

Phylogeography of Lake Malawi cichlids of the genus *Pseudotropheus*: significance of allopatric colour variation

Peter F. Smith¹ and Irv Kornfield^{2*}

¹Department of Biological Sciences, and ²School of Marine Sciences, University of Maine, Orono, ME 04469-5751, USA

One of the most compelling features of the cichlid fishes of the African Great Lakes is the seemingly endless diversity of male coloration. Colour diversification has been implicated as an important factor driving cichlid speciation. Colour has also been central to cichlid taxonomy and, thus, to our concept of species diversity. We undertook a phylogeographical examination of several allopatric populations of the Lake Malawi cichlid *Pseudotropheus zebra* in order to reconstruct the evolutionary history of the populations, which exhibit one of two dorsal fin colours. We present evidence that populations with red dorsal fins (RT) are not monophyletic. The RT population defining the northern limit of the distribution has evidently originated independently of the southern RT populations, which share a common ancestry. This evidence of species-level colour convergence is an important discovery in our understanding of cichlid evolution. It implies that divergence in coloration may accompany speciation, and that allopatric populations with similar coloration cannot be assumed to be conspecific. In addition to this finding, we have observed evidence for introgression, contributing to current evidence that this phenomenon may be extremely widespread. Thus, in species-level phylogenetic reconstructions, including our own, consideration must be given to the potential effects of introgression.

Keywords: cichlid; *Pseudotropheus zebra*; *Pseudotropheus* microsatellite; speciation; sexual selection

1. INTRODUCTION

In Lake Malawi, Africa, the number of endemic cichlid species was recently estimated to be almost 800 (Snoeks 2000). Much of the species richness in Lake Malawi is reflected by geographical colour variation, first described in detail by Ribbink *et al.* (1983). The importance of male reproductive coloration in the origin and maintenance of new cichlid species in the African Great Lakes is virtually universally accepted. Coloration has been incorporated in models suggesting both sympatric (Turner & Burrows 1995; Seehausen *et al.* 1999) and allopatric (Dominey 1984; Danley *et al.* 2000) divergence. In spite of the recognized importance of colour, little is known about the phylogeographical relationships among allopatric colour variants and, thus, the forces that might drive divergence in coloration.

Taxonomic designation of allopatric colour variants is complicated by the extreme species diversity and the lack of species-level phylogenies for most groups of cichlids. Allopatric populations with subtle variations in coloration are often assigned to the same species, but the affinities of allopatric populations are not always immediately apparent (Turner *et al.* 2001). When localized colour variants cannot be easily allied to more broadly distributed taxa, they are sometimes elevated to specific status. These species usually have distributions limited to the type locality (often a single rock outcropping) and are known as 'narrow endemics' (Ribbink *et al.* 1983). This approach provides a taxonomic framework, but it does not illumi-

nate the ecological and evolutionary events that have influenced the origin and distribution of cichlid species. Unambiguous groupings of allopatric populations with shared evolutionary histories are necessary for such insights. Because of the difficulty in effectively grouping allopatric variants, and because of the unknown evolutionary stability of coloration, molecular markers are appropriate for generating phylogenetic information. Since ancestral polymorphisms are maintained within the species of the African Cichlidae, DNA sequence variation has not been widely successful in reconstructing ancestral relationships (Moran & Kornfield 1993; Parker & Kornfield 1997); thus, we employ microsatellite markers to investigate phylogenetic relationships.

Climatic variation in Africa's rift valley is associated with dramatic fluctuations in lake levels, a recognized factor that may promote intralacustrine allopatric divergence (Kornfield & Smith 2000). The effect of fluctuations in lake level on cichlid phylogeography has been extensively documented in Lake Tanganyika (Sturmbauer *et al.* 2001). In Lake Victoria, the origin of the entire species flock of haplochromine cichlids has been estimated at just 14 600 years before present (BP), based on inferences from lake-level data (Seehausen 2002) (but see Fryer (2001) for an opposing view). In Lake Malawi, one estimate indicated that the lake level was over 100 m below its current level just 200–300 years BP, when the southern portion of the lake would have been completely dry. This implies that many of the southern endemic species and variants have evolved in the dramatically short period since lake-level rise (Owen *et al.* 1990). However, more recent evidence suggests that the latest lake-level rise occurred much earlier (10⁶ years BP) (Finney *et al.* 1996).

* Author for correspondence (irvk@umit.maine.edu).

Table 1. Genetic diversity of *P. zebra* populations examined.

(*n*: sample size; H_e : expected heterozygosity (average gene diversity); H_o : observed heterozygosity; *A*: average number of alleles per locus. See figure 1 for pictures of RT and BB phenotypes and sample locations. Global positioning system coordinates were measured using World Geodetic System 1984 map datum.)

| sample | sample location | | code | phenotype | species name | <i>n</i> | H_e | H_o | <i>A</i> |
|-------------------|-----------------|-------------|------|-----------|-----------------------------------|----------|-------|-------|----------|
| | S latitude | E longitude | | | | | | | |
| Chiofu Bay | 13° 31.773' | 34° 51.962' | CH | BB | <i>P. zebra</i> | 70 | 0.842 | 0.846 | 22.67 |
| Nkhata Bay | 11° 36.370' | 34° 17.992' | NKB | BB | <i>P. zebra</i> | 71 | 0.830 | 0.836 | 22.67 |
| Namalenje Island | 13° 43.754' | 34° 38.462' | NAM | BB | <i>P. zebra</i> | 64 | 0.785 | 0.806 | 14.56 |
| Eccles Reef | 13° 45.991' | 34° 57.687' | ER | RT | <i>P. thapsinogen</i> | 72 | 0.830 | 0.834 | 19.89 |
| Nakantenga Island | 13° 55.218' | 34° 38.546' | NAK | RT | <i>P. pyrsonotos</i> | 71 | 0.865 | 0.856 | 23.11 |
| Mpanga Rocks | 10° 25.779' | 34° 16.720' | MR | RT | <i>P. emmiltos</i> | 70 | 0.832 | 0.820 | 21.00 |
| Namaso Bay | 14° 09.580' | 35° 00.000' | NMB | RT | <i>P. sandaracinos</i> | 67 | 0.790 | 0.791 | 21.11 |
| Mpanga Rocks | 10° 25.779' | 34° 16.720' | CZA | CZ | <i>P. sp.</i> 'zebra Chilumba' | 37 | 0.830 | 0.765 | 19.56 |
| Luwino Reef | 10° 26.775' | 34° 16.966' | CZB | CZ | <i>P. sp.</i> 'zebra Chilumba' | 55 | 0.848 | 0.783 | 25.00 |

Regardless of the exact timing, the depth profile of Lake Malawi is suggestive of a north-to-south colonization trend facilitated by rising water levels at some time in the recent past. The creation of new habitats during lake-level rise would certainly have been important in the evolution of the Lake Malawi flock; whether it was accompanied by evolution of new colour variants, or simply by colonization by northern forms, is unknown. Knowledge of the phylogeographical relationships among northern and southern populations and among allopatric colour variants is important in developing an understanding of the relative roles of colonization and speciation in the invasion of new habitat by cichlids in Lake Malawi.

Pseudotropheus zebra is a rock-dwelling Lake Malawi cichlid with lake-wide distribution and considerable geographical colour variation: more than 20 species or geographical colour variants are currently recognized (Konings 2001). Within the *P. zebra* complex are two well-known forms of zebras: the 'red top' (RT) and the 'blue-black' (BB). The RT zebra has a limited and highly disjunctive distribution spanning several hundred kilometres from north to south. Four geographically separated populations of the RT zebra have each been recognized as distinct 'narrow endemic' species based on morphometric differences (Stauffer *et al.* 1997). The BB zebra has a much wider distribution, which is mutually exclusive to that of the RT zebra. Although these two forms never co-occur (Konings 2001), a hybrid population has recently been discovered in one area of the lake (P. F. Smith & I. Kornfield, unpublished data).

Two explanations could be advanced for the distribution of the RT zebra relative to its BB counterpart: (i) repeated *in situ* divergence of RT and BB forms in each area of the lake; or (ii) colonization of the south by extant differentiated northern populations (figure 1). Similar species distributions in Lake Victoria have been examined as a means of inferring divergence processes. Our first proposal is similar to that of Seehausen & van Alphen (1999), that geographically proximate populations within a new habitat have originated via (sympatric) speciation. The

second proposal is aligned with that of Bouton (2000), that such species distributions in a new habitat can be explained by colonization by extant lineages. The theory of *in situ* divergence, in the present case, requires convergent evolution of the red dorsal fin. Convergence is evident across the species flocks of the African Great Lakes (Kocher *et al.* 1993), but colour convergence at the species level has not been documented. The alternative theory of colonization posits conservation of dorsal fin colour, and suggests that ecological factors other than lineage splitting (such as habitat preference, competitive exclusion, etc.) are responsible for the current distributions. Because allopatric colour variation is abundant in all genera of rock-dwelling cichlids from Lake Malawi, the resolution of these proposals has important implications for our general understanding of mbuna evolution.

We examined four populations of the RT *P. zebra*, and investigated their relationships with each other and with neighbouring BB *P. zebra* populations using microsatellite markers. We analysed two additional populations of a distinct species in the northern limit of the *P. zebra* distribution (figure 1) in order to investigate the potential contribution of hybridization to the distribution of genetic variation in *P. zebra*. In addition to resolving these proposals, our data may allow generalized inferences to be made regarding questions of north-south biogeography, colour pattern evolution and the potential effects of introgression in the Cichlidae of Lake Malawi.

2. MATERIAL AND METHODS

(a) Study species

The *P. zebra* species complex has been the subject of taxonomic uncertainty. Within *Pseudotropheus*, the subgenus *Maylandia* was created for this group (Meyer & Foerster 1984), but *Metriaclima* was later erected for this species complex, excluding the name *Maylandia* as the generic designator (Stauffer *et al.* 1997). However, *Metriaclima* is now considered a junior synonym of *Maylandia* (Conde & Gery 1999). We follow other recent studies (Rico & Turner 2002) in using the former valid

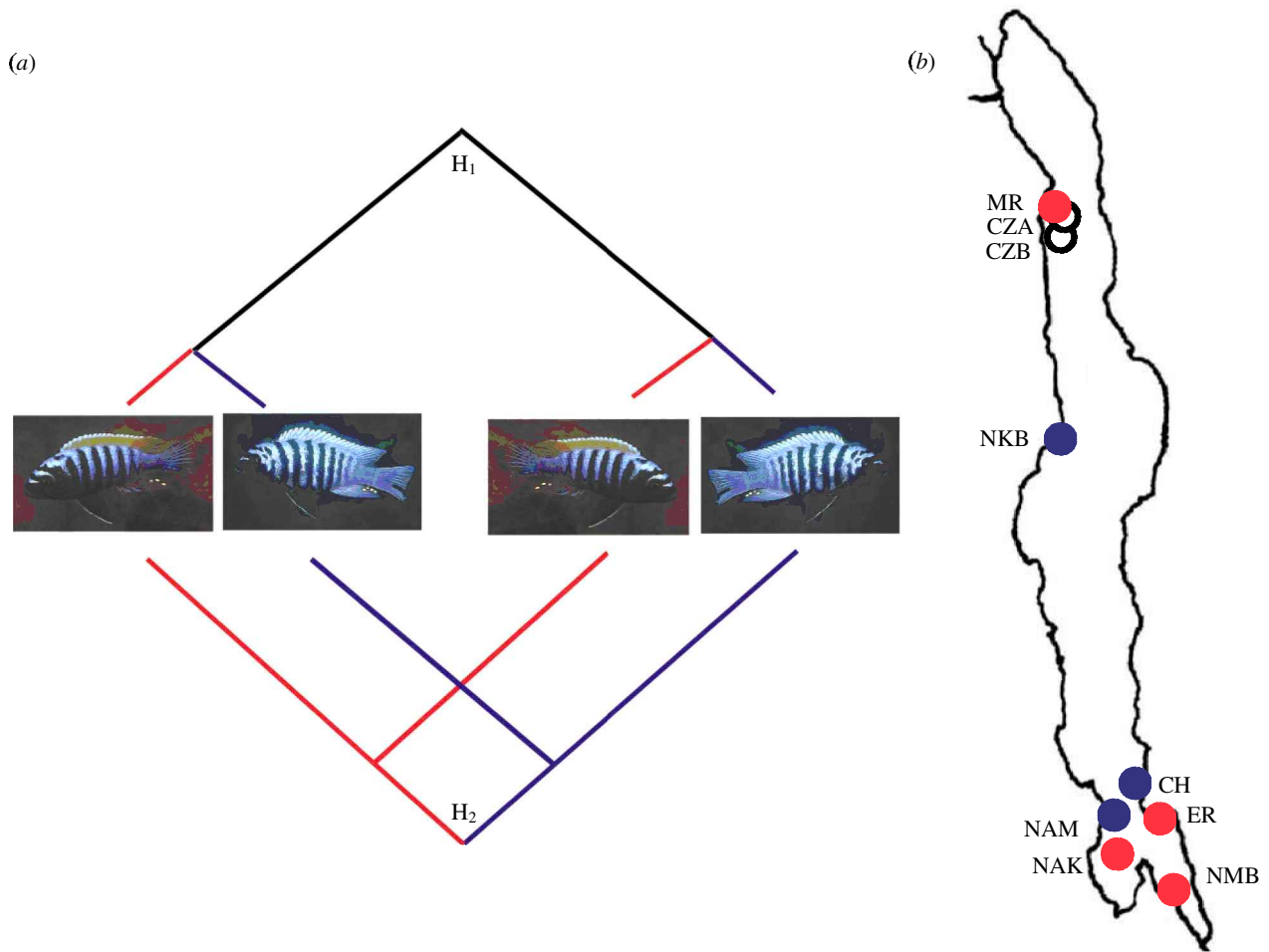


Figure 1. Colour variation in *Pseudotropheus zebra*. (a) Examples of RT and BB phenotypes and graphical representation of the hypotheses tested. H₁: disjunctive RT populations arose from respective BB neighbours by convergent evolution; H₂: RT populations are monophyletic, and southern distributions are explained by colonization of divergent lineages from the north. The BB pictured is from Thumbi West Island, and the RT is from Nakantenga Island (photographs courtesy of Ad Konings). (b) Diagram of Lake Malawi showing seven *P. zebra* and two *P. sp.* ‘zebra Chilumba’ sampling locations. Red circles indicate locations of RT samples. Blue circles indicate locations of BB samples. White circles are the sampling locations of two populations of *P. sp.* ‘zebra Chilumba’. Sample codes are explained in table 1.

Pseudotropheus (*Maylandia*) generic assignment until this taxonomic issue is fully resolved.

Pseudotropheus zebra is a widely distributed rock-dwelling cichlid endemic to Lake Malawi. The rock-dwelling cichlids of Lake Malawi are collectively known as mbuna. Several allopatric colour variants of *P. zebra* are recognized, and are regarded by some as distinct species. The classic *P. zebra* has a blue dorsal fin and six to eight vertical black bars over a blue ground coloration, and we refer to this phenotype as BB. An alternative phenotype exists, wherein the dorsal fin is red or orange instead of blue, and we refer to this phenotype as RT. BB and RT *P. zebra* never co-occur in the same rock area (Konings 2001; personal observations) and populations are not polymorphic with regard to the RT and BB phenotypes (Kornfield & Smith 2000; personal observations). Populations of each phenotype often inhabit adjacent rocky areas, which may be separated by very short distances.

RT populations were sampled from four locations in Lake Malawi: Eccles Reef (*P. thapsinogen*), Nakantenga Island (*P. pyrsonotos*), Mpanga Rocks (*P. emmitos*) and Namaso Bay (*P. sandaracinos*). Long considered conspecific, these populations were elevated to specific status (species names given in parentheses) by Stauffer *et al.* (1997). Regardless of taxonomic

rank, we seek the mechanism whereby these allopatric populations have differentiated and populated their current habitats. Therefore, to assist the readers’ understanding, we loosely refer to all populations examined as variants within the *P. zebra* species complex, and note valid species names where appropriate. Populations of BB *P. zebra* were sampled from three locations: Chiofu Bay, Nkhata Bay and Namalenje Island (figure 1). Samples of 64–72 individuals were collected for DNA analysis at each of these seven locations (table 1).

Two additional *Pseudotropheus* populations from the Chilumba region in northern Lake Malawi, at Mpanga Rocks and Luwino Reef, were also examined. These two populations belong to the undescribed *Pseudotropheus sp.* ‘zebra Chilumba’ (Ribbink *et al.* 1983), which is thought to be allied with *P. zebra*, but is considered to be outside the *P. zebra* complex *sensu strictu* (Konings 2001). This species was examined not to address our proposals but because it is sympatric with the RT population (*P. emmitos*) from Mpanga Rocks, and may play a part in introgression.

(b) Sampling and DNA preparation

Fishes were collected with monofilament gill nets during scuba-diving. Only males were sampled, to ensure proper spec-

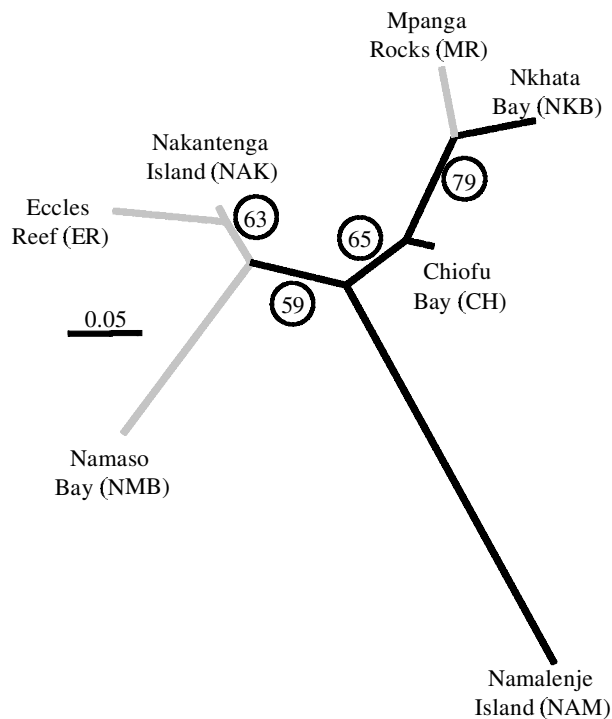


Figure 2. Neighbour-joining majority-rule consensus tree of the seven sampled populations using D_s calculated from nine microsatellite DNA markers. Proportion of 100 bootstrap replicates (re-sampled over loci) supporting each node are shown in the circles. The Mpanga Rocks RT sample is sister to the northern BB sample from Nkhata Bay, indicating an independent origin of the RT phenotype in the north (see § 4). The tree joins RT samples from the southern portion of the lake, suggesting that this recently available habitat has been invaded by pre-existing lineages, rather than by *in situ* speciation. The tree is unrooted; 59 refers to the clade formed by NAK, ER and NMB; 65 refers to the clade formed by MR, NKB and CH.

ies identification. Small portions of the caudal fin were clipped from individual fishes and preserved in 95% ethanol; the ethanol was changed after 48 h. DNA was extracted using the QiAmp tissue kit (Qiagen, Valencia, CA, USA) following the manufacturer's instructions.

Nine microsatellite loci were examined: Pzeb2, Pzeb3, Pzeb4, Pzeb5 (GenBank accession numbers X99783, X99784, X99786 and X99787; Van Oppen *et al.* (1997a)), UME002, UME003 (U14396, U14397; Parker & Kornfield 1996), UNH002, UNH231 (U17045, G12382; K. Kellog, J. Markert, J. R. Stauffer and T. D. Kucher (unpublished data)) and DXTUCA3 (U94850; H. Seultmann, H. Tichy, Y. Satta, N. Takahata and J. Klein (unpublished data)). For each sample, alleles at the nine loci were amplified by PCR in a triplex and a hexaplex using standard conditions, combined and separated on an ABI 377 DNA sequencer (Applied Biosystems International) as per the manufacturer's instructions. PCR fragments were detected using GENESCAN (v. 3.1) software, and alleles were scored using GENOTYPER (v. 2.1) software (Applied Biosystems International).

(c) Data analysis

Microsatellite data were examined for conformity to Hardy-Weinberg expectations and linkage equilibrium, using genetic data analysis (Lewis & Zaykin (2001)). Probability values for Hardy-Weinberg equilibrium were generated to test the fit of

genotype proportions to those generated by 3200 random shufflings, and were Bonferroni-corrected for comparisons across the seven *Pseudotropheus* populations ($\alpha = 0.05$, critical p value = 0.0071). An exact test for linkage disequilibrium was used to examine the association between alleles among pairs of microsatellite loci. Various combinations of these loci have been used in several other studies (Van Oppen *et al.* 1997b; Arnegard *et al.* 1999; Markert *et al.* 1999; Danley *et al.* 2000) but have not been shown to be linked. Allele frequencies, heterozygosities and mean number of alleles were calculated using the microsatellite toolkit macro for Microsoft Excel (Park 2001). Population pairwise F_{ST} values, overall F_{ST} and F_{ST} standard deviations were calculated using ARLEQUIN (Schneider *et al.* 2000).

The divergence between populations was measured with Nei's (1972) standard genetic distance (D_s) and Cavalli-Sforza & Edwards' (1967) chord distance (D_c), using the GENDIST module of the PHYLIP software package (Felsenstein 1993). Van Oppen *et al.* (2000) found evidence for size homoplasy and non-stepwise mutations of Pzeb4 in Lake Malawi cichlids. This locus (which violates the assumptions of the stepwise mutation model (SMM)) had a great effect on SMM-based distances, but a much smaller effect on infinite alleles model (IAM)-based distances. Using simulated data, Takezaki & Nei (1996) concluded that D_c was one of the most efficient distances for obtaining correct tree topology with data generated under both the SMM and the IAM. In an empirical study of genetic distances, Paetkau *et al.* (1997) concluded that D_s was an effective metric for reconstructing relationships between populations that diverged in the last 20 000 years. D_s is based on the IAM, and does not require that microsatellites conform to strict stepwise mutation. D_c , which measures the distance between populations in multidimensional space, is free of assumptions of mutational mechanism. Thus, these distances are appropriate metrics for the current dataset.

Majority-rule neighbour-joining dendrograms were generated using the NEIGHBOR and CONSENS modules of PHYLIP (Felsenstein 1993). Data were re-sampled over loci to generate a consensus tree and bootstrap support.

Introgression into the Mpanga Rocks RT population was examined with factorial correspondence analysis (FCA) performed using the program GENETIX (Belkhir 1999). The Mpanga Rocks RT and the *P. sp.* 'zebra Chilumba' samples were compared with the southernmost RT (Namaso Bay) and BB (Namalenje Island) samples.

3. RESULTS

(a) Microsatellite loci

As noted in previous studies of mbuna populations (Van Oppen *et al.* 1997b; Arnegard *et al.* 1999; Markert *et al.* 1999; Danley *et al.* 2000), the allelic diversity of microsatellites is very high. In the current study, 11–73 alleles were detected per locus, with an average heterozygosity across populations of 0.452 (Pzeb5)–0.967 (UNH231) (data not shown). Summary statistics of genetic diversity among populations are provided in table 1. Hardy-Weinberg equilibrium was frequently observed among *P. zebra* samples. One locus (Pzeb4) deviated from expected genotype proportions in two populations (Nkhata Bay and Namalenje Island), while another (UME002) deviated from expectations in a single population (Chiofu Bay Island). The Namalenje population had two other departures. All other loci conformed to expectations of

Table 2. Pairwise F_{ST} (and standard deviations) for *P. zebra* population comparisons.

| populations ^a | | F_{ST} mean ^b | F_{ST} s.d. |
|--------------------------|----------|----------------------------|---------------|
| CH (BB) | ER (RT) | 0.046 | 0.019 |
| CH (BB) | NAK (RT) | 0.023 | 0.010 |
| CH (BB) | MR (RT) | 0.023 | 0.007 |
| CH (BB) | NMB (RT) | 0.062 | 0.030 |
| NKB (BB) | ER (RT) | 0.059 | 0.031 |
| NAM (BB) | ER (RT) | 0.091 | 0.031 |
| NKB (BB) | NAK (RT) | 0.037 | 0.022 |
| NAM (BB) | NAK (RT) | 0.072 | 0.022 |
| NKB (BB) | MR (RT) | 0.018 | 0.004 |
| NKB (BB) | NMB (RT) | 0.087 | 0.055 |
| NAM (BB) | MR (RT) | 0.081 | 0.021 |
| NAM (BB) | NMB (RT) | 0.096 | 0.039 |
| CH (BB) | NKB (BB) | 0.025 | 0.012 |
| CH (BB) | NAM (BB) | 0.071 | 0.026 |
| NKB (BB) | NAM (BB) | 0.085 | 0.030 |
| ER (RT) | NAK (RT) | 0.015 | 0.005 |
| ER (RT) | MR (RT) | 0.064 | 0.026 |
| ER (RT) | NMB (RT) | 0.058 | 0.031 |
| NAK (RT) | MR (RT) | 0.040 | 0.017 |
| NAK (RT) | NMB (RT) | 0.044 | 0.024 |
| MR (RT) | NMB (RT) | 0.093 | 0.052 |

^a See table 1 for population codes.

^b All F_{ST} values are significant at the 1% level. For all populations together, $F_{ST} = 0.055$ (s.d. = 0.021).

Hardy–Weinberg equilibrium within all of the *P. zebra* populations examined. The *P. sp.* ‘zebra Chilumba’ populations were each characterized by multiple Hardy–Weinberg deviations. The population at Mpanga Rocks had two loci out of equilibrium (UME002 and UME003) and the Luwino Reef population had three (UNH231, Pzeb4 and UME002).

Linkage disequilibrium was detected sporadically, with three significant observations (Eccles Reef, Pzeb2/UME002; Namaso Bay, Pzeb2/Pzeb3 and Pzeb2/UME003) among all *P. zebra* samples other than Mpanga Rocks (RT), wherein linkage disequilibrium was detected among four of the 36 pairwise comparisons (Pzeb3/Pzeb4, Pzeb3/UME002, UNH231/Pzeb4 and Pzeb4/UME003). The genesis of this pattern is not known, but a possible explanation includes introgression from the sympatric *P. sp.* ‘zebra Chilumba’ population (see § 4). Extensive linkage disequilibrium was detected within both the Mpanga Rocks (15/36 pairwise comparisons) and Luwino Reef (21/36 pairwise comparisons) populations of *P. sp.* ‘zebra Chilumba’.

(b) Genetic relationships

Multiple analyses support previous data that populations of *P. zebra* are well differentiated. For all BB and RT samples F_{ST} (excluding *P. sp.* ‘zebra Chilumba’) was 0.055 (s.d. = 0.021), while pairwise estimates of F_{ST} ranged from 0.018 to 0.096 (table 2). A strong and highly significant correlation between genetic distance and geographical distance was detected amongst BB and RT samples (Mantel test, $g = 3.54$, $r = 0.801$, $p < 0.01$). The addition of the *P. sp.* ‘zebra Chilumba’ samples maintains this correlation among all samples in this dataset ($g = 5.14$, $r = 0.862$, $p < 0.01$). This result is highly

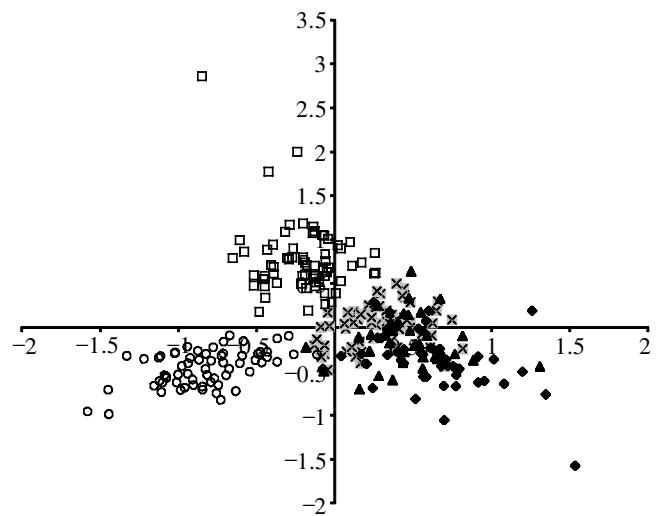


Figure 3. Factorial correspondence analysis shows strong separation of *P. sp.* ‘zebra Chilumba’ individuals from the southern BB (Namalenje Island) and RT (Namaso Bay) individuals. Mpanga Rocks RT individuals occupy much of the space intermediate between *P. sp.* ‘zebra Chilumba’ and the southern samples, and substantially overlap the cluster of *P. sp.* ‘zebra Chilumba’ individuals. The overlap of the Mpanga Rocks individuals with the sympatric *P. sp.* ‘zebra Chilumba’ individuals and their exclusion from southern *P. zebra* clusters implicate introgression in influencing the genetic makeup of the Mpanga Rocks RT population. Key: shaded crosses: Mpanga Rocks RT; open squares: Namaso Bay RT; open circles: Namalenje Island BB; filled triangles: Mpanga Rocks *P. sp.* ‘zebra Chilumba’; filled diamonds: Luwino Reef *P. sp.* ‘zebra Chilumba’.

unexpected because the *P. sp.* ‘zebra Chilumba’ populations are not believed to be part of the *P. zebra* species complex *sensu strictu* (Konings 2001).

A neighbour-joining tree was constructed for RT and BB samples from D_c (figure 2). This tree joins populations that share the same colour phenotypes with the exception of the Mpanga Rocks population. The tree shown is the majority-rule consensus of 100 bootstrap replicates; the tree constructed using D_c has the same topology (not shown), but is less resolved. The proportion of replicates supporting each node is shown on the tree. All nodes have bootstrap values of 50 or greater, indicating a well-supported tree topology.

The three RT samples that were joined on the neighbour-joining tree are all from the southern portion of the lake, indicating a single origin of the red dorsal fin in the south. The northern RT zebra from Mpanga Rocks (*P. emmiltos*) does not group with the southern RT samples, but is sister to the northernmost BB population sampled from Nkhata Bay. Thus, the results of our phylogeny suggest that the RT zebras are not monophyletic, and that the red dorsal fin has arisen more than once among *Pseudotropheus* populations via convergent evolution.

In order to differentiate more critically between ancestral and reticulate relationships, we examined the full dataset by FCA. In a comparison of nine-locus genotypes from northern and southern populations on an FCA plot (figure 3), the Mpanga Rocks (RT) data are seen to be intermediate between the *P. sp.* ‘zebra Chilumba’ points and the southern RT and BB points, but substantially overlap with

the *P. sp.* 'zebra Chilumba' cluster. Thus, the RT sample from the Chilumba region (*P. emmiltos*) apparently shares a closer genetic relationship with the sympatric *P. sp.* 'zebra Chilumba' than with geographically distant RT and BB populations.

4. DISCUSSION

To our knowledge, this is the first demonstration of convergent evolution of reproductive coloration in cichlids at the species level. The placement of the northern RT (*P. emmiltos*) in the phylogeny supports the hypothesis that the red dorsal fin has arisen multiple times within *Pseudotropheus*. However, the three southern RT samples cluster together, suggesting that they share a common ancestor prior to the colonization of their respective habitats; thus, they have not arisen independently within their southern habitats from their BB neighbours. This result indicates the invasion of newly available southern habitats by distinct lineages that were present in the deeper portion of the lake prior to lake-level rise (Bouton 2000).

Convergent evolution of reproductive coloration is an important finding for our understanding of mbuna evolution. Allopatric populations with similar coloration have long been considered conspecific (Ribbink *et al.* 1983), and differences in coloration have been considered sufficient to delimit species (Stauffer *et al.* 1997). The data presented here indicate that both of these assumptions may be untrue, as RT populations appear to be paraphyletic and the populations joined with the greatest support differ in coloration. Colour convergence is prevalent at the generic level. For example, species in other Malawi mbuna genera (e.g. *Cynotilapia afra* and *Labeotropheus fuelleborni*) have a similar body coloration to the RT examined here (Konings 2001), and have evolved independently of each other (Albertson *et al.* 1999). Our data indicate that coloration may be genetically labile, and convergence may be ubiquitous even at the species level.

The role of coloration in species divergence remains unknown, but sexual selection has been proposed as a mechanism that might drive speciation (Dominey 1984; Turner & Burrows 1995; Seehausen & van Alphen 1999; Seehausen *et al.* 1999; Turner 2000). Sibling species that diverged by sexual selection are expected to differ in selected characters of coloration. The species-level colour convergence apparent in our phylogeny indicates that change in coloration may have a role in the process of speciation, and that dorsal-fin colour holds potential interest for study as a sexually selected character. This finding suggests that speciation, marked by divergence in male coloration, may be extremely rapid, and that many cryptic species, bearing similar colour but derived from independent origins, may exist.

The phylogenetic signal observed in this dataset may be substantially influenced by introgression. The RT population at Mpanga Rocks (*P. emmiltos*) is sympatric with *P. sp.* 'zebra Chilumba', a species that is morphologically and behaviourally distinct, but bears coloration similar to the *P. zebra* populations examined here (Ribbink *et al.* 1983). We detected linkage disequilibrium among loci within the Mpanga Rocks RT population and extensively within both *P. sp.* 'zebra Chilumba' populations. Linkage disequilibrium is characteristic of populations experiencing intro-

gression (Hedrick 1999). Examination of the two populations of *P. sp.* 'zebra Chilumba' indeed reveals a remarkable genetic similarity to the sympatric Mpanga Rocks RT population. Additionally, the correlation between geographical and genetic distance among all populations is maintained with the addition of these populations. Isolation by distance among several allopatric populations, many of which are believed to represent good biological species, is highly unexpected and implies a role for gene flow across species boundaries in determining the partitioning of genetic variation among populations of Lake Malawi cichlids. This phenomenon certainly deserves further examination in future studies. Importantly for the current study, these data, together with FCA plots, suggest that the genetic variation in the Mpanga Rocks RT population may have been influenced by introgression from the sympatric *P. sp.* 'zebra Chilumba'.

The observation of introgression in the northern RT population does not invalidate the independent RT origin indicated by our microsatellite phylogeny. However, introgression might explain the sister relationship between the Nkhata Bay BB and the Mpanga Rocks RT if a cline of gene frequencies existed between Chilumba and Nkhata Bay at the time of colonization by the northern RT. Such a cline should be observable in the *Pseudotropheus* populations along the coast between Chilumba and Nkhata Bay as a test of this theory. Genetic complementation tests have been performed with northern and southern RT populations, and the results are consistent with a common genetic origin for the RT phenotype (Reneau 2001), which may be at odds with our phylogeny.

Gene flow between the two well-differentiated species at Chilumba (*P. sp.* 'zebra Chilumba' and *P. emmiltos*) suggests that selection may be operating to maintain distinct colour phenotypes within both taxa. In-depth examination of the colour phenotypes at each locality in the Chilumba region, and field and laboratory examination of mate-choice behaviours among these taxa, would be highly relevant to assessing further the nature of this potential hybrid zone. Additional phylogeographical data should help to resolve these relationships further, and will be necessary before we fully embrace the hypothesis of multiple origins of the RT phenotype within the *P. zebra* species complex.

The putative hybridization evident in this phylogeny together with recently published accounts (Ruber *et al.* 2001; Salsburger *et al.* 2002) suggest that this phenomenon may be much more widespread than previously proposed. If secondary contact often results in gene flow, attempts at species-level phylogenetic reconstruction will be challenged with unravelling the signal of historical relationships from genetic similarities generated by reticulation.

To our knowledge, this study is the first to use microsatellites to reconstruct a species-level phylogeny for cichlids, and shows the effectiveness of the methodology for this group. Although the microsatellite phylogeny produced here is robust, evidence for extensive introgression within the most critical branch of our tree demands that further examination be directed at the problems of phylogeny and colour convergences at the species level in Lake Malawi cichlids.

Malawi fieldwork was facilitated by S. Grant and A. Konings. I.K. thanks Mary Haven for encouragement. Supported by NSF DEB 97-07532.

REFERENCES

- Albertson, C. R., Markert, J. A., Danley, P. D. & Kocher, T. D. 1999 Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proc. Natl Acad. Sci. USA* **96**, 5107–5110.
- Arnegard, M. E., Markert, J. A., Danley, P. D., Stauffer, 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. (DOI 10.1098/rspb.1999.0611.)
- Belkhir, K. 1999 *GENETIX v. 4.0*. Montpellier, France: Laboratoire Genome, Populations, Interactions, CNRS UPR 9060.
- Bouton, N. 2000 Progressive invasion and allopatric speciation can also explain distribution patterns of rock-dwelling cichlids from southern Lake Victoria: a comment on Seehausen and van Alphen (1999). *Ecol. Lett.* **3**, 166–171.
- Cavalli-Sforza, L. L. & Edwards, A. W. F. 1967 Phylogenetic analysis: models and estimation procedures. *Am. J. Hum. Genet.* **19**, 233–257.
- Conde, B. & Gery, J. 1999 *Maylandia* Meyer et Foerster, 1984, un nom generique disponible (Teleostei, Perciformes, Cichlidae). *Rev. Franc. Aqu. Herpetol.* **26**, 21–22.
- Danley, P. D., Markert, J. A., Arnegard, M. E. & Kocher, T. D. 2000 Divergence with gene flow in the rock-dwelling cichlids of Lake Malawi. *Evolution* **54**, 1725–1737.
- Dominey, W. 1984 Effects of sexual selection and life history on speciation: species flocks in African cichlids and Hawaiian *Drosophila*. In *Evolution of fish species flocks* (ed. A. Echelle & I. Kornfield), pp. 231–249. Orono, ME: University of Maine Press.
- Felsenstein, J. 1993 PHYLIP (*phylogeny inference package*) v. 3.5c. See <http://evolution.genetics.washington.edu/phylip>.
- Finney, B. P., Scholz, C. A., Johnson, T. C., Trumbore, S. & Southon, J. 1996 Late quaternary lake-level changes of Lake Malawi. In *The limnology, climatology, and paleoclimatology of the east African lakes* (ed. T. C. Johnson & E. O. Odada), pp. 495–508. Toronto: Gordon and Breach.
- Fryer, G. 2001 On the age and origin of the species flock of haplochromine cichlid fishes of Lake Victoria. *Proc. R. Soc. Lond. B* **268**, 1147–1152. (DOI 10.1098/rspb.2001.1601.)
- Hedrick, P. W. 1999 *Genetics of populations*. Sudbury, MA: Jones and Bartlett.
- Kocher, T., Conroy, J., McKaye, K. & Stauffer, J. 1993 Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Mol. Phylogenet. Evol.* **2**, 158–165.
- Konings, A. 2001 *Malawi cichlids in their natural habitat*. El Paso, TX: Cichlid Press.
- Kornfield, I. & Smith, P. F. 2000 African cichlid fishes: model systems for evolutionary biology. *A. Rev. Ecol. Syst.* **31**, 163–196.
- Lewis, P. O. & Zaykin, D. 2001 Genetic data analysis: computer program for the analysis of allelic data v. 1.0 (d16c). See <http://lewis.eeb.uconn.edu/lewishome/software.html>.
- 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, M. & Foerster, W. 1984 Un nouveau *Pseudotropheus* du lac Malawi avec des remarques sur le complexe *Pseudotropheus-Melanochromis* (Pisces, Perciformes, Cichlidae). *Rev. Franc. Aqu. Herpetol.* **10**, 107–112.
- Moran, P. & Kornfield, I. 1993 Retention of an ancestral polymorphism in the mbuna species flock (Pisces: Cichlidae) of Lake Malawi. *Mol. Biol. Evol.* **10**, 1015–1029.
- Nei, M. 1972 Genetic distances between populations. *Am. Nat.* **106**, 283–292.
- Owen, R. B., Crossley, R., Johnson, T. C., Tweddle, D., Kornfield, I., Davison, S., Eccles, D. H. & Engstrom, D. E. 1990 Major low levels of Lake Malawi and their implications for speciation rates in cichlid fishes. *Proc. R. Soc. Lond. B* **240**, 519–533.
- Paetkau, D., Waits, L. P., Clarkson, P. L., Craighead, L. & Strobeck, C. 1997 An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* **147**, 1943–1957.
- Park, S. D. E. 2001 Trypanotolerance in West African cattle and the population genetic effects of selection. PhD thesis, University of Dublin, Dublin, Ireland.
- Parker, A. & Kornfield, I. 1996 Polygynandry in *Pseudotropheus zebra*, a cichlid fish from Lake Malawi. *Environ. Biol. Fishes* **47**, 345–352.
- Parker, A. & Kornfield, I. 1997 Evolution of the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of Lake Malawi, east Africa. *J. Mol. Evol.* **45**, 70–83.
- Reneau, P. 2001 *Inheritance of dorsal fin coloration in the Metriaclima species complex (Teleostei: Cichlidae) of Lake Malawi*. MSc thesis, Western Kentucky University, KY, USA.
- Ribbink, A., Marsh, B., Marsh, A., Ribbink, A. & Sharp, B. 1983 A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *S. Afr. J. Zool.* **18**, 149–310.
- Rico, C. & Turner, G. 2002 Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Mol. Ecol.* **11**, 1585–1590.
- Ruber, L., Meyer, A., Sturmbauer, C. & Verheyen, E. 2001 Population structure in two sympatric species of the Lake Tanganyika cichlid tribe Eretmodini: evidence for introgression. *Mol. Ecol.* **10**, 1207–1225.
- Salsburger, W., Baric, S. & Sturmbauer, C. 2002 Speciation via introgressive hybridization in African cichlids? *Mol. Ecol.* **11**, 619–625.
- Schneider, S., Roessli, D. & Excofier, L. 2000 ARLEQUIN. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.
- Seehausen, O. 2002 Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history of its cichlid species flock. *Proc. R. Soc. Lond. B* **269**, 491–497. (DOI 10.1098/rspb.2001.1906.)
- Seehausen, O. & van Alphen, J. J. A. 1999 Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol. Lett.* **2**, 262–271.
- Seehausen, O., van Alphen, J. J. A. & Lande, R. 1999 Color polymorphism and sex ratio distortion in a cichlid fish as an insipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**, 367–378.
- Snoeks, J. 2000 How well known is the ichthyodiversity of large east African lakes? In *Ancient lakes: biodiversity, ecology and evolution*, vol. 31 (ed. A. Rossiter & H. Kawanabe), pp. 17–38. Boston, MA: Academic Press.
- Stauffer, J. R. J., Bowers, N. J., Kellogg, K. A. & McKaye, K. R. 1997 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. *Proc. Acad. Nat. Sci. Philadelphia* **148**, 189–230.
- Sturmbauer, C., Baric, S., Salzburger, W., Ruber, L. & Verheyen, E. 2001 Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Mol. Biol. Evol.* **18**, 144–154.
- Takezaki, N. & Nei, M. 1996 Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**, 389–399.

- Turner, G. F. 2000 The nature of species in ancient lakes: perspectives from the fishes of Lake Malawi. In *Ancient lakes: biodiversity, ecology and evolution*, vol. 31 (ed. A. Rossiter & H. Kawanabe), pp. 39–60. Boston, MA: Academic Press.
- Turner, G. F. & Burrows, M. T. 1995 A model of sympatric speciation by sexual selection. *Proc. R. Soc. Lond. B* **260**, 287–292.
- Turner, G. F., Seehausen, O., Knight, M. E., Allender, C. J. & Robinson, R. L. 2001 How many species of cichlid fishes are there in African lakes? *Mol. Ecol.* **10**, 793–806.
- 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**, 387–388.
- Van Oppen, M. J. H., Turner, G. F., Rico, G., 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. (DOI 10.1098/rspb.1997.0248.)
- Van Oppen, M. J. H., Rico, C., Turner, G. F. & Hewitt, G. M. 2000 Extensive homoplasy, nonstepwise mutations and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. *Mol. Biol. Evol.* **17**, 489–498.

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