with which to investigate the initial stage of adaptive radiation. As mentioned above, a variety of species adapted to different habitats exist (Seehausen 1996), and the species flock is thought to have diverged over the last 100,000 years or less (Verheyen *et al*. 2003, Genner *et al*. 2007). Therefore, it may still be possible to infer the initial stage of adaptive radiation. However, the fact that the diversification of the species flock within Lake Victoria occurred so recently also poses a problem. Because the speciation events were so recent, genetic differentiation between species can be weak, and it is difficult to infer phylogenetic relationships within the species

flock using neutral markers (Seehausen *et al*. 2003, Nagl *et al*. 1998, Elmer *et al*. 2009). Indeed, Samonte *et al*. (2007) have suggested that gene flow between species can be as extensive as that between local populations of the individual species within Lake Victoria. However, previous population genetic studies based on many individuals of pelagic cichlid species from Lake Victoria found significant, albeit weak, genetic differentiation between these species (Maeda *et al*. 2009, Mzighani *et al*. 2010). Therefore, by sampling many individuals from several species living in different habitats, I may be able to infer diversification patterns from the initial stage of adaptive radiation using neutral markers. Alternatively, this can be achieved using many more markers with a smaller number of samples for each species. Indeed, Bezault *et al*. (2011), using amplified fragment length polymorphism (AFLP) markers, and Wagner *et al*. (2013) and Keller *et al*. (2012), using restriction-site-associated DNA (RAD) markers, have recently found significant differentiation between the cichlid fish species of Lake Victoria. Although RAD markers revealed that the cichlid fish species of Lake Victoria were reciprocally monophyletic, the initial diversification is still unclear in Lake Victoria. Moreover, because Wagner *et al*. (2013) and Keller *et al*. (2012) used only species caught in the same habitat - rocky shores, it is not clear whether the differentiation exists between the cichlid species adapted to different habitats.

In general, it is considered that differentiation between populations gradually increases and if gene flow between them ceases by some factor, these populations will become different species. Investigation of the population structure in one species provides us clues to its evolutionary history and biology. Previous studies on the cichlid fish spesies of Lake Malawi suggested that population structure differ between different species adapted to different habitats. Shaw *et al*. (2000) have shown that

each of a few pelagic cichlid species does not have recognizable population structure across the Lake Malawi that is about 500 km long at its maximum. They have suggested that because pelagic species have high mobility, each species forms a single, panmictic population across the lake. In contrast, in littoral, rock cichlid species of Lake Malawi, it has been shown that population differentiation exists between poplations isolated by only several dozen meter or several hundreds meter (Pereyra *et al*. 2004, Rico & Turner 2002, Markert *et al*. 1999, van Oppen *et al*. 1997, Arnegard *et al*. 1999). Among these works, the most notable finding was obtained by Rico & Turner (2002). They have observed significant population differentiation between populations isolated by only 35 m. It was thought that because rock cichlids have generally low mobility, they tend to have strong population differentiation. Although several such studies relating the population struture and fish mobility in the cichlids of Lake Malawi have been reported, population structure has not been well studied in the cichlids of Lake Victoria. Only a few pelagic species in Lake Victoria have been examined for genetic differentiation between species and between populations in detail (Maeda *et al*. 2009, Mzighani *et al*. 2010). Therefore, we still do not know, for example, how different habitats, such as pelagic and littoral ones, affect fish's population structure.

In this study, I genotyped populations of seven littoral cichlid species collected from the southern part of Lake Victoria using a mitochondrial marker and 12 microsatellite markers that were developed by Maeda *et al*. (2008). Furthermore, I combined these data with genetic information gathered previously from two pelagic species using the same set of markers (Maeda *et al*. 2009). By analyzing this dataset from nine species living in different habitats, I was able to address the following questions. (1) Does a hierarchical genetic structure for the cichlid fishes of Lake Victoria exist? (2) If so, is

this hierarchical structure related to habitat? (3) Finally, are there further genetic substructures within these cichlid fishes, and how might these relate to species?

3. Materials and Methods

3. 1. Ethics Statement

This study was conducted in collaboration with Dr. Norihiro Okada and members of his laboratory at Tokyo Institute of Technology and the Tanzania Fisheries Research Institute (TAFIRI), which also provided us with logistical support, including permissions for the field studies. In the field studies, we complied with local legislation and the Convention on Biological Diversity and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. The animal protocols and procedures were approved by the Institutional Animal Care and Use Committee of Tokyo Institute of Technology.

3.2. Sampling and DNA extraction

The research group of Tokyo Institute of Technology and members of TAFIRI collected specimens of seven littoral cichlid species from the southern part of Lake Victoria between September 2004 and November 2006. All specimens were collected by gill net (1.5-m height) or angling from a depth of 0–10 m. After collecting the fishes they took photographs to record live coloration and kept them in crushed ice immediately to kill the fishes without unnecessary pain. After killing the fish pectoral and pelvic fins or muscle from the right caudal peduncle were removed from each specimen and fixed in 100% ethanol. The remainder of each specimen was fixed in 10% formalin for later identification. Five to 10 mg of each ethanol-fixed tissue

sample was added to a 1.5 ml microfuge tube, and, following thorough homogenization with a sharp pair of scissors, genomic DNA was extracted using either the AquaPure Genomic DNA Isolation Kit (Bio-Rad, CA) or the DNeasy Blood & Tissue Kit (Qiagen, CA) according to the manufacturer's protocols. Identification of all specimens was verified by Drs. Mitsuto Aibara and Shinji Mizoiri of Tokyo Institute of Technology. The seven littoral species collected were *Lithochromis rubripinnis* Seehausen *et al*. 1998, *L. rufus* Seehausen *et al*. 1998, *Neochromis rufocaudalis* Seehausen *et al*. 1998, *N. greenwoodi* Seehausen *et al*. 1998, *N. omnicaeruleus* Seehausen *et al*. 1998, *Haplochromis* (*Paralabidochromis*) *sauvagei* (Pfeffer, 1896) (more specifically, this species was *H.* sp. "rockkribensis" sensu Seehausen, 1996 and; not *H. sauvagei* sensu Greenwood, 1957 and Barel *et al*., 1977; see Seegers (2008) for more information), and *Mbipia mbipi* Seehausen *et al*. 1998. In this study, I use conventional species names to simplify cross-referencing with other studies. Although *L. rufus* has been previously described as a rock-dwelling species, individuals of this species were caught in vegetation zones containing reed grass and/or papyrus. The other six species are territorial rock-dwelling species. Therefore, I consider *L. rufus* to occupy a different habitat from the other rock-dwelling species. Among the specimens, certain individuals belonging to *Lithochromis* or *Haplochromis* could not be identified at the species level, and I refer to these as *Lithochromis* sp. and *Haplochromis* sp., respectively. In addition, the samples included five specimens of *Pundamilia macrocephala*.
The total number of individuals genotyped are summarized in Table 1. Sampling details are shown in Fig. 1 and Table 2. Specimens caught in the same location were considered to be part of the same population.

3.3. Amplification of microsatellite loci and genotyping

I amplified 12 microsatellite loci from each sample using primers developed by Maeda *et al*. (2008). Forward primers used for the microsatellite markers were 5'-labeled with 6-FAM, NED, PET, or VIC dyes (Applied Biosystems, CA). Multiplex polymerase chain reactions (multiplex PCRs) were performed to amplify the target fragments using the QIAGEN Multiplex PCR Kit (Hilden, Germany). PCR amplifications were performed in a final reaction volume of 6.25 μL [3.125 μL 2× QIAGEN Multiplex PCR Master mix, 0.625 μL 10× Primer Mix (2μM), 1.5 μL RNase free water and 1 μL diluted DNA (containing <10 ng of genomic DNA)]. The PCR amplification conditions were as follows: genomic DNA was denatured for 15 min at 95°C, followed by 35 cycles of denaturation for 30 s at 94°C, annealing for 1 min and 30 s at 55°C, and extension for 1 min at 72°C. Extension was completed using a final incubation for 30 min at 60°C. The PCR products were run on an ABI3100 automated sequencer (Applied Biosystems, CA) with a GeneScan™ – 500 LIZ™ Size Standard (Applied Biosystems, CA) and genotyped using GeneMapper® Software Version 4.0 (Applied Biosystems, CA). To combine the results with those of the previous study (Maeda *et al*. 2009), the same bin sets were used for both experiments.

3.4. Mitochondrial DNA (mtDNA) amplification and sequencing

I amplified the mitochondrial control region using the primer pair SNmt-UP1 (5'-TAAAATCCTTCCTACTGCTTCA-3') and SNmt-LP1 (5'-TCAAACAAAATATGAATAACAAACA-3') as described by Nagl *et al*. (2000). These primers are specific to the $tRNA^{Pro}$ $tRNA^{Thr}$ gene and the 3'-end of the control region, respectively. The amplification products encompassed nearly the entire control region (approximately 850 bp). PCR amplification was performed using either Ex*Taq*™ (TaKaRa, Ohtsu, Japan) or GoTaq® DNA polymerase (Promega, WI) according to the manufacturer's recommendations. The PCR amplification conditions were as follows: DNA was denatured for 2 min at 94°C, followed by 30 cycles of denaturation for 40 s at 94°C, annealing for 30 s at 58°C, and extension for 1 min at 72°C. Extension was completed using a final incubation for 10 min at 72°C.

Two microliters of purified PCR product was used as a template in the cycle sequencing reactions. The primers used for sequencing were the two PCR primers SNmt-UP1 and SNmt-LP1 and two internal primers int-F (5'-CCTTTCATTTGACATCTCA-3') and int-R2 (5'-CACACGCTGGAAAGAACGCC-3'). When DNA sequencing results were ambiguous, two additional internal primers, int-F2 (5'-CCACCATCCTATTTACATCCCT-3') and int-R (5'-TCAACTGATGGTGGGCTCTT-3'), were used for further sequencing. The reaction mixture for the cycle sequencing consisted of 1.0 μL of each primer (1.6 μM), 1.25 μL Half BigDye (Genetix, New Milton, UK), 0.75 μL BigDye (Applied Biosystems, CA) and 5.0 μL of sterilized water. The annealing temperature for the cycle sequencing reactions was adjusted to 50°C. The DNA products were purified using ethanol/sodium-acetate precipitation, resuspended in 15 μL Hi-Di™ Formamide (Applied Biosystems, CA), and analyzed using an ABI PRISM 3100 capillary DNA

sequencer (Applied Biosystems, CA). All sequences obtained in this study have been deposited within the DNA Data Bank of Japan (DDBJ) under the accession numbers [DDBJ: AB762784 – AB 763333].

3.5. Data analyses

In the following analyses, I included data obtained by Maeda *et al*. (2009) for two pelagic species – *Haplochromis* (*Yssichromis*) *pyrrhocephalus* and *H.* (*Y.*) *laparogramma* – with the data from the seven species described above. For the STRUCTURE and haplotype network analyses, I also included data from the *Lithochromis* spp. and *Haplochromis* spp. specimens that could not be identified at the species level. For the remaining analyses, classification at the species level was necessary, and therefore, data from these specimens were not included. Furthermore, because the sample size for *P. macrocephala* was small (five), these data were only used for the STRUCTURE and haplotype network analyses.

First, to determine the population structure of the whole sample set, we applied a Bayesian model-based clustering method to the microsatellite data, which was implemented in STRUCTURE Version 2.3.3 (Pritchard *et al*. 2000, Falush *et al*. 2003). Briefly, the program assumed a certain number of populations (*K*) and assigned each individual to one of the populations based on its multi-locus genotype. In my analyses, I applied the admixture model, which assumed that each individual might have mixed ancestry. I assumed *K* to be between 1 and 20, and did not specify the origins of the samples. Each run consisted of 10,000 burn-in iterations, followed by 100,000 iterations to collect data. Other than these variables, the default program settings were used. I executed 20 runs for each *K* value, computed the averages of the

estimated Log probabilities of the data (ln P[*D*]), and calculated ΔK for each *K* using the method of Evanno *et al*. (2005). It has been suggested that ΔK can be used to detect the uppermost hierarchical level of genetic structure.

Sequences from the mitochondrial control region were edited and aligned by eye using the computer program Se-Al (Rambaut 1996). To this alignment, I added sequence data from 51 haplotypes of Lake Victoria, Lake Kivu and Lake Victoria obtained by Nagl *et al*. (2000) and Verheyen *et al*. (2003). Gaps were included in the sequences as information, as indels reflect the evolutionary history of the species. I constructed a haplotype network of these sequences using the program TCS (Clement *et al*. 2000). Alternative branching orders in the TCS-generated network were assessed using the maximum parsimony method and the software program PAUP* 4.0b10 (Swofford 2003). Only connections between haplotypes favored by the maximum parsimony criterion were used.

Next, to determine whether there were species-level differentiations of the mitochondrial and microsatellite loci within each group (defined in the Results section), I carried out Analysis of Molecular Variance (AMOVA) (Excoffier *et al*. 1992) on the eight species for which multiple populations were sampled. The hierarchy of the analysis was species/populations for the mtDNA and species/populations/individuals for the microsatellite markers. I used the software programs GenAlEx Version 6.4 (Peakall & Smouse 2006) for the mitochondrial data and Arlequin Version 3.5 (Excoffier & Lischer 2010) for the microsatellite data. Between populations, I also estimated R_{ST} (Slatkin 1995) for the microsatellite data and F_{ST} (Hudson *et al*. 1992) for the mitochondrial data using Arlequin Version 3.5 and DNAsp 5.0 (Librado & Rozas 2009), respectively. Under the assumption of the symmetric stepwise mutation model for the microsatellite loci, R_{ST} measures the same

quantity as F_{ST} defined by Hudson *et al*. (1992) for nucleotide sequences (Slatkin 1995). To evaluate the significance of differentiation, I used permutation tests for R_{ST} for the microsatellite data and S_{nn} (Hudson 2000) for the mitochondrial data. Significance levels for multiple testing were corrected using the Bonferroni procedure. Furthermore, I conducted the Mantel test of association (Mantel 1967) using GeneAlex to examine the relationship between linearized F_{ST} (or R_{ST}), $F_{ST}/(1-F_{ST})$ (Rousset 1997), and geographic distance. The geographic distance was measured as the shortest waterway distance between location pairs.

The basic genetic parameters of variation within species were calculated. For the mitochondrial data, I estimated nucleotide diversity π (Nei 1987), Watterson's estimator of the population mutation rate θ_W (Watterson 1975), and Tajima's *D* (Tajima 1989) using DNAsp 5.0 (Librado & Rozas 2009). I also estimated the parameters of the demographic expansion model of Schneider et al. (Schneider & Excoffier 1999) using their method, which was implemented in Arlequin. In this method, population size is assumed to increase quickly from N_0 to N_1 t generations ago, with the estimated parameters $\theta_0 = 2N_0u$, $\theta_1 = 2N_1u$ and $T_0 = 2ut$ (where u is the mutation rate). Goodness of fit for the model was evaluated using the estimated parameters. For the microsatellite data, I computed the expected heterozygosity (H_E) and Wright's inbreeding coefficient (F_{IS}) using Arlequin.

4. Results

In total, data for the 12 microsatellites and the mitochondrial control region from 906 and 748 individuals of cichlids, respectively, were used for the analysis.

4.1. Population structure inferred by STRUCTURE

To infer the population structure of the whole sample set, including the two pelagic species, I ran the STRUCTURE program (Pritchard *et al*. 2000) using data from the 12 microsatellite loci, assuming the number of populations (K) to range from 1 to 20. The estimated log probability of the data (ln P[D]) increased as K was incrementally raised from 1 to 4, stayed approximately constant until K reached 11, and then decreased rapidly as K was increased further (Fig. 2). The maximal ln P[D] value was reached when $K = 8$. The modal value for the index as defined by Evanno *et al*. (2005), ΔK, was reached when $K = 2$ (Fig. 3). Therefore, the number of populations at the uppermost hierarchical level of population structure appears to equal two (Evanno *et al*. 2005). However, further subdivisions were apparent as the number of populations was increased. Individual assignments are shown in Fig. 4 for $K = 2$, 3, and 4. With a few exceptions, individuals belonging to the same species were classified into the same emerging groups as the K value was increased. When $K = 2$, the rock-dwelling species *Haplochromis* (*Paralabidochromis*) *sauvagei* separated from the other eight species. When $K = 3$, the two pelagic species, *H.* (*Yssichromis*) *pyrrhocephalus* and *H.* (*Y.*) *laparogramma*, separated from the remaining six littoral species, although part of their genetic components were shared by some *Lithochromis*.

When $K = 4$, the six littoral species separated into two groups, one group consisting of the species belonging to the genus *Lithochromis*, *L. rubripinnis, L. rufus* and *L.* spp., and the other group consisting of four rock-dwelling species, *Neochromis rufocaudalis, N. greenwoodi, N. omnicaeruleus* and *Mbipia mbipi*. Because part of the species boundary became obscure when $K \geq 5$ (data not shown), I did not consider these cases any further. Thus, I restrict my attention to the four groups identified when $K = 4$ and refer to them as the following: (1) pelagic [*H.* (*Y.*) *pyrrhocephalus and H.* (*Y.*) *laparogramma*], (2) *Lithochromis* (*L. rubripinnis and L. rufus*), (3) rocky-shore 1 (*H.* (*P.*) *sauvagei*), and (4) rocky-shore 2 (*N. rufocaudalis, N. greenwoodi, N. omnicaeruleus* and *M. mbipi*). Although the specimens of *L. rufus* were mainly caught in vegetation zones, specimens of *L. rubripinnis* were found along rocky shores. Therefore, these two species could not be assigned to a single habitat. Thus, the nine studied species were genetically classified into four groups, three of which only contained species from a single habitat. Note that a small number of individuals classified as *Lithochromis* spp. were also grouped genetically into the *Lithochromis* group, although those classified as *Haplochromis* spp. could not be unambiguously assigned to any one group.

4.2. Haplotype network of the mitochondrial control region

The haplotype network was reconstructed using sequence data from the mitochondrial control region (Fig. 5). Ten species, including *Pundamilia macrocephala*, the *Lithochromis* spp. and the *Haplochromis* spp., are represented by 12 different colors in Fig. 5. I employed the haplotype designations defined by Verheyen et al. (2003) for

the previously characterized haplotypes. Many new haplotypes were identified in the present study, which have been numbered from k1 to k175.

The sequences from these nine species were not monophyletic, as has been previously noted for the cichlids of Lake Victoria by Verheyen et al. (2003). However, characteristic distributions of haplotypes can be observed for some of the species. For example, although the haplotypes of the two pelagic species (blue and light blue) were scattered throughout the haplotype network, the majority of them were concentrated around haplotype 77 and haplotype k1. On the other hand, haplotypes of the two *Lithochromis* species (red and orange) were mostly clustered around haplotypes 77, 92 and other closely related haplotypes. Another striking case was that of *H.* (*P.*) *sauvagei* (*H.* sp. "rockkribensis"), whose haplotypes (green) were generally located near haplotype 92 and were mostly species specific. On the other hand, the haplotypes of another rock-dwelling species, *N. rufocaudalis* (pink), were widely distributed throughout the network. Therefore, species differed in the distributions of their mitochondrial haplotypes throughout the haplotype network, and the haplotypes of *H.* (*Y.*) *laparogramma*, *H.* (*P.*) *sauvagei, L. rubripinnis* and *L. rufus* clustered in a similar manner to clusters observed with the STRUCTURE-based groupings based on the nuclear microsatellite markers.

4.3. AMOVA and analyses based on F_{ST} and R_{ST}

To evaluate genetic differentiation between species within groups, I carried out AMOVA within each group and estimated the variation both within and between species using the mitochondrial and microsatellite data. These analyses were carried

out for the pelagic, *Lithochromis* and rocky-shore 2 groups, which contained multiple species, each of which was sampled at multiple locations. These results are shown in Table 3. For all groups, the variation between species was significant. The relative variance components were 19.31~27.87% for mtDNA and 1.14~4.81% for microsatellites (all $P < 0.05$). Therefore, species appeared to be genetically divergent within each group. Indeed, pairwise F_{ST} values for the mitochondrial locus and R_{ST} values for the microsatellite loci between species were mostly significant, as shown in Table 4. Furthermore, the variation between populations within species was significant for all groups. Relative variance components were 5.56~31.52% for mtDNA and 0.33~3.32% for microsatellites (all $P < 0.01$).

I also estimated pairwise F_{ST} values for the mitochondrial locus and R_{ST} values for the microsatellite loci between populations within species, and these results are shown in Table 5. At the microsatellite loci, significant differentiation was found between one or more pairs of populations in only two species, *N. greenwoodi* and *L. rubripinnis*, and, in general, estimates of R_{ST} were lower than those for F_{ST} at the mitochondrial locus. For example, in *N. greenwoodi*, the Gabalema population was significantly differentiated, albeit weakly, from the other populations with respect to the microsatellite loci (R_{ST} = 0.065-0.148, P = 0.001-0.003). On the other hand, for the mitochondrial control region, I found significant differentiation between populations in all species, with the exception of *H. pyrrhocephalus*, *H. laparogramma* and *L. rufus*, and estimates of F_{ST} were generally high. For example, in *H.* (*P.*) *sauvagei*, seven of the 15 population pairs showed significant differentiation (F_{ST} = 0.278-0.756, P = 0.0000). For the *Lithochromis* and rocky-shore 2 species in which four or more populations were sampled, I plotted the linearized F_{ST} (or R_{ST}), $F_{ST}/(1-F_{ST})$, (shown in Fig. 6) and tested isolation by distance using the Mantel test. Isolation by distance

was found for *N. rufocaudalis* ($P = 0.043$) and *H.* (*P.*) *sauvagei* ($P = 0.051$) using the mitochondrial control region data, though it was not significant in the latter species. Weak isolation by distance was also observed for the mitochondrial region in *L. rubripinnis*, although this was not significant ($P = 0.066$).

4.4. Diversity statistics within species and inferences on expansion

I estimated various diversity statistics for each species, which are shown in Table 6. *N. rufocaudalis* had the highest nucleotide diversity in the mitochondrial control region ($\pi = 0.00481$), whereas *L. rufus* had the lowest diversity ($\pi = 0.00118$), although both species had high levels of diversity at the microsatellite loci. *H.* (*Y.*) *pyrrhocephalus* had the highest θ_W value (0.01142), whereas *H.* (*P.*) *sauvagei* had the lowest θ_W value (0.00328). F_{IS} was not significantly different from zero in any species after Bonferroni correction (data not shown). In all species, Tajima's *D* values were negative, and these results were significant for the two pelagic species, two *Lithochromis* species, *M. mbipi*, and *N. omnicaeruleus*.

As negative values for Tajima's *D* indicate recent demographic expansion, I used the method described by Schneider & Excoffier (1999) and implemented in Arlequin Version 3.5 (Excoffier & Lischer 2010) to estimate $\theta_0 = 2N_0u$, $\theta_1 = 2N_1u$ and $T_0 = 2ut$ using the mitochondrial data. The results are shown in Table 6. Assuming an evolutionary rate of 2.3×10^{-8} per year per base pair in this region, as was employed by Samonte *et al*. (2007), I estimated the absolute year of the start of expansion. For all species, current population size was estimated to be at least 30 times greater than the size of the population before expansion, and the start of expansion was estimated

to have occurred between 17,000 and 95,000 years ago. Except for *N. greenwoodi* and *H. sauvagei*, fits of the expansion model were generally good.

5. Discussion

5.1. Hierarchic genetic grouping

In the present thesis, I genetically examined nine cichlid species from Lake Victoria using mitochondrial and microsatellite markers to determine the genetic structure of cichlid populations during early adaptive speciation. More specifically, I asked whether a hierarchical genetic structure exists within cichlid fish populations in Lake Victoria, and if so, how is this structure related to habitat and species?

My analyses of microsatellite loci using STRUCTURE showed that the nine studied species could be genetically classified into four groups (Fig. 4): pelagic (*Haplochromis* (*Y.*) *pyrrhocephalus* and *H.* (*Y.*) *laparogramma*), *Lithochromis* (*L. rubripinnis* and *L. rufus*), rocky-shore 1 (*H.* (*P.*) *sauvagei*), and rocky-shore 2 (*Neochromis rufocaudalis*, *N. greenwoodi, N. omnicaeruleus* and *Mbipia mbipi*). Note that the rocky-shore species *L. rubripinnis* is included in the *Lithochromis* group but is not included in the rocky-shore group. In addition, two rocky-shore groups, the first consisting of *H.* (*P.*) *sauvagei* and the second consisting of the four remaining rocky-shore species, were differentiated genetically. Therefore, with the caveat that the present samples were limited to only nine of the approximately 700 species found within the lake, I conclude that a hierarchical genetic structure of species groups, species and populations exists in cichlid fish from Lake Victoria and that this structure is partially correlated with their respective habitats. Some of the groupings (e.g., pelagic) are consistent with the scenario proposed by Danley & Kocher (2001), which posits that adaptive radiation of cichlid fish in Lake Malawi occurred first by adaptation to different habitats. Alternatively, this pattern can be explained by the

higher level of gene flow between species that diverged earlier but still live together in the same habitat (Konijnendijk *et al*. 2011). In this case, speciation may not have occurred first by adaptation to different habitats. To discriminate between recent separation of populations and high levels of migration between them as a cause of the genetic similarity of the species in the same habitat, more detailed analyses such as those by IMa (Hey & Nielsen 2007) using multi-locus sequence data would be necessary.

Another notable feature of this grouping is that, with the exception of *Haplochromis*, individuals of a given genus were confined to individual groups. Moreover, because the pelagic group consists of only species belonging to the old genus *Yssichromis* and the rocky-shore 1 group consists of only one species belonging to the old genus *Paralabidochromis*, this grouping shows good agreement with the old classification of genera for cichlid fish in Lake Victoria by Greenwood (1981). This finding is consistent with the genus-level clustering of cichlids in Lake Victoria as shown by Bezault *et al*. (2011) using AFLP markers. Additionally, with the exception of a small number of individuals that includes those identified as either of the *Haplochromis* species, individuals belonging to the same species as judged from their morphology were classified into the same group as defined by the 12 microsatellite markers.

The differentiation of the rocky-shore 1 group from the other cichlids within Lake Victoria has been known for some time. This group consists of *H.* (*P.*) *sauvagei*, which was previously known as *H. sp.* "rockkribensis" (the species previously called *H. sauvagei* is now known as *H. fischeri* Seegers, 2008). Nagl *et al*. (2000) found that the mitochondrial haplotypes of this species belonged to subgroup VD, whereas the haplotypes of all the other cichlids in Lake Victoria belonged to subgroup VC. Moreover, Samonte *et al*. (2007) estimated that this species diverged from the other

cichlids in Lake Victoria approximately 41,300 years ago, whereas the other cichlids in the lake diverged from each other approximately 13,800 years ago, a period during which desiccation of the lake is thought to have occurred (Johnson *et al*. 1996). This result may indicate that the rocky-shore 1 group (*H.* (*P.*) *sauvagei*) has a unique evolutionary origin and that adaptation to the rocky-shore habitat during the early stage of the adaptive radiation in Lake Victoria might be represented by the rocky-shore 2 group.

I could also detect significant genetic differentiation between species within each group using AMOVA of the mitochondrial and microsatellite data, although the levels of differentiation differed between the markers and groups (Table 3). Nonetheless, the correct assignment of individuals to species groups using STRUCTURE analysis of the microsatellite data was not possible (data not shown). Although the locations of our sample collections were restricted to the Mwanza Gulf and the surrounding areas, sampling points for each species were scattered throughout the region, and different species from the same groups were occasionally sampled at the same location, as shown in Fig. 1. Therefore, differentiation between species beyond the differentiation observed between populations within species was observed at the neutral marker loci, although the levels of differentiation were generally very low at the microsatellite loci.

Our finding that significant genetic differentiation exists between species beyond what was observed between populations within species does not agree with the results of Samonte *et al*. (2007), who found that, with the exception of *H.* (*P.*) *sauvagei*, interpopulation genetic distances within species were similar to those observed between species in Lake Victoria. Additionally, Konijnendijk *et al*. (2011) have shown that allopatric conspecific populations were more strongly differentiated than

sympatric heterospecific populations of closely related species. Finally, Elmer *et al*. (2009) stated that current markers and methods were not sufficient to differentiate between biological species within Lake Victoria. Our contradictory findings might be explained by differences inherent to the species used in this study, differences in the marker type (e.g., microsatellites versus nuclear gene sequences in the case of Samonte *et al*. (2007)), numbers of markers (see Bezault *et al*. 2011, Wagner *et al*. 2013) or the differences in the geographic distances between surveyed populations. Indeed, the present population samples were generally separated by 30 km or less, whereas the populations used by Samonte *et al*. (2007) were separated by up to 350 km. Therefore, differentiation between populations may be underestimated in my study, as the samplings did not cover the entire range of each species. However, as the populations of each species were scattered throughout the studied region and were not concentrated geographically (Fig. 1). Moreover, although my samples were taken from restricted areas and only nine species out of approximately 700 species in Lake Victoria were studied here, the dominant haplotypes (haplotypes 77 and 92) in Fig. 5 corresponded with the two dominant haplotypes found by Verheyen et al. (2003) whose samples covered much wider areas of Lake Victoria (Verheyen *et al*. 2003, Nagl *et al*. 2000). Therefore, I think that differentiation in neutral marker loci between species beyond what is observed in populations is a real phenomenon in the species studied here. This finding is in agreement with a recent study by Wagner et al. (2013), who used RAD markers to show reciprocal monophyly of the species in Lake Victoria.

5.2. Population expansion

Estimates of Tajima's *D* values for the mitochondrial control region were negative for all species and highly significant in species belonging to the pelagic and *Lithochromis* groups (Table 6). The negative values of Tajima's *D* were caused by many low-frequency haplotypes that differed by one base pair from the major haplotypes (77, 92, 101 and k1 in Fig. 5). Because these results indicated recent population expansions, I estimated the time of the expansions using the method described by Schneider & Excoffier (1999). If I assume the mutation rate per base pair per year in the mitochondrial control region to be 2.3×10^{-8}, as used by Samonte *et al*. (2007) for the cichlids of Lake Victoria, the expansion time was estimated to be between 17,000 and 83,000 years ago for the species in the pelagic and *Lithochromis* groups, though we need to note that confidence intervals for the estimates from single locus data are usually very large. Also because the method (Schneider & Excoffier 1999) assumes a random mating population, which was violated in some of the studied species as shown in Fig. 6, some of the estimates might not be reliable.

Based on the microsatellite data, Elmer et al. (2009) inferred that cichlid populations in Lake Victoria began to decline approximately 18,000 years ago, and they suggested that this decline corresponded to the desiccation of Lake Victoria hypothesized by Johnson et al. (1996). This decline may in fact correspond with the beginning of the expansion previously estimated using mitochondrial markers (Maeda *et al*. 2009, Mzighani *et al*. 2010). First, my estimate of an expansion occurring between 17,000 and 83,000 years ago might be an overestimate due to an acceleration of evolutionary rates during more recent periods, possibly due to inclusion of deleterious mutations, as has been previously suggested by Ho *et al*. (2005) and Genner *et al*. (2007). Therefore, the beginning of the expansion could be closer to the estimate of 18,000

years ago proposed by Elmer *et al*. (2009). Second, a bottleneck event produces different patterns in neutrality-test statistics for mitochondrial genes compared with nuclear genes (Fay & Wu 1999). More specifically, values of Tajima's *D* for mitochondrial genes quickly become negative following a bottleneck event, whereas values for nuclear genes stay positive for some time. Therefore, for a short period after the bottleneck event, mitochondrial genes may indicate a population expansion, whereas nuclear genes may indicate a population decline. Therefore, my assessment of population expansion may indeed be consistent with the findings of Elmer *et al*. (2009).

5.3. Population structure within species

In the mitochondrial control region, I observed significant differentiation between the populations of most species, although the levels of differentiation differed between species. Strong differentiation was observed in *H.* (*P.*) *sauvagei* and *N. rufocaudalis*, showing isolation by distance, whereas differentiation in *L. rufus* and *N. omnicaeruleus* was weak (Fig. 6). Furthermore, the levels of differentiation differed even among species within the same group (e.g., *N. rufocaudalis* and *N. omnicaeruleus*). In contrast, although differentiation between populations was significant at the microsatellite loci in the AMOVA analysis, most of the pairwise R_{ST} values between populations were not significant. This indicated that the levels of differentiation at those loci were very low and could be detected only when a large number of samples were analyzed together.

In Lake Malawi, although three pelagic species show little differentiation between populations separated by more than 100 km (Shaw *et al*. 2000), rock-dwelling mbuna and non-mbuna species show much stronger differentiation (Pereyra *et al*. 2004). On the other hand, some species inhabiting rocky shores in Lake Victoria showed very weak differentiation between populations (*L. rubripinnis* and *N. omnicaeruleus*) in my study. Other authors have compiled similar results (Seehausen *et al*. 2008, Magalhaes *et al*. 2010, Magalhaes *et al*. 2013), although I do note that our samples were collected from an area approximately 30 km in diameter. The low levels of differentiation observed between populations of species living along rocky shores may indicate high mobility for those species. As species evolve lower mobility within this habitat, species may accumulate higher levels of differentiation between populations. On the other hand, these populations may show low differentiation due to the relatively recent dispersal of the species. Either way, the weak differentiation between populations of certain rocky-shore species stands in contrast to the strong differentiation found in rocky-shore species in other lakes (Pereyra *et al*. 2004) and may indicate a recent diversification of cichlids in Lake Victoria (Seehausen 2006, Verheyen *et al*. 2003).

6. Conclusions

Based on the observation of low levels of differentiation and an overlap between mitochondrial and nuclear haplotypes (Nagl *et al*. 1998, Nagl *et al*. 2000), the cichlid species of Lake Victoria have often been treated as genetically homogenous (*e.g.*, Elmer *et al*. 2009). However, as shown here, a clear hierarchical genetic structure can be seen in the cichlid fishes of Lake Victoria. Interestingly, the groupings were mostly consistent with the genus-level clustering, and some of the groups corresponded to different habitats. The habitat clustering found in some groups may be explained by the scenario proposed by Danley *et al*. (2001) for Lake Malawi in which species first diverge based on habitat, though recent gene flow between species in the same habitat can also explain the hierarchical structure. In addition, most species appear to be differentiated within each group, as has also been recently shown by Bezault *et al*. (2011) using AFLP markers and by Wagner *et al*. (2013) and Keller *et al*. (2012) using RAD markers. Finally, each species showed its own characteristic genetic structure, with either high or low levels of population differentiation. As this radiation process occurred recently, I was able to study this process more accurately than was possible with older radiations. Therefore, the cichlid fish of Lake Victoria provide a good opportunity to study adaptive radiation. Future studies that use larger numbers of nuclear markers will help us understand this process in greater detail.

Acknowledgements

I would like to express special thanks to Professor Hidenori Tachida, my supervisor, for his many valiant advices, continual supprt and warm encouragement during my study. My thanks go as well to Dr. Nobuyuki Inomata, Dr. Junko Kusumi, Dr. Midori Kobayakawa, Dr. Koichi Kamiya, Dr. Kosuke M. Teshima, Dr. Takanori Yoshida, Dr. Nakatada Wachi, Dr. Miho Tamura, Dr. Tomotaka Matsumoto, Dr. Alfred E. Szmidt and Professor Yuji Oshima for many helpful comments, sound advices and encouragement on this study.

I would also like to thank all the past and present members of our laboratory for their kind advices, many support, encoraging words and valuable discussion. I would especially like to thank Ms. Kaoru Maeda for teaching the experimental technique and giving knowledge about previous works of the cichlid fishes. I am also grateful to Dr. Koichi Kamiya for teaching experimental skills. I thank Dr. Tomotaka Matsumoto for his help in extracting DNA.

I thank Tanzania Fisheries Research Institute (TAFIRI) for collaboration in the field work and for providing us equipments and laboratory spaces. I would also like to express thanks to Professor Norihiro Okada and his laboratory members for collaboration. I thank Professor Norihiro Okada for offering valuable samples of cichlid fishes and many information about their field study with good grace. I also thank Dr. Kazuhiko Takahashi for helping me in the collection of the samples. I am also grateful to Dr. Yohey Terai for giving many useful advices, teaching me DNA extraction skills and giving continual encouragement. I appreciate Dr. Shinji Mizoiri for giving me ecological information about samples. My thanks go as well to Dr. Mitsuto Aibara for giving me many advices about taxonomy. I thank Professor Tetsu Sato for providing me many papers about ciclid fishes that were difficult to obtain by

myself. I would also like to thank Dr. Semvua I. Mzighani for instructing me detailed knowledge about genetic structure of pelagic cichlid species.

This study was partially supported by the Global COE Program (Center of Excellence for Asian Conservation Ecology as a Basis of Human-Nature Mutualism), Ministry of Education, Culture, Sports, Science and Technology, Japan.

Finally, I would like to express my gratitude to my parents, Shinji and Junko for their understanding, generous support and warm encouragement. I thank Dr. Tsuyoshi Inoshita for many discussion and advices.

References

Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ, Kocher TD (1999) Population structure and colour variation of the cichlid fish Labeotropheus fuelleborni Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. Proc R Soc Lond B 266: 119-130.

Barel CDN, van Oijen MJP, Witte F, Witte-Maas, ELM (1977) An introduction to the taxonomy and morphology of the haplochromine Cichlidae from Lake Victoria. Neth J Zool 27: 333-389.

Barlow GW (2002) The cichlid fishes-Nature's grand experiment in evolution. Perseus publishing.

Bezault E, Mwaiko S, Seehausen O (2011) Population genomic tests of models of adaptive radiation in Lake Victoria region cichlid fish. Evolution 65: 3318–3397.

Clement M, Posada D, Crandall KA: TCS (2000) a computer program to estimate gene genealogies. Mo Ecol 9: 1657-1659.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Inc.

Danley PD, Kocher TD (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. Mol Ecol 10: 1075–1086.

Elmer KR, Reggio C, Wirth T, Verheyen E, Salzburger W, Meyer A (2009) Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. Proc Natl Acad Sci U S A 106: 13404-13409.

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.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479-491.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Eco Res 10: 564-567.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Fay JC, Wu C-I (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol 16: 1003-1005.

Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF (2007) Age of cichlids: new dates for ancient lake fish radiations. Mol Biol Evol 24: 1269–1282.

Greenwood PH (1957) A revision of the Lake Victoria Haplochromis species (Pisces, Cichlidae). Part II: *H. sauvagei* (Pfeffer)., *H. prodromus* Trewavas, *H. granti* Blgr., and *H. xenognathus*, sp. n. Bull Br Mus Nat Hist (Zool) 5: 76-97.

Greenwood PH (1974) The cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock. Bull Br Mus Nat Hist. Suppl 6: 1–34.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl Acad Sci U S A 104: 2785-2790.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 22: 1561-1568.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132: 583-589.

Hudson RR (2000) A new statistic for detecting genetic differentiation. Genetics 155: 2011-2014.

Johnson TC, Scholz CA, Talbot MR, Kelts K, Ricketts RD (1996) Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. Science 273: 1091-1093.

Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A, Wittwer S, Seehausen O (2012) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Mol Ecol. Advance Access publication | DOI: 10.1111/mec.12083.

Kocher TD (2004) Adaptive evolution and explosive speciation: The cichlid fish model. Nat Rev Genet 5: 288-298.

Konijnendijk N, Joyce DA, Mrosso HDJ, Egas M, Seehausen O (2011) Community genetics reveal elevated levels of sympatric gene flow among morphologically similar but not among morphologically dissimilar species of Lake Victoria cichlid fish. Int J Evol Biol Article ID 616320. Available at http://www.hindawi.com/journals/ijeb/2011/616320/. Accessed 6 May 2013

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

Maeda K., Takeshima H, Mizoiri S, Okada N, Nishida M, Tachida H (2008) Isolation and characterization of microsatellite loci in the cichlid fish in Lake Victoria, *Haplochromis chilotes*. Mol Ecol Res 8: 428–430.

Maeda K, Takeda M, Kamiya K, Aibara M, Mzighani SI, Nishida M, Mizoiri S, Sato S, Terai Y, Okada N, Tachida H (2009) Population structure of two closely related pelagic cichlids in Lake Victoria, *Haplochromis pyrrhocephalus* and *H. laparogramma*. Gene 441: 67–73.

Magalhaes IS, Mwaiko S, Seehausen O (2010) Sympatric colour polymorphisms associated with nonrandom gene flow in cichlid fish of Lake Victoria. Mol Ecol 19: 3285-3300

Magalhaes IS, Lundsgaard-Hansen B, Mwaiko S, Seehausen O (2013) Evolutionary divergence in replicate pairs of ecotypes of Lake Victoria cichlid fish. Evol Eco Res 14: 381-401

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27: 209-220.

Markert JA, Arnegard ME, Danley PD, Kocher TD (1999) Biogeography and population genetics of the Lake Malawi cichlid *Melanochromis auratus*: habitat transience, philopatry and speciation. Mol Ecol 8: 1013-1026.

Mzighani SI, M. Nikaido M, Takeda M, Seehausen O, Budeba YL, Ngatunga BP, Katunzi EFB, Aibara M, Mizoiri S, Sato T, Tachida H, Okada N (2010) Genetic variation and demographic history of the Haplochromis laparogramma group of Lake Victoria—An analysis based on SINEs and mitochondrial DNA. Gene 450: 39–47.

Nagl S, Tichy H, Mayer WE, Takahata T, Klein J (1998) Persistence of neutral polymorphisms in Lake Victoria cichlid fish. Proc Natl Acad Sci U S A 95: 14238-14243.

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

Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press.

Oppen MJ van, Turner GF, Rico C, Deutsch JC, Ibrahim KM, Robinson RL, Hewitt GM (1997) Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. Proc R Soc Lond B 264: 1803-1812.

Peakall R, Smouse PE (2006) GENALEX6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288-295.

Pereyra R, Taylor MI, Turner GF, Rico C (2004) Variation in habitat preference eand population structure among three species of the Lake Malawi cichlid genus Protomelas. Mol Ecol 13: 2691-2697.

Pfeffer GJ (1896) Die Fische Ost-Afrikas. In: Mæbius F, editor. Deutsch-Ost-Afrika, Band III. Die Thierwelt Ost-Afrikas und der Nachbargebiete. Wirbelthiere. 4. (Separatum: 1 -72) pp. i -xviii + 1-72.

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

Rambaut A (1996) Se-Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk/.

Rico C, Turner GF (2002) Extreme microallopatric divergence in a cichlid species from Lake Malawi. Mol Ecol 11: 1585-1590.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145: 1219-1228.

Salzburger W, Meyer A (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. Naturwissenschaften 91: 277-290.

Samonte IE, Satta Y, Sato A, Tichy H, Takahata N, Klein J (2007) Gene flow between species of Lake Victoria haplochromine fishes. Mol Biol Evol 24: 2069-2080.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. Genetics 152: 1079-1089.

Seegers L (2008) The fishes collected by G. A. Fischer in East Africa in 1883 and 1885/86. Zoosyst Evol: 149-195.

Seehausen O (1996) Lake Victoria rock cichlids: taxonomy, ecology, and distribution. Zevenhuizen, The Netherlands: Verduijn Cichlids.

Seehausen O, Lippitsch E,N. Bouton N, Zwennes H (1998) Mbipi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). Icthyol Explor Fres 9: 129-228.

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

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. Proc R Soc Lond B Biol Sci 273: 1987–1998.

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HDJ, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H, Imai H, Okada N (2008) Speciation through sensory drive in cichlid fish. Nature 455: 620–626.

Shaw PW, Turner GF, Idid MR, Robinson RL, Carvalho GR (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. Proc R Soc Lond B Biol Sci 267: 2273-2280.

Slatkin M (1995) A Measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-462.

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

Terai Y, Seehausen O, Sasaki T, Takahashi K, Mizoiri S, Sugawara T, Sato T, Watanabe M, Konijnendijk N, Mrosso HDJ, Tachida H, Imai H, Shichida Y, Okada N (2006) Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. PLoS Biol 4: e433.

Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL (2001) How many species of cichlid fishes are there in African lakes? Mol Ecol 10: 793–806.

Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. Science 300: 325-328.

Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol Ecol 22: 787-798.

Watterson GA (1975) On the number of segregation sites. Theor Popul Biol 7: 256-276.

Figure legends

Figure 1. Sampling locations of the cichlids. Different species are represented by different symbols as indicated in Panel B. Panel A: Lake Victoria. Panel B: the sourhtern part of Lake Victoria. Panel C: Mwanza Gulf.

Figure 2. Mean Ln P(D) over 20 runs for each K value.

Figure 3. Δ*K* (Evanno *et al*. 2005) for each K value.

Figure 4. Results of STRUCTURE analyses of the entire sample set with *K* = 2-4. The grouping of the species is shown at the bottom. Speccies delimitation is indicated by the vertical bars above the species names.

Figure 5. Haplotype network of the mitochondrial control region. Different species are represented by different colors. *H. (Y.) pyrrhocephalus* and *H. (Y.) laparogramma* are pelagic. *L. rufus* lives in the vegetation zone and the remainig species live in rocky shores. The size of the the circle shows the number of the samples having the haplotype.

Figure 6. The relationships between geographical distance and genetic differentiation. Pane A: mitochondrial. Panel B: nuclear microsatellite loci.

Table 1. Numbers of individuals typed and sequenced in each species.

	Habitat[a]	Microsatellite loci	mtDNA control region
H. (*Y.*) *pyrrhocephalus*	P	289	166
H. (*Y.*) *laparogramma*	P	89	36
L. rubripinnis	R	61	66
L. rufus	V	112	128
M. mbipi	R	13	13
N. rufocaudalis	R	81	81
N. greenwoodi	R	77	77
N. omnicaeruleus	R	46	46
H. (*P.*) *sauvagei*	R	103	103
Haplochromis spp.	R	25	22
Lithochromis spp.	R	5	5
P. macrocephala	R	5	5
total		906	748

[a] P: pelagic, V: vegetation zone, R: rocky-shore

Table 2. Locations and numbers of samples for each species.

Location		Number of specimens								
		H. (Y.) pyrrhocephalus	*H. (Y.) laparogramma*	*L. rubripinnis*	*L. rufus*	*M. mbipi*	*N. greenwoodi*	*N. rufocaudalis*	*N. omnicaeruleus*	*H. (P.) sauvagei*
Speke Gulf										
	Kisimani	39 / 20								
	Pogopogo Speke Gulf	24 / 24								
	Mwabulugu	49 / 25								
	Nyamikoma	15 / 8								
	Namatembi Island					12 / 12				
	Bwiru Island								5 / 5	
	Mabibi Island								9 / 9	
	Ruti Island								12 / 12	
	Between Igombe and Ruti	8 / 8	32 / 20							
	Igombe fishing village		46 / 15							
	Makobe Island					1/1		18 / 18	20 / 20	14 / 14
	Juma Island									4 / 4
Mwanza Gulf										
	Bwiru Peninsula									23 / 23
	Bwiru point-Hippo	10 / 10	11/1							
	Hippo Island			5 / 5						
	Gabalema Island				3/0		20 / 20	8 / 8		
	East offshore of Kissenda	40 / 20								
	Kissenda Island			14 / 14			17 / 17	20 / 20		25 / 25
	Chankende Island				23 / 23					
	Nyegezi Rocks			3 / 3						
	Nyaruwambu				20/ 20					
	Kilimo Island	4 / 4						14 / 14		2 / 2
	Northern offshore Nyameruguyu	19 / 19								
	Kijiwe Landing Site	67 / 14								
	Offshore Nyameruguyu	14 / 14								
	Nyamatala Island			9 / 9	3/6			20 / 20		35 / 35
	Northend of Luanso Bay				7/ 7					
	Shadi Rocks			0'/2	0'/9					
	Luanso Island			4 / 6						
	Froating island				0'/9					
	Luanso 2 Island				3 / 3					
	Matumbi Island			15 / 15	4/ 4		20 / 20			
	Southend of Luanso Bay				4/ 4					
	Marumbi Island			11 / 11	43 / 43		20 / 20			

Table 3. Results of AMOVA at mitochondrial and nuclear loci.

	pelagic			*Lithochromis*			rocky-shore 2		
	df	% variation	P value	df	% variation	P value	df	% variation	P value
mtDNA									
between species	1	22.51	0.021	1	27.87	0.000	3	19.31	0.006
between populations	12	21.57	0.000	17	5.56	0.000	16	31.52	0.000
within populations	188	55.92		175	66.58		287	49.17	
microsatellite									
between species	1	1.14	0.005	1	1.15	0.001	3	4.81	0.000
between populations	12	0.33	0.001	14	1.08	0.000	15	3.32	0.000
between individuals	364	2.35	0.000	155	1.79	0.011	287	1.45	0.016
Within individuals	378	96.18		171	95.99		306	90.42	

Table 4. F_{ST} estimated by mtDNA (below diagonal) and R_{ST} estimated by microsatellites (above diagonal) between species.

		H. pyrrhocephalus	*H. laparogramma*	*L. rubripinnis*	*L. rufus*	*M. mbipi*	*N. rufocaudalis*	*N. greenwoodi*	*N. omnicaeruleus*	*H. sauvagei*
	n	289	89	61	112	13	81	77	46	103
H. pyrrhocephalus	166	------	0.012*	0.116*	0.036*	0.170*	0.127*	0.122*	0.157*	0.216*
H. laparogramma	36	0.305**	------	0.101*	0.036*	0.126*	0.107*	0.119*	0.116*	0.183*
L. rubripinnis	66	0.243**	0.567**	------	0.035*	0.063ns	0.036*	0.037*	0.086*	0.149*
L. rufus	128	0.139**	0.520**	0.278**	------	0.029ns	0.027*	0.018ns	0.043*	0.085*
M. mbipi	13	0.404**	0.643**	0.470ns	0.570**	------	0.045ns	0.041ns	0.000ns	0.090ns
N. rufocaudalis	81	0.235**	0.414**	0.301**	0.246**	0.472**	------	0.039*	0.075*	0.140*
N. greenwoodi	77	0.174**	0.481**	0.048**	0.135**	0.434ns	0.186**	------	0.068*	0.110*
N. omnicaeruleus	46	0.255**	0.517**	0.060**	0.295**	0.424ns	0.259**	0.065**	------	0.110*
H. sauvagei	103	0.323**	0.587**	0.299**	0.449**	0.282**	0.355**	0.278**	0.269**	------

ns not significant

* significant at 5%

** sinignificant at 1%

Table 5. F_{ST} estimated by mtDNA (below diagonal) and R_{ST} estimated by microsatellites (above diagonal) between populations within species.

A. *H.* (*Y.*) *pyrrhocaphalus*.

		Offshore N	Kijiwe Lan	Northern c	Kilimo Islai	East offsh	Bwiru poin	Between I	Nyamikom	Kisimani	Pogopogo	Mwaburu
		14	67	19	4	40	10	8	15	39	24	49
Offshore Nyameruguyu	14	------	0.021ns	0.000ns	0.000ns	0.025ns	0.041ns	0.000ns	0.003ns	0.020ns	0.085ns	0.008ns
Kijiwe Landing site	14	0.317ns	------	0.010ns	0.023ns	0.000ns	0.000ns	0.000ns	0.004ns	0.019ns	0.128ns	0.005ns
Northern offshore Nyameruguyu	19	-0.011ns	0.359ns	------	0.052ns	0.007ns	0.037ns	0.000ns	0.003ns	0.018ns	0.097ns	0.010ns
Kilimo Island	4	0.381ns	0.065ns	0.409ns	------	0.011ns	0.019ns	0.000ns	0.000ns	0.000ns	0.075ns	0.000ns
East offshore of Kissenda	20	0.319ns	-0.004ns	0.366ns	0.139ns	------	0.000ns	0.000ns	0.000ns	0.011ns	0.112ns	0.015ns
Bwiru point-Hippo	10	0.323ns	-0.004ns	0.360ns	0.185ns	-0.034ns	------	0.000ns	0.000ns	0.000ns	0.141ns	0.000ns
Between Igombe & Ruti	8	0.315ns	-0.030ns	0.356ns	-0.005ns	-0.004ns	0.029ns	------	0.000ns	0.000ns	0.095ns	0.000ns
Nyamikoma	8	0.450ns	0.091ns	0.482ns	-0.064ns	0.179ns	0.240ns	0.009ns	------	0.000ns	0.072ns	0.000ns
Kisimani	20	0.462ns	0.088ns	0.511ns	0.026ns	0.154ns	0.225ns	-0.005ns	0.027ns	------	0.041ns	0.007ns
Pogopogo Speke Gulf	24	0.026ns	0.388ns	0.031ns	0.383ns	0.408ns	0.417ns	0.362ns	0.471ns	0.504ns	------	0.063ns
Mwaburugu	25	0.448ns	0.080ns	0.494ns	0.004ns	0.155ns	0.223ns	-0.007ns	0.012ns	-0.020ns	0.483ns	------

B.*H.* (*Y.*) *laparogramma*.

		Igombe fis	Between I
		46	32
Igombe fishing village	15	------	0.000ns
Between Igombe & Ruti	20	0.012ns	------

C. *L. rubripinnis*.

		Marumbi I	Matumbi I	Luanso Isl	Nyamatala	Nyegezi R	Kissenda I	Hippo Islar
	n	11	15	4	9	3	14	5
Marumbi Island	11	------	0.000ns	0.045ns	0.008ns	0.059ns	0.000ns	0.022ns
Matumbi Island	15	0.005ns	------	0.090ns	0.057ns	0.103ns	0.021ns	0.053ns
Luanso Island	6	0.027ns	-0.021ns	------	0.240*	0.283ns	0.000ns	0.180ns
Nyamatala Islands	9	0.005ns	0.065ns	0.063ns	------	0.002ns	0.064ns	0.000ns
Nyegezi Rocks	3	0.680ns	0.048*	0.133ns	0.286ns	------	0.060ns	0.000ns
Kissenda Island	14	0.023ns	0.077ns	0.101ns	-0.043ns	0.432ns	------	0.039ns
Hippo Island	5	0.306*	0.381**	0.351ns	0.085ns	0.000ns	0.168ns	------

D. *L. rufus*.

		Marumbi I	South end	Matumbi	Luanso 2 I	North end	Nyaruwan	Chankend	Nyamatala
		43	4	4	3	7	20	23	3
Marumbi Island	43	------	0.275ns	0.000ns	0.002ns	0.000ns	0.004ns	0.007ns	0.000ns
South end of Luanso Bay	4	-0.057ns	------	0.000ns	0.000ns	0.024ns	0.189ns	0.195ns	0.000ns
Matumbi Island	4	-0.022ns	0.000ns	------	0.092ns	0.000ns	0.043ns	0.031ns	0.012ns
Luanso 2 Island	3	-0.024ns	0.000ns	-0.250ns	------	0.000ns	0.002ns	0.000ns	0.000ns
North end of Luanso Bay	7	0.019ns	0.033ns	-0.233ns	-0.071ns	------	0.000ns	0.000ns	0.000ns
Nyaruwambu	20	-0.009ns	-0.026ns	-0.048ns	-0.036ns	0.005ns	------	0.000ns	0.000ns
Chankende Island	23	0.039ns	0.021ns	-0.067ns	-0.142ns	0.040ns	0.010ns	------	0.000ns
Nyamatala Islands	6	-0.008ns	0.000ns	-0.143ns	0.000ns	-0.095ns	-0.010ns	0.084ns	------

E. *N. greenwoodi*.

		Marumbi I	Matumbi I	Kissenda	Gabalema
		20	20	17	20
Marumbi Island	20	------	0.011ns	0.035ns	0.093*
Matumbi Island	20	0.277***	------	0.037ns	0.148*
Kissenda	17	0.266***	0.073ns	------	0.065*
Gabalema Islands	20	0.356***	0.096***	0.095***	------

F. *N. rufocaudalis*.

		Nyamatala	Kilimo	Kissenda	Gabalema	Makobe Isl
		20	14	20	8	18
Nyamatala Islands	20	------	0.043ns	0.028ns	0.000ns	0.014ns
Kilimo	14	-0.012ns	------	0.000ns	0.000ns	0.028ns
Kissenda	20	0.010*	0.021*	------	0.000ns	0.026ns
Gabalema Islands	8	0.253ns	0.168ns	0.336ns	------	0.000ns
Makobe Island	18	0.558***	0.500***	0.572***	0.378***	------

G. *N. omnicaeruleus* .

		Makobe Is	Ruti Island	Mabibi Isla	Bwiru Islan
		20	12	9	5
Makobe Island	20	------	0.030ns	0.004ns	0.047ns
Ruti Island	12	0.014ns	------	0.000ns	0.022ns
Mabibi Island	9	0.059***	0.023ns	------	0.037ns
Bwiru Island	5	0.209ns	0.254*	0.028ns	------

H. *H. (P.) sauvage* i

		Nyamatala	Kilimo Islai	Kissenda I:	Bwiru Poin	Makobe Is	Juma Islan
		35	2	25	23	14	4
Nyamatala Islands	35	------	0.169ns	0.003ns	0.023ns	0.030ns	0.000ns
Kilimo Island	2	0.127ns	------	0.082ns	0.287ns	0.365ns	0.175ns
Kissenda Island	25	0.278**	0.055ns	------	0.040ns	0.051ns	0.000ns
Bwiru Point	23	0.735**	0.019ns	0.680**	------	0.024ns	0.000ns
Makobe Island	14	0.756**	0.109ns	0.704**	-0.004ns	------	0.007ns
Juma Island	4	0.717**	0.241ns	0.667**	0.508ns	0.481ns	------

ns not significant
* significant at 5%
** significant at 1%

Table 6. Statistics of population diversity and estimates for population size change for the 9 species.

	Hp[a]	Hl[a]	Lrub[a]	Lruf[a]	Mm[a]	Nr[a]	Ng[a]	No[a]	Hs[a]
mtDNA									
n	166	36	66	128	13	81	77	46	103
S	57	29	18	24	10	28	23	23	15
π	0.00266	0.00226	0.00145	0.00118	0.00221	0.00481	0.00271	0.00263	0.00236
θ_W	0.01165	0.00823	0.00430	0.00504	0.00367	0.00642	0.00533	0.00595	0.00328
Tajima's D	-2.412**	-2.504***	-2.071*	-2.317**	-1.834*	-0.862ns	-1.640ns	-1.903*	-0.905ns
θ_0[b]	0.353	0.005	0.030	0.000	0.000	0.005	0.012	0.000	0.000
θ_1[b]	17.412	∞	7.437	10.833	∞	7.668	∞	∞	6.455
T_0 (=2ut)	3.342	2.199	1.535	0.719	2.295	6.207	2.277	2.250	3.840
time (years)[c]	82279	54139	37791	17702	56502	152814	56059	55394	94539
goodness of fit[d]	0.889	0.633	0.997	0.137	0.453	00.489	0.014	0.546	0.067
microsatellite									
n	289	89	61	112	13	81	77	46	103

heterozygosity	0.750	0.745	0.731	0.779	0.744	0.741	0.729	0.768	0.697

[a] Hp, *H*. (*Y*.) *pyrrhocephalus*: Hl, *H*. (*Y*.) *laparogramma*: Lrub, *L*. *rubripinnis*: Lruf, *L*. *rufus*: Mm, *M*. *mbipi*: Nr, *N*. *rufocaudalis*: Ng, *N*. *greenwoodi*: No, *N*. *omnicaeruleus*: Hs, *H*. (*P*.) *sauvagei*

[b] Parameters of the model by Schneider and Excoffier (1999).

[c] $u = 2.3 \times 10\text{-}8$ per year per base pair was assumed.

[d] Results of goodness of fit for the predicted expansion model.

* significant at 5%, ** significant at 1%. *** significant at 0.1%

Figure 1

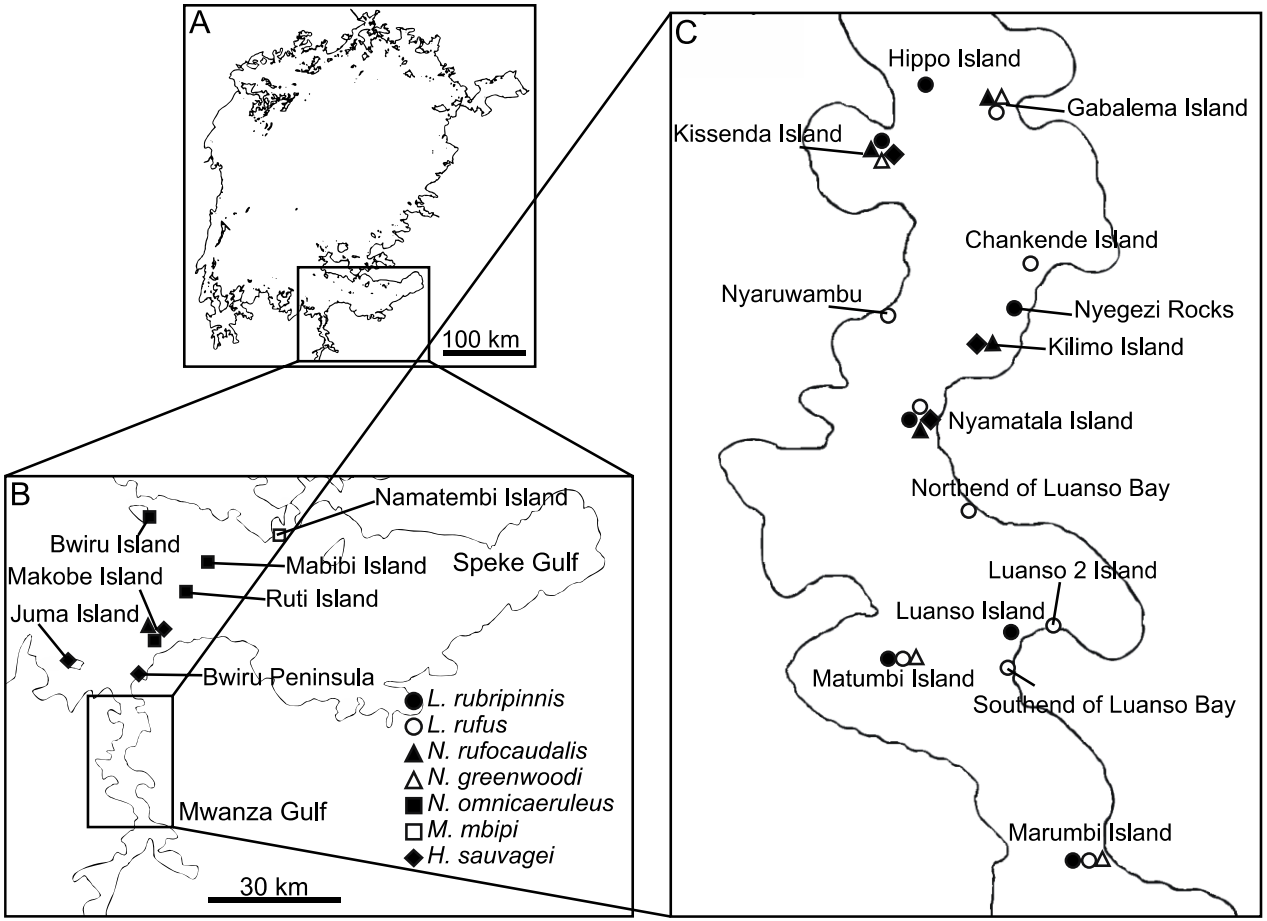

A
100 km
B
Namatembi Island
Bwiru Island
Mabibi Island
Speke Gulf
Makobe Island
Ruti Island
Juma Island
Bwiru Peninsula
Mwanza Gulf
30 km
L. rubripinnis
L. rufus
N. rufocaudalis
N. greenwoodi
N. omnicaeruleus
M. mbipi
H. sauvagei
C
Hippo Island
Gabalema Island
Kissenda Island
Chankende Island
Nyaruwambu
Nyegezi Rocks
Kilimo Island
Nyamatala Island
Northend of Luanso Bay
Luanso 2 Island
Luanso Island
Matumbi Island
Southend of Luanso Bay
Marumbi Island

Figure 2

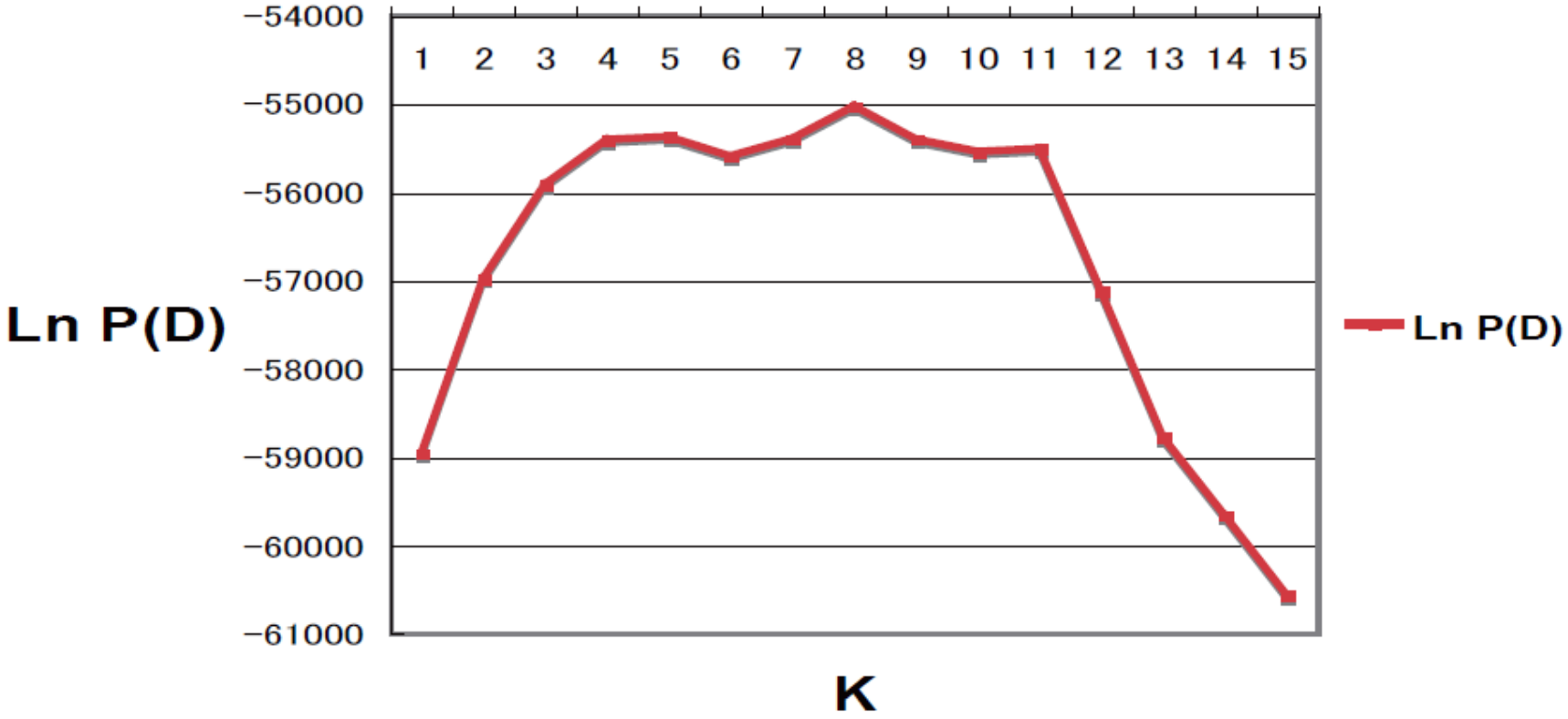

−54000
−55000
−56000
−57000
−58000
−59000
−60000
−61000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Ln P(D)
K
Ln P(D)

Figure 3

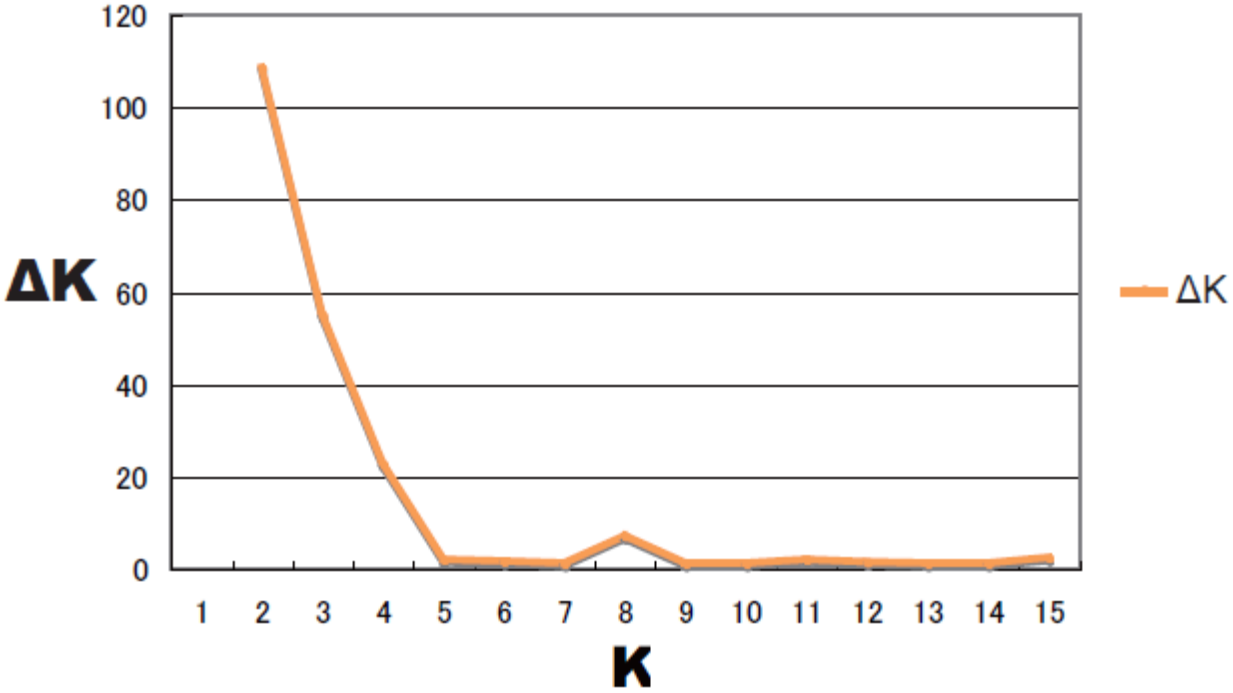


Figure 4

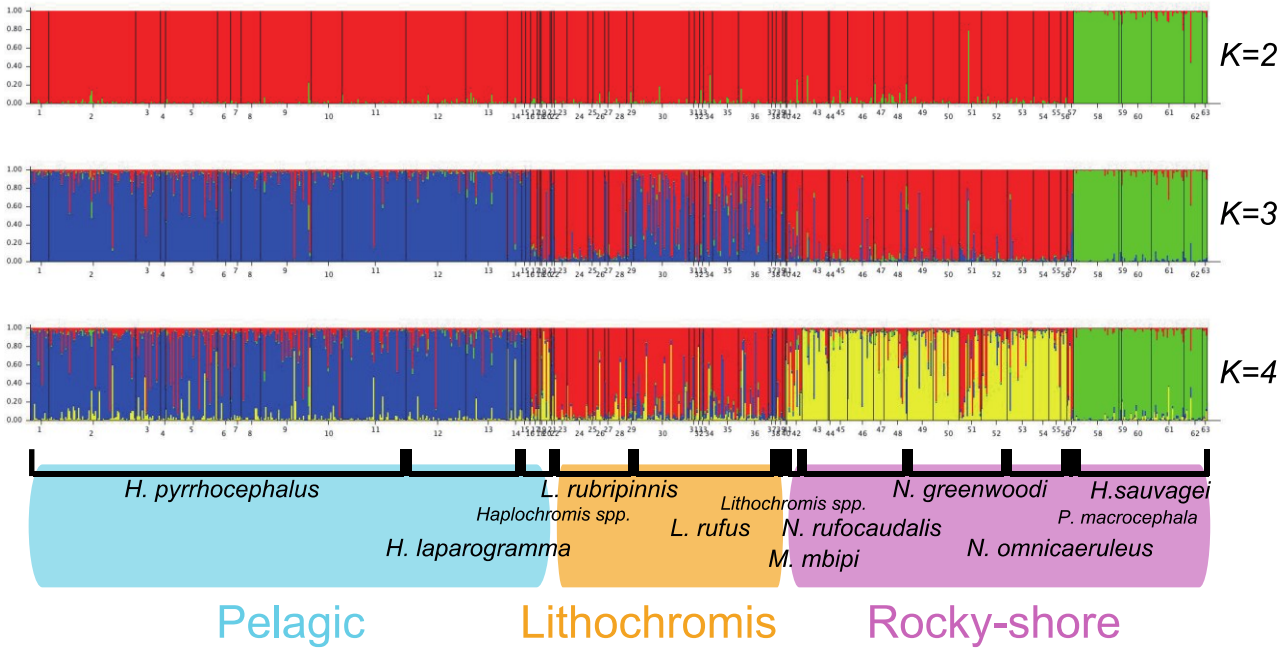

K=2
K=3
K=4
H. pyrrhocephalus
Haplochromis spp.
H. laparogramma
L. rubripinnis
L. rufus
Lithochromis spp.
M. mbipi
N. greenwoodi
N. rufocaudalis
N. omnicaeruleus
H.sauvagei
P. macrocephala
Pelagic
Lithochromis
Rocky-shore

Figure 5

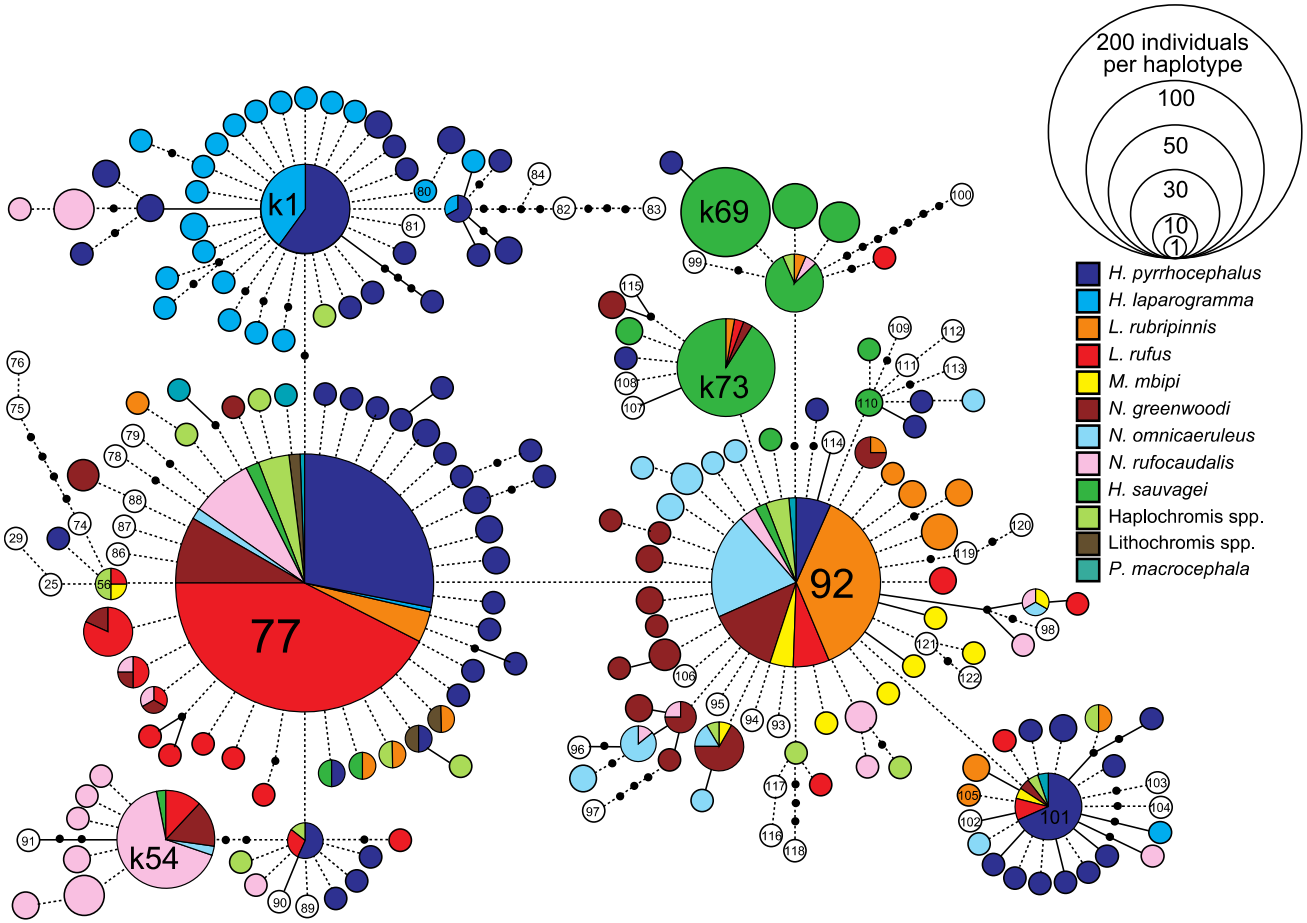
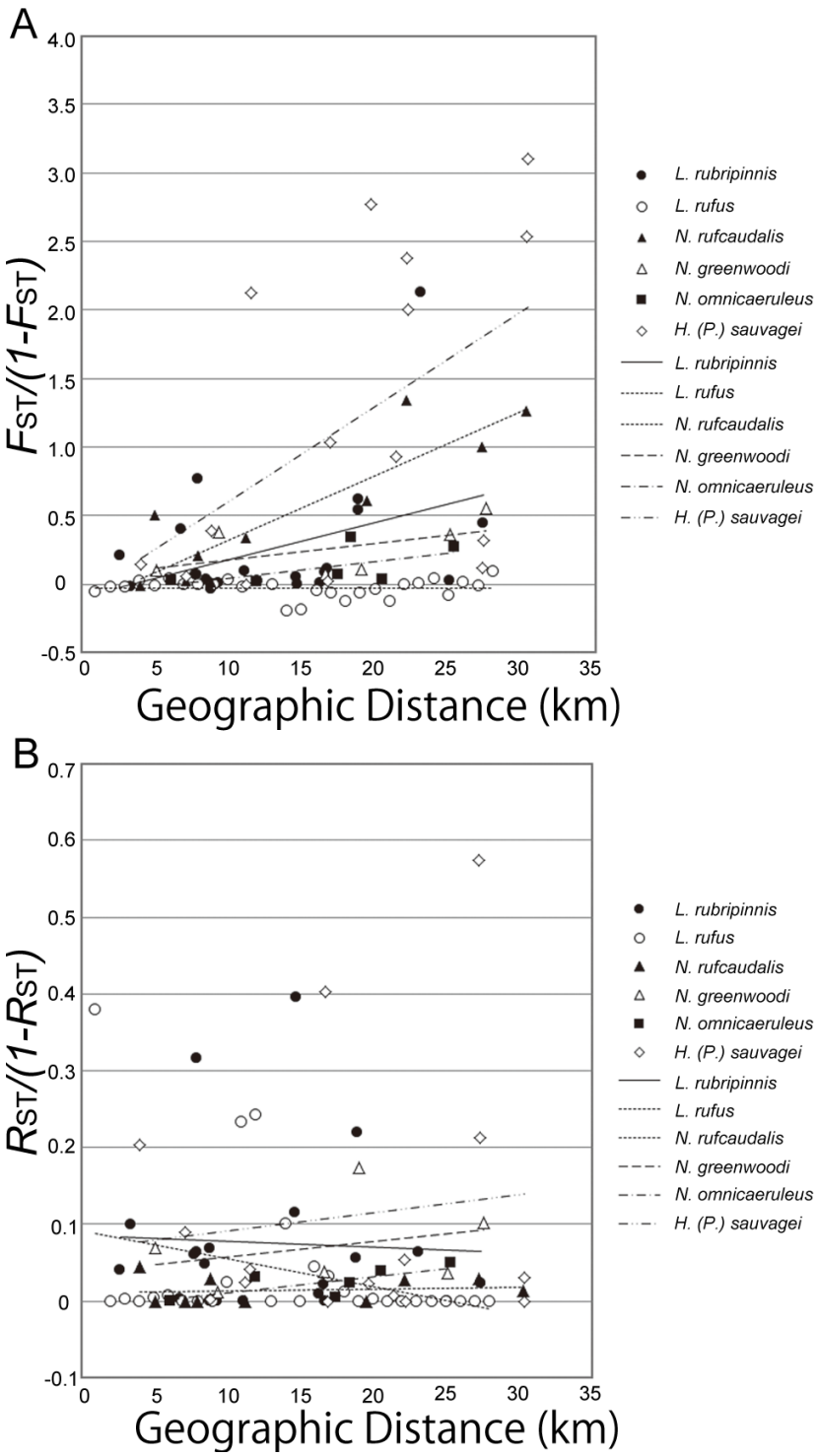

200 individuals per haplotype
100
50
30
10
1
H. pyrrhocephalus
H. laparogramma
L. rubripinnis
L. rufus
M. mbipi
N. greenwoodi
N. omnicaeruleus
N. rufocaudalis
H. sauvagei
Haplochromis spp.
Lithochromis spp.
P. macrocephala
k1
k69
k73
77
92
k54
101

Figure 6


A
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
-0.5
0
5
10
15
20
25
30
35
FST/(1-FST)
Geographic Distance (km)
L. rubripinnis
L. rufus
N. rufcaudalis
N. greenwoodi
N. omnicaeruleus
H. (P.) sauvagei
L. rubripinnis
L. rufus
N. rufcaudalis
N. greenwoodi
N. omnicaeruleus
H. (P.) sauvagei
B
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
5
10
15
20
25
30
35
RST/(1-RST)
Geographic Distance (km)
L. rubripinnis
L. rufus
N. rufcaudalis
N. greenwoodi
N. omnicaeruleus
H. (P.) sauvagei
L. rubripinnis
L. rufus
N. rufcaudalis
N. greenwoodi
N. omnicaeruleus
H. (P.) sauvagei