Spring 1998

Population structure and phylogenetic history of the Lake Malawi cichlid species flock

Jeffrey Alan Markert
University of New Hampshire, Durham

Population Structure and Phylogenetic History of the Lake Malawi Cichlid Species Flock

by

Jeffrey Alan Markert
B. A. Hampshire College, 1989
M. S. University of Vermont, 1995

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Zoology

May, 1998

This dissertation has been examined and approved.

Dissertation Director, Dr. Thomas D. Kocher,
Associate Professor of Zoology

Dr. Ann C. Bucklin,
Research Professor of Zoology

Dr. C. Sarah Cohen,
Adjunct Professor of Zoology

Dr. Irving Kornfield,
Professor of Zoology, University of Maine

Dr. Michelle P. Scott,
Associate Professor of Zoology

5/1/98
Date

Table of Contents

List of Tables

List of Figures

Abstract

Population Structure and Phylogenetic History of the
Lake Malawi Cichlid Species Flock

by

Jeffrey Alan Markert
University of New Hampshire, May, 1998

The cichlid species "flocks" which are endemic to the East African Rift Valley are characterized by frequent lineage splitting events which have led to the rapid evolution of high levels of taxonomic diversity. Changes in water level cause these habitat patches to be chronically unstable, hypothetically speeding the process of genetic differentiation through the combined effects of genetic drift and selection. Allele frequencies at four simple sequence repeat loci indicate low levels of gene flow in two rock dwelling fish species, *Melanochromis auratus* and *Labeotropheus fuelleborni*, collected from the Nankumba Peninsula in southern Lake Malawi. Small interruptions in habitat cause low, but statistically significant genetic differentiation among populations. The highest levels of interpopulation heterogeneity were observed between populations separated by deep troughs of open water. Differences in habitat usage cause the absolute magnitude of interpopulation heterogeneity to be higher among *M. auratus* populations than among *L. fuelleborni* populations. A correlation exists between allelic diversity at a locus and the relative age of a habitat, suggesting that mild bottlenecks are associated with colonization. Simulation studies indicate that the level of differentiation observed among these populations is unlikely to be merely an artifact of modest sample sizes and highly polymorphic loci.

Philopatry alone is not sufficient to drive speciation. Populations must become reproductively isolated as well. A series of mate choice experiments indicated that mate recognition is nearly perfect among the congeners *M. auratus* and *M. heterochromis*. When F_1 hybrid females were included in these experiments, they preferentially mated with hybrid males.

An estimate of the relationships among Lake Tanganyika and Lake Malawi species was obtained by surveying the genome for SINE (retrotransoposon) insertions by using a modified AFLP protocol which incorporates a SINE specific primer. The resulting phylogeny estimate was consistent with other molecular and morphological data sets for the older lineages in Lake Tanganyika, and indicated that the Lake Malawi species flock has a common ancestor with the Tanganyikan tribe Tropheini. Resolution among the Lake Malawi species was poor due to the incomplete lineage sorting which is characteristic of this extremely rapidly evolving lineage.

Introduction

The Great Lakes of East Africa are one of the few places that dramatically illustrate the dynamic nature of species, and that allow us to glimpse the mechanisms responsible for the generation of taxonomic diversity. These lakes contain well known "flocks" of fish species, most notably in the family Cichlidae, although other fishes in these lakes have undergone less dramatic radiations (Banister and Clarke 1980). Lakes Malawi, Tanganyika, and Victoria each contain hundreds of fish in the family Cichlidae, the majority of which are endemic to only one of these lakes (Greenwood 1991). More striking than the sheer number of species is the fact that many of these taxa have arisen quite recently. Lake Victoria was completely desiccated some 12,000 yr ago (Johnson *et al.* 1996). Lakes Malawi and Victoria have held water for millions of years but periodic variation in rainfall routinely causes fluctuations of hundreds of meters, leading to cycles of habitat destruction and creation (Fryer 1959; Ribbink *et al.* 1983; McKaye and Gray 1984). A well documented recent decrease in levels in Lake Malawi between the years 1500 and 1850 led to a decrease of 120 m (Owen *et al.* 1990). This decrease rendered much of the southern end of the lake dry land. Many of the sites which have now been reflooded contain extreme local endemics - fish taxa which are known only from a single site (Stauffer *et al.* 1997). Presumably at least some of these locally endemic species evolved *in situ* during the last century and a half. If this is true, then speciation has occurred subsequent to the first scientific explorations of the lake (Günther 1864)!

What forces are responsible for the rapid generation of this taxonomic diversity? A number of different models have been proposed to explain the species diversity of the African Great Lakes, and several of these are summarized in Table 1. These processes may roughly be divided into those that focus primarily on the role of environmental forces and those that focus of factors intrinsic to the cichlid lineage itself.

A number of authors have focused on the role of environmental factors unique to the E. African great lakes. Worthington (1937; 1954) suggested that speciation was aided by short tropical generation times, the creation of new niches as a result of water level fluctuations within the lakes, and a reduced number of large predators in some of the lakes. Water level fluctuations were also thought to be important by Trewavas (1947),

who suggested that fluctuations led to changes in breeding habitats which could eventually lead to phenotypic divergence. In a textbook summary, de Beaufort (1951) suggested that the sheer size of the great lakes generated a large number of available niches which facilitated evolutionary divergence. Brooks (1950) proposed a role for habitat fragmentation as part of an allopatric speciation process, an idea later formalized and expanded by Fryer (1959b) who combined several lines of speculation and natural history data to construct a microallopatric model for the rock-dwelling "mbuna" of Lake Malawi. Fryer suggested a *rassenkreis*-like arrangement of semi-isolated populations along the rocky shoreline. Under this model, the proximate causes of reduced gene flow among populations could be predation on fishes who stray from the shelter of the rocky habitat, stochastic elimination of intermediate populations in the *rassenkreis*, or changes in water level which subdivide continuous habitats.

Table 1

A history of proposed speciation mechanisms in the East African Lakes

Date - Authors	Mechanism
1921 - Regan	Trophic adaptations.
1935 - Worthington	Short tropical generation time, empty niches resulting from lake level fluctuations and absence of predators in certain lakes.
1947 - Kosswig	Selective mating and monogamy.
1947 - Trewavas	Adaptation to changes in habitat resulting from lake level fluctuations.
1950 - Brooks	Allopatry resulting from discontinuous habitat distribution
1951 - de Beaufort	Adaptation to available niches resulting from the sheer size and depth of lakes.
1955 - Jackson	Space available in lakes enables speciation
1959 - Fryer	A rassenkreis-like arrangement of semi-isolated populations, isolation possibly enhanced by predation on unsheltered fishes and/or elimination of intermediate populations, influence of lake level fluctuations on distribution of food
1974 - Liem	An adaptable pharyngeal apparatus enables the exploitation of many niches
1984 - Dominey	Sexual selection and extreme philopatry lead to development of different SMRS's at isolated sites.

In the 1920's Regan proposed that much of the cichlid radiation could be explained by adaptation to distinct trophic niches, an idea later expanded by Liem (1974) who suggested that the evolutionary flexibility of the cichlid jaw apparatus permits cichlids to expand into new trophic niches. A role for reproductive behavior in the speciation process was initially proposed by Kosswig (Kosswig 1947) who suggested that the monogamy he

believed to be common among cichlids was responsible for lineage splitting through the mechanism of selective mating. In 1984, Dominey suggested that runaway sexual selection combined with extreme philopatry might be responsible for lineage splitting in E. African cichlids paralleling then recent work on speciation in *Drosophila*. Under Dominey's model, runaway sexual selection leads to the evolution of a distinct Specific Mate Recognition System (SMRS) within an isolated population. Each population may become fixed for its own SMRS because low migration rates between populations prevent the spread newly evolved SMRS's to other populations or the dilution of an SMRS within a population. Indirect support was derived from the bright color patterns exhibited by the males of several E. African cichlids, and the model was eventually expanded to include the sand dwelling species of Lake Malawi, whose males construct elaborate breeding platforms (McKaye 1990).

The "synthetic" model proposed by Dominey incorporates several of the elements from earlier hypotheses, with the a strong emphasis on reproductive behavior and habitat fidelity. Under the Dominey's model, runaway sexual selection leads to the development of distinct SMRS's within each population. Low inter-population migration rates lead to the development of reproductive isolation because unique SMRS's which evolve within populations do not spread to other populations, nor are they diluted by immigrants to that population.

This model is compatible with a number of the hypothesized speciation drivers listed in Table 1, particularly those in which environmental factors restrict migration between populations. The most thoroughly studied example of biogeographic restrictions to migration in E. African cichlids may be the rock-dwelling cichlids of Lake Malawi, a group of fish often referred to by their Chitonga name *mbuna,* which are found in littoral rocky habitats throughout the lake (Fryer 1959a or b). The majority of their habitat is located near the shoreline, which is primarily an alternating series of rocky and sandy substrate (McKaye and Gray 1984)

Dominey's model makes two specific testable predictions:

- Gene flow between populations is limited
- Mate recognition systems are established, such that allopatric forms will select appropriate mates when brought into sympatry.

This dissertation consists of a series of related projects designed primarily to test the predictions derived from the Dominey model. Additional chapters were added to address issues which arose as a result of these efforts, or to establish a phylogenetic

context for these projects. Each chapter described below is intended to stand as an independent unit, yet all aim to enhance our understanding of the processes which have led to the evolution Lake Malawi cichlid flock.

Chapter 1:

Biogeography and Population Genetics of the Lake Malawi Cichlid Melanochromis auratus*: Habitat Transience, Philopatry, and Speciation*

This chapter presents an analysis of population structure in the widely distributed mbuna *Melanochromis auratus*. Population samples were systematically collected from an archipelago of habitat patches in the southern end of the lake. Four locus simple sequence repeat (SSR or microsatellite) genotypes were determined for each individual to estimate migration rates between habitat patches.

Chapter 2:

Biogeography and Population Structure in the evolution of the East African Cichlids: Evidence from DNA Fingerprinting

Recently, SSR loci have been used to analyze population structure in a number of East African cichlid species. In this chapter the studies are summarized and the broader evolutionary implications of these data sets are discussed.

Chapter 3:

Sample sizes needed to estimate population structure using highly polymorphic loci

Sample sizes in Chapter 2 averaged 37 individuals per population. The highly polymorphic character of these loci (~25-30 alleles observed at each locus) raised concerns about the possibility of spuriously high estimates of sample divergence. In this chapter, simulation techniques were used to explore the relationship between divergence estimates, sample size, allele distributions, number of loci surveyed.

Chapter 4:

Mate choice in two Malawian congeners

In order to test the strength of mbuna mate recognition, a series of free mate choice experiments was designed. These involved allowing males from two *Melanochromis* congeners to establish territories in either aquaria or pools containing standardized structural elements (flower pots, ceramic tiles). After territories were established, females of both species were introduced into the mesocosm. Fry and juveniles were collected and SSR loci were used to determine parentage.

Chapter 5:

An AFLP Based Method of Rapidly Detecting the Insertion of SINE Elements for use as Clade Markers

Relationships among taxa within Lake Malawi have been difficult to determine due to the recent and rapid evolution of these species. In this chapter, SINE element integration was detected at orthologous sites using a modified version of the AFLP protocol to estimate the phylogenetic relationships among taxa from both Lake Malawi and the hypothesized ancestral taxa from Lake Tanganyika.

Chapter 1

Biogeography and Population Genetics of the Lake Malawi Cichlid *Melanochromis auratus*: Habitat Transience, Philopatry and Speciation

Abstract

Migration rates among populations of the Lake Malawi cichlid *Melanochromis auratus* were estimated by surveying allele frequencies at four simple sequence repeat (SSR) loci among 10 populations from a 42 km stretch of habitat in the southern end of the lake. The data suggest migration rates among populations are in general quite low, with an among population F_{st} estimate of 0.14 ($p<0.0020$). A biogeographic survey suggests that the highest levels of genetic differentiation exist between populations separated by stretches of deep water, and that migration is common between populations separated by shallower water. Water levels in Lake Malawi have risen dramatically in the past few centuries creating many new *M. auratus* habitats. Reduced allelic diversity was observed at more recently created habitat patches, suggesting that genetic drift resulting from bottlenecks is associated with the colonization of new habitat patches. 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.

Introduction

The flocks of endemic fish species in the East African Great Lakes are well known examples of "explosive" cladogenesis (Greenwood 1964). The cichlid 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 explorations of the rift valley lakes by European naturalists and explorers (Günther 1864).

In the century following the first formal description of the first rift valley taxa (see Boulenger (1915) for early citations), many explanations for the rapid evolution of taxonomic diversity have been suggested. Explanations invoking selective mating and brood care (Kosswig 1947), adaptation to changes in habitat arising from changes in lake level (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 1959b) have been proposed. 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 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. An assessment of both the spatial scale of genetic differentiation and the identification of features of the physical environment which constrain population differentiation are important if we are to evaluate models offered to explain the speciation of 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. McKaye and Gray (1984) described three types of substrate in the nearshore habitats of Lake Malawi. In the southern end of the lake 73% of the shoreline habitat is gently sloping stretches of either bare sand or weed beds (*Vallisneria* sp.). The remaining 27% of the habitat consists of steeply sloping jumbles of boulders, rocks, and cobbles. The bottom of these rocky slopes intersects the flat sandy bottom which forms the bulk of the lake floor (Figure 1c). The shoreline is a mosaic of habitat types with rocky stretches existing as habitat islands separated by long stretches of sandy or weedy substrate.

Rocky habitats also exist along the shores of several small islands, and as completely submerged offshore rocky outcrops.

The configuration of shallow water habitats along the shores of Lake Malawi is not a permanent 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 (Owen *et al* 1990; McKaye and Gray 1984; Scholz & Rosendahl 1988; Fryer, 1959a). 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.

The distribution and continuity of habitats along the lake shore is influenced by the amount of water in the lake basin. Decreased rainfall can lead to the loss of rocky habitat patches when water levels fall below the rock-sand interface. 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 habitats depending on local topography (see McKaye and Gray (1984)).

The rocky areas are the primary habitat for a well-studied guild of small, brightly colored fish known collectively as *mbuna* (Fryer 1959a). The upper surfaces of these rocks in the shallows are covered with a thick biofilm, known as the *Aufwuchs,* which forms the bulk of mbuna diets (Reinthal 1990). The crevices between the rocks provide shelter from predators and serve as the focus of male breeding territories in many of these species (Fryer 1959a or b; Danley IN PREP). 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 heterogeneity - a probable precondition for lineage splitting - may be explored.

Genetic data suggest 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 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 1993) with their high mutation rates and consequently high allelic diversity provides a tool with the resolution needed to detect fine-scale population differentiation among recently established local populations (van Oppen *et al* 1997).

The newly flooded rocky habitats in the southern end of the lake provide an opportunity to examine gene flow, colonization, and migration of mbuna species. By combining fine-scale population sampling and a detailed survey of habitat distribution with an analysis of allele frequencies at SSR loci we evaluate the influence of biogeographic forces on population structure in the Malawi cichlid species flock. These data are used to make inferences concerning the role of biogeographic forces in the speciation process and the magnitude of the selective processes required for the evolution of phenotypic diversity.

Methods

Study Species

Melanochromis auratus (Boulenger 1897) is an easily recognized species that is widely distributed in the southern end of Lake Malawi, making it a good subject for investigations of population structure. It is a small (4-9 cm), sexually dimorphic fish. Females (Figure 1.1) have a bright yellow ground color with black or dark brown horizontal stripes. Males are slightly larger than females and have a dark brown or blue/black ground color with yellow-gold stripes (Bowers 1993). *M. auratus* show little systematic variation in color pattern from locality to locality although slight intrapopulation variation in color 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). They are almost never observed over sandy substrate, although Bowers (1993) suggests that they spend time foraging at the rock-sand interface.

Study Area

Melanochromis auratus were sampled from areas in the proximity of the Nankumba Peninsula which divides the southern end of Lake Malawi into two shallow basins (Figure 1.2 a & b). Because water levels were 121 m below their present level between 1500 and 1850, (Owen *et al.* 1990) most of the rocky areas in the southern end of the lake were dry land, providing an upper limit to the age of the habitat in this area.

The steeply sloping rocky habitats adjacent to the shores of the Nankumba Peninsula all intersect the flatter, sandy lake bottom (Figure 1.2 c). 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. Deeper habitats became available first while the shallower habitats became available more recently (See Ribbink *et al.* (1983); McKaye & Gray (1984); and Figure 1.2 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) (Figure 1.2b). 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 GPS unit and nautical maps (Malawi Government 1977; Tripp *et al.* 1957).

Sample Collection

A total of 372 individual *M. auratus* were sampled at 10 sites in southern Lake Malawi. Sample sizes at each locality are shown in Table 1.1. Fish were captured by SCUBA divers using monofilament gill nets. Divers usually worked within 50 m of each other to avoid possible complications from the Wahlund effect, with the exception of the site designated Shallow Reef. At this site, fish were collected some 800 m apart, at either end of a complex sprawling aggregation of rocky habitats in a sand/gravel matrix some 2-300 m off shore. This was necessary due to the relatively low densities of *M. auratus* at these sites which did not permit the sampling of an adequate number of fish at any single spot.

Tissue for this study consisted of fin clips (*ca.* 0.5 - 1 cm^2) obtained from one of the unpaired fins (for fish collected in Lake Malawi National Park) or from pectoral fins. Fish collected in Lake Malawi National Park were clipped and released per collecting permit no. # 684658. 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 *ca.* -15° C pending transport to the United States.

Locus Isolation & 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 in (Lee and Kocher 1996). All four loci are perfect dinucleotide repeats. Locus 231 was cloned from *Oreochromis niloticus*, the remaining loci were developed from an *M. auratus* library. Primer sequences, GenBank accession numbers, and fragment size ranges are provided in Table 1.2.

DNA Preparation & Amplification

DNA samples were extracted and amplified using the methods outlined in Kellogg *et al.*, 1995. Optimal PCR conditions were determined empirically, and a summary of primer sequences, annealing temperatures and locus characteristics is shown in Table 2. Samples were electrophoresed on a 6% denaturing acrylamide gel at 30 W on an ABI 373-A DNA sequencer for 8.25 hours.

Scoring and Binning of Alleles

GeneScan Analysis software (Applied Biosystems, Foster City, California) 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 which typically differed by two base pairs. To facilitate this process, 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. In instances where the limits of a bin were ambiguous, individuals at both extremes of that bin and from neighboring bins were re-run on a single gel to insure the integrity of the allele size estimates.

Detection of Null Alleles

The possibility of "null" alleles - alleles which 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) report 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 in which PCR products could be generated for only three of the four loci were assumed to be homozygotes for a 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).

Estimators of Between Population Heterogeneity.

Estimates of population heterozygosity, allele frequencies and F-statistics were estimated with the aid of Goudot's (1995) program FSTAT. This package estimates F-statistics using method of (Weir and Cockerham 1984), and calculates confidence intervals for these estimates using a resampling algorithm which permits both jackknifing among loci and bootstrapping among 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 adjacent populations. Confidence intervals were estimated by performing 5000 resamplings for the overall F_{st} estimate and 2000 resamplings for each of the pairwise comparisons.

Barton and Slatkin's (1985) rare allele based estimate of N_m was calculated using GenePop 3.1 (Raymond and Rousset 1995). This divergence estimator should be less sensitive to bias arising from the non-equilibrium state of these recently established populations.

Pairwise Nei's Distance [D_N] was calculated for all pairs of populations using the program Microsat [vers. 1.4d] (Minch *et al.* 1995; 1996). Nei's Distance was determined to be a more appropriate measure of divergence between these populations than Delta μ (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, it seems likely that the allele frequency distributions observed in the sampled populations are the result of recent historical sampling processes rather than post-divergence mutation, given the extremely recent origin of the habitats in the area surveyed (Owen *et al.* 1990).

Results

Distribution of Habitats

Habitat depth estimates and brief habitat descriptions are reported in Table 1.1. The distribution of habitat is shown in Figure 1.1 b. 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 1.4.

Allelic Diversity & Heterozygosity

A total of 30, 24, 21, and 29 alleles were observed at loci UNH-001, UNH-002, UNH-050 and UNH-231 respectively. The total number of dinucleotide repeats varied between 14 and 103 at locus UNH-001 and locus UNH-050 respectively. The average pooled heterozygosity is 0.671. These results are summarized in Table 1.2.

A strong positive correlation was observed between the maximum depth of rocky substrate at a site and the observed heterozygosity at that site ($r^2 = 0.803$, $p=0.003$), Figure 1.4.

The estimated frequency of null alleles within each population ranged from 0.0 to 0.31 (Table 1.3). These estimates are likely to overstate the frequency of null alleles because PCR reactions can fail for a variety of reasons other than primer incompatibility. True null alleles should be associated with an excess of homozygosity and might distort estimates of population differentiation. To test for an excess of homozygosity, F_{is} values were calculated. Populations with F_{is} values which are significantly different from zero are indicated in Table 1.3. Populations with high null allele frequency estimates did not necessarily have F_{is} values which were significantly different from zero. The impact of putative null alleles on the estimates of genetic divergence was assessed by jackknifing over loci. The different loci yielded similar estimates of genetic diversity, and the pairwise estimates of F_{st} values were typically narrow, suggesting that these nonamplifying alleles do not bias our conclusions. Mean estimates of genetic divergence and standard deviations based on jackknifing loci are shown in Table 1.4.

Population differentiation

Individuals sampled from the two sites 0.8 km apart in the extensive habitat designated shallow reef appeared to be from identical gene pools. Nei's distance between the two samples is 0.053 with a standard error of 0.046. Because of the genetic similarity

of these two populations and the continuous nature of the habitat these two subsamples were pooled to simplify subsequent analyses.

A high level of population structure is observed between populations. The overall F_{st} estimate is 0.14 (95% CI = 0.121 to 0.179). Pairwise population statistics and distance estimates between adjacent collection sites are summarized in Table 1.4. Pairwise F_{st} values between adjacent populations range from 0 to 0.15. Pairwise estimates of N_m between adjacent sites range from 1.71 to 5.19. The estimated pairwise N_m between the two terminal sites is 0.32. Nei's D estimates between adjacent sites range from 0.024 to 0.700 with a calculated value of 2.629 between the two terminal sites.

In general, the four loci generated congruent estimates of population differentiation. Table 1.5 shows F_{ST} estimates for all four loci between adjacent sites. Inter-locus F_{ST} estimates differ most between Nkhudzi Point and two adjacent sites, Mazinzi Reef and Shallow Reef. In both cases, F_{ST} estimates derived from locus UNH-001 and UNH-002 were much lower than those derived from UNH-050 and UNH-231. Null allele frequency estimates are high at this population for UNH-001, UNH-002, and UNH-050 (Table 1.3), and there is no obvious relationship between null allele frequency estimates and the magnitude of the F_{ST} values calculated.

Table 1.6 shows four locus F_{ST} estimates for all population pairs. As would be expected, the highest pairwise F_{ST} estimates were obtained between population samples obtained fromt the two terminal sites, Mumbo Island and Mphande Island.

Discussion

For speciation to occur, gene flow between incipient species must be low enough to permit them to follow independent evolutionary trajectories. Taxonomic divergence may arise either from the deterministic processes of selection, through the stochastic forces of genetic drift, or by some combination of these two forces. There are two patterns which have emerged from the analysis of locus frequency data in *Melanochromis auratus*. The first of these is an extremely high level of genetic differentiation among populations, and the second is a trend towards reduced heterozygosity at more recently available habitat patches. The extreme philopatry observed in *M. auratus* coupled with the patchy distribution of ephemeral habitats provides many opportunities for both selection and drift to contribute to the evolution of taxonomic diversity.

Melanochromis auratus in southern Lake Malawi show a surprisingly high level of population structure. An overall F_{st} value of 0.14 ($p<0.0002$) was observed among all sites along a 42 km transect. The estimated N_m between the two terminal sites is 0.32 migrants / generation. Barton and Slatkin's (1986) N_m is believed to be a more reliable index of population structure than F_{st} in this instance because the extremely recent availability of most of the habitat sites surveyed makes it unlikely that these populations have had sufficient time to reach a state of equilibrium. 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 artifact of a single major barrier to gene flow within the area surveyed.

The level of population differentiation is strongly influenced by the nature of the intervening substrate. 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 considerably lower levels of differentiation. The lowest N_m estimates occur between Mumbo Island and Ilala Gap which are separated by approximately 10 km of deep open water. The rock-sand interface occurs at 45 m at Mumbo Island and 36 m at Ilala Gap, and 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 Point. (Marsh and Ribbink 1981) and (Hill and Ribbink 1978) have demonstrated experimentally that other mbuna taxa are unable to control their buoyancy in waters greater than 40 m deep and that the maximum daily rate of depth acclimation for fish in the mbuna genus *Petrotilapia* is less than 4 m (Hill and Ribbink 1978; Marsh and Ribbink 1981) suggesting that substrate hugging mbuna are

physiologically incapable of crossing long stretches of deep water. Our data supports the hypothesis articulated by Ribbink (1986; 1983) that deep water can serve as a barrier to migration.

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 substrate at Mvunguti Village. These two sites are 8.2 km apart and show considerably less differentiation than the sites separated by similar stretches of deep 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 occurs between these sites.

Shallow sandy shoreline also appears to facilitate dispersal. Mphande Island is located in a shallow bay about 5.6 km south-east of Nkhudzi Point. The shoreline between the two sites is apparently free of classical mbuna habitat and yet the estimated N_m value of 4.52 suggests very little differentiation between these two populations. A similar pattern is observed between Mazinzi Reef and Shallow Reef which have the highest pairwise estimate of migration rate observed in this study ($N_m = 5.19$). This last observation contrasts with the low migration rates observed between Mazinzi Reef and other neighboring 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 Point (2.8 km vs 7.6 or 5 km respectively), it seems likely that other geographic features influence the facilitate 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, extending about 400 m out into the lake. 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. A similar series of submerged stepping stones might exist between Mphande Island and Nkhudzi Point. 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. Given the apparent ease with which *M. auratus* disperse across shallow stretches of sand, this observation underscores the recent founding of these habitats.

These results are consistent with mtDNA haplotype data (Bowers 1993), who found 5 haplotypes within *M. auratus* in southern Lake Malawi. Bowers' samples were collected from a more widely distributed set of habitat islands separated from each other by deep troughs. Three of the sites she sampled were on the Nankumba Peninsula. All the Nankumba populations were fixed for the haplotype AUR1. More distant sites possessed unique haplotypes, and a single site, Chidunga Rocks, an island about 10 km n.w. of Mumbo Island, possessed three haplotypes, among them AUR1. The presence of endemic haplotypes at three of Bowers' collection sites suggests that migration among these sites may be almost nonexistent, although her sample sizes for some of these sites were small enough that some haplotype diversity could have been missed.

The observed pattern of population differentiation permits inferences about the process of recolonization during the most recent refilling of the lake. This pattern may have implications important to the evolution of the species flock as well. For example, the high level of genetic divergence between the two apparent genetic units Mazinzi Reef-Shallow Reef and Nkhudzi Point-Mphande Island suggests that these two population pairs are more diverged than we might expect if Nkhudzi Point were colonized by migrants from Mazinzi Reef as water levels rose. Indeed, one would expect Nkhudzi Point to be about as similar to Mazinzi Reef as Shallow Reef is. One likely explanation for this discrepancy is that these populations belonged to separate genetic units *before* one or both their current habitats were founded. Several isolated sites in the lake could have served as refugia when lake levels were lower. 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.

At least two potential refugia exist in the area which 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 *ca.* 40 m below the lake surface at its shallowest point (personal observation). *M. auratus* is not known to exist at either of these sites currently, however 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 South-East arm of Lake Malawi are known to fishermen or are shown on navigational maps (Tripp *et al.*, 1957). Jerusalem Reef and other similar structures in could represent former habitats which became less and 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 Point was most likely colonized from now submerged habitats to the east or south.

In general shallower sites show reduced heterozygosity relative to sites with deeper habitat (Figure 1.3). This reduction in heterozygosity suggests a series of bottlenecks with each colonization event. The four southernmost sites illustrate this process particularly well. Mazinzi Reef is a completely submerged and isolated rocky habitat located about 2.9 km offshore in Madzidzi Bay. The Shallow Reef complex lies in shallow water just offshore some 2.7 km east of Mazinzi Reef. Mazinzi Reef is the deeper of the two sites with a maximum habitat depth of 13 m which compares to a maximum depth of about 3 m at Shallow Reef. Ribbink (1983) cites evidence that water levels were 7 m lower than they are now early in the twentieth century. This suggests that habitat at Shallow Reef became available very recently. As alluded to earlier, these sites are quite similar genetically. Pairwise D_N and F_{st} values calculated between these populations are not significantly different from zero, however the four locus heterozygosity at Mazinzi Reef is 0.57 whereas at Shallow Reef it is 0.499. Similarly, Nkhudzi Point with a habitat depth of 11 m has a heterozygosity of 0.569 whereas MPH with a habitat of 5 m has a heterozygosity of 0.500. Like Mazinzi Reef and Shallow Reef, pairwise estimates of F_{st} and D_n. are indistinguishable from zero.

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 (1983) has suggested that these fluctuations in water level may play a generative role in speciation as "fluctuations in water level would increase or decrease the size of rocky zones, expose or drown areas, and fragment or unite similar habitats.", thereby potentially accelerating the process of genetic differentiation among populations.

The frequent fluctuations in water level within the lake basin constantly rearrange the configuration of available habitat (*sensu* Ribbink *et al*, (1983); McKaye & Gray (1984). Because of this chronic habitat instability, many opportunities have existed for adaptation to distinct local environments and for changes in allele frequencies due solely to genetic drift. Limited migration between isolated habitats could permit adaptation to local conditions. Each newly available habitat patch possesses its own unique collection of fauna and it's own set of physical conditions. Each founding event might produce a new

combination of genotypes, causing each population to have a different potential response to selective forces.

Conclusion

This study documents population differentiation on an extremely fine scale in the Lake Malawi cichlid *Melanochromis auratus*. Data are emerging which suggest that this may be a general feature of mbuna biology. Arnegard *et al.*(IN PREP) and van Oppen *et al.* (1997) have demonstrated high levels of population structure in five additional species using SSR loci. When combined with data from allozyme and mtDNA sequences (McKaye *et al.* 1984; Bowers *et al.* 1994) a pattern of extreme philopatry emerges.

The low level of migration in mbuna species combined with the isolated nature of many of the rocky habitat patches within Lake Malawi provides many opportunities for evolutionary divergence. Van Oppen *et al.* (1997) have suggested that mbuna species are divided into thousands of genetically isolated units. This division provides numerous opportunities for allopatric speciation. The *Melanochromis auratus* data presented here are consistent with this suggestion.

Speciation is also likely to be influenced by the dynamic nature of Lake Malawi itself. In mbuna the speciation process is likely 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 social 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.* Johnson *et al.* and Scholz & Rosendahl 1988). 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.

Acknowledgments

This work would not have been possible without the support and assistance of my advisor, Thomas Kocher and of my collaborators, Matthew Arnegard and Patrick Danley. 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. This work was supported by grants from the National Geographic Society (JAM, TDK), the National Science Foundation (JAM, TDK), the Fulbright Commission (PDD) and the Rotary Foundation (MEA). Additional insights were provided by my dissertation committee, Ann C. Bucklin, C. Sarah Cohen, Irving Kornfield, and Michelle Scott.

Figures and Tables

Figure 1.1: A female *Melanochromis auratus* as illustrated in Boulenger (1915).

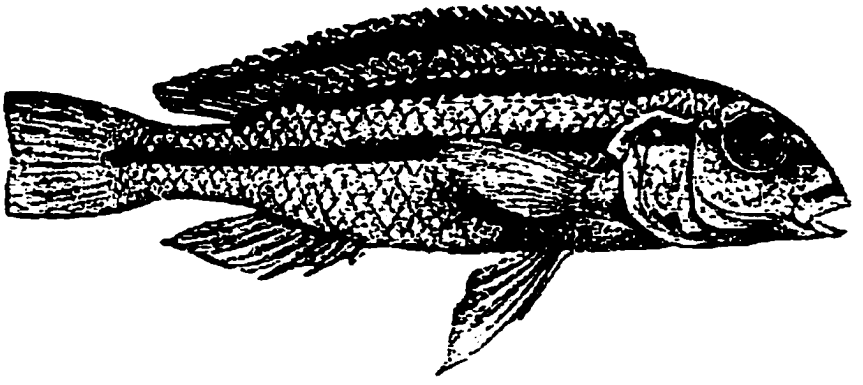


Figure 1.2: Habitat distribution along the shores of the Nankumba Peninsula.

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. 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 showing the typical arrangement of rocky habitat along the shoreline. Rocks and boulders (gray) 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 shallow water fishes.

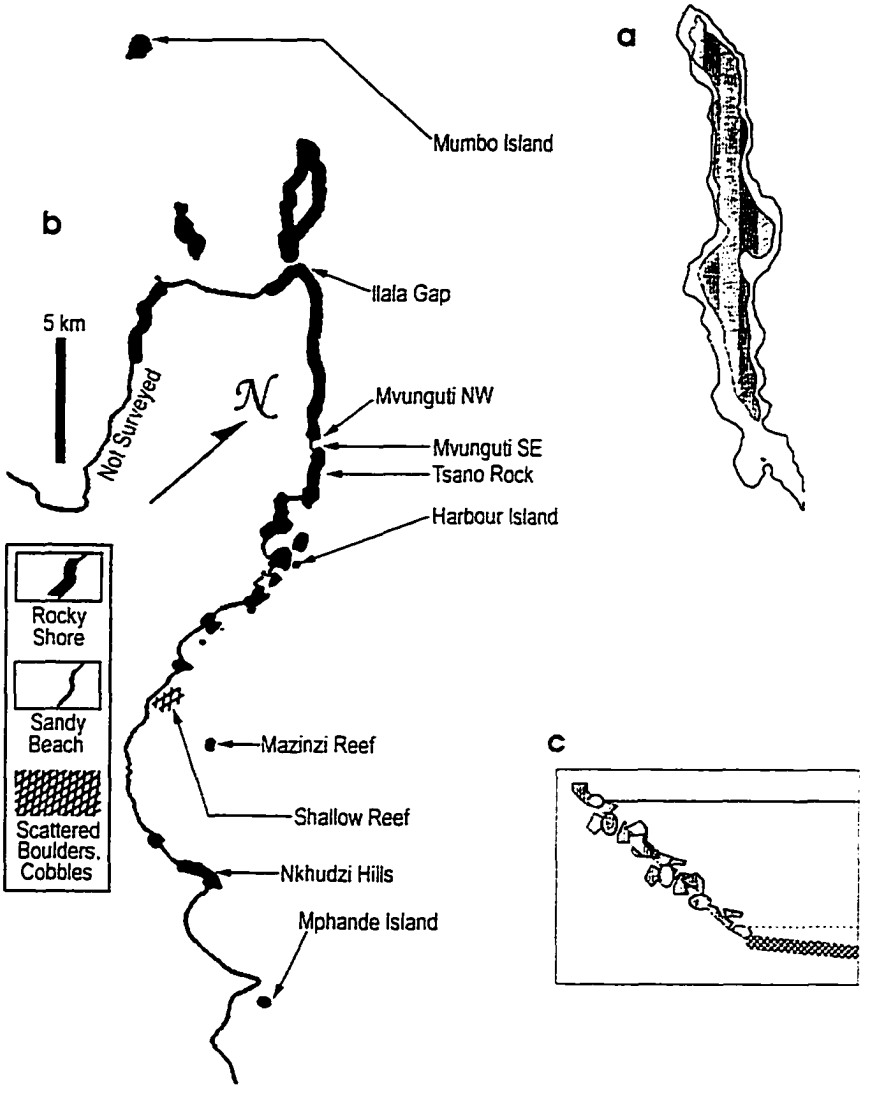


Figure 1.3: The relationship between habitat depth and average population heterozygosity.

Heterozygosity is highly correlated with the depth of the rock-sand interface ($r^2 = 0.803$, $p=0.003$), suggesting that deeper sites are older and more stable than shallower sites, and that colonization is achieved through a series of mild bottlenecks. The deepest site, Mumbo Island (inside square), has lower than expected heterozygosity, possibly a result of the long-term isolation of this site. The two points enclosed by circles represent Mazinzi Reef and Shallow Reef. Allele frequency and biogeographic data suggest that Shallow Reef was founded by migrants from Mazinzi Reef. This hypothesized colonization appears to have been accompanied by a decrease in heterozygosity.

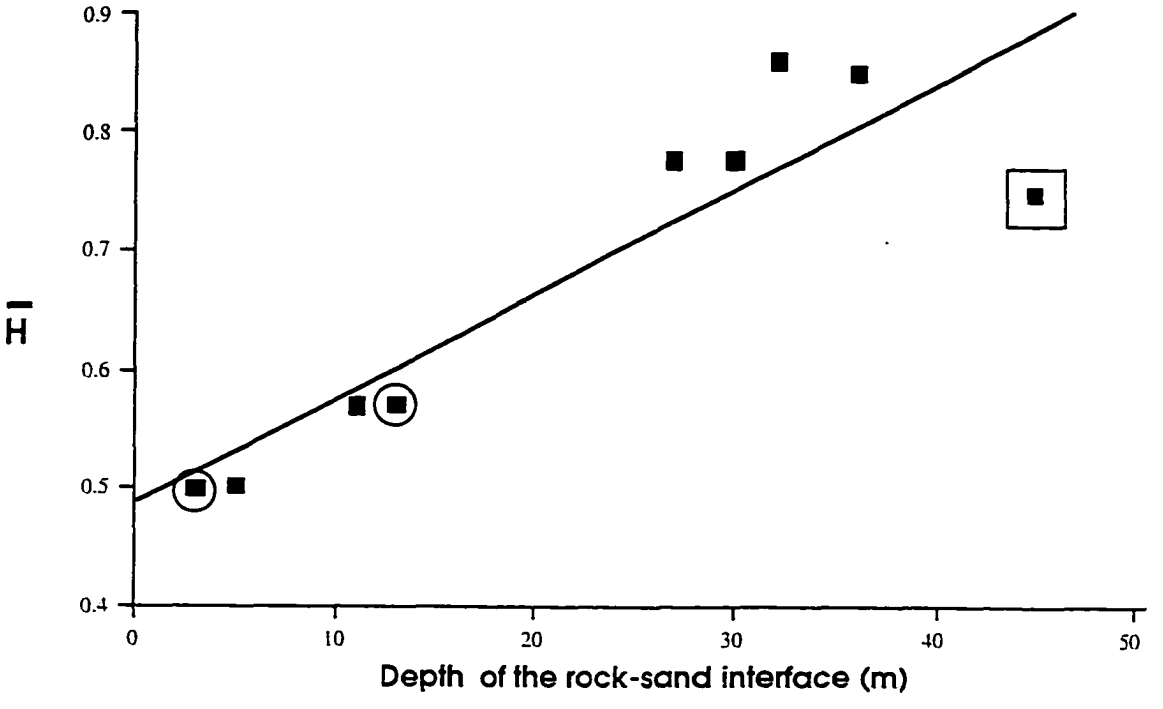


Table 1.1

Sample sizes and brief habitat descriptions for each of the collection localities sampled

Collection Site	Number of fish Sampled	Depth of Habitat	Description
Harbour Island	74	30	An emergent landmark immediately adjacent to the shore near Monkey Bay. The habitat adjacent to the island is primarily large rocks and boulders and is continuous with similar habitat on the peninsula. The 30 m wide channel between the island and the peninsula contains a mixture of lightly vegetated sand, gravel, and rock.
Ilala Gap	55	36	The northern end of the longest stretch of continuous rocky habitat in the southern end of the lake. The substrate is sediment free to a depth of 28 m, and consists primarily of very large rounded boulders which form many caves and crevices. Fish were sampled at the northern tip of the Nankumba Peninsula, about 0.4 km N.W. of the narrow channel which separates Domwe Island from the Nankumba Peninsula.
Mazinzi Reef	38	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	27	4	An isolated shallow habitat along lakeward side of Mphande Island consisting of sedimented cobbles, small (<1m diameter) boulders and some large (>3 m) boulders at the rock / sand interface.
Mumbo Island	37	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 - NW	4	27	The southern end of a continuous stretch of rocky habitat which extends northward toward Ilala Gap. Fish were collected just N.W. of Mvunguti Bay. A heterogeneous habitat composed of all size-classes of rocky substrate; sand and gravel are absent. A light sediment layer covers the rocks at this site.
Mvunguti - SE	12	36	A heterogeneous habitat just S. E. of Mvunguti Bay composed of all size-classes of rocky substrate; sand and gravel are absent. Below 21m a moderate sediment layer covers the rocks; at shallower depths, the rocks are covered by a lighter silt layer.
Nkhudzi Point	35	11	A heterogeneous habitat composed of all sizes of rocks and sand. Patches of open sand are present, and boulders and cobbles are often fairly widely spaced in a sand/gravel matrix. Rocks are covered with a moderate layer of silt below 2.5 m.
Shallow Reef	36	3	The substrate at this sprawling site is a sand/gravel matrix with many small patches of rocky habitat. Our collection efforts were concentrated at two rocky ledges about 0.8 km apart and 2-300m from shore. The gravel field extends lakeward for 0.4 km.
Tsano Rock	54	32	Tsano Rock is a huge (>40 m diameter)boulder located about 20 m from shore. The substrate in the vicinity is primarily rocky, lightly sedimented in sheltered areas, with small patches of sand and gravel in the shallow areas.

Table 1.2

Summary data for the four simple sequence repeat loci surveyed

Locus	Primer sequences	GenBank Accession #	Annealing Temperature	Repeat Range	Total # of alleles	Average Heterozygosity
UNH-001	GATTAACTCTGTCCCTGTCT CTGAAGTGTTAAAAATATTGTT	U17044	56° C	14-69	30	0.650
UNH-002	TTATCCCAACTTGCAACTCTATTT TCCATTTCCTGATCTAACGACAAG	U17045	54° C	19-48	24	0.549
UNH-050	GTCATCCCACTCACTAACAT AGAACAAACACAGGAAACTAT	AF036714	56° C	67-103	21	0.707
UNH-231	GCCTATTAGTCAAAGCGT ATTTCTGCAAAAGTTTTCC	G12382	56° C	51-87	27	0.777

Table 1.3

Maximum-Likelihood estimates of null allele frequencies for each locus within each population. These estimates were derived by counting the number of individuals with single locus amplification failures within populations. Because PCR amplifications can fail for a variety of reasons, these estimates represent maximum estimates of null alleles. True null alleles would be expected to cause heterozygosity to be below Hardy-Weinberg expectations. Null allele frequency estimates for loci within populations with F_{is} values which are significantly greater than zero are indicated with an asterisk. Several of the highest estimated null allele frequencies occur in populations which do not show decreased heterozygosity.

Collection Locality	Locus			
	UNH-001	UNH-002	UNH-050	UNH-231
Mumbo Island	0.09	0.16*	0.07	0.00
Ilala Gap	0.13	0.14*	0.04	0.09
Tsano Rock	0.02	0.10	0.12	0.08*
Harbour Island	0.10	0.18	0.08	0.10
Mazinzi Reef	0.13*	0.14	0.11	0.14
Shallow Reef	0.25*	0.10*	0.06*	0.00
Nkhudzi Point	0.23	0.13*	0.21	0.05
Mphande Island	0.18*	0.16*	0.17	0.31

Table 1.4

Pairwise population comparisons. All comparisons are from adjacent collection localities except for Ilala Gap / Tsano Rock and Mumbo Island / Mphande Island and the two terminal sites Mumbo Island / Mphande Island. The distance between collection localities was estimated by calculating the dispersal distance along the shoreline where appropriate. Nei's D values and standard errors were calculated with Microsat version 1.4 d (Minch *et al.*, 1995). Mean pairwise population F_{st} values and standard deviations were estimated with the aid of FSTAT (Goudet, 1995) and were estimated by jackknifing over loci. Significance values were estimated by bootstrapping (2000 replicates) and represent the probability that F_{st} is not ≥ 0 (Bonferroni corrected $\alpha = 0.0038$). N_m estimates were calculated using Barton and Slatkin's (1986) private alleles method which should be less sensitive to the extremely recent divergences of these populations. The N_m estimates were calculated using Genepop 3.1 (Raymond & Rousset, 1995). N_m could not be estimated between Shallow Reef and Nkhudzi Point due to a lack of private alleles between these two sites.

Collection Sites	Distance between collection points (km)	Intervening substrate	Nei's D (Standard Error)	Fst (Standard Deviation)	p Fst ≤ 0	Nm
Mumbo Island - Ilala Gap	10.4	Deep water, sandy lake bottom	0.700 (0.12)	0.104 (0.019)	0.001	1.71
Ilala Gap - Mvunguti NW	6.6	Rocky coastline	0.155 (0.10)	0.017 (0.018)	n.s.	2.56
Ilala Gap - Tsano Rock	8.1	Rock and a ~ 350 m sandy bay at Mvunguti Village	0.226 (0.04)	0.029 (0.003)	0.0005	4.10
Mvunguti NW - Mvunguti SE	0.6	~350 m sandy bay at Mvunguti Village	0.422 (0.16)	0.047 (0.026)	n.s	1.12
Mvunguti SE - Tsano Rock	0.9	Rocky coastline	0.072 (0.08)	0.013 (0.014)	n.s	5.83
Tsano Rock - Harbour Island	3.7	Rock and a sandy channel (> 24 m deep) at Monkey Bay	0.364 (0.11)	0.058 (0.013)	0.0005	4.46
Harbour Island - Mazinzi Reef	7.6	Open water, sandy lake bottom	0.335 (0.25)	0.095 (0.032)	0.0005	1.84
Harbour Island - Shallow Reef	7.7	Alternating sandy and rocky shoreline	0.307 (0.21)	0.113 (0.031)	0.0005	2.50
Shallow Reef - Mazinzi Reef	2.8	Shallow to deep sandy lake bottom	0.031 (0.04)	0.016 (0.014)	n.s	5.19
Shallow Reef - Nkhudzi Point	6.8	Sandy shoreline	0.241 (0.19)	0.140 (0.061)	0.0005	n.a
Mazinzi Reef - Nknudzi Point	5	Open water, sandy lake bottom	0.405 (0.32)	0.158 (0.081)	0.0005	1.81
Nkhudzi Point - Mphande Island	5.6	Sandy shoreline	0.024 (0.03)	0.150 (0.011)	n.s	4.52
Mumbo Island - Mphande Island	42.4	Open water, sandy shoreline, rocky shoreline	2.629 (0.93)	0.308 (0.079)	0.0005	0.32

Table 1.5

Pairwise F_{st} values for each locus. Average estimates which are statistically distinguishable from zero are indicated with an asterisk.

Collection Sites	UNH-001	UNH-002	UNH-050	UNH-231	Average F_{st} (Standard Deviation)
Mumbo Island - Ilala Gap	0.088	0.121	0.146	0.060	0.104 (0.019)*
Ilala Gap - Mvunguti NW	0.054	0.034	-0.032	0.011	0.017 (0.018)
Ilala Gap - Tsano Rock	0.037	0.028	0.020	0.032	0.029 (0.003)*
Mvunguti NW - Mvunguti SE	0.095	0.075	-0.025	0.042	0.047 (0.026)
Mvunuti SE - Tsano Rock	0.055	-0.005	-0.001	-0.004	0.013 (0.014)
Tsano Rock - Harbour Island	0.095	0.057	0.034	0.047	0.058 (0.013)*
Harbour Island - Mazinzi Reef	0.068	0.118	0.157	0.008	0.095 (0.032)*
Harbour Island - Shallow Reef	0.089	0.172	0.148	0.020	0.113 (0.031)*
Shallow Reef - Mazinzi Reef	-0.002	0.014	0.047	-0.005	0.016 (0.014)
Shallow Reef - Nkhudzi Point	0.003	0.018	0.248	0.18	0.140 (0.061)*
Mazinzi Reef - Nknudzi Point	0.003	0.027	0.335	0.152	0.158 (0.081)*
Nkhudzi Point - Mphande Island	0.013	-0.007	-0.009	0.033	0.015 (0.011)

Table 1.6

A matrix of four locus average F_{ST} values for all populations

	Harbour Island	Shallow Reef	Ilala Gap	Mphande Island	Mazinzi Reef	Mumbo Island	Mvunguti SE	Nkhudzi Point	Tsano Rock
Shallow Reef	0.1120								
Ilala Gap	0.1112	0.2292							
Mphande Island	0.1679	0.1497	0.2464						
Mazinzi Reef	0.0932	0.0156	0.1844	0.1711					
Mumbo Island	0.1631	0.3049	0.1038	0.3080	0.2641				
Mvunguti SE	0.0916	0.2510	0.0573	0.2788	0.2121	0.1582			
Nkhudzi Point	0.1414	0.1225	0.2171	0.0128	0.1513	0.2859	0.2453		
Tsano Rock	0.0584	0.1788	0.0291	0.2129	0.1437	0.0949	0.0133	0.1855	
Mvunguti NW	0.0668	0.2164	0.0171	0.2870	0.1457	0.0953	0.0471	0.2370	0.0007

Chapter 2

Biogeography and Population Structure in the evolution of the East African Cichlids: Evidence from DNA Fingerprinting

Summary

In the 130 years following the first European scientific exploriations of the E. African Rift Valley lakes, a number of models have been proposed to explain the rapid evolution of taxonomic diversity which characterizes the cichlid species "flocks" endemic to these lakes. The most comprehensive of these models were assembled by Fryer (1959) and Dominey(1984) who emphasized the roles of biogeography and sexual selection respectively. Anecdotal and experimental evidence has long been available which suggests a role for sexual selection. Recently, DNA fingerprinting techniques and systematic population sampling have produced data sets which quantify the role biogeography has played in the most extensive of the vertebrate adaptive radiations.

Introduction

The adaptive radiation of the E. African cichlid fishes is one of the most dramatic in vertebrate evolution. Estimates of endemic species within each of the great lakes range from the hundreds to over 1,000 (Poll 1986; Greenwood 1991). Geological evidence suggests that in Lakes Victoria and Malawi, this taxonomic diversity arose very recently. Lake Victoria was completely desiccated 12,000 years ago (Johnson *et al.* 1996), and sites in Lake Malawi harbor local endemic species in places which were dry land as recently as 150 years ago (Owen *et al.* 1990). If these species arose *in situ*, then speciation has occurred during within the last century and a half, in some cases perhaps *after* the first scientific explorations of the lakes in the mid-nineteenth century (Günther 1864). Even if it is ultimately demonstrated that the local endemic species in southern Lake Malawi evolved elsewhere and migrated to their current habitats, the evolution species in these lakes is undeniably rapid.

The role of sexual selection

The "synthetic" model proposed by Dominey incorporates several of the elements from earlier hypotheses, and places a strong emphasis on reproductive behavior and habitat fidelity. Under the Dominey's model, runaway sexual selection leads to the development of distinct SMRS's within each population. Low between population migration rates leads to the development of reproductive isolation because unique SMRS's which evolve within populations do not spread to other populations, nor are they diluted by immigrants to that population.

Until recently, the presence of sexual selection in E. African cichlids has been more thoroughly explored than the role of philopatry. Hert (1989), for example, has shown that the number of "egg dummies", small yellow spots on the anal fins of some fish, determines reproductive success in males. Direct evidence supporting the importance of color morphology in the SMRS has been provided by Seehausen *et al.* (1998) who demonstrate the importance of male coloration to female mate choice by manipulating ambient light, and a "natural" experiment in sister taxa from Lake Victoria which demonstrates that isolation by sexual selection may be breaking down as the water becomes increasingly turbid as a result of environmental degradation (Seehausen *et al.* 1997). Genetic support

for variance in male reproductive success has been provided by Kellogg *et al* (1995) and Parker and Kornfield (1996) who have demonstrated that female mbuna routinely use more than one male to fertilize their broods, suggesting an element of female choice. Sexual selection is also well supported in the sand dwelling fishes of Lake Malawi who have permanent leks and who build elaborate breeding platforms (McKaye 1990).

Population Structure and Biogeography - Evidence from SSR Loci

DNA fingerprinting techniques are now producing data which can be used to estimate the extent of population structuring in these fish. Highly polymorphic simple sequence repeat (microsatellite) loci have been used to quantify levels of migration among populations of 6 mbuna species. In the first of these papers, van Oppen *et al* (1997) used 6 SSR loci to estimate levels of gene flow in 4 species from sites near Nkhata bay, along the central-western shore of Lake Malawi (Figure 2.1). They found extremely low levels of migration among these fish as evidenced by low F_{ST} estimates among populations separated by less than 7 km. This striking result suggests population structure on smaller scales than had been expected. The low levels of migration observed among populations provides an opportunity for modest selection pressures alter phenotypes. Among the four species surveyed by van Oppen *et al* (1997), an overall estimate of 7 migrants for every 10,000 territorial males may be obtained. The evolutionary implication of this result is that moderate levels of selection (natural or sexual) could easily overcome the homogenizing effects of migration, leading to phenotypic divergence and/or reproductive isolation.

Recently, we have analyzed four locus SSR genotypes from two additional Lake Malawi species which illuminate the role that biogeographic forces play in the maintenance of population structure in both *Melanochromis auratus* and *Labeotropheus fuelleborni.* (Chapter 1 and Arnegard *et al.* IN PREP). We sampled populations of these mbuna species from an "archipelago" of rocky habitats in the vicinity of the Nankumba Peninsula which subdivides the southern end of the lake into two shallow basins (Figure 2.1). Populations of both species were systematically sampled from a number of sites along the eastern shore of the peninsula, the shores of several islands in the area, and a completely submerged rocky "reef" about 3 km from the shore.

The southern end of Lake Malawi is a special environment in which to study population structure because the habitat in this area has become available for colonization by mbuna very recently. The shoreline alternates between rocky and sandy stretches of

varying lengths. A xeric period which ended 150 years ago led to a decrease in water levels of 150 m. At its peak, the southern basins were dry land (Owen *et al.* 1990) (Figure 2.1). The rocky habitats along the shoreline which form the core of present day habitats typically slope at steep angle and eventually intersect the flat sandy lake bottom. The depth of this rock-sand interface determines the age and ultimately the temporal stability of the habitat.

The two species surveyed are both confined to the shallower portions of these rocky habitats, however they utilize the rocky substrate in different ways. *Melanochromis auratus* is a small (6-8 cm) fusiform fish that is present at depths between 0 and 40 meters (Boulenger 1897; Ribbink *et al.* 1983). In contrast, *L. fuelleborni* is a somewhat larger and rather stout fish which is rarely observed at depths >2m, and which is most abundant over wave washed rocks in the extreme shallows (Figure 2.2) (Ribbink *et al.* 1983). *L. fuelleborni*, with its subterminal mouth and robust bodyplan seems well adapted to scraping algae off the horizontal wave washed surfaces where it is most common. *M. auratus* on the other hand is perhaps more of a generalist, plucking algae off a variety of surfaces and venturing to greater depths. *M. auratus* is rarely encountered in the extreme shallows dominated by *L. fuelleborni* (Figure 2.2). Because the two fish species studied are in the same geographic area, we can use these data to infer the relative importance of various environmental forces on population structuring in cichlids and begin to understand the range of responses possible in these closely related but phenotypically distinct species.

As the lake basin refilled, sites with a deeper rock-sand interface became available for colonization sooner than sites with a shallower intersection. The depth of the rock-sand interface varies from site to site, ranging from less than a meter to over 45 meters. Sites with a deep rock sand interface became available for colonization sooner than sites with a shallow interface. *L. fuelleborni*, with its shallow habitat preference appears to be capable of colonizing rocky habitats as soon as they become available. Arnegard *et al* found *L. fuelleborni* at some sites with a rock / sand interface <<1 m. In contrast, *M. auratus* were rarely observed at sites with a rock / sand interface < 2 m. The level of population structuring in these species is probably related to the occupation of distinct niches in the rocky littoral region. F_{ST} values were typically much lower for *L. fuelleborni*, than for *M. auratus* populations from identical pairs of sites (Figure 2.3). These data suggest that *L. fuelleborni* is an early colonizer and is capable of becoming established as soon as a rocky outcrop becomes submerged. This capacity enables *L. fuelleborni* to use a number of tiny habitat patches as stepping stones between larger patches of habitat, which may facilitate gene flow among more distant sites. In contrast, *M. auratus* does not thrive in the extreme

shallows and cannot utilize these intermediate patches, leading to a higher level of genetic heterogeneity among populations.

Deep Water as a Migration Barrier

The lowest level of migration occurs between sites separated by deep troughs. In *M. auratus* the lowest pairwise migration rates (expressed as Barton and Slatkin's N_m (Barton and Slatkin 1985)) occur between Mumbo Island and Ilala Gap, two sites which are 10.4 km apart and separated by a trough at least 46m deep. Similar migration estimates were obtained between Mazinzi Reef and Harbour Island and between Mazinzi Reef and Nkhudzi Point. Mazinzi Reef is an isolated habitat 2.8 km from shore. At its highest point, the reef is 3 m below the surface of the lake and the rock-sand interface occurs at 13 m. The trough separating Harbour Island and the reef is at least 30 m deep.

Although the trough dividing Mazinzi Reef from Nkhudzi Point is shallower, migration rates from the reef are still very low. Marsh and Ribbink (1981) have shown that that these essentially benthic mbuna have a limited capacity to adjust to depth related changes in pressure. Pressure chamber experiments indicate that mbuna species can compensate for pressure changes equivalent to a depth change of about 4 m / day. Beyond that point, their swim bladders lose the capacity to regulate buoyancy. Mbuna attempting to cross the trough separating Mumbo Island and Ilala gap would require weeks to achieve the vertical component of the migration, wherease fish leaving Mazinzi Reef for Nkhudzi Point might require only a few days.

Other Ecological Forces Reducing Gene Flow

Deep troughs are not the only force limiting migration. Smaller, but statistically significant, genetic heterogeneity is also present between sites separated by shallow water or a stretches of sandy shoreline. In these case, predation may be operating to limit gene flow. Mbuna rely on the rocky substrate for shelter from predators. Trendall (1988) has shown that newly released fry are extremely vulnerable to predation and rely on the acquisition of a rocky shelter survive. We have observed a similar pattern in an unpublished pool experiment in which 9 sets of 3 11 x 11 cm bathroom tiles were arranged to form an 11 x 1 x 0.5 cm tunnel. Several dozen adult fish from the *Pseudotropheus* and *Melanochromis* genera were introduced into the pool. After several months, a dip net was used to recover juveniles present under these shelters. Each shelter harbored a single juvenile. No unsheltered juveniles were observed, suggesting that juveniles who could not obtain and defend a shelter were preyed upon by the adults in the

pool. Fryer (1959) has suggested a similar situation may exist within the lake itself. A number of large pelagic predators are known to cruise the rock sand interface, preying on mbuna who stray far from shelter and incidentally reducing the ability of these fish to migrate between habitat patches, in a manner similar to the adult mbuna in our pool experiment who opportunistically fed on juveniles.

Evolution of Phenotypic Diversity

Low migration rates among populations is a necessary but not sufficient component of speciation. Reduced gene flow must be accompanied by phenotypic change for permanent reproductive isolation to evolve. Given the difference in average gene flow in these two species, we might expect *M. auratus* to show more site to site phenotypic variation than *L. fuelleborni*. Surprisingly, the opposite pattern was observed. *L. fuelleborni* show interpopulation differences in both gular and fin coloration in our study area, and a range of color variation elsewhere (Ribbink, Marsh et al. 1983). A phylogenetic analysis of the *L. fuelleborni* populations surveyed suggests an explanation; the morphological differences may have evolved in the past when water levels were lower and the shallow stepping stone habitats which connect these habitats did not exist, and that the present situation is a result of secondary contact. A number of possible deep refuges exist in the area, including Boadzulu Island which approximately 12 km south of our study area and a number of offshore rocky "reefs". Although these reef structures are currently covered by several meters of water and do not currently support *Labeotropheus* populations, they would have been islands with rocky shores during xeric periods.

Water Level Flucturations and Evolution

In both the *L. fuelleborni* and *M. auratus* data sets a negative correlation between heterozygosity and habitat depth was observed. The reduced heterozygosity at shallower sites suggests the serial dilution of genetic diversity as newly available habitats are colonized by migrants from nearby deeper sites with a loss of alleles due to sampling. As the southern basins refill, the cycle is repeated compounding the loss of allelic diversity. Although SSR markers are believed to be selectively neutral, the differences in allele frequencies at these loci underscore the potential for the uneven distribution of alleles responsible for maintaining the SMRS's and other phenotypic variation, ultimately determining whether reproductive isolation or continuity exists among populations.

Water level fluctuations which lead to the creation of new habitats, the destruction of existing habitats, unification of isolated habitat patches or the division of continuous

stretches of habitat (and interruptions in Fryer's (1959) *rassenkreis*) could accelerate the process of speciation by driving stochastic changes in allele frequency (Ribbink *et al.* 1983; McKaye and Gray 1984). The chronic instability of these habitats may keep mbuna species in the "turnover" state described in Vrba's Turnover-Pulse hypothesis which suggests that phenotypic change is most likely to occur during periods of rapid environmental change (Vrba 1985).

Water level fluctuations may also lead to the creation of a variety of selective environments. Each time a new habitat patch becomes available, it is colonized by a different assemblage of species. Although many fish are common to many sites, almost every habitat patch has its own constellation of taxa and a unique assemblage of competitors. Further, changes in water level may drive the evolution of species on a habitat patch. For example, shallow water species living on the shores of a small island must adapt, migrate or perish as water levels increase.

As additional information becomes available, it is becoming clear that the E. African cichlid radiation is the result of a combination of environmental forces unique to the Rift Valley lakes and a number of forces intrinsic to the cichlid lineage itself. As data continue to accumulate, it will eventually be possible to determine which forces have been responsible for causing specific lineage splitting events.

Species Concepts and Malawi Cichlids

The Biological Species Concept (BSC) (Mayr 1963) defines species as assemblages of actually or potentially interbreeding populations of individuals, and speciation scenarios derived from it often involve the evolution of phenotypic divergence as a result of different selective pressures operating on allopatric populations. If barriers to migration fall, barriers to hybridization are expected to arise if hybrid offspring show lower fitness than the parent populations. The BSC is philosophically problematic in cases where allopatry exists and little interpopulation migration occurs. Such populations may potentially interbreed if migration barriers fall, even if no gene flow exists between them at the present time. The dynamic nature of the habitats within Lake Malawi suggests that the BSC cannot be too strictly applied within this system.

Ribbink (1986) has pointed out the utility of Patterson's (1985) "Recognition Concept" in this context. The evidence accumulated to date supporting the Dominey model are quite compatible with the recognition concept in which species are defined as genetically continuous units in which genetic cohesion is maintained by a common mate recognition system consisting of a variety of phenotypic characters which are responsible for

synchronizing and facilitating reproduction among the members of a species. The patchy distribution of habitats within Lake Malawi combined with modest levels of migration among them suggests that the recognition concept, which defines a species as a group of populations with a common SMRS, avoids the ambiguities of the BSC in which species are defined as actually or potentially interbreeding populations (Mayr, 1963). Under the BSC, we might be tempted to classify every population at an isolated habitat patch as a separate species, particularly if it can be demonstrated that gene flow to and from that patch are low. However, the unstable nature of many habitat patches in the lake suggests species defined in this way would be extremely ephemeral given the transient nature of many habitat patches. By relying on the Recognition Concept, with its emphasis on a common mate recognition system, this difficulty is avoided.

Figure 2.1: Lake Malawi past and present

a - During the most recent xeric period which ended 150 years ago, water levels within Lake Malawi were 120 m below their present level. Such a decrease in water level would reduce the lake to the gray area shown above, rendering the southern end of the lake dry land. b - Mbuna live in rocky habitats along the shore and on a few submerged rocky reefs. These rocky habitats slope steeply, eventually intersecting the flatter, sandy lake bottom (checkered area). When water levels fall (dotted line), the habitat becomes unusable or disappears completely. The depth of the rock-sand interface determines both the age and stability of rocky-habitat patches, especially in the southern end of the lake.

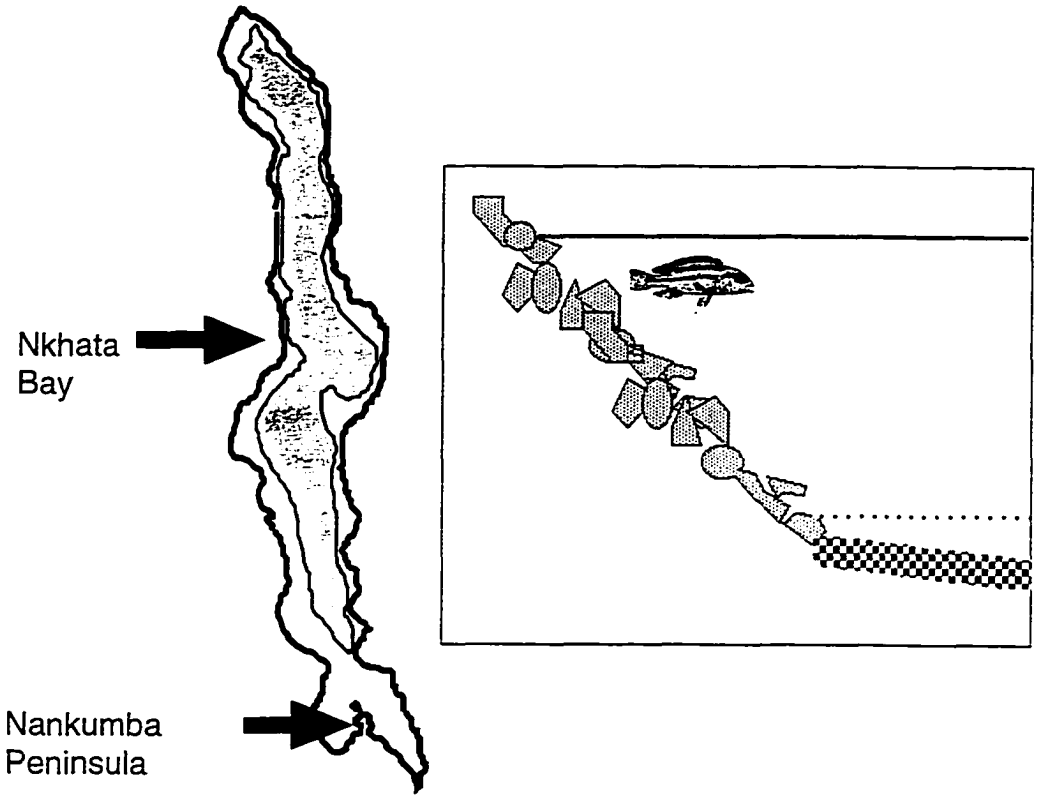


Figure 2.2: Niche partitioning in two mbuna

Although *Melanochromis auratus* (left) and *Labeotropheus fuelleborni* live in close proximity to each other in rocky habitats along the lake shore, population density data suggests that they are using different areas of the habitat. *M. auratus* inhabit a wide range of depths whereas *L. fuelleborni* are confined mainly to the shallows where their subterminal mouth and robust bodies helps them scrape algae of flat, wave-washed rocks in the shallows.

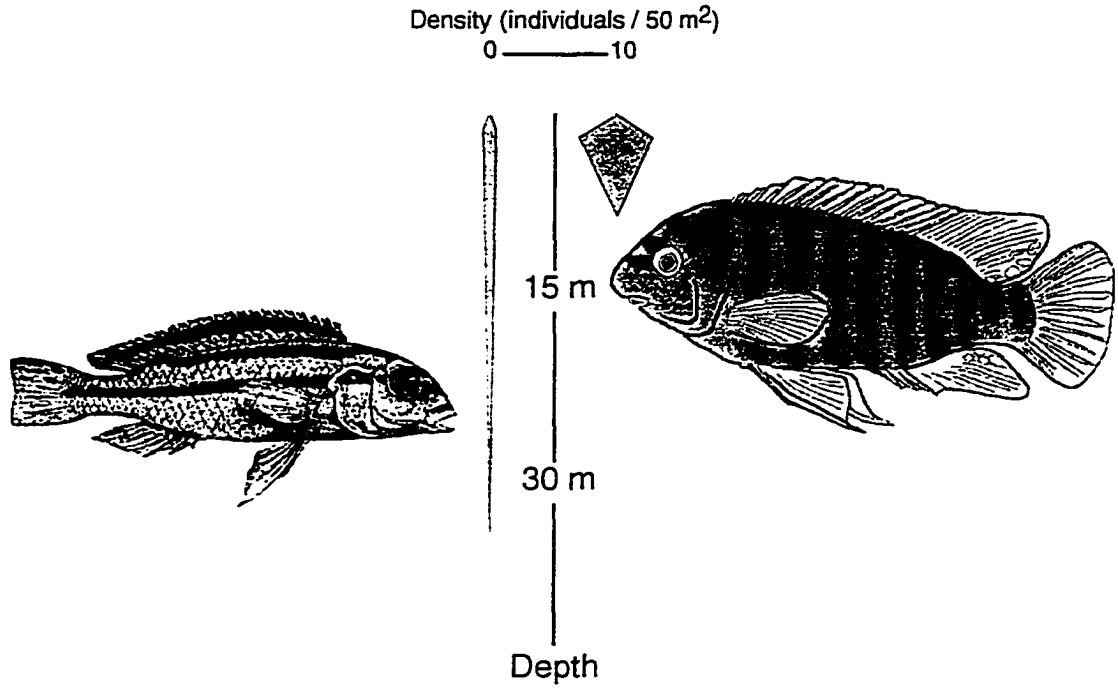
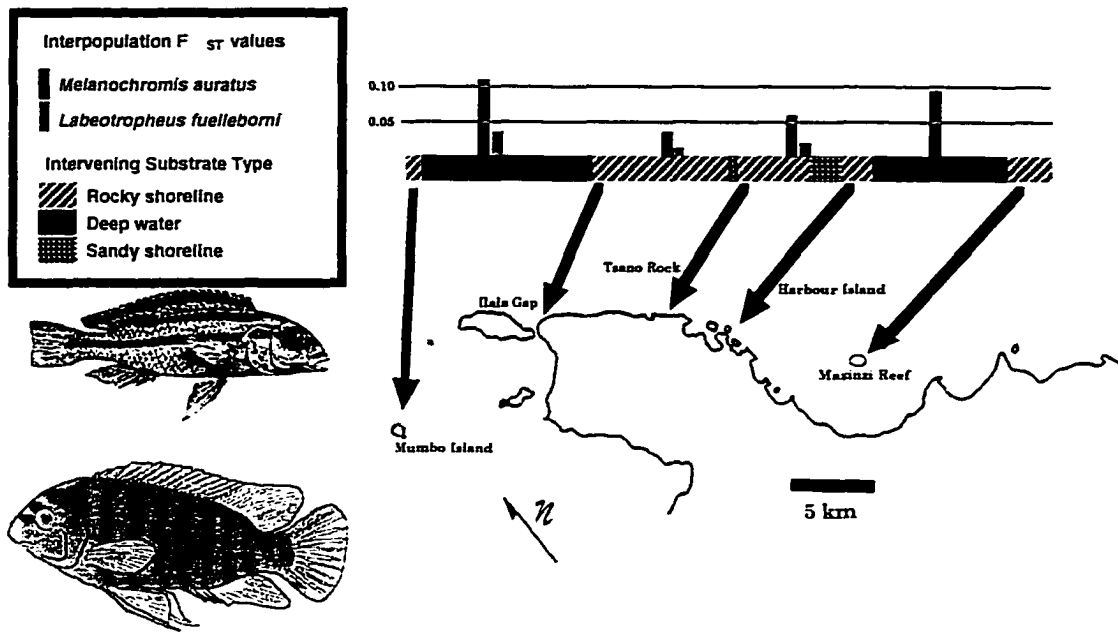


Figure 2.3: Barriers to migration.

The highest F_{ST} values were observed between population separated by deep troughs, although rocky habitats separated by shallower stretches of sandy shoreline also contain populations which are genetically distinct. In all cases, *M. auratus* populations are more differentiated than *L. fuelleborni* populations. Mazinzi Reef has few if any *L. fuelleborni*.



Chapter 3

Sample sizes needed to estimate population structure using highly polymorphic loci

Abstract

The large number of alleles found at some simple sequence repeat loci raises questions about the size of samples needed to analyze population structure. Here we present some simulations which quantify the accuracy and precision of two estimators (D_N and F_{st}) for different sampling schemes. Previous studies have shown that large numbers of individuals must be sampled at a large number of loci to accurately estimate the historical relationships among populations. In contrast, when the goal is to quantify population structure, we find that the number of samples needed to reliably estimate F_{st} and D_N is much smaller. Samples of 30 individuals scored for 5 microsatellite loci may be sufficient to estimate the magnitude of population differentiation in many species.

Introduction

Simple sequence repeat (SSR, microsatellite) loci are now widely used in studies of genetic divergence (Lehmann, *et al.* 1997; Allen *et al.* 1995; Paetkau and Strobeck 1994), pedigree analysis (Jones and Avise 1997; Kellogg et al. 1997), and, systematics (Kornfield and Parker 1997; Roy *et al.* 1994). It has been suggested that the high mutation rate at these loci makes them an ideal tool for estimating historical relationships among closely related taxa (Tautz 1989), and that they might be particularly useful in cases where allozyme or mtDNA markers are not polymorphic enough to resolve relationships (Kornfield and Parker 1997). The popularity of these markers has led to a number of simulation studies which address sampling issues for these highly polymorphic loci, with particular emphasis on requirements for phylogeny estimation. Relatively little attention has been paid to the separate issue of sampling requirements for the detection of population structure (Ruzzante IN PRESS).

This distinction is important. Phylogeny reconstruction requires reliable *rankings* of pairwise estimates of genetic distances between taxa. Any variability in the estimates which alters the rank of pairwise distances will alter the topology of the phylogeny. In the heyday of allozyme markers, Nei (1978) and Gorman and Renzi (1979) both suggested that surveying a large number of loci in a modest number of individuals produced estimates of genetic differentiation reliable enough to use for phylogeny estimation, although the practical utility of this approach was later seriously questioned by Archie *et al.* (1989). Recently, a number of authors have focused on similar issues for highly polymorphic SSR loci. Zhivotsky and Feldman (1995) and Takezaki and Nei (1996) have suggested that genotypes at hundreds of loci may be necessary for reliable phylogeny reconstruction. Further, Ruzzante (IN PRESS) has indicated that sample sizes of 50 to 100 individuals are required for reliable distance estimates and Kornfield and Parker (1997) have indicated that sample sizes on the order of 100 individuals may be required to accurately estimate genetic distances when 25 to 30 alleles are present at a locus. Together, these studies suggest that sampling requirements for highly polymorphic SSR loci are so restrictive that their utility as a phylogenetic tool may be limited to those rare instances where large numbers of individuals are available and large numbers of loci may be surveyed.

In studies of population structure, less stringent requirements apply. When the goal is merely to detect population structure, (deviations from the null hypothesis of no differentiation), accurate ranking of divergence estimates is not required. When the relative strength of different migration barriers is to be tested, the precision in distance estimates needs to be high enough to differentiate between barriers. If the hypothesized difference in the strength of these barriers is large, then moderate variability in distance estimates may be tolerable. Further, relatively small sample sizes may suffice to reject a null hypothesis of no differentiation.

In this report we quantify how variability in estimates of genetic differentiation is influenced by the number of individuals sampled, number of loci surveyed, and allelic distribution at these loci. We selected two widely used estimators of genetic differentiation, D_N (Nei 1987) and Weir and Cockerham's (1984) ANOVA based estimator of F_{ST}.

Methods

Two allele distributions were created. The first contained 30 alleles in equal frequency (the effective number of alleles $[n_e] = 30$). The second distribution contained 29 alleles in frequencies ranging from 0.006 to 0.137, with the four most common alleles representing 40% of the total . This distribution was based on the sample distribution observed by Arnegard *et al.* (IN PREP) for a population sample of 85 individuals from a single locality in southern Lake Malawi. The effective number of alleles for this distribution is 15.53, slightly more than half the value of in the first distribution.

Genotypes were simulated at thirty loci for population samples ranging from 15 to 100 individuals One hundred samples were drawn for each sample size from each allele distribution, permitting 50 independent pairwise estimates of the interpopulation parameters D_N and F_{ST} for each sampling condition. Estimates were calculated using the first locus in each individual, the first three loci, the first five loci, and so on up to 30 loci. This method was used to simplify calculations and, as Gorman and Renzi (1979) have suggested, represents a scientifically realistic situation in which additional loci are added to existing population samples. Sampling and calculations were carried out with the aid of Microsoft Excel 5.0 (Microsoft Corporation, Redmond, WA), GenePop version 3.1 (Raymond and Rousset 1995) and FSTAT version 1.2 (Goudet 1995).

In order to determine how the results of these simulations are influenced by the number of replications performed, an additional 500 pairwise F_{ST} estimates were calculated from sample pairs drawn from the *L. fuelleborni* distribution. The results are shown in Figure 3.3.

Because each sample in a pair was drawn from an identical distribution, F_{ST} and D_N values of zero are expected. The average value of these pairwise estimates indicates the level of accuracy obtained for each sampling condition, and the deviation from this average represents their precision.

Results and Discussion

Median values for F_{ST} and D_N are close to zero, regardless of sample size, allelic distribution, or number of loci sampled, with about half the estimates < 0 (Figure 3.1-A and -C). This is possible because, although these parameters by definition must be positive, the bias corrected estimators for them can be negative (*cf* Nei, 1987, pp. 224). On average, both estimators are accurate even under the most extreme conditions tested (15 individuals, 1 locus, n_e = 30), however, the range (precision) of the 50 pairwise estimates is strongly influenced by the size of the sample.

As would be expected, estimates calculated from smaller samples (of either individuals or loci) are more variable than those calculated from larger samples. Figure 3.1-A shows notch plots of 50 pairwise comparisons of F_{ST} estimates for different sample sizes. The trend is for the range of the pairwise estimates to decrease as sample size increases. Figure 3.1-B shows the relative improvement in precision as sample sizes are increased, in terms of the most extreme positive outlier (open squares), and in terms of the value of the 95th percentile (closed circles). Over the range of sample sizes tested, a doubling in number of individuals sampled decreases the value of the 95th percentile by about one half. D_N and F_{ST} respond almost identically to enhanced sampling and the distributions of the two estimators for a given sample size are concordant in their response to improved sampling. This is not surprising given the high correlation between these estimators.

Increasing the number of loci sampled also steadily decreases the variability in the estimates, but with steeply diminishing returns. Figures 3.1 - C and -D illustrate the effect of increasing the number of loci surveyed for a sample size of 25 individuals drawn from the *L. fuelleborni* distribution. The most dramatic increase in the precision results from the addition of the first few loci (Figure 3.1-C). The maximum variability is observed when a single locus is sampled. For five loci, the value of the 95th percentile is about half the value observed for a single locus. Twenty five additional loci must be surveyed to achieve another halving of the 95th percentile value.

Allele distributions also affect the reliability of these estimators (Figure 3.2). Estimates of D_N for samples drawn from the n_e = 30 distribution are more variable than those drawn from the n_e = 15.53 distribution. This is the expected result because smaller sample sizes are required to accurately estimate allele frequencies when the number of alleles present is small. In contrast, the variation in F_{ST} is higher for samples drawn from the *L. fuelleborni* distribution than for samples drawn from the distribution with 30 alleles present in equal frequency. The relationship is not obvious, but where the number of alleles in a sample is smaller than twice the number of individuals sampled, the variance in F_{ST} is proportional to the inverse of the square of the number of alleles (Weir and Cockerham 1984). For small sample sizes, as the number of alleles decreases, the variability of F_{ST} estimates increases. The impact of this effect decreases as sample sizes are increased (Figure 3.2).

There are two possible negative outcomes of inadequate sampling in population studies; population structure might be statistically indicated when no such structure actually exists (Type I error under a null of no differentiation), or existing structure might

not be detectable (Type II error). In the simulations presented here, the distribution of observed pairwise distances for a given sampling condition suggests the minimum level of detectable divergence when either permutation (FSTAT) or exact tests (GenePop) are used to determine whether two populations are statistically different. For example, for samples of 30 individuals at 5 loci, the most extreme pairwise F_{ST} value observed is 0.0072, and 95% of the values are below 0.0028. Values lower than the 95th percentile for samples drawn from the same distribution should in theory be statistically undetectable when permutation tests are used to determine statistical significance. The 95th percentile values obtained in these simulations suggest the minimum detectable levels of interpopulation genetic structure when the underlying allele distributions are similar to the one used here. For a sample size of 15 individuals, 5 loci, and ne=15.53, 95% of the values are below 0.11 or 0.007 for D_N and F_{ST} respectively. When the sample size is doubled to 30, 95% of the values are below 0.04 or 0.0049, a level of precision adequate to detect population structure in many species.

Two widely used analysis packages provide empirical support for this view, and both are conservative in rejecting a null hypothesis of no interpopulation differentiation. When the samples used to generate the most extreme observed F_{ST} value were reanalyzed using the permutation tests in the FSTAT package, using 1000 permutations of genotypes over populations to determine whether the F_{ST} value is significantly different from zero, the calculated F_{ST} value was found to be statistically significant ($p = 0.011$). The exact tests of GenePop, which determines the significance of interpopulation allele frequency differences, also found these two sample to be statistically significant at a probability level of 0.029. However, when the sample pair which generated the next most extreme value (F_{ST} = 0.0034, D_N = 0.045) was analyzed in the same way, no statistical significance was indicated. We can infer from this, that samples of 30 individuals at 5 loci with a distribution similar to that used here might be inadequate to detect differentiation below an F_{ST} value of 0.003, and that Type I error, manifested as spuriously high differentiation index values, is at an expected and tolerable level whether permutation or exact tests are used.

The values shown above apply only in cases where the allele distribution closely resembles the one used in these simulations, however this distribution is not unlike that found for dinucleotide repeats in other fish species (*cf* Ruzzante, IN PRESS or van Oppen, 1997). In cases where pilot studies have indicated that allele distributions are very different, spreadsheet based simulation studies can be a valuable tool in the design of sampling strategies.

Figures

Figure 3.1: Notch plots (Systat, Inc., Evanston IL) showing the relationship between sample size and the distribution of 50 independent pairwise estimates of F_{ST}.

The number of loci surveyed is 5 for all sample sizes and the samples were drawn, with replacement from an allele distribution equivalent to that found at locus UNH-001 in *Labeotropheus fuelleborni* at a single collection site. Notch plots show the median (central horizontal line), and central 50% of observations (area within boxes) The "whiskers" represent +/- 3 times the range between the hinges. Values outside this range are indicated with asterisks. The 95% confidence area for the median is indicated by the notch within the box. ***B*** - Number of individuals sampled vs. either the most extreme value observed in 50 pairwise estimates (open squares) or the 95th percentile value for the same set of estimates (closed circle). ***C*** - A notch plot showing the relationship between number of loci sampled and the distribution of 50 pairwise estimates for samples of 25 individuals drawn from the *L. fuelleborni* distribution. ***D*** - Number of loci surveyed vs extreme and 95th percentile values.

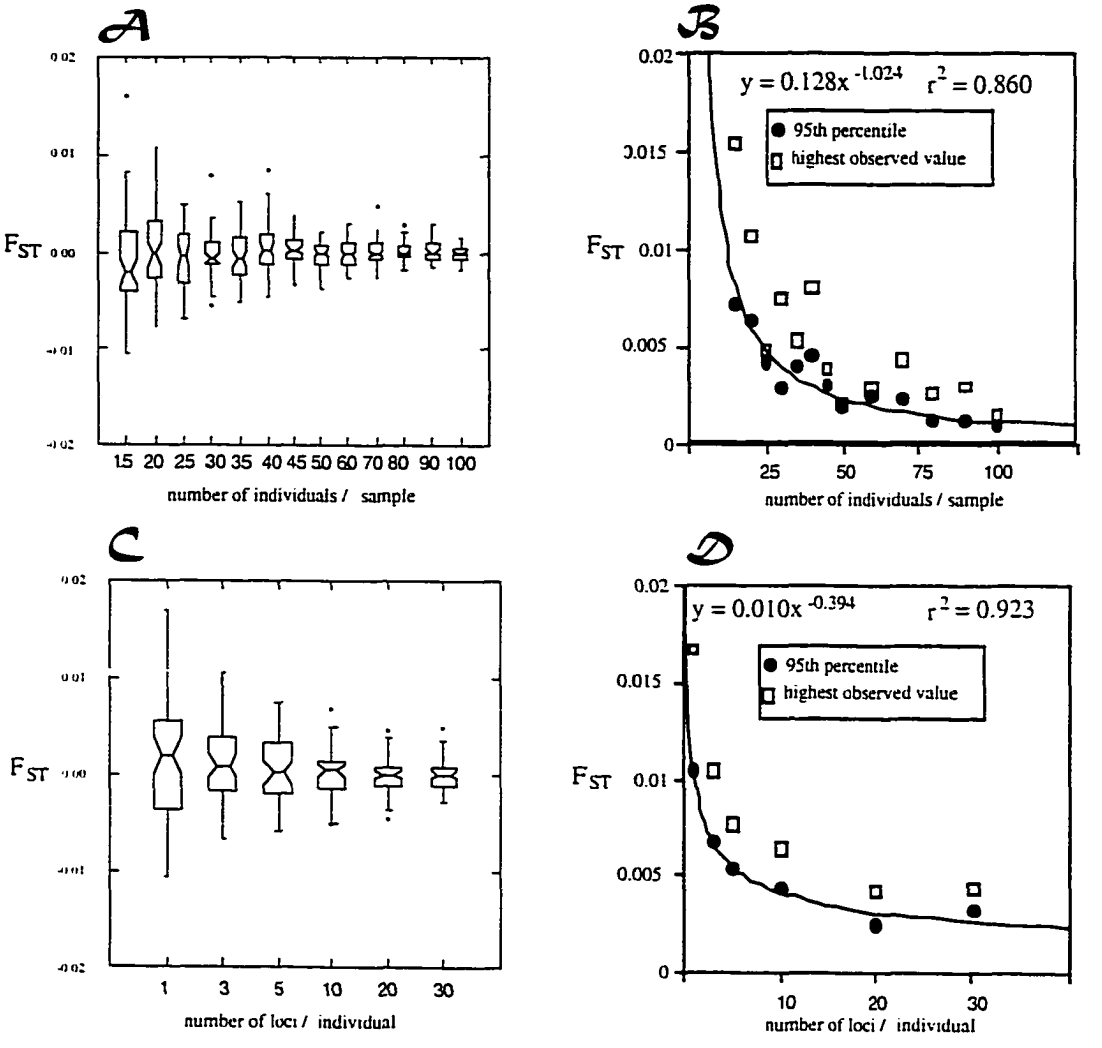


Figure 3.2: The variability of Nei's D and F_{ST} estimates is influenced by the underlying distribution of alleles.

Box plots for D_N and F_{ST} are shown here for sample sizes of 15, 30 or 50 individuals drawn from either the uneven *L. fuelleborni* distribution (shaded boxes) or a distribution in which 30 alleles are present in equal frequency (white boxes). Nei's Distance shows more variability for samples drawn from the distribution with the higher n_e. F_{ST} is more variable for sample pairs drawn from the *L. fuelleborni* distribution, although the difference becomes less pronounced as the sample size is increased.

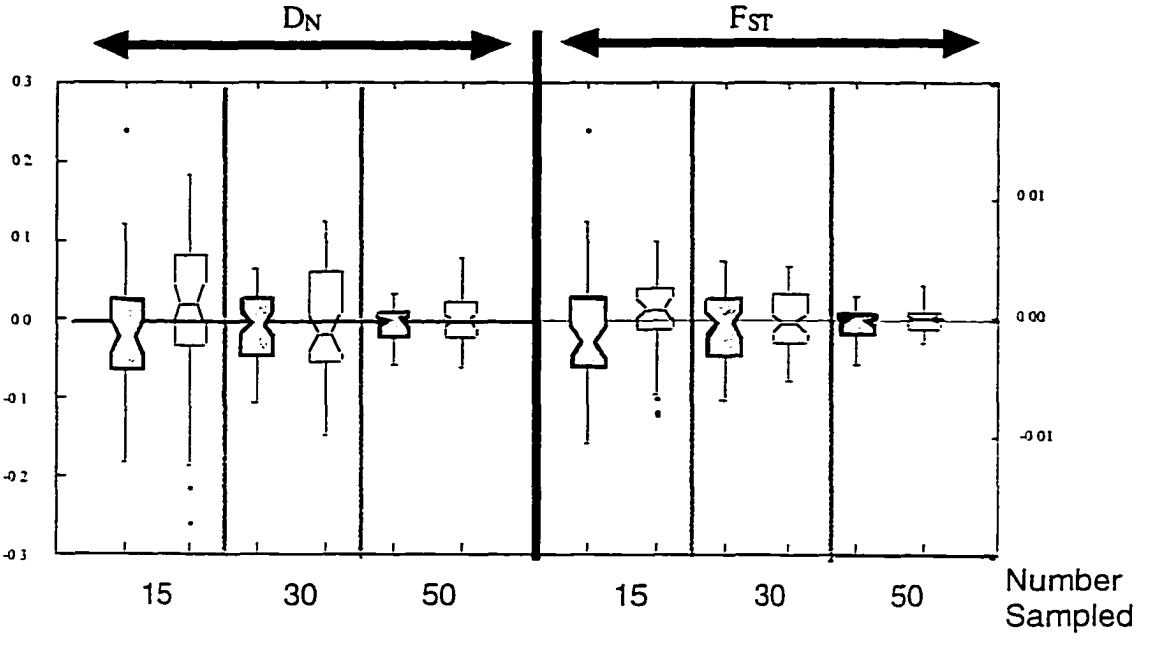
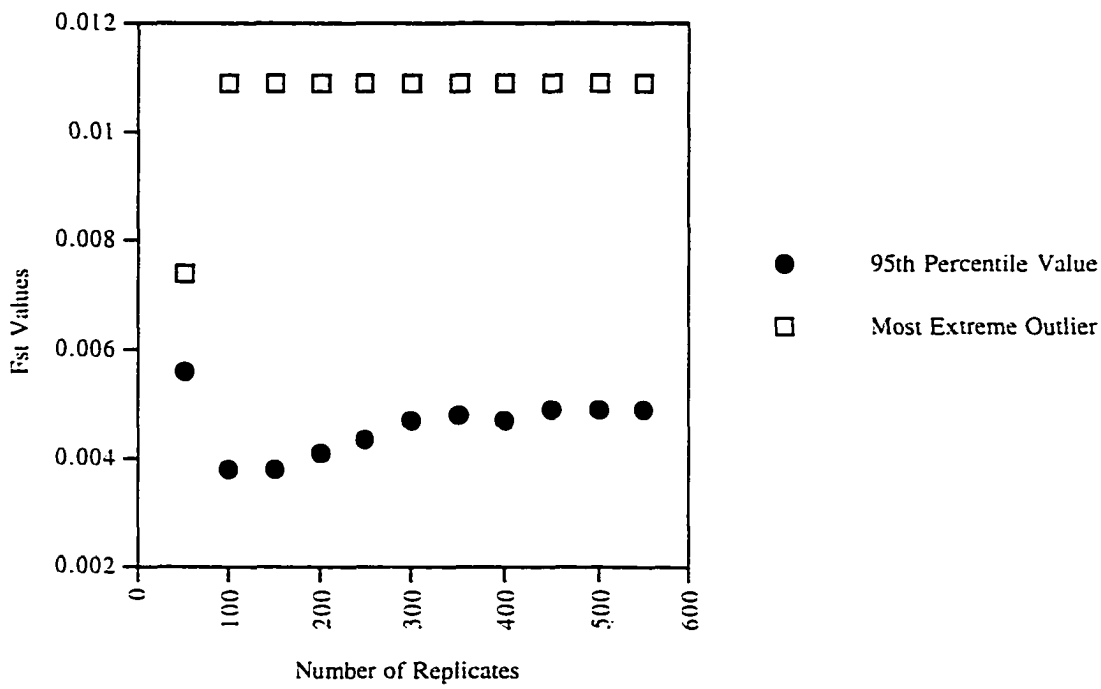


Figure 3.3

Replicate number vs 95th percentile value and the value of the most extreme outlier. In order to determine the effect of the total number of replicates on the F_{ST} estimates obtained in these simulation studies, an additional 500 pairs of 5 locus, 30 individual population samples were simulated. These additional samples were added to the original 50 pairs of samples in increments of 50 sample pairs; the data shown for 100 replicates includes the original 50 pairwise comparisons and an additional 50 comparisons. The F_{ST} value at the 95th percentile became stable after 350 individuals had been sampled at a value of approximately 0.0049. When the samples pairs which generated values of 0.0049 were analyzed using the permutation tests in FSTAT, they were found to be statistically insignificant at a p-value of approximately 0.06. This suggests that FSTAT is slightly less likely to reject a null hypothesis of no differentiation than might be expected under these conditions. The value of the most extreme outlier observed was 0.0109. This sample pair happened to occur in the second set of 50 sample pairs simulated.



Acknowledgments

This project benefited from the efforts of Karen Carleton who generously used her time helping me understand the impact of allele distributions on F_{ST} estimates. This manuscript was improved by incorporating comments from Matthew Arnegard, Ann Bucklin, Jason Curole, Karen Kellogg, Irv Kornfield, and Leslie Pray, who read earlier versions of this work.

Chapter 4

Mate choice in two Malawian Congeners

Introduction

The Dominey hypothesis (1984) holds that cichlid diversity is driven by the synergistic interaction between reduced migration and a rapidly evolving mating system. The model predicts that reproductively isolated populations develop unique mate recognition systems as a result of runaway sexual selection. If the sexually selected trait (and/or the preference for it) is initially a result of random mutation, then each population could wind up following its own distinct evolutionary trajectory.

Dominey identifies a set of population genetic conditions which are required for speciation to occur. These conditions include the evolution of distinct gene pools, genetic differentiation of lineages, and ultimately the establishment of reproductive isolation. Biogeographic and migratory forces which serve to establish the first two conditions were discussed in Chapters 1 and 2, but these factors alone are not sufficient for speciation to occur. Genetically differentiated populations become distinct species only when reproductive isolation is established. Mutations which alter the Specific Mate Recognition System (SMRS) (Patterson 1985) can lead to the evolution of reproductive isolation between groups and ultimately speciation.

Dominey (following West-Eberhard (1983) and Thornhill and Alcock (1983) reasoned that the components of the SMRS which are most evolutionarily pliable (and therefore more likely to cause rapid speciation) are those associated with sexual selection. The plasticity of sexually selected traits is a consequence of the variance in reproductive success. Under Dominey's model, runaway sexual selection leads to the rapid establishment of reproductive isolation, and the model assumes that the runaway process is started by random mutations and possibly by rare-male effects.

Circumstantial evidence for the importance of sexual selection abounds in the cichlids. The best documented in Lake Malawi involves sand dwelling species who build elaborate breeding platforms. Field experiments and observations have demonstrated that the size, shape and position of this extended phenotype determines male reproductive success (McKaye 1990). Many species show variation in color morphology throughout their range (*cf.* Ribbink *et al.* 1983). Seehausen *et al.* (1998) have recently documented a number of red/blue species pairs - sibling taxa which are presumed to have developed divergent mate recognition systems based on color. This observation was tested experimentally by changing the ambient light conditions so that red and blue fish were indistinguishable; under these conditions, species-specific mate recognition broke down.

Documentation of the scale of differentiation in mate recognition systems is critical to understanding the high rate of lineage splitting which characterizes this system. Ideally, mate choice experiments would be conducted on the populations surveyed in Chapter 2. Because live specimens from these populations were not readily available, I decided to test the relative strength of mate recognition for two congeners, *Melanochromis auratus*, and *M.* cf *heterochromis* (Bowers and Stauffer 1997) Because these species are sympatric, almost perfect mate recognition would be expected (hybridization is believed to be rare in the wild). The frequency of incorrect matings in a laboratory experiment can be used to estimate the baseline error rate resulting from an artificial environment.

Methods

M. cf *heterochromis* (Bowers and Stauffer 1997) and *M. auratus* are comparably sized rock-dwelling fishes from Lake Malawi. *M. auratus* males have a distinct horizontal gold stripe on a deep blue to black ground color. *M. heterochromis* are lighter blue to brown and occasionally have a pale blue horizontal stripe. In the laboratory, *M. heterochromis* appear to be less aggressive and outgoing, often retreating to the back of their aquaria when humans are present. Recently handled *M. auratus* exhibit a similar behavior, but they are less shy of humans after only a few days.

Two different types of mate choice experiments were performed. In the first type, the ability of females to detect conspecifics in smaller aquaria (50 gallons) was tested. In the second set of experiments, large (440 gallon), round pools were used. In the pool experiments, both species were present and *M. auratus* x *M. heterochromis* hybrids were included in the experiments. The strategy for both experiments was to place two or three males into a "mesocosm" which was outfitted with a shelter for each male which could serve as the focus of a breeding territory. These shelters were either 4" clay flowerpots or cubes constructed of 12" ceramic floor tiles and which were open on three sides. Shelters were arranged symmetrically in the aquarium so that the territories appeared to be equivalent with respect to light, food distribution, level of disturbance etc. Laboratory raised fish were used these experiments. In all cases, these were first generation fish derived from wild stocks.

In the first set of experiments, one male from each species was placed in a 55 gallon aquarium. The fish were then allowed one week to establish territories. After 1 week, three females of each species were introduced into the aquarium. After approximately one month, adults and progeny were removed from the pool and two locus SSR genotypes were generated for all individuals in order to determine maternity and paternity using the methods described in Chapter 2.

The general methods of the second experiment were similar to the first except that three males were introduced into a 440 gallon pool. In addition to *M. auratus* and *M. heterochromis* males which were hybrids between two species were also included. After one week, 9 female fish were added to the tank; 3 *M. auratus*, 3 *M. heterochromis*, and 3 hybrids. This experiment ran for several months. Fry and juveniles were periodically netted from the pool.

Results

A total of 117 juveniles were scored from 7 independent experiments. The species of both parents could be determined in 112 of these. Both species almost always mated with conspecifics. In the 50 gallon aquaria, only one of the two species present mated successfully. In 440 gallon pools, only two of the three types present mated successfully. Surprisingly, hybrids also mated primarily with hybrids. The results are summarized in Table 4.1.

Within the 50 gallon aquaria, all fry present appeared to be from a single brood. The pool experiments were run for several months, and females had sufficient time to produce multiple broods. The fry surveyed in mesocosm # 7 appeared to have been the offspring of single pairs, even though they may represent multiple broods. This is also the case for the hybrid offspring in mesocosm 6. Because two of *M. auratus* females in this pool shared alleles, it is not possible to reach a similar conclusion for this species.

Discussion

The fact that two locally sympatric congeners demonstrate nearly perfect mate recognition is not surprising. It is not clear whether the mismatings observed here represent a laboratory artifact or whether low levels of hybridization occur in nature.

In the 4 of the 5 two way aquarium experiments, only *M. heterochromis* were observed to breed, in the fifth aquarium only *M. auratus* individuals bred. Individuals within an aquarium were carefully matched for size and age, but it is possible that *M. heterochromis* mature at a smaller size than *M. auratus*. In contrast, the pool only 3 *M. heterochromis* progeny were produced. The majority of the offspring were pure *M. auratus* or hybrid x hybrid offspring. In addition, it appears that only a single female of each species was breeding in each mesocosm.

These results suggest that some aspect of the laboratory environment may be preventing certain individuals from mating at all. The densities of fishes in vessels used in this study are much higher than they are in nature, so it is possible that dominant individuals are preventing subordinates from breeding, perhaps by dominant males preventing subordinate males from displaying to females, regardless of their species. If this is the case, then the apparently "timid" *M. heterochromis* males may in fact establish dominance over *M. auratus* males.

The tendency of hybrids to mate with each other has also been observed in other cichlids (Crapon de Caprona 1986), and the explanation for this phenomenon is unclear. In sticklebacks, (Hatfield and Schluter 1996) hybrid males have an intermediate level of reproductive success, that is their reproductive success is equal to their

Table 4.1

Results of Mate-Choice Experiments Conducted in 7 Mesocosms. For experiments performed in 50-gallon aquaria, only the total number of progeny produced of each type is shown. Complete parental genotypes and numbers of fry produced by each parent are shown for experiments performed in 440 gallon pools. (a) - indicates fry for which maternity not be unambiguously assigned.

	Mesocosm #		Available males	
		Females	*M. auratus*	*M. heterochromis*
50 gallon aquaria	1	*M. auratus*	12	0
		M. heterochromis	0	0
	2	*M. auratus*	0	0
		M. heterochromis	0	4
	3	*M. auratus*	0	0
		M. heterochromis	0	16
	4	*M. auratus*	0	0
		M. heterochromis	0	7
	5	*M. auratus*	0	0
		M. heterochromis	0	8

440 Gallon Pools		Male Genotypes @ Loci UNH-001&231		
		Hybrid 1212/0202	*M. het.* 0720/1725	*M. aur.* 1313/0219
Mesocosm #6 Females	Female Genotypes @ locus UNH-001/231			
Hybrid	0126/0919	18		
Hybrid	1217/0924	16		
Hybrid	1621/0420			
M. het.	0421/2526			
M. het.	1315/2230			
M. het.	1619/2534			
M. aur.	1417/0215			
M. aur.	1429/0122			16[a]
M. aur.	1429/0122			
Mesocosm #7 Females		Male Genotypes @ Loci UNH-001&231		
		1717/1824	0104/0211	1315/2020
Hybrid	0117/0222			
Hybrid	0117/0219		1[a]	
Hybrid	1226/0219			
M. het.	0000/0317		3	
M. het.	0101/1016			
M. het.	0618/1220			
M. aur.	1315/2140			
M. aur.	1321/2020	1		10
M. aur.	1819/2121			

frequency in the population. In these experiments, hybrid females preferentially mated with hybrid males over either of the parental types. There are several explanations for these results which cannot be differentiated at the present time. Many of these explanations represent confounding factors which must be carefully considered in the design of future mate-choice experiments.

1) Hybrids have a novel SMRS composed of aspects of each parental SMRS. Female hybrids are choosing a unique assemblage of phenotypic traits which is only complete in hybrid males. This pattern suggests that the traits involved in the SMRS are dominant traits, and that they may be different loci in *M. heterochromis* and *M. auratus*.

2) Juvenile hybrids develop their SMRS through imprinting. The fish used in were raised with siblings in a small aquarium. If a search image for mates is developed during this time (possibly visual, chemical or acoustic), then hybrids would be expected to develop a SMRS specific for other hybrids purely as a result of culture conditions. If this postulated imprinting occurs after juveniles are released from their mother's mouths, then rearing fry together might reduce this effect. Such an experiment would require careful genotyping of the parents of the broods because these species are difficult to distinguish when these fish are immature.

3) Reproductive asynchrony among taxa is also a possibility. The fish used in this experiment were mostly newly mature individuals. If some of the females used in this experiment were not fully mature then they might not breed, causing the majority of the progeny produced to be derived from a single taxon. This could explain both the nearly perfect mate choice observed in these experiments and the tendency of hybrids to mate primarily with other hybrids.

The appearance of a reproductively dominant female also complicates the interpretation of these results. One possible explanation within the 440 gallon pools is that many several females mated but only the first brood to be released survived. Shelter within the pool was limited to the areas near the tile breeding territories, and it has been noted that in the wild shelter is limiting and larger juveniles readily exclude smaller individuals from available shelter (Trendall 1988). Alternatively, a dominance hierarchy may exist among females within a species.

Future experiments should be designed which can control for some of the factors which may be generating artifacts in these experiments, specifically, mesocosms must be either larger, less densely stocked, or more structurally complex to reduce male-male contact, dominance effects, and predation on juveniles. Additional care must be taken to ensure that all specimens within an aquarium are reproductively mature and that female

brood production is synchronized, and experiments must be undertaken to address the role of imprinting in the development of the SMRS.

Chapter 5

An AFLP Based Method of Rapidly Detecting the Insertion of SINE Elements for use as Clade Markers

Abstract

Phylogeny estimation among the cichlid fishes of the East African Rift Valley is challenging due to the extremely recent and rapid divergence of many of these taxa. In order to survey a large number of cladistically informative nuclear markers characters, AFLP methods were combined with a primer specific for a transposable element in the cichlid AFC - SINE family. Because SINE elements do not move after they are integrated into the genome, the position of this insertion can serve as a cladistic character with two states determined by the presence or absence of a SINE element. By using this method, we were able to obtain a phylogeny estimate for Lake Tanganyika cichlids which is similar to estimates obtained by conventional morphological and molecular methods. A close relationship between the Tanganyikan tribe Tropheini and the Malawian cichlids was observed however method was not capable of resolving relationships among the extremely recently diverged Lake Malawi species flock.

Introduction

The fish family Cichlidae is often cited as one of the most dramatic adaptive radiations in the history of vertebrate evolution (Fryer *et al.* 1972). This family contains over 1500 freshwater fish species in tropical Africa, America, and parts of Asia . The majority of this taxonomic diversity can be found in the lakes of the African Rift Valley (Ribbink 1991). It has been estimated that the three largest lakes, Malawi, Tanganyika, and Victoria may contain more than 1,200 species, the majority of which are endemic to a single lake (Ribbink 1983; Meyer 1993; Seehausen *et al.* 1997). Historical data on water level fluctuations within these basins provide evidence for the most striking aspect of the E. African Cichlid radiation, its extreme recency.

Fluctuations in rainfall have caused water levels in all of these lakes to vary dramatically. Lake Victoria, the shallowest of these lakes, was completely desiccated 12,400 yr b.p. (Johnson *et al.* 1996). Lakes Malawi and Tanganyika are much deeper, but xeric periods sill cause dramatic decreases in water level (Scholz and Rosendahl 1988). Lake Malawi experienced a 200 m drop in water level between 500 and 150 yr b.p (Owen, *et al.* 1990). If these species evolved *in situ,* then speciation in these lakes occurs on historical, rather than geological time scales.

Within Lake Malawi, the frequency of lineage splitting events makes phylogeny estimation using traditional genetic markers difficult because alleles do not have time to become fixed within a species before the next lineage splitting event occurs, leading to the "retention of ancestral polymorphisms" in descendant species (Moran & Kornfield 1993; Parker and Kornfield 1997). The scarcity of phylogenetically informative characters within Lake Malawi, complicates other inter- and intralacustrine phylogenetic analyses and has led to the use of a variety of traditional and more novel molecular tools. One hypothesis which is generally well supported by the available data is that Lake Tanganyika is a reservoir of lineage diversity (Nishida 1991), and that taxa within this lake are ancestral to those in both Lake Victoria and Lake Malawi. Nishida used allozyme frequency data to construct a phylogeny which suggested the presence of an "H" (Haplochromine) lineage within the lake which is ancestral to the species flocks in the other great lakes. Data from sequenced RAPD bands (Sültmann *et al.* 1995), and mtDNA sequences (reviewed in Meyer 1993) generally support the hypothesis that a Lake Tanganyika lineage is ancestral to the flocks in the other two lakes, but they differ as to which of the lineages within Lake Tanganyika is the actual ancestor.

Recently, (Takahashi *et al.* 1998) have demonstrated that the insertions of Short Interspersed Nuclear Elements (SINEs), a type of retrotransposon, into genomes may be a

useful tool for estimating phylogenies in East African cichlids. They argue (following Murata *et al.* 1996)) that the presence of a SINE element within a specific orthologous locus can be a powerful tool for phylogeny reconstruction because these transposable elements are not excised from the genome, and they are inserted apparently at random. Thus the presence of a SINE element at an orthologous site in two different taxa can be interpreted to be a true synapomorphy (Cook and Tristem 1997). The character in this case is a specific site in the genome with two possible stated defined by either the presence or absence of a SINE element at that site. Insertion of a SINE element at a specific site in a genome is a rare event, therefore, in cases where two taxa share a SINE insertion, it can safely be assumed that the this insertion is identical by descent as it is extremely improbably for insertion to occur independently in exactly the same place. Using this reasoning, Takahashi *et al* (1998) were able to determine the monophyly of several of Poll's (1986) tribes. Due to a lack of insertions within Lake Malawi taxa they were unable to determine which lineage within Lake Tanganyika is ancestral to the Lake Malawi flock.

The methodology involved in cloning SINE elements, (Murata *et al.* 1996) determining the sequence of flanking regions, and using PCR to detect the presence of SINEs at a specific locus is fairly tedious, and might prevent the widespread adaptation of the use of SINEs as clade markers. However, by replacing cloning with portions of the AFLP (Vos *et al.* 1995) protocol, it is possible to obtain information on the location of several SINE integrations simultaneously.

Methods

Specimens

DNA extractions were prepared using standard techniques. DNA samples were obtained from 68 individual fish in 32 taxa from Lake Malawi and Lake Tanganyika (Table 5.1). Single representatives were available for each Lake Tanganyika species and multiple individuals were available for many of the Lake Malawi fish.

SIFLP

The AFLP technique is a DNA fingerprinting method typically used to generate a set of "anonymous" DNA fragments which may be used as map markers or for use in paternity studies (Vos *et al.* 1995). The procedure involves digesting genomic DNA with two restriction enzymes to produce a population of DNA fragments cut with one or both of the restriction enzymes. DNA adapters containing PCR primer sites are ligated onto these fragments. A preselective round of PCR is performed using primers which recognize the adapter pairs. A second "selective" round of PCR is performed using primers similar to the preselective primers, except that an additional one or more nucleotides are attached to the 3' end of the primers. This permits a fraction of the fragments produced in the preselective amplification to be amplified. One of the primers is also labeled with a fluorescent dye which can be detected by an ABI DNA sequencing machine. By altering the number and composition of nucleotides at the 3' end of the primers, (either preselective or selective) the number of bands visualizable fragments may be adjusted. Bands may then be accurately and reproducibly sized on an ABI DNA sequencer.

For this study, the basic AFLP protocol was modified by replacing the labeled AFLP primer with a primer designed to recognize an AFC family SINE element isolated from Tanganyikan cichlids (Figure 5.1) (Takahashi et al 1998). The SINE specific primer was the 17-mer 5'-GCAACCTTCCGATTACA. The instructions provided with Applied Biosystems AFLP kit (Part # 402083) were followed with some minor modifications (Figure 5.2).

Digestion/Ligation Reactions

Digestion/Ligation reactions were performed as follows: The reaction mixture contained 1 unit of the 4 base cutter Mse I, 5 units of the 6 base cutter Eco RI, 1 unit of T4 DNA Ligase, 0.045 mg/ml BSA, 1µl 0.5 *M* NaCl, and 1 µl T4 ligase buffer with dNTP's, and 1 µl of each of the adapter pairs, and 5.5 µl of the DNA extraction was added to this

mixture. After 2 h incubation at 37° for exactly 2 h in a PEC 9600 thermocycler the digestion/ligation reaction was stopped by adding 190 µl of 1/10 TE.

Preselective PCR Reaction

Four microliters of the diluted restriction/ligation product was used in a preselective amplification which contained 0.5 µl of each of the preselective primers, 2µl 10x Thermo buffer (Promega, Inc., Madison WI), 1.5 µl 10 µm dNTP mixture, 10 µM MgCl2, and 0.2 µl Taq polymerase in a total volume of 20 µl. PCR conditions were as described in the AFLP plant mapping protocol provided by PE-Applied Biosystems (Foster City, CA). Each preselective primer recognizes the adapter and an additional 3' nucleotide, producing ~ 1 PCR product for every 16 restriction fragments.

Selective PCR Reactions

Selective PCR was carried out as described above, except that one of the preselective primers was replaced with a 6-FAM labeled primer, SINE-1, which recognizes a portion of the sequence published by Takahashi *et al* (1998). For each sample, two different PCR reactions were carried out, one with the labeled SINE-1 primer and the Preselective-Forward primer, and one with the labeled SINE-1 primer and the Preselective-Reverse primer. These two reactions create labeled fragments containing either a portion of the SINE element and an *Eco* RI site, or a portion of the SINE element and an *Mse* I site, respecitvely.

PCR products were electrophoresed on an ABI 373-A DNA sequencer. Applied Biosystem's GS-500 Tamra size standard was loaded along with each sample and electrophoresed for 9h. These conditions permit the detection of fragments ranging approximately from 75 to 450 b.p. in length. To enhance reproducibility, extra care was taken to standardize gel polymerization time, buffer and sample volumes, and other variables which could influence fragment migration.

Data Analysis

These data were analyzed using the Dollo parsimony option of Paup (version 3.1.1; Swofford 1993). In this procedure, AFLP fragments were interpreted to be cladistically informative characters with two states; presence or absence of a sequence complementary to the SINE-1 primer. We assume positional homology for SINE elements which appear in AFLP fragments of similar size. Although positional homology is not as strongly supported for SIFLP fragments as it would be if flanking regions for individual SINE elements were cloned (as in Takahashi *et al* 1998), the modest number of SIFLP fragments present in a taxon leads us to assume that falsely convergent characters will occur only rarely.

SIFLP fragments are produced only when a SINE element inserts into the genome within 450 b.p. of an appropriate restriction site with the appropriate orientation. SIFLP fragments may be lost due to mutations in the restriction site or the primer recognition site within the SINE element itself. Accordingly, Dollo parsimony, which assumes a single origin and multiple losses of cladistically informative characters was determined to be the most appropriate model for the analysis of these data.

Fragments produced from either the SINE and *Eco* RI primer or the SINE and *Mse* I primer were analyzed separately and to address distinct phylogenetic questions. Fragments produced from SINE and *Eco* RI primers were expected to be approximately 16 times rarer than those produced from the SINE and *Mse* I combination because the forward primer recognizes restriction fragments produced by the 6-base cutter *Eco* RI whereas the SINE AFLP reverse amplification products were produced from the 4 base cutter *Mse* I. The rare bands were used to estimate a phylogeny containing all the taxa using *Bathybates* sp. as an outgroup (following Meyer 1993). The resulting phylogeny was then used to determine an appropriate outgroup to use to further resolve the relationships within Lake Malawi using the SINE / *Mse* I data set which was expected to contain more phylogenetically informative characters. Because of the large number of taxa analyzed, consensus trees are presented here. Bootstrap analysis is not practical with so many taxa. In OTU's where many individuals were available, the presence of a fragment in any of the individuals within a taxon was interpreted as presence within that species (or population for *Mel. auratus*). Finally, a set of Lake Malawi haplotypes were analyzed using a synthetic outgroup which contained no SINE insertions to address the possibility of incomplete lineage sorting by determining whether haplotypes sort into phylogenetically credible groups.

Results

A total of 141 variable characters were observed among all taxa for the SINE *Eco RI* primer, 74 of these are phylogenetically informative among the taxa surveyed. The SINE / *Mse I* primer pair produced 237 different amplification fragments in the same set of taxa. The total number of bands observed within an individual was not far from the expected range. Takahashi *et al* (1998) estimate that between 1,000 and 10,000 copies of SINE elements are present in cichlid genomes. *Eco* RI sites are expected to occur every 5000 base pairs. Detectable SIFLP fragments in this study must be within 75 to 450 b.p. of a restriction site, therefore it is expected that about 7.5% of the SINE elements are the appropriate distance an *Eco*RI site and oriented in an amplifiable direction. Because the

preselective primers used each contain an additional 3' nucleotide beyond the adapter, only 1 in 16 fragments are expected to be amplified so ~6 to 60 fragments are expected per individual for the SINE / *Eco* RI reactions and ~100 to 1000 fragments for the SINE / *Mse* I reactions. The SINE / *Eco* RI combination produced an average of 11.25 detectable fragments / individual in fish from both lakes. The SINE / *Mse* I primer combination produced 25.35 fragments / individual for individuals from Lake Tanganyika and only 4.78 fragments / individual for individuals from Lake Malawi.

An analysis of all the taxa using SINE / *Eco* RI fragments resulted in the phylogeny estimate shown in Figure 5.3. This phylogeny is in general agreement with Poll's (1986) tribes with a few exceptions. The branch containing species from the tribe Limnochromini also contains a single representative of the tribe Tropheini (*Cyphotilapia frontosa*) and also contains both species in the tribe Perissodini. The remaining tribes in which more than a single individual was analyzed (Ectodini, Lamprologini, and the remaining Tropheini) form monophyletic assemblages which are distinct from the two tribes represented by single individuals (Eretmodini and Bathybatini). The Lake Malawi fish *Protomelus insignis* falls within the tribe Lamprologini. The remaining Lake Malawi taxa are in a monophyletic assemblage which also contains the remaining Tropheini species. The node leading to Lake Malawi / Tropheini group is defined by 4 SIFLP fragments.

In order to estimate the phylogeny within Lake Malawi, fragments resulting from the SINE / *Mse I* primers were analyzed using two Lamprologine taxa (*N. brichardi* and *J. marlieri*) as an outgroup. Because of the relatively small number of additional characters produced by either primer combination for Lake Malawi taxa, data from both primer sets were combined and analyzed. Fifteen equally parsimonious trees resulted and these phylogeny estimates have little bootstrap support.

When Lake Malawi haplotypes were analyzed as individual OTU's, no detectable phylogenetic signal resulted. The analysis was stopped after running for 12 hours on a PowerMacintosh with a 180 mhz processor. A consensus of the 15,000 equally parsimonious trees generated produced an essentially random intermingling of haplotypes.

Discussion

By combining AFLP methods with a SINE-specific primer, it is possible to rapidly survey a number of loci for the presence of a SINE element. This method successfully recovers phylogenetic information consistent with other data sets for older lineages within Lake Tanganyika, but is confounded by lineage sorting which has frustrated other

attempts at estimating phylogenies within Lake Malawi (Moran and Kornfield 1993; Parker and Kornfield 1997).

The Tropheini and all but a single Lake Malawi taxon form a well supported monophyletic assemblage defined by 4 characters (Figure 5.3). This is compatible with other data sets including Nishida's allozyme (1978), and mtDNA data (Kocher et al. 1993; Meyer 1993). The exception to this is *Protomelus insignis* which falls in the Lamprologini clade. Its position in this clade is only weakly supported however, as it is linked to the Lamprologine *Telmatochromis temporalis* by a single synapomorphy. This could be a spurious affinity resulting from an incomplete estimate of the SINE elements in this taxon, or it could suggest that the Lake Malawi flock is polyphyletic. This pattern could also be a result incomplete lineage sorting within the ancestral Tanganyikan flock. Alternatively, the fragments uniting these two taxa might be non-homologous.

The Limnochromini and Perissodini tribe representatives form a monophyletic assemblage along with *Cyphotilapia frontosa*. Mitochondrial DNA sequence data (Kocher *et al.* 1995) also indicate that *C. frontosa* is not a member of the Tropheini and is allied with the Limnochromini, whereas Sültmann *et al* (1995), who did not include other Tropheini in their analysis, found *C. frontosa* to be basal to the lineage containing the Malawi and Victoria flocks.

Phylogeny within lake Malawi was not well resolved with the SINE1 / *Eco* RI combination. In an attempt to further resolve relationships within the lake, SINE / *Mse I* fragments were analyzed. Because the reverse primer recognizes adapters attached to an *Mse* I site (a four base cutter) as opposed to an *Eco* RI site (a six base cutter), approximately 16 times more fragments were expected. The total number of fragments observed for this primer combination was larger for individuals from Lake Tanganyika, but fell short of the expected 16 fold increase (mean = 25.35 bands/individual within Lake Tanganyika). Surprisingly, the Lake Malawi taxa showed a decrease in observed bands, with an average of only 4.78 detectable fragments / individual. The shortfall in Lake Tanganyika might be due partly to a large number of fragments smaller than the lower detection limit of 75 b.p., however the drop in observed fragments for Lake Malawi taxa is puzzling. Explanations involving a reduced number of SINE elements within Lake Malawi are implausible given the identical number of fragments detected when the SINE / *Eco RI* primer combination was used. Methodological biases are also unlikely as individuals from both lakes were often run on the same gels.

The small number of detectable fragments produced a weak phylogenetic signal within Lake Malawi. The dendrogram shown in generated using all data for the Lake Malawi taxa shows some evidence of taxonomic signal, however little bootstrap support is

shown for the nodes in this phylogeny and there are some contradictions within the estimate. For example, the species from the genus *Metriaclima* are not sister taxa, the sand-dwelling predator *D. compressiceps* is a sister taxon the rock dwelling *Melanochromis* species and the two representatives of the Tanganyikan tribe Tropheini are embedded within the Malawi clade. The position of the two Tanganyikan taxa underscores the difficulties associated with phylogeny estimation within Lake Malawi. Although the Tropheini would be expected to form a paraphyletic group with the Malawian taxa, that pattern is not observed in these data. If phylogenetic information could be recovered with SIFLP data within the Lake Malawi flock, we would expect that it would at a minimum be able to distinguish between lineages from the two different lake basins.

Can SIFLP methods ever overcome incomplete lineage sorting? Possibly, but the practical utility of this approach may be severely limited. Incomplete lineage sorting occurs when lineage splitting occurs before genomic markers become fixed in a taxon, thus ancestral polymorphisms are inherited by descendant taxa. In rapidly speciating lineages like the Lake Malawi cichlids, this tendency has made it impossible to produce reliable phylogeny estimates (*cf* Parker and Kornfield 1997). However, incomplete lineage sorting does not apply to the entire genome, after all, taxa are recognized by a suite of genetically controlled morphological characters, even within Lake Malawi. Some genes must become fixed for taxonomic diversification to occur. If a SINE element insertion occurs near one of these genes, then it could be swept to fixation as part of the speciation process. This is only useful however if SINE element insertions occur frequently and if their density in the genome is high enough for them to stand a good chance of being associated with genes which define speciation. Thus if enough SINE insertions are visualized by varying primer and restriction enzyme combinations then phylogenetic signal within Lake Malawi might yet be recovered. Further, cladistically informative SINE insertions might be useful as map markers for the genes which define lineages. SIFLP may also be used to rapidly screen a number of candidate SINE insertions for the development PCR primers which recognize the flanking region of the SINE insertion, permitting the rapid development of taxon specific assays or for phylogenetic analysis beyond a set of taxa initially screened with SIFLP.

SINE insertions provide a credible phylogenetic signal for the older lineages of Lake Tanganyika and provide a rapid method of assaying set of cladistically informative nuclear markers scattered broadly throughout the genome. These and similar markers may be a useful alternative to sequence based approaches to phylogeny estimation in closely related taxa.

Figure 5.1

A SINE element described by Takahashi *et al* (1998) isolated from the Tanganyikan cichlid *Julidochromis transcriptus*

The region complementary to the primer SINE-1 is underlined, and the primer sequence is shown in capital letters at the bottom of the figure.

ctgagggttgactggaaatgggaaacagggtaactaggaaacatggtgaaaataatcaagacaaacgcaa
agaaggaagtaaaattgacacaaggtatgaaaaaaaactgtagctagaactcgtgtaataaaagagagcc
tttggcgattgtggctcaagagttgggagttcgcttgtaatcggaaggttgccggttcgagccccggttt
ggacagtctctgtcgttgtgtccttgggcaagacacttcacccgttgcctactggtggtggtcagagggc
ccggtggcgccagtgtccggcagcctcgcctctgtcagtgcaccccagggtggctgtggctacaatgtag
ctgccatcaccagtgtgtgaatgtgtgtgtgtgaatgggtgaatgactggatgtagtgtaaagcgctttg
gggtccttagggactagaaaagcgctatacaaatacaggccatttaccatttgcctcagtattataaaat
atgaaaaaaaaaaaattcaaatacagaaaaacctaaacacaaatccaaaactcacatttgttaactaaag
attcaagttttagacacagcacataatcctgagac

5'-GCAACCTTCCGATTACA

Figure 5.2

SIFLP. This process is a variation of the AFLP methods described by Vos *et al.* (1995). A) Genomic DNA is digested with two restriction enzymes. B) Adapters containing PCR primer sites are ligated onto the restriction fragments. C) A "nonselective" PCR reaction is performed using primers which recognize the ligated adapters. D) The product from step C is used in a "selective" PCR reaction using a labeled primer which recognizes a SINE element and an unlabeled primer that recognizes one of the ligated adapters. E) Amplification products which contain a section of a SINE element are visualized on an ABI DNA sequencer using ABI's GeneScan software. Lanes 1-4 are *Lethrinops gossi*, Lanes 5-8 are *Protomelus spilopterus*, lanes 9-11 are *Metriaclima zebra* cf Mazinzi Blue. *Lethrinops* and *Protomelus* are pelagic genera. *Metriaclima* are found in shallow, rocky habitats. The white bands in this image are SINE containing PCR products, the dark gray bands are size standards which are loaded in every lane.

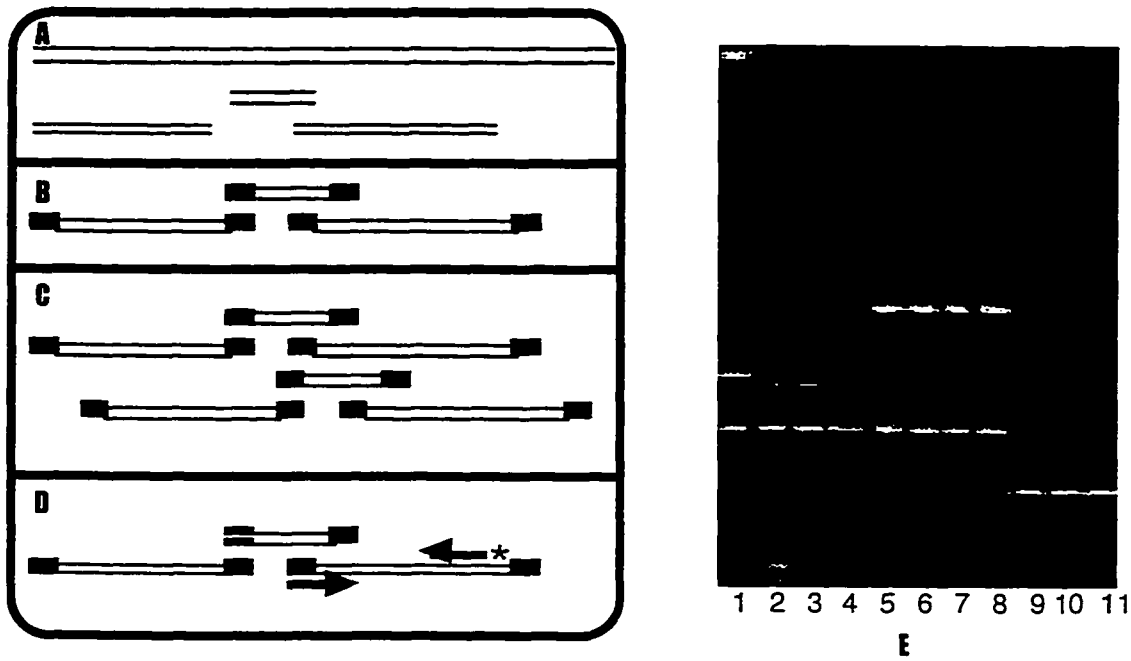


Figure 5.3

A majority rule consensus of 6 equally parsimonious cladograms constructed using SIFLP fragments generated by using the SINE / *Eco* RI primer combination. *Bathybates* sp. was used as an outgroup. The text to the right of the bars indicates Tribe affiliation for the fish from Lake Tanganyika.

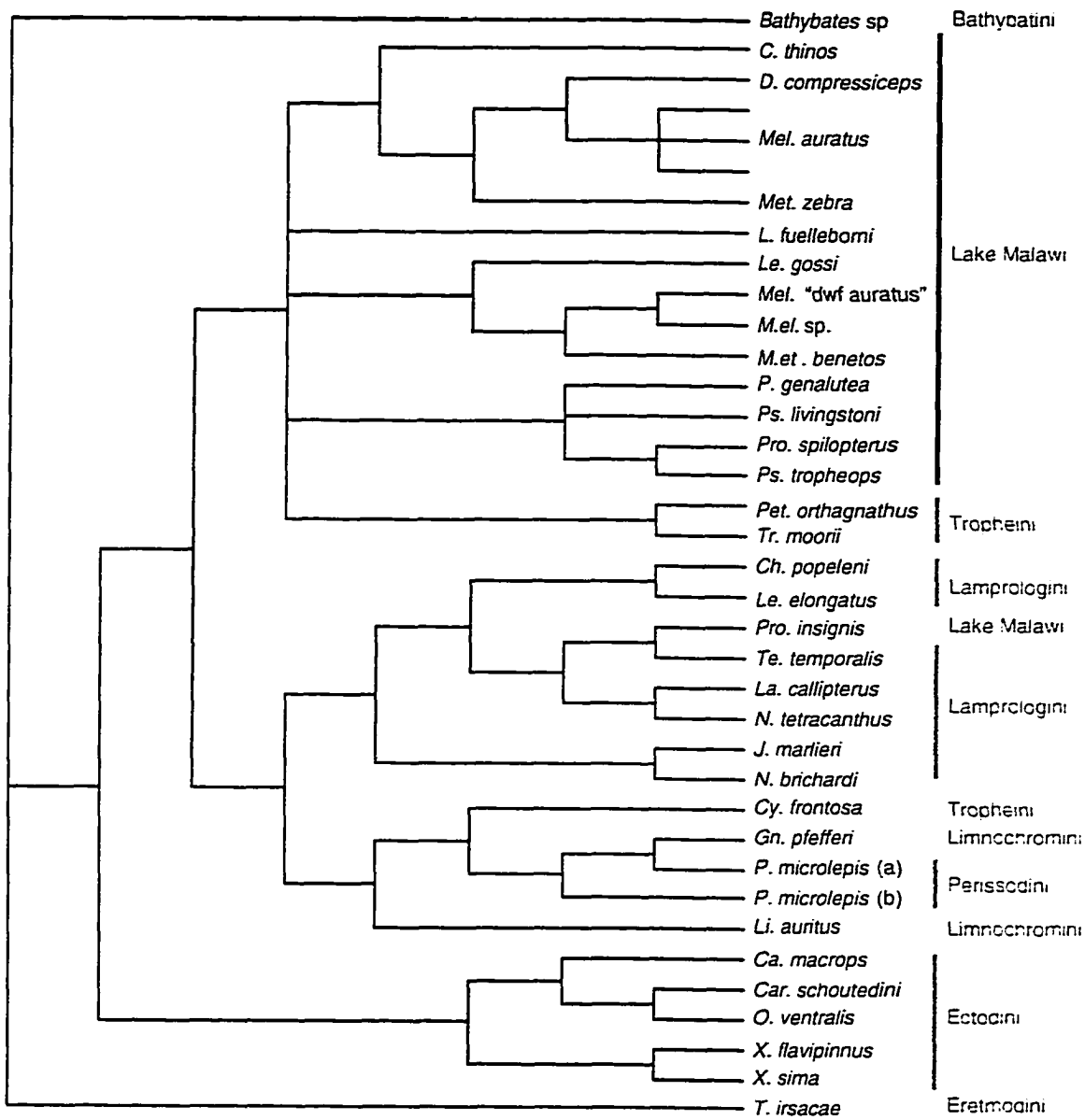


Table 5. 1

Taxa and sample sizes

Lake	Tribe/Guild	Species	Number sampled
Tanganyika	Bathybatini	*Bathybates* sp.	1
	Ectodini	*Callochromis macrops*	1
		Cardiopharynx schoutedeni (sp?)	1
		Opthalmotilapia ventralis	1
		Xenotilapia flavipinnus	1
		X. sima	1
	Eretmodini	*Tanganicodus irsacae*	1
	Lamprologini	*Chalinochromis popeleni*	1
		Julidochromis marlieri	1
		Lamprologus callipterus	1
		Lamprologus callipterus	1
		N. brichardi	1
		Neolamprologus tetracanthus	1
		Telmatochromis temporalis	1
	Limnochromini	*Gnathochromis pfefferi*	1
		Limnochromis auritus	1
	Perissodini	*Perissodus microlepis a*	1
		Perissodus microlepis b	1
	Tropheini	*Cyphotilapia frontosa*	1
		Petrochromis orthognathus	1
		Tropheus moorii	1
Malawi	Sand dwellers	*Copadichromis thinos*	4
		Dimidiochromis compressiceps	6
		Lethrinops gossi	6
		Protomelus insignis	
		Protomelus spilopterus	4
	Rock dwellers	*Melanochromis auratus* (Chipoka)	3
		M. auratus (Shallow Reef)	4
		M. cf heterochromis	5
		Metriaclima zebra	4
		Met. benetos	3
		Pseudotropheus tropheops	5
		Ps. livingstoni	2
		Petrotilapia genalutea	3

Appendix

The data presented on the following pages represent coded four locus genotypes for all *Melanochromis auratus* individuals analyzed in this study. Each line represents an individual, and genotypes are indicated by four digit numbers - the first two digits represent the state of the smaller SSR allele at that locus, the second two digits represent the larger SSR allele at that locus. Missing data is represented by "0000". These two digit allele codes may be translated to approximate fragment lengths by referring to the chart below. They may be translated into estimated repeat numbers by referring to *Chapter 1.*

UNH-001		**UNH-002**		**UNH-050**		**UNH-231**	
Bin Range	*Code*	*Bin Range*	*Code*	*Bin Range*	*Code*	*Bin Range*	*Code*
156.87-156.99	01	174.78-174.78	01	292.15-292.65	01	191.82-194.22	01
162.28-163.64	04	181.72-182.52	04	316.19-316.66	13	195.8-196.73	02
168.52-169.67	07	185.68-186.40	06	321.15-321.15	16	199.66-199.79	03
171.66-172.10	08	187.77-188.99	07	324.50-325.17	18	201.78-201.79	04
173.10-173.91	09	189.61-190.36	08	328.55-329.57	20	205.65-205.91	06
175.40-176.39	10	191.55-192.30	09	330.47-331.77	21	207.64-208.10	07
177.46-178.03	11	192.88-194.08	10	333.30-333.94	22	211.57-212.01	08
179.93-179.75	12	195.66-196.13	11	334.42-336.33	23	213.16-213.96	09
180.84-182.71	13	197.33-198.14	12	337.20-337.87	24	217.22-218.14	11
183.10-183.88	14	199.21-200.21	13	339.28-340.53	25	219.19-220.12	12
185.09-185.78	15	200.44-202.12	14	341.20-342.66	26	221.52-222.06	13
186.45-187.72	16	203.31-203.97	15	344.07-344.86	27	223.16-224.18	14
188.73-189.70	17	204.49-206.68	16	345.76-347.10	28	225.17-226.25	15
190.05-190.37	18	207.41-207.90	17	347.99-348.97	29	227.06-228.22	16
190.94-191.58	19	209.25-210.13	18	350.00-351.43	30	229.06-230.04	17
192.90-193.50	20	211.72-214.02	19	352.00-353.42	31	231.09-232.16	18
194.63-195.48	21	214.41-215.83	20	355.00-355.78	32	233.08-233.62	19
196.11-197.41	22	216.94-219.90	21	357.44-357.92	33	234.06-235.99	20
198.60-199.90	23	223.37-223.91	23	363.51-364.26	36	237.10-238.00	21
200.53-201.66	24	225.36-225.75	24			239.18-240.18	22
202.41-202.92	25	227.29-227.69	25			241.35-241.98	23
204.64-204.87	26	229.36-229.76	26			243.06-243.68	24
206.50-207.11	27	235.65-235.70	29			244.00-245.48	25
208.81-208.90	28					246.99-247.88	26
210.79-210.79	29					249.42-249.54	28
212.54-213.17	30					251.24-251.89	29
214.58-215.07	31					252.98-253.27	30
218.76-219.8	33						
223.74-223.75	35						
243.60-243.60	55						

Four locus genotypes for individuals from 10 populations in southern Lake Malawi.

Harbour Island

Locus UNH- 001	002	050	231
1317	1218	0000	0000
1417	1616	0000	0000
1321	0716	2328	1214
1717	1819	0000	0000
2131	1619	2328	1430
1721	0719	0000	0000
1730	1623	2528	1417
1733	1219	2128	0000
2130	1216	2532	1218
1331	0716	0000	0000
2133	0812	2325	1214
2133	1216	2023	0116
1717	1216	0121	1214
1621	1219	2129	1414
1421	1616	2628	1220
1617	1219	2329	1416
1111	1919	0121	0116
2127	1825	0123	1216
1721	1616	2124	0000
1121	1218	0125	1229
1717	1919	0000	1616
1517	0716	2828	1229
1721	1219	2323	1214
2121	0719	2324	1214
1114	1819	2628	1420
1721	0712	2829	0114
0000	1219	2829	1414
0000	0707	2429	1416
1723	1619	2428	1414
1719	1618	2323	1420
1617	0000	2123	1417
1117	0707	2323	1230
0000	0719	0101	1416
2121	0719	2428	1214
1721	1619	0123	1414
1717	1216	0128	1414
1621	1619	2128	1420
0000	1219	2125	1414
2127	0719	2129	1416
1717	1616	0000	1414
2121	1616	2323	1214

Harbour Island (continued)

Locus UNH- 001	002	050	231
1717	0000	2828	1212
2121	1616	2828	1421
1721	0000	2128	1114
1730	1618	0121	1320
1421	0000	2328	1417
1721	0000	2128	1418
1113	0000	2123	1414
1121	0716	2123	1214
1721	1616	2424	1414
1731	1619	2424	1214
1427	0716	2829	1618
1731	2020	2424	0000
2130	0819	2029	0000
1717	1216	2324	1214
1721	0712	2425	1420
1717	1619	0123	1214
1021	0911	2128	1625
1721	1219	0129	1214
1721	0716	2124	1214
1621	0712	2324	1730
1617	1616	0121	0112
1317	1819	0128	1414
1717	0719	2325	1430
1619	1919	0101	1212
1116	0000	2324	1214
1721	1418	2128	1214
2121	1619	2424	1414
1121	0816	2329	2630
2121	0000	2424	1317
1121	0716	2829	1216
0116	0707	0121	1220
1516	0000	0000	1214

Shallow Reef

Locus UNH-001	002	050	231
1717	1616	2021	1416
1717	1616	2021	1214
1721	0716	2021	1214
1717	1616	2125	1414
1717	1616	2136	1214
0000	0000	2126	1416
1721	1616	2125	1414
2121	1616	2136	1417
1717	1616	2036	1217
1721	1212	0000	1414
1717	1616	2125	1214
1721	0716	2021	1216
1717	1616	2021	1416
1717	1616	2020	1214
1919	0707	2021	1414
1724	1616	2020	1215
1731	1316	2021	1414
1717	1616	2023	1214
3135	1316	2021	1414
1717	1616	2532	1414
1717	1617	2025	1214
1720	1616	2136	1214
3535	1621	2020	1414
1717	1616	2532	1415
0000	0719	2021	1414
1717	1616	2122	1414
1717	1616	2020	1416
2122	1616	2125	1224
5555	1616	2036	1214
1717	0000	2020	0000
0000	1616	2121	1214
1717	1621	2021	1414
1717	0716	2121	1414
1919	1616	2123	1213
2031	1516	2020	1414

Ilala Gap

Locus UNH-001	002	050	231
0713	0000	1621	0109
1617	0000	1830	0616
1414	0000	0000	0000
1320	1218	2324	0000
1314	1126	1829	0416
1011	1112	0000	0000
1314	1012	0000	0000
1321	1229	2324	0000
1417	0811	2021	0620
1314	1117	3031	0000
0000	1117	0000	0000
1313	1117	2330	0118
1321	1719	2330	1518
1313	0717	2029	1623
1317	1125	2230	1616
1421	1111	2930	1617
1316	0909	1829	1824
1621	1526	2023	1818
1014	0911	2020	1315
1314	1515	2123	0114
1321	1729	2628	0414
1321	0911	2021	0116
1616	0809	0000	0000
1416	0406	0000	1919
1314	0911	2930	1316
1317	1125	2331	0314
1416	0000	2026	0918
2121	1117	2024	1317
1319	0609	2029	0120
1320	0917	2130	1624
1319	0411	2023	1924
1313	0909	1820	0124
1317	1125	2329	1424
1313	0707	2530	1117
1420	1111	2029	0916
1313	0404	2323	1624
1317	1619	2123	1624
1214	1115	2929	1618
1421	1217	2330	1718
1320	0811	1820	1626
0000	0917	2930	0916
1313	0000	2026	1818
2020	0000	0000	0000
1313	0712	2930	0616
0000	0919	2229	1424
1621	1226	2530	0916
1316	0811	2323	0915
0000	1113	2029	0118
1321	0909	2929	0112

Ilalla Gap (continued)

Locus UNH-			
001	002	050	231
0000	0911	2023	0914
1317	1217	2027	1824
0000	0000	2030	0809
1314	0909	2023	0914
0708	0000	0000	0000
0000	0709	2930	2425

Mphande Island

Locus UNH-			
001	002	050	231
1717	1416	2123	1221
1717	0000	2123	2426
0418	1616	2121	1124
0425	1416	2123	1226
1717	0000	0000	0000
0418	1616	2021	1221
1833	1616	2121	2626
0421	1616	2121	1226
1717	0916	2121	2126
0421	1616	2123	1226
1717	1616	2123	1226
1731	1616	2121	1214
1818	1416	2121	1226
1818	0916	2121	1226
1717	1616	2121	1426
1730	1717	2121	1224
1721	1619	2323	1826
2125	0916	2121	1226
1717	1616	2121	2626
1720	1616	2123	1212
1717	1616	2021	1426
1717	0000	2123	0000
1717	1616	2123	1217
1733	1616	2121	1417
2633	1616	2121	1212
0404	1616	2121	1226
1717	1616	2121	2426

Mazinzi Reef

Locus UNH-			
001	002	050	231
00,00	16,16	00,00	00,00
17,17	16,17	23,32	14,14
21,25	14,16	00,00	1212
1721	0716	2123	1415
1722	1616	2020	1214
1717	1616	2228	1414
2131	1616	2021	0000
1717	0000	2022	1414
1818	0714	2021	1224
0000	1616	0000	1212
0000	0000	2026	1414
1818	0716	2020	1214
1721	1416	2021	1414
1717	1314	2536	1416
1722	1316	2021	1212
0000	1417	2020	1214
1731	1616	2123	1214
1722	0716	2020	1414
1722	0716	2020	1414
1717	1617	2023	1414
1717	1617	2020	1414
1717	1316	2020	1214
1722	1616	2020	1414
1717	1616	2326	1214
1731	1616	2326	1214
1717	1617	2326	0312
1319	0708	2122	0416
1717	1616	2022	1214
1414	0707	2026	2022
1014	0708	2030	2323
1717	1616	2026	1214
1727	1616	2122	1214
1717	1616	2020	1416
1717	0000	2020	1214
2022	1316	2323	1414
1721	1616	2021	1415
1721	1616	0000	1415
1721	1616	2022	0000

Mumbo Island

Locus UNH-			
001	002	050	231
1010	1313	2929	1717
0000	1010	2929	1824
1621	0913	2829	1820
0812	1315	2930	1724
1021	0913	2929	1420
1010	0915	2933	1421
1621	1515	2833	1417
0000	0913	2929	1722
1016	1313	2929	1717
1213	1515	2933	1717
1216	1313	2728	1824
1016	0909	2829	1725
1010	0913	2929	1424
1316	0913	2729	1724
1010	0909	3033	1724
2124	1313	2929	1617
0816	0913	2729	1824
1616	1313	2033	1724
1013	0909	2829	1617
1417	1414	2829	1424
1313	0913	2829	1724
1621	1315	2729	1424
1316	1313	2829	1720
1322	1414	2729	2024
0816	1414	2929	2424
1216	1515	2729	1722
1324	1416	2933	1718
1016	0913	2829	2025
1013	1014	2929	1417
1016	0915	2829	2022
1621	0909	2829	1722
1621	1315	2829	1620
1321	1315	2933	2021
1121	0913	2833	1718
1016	1414	2829	1422
1022	1014	2729	1622
1013	1315	2831	1617

Mvunguti East

Locus UNH-			
001	002	050	231
0814	1219	2121	1616
1116	1112	2125	1416
0000	1515	2028	0101
1414	0712	2021	1414
1419	1215	2029	0101
1417	1826	2124	1418
1313	0911	2324	1819
1414	0718	2428	0112
1414	1819	2430	1116
1416	0811	2125	1416
1127	1921	2828	1616
1416	1919	2224	1616

Nkhudzi Point

Locus UNH-			
001	002	050	231
0000	0916	2123	0000
1725	1618	2121	1826
1725	1616	2121	1218
1730	1616	2123	1221
0421	1616	2121	2526
1721	1616	2121	1418
0000	0917	2121	1726
1720	1617	2123	1718
1717	1616	2123	1824
1717	1616	0000	1212
1717	0916	2123	1717
1721	1616	2121	1214
1717	1616	2121	1226
1921	1717	2124	1226
2121	1616	2123	1418
1725	0000	2424	0000
1725	1717	2121	1726
1717	1618	2123	1617
1717	1616	2021	0000
0000	0000	1321	1426
1517	1616	1323	1818
0000	1616	2121	1718
0000	0000	2122	1426
0000	1416	2121	1224
1717	1616	2121	1821
1723	1616	2121	1726
1725	1717	2123	1417
1617	1616	2121	1226
2130	1616	0000	1215
0000	1616	2121	1226
1317	1616	2121	2526
0000	0916	2121	1418
1721	1616	2121	1224
0000	0000	0000	0820
2331	1616	0000	1226

Tsano Rock

Locus UNH-			
001	002	050	231
1414	0406	2127	0111
1322	1121	2327	1212
1314	1919	2124	1116
0000	0000	2426	1416
0000	0000	2021	1418
1421	0715	2021	1618
1416	0916	2229	1414
1616	1818	2022	1418
1316	1921	2930	0112
0000	0000	2829	1214
0000	0000	2730	0000
1316	0709	2425	1416
1330	0000	2324	0101
1316	0809	2129	1416
1416	1619	2028	0112
0000	0911	2023	0000
0000	0000	2228	1114
0913	0000	2229	0000
1317	0708	2024	0114
1617	1316	2429	1720
1319	0707	2428	1416
1616	1111	2025	1216
1313	0000	2324	0112
1416	0711	2430	0000
1322	0919	2930	0000
1720	0000	0000	0000
1316	1818	2125	1516
1017	1111	2024	1114

Tsano Rock (continued)

Locus UNH-			
001	002	050	231
1416	1417	1824	1416
1417	1214	0000	1417
1416	1121	2930	1617
1621	0719	2129	1619
0816	0718	2429	1719
1414	0716	2129	1218
1617	0915	2121	0119
1930	0923	2323	0111
1013	1213	2930	1417
1517	0919	2426	1119
1021	0919	2123	1620
1417	1415	0000	0114
1415	1111	2429	1414
1720	1821	2127	1416
1113	1215	2020	1221
1717	0918	2029	1116
1317	1119	0000	0101
1017	0707	2425	1416
1416	1221	2626	1114
1416	0709	2020	0101
1416	1526	2429	0202
1313	0713	0000	1820
1621	0719	2128	0118
1621	1319	2331	1418
1416	0719	0000	0114
1628	1219	2125	0116

References

Allen, P. J., W. Amos, P. P. Pomeroy and S. D. Twiss. (1995). "Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two British breeding colonies." Molecular Ecology **4**: 653-662.

Banister, K. E. and M. A. Clarke (1980). "A revision of the large *Barbus* (Pisces, Cyprinidae) of Lake Malawi with a reconstruction of the history of the southern African Rift Valley lakes." Journal of Natural History **14**: 483-542.

Barton, N. H. and M. Slatkin (1985). "A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population." Heredity **56**: 409-415.

Boulenger, G. A. (1897). "Description of a new fish from Lake Nyassa." Annals and Magazine of Natural History **6**(xix): 155.

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

Bowers, N. and J. J R Stauffer (1997). "Eight new species of rock-dwelling cichlids of the genus *Melanochromis* (Teleostei: Cichlidae) from Lake Malawi, Africa." Ichthyological Explorations of Freshwaters **8**(1): 49-70.

Bowers, N., J. R. Stauffer, and T. D. Kocher. (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.

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

Brooks, J. L. (1950). "Speciation in ancient lakes." Quarterly Review of Biology **25**(2): 131-176.

Crapon de Caprona, M. D. (1986). The use of fertile hybrids for the study of the accuracy of species recognition in cichlids. Proceedings of the 3rd European Workshop on Cichlid Biology, Tevuren, Belgium, Musee Royal de L'Afrique Centrale.

Cook, J. M. and M. Tristem (1997). "'SINEs of the times' - transposable elements as clade markers for their hosts." TREE **12**(8): 295-297.

de Beaufort, L. F. (1951). Zoogeography of the Land and Inland Waters. London, Sidgwick and Jackson, Ltd.

Dominey, W. J. (1984). Effects of sexual selection and life history on speciation: species flocks in African cichlids and Hawaiian Drosophilia. Evolution of Species Flocks. A. A. Echelle and I. Kornfield. Orono, University of Orono Press: 231-249.

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.

Fryer, G. (1959). "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**(2): 237-279.

Fryer, G. (1959). "Some aspects of evolution in Lake Nyasa." Evolution **13**: 440-451.

Fryer, G., T. D. Iles (1972). The cichlid fishes of the great lakes of Africa: their biology and evolution. Edinburgh, Oliver and Boyd.

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza, and M. W. Feldman (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." The Journal of Heredity **86:** 485-486.

Greenwood, P. H. (1964). "Explosive speciation in African lakes." Proceedings of the Royal Institute of Great Britain **40:** 256-269.

Greenwood, P. H. (1991). Speciation. Cichlid fishes: Behaviour, ecology and evolution. M. H. A. Keenleyside. London, Chapman and Hall: 86-102.

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: 303-314.

Hatfield, T. and D. Schluter (1996). "A test for sexual selection on hybrids of two sympatric sticklebacks" Evolution **50:** 2429-2434.

Hert, E. (1989). "The function of egg spots in an African mouthbrooding cichlid fish." Animal Behaviour **37**: 726-732.

Hill, B. J. and A. J. Ribbink (1978). "Depth equilibration of a shallow-water cichlid fish." Journal of Fish Biology **13**: 741-745.

Jackson, P. B. N. (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, T. C., C. A. Scholz, Talbot, M. R. et al. (1996). "Late Pleistocene dessication of Lake Victoria and rapid evolution of cichlid fishes." Science **273**: 1091-1093.

Kellogg, K. A., J. A. Markert, J. R. Stauffer, jr., and T. D. Kocher (1995). "Microsatellite variation demonstrates multiple paternity in lekking cichlid fishes from Lake Malawi, Africa." Proceedings of the Royal Society of London, Series B. **260**: 79-84.

Kocher T.D., J. A. Conroy, K. R. McKaye, J. R. Stauffer, jr, S. F. Lockwood (1995). "Evolution of NADH dehydrogenase subunit 2 in east African cichlid fish." Molecular Phylogenetics and Evolution **4:** 420-432.

Kocher, T. D., J. A. Conroy, et al. (1993). "Similar morphologies of cichlids in lakes Tanganyika and Malawi are due to convergence." Molecular Phylogenetics and Evolution **2**: 158-165.

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

Lee, W.-J. and T. D. Kocher (1996). "Microsatellite DNA markers for genetic mapping in Oreochromis niloticus." Journal of Fish Biology **49**: 169-171.

Lehman, T., W. A. Hawley, L. Kamau, T. Fontenille, F. Simard and F. H. Collins. (1996). "An evaluation of evolutionary constraints on microsatellite loci using null alleles." Genetics **144**: 1155-1163.

Liem, K. F. (1974). "Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws." Systematic Zoology **22**: 425-441.

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

Marsh, A. C. and A. J. Ribbink (1981). "A comparison of the abilities of three sympatric species of *Petrotilapia* (Cichlidae, Lake Malawi) to penetrate deep water." Environmental Biology of Fishes **6**(3/4): 367-369.

Mayr, E. (1963). Animal Species and Evolution. Harvard University Press, Cambridge, Massachusetts.

McKaye, K. R. (1990). Sexual selection and the evolution of the cichlid fishes of Lake Malawi, Africa. Behavior, Ecology, and the Evolution of Cichlid Fishes. M. Keenleyside. London, Chapman Hall: 241-257.

McKaye, K. R. and W. N. Gray (1984). Extrinsic barriers to gene flow in rock-dwelling cichlids of Lake Malawi: Macrohabitat heterogeneity and reef colonization. Evolution of Fish Species Flocks. A. A. Echelle and I. Kornfield. Orono, University of Maine at Orono Press.

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

Meyer, A. (1993). "Phylogenetic Relationships and Evolutionary Processes in East African Cichlid Fishes." TREE **8**: 279-284.

Minch, E., A. Ruiz-Linares, D. Goldstein, M. Feldman, and L. L. Cavalli-Sforza (1995-1996) "Microsat (version 1.4d): a computer program for calculating various statistics on microsatellite allele data". WWW: http://lotka.stanford.edu/microsat.html.

Moran, P. and I. Kornfield (1993). "Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi." Molecular Biology and Evolution **10**(5): 1015-1029.

Murata, S., N. Takasaki, M. Saitoh, et al. (1996). "Details of Retropositional Genome Dynamics that provide a rationale for a Generic Division: The distinct branching of all the Pacific salmon and trout (Oncorhynchus) from the Atlantic salmon and trout." Genetics **142**: 915-926.

Nishida, M. (1991). "Lake Tanganyika as an evolutionary reservoir of old lineages of East African cichlid fishes: Inferences from allozyme data." Experientia **47**: 974-979.

Owen, R. B., R. Crossley, T. C. Johnson, D. Tweddle, I. Kornfield, S. Davidson, D. H. Eccles, D. E. Engstrom. (1990). "Major low lake levels of Lake Malawi and their implications for speciation rates in cichlid fishes." Proceedings of the Royal Society of London, Series B **240**: 519-553.

Parker, A. and I. Kornfield (1996). "Polygynandry in *Pseudotropheus zebra*, a cichlid fish from Lake Malawi." Environmental Biology of Fishes **47**: 345-352.

Parker, A. and I. Kornfield (1997). "Evolutionof the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of Lake Malawi." Journal of Molecular Evolution **45**(1): 70-83.

Patterson, H. E. H. (1985). The recognition concept of species. Species and Speciation. E. S. Vrba. Pretoria, S.A., Transvaal Museum. **4**: 21-34.

Poll, M. (1986). "Classification des cichlidae du lac Tanganika tribus, genres et especes." Mémoires de la Classe des Sciences **45**: 5-163.

Raymond, M. and F. Rousset (1995). "GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism." Journal of Heredity **86**(3): 248-249.

Regan, C. T. (1921). "The cichlid fishes of Lake Nyassa." Proceedings of the Zoological Society of London **1921**: 675-727.

Reinthal, P. N. (1990). "The feeding habits of a group of herbivorous rock-dwelling cichlid fishes (Cichlidae: Perciformes) from Lake Malawi, Africa." Env. Bio. Fish. **27**: 215-233.

Ribbink, A. J. (1983). "The zoogeography, ecology and taxonomy of the genus *Labeotropheus* Ahl, 1927, of Lake Malawi (Pisces: Cichlidae)." Zoological Journal of the Linnean Society **79**: 223-243.

Ribbink, A. J. (1986). "The species concept, sibling species and speciation". Proceedings of the 3rd European Workshop on Cichlid Biology, Tevuren, Belgium, Musee Royal de L'Afrique Centrale. **251**: 109-116.

Ribbink, A. J. (1991). Distribution and ecology of the cichlids of the African Great Lakes. Cichlid fishes. Behavior, ecology and evolution. M. H. A. Keenleyside. London, Chapman and Hall.

Ribbink, A. J., A. C. Marsh, B. A. Marsh, and B. J. Sharp. (1983). "A preliminary survey of the cichlid fishes of the rocky habitats in Lake Malawi." South African Journal of Zoology **18**(3): 149-310.

Ruzzante, D. E. (IN PRESS). "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.

Scholz, C. A. and B. R. Rosendahl (1988). "Low lake stands in Lake Malawi and Tanganyika, East Africa, delineated with multifold seismic data." Science **240**: 1645-1648.

Seehausen, O., J. J. M. van Alphen and F. Witte. (1997). "Cichlid fish diversity threatened by eutrophication that curbs sexual selection." Science **277**: 1818-1811.

Seehausen, O., J. J. M. van Alphen, and F. Witte. (IN PRESS). "The effect of male coloration on female mate choice in closely related Lake Victoria cichlids." Behavioral Ecology and Sociobiology.

Stauffer, J. R., jr, N. J. Bowers, K. A. Kellogg, and K. R. McKaye. "A revision of the blue-black *Pseudotropheus zebra* (Teleostei: Cichlidae) complex from Lake Malawi, Africa, with a description of a new genus and ten new species". Proceedings of the Academy of Natural Sciences of Philadelphia. **148:** 189-230.

Sültmann, H., W. E. Mayer, F. Figueroa, H. Tichy, and J. Klein (1995). "Phylogenetic analysis of cichlid fishes using nuclear DNA markers" Molecular Biology and Evolution **12:** 1033-1047.

Takahashi, K., Y. Terai, M Nishida, and N Okada. (1998). "A novel family of short interspersed repetitive elements (SINEs) from cichlids: The patterns of insertion of SINES at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika." Molecular Biology and Evolution **15**(4): 391-407.

Tautz, D. (1993). Notes on the definition and nomenclature of tandemly repetitive DNA sequences. DNA Fingerprinting: State of the Science. D. D. J. Pena, R. Chakraborty, J. T. Epplen and A. J. Jeffreys. Basel-Boston-Berlin, Birkhäuser Verlag. **EXS 67:** 21-28.

Thornhill, R. and J. Alcock (1983). The Evolution of Insect Mating Systems. Cambridge, Harvard University Press.

Trendall, J. (1988). "Recruitment of juvenile mbuna (Pisces: Cichlidae) to experimental rock shelters in Lake Malawi, Africa." Environmental Biology of Fishes **22**(2): 117-131.

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

Tripp, R. T., R. T. Bailey, A. Crossby and R. Farrar (1957). Chipoka to Monkey Bay.

van Oppen, M. J. H., J. C. Deutsch, G. F. Turner et al. (1997). "Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes." Proceedings of the Royal Society, London: Series B **264**: 1803-1812.

Vos, P., R. Hogers, M. Bleeker et al. (1995). "AFLP: a new technique for DNA fingerprinting." Nucleic Acids Research **23**(21): 4407-4414.

Vrba, E. S. (1985). "Environment and evolution: alternative causes of the temporal distribution of evolutionary events." South African Journal of Science **81**(5): 229-236.

Weir, B. S., and C.C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." Evolution **38**(6): 1358-1370.

West-Eberhard, M. J. (1983). "Sexual selection, social competetion, and speciation." The Quarterly Review of Biology **58**(2): 155-183.

Worthington, E. B. (1937). "On the evolution of fish in the great lakes of Africa." Int. Revue d. ges. Hydrob. u. Hydrogr. **35**: 305-317.

Worthington, E. B. (1954). "Speciation of fishes in African lakes." Nature **173**: 1064-1067.