

Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids

Paul W. Shaw^{1*}, George F. Turner², M. Rizman Iddid^{1†}, Rosanna L. Robinson² and Gary R. Carvalho¹

¹*Department of Biological Sciences, University of Hull, Hull HU6 7RX, UK*

²*School of Biological Sciences, University of Southampton, Southampton SO16 7PX, UK*

Allopatric processes of speciation have routinely been presented to explain the extraordinary radiation of the East African Great Lakes cichlid fish species flocks. The 21 or more species of pelagic cichlids within the Lake Malawi flock appear to have lake-wide distributions that challenge such a concept. Data from six microsatellite DNA loci indicate single, panmictic populations across the lake of three *Diplotaxodon* species. Levels of variability at these loci suggest that populations have been large and stable. Mitochondrial DNA sequence data (872 bp of control region + 981 bp of the NADH-2) from 90 species, representing all major clades within the Lake Malawi flock, indicate reciprocal monophyly of the pelagic clade. We suggest that these data support a hypothesis that speciation in sympatry is more plausible (and widespread) within the cichlid species flocks than previously thought.

Keywords: Lake Malawi cichlids; microsatellites; mitochondrial DNA sequence; phylogeny; sympatric speciation; population structure

1. INTRODUCTION

Mechanisms of speciation are of fundamental interest to biologists as they underpin the whole process of macroevolution. The process of genetic divergence (and eventually speciation) of populations isolated physically (allopatry) is well described, and documented (e.g. Coyne 1992). The alternative process where divergence occurs without isolation (sympatry) has proved to be intractable theoretically and with few good natural examples; a number of instances where monophyletic species assemblages occur in sympatry, e.g. sticklebacks (Taylor & McPhail 1999), lizards (Losos *et al.* 1998) and senecios (Knox & Palmer 1995), suggest sympatric speciation. One of the most convincing studies concerns cichlid fishes in crater lakes (Schliewen *et al.* 1994), where monophyly discounts repeated invasions and recent radiation in a continuous habitat suggests sympatric divergence. In extension of this latter case, the species flocks of cichlid fishes within the East African Great Lakes are good candidates for assessing the contribution of sympatric processes of speciation, as large numbers of species appear to have radiated within single water bodies. Molecular studies have confirmed that these radiations have occurred by intralacustrine speciation from a single common ancestor by demonstrating monophyly within lakes (Meyer *et al.* 1990; Moran *et al.* 1994).

Lake Malawi (Nyasa) contains hundreds of species of endemic haplochromine cichlid fishes, representing one of the world's most spectacular examples of explosive speciation. It has been suggested that allopatric mechanisms of speciation can account for the observed rapid speciation and radiation (Fryer & Iles 1972; Ribbink *et al.* 1983).

Recent molecular genetic evidence of substantial substructuring of populations linked to habitat discontinuities over very small geographical scales (Van Oppen *et al.* 1997b; Arnegard *et al.* 1999; Markert *et al.* 1999), appears to support such a hypothesis of allopatric speciation within the 400 or more species of habitat specialists belonging to the 'mbuna' group. Evidence from other lakes suggests that subdivision (Lake Tanganyika; Coulter 1991) or isolation of peripheral lagoons (Lake Nabugabo; see Greenwood 1965), resulting from water level changes, may have contributed to speciation through allopatric mechanisms. Theoretical considerations, however, suggest a possible role for sympatric mechanisms, applicable to the haplochromine cichlids, through the action of sexual selection (Turner & Burrows 1995; Higashi *et al.* 1999) or natural selection (Dieckmann & Doebeli 1999; Kondrashov & Kondrashov 1999).

The most likely candidates for the operation of non-allopatric mechanisms in subpopulation divergence will be species with widespread distributions occupying continuous habitats with no apparent barriers to migration. The Lake Malawi haplochromine species flock contains the only known truly pelagic cichlid species, piscivores and zooplanktivores of the genera *Rhamphochromis*, *Diplotaxodon* and *Pallidochromis*. Our studies of the distribution and ecology of these species indicate that all occur throughout this 500 km long lake in habitats that appear to be uniform throughout the distribution (R. L. Robinson, unpublished data). Morphological studies show no systematic variation within species with respect to geographical region, depth or inshore-offshore distributions, and no regional variation in male breeding colours (R. L. Robinson, unpublished data). All data thus far suggest that populations of the pelagic cichlids are potentially single panmictic units within the lake, with little, if any, opportunity for or indication of allopatric isolation and genetic divergence. Our studies indicate there to be at least 21

* Author and address for correspondence: School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK (p.shaw@rhnc.ac.uk).

† Present address: Department of Zoology, University of Malaya, Lembah Pantai, 59100 Kuala Lumpur, Malaysia.

species of pelagic cichlids within Lake Malawi: nine *Rhamphochromis*, 11 *Diplotaxodon* and one *Pallidochromis* (Turner 1996; G. F. Turner and R. L. Robinson, unpublished data). Molecular genetic studies have suggested that *Rhamphochromis* and *Diplotaxodon* are 'oligotypic' clades within the Lake Malawi flock, being distinct phylogenetically from the other more speciose clades (Moran *et al.* 1994). Unfortunately, the use of only single individuals to represent the pelagic genera in this work precludes the conclusion that the large number of pelagic species we now know to exist form monophyletic clades exclusive of all other ecotypes.

The aim of the present study was to establish, using molecular genetic markers, if substructuring occurs within the current populations of pelagic cichlids in Lake Malawi, and therefore if the potential for allopatric processes of genetic divergence exists for these taxa. An important additional aim was to establish whether the pelagic species form a fully inclusive monophyletic group, exclusive of all non-pelagic species. Such a demonstration would suggest that radiation of the pelagics has occurred within the pelagic realm, and therefore that speciation most likely occurred under conditions of population structuring concurrent with present populations, i.e. in sympatry. We establish that populations of three species of pelagic cichlids of the genus *Diplotaxodon* are essentially single panmictic units within the lake, and do not indicate the substructuring prerequisite for allopatric genetic divergence during speciation. We also demonstrate that extant species of both pelagic genera (*Pallidochromis* falls within the *Diplotaxodon*) form monophyletic clades, and that there is evidence for all pelagics forming an exclusively pelagic clade within Lake Malawi. We suggest the data represent strong, if indirect, evidence for the action of sympatric speciation on a scale not reported previously.

2. METHODS

(a) Population structuring

Levels of genetic diversity within and between samples of Lake Malawi pelagic cichlids were assessed using variation at microsatellite loci. Eleven samples, representing the geographical range, plus both inshore and offshore populations, of three species of *Diplotaxodon* (*D. limnothrissa*, *D. macrops* and *D.* 'offshore') were screened (see table 1 and figure 1 for details and locations). These three species were used as they could be obtained in large numbers from single trawls, although we consider them representative of pelagic cichlids in general. No *Rhamphochromis* species were screened as sample sizes per location sampling event were generally too small to provide accurate data with such variable marker loci, and because the only abundant species (*Rhamphochromis longiceps*) was found to comprise a species complex whose individuals could not be unambiguously separated on site (R. L. Robinson, unpublished data), which would render population genetic data unreliable.

Six microsatellite loci were screened: Pzeb1, Pzeb2, Pzeb3 and Pzeb4 (Van Oppen *et al.* 1997a), to allow comparison with data from a previous study of Lake Malawi rocky shore mbuna cichlids (Van Oppen *et al.* 1997b), plus two loci (UNH130 and UNH154; Lee & Kocher 1996) showing most consistent amplification in an initial survey of 15 cichlid-derived loci. Total DNA was extracted from ethanol-preserved fin clips using a salting-out method modified from Bruford *et al.* (1992). Genotypes of

each individual at each locus were revealed by polymerase chain reaction (PCR) amplification of a 1/100 dilution of total DNA with locus-specific primers, one of which was labelled with fluorescent dye (Cy5), using the following protocol: 2 min at 93 °C, followed by 35 cycles of 30 s at 91 °C, 45 s at 55 °C, 10 s at 72 °C. PCR products were run out on an ALFexpress™ automated sequencer (Pharmacia Biotech, Amersham, UK) and relative mobility (in base pairs) of alleles scored against internal size markers and standard individuals with Fragment Manager™ software (Pharmacia Biotech).

Standard population genetic analyses were performed to check all samples for evidence of linkage between loci and departure from random (outcrossing) genotypic expectations. Tests for significant genetic heterogeneity between samples were performed using Exact Tests of allele and genotype frequencies, and via the departure of measures of genetic differentiation (F_{ST} and R_{ST}) from zero, using Genepop (Raymond & Rousset 1995), FSTAT (Goudet 1995) and RST-CALC (Goodman 1997) analysis packages. Estimates of R_{ST} were calculated in order to test the prediction (Slatkin 1995) that, in the case of significant genetic differentiation among samples, non-stepwise mutation model estimators (F_{ST}) would underestimate differentiation as measured by microsatellite loci.

(b) Phylogenetic methods

An important issue concerning elucidation of speciation mechanisms in a group such as the Lake Malawi cichlids is the establishment of monophyly of the particular groups of interest, in the present case the pelagic species. To test whether all recognized pelagic species within Lake Malawi do indeed form monophyletic lineages corresponding to the two recognized genera *Rhamphochromis* and *Diplotaxodon*, we carried out a phylogenetic reconstruction using mitochondrial DNA (mtDNA) sequence data. As Lake Malawi cichlids are known to have limited informative variability in the hypervariable part of the control region (CR) (Meyer *et al.* 1990), to maximize phylogenetic signal we sequenced two different regions: 872 bp encompassing the complete CR; and 981 bp of the NADH-2 (ND2) gene region. Full details of the sequencing study will be presented elsewhere but, briefly, mtDNA gene trees of the Lake Malawi haplochromine species flock were obtained as follows.

In order to confirm true monophyly of the pelagics we attempted to include as many different species as possible, represented by several individuals collected from geographically disparate sites within the lake. We also included as many different species of other Lake Malawi cichlids, covering as broad a range of morphotypes and ecotypes as possible, to clearly establish the position of the pelagics within the Malawi flock. In total 311 (CR)/113 (ND2) individual sequences were obtained, representing 77/18 *Rhamphochromis* (eight species), 81/26 *Diplotaxodon*/*Pallidochromis* (11 species), and 142/58 other Lake Malawi cichlids (71 species, 35 genera).

(c) Sequencing

An initial round of PCR amplification of the target region was conducted to produce template DNA suitable for sequencing. For CR a 1.1 kb region was amplified using primers FISH L15926 (5'-gAgCgCCggTCTTgTAAKCC, modified from L15926 of Kocher *et al.* (1989)) and H00650 (5'-TgATAgTAAAgTCAggACCAAgC, modified from H00651 of Kocher *et al.* (1989)), and for ND2 a 1.2 kb region was amplified using primers ASN and GLN (Kocher *et al.* 1995), in a protocol of 3 min at 94 °C followed by 30 cycles of 30 s at 93 °C, 45 s at

Table 1. Genetic variability within samples of three species of *Diplotaxodon* at six microsatellite loci

(Sample sites are indicated in figure 1.)

sample	n	locus											
		Pzeb1		Pzeb2		Pzeb3		Pzeb4		UNH130		UNH154	
		A	H _o	A	H _o	A	H _o	A	H _o	A	H _o	A	H _o
<i>D. limnothrissa</i>													
D.lim14	95	32	0.79	33	0.72	12	0.61	16	0.75	38	0.93	34	0.83
D.lim16	93	38	0.83	35	0.71	13	0.66	16	0.67	40	0.91	40	0.89
D.lim4	97	40	0.91	35	0.79	10	0.64	18	0.64	37	0.91	34	0.82
D.lim37	95	39	0.87	39	0.82	10	0.70	12	0.66	41	0.94	40	0.85
D.lim38	95	37	0.82	34	0.76	10	0.62	20	0.75	38	0.89	38	0.78
overall	475	48	0.85	46	0.76	17	0.65	23	0.69	45	0.91	50	0.83
allele range (bp)	—	120–230		217–325		308–356		103–155		162–252		93–195	
<i>D. macrops</i>													
D.mac39	95	43	0.81	30	0.76	7	0.65	15	0.73	41	0.90	31	0.87
D.mac40	60	41	0.83	28	0.76	9	0.58	11	0.83	35	0.82	26	0.89
D.mac42	86	42	0.91	28	0.77	9	0.64	15	0.74	37	0.90	33	0.92
overall	241	51	0.85	35	0.77	9	0.63	19	0.77	43	0.88	38	0.90
allele range (bp)	—	126–248		215–299		312–332		103–171		164–256		93–193	
<i>D. 'offshore'</i>													
D.off9	88	44	0.86	35	0.73	9	0.65	17	0.78	40	0.93	29	0.92
D.off7	60	41	0.85	29	0.75	6	0.47	15	0.88	36	0.85	32	0.87
D.off1	79	43	0.83	29	0.66	7	0.54	19	0.79	38	0.92	37	0.90
overall	227	51	0.85	39	0.71	10	0.57	21	0.81	45	0.91	44	0.90
allele range (bp)	—	126–230		209–299		308–330		109–213		162–270		93–187	

65 °C/55 °C (CR/ND2) and 30 s at 72 °C. Template DNA was isolated using Qiagen PCR-cleanup kits (Qiagen, Dorking, UK). Two hundred nanograms of template DNA were cycle sequenced employing Cy5-labelled primers: CR, THR2 (5'-CCCCTAACTCCCAAAGCTAg, modified from THR of Kocher *et al.* (1993); L16500 (5'-ATTATTCTTggCATCTggTTCC); H00650 (as above) and TDKD (Kocher *et al.* 1993); ND2, ND2b, TRP and ND2a (Kocher *et al.* 1995); cycle time 2 min at 93 °C, followed by 20 cycles of 15 s at 91 °C, in 30 s at 65 °C, and 45 s at 72 °C. Products were run out and scored on an ALFexpressTM automated sequencer.

All sequence profiles for a particular region from an individual were aligned within the ALFWINTM sequencing package (Pharmacia Biotech) and a consensus produced. All individual consensus sequences were aligned using ESEE (Cabot & Beckenbach 1989). Trees were generated using neighbour joining (NJ) using Kimura–Nei distance, maximum-parsimony (MP) and maximum-likelihood (ML) algorithms within PAUP* 4.0 (Swofford 1998). Statistical support for branches was estimated using bootstrap resampling for NJ and MP or quartet puzzling for ML (1000 iterations).

3. RESULTS

(a) Microsatellite variation

Levels of genetic variability at the six microsatellite loci screened were high, and remarkably consistent both between samples within species and among the three species (table 1). Allele size ranges were consistent across species at all six loci, other than for rare large alleles distorting the upper size limit occasionally (table 1). Global estimates of mean number of alleles observed (*A*),

 Table 2. Global estimates (all samples combined) of *A*, *H_o* and *H_e* for the three species

species	n	A	H _o	H _e
<i>D. limnothrissa</i>	475	37.7	0.78	0.88
<i>D. macrops</i>	241	32.3	0.79	0.87
<i>D. 'offshore'</i>	227	34.7	0.79	0.89

observed heterozygosity (*H_o*) and expected heterozygosity (*H_e*) for the three species are shown in table 2.

No evidence of linkage between alleles at different loci was found, so all loci can be considered independent markers. Fifteen out of 66 locus by sample combinations displayed significant departures from Hardy–Weinberg (H–W) genotype expectations (all heterozygote deficits) at *p* < 0.05 (Bonferroni corrected). However, 11 of these departures are at loci (Pzeb1 and Pzeb2) known to possess null (non-amplifying) alleles in Lake Malawi cichlids (Van Oppen *et al.* 1997b), and a further three are at loci (UNH130 and UNH154) developed from a Tilapiine cichlid. It may be assumed therefore that the deficits of heterozygotes most likely result from the presence of null alleles. The presence of low frequencies of non-amplifying alleles does not affect the test for population differentiation under the null hypothesis of allele frequency homogeneity, although all subsequent testing of population data was conducted through permuting genotypes rather than alleles, as suggested by Weir (1990), to take into account the departures from H–W expectations.

Table 3. *Exact Test probabilities of allele frequency homogeneity and F_{ST} -values, at six microsatellite loci and over all loci combined, among samples of three *Diplotaxodon* species*

(a) Exact Test probabilities of allele frequency homogeneity							
species	locus						
	Pzeb1	Pzeb2	Pzeb3	Pzeb4	UNH130	UNH154	combined
<i>D. limnothrissa</i>	0.0874	0.0570	0.4067	0.1435	0.4095	0.0010	0.0074
<i>D. macrops</i>	0.0507	0.7547	0.9818	0.0033	0.6168	0.7207	0.0748
<i>D. 'offshore'</i>	0.3246	0.0507	0.4340	0.7685	0.6326	0.2696	0.3041

(b) F_{ST} -values								
species	locus							p^a
	Pzeb1	Pzeb2	Pzeb3	Pzeb4	UNH130	UNH154	overall	
<i>D. limnothrissa</i>	0.0007	0.0014	0.0018	0.0012	0.0000	0.0022	0.0012	0.017
<i>D. macrops</i>	0.0015	-0.0020	-0.0045	0.0081	-0.0004	-0.0005	0.0005	0.261
<i>D. 'offshore'</i>	-0.0006	0.0005	0.0079	-0.0034	-0.0009	0.0017	0.0003	0.291

^a p = probability of overall F_{ST} not > 0, for 2500 permutations.

(b) Population structuring

Exact Tests of allele frequencies, and estimates of genetic differentiation (F_{ST} and R_{ST}) among samples indicate that there is no substantial genetic structuring within populations of any of the three *Diplotaxodon* species examined. In most cases estimates of R_{ST} were an order of magnitude greater than F_{ST} -values, but still extremely low and not significantly different from zero, and so are not presented unless they indicate contrary to F_{ST} results. In summary the main results were as follows.

- (i) *D. limnothrissa*. Over all loci, both Exact Tests (table 3a, $p = 0.0074$) and F_{ST} estimates (table 3b, $F_{ST} = 0.0012$, p that F_{ST} not > 0 = 0.017) indicate statistically significant differences among the five samples screened, although R_{ST} indicates no significant heterogeneity ($R_{ST} = 0.0004$, p that R_{ST} not > 0 = 0.364). There is, however, little evidence of substantial and systematic genetic differences among samples indicating population structuring. The Exact Test is extremely conservative, and examination of single locus results indicates the overall value to be due solely to one locus (table 3a, UNH154). Finally, examination of pairwise F_{ST} -values between samples (table 4a) shows only a single significant difference (F_{ST} between D.lim4 and D.lim38), which becomes non-significant after Bonferroni correction of probabilities for multiple tests, and no obvious systematic pattern of genetic differences between samples based on such divisions as northern versus southern, or inshore versus offshore populations.
- (ii) *D. macrops*. Over all loci, both Exact Tests (table 3a, $p = 0.0748$) and F_{ST} estimates (table 3b, $F_{ST} = 0.0005$, p that F_{ST} not > 0 = 0.261) indicate there to be no significant differences among the samples screened. Pairwise tests (table 4b) also confirm a complete lack of genetic differentiation among samples.
- (iii) *D. 'offshore'*. Over all loci, both Exact Tests (table 3a, $p = 0.3041$) and F_{ST} estimates (table 3b, $F_{ST} = 0.0003$, p that F_{ST} not > 0 = 0.291) indicate there to be no

Table 4. *Pairwise F_{ST} -values, over all six microsatellite loci combined, for samples of *D. limnothrissa*, *D. macrops* and *D. 'offshore'**

(a) <i>D. limnothrissa</i>				
	D.lim14	D.lim16	D.lim4	D.lim37
D.lim14	—	—	—	—
D.lim16	0.0015	—	—	—
D.lim4	0.0020	0.0005	—	—
D.lim37	0.0013	0.0009	0.0012	—
D.lim38	0.0005	0.0009	0.0026 ^a	0.0009

(b) <i>D. macrops</i>		
	D.mac39	D.mac40
D.mac39	—	—
D.mac40	0.0002	—
D.mac42	0.0000	0.0020

(c) <i>D. 'offshore'</i>		
	D.off	9 D.off7
D.off9	—	—
D.off7	0.0015	—
D.off1	0.0004	0.0000

^a $p < 0.05$ that F_{ST} not > 0 for 2500 permutations; no tests are significant at $p < 0.05$ after Bonferroni correction for multiple tests within half matrices.

significant differences among the three samples screened. Pairwise tests (table 4c) also confirm a complete lack of genetic differentiation among samples.

Figure 1 displays sample allele frequency profiles at locus Pzeb4 in all samples screened for the three species. Locus Pzeb4 has been found to be a sensitive indicator of

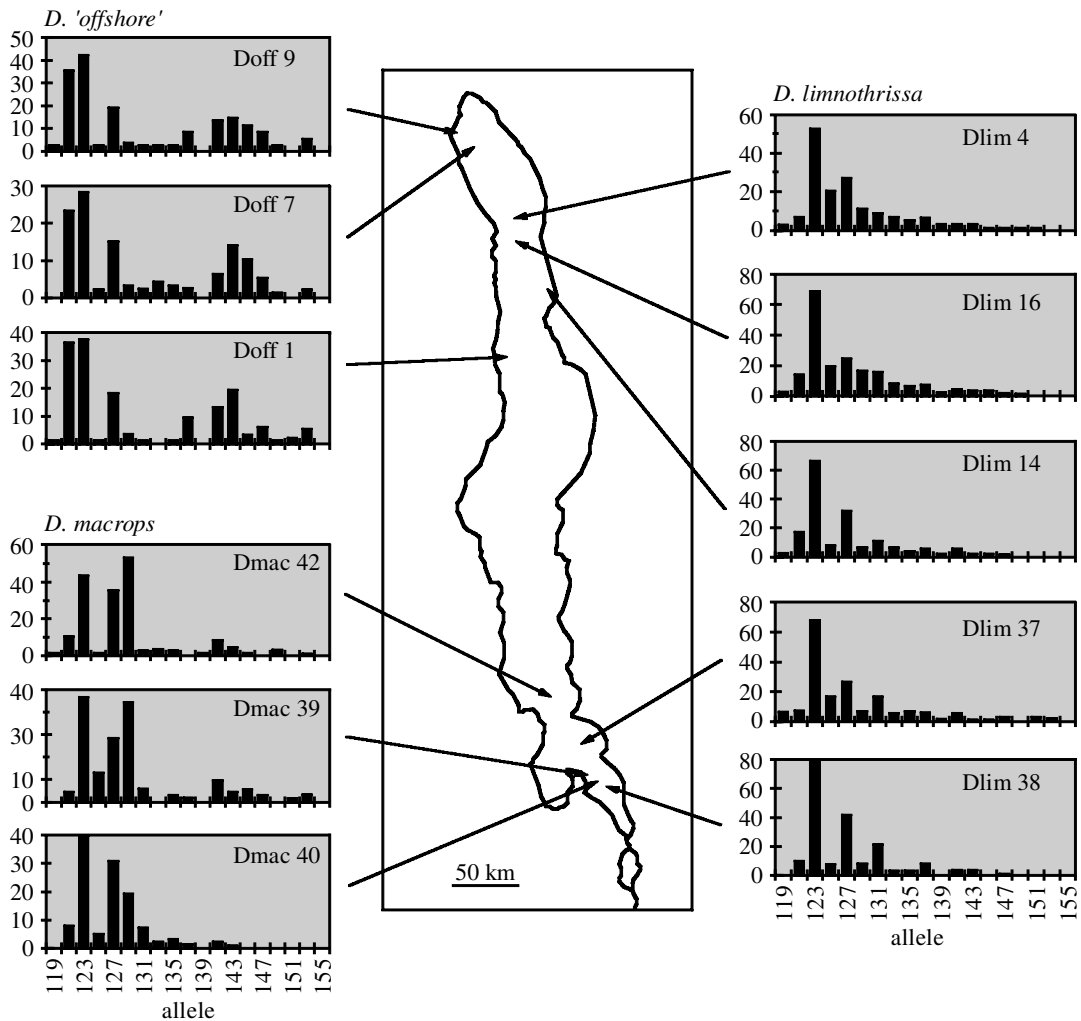


Figure 1. Allele frequencies at locus Pzeb4 for samples of *D. limnothrissa*, *D. 'offshore'* and *D. macrops* from Lake Malawi. Allele sizes (bp) are matched to those of Van Oppen *et al.* (1997a) by cross-referencing of samples between the two studies. Sample sizes are given in table 1.

population substructuring, where present, in other Lake Malawi cichlid species (Van Oppen *et al.* 1997b; P. W. Shaw, unpublished data), and indeed displays species-specific frequency profiles among the three species examined in this study. From figure 1 it is clear that within species all samples exhibit remarkable conformity in allele composition and frequency throughout the geographical range sampled—this pattern is also true for the other five loci (not shown). Substantial and significant differences are observed in allele frequencies between the three species at five out of six loci (Pzeb2 shows no significant difference), as illustrated by the multilocus estimates of F_{ST} and R_{ST} (table 5).

(c) Phylogeny of Lake Malawi pelagic cichlids

For the purposes of the present study we only wish to consider whether the pelagic species form reciprocally monophyletic clades with the rest of the Lake Malawi cichlid species flock—details of the full sequencing study will be presented elsewhere (sequences used are deposited in GenBank).

Figure 2 shows a cladogram constructed from a consensus of trees derived from NJ, MP and ML analyses of 88 individuals for which both CR and ND2 gene region

Table 5. Pairwise F_{ST} -values (below diagonal) and R_{ST} -values (above diagonal), over all six microsatellite loci combined, between the three *Diplotoxodon* species

	<i>D. limnothrissa</i>	<i>D. macrops</i>	<i>D. 'offshore'</i>
<i>D. limnothrissa</i>	—	0.1247 ^a	0.1207 ^a
<i>D. macrops</i>	0.0110 ^a	—	0.0509 ^a
<i>D. 'offshore'</i>	0.0116 ^a	0.0150 ^a	—

^a $p < 0.001$ that F_{ST}/R_{ST} not > 0 for 2500 permutations, after Bonferroni correction for multiple tests within half matrices.

sequences were obtained. For all three analyses (NJ, MP, ML) the two DNA region sequences were combined, without differential weighting given to transversions or transitions or to codon positions in the ND2 sequence; previous analyses (M. R. Idid, unpublished data) had demonstrated that use of either region separately and/or weighting did not significantly affect definition of the six main clades within the Malawi flock. The consensus tree clearly illustrates the divergence (100% support for NJ and MP, 85% for ML) of the Lake Malawi flock from representative cichlids of Lake Tanganyika (*Bathybates*, *Julidochromis*, *Tropheus*, *Lobochilotes*) and generalized riverine

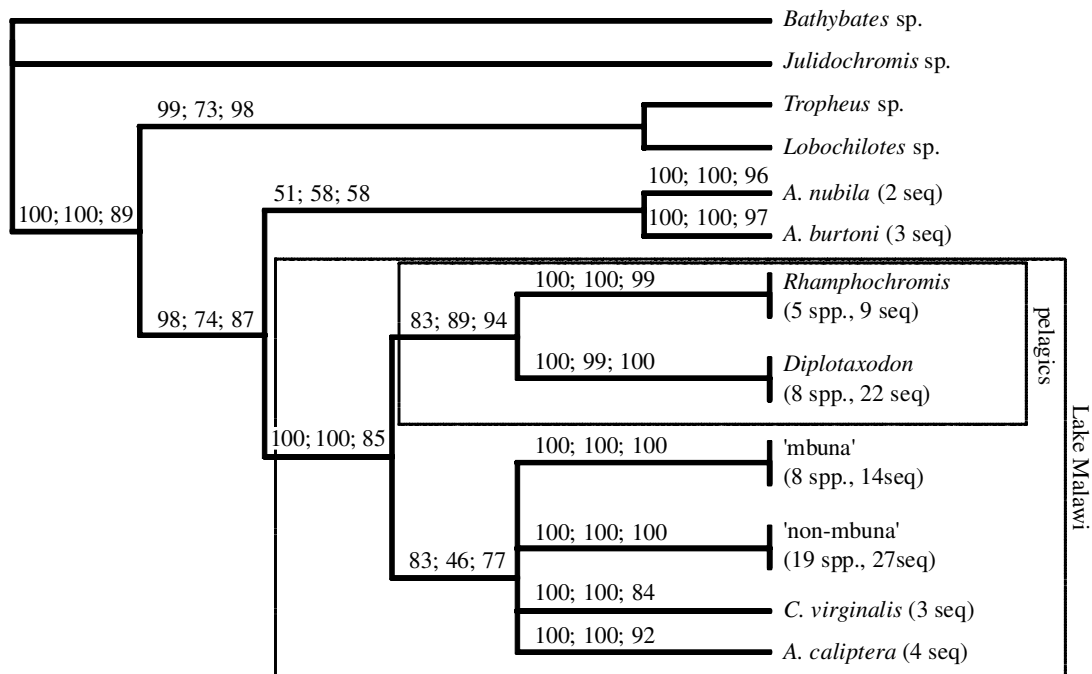


Figure 2. Consensus cladogram illustrating the phylogenetic relationship of the pelagic cichlid genera to other Lake Malawi cichlid major clades, and outgroups representing Lake Tanganyika (*Bathybates*, *Julidochromis*, *Tropheus*, *Lobochilotes*) and generalized riverine haplochromine cichlids (*Astatotilapia nubilus*, *Astatotilapia burtoni*). The schematic represents a consensus of trees derived from NJ, MP and ML analyses of 88 combined mtDNA CR and ND2 gene region sequences (see § 2). Figures above branches indicate support derived from per cent bootstrap (NJ of Tamura–Nei distance, 1000 replicates), per cent bootstrap (MP, 250 replicates) and per cent puzzling quartet (ML, 1000 replicates). All sequences employed have been deposited in GenBank.

haplochromine cichlids (*Astatotilapia nubilus*, *Astatotilapia burtoni*). The tree also shows very strong support for the clear definition of six main clades within the Malawian flock: the two pelagic genera *Rhamphochromis* and *Diplotaxodon* (the single *Pallidochromis* falls within the latter clade); the mbuna group (mainly shallow-water, rock-dwelling species, corresponding to clade B of Meyer *et al.* (1990) and Moran *et al.* (1994)); the non-mbuna group (mostly sand-dwelling demersal species, corresponding to clade A of Meyer *et al.* (1990) and Moran *et al.* (1994)); members of the *Copadichromis virginalis* complex; the generalized riverine cichlid found within Lake Malawi and surrounding areas, *Astatotilapia caliptera*. The important point is that no individuals of pelagic species fall outside their genus clade, and no individuals of other species fall within the pelagic clades. This reciprocal monophyly is not challenged whichever individuals are included in the analysis, or if all individuals from the larger data set for each DNA region separately are used (support for the pelagic clades is always > 96%). Figure 2 also shows that the two pelagic genera fall on a single branch separate, with reasonably high support (83–94%), from the other four Malawi clades. The branch uniting the two pelagic genera is confirmed in all analyses (NJ, MP, ML) of each DNA region separately with support > 60%.

4. DISCUSSION

The first of two major findings of the present study is that data indicate populations of the pelagic cichlid species of Lake Malawi comprise effectively of single panmictic units within the lake. All three of the species

studied here have lake-wide distributions, as do all of the other 21 species of Lake Malawi pelagic cichlids identified to date (R. L. Robinson, unpublished data). All three *Diplotaxodon* species exhibited exceedingly small estimates of genetic differentiation (F_{ST}), which were not significantly different from zero among samples. Although not the most robust measure of gene flow, N_m estimates derived from F_{ST} -values ($N_m = 208.2–833.1$) indicate effective migration rates among areas within species to be substantially above 1, levels which would normally be assumed to indicate panmixia (Wright 1969). A slight indication of genetic differentiation among samples was found in one species (*D. limnothrissa*), but differences showed no biologically meaningful distribution with regard to geography or limnology (see Hedrick 1999). Comparison of overall F_{ST} -values with those observed in other Lake Malawi cichlid species, which display distinct population substructuring, emphasizes how insignificant any indications of differentiation are in pelagic species, especially in view of the difference in geographical scale over which the samples were collected (overall $F_{ST} = 0.0003–0.0012$, over a range of 460 km in this study); overall $F_{ST} = 0.151$ over 0.9–42.4 km in *Melanochromis auratus* (Markert *et al.* 1999); overall $F_{ST} = 0.079$ over 0.6–10.4 km in *Labeotropheus fuelleboni* (Arnegard *et al.* 1999); $F_{ST} = 0.007–0.016$ over 3 km in four species of the *Pseudotropheus* complex (Van Oopen *et al.* 1997b).

One possible bias in comparison of F_{ST} -values among species is that this estimator is affected by levels of variability within samples, such that if one species exhibits significantly higher variability than a second species the maximum possible value of F_{ST} is lower in the

former. As the figures in table 1 show, levels of variability at the six microsatellite loci screened are very high in the pelagic cichlids ($H_e = 0.87\text{--}0.89$, $A = 32.3\text{--}37.7$). Comparison with the studies of other cichlid species noted above, however, shows that levels of variability are equally high in these species: mean H_e across the four species in Van Oppen *et al.* (1997b) for the four loci screened in common with the present study were 0.95, 0.94, 0.74 and 0.56 compared with 0.85, 0.75, 0.62 and 0.76 reported here. Allele size ranges observed in the pelagics are not substantially different, and almost completely overlapping those observed at the same loci in mbuna species (Van Oppen *et al.* 1997b). So it would appear that between-species variability bias is not a factor in the present results.

The second major finding of the present study is that the 21 or more species of pelagic cichlids of Lake Malawi fall within two monophyletic clades, with no pelagics falling outside these lineages and no non-pelagic species falling within them. There is also some support for the two clades having a common ancestor separate from the other clades within the Lake Malawi cichlid flock, i.e. that the pelagics as a whole have arisen from a single lineage. This phylogeny suggests that the radiation leading to the current group of pelagic species has taken place within the pelagic realm.

Although our wider sequencing studies covered 90% of the pelagic species, and *ca.* 30% of species and 85% of genera of the non-mbuna non-pelagic haplochromine cichlids of Lake Malawi, we have only sequenced about 15–20% of endemic species thought to occur in this flock—is it possible that further sampling could refute monophyly of the pelagics? We only included nine species of the 350–400 known of the mbuna group, but previous studies have consistently placed all such species in a clade that excludes pelagic species (our mbuna clade corresponding to clade B in Meyer *et al.* (1990) and Moran *et al.* (1994)). Similarly, of the nine genera of non-mbuna not covered here, four were placed in the non-mbuna group (clade A) by Moran *et al.* (1994) and the remaining five represent only five to seven species with close affinities to non-mbuna we have examined. We therefore think it unlikely that further sampling will compromise the monophyly of the Lake Malawi pelagics. One final caveat is that our phylogeny is based upon, effectively, a single locus (mtDNA)—corroborative phylogenies from nuclear genes will be desirable to confirm the conclusions presented here.

The conclusion of the current study is therefore that extant pelagic species show no evidence of population structuring within the lake, such that possibilities for divergence under allopatry do not occur. In addition, all pelagic species form an endemic clade, which it is most parsimonious to believe resulted by radiation from an ancestral pelagic taxon, and that speciation therefore proceeded under conditions of sympatry. Are there other possible historical conditions under which allopatric processes may have generated the current pelagic radiation? Water level changes within the African Great Lakes may have resulted in the formation of separate basins, as suggested for Lake Tanganyika (Coulter 1991), which would allow divergence of populations. Several studies of Lake Tanganyikan cichlids, however, have not indicated a

complete link between ancestral genetic divergence and current species boundaries (Meyer *et al.* 1996; Ruber *et al.* 1999), suggesting that isolation is not necessarily associated with speciation. Lake Malawi is known to have experienced a fall in water level of *ca.* 300 m around 40 000 years ago, but it would have taken a drop of more than 400 m to split the lake into two basins, the maximum number possible (Meyer *et al.* 1990). There is no other evidence that Lake Malawi has been subdivided historically in a manner that could have generated the currently observed number of pelagic species. Alternatively, allopatric speciation might occur in isolated peripheral lagoons, as is believed to have happened with Lake Nabugabo at the margin of Lake Victoria (Greenwood 1965). Juveniles of several *Rhamphochromis* species are found in Lake Malombe and Chia Lagoon, the only large bodies of still water connected to Lake Malawi. But, while isolation of *Rhamphochromis* populations in such peripheral water bodies seems possible, this is far less likely to have occurred with *Diplotaxodon*, which appear to be confined to deep-water offshore habitats at all stages in their life histories. Finally, population-specific homing to separate breeding and spawning sites within the lake could allow allopatric divergence, with free mixing of individuals throughout the lake outside breeding seasons giving the appearance of a single homogeneous population, as observed in some marine pelagic fishes (Nesbø *et al.* 2000). At present this hypothesis cannot be excluded, but individuals of *D. limnothrissa* in mature breeding condition or carrying eggs (they are maternal mouth-brooders) have been observed throughout the offshore areas of the lake at all times of year (G. F. Turner, unpublished data), so specific inshore breeding sites appear unlikely.

Theoretical studies are suggesting that sympatric speciation is plausible under realistic conditions of ecology and population genetic variation, and may be driven by disruptive sexual selection (Turner & Burrows 1995; Higashi *et al.* 1999), disruptive natural (ecological) selection plus assortative mating (Kondrashov & Kondrashov 1999; Dieckmann & Doebeli 1999) or a combination of disruptive natural and sexual selection (Van Doorn *et al.* 1998). The small species flocks (nine to 11 species) within crater lakes (Schliewen *et al.* 1994) are regarded as among the most convincing demonstrations in support of sympatric speciation in any animal group. We consider our demonstration of panmictic population structures of species with lake-wide distributions in an apparently unstructured habitat, within a monophyletic clade of pelagic species, is similarly strong evidence that sympatric processes have also been involved in the large-scale explosive speciation events of the East African lakes.

This work was supported by a grant from the Environment Research Programme of the Department for International Development (UK) and a Malaysian Government scholarship (to M.R.I.). The Department of Fisheries (Malawi), the Tanzanian Fisheries Research Institute, the Instituto de Investigação Pesqueira (Mozambique), and the Lake Malawi/Niassa GEF Biodiversity Project provided clearance and collaboration. We would like to thank M. Harley for technical support, and C. Cruz, W. R. T. Darwall, M. Day, M. J. Genner, M. Hanssens, J. Kihedu, D. Mandere, J. Mwambungu, B. P. Ngatunga, H. Ngulube, A. J. Ribbink and J. Snoeks for assistance with sample collection and logistics.

REFERENCES

- Arnegard, M. E., Markert, J. A., Danley, P. D., Stauffer Jr, J. R., Ambali, A. J. & Kocher, T. D. 1999 Population structure and colour variation of the cichlid fish *Labeotropheus fuelleborni* Ahl along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proc. R. Soc. Lond. B* **266**, 119–130.
- Bruford, M. W., Hanotte, O., Brookfield, J. F. Y. & Burke, T. 1992 Single locus and multilocus DNA fingerprinting. In *Molecular genetic analysis of populations—a practical approach* (ed. A. R. Hoelzel), pp. 227–229. Oxford, UK: IRL Press.
- Cabot, E. L. & Beckenbach, A. T. 1989 Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Comp. Appl. Biosci.* **5**, 233–234.
- Coulter, G. W. 1991 *Lake Tanganyika and its life*. Oxford University Press.
- Coyne, J. A. 1992 Genetics and speciation. *Nature* **355**, 511–515.
- Dieckmann, U. & Doebeli, M. 1999 On the origin of species by sympatric speciation. *Nature* **400**, 354–357.
- Fryer, G. & Iles, T. D. 1972 *The cichlid fishes of the Great Lakes of Africa*. Edinburgh, UK: Oliver & Boyd.
- Goodman, S. J. 1997 RST CALC: a collection of computer programs for calculating unbiased estimates of genetic differentiation and gene flow from microsatellite data and determining their significance. *Mol. Ecol.* **6**, 881–885.
- Goudet, J. 1995 F-STAT (version 1.2): a computer program to calculate F-statistics. *J. Hered.* **86**, 485–486.
- Greenwood, P. 1965 The cichlid fishes of Lake Nabugabo, Uganda. *Bull. Br. Mus. Nat. Hist. Zool.* **12**, 315–357.
- Hedrick, P. W. 1999 Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313–318.
- Higashi, M., Takimoto, G. & Yamamura, N. 1999 Sympatric speciation by sexual selection. *Nature* **402**, 523–526.
- Knox, E. B. & Palmer, J. D. 1995 Chloroplast DNA variation and the recent radiation of giant senecios (Asteraceae) on the tall mountains of Eastern Africa. *Proc. Natl Acad. Sci. USA* **92**, 10 349–10 353.
- Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, A., Villablanca, F. X. & Wilson, A. C. 1989 Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl Acad. Sci. USA* **86**, 6196–6200.
- Kocher, T. D., Conroy, J. A., McKaye, K. R. & Stauffer, J. R. 1993 Similar morphologies of cichlid fishes in lakes Tanganyika and Malawi are due to convergence. *Mol. Phyl. Evol.* **2**, 158–165.
- Kocher, T. D., Conroy, J. A., McKaye, K. R., Stauffer, J. R. & Lockwood, S. F. 1995 Evolution of NADH dehydrogenase subunit-2 in East-African cichlid fish. *Mol. Phyl. Evol.* **4**, 420–432.
- Kondrashov, A. S. & Kondrashov, F. A. 1999 Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**, 351–354.
- Lee, W.-J. & Kocher, T. D. 1996 Microsatellite DNA markers for genetic mapping in *Oreochromis niloticus*. *J. Fish Biol.* **49**, 169–171.
- Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. 1998 Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–2118.
- Markert, J. A., Arnegard, M. E., Danley, P. D. & Kocher, T. D. 1999 Biogeography and population genetics of the Lake Malawi cichlid *Melanochromis auratus*: habitat transience, philopatry and speciation. *Mol. Ecol.* **8**, 1013–1026.
- Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. 1990 Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**, 550–553.
- Meyer, A., Knowles, L. L. & Verheyen, E. 1996 Widespread geographical distribution of mitochondrial haplotypes in rock-dwelling cichlid fishes from Lake Tanganyika. *Mol. Ecol.* **5**, 341–350.
- Moran, P., Kornfield, I. & Reinthal, P. N. 1994 Molecular systematics and radiation of the haplochromine cichlids (Teleostei: Perciformes) of Lake Malawi. *Copeia* **1994**, 274–288.
- Nesbø, C. L., Rueness, E. K., Iversen, S. A., Skagen, D. W. & Jakobsen, K. S. 2000 Phylogeography and population history of Atlantic mackerel (*Scomber scombrus* L.): a genealogical approach reveals genetic structuring among the eastern Atlantic stocks. *Proc. R. Soc. Lond. B* **267**, 281–292.
- Raymond, M. & Rousset, F. 1995 Genepop (version-1.2)—population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249.
- Ribbink, A. J., Marsh, B. A., Marsh, A. C., Ribbink, A. C. & Sharp, B. J. 1983 A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *S. Afr. J. Zool.* **18**, 149–310.
- Ruber, L., Verheyen, E. & Meyer, A. 1999 Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. *Proc. Natl Acad. Sci. USA* **96**, 10 230–10 235.
- Schliwien, U. K., Tautz, D. & Paabo, S. 1994 Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**, 629–632.
- Slatkin, M. 1995 A measure of population subdivision based in microsatellite allele frequencies. *Genetics* **139**, 457–462.
- Swofford, D. L. 1998 *PAUP**. *Phylogenetic analysis using parsimony (* and other methods)*, v. 4. Sunderland, MA: Sinauer Associates.
- Taylor, E. B. & McPhail, J. D. 1999 Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA. *Biol. J. Linn. Soc.* **66**, 271–291.
- Turner, G. F. 1996 *Offshore cichlids of Lake Malawi*. Lauenau, Germany: Cichlid Press.
- Turner, G. F. & Burrows, M. T. 1995 A model of sympatric speciation by sexual selection. *Proc. R. Soc. Lond. B* **260**, 287–292.
- Van Doorn, G. S., Noest, A. J. & Hogeweg, P. 1998 Sympatric speciation and extinction driven by environment dependent natural selection. *Proc. R. Soc. Lond. B* **265**, 1915–1919.
- Van Oppen, M. J. H., Rico, C., Deutsch, J. C., Turner, G. F. & Hewitt, G. M. 1997a Isolation and characterization of microsatellite loci in the cichlid fish *Pseudotropheus zebra*. *Mol. Ecol.* **6**, 185–186.
- Van Oppen, M. J. H., Turner, G. F., Rico, C., Deutsch, J. C., Ibrahim, K. M., Robinson, R. L. & Hewitt, G. M. 1997b Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proc. R. Soc. Lond. B* **264**, 1803–1812.
- Weir, B. 1990 *Genetic data analysis*. Sunderland, MA: Sinauer Associates.
- Wright, S. 1969 *Evolution and genetics of populations. II. The theory of gene frequencies*. University of Chicago Press.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.