Molecular Ecology (1999) **8**, 1013–1026

# Biogeography and population genetics of the Lake Malawi cichlid *Melanochromis auratus*: habitat transience, philopatry and speciation

JEFFREY A. MARKERT,* MATTHEW E. ARNEGARD,† PATRICK D. DANLEY* and THOMAS D. KOCHER*

**Department of Zoology, University of New Hampshire, Durham NH 03824, USA, †Field of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853, USA*

## Abstract

**Migration rates among nine populations of the endemic Lake Malawi cichlid *Melanochromis auratus* were estimated along a 42-km stretch of habitat in the southern end of the lake. Allele frequencies were surveyed at four simple sequence repeat (SSR) loci. The data suggest migration rates among populations are quite low. Exact tests indicate that statistically detectable allele frequency differences exist between many adjacent populations in the study. The $F_{ST}$ value among all populations was estimated to be 0.151 ($P < 0.0002$). A biogeographic survey suggests that the highest levels of genetic differentiation exist between populations separated by stretches of deep water. Migration is more common between populations separated by shallower water or with shoreline dispersal routes. Reduced allelic diversity was observed at more recently created habitat patches, suggesting that either bottlenecks are associated with the colonization of new habitat patches or that these shallower sites were all founded by genetically depauperate ancestral populations. The extreme philopatry of *M. auratus*, coupled with the patchy distribution and transient nature of its preferred habitat, provides opportunities for both selection and genetic drift to produce genetic differentiation among populations. Both processes may be important to the evolution of taxonomic diversity in the East African cichlid species flocks.**

*Keywords*: biogeography, cichlids, Lake Malawi, Mbuna, microsatellites, migration

*Received 5 August 1998; revision received 22 January 1999; accepted 22 January 1999*

## Introduction

The flocks of endemic fish species in the East African Great Lakes are well known examples of 'explosive' cladogenesis (Greenwood 1964). The fishes of Lake Malawi are a dramatic example, with an estimated 500 endemic species, virtually all of which are in the Teleost family Cichlidae (Ribbink *et al*. 1983). This extraordinary taxonomic diversity has fascinated and challenged evolutionary biologists since the earliest description of the rift valley lakes by European naturalists and explorers in the 19th century (Günther 1864).

Since the first formal description of the Rift Valley taxa (see Boulenger (1915) for early citations), many explanations for the rapid evolution of taxonomic diversity have been suggested. Proposed explanations include selective mating and brood care (Kosswig 1947), adaptation to changes in habitat arising from lake level dynamics (Trewavas 1947), adaptation to a postulated diversity of habitats resulting from the sheer size of the lakes (Jackson 1955), restricted migration due to predation (Fryer 1965) and microadaptation to fragmented habitats (Fryer 1959a, b). The most recent synthetic model was proposed by Dominey (1984) who suggested that a combination of extremely low gene flow among populations, coupled with sexual selection, could lead to the fixation of distinct mate recognition systems (*sensu* Patterson (1985)) within local populations.

Most of these models stress the importance of selection

Correspondence: J. A. Markert. Fax: +01-603-862 3784; E-mail: jmarkert@cisunix.unh.edu

© 1999 Blackwell Science Ltd

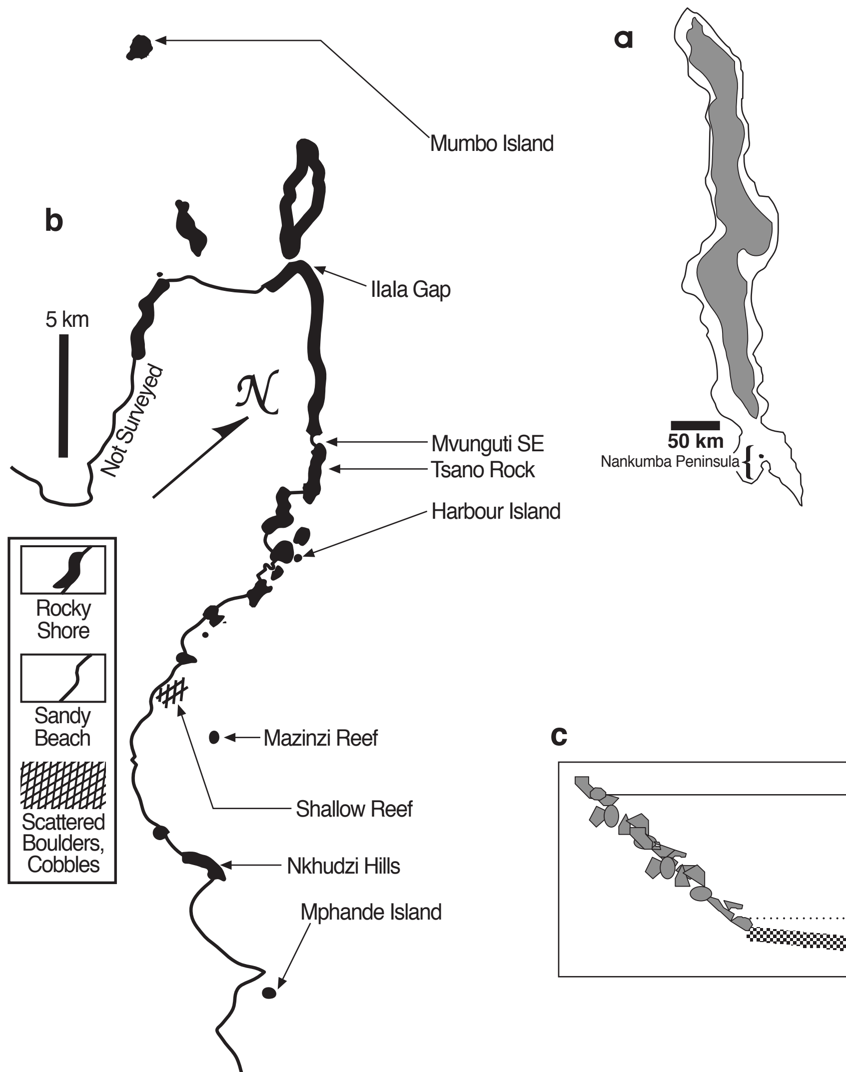


**Fig. 1** Habitat distribution along the shores of the Nankumba Peninsula. (a) The peninsula divides the southern end of Lake Malawi into two shallow basins. The shaded area represents parts of the lake which are > 200 m deep (adapted from Owen *et al*. 1990). (b) The distribution of rocky habitats along the shores of the peninsula and the location of sites from which fish were sampled. (c) A schematic diagram showing the typical arrangement of rocky habitat along the shoreline. Rocks and boulders (stippled) slope away from the shore, eventually intersecting the flatter sandy lake bottom (checks) When water levels are high (solid line) mbuna can use the rocks as habitat. Periodic recessions in lake level can destroy this habitat (dashed line). Increases in water level can completely submerge some rocky areas, making the site inhospitable to obligate rock-dwelling fishes with shallow depth preferences.

to the evolution of lineage divergence. However, selective forces must overcome the effects of gene flow between diverging gene pools or local adaptations will not develop. It is important to assess both the spatial scale of genetic differentiation and identify the physical features of the environment that constrain population differentiation if we are to evaluate models proposed to explain the rapid evolution of taxonomic diversity in the Lake Malawi cichlid species flock.

Many Malawi cichlids have a high level of habitat fidelity which, in combination with the patchy distribution of habitat types, could contribute to the evolution of lineage divergence. The shoreline is a mosaic of habitat types with rocky stretches existing as habitat islands separated by long stretches of sandy or weedy substrate (McKaye & Gray 1984). Rocky habitats also exist along the shores of several small islands and as completely submerged offshore rocky outcrops. The configuration of substrate types along the shores of Lake Malawi is not a fixed feature of the physical environment. The extent and distribution of sandy and rocky habitats is influenced by the rapid and dramatic changes in water level which are typical in Lake Malawi (Fryer 1959a; McKaye & Gray 1984; Scholz & Rosendahl 1988; Owen *et al*. 1990). Fluctuations in water level occur on both geological and historical time scales. Owen *et al*. (1990) document three climatically controlled late-Pleistocene decreases in water level, the most recent of which occurred between the years 1500 and 1850. During this period, water levels were at least 121 m below their present level, and the two southern basins were mostly dry land.

Decreased rainfall can lead to the loss of rocky habitats when water levels fall below the rock–sand interface (Fig. 1). Increases in water level can open newly flooded habitat patches for colonization and can alter the nature of existing habitat patches by increasing their depth. Variation in water levels can unite previously isolated habitat patches or subdivide continuous stretches of

habitat, depending on local topography (see McKaye & Gray (1984) or Ribbink *et al*. (1983)).

The rocky areas are the primary habitat for a well-studied guild of small, brightly coloured fish known collectively as mbuna (Fryer 1959a). The lithophilic nature of most mbuna, combined with the patchy distribution of rocky habitats within Lake Malawi, form a system in which the development of interpopulation genetic heterogeneity (a probable precondition for lineage splitting) may be explored. Early genetic data suggested that migration rates among mbuna populations might be very low. McKaye *et al*. (1984) found evidence of genetic differentiation at some allozyme loci among four populations of the widely distributed *Pseudotropheus zebra*. Bowers *et al*. (1994) found differences in mitochondrial DNA haplotype frequencies between populations in the southern end of the lake for two different species in the mbuna genus *Melanochromis*. The recent availability of simple sequence repeat (SSR) loci (Tautz 1989) with their high allelic diversity provides a tool with the resolution needed to detect fine-scale population differentiation among recently established populations.

Recently, van Oppen *et al*. (1997) have used SSR loci to detect population structure in several mbuna species across very small gaps at more stable rocky habitats near Nkhata Bay on the western shore of Lake Malawi. The newly flooded rocky habitats in the southern end of the lake, many of which are less than 200-years old, provide an opportunity to further explore the spatial and temporal scale of population structure in Malawi Cichlids. We combine fine-scale population sampling with a detailed survey of habitat distribution and an analysis of allele frequencies at SSR loci to evaluate the influence of biogeographic forces on population structure in the rock-dwelling Malawi cichlid *Melanochromis auratus*.

## Materials and methods

### *Study species*

*Melanochromis auratus* (Boulenger 1897) is an easily recognizable species that is widely distributed in the southern end of Lake Malawi. It is a small (4–9 cm), sexually dimorphic fish. Females have a bright yellow ground colour with black or dark-brown horizontal stripes. Males are slightly larger than females and have a dark-brown or blue/black ground colour with yellow-gold stripes (Bowers 1993). Unlike many other mbuna species in the vicinity of the Nankumba Peninsula, *M. auratus* shows little systematic variation in colour pattern from locality to locality, although slight intrapopulation variation in colour intensity is common. They are most common at depths between 1.5 and 10 m, although their full range extends from the surface to a depth of 40 m (Ribbink *et al*. 1983). Similar to most other mbuna, they are almost never observed over a sandy substrate.

### *Study area*

*M. auratus* were sampled from areas in the proximity of the Nankumba Peninsula which divides the southern end of Lake Malawi into two shallow basins (Fig. 1). Mbuna habitats in this vicinity are probably quite young. Much of the southern end of Lake Malawi was dry land between the years 1500 and 1850 when lake levels are estimated to have been 121 m lower than they are now (Owen *et al*. 1990), providing an upper boundary to the age of populations at most habitats.

The steeply sloping rocky habitats adjacent to the shores of the Nankumba Peninsula all intersect the flatter, sandy lake bottom (Fig. 1c). The depth of this present-day rock–sand interface should determine the order in which sites became available for colonization as the southern basins refilled (Table 1). Deeper habitats became available first while the shallower habitats became available more recently (See Ribbink *et al*. 1983; McKaye & Gray 1984; and Fig. 1 for details of this process).

The study area ranges from Mumbo Island (to the north and west of the peninsula) to Mphande Island (to the south and east) (Fig. 1b). Fish were sampled from waters adjacent to both these islands, several habitats adjacent to the eastern shore of the peninsula itself, and two submerged offshore rocky 'reefs'. Divers assessed the depth of the rock–sand interface at each site. The length of each habitat and the distances between habitats were estimated with the aid of a global positioning system (GPS) receiver and nautical maps (Tripp *et al*. 1957; Malawi Government 1977).

### *Sample collection*

A total of 368 individual *M. auratus* was sampled at nine sites in southern Lake Malawi. Sample sizes at each locality are shown in Table 2; the mean number of fish sampled/site is 41. Sample sizes at these sites, many of which are within the boundaries of Lake Malawi National Park, were restricted under the conditions of our collecting permit.

Fish were captured by SCUBA divers using monofilament nets. Divers usually worked within 50 m of each other to avoid possible complications from the Wahlund effect. An exception was made at the site designated Shallow Reef where fish were collected from two exposed rocky ledges ≈ 800 m apart, at either end of a sprawling aggregation of rocky habitats in a sand and gravel matrix generally 100–300 m from shore. This was necessary due to the relatively low densities of *M. auratus* at these sites.

Tissue for this study consisted of fin clips (≈ 0.5–1 $cm^2$) obtained from one of the unpaired fins (for fish collected

**Table 1** A description of collection localities along the Nakumba Peninsula. Habitat depth is defined as the depth of the rock–sand interface

| Collection site | Depth of habitat (m) | Habitat description |
|---|---|---|
| Harbour Island | 30 | Large boulders with some gravel and cobbles. Continuous with shoreline habitat. |
| Ilala Gap | 36 | A shoreline habitat composed primarily of large rounded boulders. Fish were collected just south of the channel between the peninsula and Domwe Island. |
| Mazinzi Reef | 13 | An isolated and submerged rocky outcrop about 3 km from shore. The total habitat area is an estimated 20 000 m$^2$ in area and is composed of all size-classes of rocky substrate. The highest point of this structure is 3 m below the lake surface. |
| Mphande Island | 4 | An isolated shallow habitat along the lakeward side of Mphande Island consisting of sedimented cobbles, small (< 1 m diameter) boulders and some large (> 3 m) boulders at the rock/sand interface. |
| Mumbo Island | 46 | An isolated habitat about 6.5 km west of the Nankumba Peninsula. Habitat at this site is primarily large rocks although some sand and gravel patches are present. |
| Mvunguti—SE | 36 | A heterogeneous habitat composed of all sizes of rocky substrate. |
| Nkhudzi Point | 11 | A heterogeneous habitat composed of all sizes of rocks in addition to patches of gravel and sand. |
| Shallow Reef | 3 | The substrate at this sprawling site is a sand/gravel matrix with many small patches of rocky habitat. The gravel field extends approximately 0.4 km from shore along at least one transect. Two large rocky ledges are found nearshore. |
| Tsano Rock | 32 | The substrate in the vicinity is primarily rocky, lightly sedimented in sheltered areas, with small patches of sand and gravel in the shallow areas. |

in Lake Malawi National Park) or from pectoral fins. Fish collected in Lake Malawi National Park were clipped and released. Fish from other sites were preserved as voucher specimens. The fin tissue was preserved in 70–100% EtOH (undenatured), and the samples were then stored at *c.* –15 °C pending transport to the USA.

### *Locus isolation and characterization*

Two of the loci used in this work (UNH-001 and UNH-002) were used previously by Kellogg *et al.* (1995) for paternity analysis. These and two additional loci, UNH-050 and UNH-231, were isolated using methods described by Lee & Kocher (1996). All four loci are perfect dinucleotide repeats. Locus 231 was cloned from *Oreochromis niloticus* and the remaining loci were developed from an *M. auratus* library. Primer sequences, GenBank accession numbers, and fragment size ranges are provided in Table 3.

### *DNA preparation and amplification*

DNA samples were extracted and amplified using the methods outlined in Kellogg *et al.* (1995). Optimal PCR conditions were determined empirically. UNH-001, UNH-050 and UNH-231 were usually amplified by 25 cycles of 94 °C (0:20), 56 °C (0:45), 72 °C (0:45). For locus UNH-002 an annealing temperature of 54 °C was substituted and 30 thermal cycles were used. Fragment size ranges are listed in Table 3. Samples were electrophoresed on a 6% denaturing acrylamide gel at 30 W for 8.25 h using an ABI 373A DNA sequencer.

### *Scoring and binning of alleles*

GeneScan Analysis software (Applied Biosystems, Foster City, California, USA) provides highly repeatable estimates of fragment size. Because of differences in base-pair composition between ABI's GeneScan-500 size standard and the PCR products, these estimates were almost never integers. In order to determine fragment homology, fragment size estimates at each locus were sorted by size and 'binned' into allele size estimates that typically differed by 2 bp. Allele size estimates were sorted by size and ranked. These rank scores were then plotted against allele size to provide a visual representation of the bins. If bin limits were ambiguous, individuals at both extremes of that bin and from neighbouring bins were re-run on a single gel.

### *Detection of null alleles*

The possibility of 'null' alleles (alleles that cannot be visualized due to mutations in the PCR primer site) complicates the analysis of SSR data. The frequency of these alleles can be quite high (see Lehman *et al.* (1996) and Allen *et al.* (1995) for recent examples). In their survey of populations of several *Pseudotropheus* species from Lake Malawi, van Oppen *et al.* (1997) reported that a true breeding null allele is present at locus UNH-002 in some of the mbuna species they studied.

To estimate the frequency of null alleles in our data set, individuals for which PCR products could be generated at only three of the four loci after two or more attempts were provisionally assumed to be homozygous for a

© 1999 Blackwell Science Ltd, *Molecular Ecology*, 8, 1013–1026

**Table 2** Sample size and summary statistics for each locus and population

| | | UNH-001 | | | | | | UNH-002 | | | | | | UNH-050 | | | | | | UNH-231 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | No. of Alleles | $n_e$ | $H_O$ | $H_E$ | Null | $F_{IS}$ | No. of Alleles | $n_e$ | $H_O$ | $H_E$ | Null | $F_{IS}$ | No. of Alleles | $n_e$ | $H_O$ | $H_E$ | Null | $F_{IS}$ | No. of Alleles | $n_e$ | $H_O$ | $H_E$ | Null | $F_{IS}$ |
| Harbour Island | 74 | 15 | 4.64 | 0.767 | 0.790 | 0.11 | n.s. | 12 | 4.92 | 0.767 | 0.805 | 0.18 | n.s. | 10 | 6.63 | 0.785 | 0.856 | 0.08 | n.s. | 14 | 3.95 | 0.782 | 0.754 | 0.10 | n.s. |
| Ilala Gap | 55 | 12 | 4.60 | 0.772 | 0.791 | 0.14 | n.s. | 17 | 7.68 | 0.809 | 0.878 | 0.15 | n.s. | 14 | 7.02 | 0.895 | 0.866 | 0.04 | n.s. | 20 | 10.20 | 0.911 | 0.912 | 0.09 | n.s. |
| Mazinzi Reef | 38 | 12 | 2.81 | 0.528 | 0.652 | 0.13 | 0.192 | 6 | 2.25 | 0.514 | 0.564 | 0.14 | n.s. | 10 | 3.48 | 0.675 | 0.722 | 0.11 | n.s. | 10 | 2.67 | 0.573 | 0.636 | 0.14 | n.s. |
| Mphande Island | 27 | 10 | 3.39 | 0.481 | 0.719 | 0.17 | n.s. | 5 | 1.49 | 0.292 | 0.337 | 0.16 | n.s. | 3 | 1.59 | 0.385 | 0.377 | 0.17 | n.s. | 8 | 3.79 | 0.840 | 0.751 | 0.31 | n.s. |
| Mumbo Island | 37 | 11 | 5.53 | 0.829 | 0.832 | 0.09 | n.s. | 6 | 4.07 | 0.513 | 0.766 | 0.16 | 0.332 | 7 | 2.82 | 0.757 | 0.654 | 0.07 | n.s. | 9 | 5.72 | 0.891 | 0.836 | 0.00 | n.s. |
| Mvunguti S.E. | 12 | 8 | 3.41 | 0.637 | 0.741 | 0.18 | n.s. | 10 | 7.58 | 0.833 | 0.905 | 0.05 | n.s. | 9 | 6.13 | 0.833 | 0.872 | 0.02 | n.s. | 7 | 4.17 | 0.499 | 0.793 | 0.19 | n.s. |
| Nkhudzi Point | 35 | 12 | 3.05 | 0.691 | 0.683 | 0.23 | n.s. | 5 | 1.68 | 0.259 | 0.409 | 0.15 | 0.374 | 6 | 1.74 | 0.419 | 0.431 | 0.22 | n.s. | 12 | 6.71 | 0.907 | 0.865 | 0.05 | n.s. |
| Shallow Reef | 36 | 9 | 2.31 | 0.313 | 0.576 | 0.25 | n.s. | 8 | 1.64 | 0.303 | 0.397 | 0.10 | n.s. | 8 | 3.46 | 0.765 | 0.721 | 0.06 | n.s. | 7 | 2.28 | 0.617 | 0.570 | 0.00 | n.s. |
| Tsano Rock | 54 | 15 | 6.32 | 0.856 | 0.851 | 0.02 | n.s. | 13 | 10.00 | 0.821 | 0.909 | 0.10 | n.s. | 13 | 8.81 | 0.896 | 0.897 | 0.15 | n.s. | 12 | 6.96 | 0.855 | 0.866 | 0.09 | n.s. |

$N$ = number of fish sampled/site; No. of Alleles = total number of alleles observed in a population; $n_e$ = effective number of alleles; $H_O$ = observed heterozygosity; $H_E$ = expected heterozygosity; Null = a maximum likelihood estimate of the null allele frequency in a population; $F_{IS}$ values are reported only when they are significantly distinguishable from zero.
n.s. = not significant.

**Table 3** Information on the four simple sequence loci

| Locus | Primer sequences | GenBank Accession no. | Annealing temperature | Fragment size range (bp) | Total no. of alleles | Average heterozygosity |
|---|---|---|---|---|---|---|
| UNH-001 | GATTAACTCTGTCCCTGTCT<br>CTGAAGTGTTAAAAATATTGTT | U17044 | 56 °C | 157–243 | 30 | 0.650 |
| UNH-002 | TTATCCCAACTTGCAACTCTATTT<br>TCCATTTCCTGATCTAACGACAAG | U17045 | 54 °C | 175–236 | 24 | 0.549 |
| UNH-050 | GTCATCCCACTCACTAACAT<br>AGAACAAACACAGGAAACTAT | AF036714 | 56 °C | 292–364 | 21 | 0.707 |
| UNH-231 | GCCTATTAGTCAAAGCGT<br>ATTTCTGCAAAAGTTTTCC | G12382 | 56 °C | 192–253 | 27 | 0.777 |

null allele. These frequencies were used to estimate the frequency of the null allele in each population using the maximum-likelihood algorithm in GENEPOP 3.1 (Goudet 1995). Heterozygote deficiency was estimated using the score test (Rousset & Raymond 1995) of GENEPOP to estimate $F_{IS}$ values and their significance level.

### *Estimators of between-population heterogeneity*

The statistical significance of interpopulation allele frequency differences was determined by using exact tests in Raymond & Rousset's (1995) GENEPOP package which follows the methods described in Rousset & Raymond (1995). Estimates of population heterozygosity, allele frequencies and *F*-statistics were estimated with the aid of Goudet's (1995) program FSTAT. This package estimates *F*-statistics using the method of Weir & Cockerham (1984) and calculates confidence intervals for these estimates using a resampling algorithm which permits permutation among loci, alleles, and populations.

Two sets of $F_{ST}$ calculations were performed. The first estimated the overall $F_{ST}$ among all populations. The second set estimated pairwise $F_{ST}$ values for all pairs of populations. Confidence intervals were estimated for $F_{ST}$ estimates between adjacent populations and between the two terminal populations by performing 5000 resamplings for the overall $F_{ST}$ estimate and 2000 resamplings for each of the pairwise comparisons.

Barton & Slatkin's (1985) rare allele based estimate of $N_m$ was calculated using GENEPOP 3.1 (Raymond & Rousset 1995). This divergence estimator should be less sensitive than Weir & Cockerham's (1984) $F_{ST}$ to biases which result from the relative youth of the populations surveyed.

Pairwise Nei's Distance ($D_N$) and $(\delta\mu)^2$ were calculated for all pairs of populations using the program MICROSAT [version 1.4d] (Minch *et al.* 1995–96). Nei's distance was determined to be a more appropriate measure of divergence between these populations than $(\delta\mu)^2$ (Goldstein *et al.* 1995) which assumes a single step mutation model for SSR loci. Although evidence is accumulating which indicates that stepwise mutation is probably responsible for generating the global array of alleles present at these loci (but see Farrall & Weeks 1998), it seems likely that the allele frequency distributions observed in the sampled populations are the result of recent historical sampling processes rather than postdivergence mutation, given the extremely recent origin of the habitats in the area surveyed (Owen *et al.* 1990). The mutation-based statistic $(\delta\mu)^2$ is therefore included for comparison only and is not intended to represent a reliable index of genetic differentiation.

An analysis of isolation by distance was conducted by regressing $\log_{10}M$, where $M = 1/4((1/F_{ST}) - 1)$ against $\log_{10}$ distance following Slatkin (1993). The statistical significance of the relationship between distance and the parameter $F_{ST}/(1 - F_{ST})$ was estimated by using the Mantel test in GENEPOP 3.1.

### *Linkage disequilibrium and locus correlation*

To determine whether the four loci surveyed represent independent samples of the *M. auratus* genome, GENEPOP was used to perform Fisher's exact test under a null hypothesis of no association between genotypes at different loci.

To test whether there was a concordance in $F_{ST}$ estimates at all four loci, a matrix of Pearson product moment correlations and their significance levels was calculated for all pairwise $F_{ST}$ estimates.

## Results

### *Distribution of habitats*

Habitat depth estimates and brief habitat descriptions are reported in Table 1. Collection sites and habitat distribution are shown in Fig. 1b. In general, the deepest habitats surveyed are at sites in the north and west of the sampled area, whereas the shallower sites are in the south and east. A description of the intervening substrate between collection sites is shown in Table 4.

**Table 4** Estimators of genetic differentiation between adjacent, nearly adjacent, and terminal populations. Exact test significance values below a Bonferroni corrected P-crit. of 0.0013 are underlined

| Collection sites | Distance between collection points (km) | Intervening substrate | Nei's $D$ (standard error) | Single locus $F_{ST}$ estimates | | | | Four locus $F_{ST}$ (standard deviation) | $P$ $F_{ST} \leq 0$ | $N_m$ | Exact test $P$-values | | | | $(\partial\mu)^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | UNH-001 | UNH-002 | UNH-050 | UNH-231 | | | | UNH-001 | UNH-002 | UNH-050 | UNH-231 | |
| Mumbo Island—Ilala Gap | 10.4 | Deep water, sandy lake bottom | 0.791 (0.12) | 0.074 | 0.121 | 0.146 | 0.060 | 0.104 (0.019) | 0.0010 | 1.71 | <0.00001 | <0.00001 | <0.00001 | <0.00001 | 10.393 |
| Ilala Gap—Tsano Rock | 8.1 | Rock and an ~ 350 m sandy bay at Mvunguti Village | 0.226 (0.04) | 0.037 | 0.028 | 0.020 | 0.032 | 0.029 (0.003) | 0.0005 | 4.10 | 0.00020 | <0.00001 | 0.00312 | <0.00001 | 1.399 |
| Mvunuti SE—Tsano Rock | 0.9 | Rocky coastline | 0.004 (0.079) | 0.055 | 0.000 | 0.000 | 0.004 | 0.013 (0.014) | n.s. | 5.83 | 0.06166 | 0.70178 | 0.66528 | 0.82534 | 0.371 |
| Tsano Rock—Harbour Island | 3.7 | Rock and a sandy channel (> 24 m deep) at Monkey Bay | 0.364 (0.11) | 0.095 | 0.057 | 0.034 | 0.047 | 0.058 (0.013) | 0.0005 | 4.46 | <0.00001 | <0.00001 | <0.00001 | <0.00001 | 6.294 |
| Harbour Island—Mazinzi Reef | 7.6 | Deep water, sandy lake bottom | 0.335 (0.25) | 0.068 | 0.112 | 0.157 | 0.008 | 0.095 (0.032) | 0.0005 | 1.84 | <0.00001 | <0.00001 | <0.00001 | 0.01130 | 0.446 |
| Harbour Island—Shallow Reef | 7.7 | Altemating sandy and rocky shoreline | 0.307 (0.21) | 0.089 | 0.172 | 0.148 | 0.020 | 0.113 (0.031) | 0.0005 | 2.50 | <0.00001 | <0.00001 | <0.00001 | 0.09454 | 1.308 |
| Shallow Reef—Mazinzi Peef | 2.8 | Shallow sandy lake bottom | 0.023 (0.032) | 0.000 | 0.014 | 0.047 | 0.000 | 0.016 (0.014) | n.s. | 5.19 | 0.03388 | 0.02398 | 0.00300 | 0.59166 | 0.656 |
| Shallow Reef—Nkhudzi Point | 6.8 | Sandy shoreline | 0.255 (0.19) | 0.043 | 0.018 | 0.205 | 0.180 | 0.140 (0.061) | 0.0005 | n.a. | 0.01496 | <0.00001 | <0.00001 | <0.00001 | 6.312 |
| Mazinzi Reef—Nknudzi Point | 5.0 | Open water, sandy lake bottom | 0.435 (0.35) | 0.003 | 0.027 | 0.335 | 0.152 | 0.158 (0.081) | 0.0005 | 1.81 | 0.01314 | 0.00004 | <0.00001 | <0.00001 | 6.203 |
| Nkhudzi Point—Mphande Island | 5.6 | Sandy shoreline | 0.016 (0.19) | 0.065 | 0.000 | 0.000 | 0.033 | 0.015 (0.011) | n.s. | 4.52 | 0.00516 | 0.23412 | 0.38190 | 0.01560 | 1.057 |
| Mumbo Island—Mphande Island | 42.4 | Open water, sandy shoreline, rocky shoreline | 2.778 (0.93) | 0.209 | 0.405 | 0.473 | 0.173 | 0.308 (0.073) | 0.0005 | 0.32 | <0.00001 | <0.00001 | <0.00001 | <0.00001 | 19.772 |
| Among all populations | — | — | — | 0.141 | 0.191 | 0.160 | 0.110 | 0.151 (0.017) | 0.0002 | 3.67 | — | — | — | — | — |

**Table 5** Population allele frequencies at four simple sequence repeat loci

Locus and Population / Alleles (estimated no. of CA repeats)

| UNH-001 | 14 | 18 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 44 | 45 | 47 | 49 | 69 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harbour Island | 0.01 | | | | | 0.01 | 0.07 | | 0.04 | 0.04 | 0.01 | 0.07 | 0.33 | | 0.01 | | 0.30 | | 0.01 | | | | 0.02 | | 0.03 | 0.03 | 0.02 | | |
| Ilala Gap | | | 0.02 | 0.01 | | 0.02 | 0.01 | 0.01 | 0.40 | 0.17 | | 0.09 | 0.07 | | 0.02 | 0.06 | 0.11 | | | | | | | | | | | | |
| Mazinzi Reef | | | | | | 0.01 | | | 0.01 | 0.04 | | | 0.57 | 0.06 | 0.01 | 0.01 | 0.10 | 0.09 | | | 0.01 | | 0.01 | | | 0.04 | | | |
| Mphande Island | | 0.13 | | | | | | | | | | | 0.50 | 0.13 | | 0.02 | 0.07 | | | | 0.04 | 0.02 | | | 0.02 | 0.02 | 0.06 | | |
| Mumbo Island | | | | 0.04 | | 0.26 | 0.01 | 0.06 | 0.16 | 0.01 | | 0.26 | 0.01 | | | | 0.13 | 0.03 | | 0.03 | | | | | | | | | |
| Mvunguti S.E. | | | | 0.05 | | | 0.09 | | 0.09 | 0.50 | | 0.14 | 0.05 | | 0.05 | | | | | | | 0.05 | | | | | | | |
| Nkhudzi Point | | 0.02 | | | | | | | 0.02 | | 0.02 | 0.02 | 0.54 | | 0.02 | 0.02 | 0.15 | | 0.04 | | 0.10 | | | | 0.04 | 0.02 | | | |
| Shallow Reef | | | | | | | | | | | | | 0.64 | | 0.06 | 0.03 | 0.11 | 0.02 | | 0.02 | | | | | | 0.05 | | 0.05 | 0.03 |
| Tsano Rock | | | | 0.01 | 0.01 | 0.04 | 0.01 | | 0.19 | 0.19 | 0.02 | 0.25 | 0.14 | | 0.02 | 0.02 | 0.05 | 0.02 | | | | | | 0.01 | 0.02 | | | | |

| UNH-002 | 19 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 39 | 40 | 43 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harbour Island | | | 0.17 | 0.02 | 0.01 | | 0.01 | 0.14 | | 0.01 | | 0.30 | | 0.07 | 0.23 | 0.02 | | 0.01 | 0.01 | | |
| Ilala Gap | 0.04 | 0.02 | 0.05 | 0.04 | 0.19 | 0.01 | 0.24 | 0.09 | 0.01 | | 0.04 | 0.01 | 0.12 | 0.01 | 0.03 | | | | 0.03 | 0.03 | 0.02 |
| Mazinzi Reef | | | 0.13 | 0.03 | | | | | 0.06 | 0.07 | | 0.64 | 0.07 | | | | | | | | |
| Mphande Island | | | | | 0.06 | | | | | 0.06 | | 0.81 | 0.04 | | 0.02 | | | | | | |
| Mumbo Island | | | | | 0.26 | 0.05 | | | 0.35 | 0.15 | 0.18 | 0.01 | | | | | | | | | |
| Mvunguti S.E. | | | 0.08 | 0.04 | 0.04 | | 0.13 | 0.17 | | | 0.13 | | | 0.13 | 0.21 | | 0.04 | | | 0.04 | |
| Nkhudzi Point | | | | | 0.06 | | | | | 0.02 | | 0.76 | 0.13 | 0.03 | | | | | | | |
| Shallow Reef | | | 0.09 | | | | | 0.03 | 0.03 | | 0.02 | 0.77 | 0.02 | | 0.02 | | 0.03 | | | | |
| Tsano Rock | 0.01 | 0.01 | 0.17 | 0.02 | 0.12 | | 0.12 | 0.06 | 0.04 | 0.03 | 0.06 | 0.04 | 0.01 | 0.08 | 0.14 | | 0.06 | 0.01 | | 0.01 | |

| UNH-005 | 67 | 79 | 82 | 84 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 102 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harbour Island | 0.12 | | | | 0.02 | 0.15 | | 0.20 | 0.15 | 0.05 | 0.02 | | 0.20 | 0.08 | | | 0.01 | | |
| Ilala Gap | | | 0.01 | 0.05 | 0.20 | 0.06 | 0.02 | 0.18 | 0.03 | 0.02 | 0.03 | 0.01 | 0.01 | 0.18 | 0.16 | 0.02 | | | |
| Mazinzi Reef | | | | | 0.49 | 0.13 | 0.09 | 0.13 | | 0.01 | 0.09 | | 0.01 | | 0.01 | | 0.01 | | 0.01 |
| Mphande Island | | | | | 0.04 | 0.77 | | 0.19 | | | | | | | | | | | |
| Mumbo Island | | | | | 0.01 | | | | | | | 0.09 | 0.20 | 0.54 | 0.03 | 0.01 | | 0.11 | |
| Mvunguti S.E. | | | | | 0.13 | 0.25 | 0.04 | 0.04 | 0.21 | 0.08 | | | 0.17 | 0.04 | 0.04 | | | | |
| Nkhudzi Point | | 0.03 | | | 0.02 | 0.74 | 0.02 | 0.15 | 0.05 | | | | | | | | | | |
| Shallow Reef | | | | | 0.38 | 0.35 | 0.01 | 0.03 | | 0.10 | 0.01 | | | | | | 0.03 | | 0.07 |
| Tsano Rock | | | | 0.01 | 0.14 | 0.15 | 0.04 | 0.08 | 0.17 | 0.05 | 0.04 | 0.04 | 0.05 | 0.16 | 0.06 | 0.01 | | | |

| UNH-231 | 51 | 52 | 53 | 54 | 56 | 58 | 59 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 79 | 80 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harbour Island | 0.03 | | | | | | | 0.01 | 0.23 | 0.02 | 0.43 | | 0.09 | 0.04 | 0.02 | | 0.05 | 0.01 | | | | 0.01 | 0.01 | 0.02 | 0.04 |
| Ilala Gap | 0.09 | | 0.01 | 0.02 | 0.03 | 0.01 | 0.10 | 0.01 | 0.01 | 0.03 | 0.08 | 0.03 | 0.19 | 0.04 | 0.13 | 0.03 | 0.02 | | | 0.01 | 0.11 | 0.01 | 0.01 | | |
| Mazinzi Reef | | | 0.01 | 0.01 | | | | | 0.27 | | 0.54 | 0.04 | 0.04 | | | | 0.01 | | 0.01 | 0.03 | 0.01 | | | | |
| Mphande Island | | | | | | | | 0.02 | 0.34 | | 0.08 | | | 0.04 | 0.02 | | | 0.06 | | | 0.08 | | 0.36 | | |
| Mumbo Island | | | | | | | | | | | 0.11 | | 0.07 | 0.31 | 0.08 | | 0.11 | 0.03 | 0.08 | | 0.19 | 0.03 | | | |
| Mvunguti S.E. | 0.21 | | | | | | | 0.04 | 0.04 | | 0.21 | | 0.38 | | 0.08 | 0.04 | | | | | | | | | |
| Nkhudzi Point | | | | | | 0.02 | | | 0.20 | | 0.11 | 0.02 | 0.02 | 0.14 | 0.17 | | 0.02 | 0.03 | | | 0.05 | 0.03 | 0.20 | | |
| Shallow Reef | | | | | | | | | 0.22 | 0.01 | 0.62 | 0.03 | 0.07 | 0.03 | | | | | | | 0.01 | | | | |
| Tsano Rock | 0.18 | 0.02 | | | | | | 0.08 | 0.09 | | 0.24 | 0.01 | 0.17 | 0.05 | 0.07 | 0.04 | 0.03 | 0.01 | | | | | | | |

### Allelic diversity and heterozygosity

A total of 29, 21, 19, and 25 alleles was observed at loci UNH-001, UNH-002, UNH-050 and UNH-231, respectively. The average population heterozygosity at all four loci is 0.671. These results and other estimates of within-population genetic diversity are shown in Table 2. Allele frequencies for all loci in all populations are shown in Table 5.

A strong positive correlation was observed between the maximum depth of rocky substrate at a site and the observed heterozygosity at that site ($r^2 = 0.803$, $P = 0.003$). However, this pattern may be a result of population ancestry rather than habitat age because all the shallow sites are located in the southern end of our survey area.

The maximum likelihood estimate of null allele frequencies within each population ranged from 0.0 to 0.31 (Table 2). These estimates will probably overstate the frequency of null alleles because PCR reactions can fail for a variety of reasons other than primer incompatibility (Brookfield 1996); however, at least one additional amplification attempt was made on individuals that did not amplify initially. We expect that true null alleles should be associated with an excess of homozygosity. Populations with $F_{IS}$ values which are significantly different from zero are indicated in Table 2. Populations with high null allele frequency estimates did not necessarily have $F_{IS}$ values significantly different from zero.

### Population differentiation

A high level of population structure was observed. The overall $F_{ST}$ estimate is 0.151 (95% CI = 0.121 – 0.186, bootstrapping over loci). Pairwise population statistics and distance estimates between adjacent collection sites are summarized in Table 4. Pairwise $F_{ST}$ values between adjacent populations range from 0 to 0.158. Pairwise estimates of $N_m$ (based on the rare alleles method) between adjacent sites range from 1.71 to 5.83.

Gene flow between the two terminal sites is very low. Barton & Slatkin's (1986) rare alleles method estimates the genetic equivalent of 0.32 migrants/generation. The $F_{ST}$-based estimate is slightly higher at 0.56 migrants/generation. Both methods produce concordant estimates of relative levels of migration (Kendall's tau = 0.722, $P = 0.01$) in adjacent populations. The only disagreement in rank scores occurred in population comparisons in which the $F_{ST}$ value was statistically indistinguishable from zero, leading us to conclude that either estimator is a good relative index of gene flow between habitat patches despite the recent establishment of these populations.

All pairwise $D_N$ and $F_{ST}$ estimates are shown in Table 6. The highest level of differentiation was observed between the two terminal sites, Mumbo and Mphande Islands. Nei's $D$ value between these two sites is 2.778 and $F_{ST}$ is 0.308 ($P = 0.0005$). The lowest $N_m$ and highest $(\delta\mu)^2$ estimates were also observed between the two terminal sites.

With a Bonferroni-corrected $P$ (critical) of 0.0013, significant differences in allele frequency were detected at all four loci in the following adjacent population pairs: Mumbo Island/Ilala Gap and Tsano Rock/Harbour Island. Differentiation was detected at three of the four loci between Ilala Gap/Tsano Rock, Harbour Island/Mazinzi Reef, Harbour Island/Shallow Reef, Shallow Reef/Nkhudzi Hills, and Mazinzi Reef/Nkhudzi Hills. No evidence of differentiation was detected between Mvunguti SE/Tsano Rock, Mazinzi Reef/Shallow Reef or Nkhudzi Hills/Mphande Island (Table 4).

### Linkage disequilibrium and locus correlation

No evidence of linkage between loci was found, suggesting that each locus is an independent estimate of population structure.

Pearson product moment correlations between $F_{ST}$ estimates at different loci range from 0.4162 [UNH-002 and UNH-050] to 0.8750 [UNH-002 and UNH-231], with an average of 0.6321. The correlation matrix is statistically significant ($P < 0.001$), suggesting different loci generally yield concordant estimates of genetic differentiation.

**Table 6** A matrix of four locus average $F_{ST}$ values (below diagonal) and $D_N$ values (above diagonal)

| | Harbour Island | Shallow Reef | Ilala Gap | Mphande Island | Mazinzi Reef | Mumbo Island | Mvunguti S. E. | Nkhudzi Point | Tsano Rock |
|---|---|---|---|---|---|---|---|---|---|
| Harbour Island | — | 0.307 | 0.939 | 0.522 | 0.335 | 1.348 | 0.471 | 0.490 | 0.364 |
| Shallow Reef | 0.113 | — | 1.372 | 0.286 | 0.023 | 2.370 | 0.897 | 0.255 | 0.734 |
| Ilala Gap | 0.111 | 0.229 | — | 1.684 | 1.287 | 0.700 | 0.262 | 1.490 | 0.226 |
| Mphande Island | 0.168 | 0.150 | 0.246 | — | 0.416 | 2.778 | 1.414 | 0.016 | 1.106 |
| Mazinzi Reef | 0.095 | 0.016 | 0.184 | 0.171 | — | 2.138 | 0.996 | 0.435 | 0.753 |
| Mumbo Island | 0.163 | 0.305 | 0.104 | 0.308 | 0.264 | — | 1.034 | 2.465 | 0.653 |
| Mvunguti SE | 0.092 | 0.251 | 0.057 | 0.279 | 0.212 | 0.158 | — | 1.270 | 0.004 |
| Nkhudzi Hills | 0.141 | 0.140 | 0.217 | 0.015 | 0.158 | 0.286 | 0.245 | — | 1.000 |
| Tsano Rock | 0.058 | 0.179 | 0.029 | 0.213 | 0.144 | 0.095 | 0.013 | 0.186 | — |

### *Isolation by distance*

Although geological evidence suggests that the populations in the southern end of Lake Malawi were founded very recently, a strong pattern of isolation by distance is observed, with the slope of the regression line being $-0.996$, $r^2 = 0.694$. This relationship between distance and genetic differentiation is statistically significant ($P < 0.00001$). This pattern is consistent with a one-dimensional stepping-stone model at equilibrium (Hellberg 1995).

## Discussion

SSR loci reveal evidence of genetic substructuring in Lake Malawi cichlids on very fine geographical scales. The first published study on this subject (van Oppen *et al.* 1997) found evidence of kilometre-scale structuring in four mbuna species (genus *Metriaclima* and genus *Pseudotropheus*) from four habitat patches along the western shore of the lake. We observe an even higher level of genetic differentiation among *M. auratus* populations from the Nankumba Peninsula with an overall $F_{ST}$ value an order of magnitude higher than the average $F_{ST}$ estimated by van Oppen *et al.* (1997). Evidence of fine-scale genetic substructuring exists for other mbuna species including *Labeotropheus fuelleborni* (Arnegard *et al.* 1999), and additional *Metriaclima* species complex taxa (P. Danley *et al.*, in preparation). These data suggest that van Oppen *et al.*'s (1997) assertion that mbuna populations may represent thousands of genetically divergent subunits has strong empirical support. By systematically surveying habitats in a small area in southern Lake Malawi and sampling *M. auratus* from these habitats we are able to determine which components of the physical environment constrain migration and further explore the spatial and temporal scales of genetic differentiation in Malawi cichlids.

*M. auratus* in southern Lake Malawi show a high level of population differentiation, even relative to the other mbuna species surveyed to date. An overall $F_{ST}$ value of 0.151 ($P < 0.0002$) was observed among all sites along a 42-km transect. Migration between the two terminal sites is estimated to be the genetic equivalent of one migrant every second generation or less. High pairwise $F_{ST}$ and low $N_m$ values were observed between several adjacent populations, suggesting that philopatry is a general feature of *M. auratus* biology rather than an artefact of a single major barrier to gene flow within the area surveyed.

The highest level of differentiation was observed between populations separated by long stretches of deep water. Conversely, collection sites separated by long stretches of rocky or sandy shoreline show much lower levels of differentiation. These observations are consistent with Ribbink's (1986) hypothesis that deep water can serve as a strong barrier to gene flow. The lowest $N_m$ estimates between adjacent sites are found between Mumbo Island and Ilala Gap which are separated by approximately 10 km of deep, open water. The rock–sand interface occurs at a depth of 45 m at Mumbo Island and 36 m at Ilala Gap. The intervening trough is about 100 m deep (Tripp *et al.* 1957). Similarly low $N_m$ estimates were calculated between the Mazinzi Reef, a submerged offshore rocky outcrop, and two nearby shoreline sites; Harbour Island and Nkhudzi Hills. Although the lake bottom between these two pairs of sites is not nearly as deep as it is between Mumbo Island and the peninsula, depth may still be a formidable barrier for *M. auratus* which are most common < 10 m below the surface at all the sites we surveyed. Marsh & Ribbink (1981) and Hill & Ribbink (1978) have demonstrated experimentally that some mbuna taxa are unable to control their buoyancy in waters greater than 40 m deep and that the maximum rate of depth acclimation for fish in the mbuna genus *Petrotilapia* is < 4 m/day, leading the authors to suggest that substrate-hugging mbuna are physiologically incapable of crossing long stretches of deep water. Our data are consistent with Ribbink *et al.*'s (1986) hypothesis and the Ribbink *et al.* (1983) hypothesis that deep water is a barrier to migration due to the physiological limits of mbuna swim bladders.

The highest levels of gene flow were inferred for samples collected from either end of continuous stretches of rocky habitat or from rocky patches separated by shallow sandy shoreline. Ilala Gap and Tsano Rock lie at opposite ends of a nearly continuous stretch of rocky coastline which is interrupted only by an approximately 350 m stretch of sandy shoreline at Mvunguti Village. These two sites are 8.2 km apart yet show considerably less differentiation than the sites separated by similar stretches of open water. The estimated $F_{ST}$ value of 0.029 between these two sites was significantly different from zero (at a Bonferroni-corrected $P < 0.0038$), but this is low relative to the other significant $F_{ST}$ values. The $N_m$ estimate of 4.10 suggests that migration is common between these two sites.

Shallow sandy shoreline also appears to facilitate dispersal. Mphande Island is located in a shallow bay about 5.6 km southeast of Nkhudzi Hills. The shoreline between the two sites is apparently free of classical mbuna habitat yet the estimated $N_m$ value of 4.52, absence of detectable allele frequency differences at any of the four loci and an $F_{ST}$ value statistically indistinguishable from zero suggest that migration occurs between these two sites. A similar pattern is observed between Mazinzi Reef and Shallow Reef which have the second-highest pairwise estimate of migration observed between adjacent sites in this study

($N_m = 5.19$), second only to the rate between Tsano Rock and Mvunguti S.E. which are separated by < 1 km of rocky habitat. In contrast, low migration rates are estimated between Mazinzi Reef and other neighbouring populations. Although this difference in migration rates might be partly explained by the fact that Shallow Reef is much closer to Mazinzi Reef than either Harbour Island or Nkhudzi Hills (2.8 km vs. 7.6 or 5.0 km, respectively), it seems likely that other geographical features influence the dispersal of fish from Mazinzi Reef to Shallow Reef. Unlike the other more compact habitats we surveyed, Shallow Reef is a sprawling complex of small rocky habitats in a sand/gravel matrix. This complex extends 0.4 km from shore along at least one transect. It is possible that undetected habitat patches similar to the habitat at Shallow Reef form a series of stepping stones between Mazinzi Reef and Shallow Reef. Alternatively, the similarity between these two pairs of sites might be a result of recent founding of the two shallower sites (Shallow Reef and Mphande Island) by populations with only modest levels of genetic diversity.

An alternative explanation for the distribution of genotypes at current habitats might be that they reflect older patterns of differentiation rather than current patterns of migration. We see an example of this process among the southern populations. Allele frequencies at Shallow Reef are most similar to those at Mazinzi Reef, and are distinct from Harbour Island, suggesting that gene flow has not occurred along the shoreline between Harbour Island and Shallow Reef. These sites are separated by a series of shallow sandy bays punctuated with a number of rocky habitats which might be expected to serve as stepping stones. The fact that migration is not detected along these stepping stones may be a result of the recent availability of these very shallow (mostly < 2.5 m) habitats which have become stepping stones too recently for migration along this route to homogenize allele frequencies between Harbour Island and Shallow Reef.

In a similar fashion, the high level of genetic divergence between the two apparent genetic units (1) Mazinzi Reef-Shallow Reef and (2) Nkhudzi Hills-Mphande Island suggests that these two population pairs are more diverged than we might expect if Nkhudzi Hills were colonized by migrants from Mazinzi Reef as water levels rose. Based solely on the distance between sites, the expectation is that Nkhudzi Hills would be only slightly more diverged from Mazinzi Reef than Shallow Reef. The observed pattern could result if genetic divergence evolved before the current habitats were colonized. Several isolated sites in the lake could have served as refugia when lake levels were lower (see Arnegard *et al*. 1999). If these refugia were as isolated as Mazinzi Reef or Mumbo Island are today, then populations at these sites might become evolutionarily detached from other populations in the lake, and serve as a source of colonists as new rocky outcrops are inundated. There are at least two potential refugia in the area that could serve as *M. auratus* habitat during moderate recessions in lake level. The waters adjacent to Boadzulu Island, some 13 km south of Mphande Island, contain rocky habitat down to at least 40 m (Ribbink *et al*. 1983). Jerusalem Reef, 7 km east of Mazinzi Reef, is an isolated rocky outcrop, the top of which is ≈ 40 m below the lake surface at its shallowest point. *M. auratus* is not known to exist at either of these sites currently, but it is known to exist at sites south of Boadzulu Island. The habitat in these areas appears similar to that at sites where *M. auratus* are abundant, except for the greater depth of habitat at Jerusalem Reef. Several other deep reefs in the southeast arm of Lake Malawi are known to fishermen or are shown on navigational maps (Tripp *et al*. 1957). Jerusalem Reef and other similar structures could represent former habitats which became progressively less suitable for *M. auratus* as water levels increased. The observed pattern of genetic divergence suggests that while Mazinzi Reef could have been colonized by migrants from Harbour Island, Nkhudzi Hills may have been colonized from now submerged habitats to the east or south.

The four shallow southern sites show lower levels of heterozygosity compared to the five deeper northern sites. This reduction in heterozygosity may suggest a series of bottlenecks associated with the colonization of newly available habitats. Alternatively, the reduced allelic diversity may simply reflect smaller populations at these shallower sites because both habitat area and population density are reduced relative to the adjacent deeper sites. A third possibility is that the southern populations were founded by a genetically depauperate population from deep water sites to the south and east of the study area and that the pattern of differentiation is more a result of the historical distribution of genotypes than of the topography of individual habitat patches.

How likely is it that observed migration estimates are merely an artefact of small sample sizes relative to the total number of alleles, or that they are a result of surveying a small number of loci? Resampling from a simulated population with an allele distribution similar to locus UNH-001 for all populations combined suggests that $F_{ST}$ estimates between small samples drawn from a single panmictic population are almost two orders of magnitude smaller than those observed between the most isolated *M. auratus* populations (Markert 1998). For samples of only 25 individuals, 95% of pairwise $F_{ST}$ estimates are < 0.007 and only the highest 5% of pairwise $F_{ST}$ estimates were found to be statistically distinguishable from zero using either the resampling algorithms of FSTAT or the exact tests for allele frequency distribution in GENEPOP. Recently, Ruzzante (1998) has demonstrated that sample variances in Weir & Cockerham's (1984) $F_{ST}$ estimator are

low relative to other estimators, that $F_{ST}$ is unbiased even at small sample sizes, and that the sampling variance in $F_{ST}$ actually decreases with an increasing number of alleles for a given sample size. The main advantage of increasing sample sizes would be to enhance our statistfical power, enabling the detection of smaller levels of genetic differentiation and finer discrimination among migration barriers.

Although adding loci would be expected to enhance the accuracy of the distance estimates, the Pearson correlation matrix indicates that the four loci surveyed in this study are generally concordant with respect to estimates of interpopulation differentiation. Single-locus $F_{ST}$ estimates and the statistical significance of allele frequency differences for adjacent pairs of populations are shown in Table 4. In the instances where interpopulation heterogeneity is detectable at only three of the four loci in adjacent populations, a single allele is present at a high frequency in both populations being compared.

The rapid evolution of the Lake Malawi cichlid flock may be driven by the fragmented and chronically unstable rocky habitats. The importance of habitat fragmentation and transience in the evolution of the Lake Malawi cichlid species was first emphasized by Trewavas (1947) and later elaborated by Fryer (1959b). For the mbuna, Fryer suggested that populations on isolated rocky outcrops are free to pursue independent evolutionary trajectories. The course of these trajectories may be set either by drift or by adaptation to local physical, social, or ecological conditions. These local conditions are modified continuously as a result of the frequent changes in water level. Ribbink *et al*. (1983) have suggested that fluctuations in water level may play a generative role in speciation by altering the size of rocky zones, exposing or drowning inhabited areas, and fragmenting or uniting habitat patches. Although this process may lead to local extinctions, a perpetually shifting evolutionary landscape could drive speciation with each newly available habitat patch possessing unique physical conditions, collections of fauna, and evolutionary challenges which may lead to evolutionary divergence among mbuna populations in the absence of substantial interhabitat gene flow.

## Conclusions

The low level of migration in mbuna species combined with the isolated nature of many of the rocky habitat patches within Lake Malawi provide many opportunities for evolutionary divergence. van Oppen *et al*. (1997) have suggested that mbuna species are divided into thousands of genetically isolated units and that this division provides numerous opportunities for allopatric speciation. The documentation of fine-scale population differentiation in *M. auratus* presented here is consistent with this suggestion.

Data are emerging which suggest that fine-scale genetic substructuring may be a general feature of mbuna biology. van Oppen *et al*. (1997) and Arnegard *et al*. (1999) have demonstrated high levels of population structure in five additional species using SSR loci. When combined with data from allozyme and mitochondrial DNA sequences (McKaye *et al*. 1984; Bowers *et al*. 1994), a widespread pattern of philopatry emerges.

Speciation is also likely to be influenced by the dynamic nature of Lake Malawi itself (*sensu*. Vrba 1985). In mbuna the speciation process is perhaps enhanced by the chronic instability of the rocky habitats along the shores of the lake. While rapid and frequent changes in lake level alter components of the physical environment, genetic and environmental differences between populations also develop. Colonization of a newly flooded habitat patch would be expected to be accompanied by stochastic changes in both allele frequencies (due to founder effects) and community structure. This may differentiate populations with respect to both the potential response to selection (as a result of allele frequency differences), and the selective environment itself (as a result of differences in community structure or the physical environment).

The other African Rift Valley Lakes have also experienced dramatic climatically driven changes in water level during the Pleistocene (cf. Scholz & Rosendahl (1988); Johnson *et al*. (1996)). If habitat fidelity is a general feature of the biology of the East African cichlids then philopatry and habitat instability may help explain the rapid evolution of biodiversity observed in the African cichlid species flocks.

## Acknowledgements

We wish to express our appreciation to Julie Baldizar, Karen Kellogg, and Jay R. Stauffer, Jr. (who provided underwater field support for our research), Aggrey Ambali and Harvey Kabwazi (for their advice and support in Malawi), and Janet Conroy and Woo-Jai Lee (who isolated the microsatellite loci used in this study). Christopher Bvalani, Wykliff Louis, and Timoth Mponda endured long hours of outfitting and transporting SCUBA divers during the field portion of this study. We also gratefully acknowledge the Malawi Parks Department for providing permission to work in Lake Malawi National Park (Permit no. 684658). This work was supported by grants from the National Geographic Society (J.A.M., T.D.K.), the National Science Foundation (J.A.M., T.D.K.), the Fulbright Commission (P.D.D.) and the Rotary Foundation (M.E.A.). Additional insights were provided by Ann C. Bucklin, C. Sarah Cohen, Irving Kornfield, Michelle Scott, Monica Stevens-Sommer and Lizhao Wu. Three anonymous reviewers and Louis Bernatchez provided useful comments on an earlier version of this manuscript.

## References

Allen PJ, Amos W, Pomeroy PP, Twiss SD (1995) Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of

genetic differentiation between two British breeding colonies. *Molecular Ecology*, **4**, 653–662.

Arnegard ME, Markert JA, Danley PD, Stauffer Jr JR, Ambali AJD, Kocher TD (1999) Population structure and color variation in the lithophilous cichlid *Labeotropheus fuelleborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proceedings of the Royal Society of London, Series B*, **266**, 119–130.

Barton NH, Slatkin M (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity*, **56**, 409–415.

Boulenger GA (1897) Description of a new fish from Lake Nyassa. *Annals and Magazine of Natural History*, **6**, 155.

Boulenger GA (1915) *Catalogue of the Fresh-Water Fishes of Africa in the British Museum (Natural History)*. British Museum (Natural History), London, UK.

Bowers NJ (1993) A revision of the Genus *Melanochromis* (Teleostei: Cichlidae) from Lake Malawi, Africa, using morphological and molecular techniques. PhD Thesis, The Pennsylvania State University.

Bowers N, Stauffer Jr JR, Kocher TD (1994) Intra- and interspecific mitcochondrial DNA sequence variation within two species of rock-dwelling cichlids (Teleostei: Cichlidae) from Lake Malawi, Africa. *Molecular Phylogenetics and Evolution*, **3**, 75–82.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Dominey WJ (1984) Effects of sexual selection and life history on speciation: species flocks in African cichlids and Hawaiian Drosophilia. In: *Evolution of Species Flocks* (eds Echelle AA, Kornfield I), pp. 231–249. University of Orono Press, Orono, USA.

Farrall M, Weeks DE (1998) Mutational mechanisms for generating microsatellite allele-frequency distributions: an analysis of 4,558 markers. *American Journal of Human Genetics*, **62**, 1260–1262.

Fryer G (1959a) The ecology and evolution of a group of rock-frequenting Nyasan cichlid fishes known as the 'Mbuna'. *Proceedings of the Zoological Society of London*, **132**, 237–279.

Fryer G (1959b) Some aspects of evolution in Lake Nyasa. *Evolution*, **13**, 440–451.

Fryer G (1965) Predation and its effects on migration and speciation in African fishes: a comment. *Proceedings of the Zoological Society of London*, **144**, 301–322.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use within microsatellite loci. *Genetics*, **139**, 463–471.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Greenwood PH (1964) Explosive speciation in African lakes. *Proceedings of the Royal Institute of Great Britain*, **40**, 256–269.

Günther A (1864) Report on a collection of reptiles and fishes made by Dr Kirk in the Zambezi and Nyasa regions. *Proceedings of the Zoological Society of London*, **1864**, 303–314.

Hellberg ME (1995) Stepping-stone gene flow in the solitary coral *Balanophyllia elegans*: equilibrium and nonequilibrium at different spatial scales. *Marine Biology*, **123**, 573–581.

Hill BJ, Ribbink AJ (1978) Depth equilibration of a shallow-water cichlid fish. *Journal of Fish Biology*, **13**, 741–745.

Jackson PBN (1955) A new fish of the genus *Clarias* Gronov. from Lake Nyasa, with notes on the distribution of the Clariidae and other catfishes in the lake. *Proceedings of the Zoological Society of London*, **125**, 681–684.

Johnson TC, Scholz CA, Talbot MR *et al.* (1996) Late Pleistocene dessication of Lake Victoria and rapid evolution of cichlid fishes. *Science*, **273**, 1091–1093.

Kellogg KA, Markert JA, Stauffer Jr JR, Kocher TD (1995) Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa. *Proceedings of the Royal Society of London*, **260**, 79–84.

Kosswig C (1947) Selective mating as a factor for speciation in cichlid fish of east African lakes. *Nature*, **159**, 604.

Lee W-J, Kocher TD (1996) Microsatellite DNA markers for genetic mapping in *Oreochromis niloticus*. *Journal of Fish Biology*, **49**, 169–171.

Lehman T, Hawley WA, Collins FH (1996) An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics*, **144**, 1155–1163.

Malawi Government (1977) Nankumba Peninsula Sheet 1434B2 & Part of 1334D4. Dept of Surveys, Blantyre, Malawi.

Markert JA (1998) *Population structure and phylogenetic history of the Lake Malawi cichlid species flock*. PhD Thesis, University of New Hampshire.

Marsh AC, Ribbink AJ (1981) A comparison of the abilities of three sympatric species of *Petrotilapia* (Cichlidae, Lake Malawi) to penetrate deep water. *Environmental Biology of Fishes*, **6**, 367–369.

McKaye KR, Gray WN (1984) Extrinsic barriers to gene flow in rock-dwelling cichlids of Lake Malawi: Macrohabitat heterogeneity and reef colonization. In: *Evolution of Fish Species Flocks* (eds Echelle AA, Kornfield I), pp. 169–181. University of Maine at Orono Press, Orono, USA.

McKaye KR, Kocher T, Reinthal P, Harrison R, Kornfield I (1984) Genetic evidence of allopatric and sympatric differentiation among color morphs of a Lake Malawi cichlid fish. *Evolution*, **38**, 215–219.

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1995–96) MICROSAT (Version 1.4d): a computer program for calculating various statistics on microsatellite allele data *WWW: http://lotka.stanford.edu/microsat.html*.

van Oppen MJH, Deutsch JC, Turner GF *et al.* (1997) Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes *Proceedings of the Royal Society of London, Series B*, **264**, 1803–1812.

Owen RB, Crossley R, Johnson TC *et al.* (1990) Major low lake levels of Lake Malawi and their implications for speciation rates in cichlid fishes. *Proceedings of the Royal Society of London*, **240**, 519–553.

Patterson HEH (1985) The recognition concept of species. In: *Species and Speciation* (ed. Vrba ES) Pretoria, S.A., Transvaal Museum, **4**, 21–34.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Ribbink AJ (1986) The species concept, sibling species and speciation. *Proceedings of the 3rd European Workshop on Cichlid Biology*. Musee Royal de L'Afrique Centrale, Tevuren, Belgium.

Ribbink AJ, Marsh AC, Marsh BA, Sharp BJ (1983) A preliminary survey of the cichlid fishes of the rocky habitats in Lake Malawi. *South African Journal of Zoology*, **18**, 149–310.

Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. *Genetics*, **140**, 1413–1419.

Ruzzante DE (1998) A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1–14.

Scholz CA, Rosendahl BR (1988) Low lake stands in Lake Malawi and Tanganyika, East Africa, delineated with multifold seismic data. *Science*, **240**, 1645–1648.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, **17**, 6463–6471.

Trewavas E (1947) Speciation in cichlid fishes of East African Lakes. *Nature*, **160**, 96–97.

Tripp RT, Bailey RT, Crosby A, Farrer R (1957) Chipoka to Monkey Bay.

Vrba ES (1985) Environment and evolution: alternative causes of the temporal distribution of evolutionary events. *South African Journal of Science*, **81**, 229–236.

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

---

This work is one of a series of papers which explore population structure and phylogeographic history in mbuna species along the shores of Lake Malawi's Nankumba Peninsula. Jeffrey Markert is a postdoctoral research associate who uses molecular population genetics to understand the processes which have led to the evolution of the taxonomic richness which characterizes the cichlids of the African Rift Valley lakes. This publication represents a portion of his doctoral dissertation at the University of New Hampshire. Matthew Arnegard is a PhD student whose interests include animal communication, species recognition, and the evolution of genetic structure in populations. Patrick Danley is a PhD student studying the genetic basis of reproductive isolation in cichlids. Thomas Kocher is Professor of Zoology at the University of New Hampshire, and is interested in genomic approaches to understanding the genetic basis of speciation and the evolution of species differences.

---

© 1999 Blackwell Science Ltd, *Molecular Ecology*, 8, 1013–1026