



How does the taxonomic status of allopatric populations influence species richness within African cichlid fish assemblages?

Martin J. Genner^{1*}, Ole Seehausen², Daniel F.R. Cleary¹, Mairi E. Knight³, Ellinor Michel¹ and George F. Turner²

¹Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands, ²Department of Biological Sciences, University of Hull, Hull, UK and ³Division of Biodiversity and Ecology, University of Southampton, Southampton, UK.

ABSTRACT

Aim Current estimates of species richness within rapidly evolving species flocks are often highly dependent on the species status of allopatric populations that differ in phenotypic traits. These traits may be unreliable indicators of biological species status and systematists may have inconsistently assigned species among lineages or locations on the basis of these traits, thus hampering comparative studies of regional species richness and speciation rates. Our aim was to develop a method of generating standardized estimates of regional species richness suitable for comparative analysis, and to use these estimates to examine the extent and consistency of species assignment of allopatric populations within rapidly evolving cichlid fish flocks present in three east African lakes.

Location Lakes Malawi, Victoria and Tanganyika.

Methods Using published taxon co-occurrence data, a novel approach was employed to calculate standardized 'minimum' estimates of regional species richness for hard substrate associated complexes of cichlids within each of the lakes. Minimum estimates were based on an explicit assumption that if taxa present on equivalent habitats have disjunct distributions, then they are allopatric forms of the same species. These estimates were compared with current observed 'high-end' regional species richness estimates for those complexes to determine the consistency of species assignment of allopatric populations between lineages within a lake. A 'sympatry' index was developed to enable comparisons of levels of species assignment of allopatric populations between-lakes to be made.

Results Within each lake, the minimum and high-end estimates for species richness were significantly correlated across complexes, indicating that the complexes that contain more recognized species contain the most genuine biological species. However, comparisons of complexes among lakes revealed considerable differences. For equivalent geographical areas, substantially higher proportions of recognized species were totally allopatric within the studied Lake Malawi and Lake Victoria complexes, than those of Lake Tanganyika.

Main Conclusions Among African lakes, levels of assignment to species status of allopatric populations were found to be distinctly different. It is unclear whether the discrepancies are a consequence of differences between the lake faunas in degrees of phenotypic divergence among allopatric populations, or are simply the result of inconsistent taxonomic practices. In either case, these results have considerable wider relevance for they emphasize that quantitative measures of regional and beta diversity are critically dependent on the species status of allopatric populations, an issue usually neglected in comparative studies of species richness. The technique introduced here can be used to standardize measures of regional diversity of lineages for comparative analyses, potentially enabling more accurate identification of processes influencing rates of speciation.

*Correspondence: Martin J. Genner, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, PO Box 94766, 1090 GT Amsterdam, The Netherlands.
E-mail: genner@science.uva.nl

Keywords

Allopecies, vicariance, biodiversity, species concepts, mbuna, mbipi.

INTRODUCTION

Estimates of regional (γ) species richness are critically dependent on systematic procedures used to delimit species. Traditionally phenetic or morphological species concepts have been used for this purpose. Characters employed include, for example, 'functional' morphological characters including beak, teeth and jaw structures, or traits more commonly associated with camouflage, communication or reproduction including coloration, sound production or exaggerated secondary sexual anatomy. However, assignments of species status based on such markers are now frequently questioned by many evolutionary biologists studying live animals who prefer 'biological' species concepts, ultimately based on interbreeding (Turner *et al.*, 2001a). This preference is partly because of the fact that geographical variation in such traits is common and morphological species concepts alone cannot account for discrete intraspecific polymorphisms or reliably take into consideration phenotypic differences along geographical clines (Mayr & Diamond, 2001). Thus, without supporting behavioural, ecological or molecular information, that may also possess spatial variance, the use of morphologically defined species can provide inaccurate assessments of regional species richness.

Geographical phenotypic variation is particularly well-recognized among 'island' assemblages where populations are separated by habitat discontinuities. Such barriers may range in form from tracts of deep ocean (Mayr & Diamond, 2001) to lowlands between forested mountain ranges (Masta, 2000), or narrow stretches of sandy lake bottom between rocky outcrops (van Oppen *et al.*, 1997). Examples of such assemblages include *Anolis* lizards of the Caribbean (Jackman *et al.*, 1999), Darwin's finches from the Galapagos (Grant *et al.*, 2000) and *Nesotes* beetles of the Canary Islands (Rees *et al.*, 2001). Schluter (2000) showed that island assemblages tend to possess higher currently recognized species richness than mainland relatives. Thus, given high potential numbers of taxa and geologically measurable time-scales of divergence (e.g. Johnson *et al.*, 1996), such lineages are excellent candidates for comparative phylogenetic analyses aimed at identifying life-history, ecological or genetic traits that are associated with speciation (see Agapow & Isaac, 2002; Orme *et al.*, 2002; Isaac *et al.*, 2003).

However, there are seldom rules concerning the types and degrees of difference in phenotypic or indeed genotypic traits that are used to distinguish species within these assemblages. Thus, systematists often utilize trait information with differing degrees of importance, and this can lead to inconsistent assignment of populations to categories such as species, subspecies, megasubspecies or allopecies, both within and between phylogenetic lineages. Given the problems of species assignment to allopatric populations (see for example Mayr &

Diamond, 2001), it would be useful to have a quantitative procedure enabling estimates to be standardized for comparative work. Here we introduce one such procedure that generates a measure of minimum regional species richness, as a standardized alternative, but not a replacement for observed high-end measures. The approach takes biogeographical presence-absence data of currently identified taxa and disregards information associated with colour or morphological differentiation. It makes one major assumption, that currently recognized taxa within defined genera and subgenera, hereafter referred to as 'complexes', are allopatric populations of the same species, unless they co-occur at least at one location.

This approach was applied to biogeographical data collected on radiations of cichlid fishes within the three largest lakes of East Africa, Malawi, Victoria and Tanganyika. Generic and subgeneric divisions of these cichlids are predominantly based upon contrasting morphological characters, rather than on molecular information. In these lakes, many currently recognized, but rather inconsistently defined 'species' are found in allopatry, and thus regional species richness is difficult to estimate (Turner *et al.*, 2001a). High-end estimates of endemic richness within lakes Malawi and Victoria are placed at over 500 species, while Lake Tanganyika is suggested to be comparatively depauperate with a mere 250 species (Turner *et al.*, 2001a). These estimates have been reached through two processes of species delimitation: first, traditional taxonomy, typically utilizing information on tooth, jaw and head characters, and secondly male breeding coloration. The use of coloration in species delimitation has been adopted within the last 25 years, particularly within lakes Malawi and Victoria, and has been based on the assumption that male breeding colours are important in mate recognition and thus colour differences represent formidable prezygotic barriers to gene flow (see Ribbink *et al.*, 1983; Seehausen, 1996; Konings, 1998, 2001). Experimental (Knight *et al.*, 1998; Seehausen & van Alphen, 1998), field observational (Seehausen *et al.*, 1998a) and molecular (van Oppen *et al.*, 1998) evidence now strongly supports this assumption for sympatric species, and we can be confident that estimates of local (α) species richness are reliable.

However, whether different breeding colours would result in prezygotic isolation in the event of allopatric populations being united is yet to have empirical support and as such, null hypotheses of conspecific status of allopatric taxa cannot be rejected. Even if sexual selection is in operation, colour differences among these populations need not result in barriers to gene flow upon unification. One example is when populations with different breeding colours inhabit environments with different ambient light regimes (see McNaught & Owens, 2002). In turbid water, where the light spectrum is narrow and long-wave shifted, an individual may prefer red mates that are

more reflective under such conditions than blue mates, while the same individual placed in the broad light environment of clearer water may exhibit no preference. Hence, gene flow need not always be prevented should allopatric red/blue colour forms be united.

The issue of species status of allopatric populations within these lakes is significant. A considerable proportion of currently identified species have small distribution ranges and assignments of species status have been inconsistent (Turner *et al.*, 2001a). This is particularly the case with cichlids associated with hard substrates that make up a large proportion of the endemic richness of each lake (e.g. Seehausen, 1996; Konings, 1998, 2001). These taxa exhibit considerable variation in colour between locations and show severely restricted gene flow between disjunct habitats because of strong habitat preferences and a lack of a pelagic dispersal phase (e.g. van Oppen *et al.*, 1997). At present, molecular markers cannot be used to delimit species boundaries among allopatric populations (Kornfield & Smith, 2000). Moreover, differences in functional morphology are unlikely to be able to resolve species status, as substantial differences in such traits among geographically separated populations currently classified as the same species are commonplace (e.g. Stauffer & Hert, 1992; Bouton *et al.*, 2002), and phenotypic plasticity has frequently been demonstrated in cichlid fishes (e.g. Witte *et al.*, 1997).

Here we use published biogeographical presence–absence data from complexes of hard substrate-associated cichlids from all three lakes to calculate minimum species richness estimates in order to compare it with high-end measures. Additionally, we examine the influence of geographical scale on these estimates, and the proportions of allopatric populations that have been classified as species within each lake.

MATERIAL AND METHODS

Source data

(1) *Lake Malawi*: Species presence–absence data were taken from Ribbink *et al.* (1983), who reported data from 79 transects surveyed on the Malawi shoreline of the lake and recognized 196 hard substrate associated species, locally known as ‘mbuna’. These taxa were placed into 12 generally accepted generic and subgeneric groupings (complexes): *Cynotilapia*, *Cyathochromis*, *Genyochromis*, *Gephyrochromis*, *Iodotropheus*, *Labidochromis*, *Labeotropheus*, *Melanochromis*, *Petrotilapia*, *Pseudotropheus (Maylandia)*, *Pseudotropheus (Tropheops)* and *Pseudotropheus (Pseudotropheus)*. Latitude and longitude coordinates were taken from maps provided by Ribbink *et al.* and were converted into Universal Transverse Mercator (UTM) coordinates, enabling Euclidean distances between sites to be calculated. Maximum geographical distance between survey locations was 449.6 km. (2) *Lake Victoria*: Seehausen (1996) surveyed 47 sites in southern Lake Victoria and recognized 57 species within four of the largest genera of hard substrate associated species, locally known as ‘mbipi’; *Mbipia*, *Lithochromis*, *Pundamilia* and *Neochromis*. We used a

previously published presence–absence matrix (Seehausen & van Alphen, 1999), modified to include all species occurring at peripheral sites, and one additional site with three new species (Kunene Islands surveyed in 2001; O. Seehausen, unpublished data). We digitized a map of the locations (Seehausen *et al.*, 1998b) and marked these with points digitally using the package TPS-dig (Rohlf, 2001). The horizontal and vertical distances to points from a central origin were then calibrated to scale, thereby enabling Euclidean distances between each pair of points to be calculated. Maximum geographical distance between locations was estimated to be 127.9 km. (3) *Lake Tanganyika*: Presence–absence data for hard substrate associated species in Lake Tanganyika were obtained from two independent data sets. Alin *et al.* (1999) surveyed three northern sites, while the Lake Tanganyika Biodiversity Project (LTBP) surveyed five stretches of habitat ranging in size between 3 and 80 km on the eastern and southern shorelines (LTBP, 1998a,b, 2000a,b). These two data sets when compiled resulted in a total of 90 species from 19 hard substrate associated complexes. Latitude and longitude of sites were extracted from the text and maps of publications, converted to UTM coordinates, enabling the Euclidean distance between them to be calculated. A central location was estimated for the locations surveyed in the LTBP surveys. Maximum geographical distance between survey locations was estimated to be 620 km. Authorities for all generic names are provided in Table 1.

Estimating minimum species richness

The procedure is founded upon the null hypothesis that conspecific status of two currently identified taxa within recognized complexes can only be rejected if they occur in sympatry. Where taxa within a complex have allopatric distributions totally, they are treated as geographically disjunct forms of the same species. The approach thereby potentially reduces regional and beta diversity, yet maintains alpha species richness. This approach is more powerful than straightforward estimates of maximum alpha diversity in lineages as, for example, it will identify three species in a trio of taxa where a maximum of two occur in sympatry, if different pairs are found together at different sites.

The approach proceeds as follows: species presence–absence matrices are taken based on current classifications and minimum regional species richness is calculated from these data using an iterative procedure. Within each species assemblage (i.e. lake), complexes are identified on the basis of the best available taxonomic information, and within each, putative species are sorted in descending order by the number of sites in which they are present. The species present at the largest number of sites (A) is selected and if the next most frequently occurring taxon (B) is sympatric with A, then it is considered to be a distinct species. Otherwise it is considered to be an allopatric population of A, and the distribution of B is added to that of A. Next, species are resorted in descending order by the number of sites at which they were present and the iterative process continued with the next taxon, if present.

Lake	Complex	Survey sites		Entire lake	
		Observed	Minimum	Predicted	Minimum
Malawi	<i>Cyathochromis</i> Trewavas	1	1	1	1
	<i>Cynotilapia</i> Regan	10	4	13	6
	<i>Genyochromis</i> Trewavas	1	1	1	1
	<i>Gephyrochromis</i> Boulenger	1	1	1	1
	<i>Iodotropheus</i> Oliver & Loiselle	1	1	1	1
	<i>Labeotropheus</i> Ahl	2	2	2	2
	<i>Labidochromis</i> Trewavas	17	5	18	6
	<i>Melanochromis</i> Trewavas	20	9	24	11
	<i>Petrotilapia</i> Trewavas	17	4	19	6
	<i>Pseudotropheus</i> (<i>Maylandia</i>) Regan	27	7	28	9
	<i>Pseudotropheus</i> (<i>Tropheops</i>) Regan	34	7	37	10
	<i>Pseudotropheus</i> (<i>Pseudotropheus</i>) Regan	65	8	80	12
Total		196	50	225	66
Victoria	<i>Neochromis</i> Regan	14	6	22	8
	<i>Mbipia</i> Lippitsch & Seehausen	8	3	13	5
	<i>Pundamilia</i> Seehausen & Lippitsch	23	6	35	10
	<i>Lithochromis</i> Lippitsch & Seehausen	15	5	45	14
	Total		60	20	115
Tanganyika	<i>Altolamprologus</i> Poll	3	3	4	3
	<i>Chalinochromis</i> Poll	1	1	1	1
	<i>Cyphotilapia</i> Regan	1	1	1	1
	<i>Eretmodus</i> Boulenger	1	1	1	1
	<i>Julidochromis</i> Boulenger	3	3	4	3
	<i>Lamprologus</i> Schilthuis	12	7	13	8
	<i>Lepidiolamprologus</i> Pellegrin	9	7	9	8
	<i>Lobochilotes</i> Boulenger	1	1	1	1
	<i>Neolamprologus</i> Colombe & Allgayer	26	17	26	18
	<i>Perissodus</i> Boulenger	2	2	2	2
	<i>Petrochromis</i> Boulenger	7	7	8	7
	<i>Plecodus</i> Boulenger	4	3	5	4
	<i>Pseudosimochromis</i> Nelissen	1	1	1	1
	<i>Simochromis</i> Boulenger	5	4	6	5
	<i>Spathodus</i> Boulenger	2	1	3	2
	<i>Tanganicodus</i> Poll	1	1	1	1
	<i>Telmatochromis</i> Boulenger	6	6	7	6
	<i>Tropheus</i> Boulenger	4	4	5	4
	<i>Variabilichromis</i> Colombe & Allgayer	1	1	1	1
	Total		90	71	99

Table 1 Estimates of regional species richness of hard substrate associated cichlid complexes. Observed high-end values are currently recognized high-end estimates for the study region from source data. The calculated minimum species richness was significantly lower than observed species richness in non-monotypic complexes in all three lakes (one-tailed Wilcoxon matched pairs tests: Malawi, $n = 8$, $P < 0.01$; Victoria, $n = 4$, $P < 0.05$; Tanganyika, $n = 12$, $P < 0.05$). The high-end predicted lake-wide estimates were calculated from extrapolated cumulative species–area relationships. Minimum lake-wide estimates were calculated using procedures outlined in the main text.

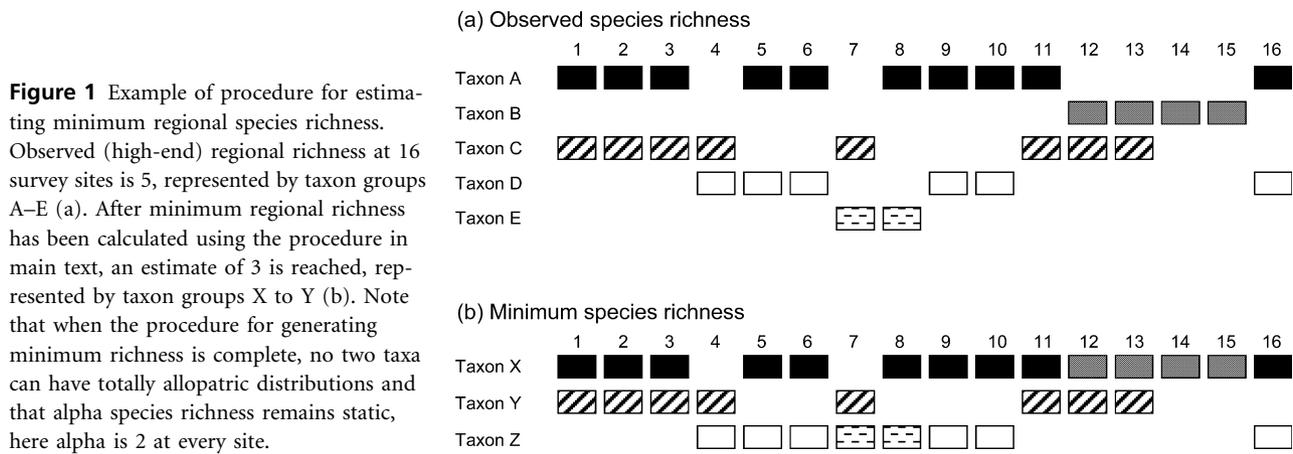
The process is continued until it has been applied to all taxa within the evolutionary lineage, so no two taxa can have totally allopatric distributions. Where two taxa are present at the same number of sites, then the procedure was applied to taxa in alphabetical order. An example of the procedure is given in Fig. 1. The procedure was performed on the source data described above.

Influence of spatial scale on high-end and minimum estimates of diversity

In order to compare high-end and minimum estimates of total richness within each lake over increasing geographical scales, values of mean cumulative diversity over predetermined

geographical distances were calculated for each complex from 10 iterations. Each iteration started from a randomly chosen survey site.

The effect of spatial scale on the high-end and minimum estimates of each complex was further investigated by formulating a ‘Sympatry Index’ that compared the observed high-end and minimum estimates of species richness of each complex within a set geographical area. First, the high-end richness within that radius was noted for each complex, and then the minimum richness for each complex was calculated. Minimum richness was then divided by high-end richness for each complex. Lower values of this index, potentially ranging between 0 and 1, indicate that more species have totally allopatric distributions within the survey area. The index,



herein, was calculated for each complex in each lake using radii of 8, 16, 32, 64, 128, 256 and 512 km, where possible. The mean values from 10 iterations of this procedure for each defined distance enabled good estimates of effects of geographical scale on this index to be made. To compare sympatry indices between lakes over the increasing spatial scales, we used repeated measure ANOVA. Sympatry indices at increasing spatial scales were considered to be repeated measures because data from each spatial scale were not independent from other scales (*sensu* Fuhlendorf *et al.*, 2002).

The sympatry index is both a function of taxonomy (how many allopatric populations have been assigned species status), and of biology (how many species occur in sympatry). The influence of taxonomy on the index increases with spatial scale. In order to compare the biological effects on sympatry indices between lakes we performed an ANCOVA on the sympatry indices obtained for the largest sampled spatial scale in each lake, using lake as factor and observed high-end richness as covariate.

Total lake-wide high-end estimates for species richness of each complex were calculated by extrapolating the relationship between increasing geographical distance and cumulative species richness of each complex to the maximum geographical radius of the lake. The relationships were fitted with Morgan–Mercer–Flodin (MMF) sigmoidal growth models [$y = (a * b + c * x^d)/(b + x^d)$] in CurveExpert 1.3 (<http://curveexpert.webhop.biz/>). The two most distant survey locations within each data set were chosen as starting points. The average of these two high-end estimates for the maximum geographical radii of the lakes was rounded up to the nearest integer. Maximum geographical radii used were 600 km for Lake Malawi, 650 km for Lake Tanganyika and 500 km for Lake Victoria. The minimum lake-wide species richness of each complex was calculated by multiplying the predicted high-end species richness by a predicted sympatry index of that complex for the maximum geographical radius of the lake (see Fig. 3), and rounding up to the nearest integer. The predicted sympatry index for each non-monotypic complex was calculated by fitting an Exponential Association (3) model: $y = a[b - \exp(-cx)]$ to the relationship between geographical

radius and the sympatry index of a complex using CurveExpert 1.3. This equation was then used to predict the sympatry index value at the maximum geographical radius of the lake.

We recognize that reliable phylogenetic evidence would be required to define nested monophyletic complexes to generate statistical comparisons that are free of bias that might be introduced by above species-level taxonomy. However, at present such comprehensive and standardized data are currently not available for African lake cichlids, so in the statistical analyses conducted, we made the assumption that the studied complexes within each lake are equally phylogenetically independent from one another.

RESULTS

Using the Lake Malawi survey data of Ribbink *et al.* (1983) the observed ‘high-end’ regional estimate of 196 species belonging to 12 complexes reduced to a minimum estimate of 50 species (Table 1). From the Lake Victoria survey of Seehausen & van Alphen (1999) and the more recent survey of the Kunene Islands, the estimated regional species richness in the four complexes of 60 reduced to 20 (Table 1). From the surveys of Lake Tanganyika, the regional estimate of 90 species belonging to 19 complexes reduced to 71 species (Table 1). In each case, there were strong positive correlations between observed high-end species richness and minimum species richness where each complex with > 1 species represented a single independent data point (Lake Malawi, $n = 8$, $r^2 = 0.47$; Lake Victoria, $n = 4$, $r^2 = 0.65$; Lake Tanganyika, $n = 12$, $r^2 = 0.96$; across lakes, $n = 24$, $r^2 = 0.30$). Comparisons of beta diversity using species–area curves showed that in all three lakes asymptotic values were attained over much smaller geographical areas in the minimum richness estimates than the high-end species richness estimates. Thus beta diversity was considerably reduced (Fig. 2). The curves for Lake Malawi and Lake Victoria showed greater contrast between observed high-end and minimum richness estimates than those of Lake Tanganyika.

The sympatry index of the non-monotypic complexes was heavily scale-dependent and overall showed a decline with increasing geographical area in all three data sets. Thus, with

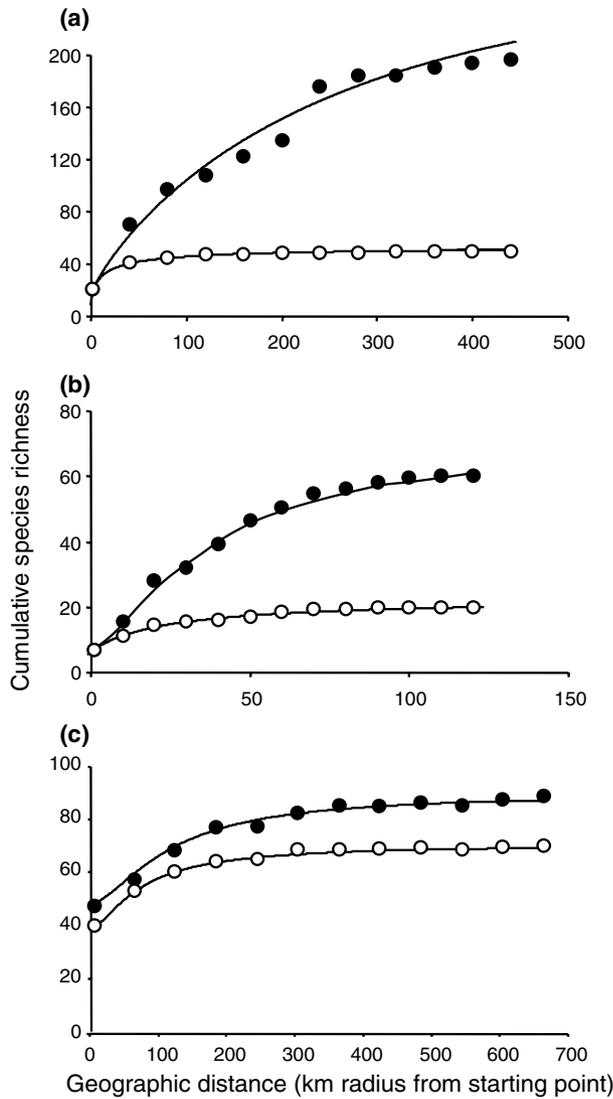


Figure 2 Cumulative species richness of observed high-end estimates (black circles) and minimum estimates (white circles) for (a) Malawi, (b) Victoria and (c) Tanganyika. Each datum represents the mean of 10 iterations. See text for data sources. Lines fitted with Morgan–Mercer–Flodin (MMF) sigmoidal growth models: $y = (a * b + c * x^d) / (b + x^d)$. Lake Malawi: high-end $a = 4.97, b = 98.59, c = 907.88, d = 0.54$; minimum $a = 14.08, b = 6.21, c = 53.15, d = 0.71$. Lake Victoria: high-end $a = 7.04, b = 185.33, c = 71.39, d = 1.45$; minimum $a = 6.99, b = 27.68, c = 23.22, d = 1.01$. Lake Tanganyika: high-end $a = 48.21, b = 2342.83, c = 91.57, d = 1.61$; minimum $a = 40.97, b = 512.15, c = 72.25, d = 1.43$. For all models $r^2 > 0.99$.

increasing area, more species tend to have totally allopatric distributions (Fig. 3a). The curvature of the sympatry index curves varied between genera in Lake Malawi more than between genera in the other lakes (see below), but were generally similar for the Lake Malawi and Victoria data sets. In contrast, the slope for the Tanganyika data was much shallower, reaching an asymptote at a much higher value. Within the radii covered by all three data sets (8, 16, 32, 64 and

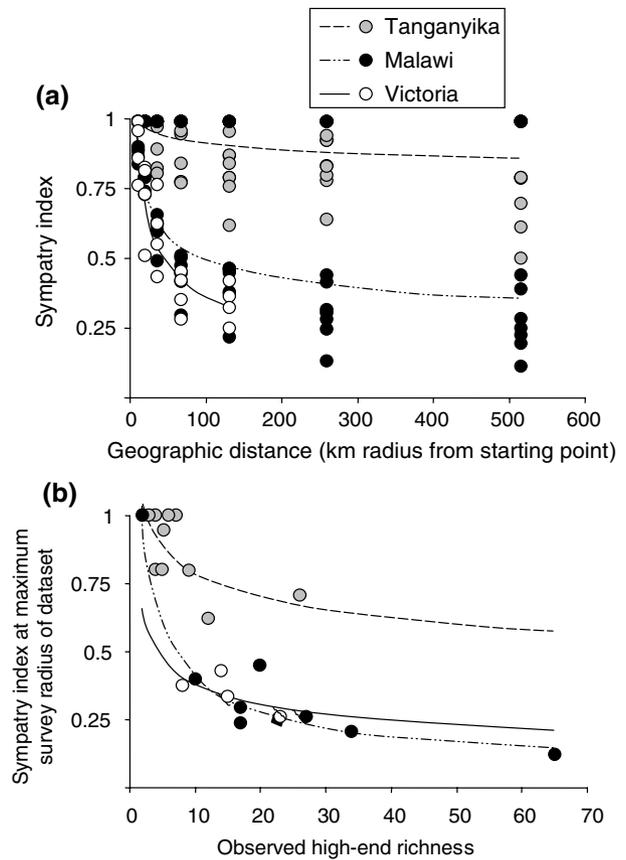


Figure 3 (a) The sympatry index with increasing geographical scale. Each datum represents the mean of 10 iterations for a single non-monotypic complex. Lower values of the index indicate that more species have totally allopatric distributions in original data. Lines fitted with the Harris nonlinear regression model: $y = 1 / (a + bx^c)$. Lake Tanganyika: $a = 0.74, b = 0.19, c = 0.13, r^2 = 0.23$. Lake Malawi: $a = -0.56, b = 1.12, c = 0.18, r^2 = 0.53$. Lake Victoria: $a = 0.53, b = 0.19, c = -0.54, r^2 = 0.81$. Note that the maximum radius for the Lake Victoria data is 128 km. (b) The sympatry index as a function of species complex size. Each datum represents the sympatry index of a single non-monotypic species complex at the largest sampled geographical scale. Lines fitted with power model: $y = ax^b$. Lake Tanganyika: $a = 1.16, b = -0.16, r^2 = 0.51$; Lake Malawi: $a = 1.45, b = -0.53, r^2 = 0.93$; Lake Victoria: $a = 0.71, b = -0.27, r^2 = 0.38$.

128 km), a global test revealed highly significant differences in the sympatry indices of complexes between the lakes (repeated measures ANOVA, Wilk's $\lambda = 0.124, F_{10,34}, P < 0.001$). *Post hoc* analyses showed that while at all geographical radii there were significant differences between the Lake Malawi and Lake Tanganyika complexes (Tukey's HSD tests, all $P < 0.05$), and between the Lake Victoria and Lake Tanganyika complexes (all $P < 0.05$), there were no significant differences between Lake Malawi and Lake Victoria complexes (all $P > 0.10$). These data indicate that overall, within the lineages examined, for the same spatial scales the proportions of species with totally allopatric distributions are similar in the Lake Victoria data (from Seehausen & van Alphen, 1999; O. Seehausen, unpublished

data) and the Lake Malawi data of Ribbink *et al.* (1983). However, for the same spatial scales considerably smaller proportions of species have totally allopatric distributions in the Lake Tanganyika data of Alin *et al.* (1999) and the LTBP. For the studied complexes, minimum whole lake richness estimates are 70.7%, 67.8% and 22.2% lower than high-end whole lake estimates for Lakes Malawi, Victoria and Tanganyika, respectively (Table 1).

Across lakes the sympatry index of species complexes at largest sampled spatial scale (excluding monotypic complexes) was strongly correlated with observed high-end richness ($n = 24$, $r^2 = 0.73$). This relationship explained the larger variance in sympatry index in the Lake Malawi data set (Fig. 3b). However, whereas asymptotic sympatry indices arising from this relationship were very similar in the Lake Malawi and Victoria data sets, it was much higher in the Lake Tanganyika data set. ANCOVA revealed highly significant effects of lake ($F = 16.2$, 2 d.f., $P < 0.001$) and of high-end richness ($F = 15.6$, 1 d.f., $P < 0.01$). *Post hoc* analyses showed significant differences between the Lake Malawi and Lake Tanganyika data sets (Tukey's HSD test, $P < 0.01$) and between the Lake Victoria and Lake Tanganyika data sets ($P < 0.01$), whereas there was no significant difference between the Lake Malawi and Lake Victoria data sets ($P > 0.90$).

DISCUSSION

Within this study, we examined levels of species richness in hard substrate associated cichlid fishes of three African lakes. Utilizing a biological species concept and the assumption that sympatric co-occurring taxa within the studied complexes represent reproductively isolated populations, we were able to generate a minimum species richness estimate of each species complex. We then compared these with high-end measures that frequently strike us as being derived somewhat inconsistently. Our results showed that within each lake there was substantial variation among complexes in proportions of allopatric taxa that have been identified as species for the same spatial scales. This between-genus variation is more extreme in the data set from Lake Malawi. For example, while Ribbink *et al.* identified 34 species of *Pseudotropheus* (*Tropheus*), minimum diversity was estimated to be 7, so over 79% of taxa were allopatric. In contrast, only two species of *Labeotropheus* were recognized by Ribbink *et al.* and these were found in sympatry in most locations. Thus none was allopatric. This resulted in a strong negative correlation between high-end species richness and sympatry index within and across lakes: the more species are recognized in a complex, the smaller the proportion of species that occur sympatrically. Nevertheless given reasonably strong positive correlations between observed high-end and minimum diversity within all three lakes, it would appear that the general natural history information used by systematists to partition allopatric populations into species has been applied reasonably constantly between complexes within a lake. Thus differences in diversity of complexes are probably not primarily a consequence of inconsistent system-

atics. Rather these data suggest that complexes with higher phenotypic diversity among allopatric populations, which are thus regarded as allopatric species, also tend to be the complexes that contain more genuine biological species.

Partitioning of allopatric populations into species was more consistent among complexes within single lakes than across lakes. We recognize that this may in part be the result of 'area shape' or habitat size effects. Between lakes, the radii may encompass differing amounts of shoreline, islands, land area and suitable habitat for the focal taxa. In addition, the sampling sites in the Lake Tanganyika data set tended to be larger than those in the other two data sets, possibly generating more opportunity for 'sympatry'. However, this alone cannot explain the large difference in sympatry index between Lake Tanganyika and the other lakes, and such effects of differing topography will be present in any comparative study of biodiversity using independent survey data from different sites where controls for the effect have not explicitly been implemented. Nevertheless, our analyses do suggest that within the complexes examined, many more allopatric populations have been recognized as distinct species in Lake Malawi and Victoria than in Lake Tanganyika. One possible explanation is that different classification systems have been used. Few taxonomists have worked on taxa from all three lakes. Another possibility is that in each lake similar levels of phenotypic diversity have been considered synonymous with similar levels of species diversity by systematists, and that cichlids from the studied complexes within lakes Malawi and Victoria possess higher within-complex phenotypic diversity than those of Lake Tanganyika. In the latter case, the higher within-complex diversity in Lake Victoria and Lake Malawi would not have led to any associated increase in alpha diversity. Comparative analysis of phenotypic variation within each of the complexes may explain the contrasting patterns observed.

The process described here for identifying the proportions of allopatric populations that have been delimited as species is heavily dependent upon the definition of a species complex. We took the approach of identifying complexes using the current systematic classifications that are based exclusively upon anatomical characters. Given the propensity for anatomical divergence to arise rapidly within African lake cichlids, and the many known cases of convergent evolution between lakes, complex delimitation would benefit from a more rigorous approach derived from additional phylogenetic evidence. Specifically, the sympatry index will be sensitive to three potential sources of error: polyphyly, nesting and variation in inclusiveness of species complexes. It will overestimate minimum species richness for complexes that are polyphyletic and underestimate it for complexes that are incomplete because members are wrongly assigned to another complex. If complexes are not reciprocally monophyletic but are nested in each other, the method will underestimate minimum richness in the paraphyletic complex, and overestimated it in the complex nested within it. Finally, variation associated with the inclusiveness of complexes overestimates minimum richness in less inclusive, and underestimates it in

more inclusive complexes. Partitioning the variation in sympatry index into effects of complex size (observed high-end richness at maximum geographical radius) and effects of lake thus provided us with a conservative way of controlling for effects of variation in inclusiveness on between-lakes comparisons.

One possible technique to generate the robust phylogenies that are necessary to account for these effects would be to employ molecular markers. However, given the considerable discrepancies in the levels of molecular differentiation between taxa both within and among lakes (McCune & Lovejoy, 1998; Kornfield & Smith, 2000), applying a standard procedure based on levels of divergence of molecular markers may be difficult. Moreover, patterns of molecular differentiation between taxa within lakes based on single loci (such as mitochondrial DNA) are also likely to be confounded by the retention of common ancestral polymorphisms within populations that are now reproductively isolated (Kornfield & Smith, 2000). Evidence also suggests introgressive hybridization may be common among hard substrate associated taxa within all three lakes (Ruber *et al.*, 2001; Salzburger *et al.*, 2002; Smith & Kornfield, 2002; Seehausen *et al.*, 2003), but at present it is unclear how extensive this may be. It is possible that problems with delimiting taxa using molecular information may be overcome to some extent by the analysis of multiple polymorphic nuclear loci in place of single locus markers. For example, amplified fragment length polymorphisms have been shown to be useful in reconstructing phylogenetic relationships of closely related species (Albertson *et al.*, 1999; Seehausen *et al.*, 2003), but to date levels of within and between population variability have to be assessed. Taken together, it would appear that while genetic markers may be useful for practical delimitation of cichlid taxa, more extensive work needs to be conducted to determine their general viability.

It would be desirable to generate both high-end and minimum estimates of the entire species richness of each lake for comparative purposes. However, it would be misleading to extrapolate an average sympatry index developed from the complexes studied here to other complexes within each lake that may have considerably different levels of geographical variability in phenotypic traits, and thus different pathways of species delimitation may have been applied. Such groups of taxa may include, for example, the pelagic cichlid species of Lake Malawi that show little evidence of intraspecific population subdivision or morphological variability over lake-wide spatial scales (Shaw *et al.*, 2000). However, it is possible to generate minimum and high-end species richness estimates of the whole lake by adding the richness estimates of the included complexes in this study to the current estimates of total species richness of the non-included complexes. A reasonable high-end richness estimate of Lake Malawi cichlids belonging to complexes not included in this study is 375 species (data from Turner, 1996; Konings, 2001; Turner *et al.*, 2001b). These, together with an estimated 66–225 species from included complexes (Table 1) would result in Lake Malawi housing between 451 and 600 species, depending upon the status of

allopatric populations of species in included complexes. A total estimate of non-rock-associated species from Lake Victoria was 271 in 1991 (Witte *et al.*, 1992). Around 30 non-rock associated, and 120 rock-associated species that do not belong to the complexes included in this study, were further listed by Kaufman & Ochumba (1993); Seehausen (1996) and Seehausen *et al.* (1997), thus resulting in *c.* 420 species. Together with the estimated 37–115 rock-associated taxa from the four studied complexes (Table 1), a total estimate of between 447 and 535 species for Lake Victoria is reached. A reasonable high-end species richness estimate of Lake Tanganyika cichlid complexes not included in this study is 85 (Konings, 1998). These, together with an estimated 77–99 species belonging to included complexes (Table 1) results in Lake Tanganyika housing between 162 and 184 cichlid species. Thus even with lower estimates based on the hypothesis of conspecific status of allopatric taxa currently recognized as species, these data still suggest that higher levels of species richness are, or were, present in lakes Malawi and Victoria than in Lake Tanganyika. However, it should be mentioned that in all three of the lakes, populations within complexes that have not been included in this analysis also show allopatric variation in phenotypic traits upon which species status has been assigned. Thus assemblage-wide minimum species richness estimates may possibly be lower than estimated here. In contrast however, more accurate delimitation of complexes, further intensive survey of understudied sites and more extensive taxonomic investigation may result in higher estimates. For example, such studies may reveal cases where minimum estimates are too low because competitive exclusion has generated the lack of overlap between distributions of some species belonging to the same complexes.

Our analyses suggest that current measures of regional and beta diversity are critically dependent on the species status of allopatric populations. As such between-lineage and between-habitat comparisons of species richness may be highly vulnerable to these effects. This simple fact has been widely neglected in classic biogeography and biodiversity texts, but has long been recognized. For example, Darwin (1872) addresses this issue in his discussion of 'doubtful species' when he remarks, 'A wide distance between the homes of two doubtful forms leads many naturalists to rank both as distinct species; but what distance, it has been well asked, will suffice?' The procedure described herein enables standardized alternatives to observed high-end estimates to be calculated for use in comparative studies. However these standardized minimum estimates, like any measure of biological diversity, will be most useful where sampling effort for comparative work has been equivalent, and skills of surveyors have been consistent. Moreover, minimum estimates will be most accurately made using data sets in which sites are equivalent in size and habitats, as reproductively isolated taxa frequently have parapatric distributions simply because they possess different environmental preferences.

In conclusion, in this study we found that quantitative measures of regional and beta diversity within complexes of cichlid fishes from three African lakes were critically dependent on the species status of allopatric populations. At present we

are unable to determine whether this is a consequence of inconsistent taxonomy or differences in levels of phenotypic divergence between the studied complexes. Nevertheless, on the basis of this study we propose that between-habitat and between-lineage comparisons of species richness should account for potentially confounding differences in the species status of allopatric populations.

ACKNOWLEDGEMENTS

This paper arose from conversations with many colleagues on the problems associated with estimating cichlid species richness within African lakes. The authors thank Martin Taylor, Jonathan Todd and Fred Schram. MJG is supported by European Commission Marie Curie Fellowship HPMF-CT-2000-798.

REFERENCES

- Agapow, P.M. & Isaac, N.J.B. (2002) MacroCAIC: Revealing correlates of species richness by comparative analysis. *Diversity and Distributions*, **8**, 41–43.
- Albertson, R.C., Markert, J.A., Danley, P.D. & Kocher, T.D. (1999) Phylogeny of a rapidly evolving complex: the cichlid fishes of Lake Malawi, East Africa. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 5107–5110.
- Alin, S.R., Cohen, A.S., Bills, R., Gashagaza, M.M., Michel, E., Tiercelin, J.J., Martens, K., Coveliers, P., Mboko, S.K., West, K., Soreghan, M., Kimbadi, S. & Ntakimazi, G. (1999) Effects of landscape disturbance on animal communities in Lake Tanganyika, East Africa. *Conservation Biology*, **13**, 1017–1033.
- Bouton, N., de Visser, J. & Barel, C.D.N. (2002) Correlating head shape with ecological variables in rock-dwelling haplochromines (Teleostei: Cichlidae) from Lake Victoria. *Biological Journal of the Linnean Society*, **76**, 39–48.
- Darwin, C. (1872) *The origin of species*, 6th edn. John Murray, London, UK.
- Fuhlendorf, S.D., Woodward, A.J.W., Leslie, D.M. Jr & Shackford, J.S. (2002) Multi-scale effects of habitat loss and fragmentation on lesser prairie-chicken populations of the US Great Plains. *Landscape Ecology*, **17**, 617–628.
- Grant, P.R., Grant, B.R. & Petren, K. (2000) The allopatric phase of speciation: the sharp-beaked ground finch (*Geospiza difficilis*) on the Galapagos islands. *Biological Journal of the Linnean Society*, **69**, 287–317.
- Isaac, N.J.B., Agapow, P.M., Harvey, P.H. & Purvis, A. (2003) Phylogenetically nested comparisons for testing correlates of species richness: a simulation study of continuous variables. *Evolution*, **57**, 18–26.
- Jackman, T.R., Larson, A., de Queiroz, K. & Losos, J.B. (1999) Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. *Systematic Biology*, **48**, 254–285.
- Johnson, T.C., Scholz, C.A., Talbot, M.R., Kelts, K., Ricketts, R.D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J.W. (1996) Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science*, **273**, 1091–1093.
- Kaufman, L. & Ochumba, P. (1993) Evolutionary and conservation biology of cichlid fishes as revealed by faunal remnants in northern Lake Victoria. *Conservation Biology*, **7**, 719–730.
- Knight, M.E., Turner, G.F., Rico, C., van Oppen, M.J.H. & Hewitt, G.M. (1998) Microsatellite paternity analysis on captive Lake Malawi cichlids supports reproductive isolation by direct mate choice. *Molecular Ecology*, **7**, 1605–1610.
- Konings, A. (1998) *Tanganyika cichlids in their natural habitat*. Cichlid Press, El Paso, TX.
- Konings, A. (2001) *Malawi cichlids in their natural habitat*, 3rd edn. Cichlid Press, El Paso, TX.
- Kornfield, I. & Smith, P.F. (2000) African cichlid fishes: model systems for evolutionary biology. *Annual Reviews in Ecology and Systematics*, **31**, 163–196.
- LTBP (1998a) *Biodiversity Special Study (BIOSS) Aquatic habitats and associated biodiversity of the Kigoma Area, Tanzania*. UNDP/GEF. (RAF/92/G32). <http://www.ltbp.org/>.
- LTBP (1998b) *Biodiversity Special Study (BIOSS) Survey of aquatic habitats and associated biodiversity adjacent to Gombe Stream National Park, Tanzania*. UNDP/GEF. (RAF/92/G32). <http://www.ltbp.org/>.
- LTBP (2000a) *Biodiversity Special Study (BIOSS) Mahale National Park, Tanzania. March/April 1999 Aquatic Survey*. UNDP/GEF. (RAF/92/G32). <http://www.ltbp.org/>.
- LTBP (2000b) *Biodiversity Special Study (BIOSS) Nsumbu National Park, Zambia. July/August 1999 Aquatic Survey*. UNDP/GEF. (RAF/92/G32). <http://www.ltbp.org/>.
- McCune, A.R. & Lovejoy, N.R. (1998) The relative rate of sympatric and allopatric speciation in fishes. *Endless forms: species and speciation* (ed. by D.J. Howard and S.H. Berlocher), pp. 172–185. Oxford University Press, New York, NY.
- McNaught, M.K. & Owens, I.P.F. (2002) Interspecific variation in plumage colour among birds: species recognition or light environment? *Journal of Evolutionary Biology*, **15**, 505–514.
- Masta, S.E. (2000) Phylogeography of the jumping spider *Habronattus pugillis* (Araneae: Salticidae): recent vicariance of sky island populations? *Evolution*, **54**, 1699–1711.
- Mayr, E. & Diamond, J. (2001) *The birds of Northern Melanesia: species, ecology, and biogeography*. Oxford University Press, New York, NY.
- van Oppen, M.J.H., Turner, G.F., Rico, C., Deutsch, J.C., Ibrahim, K.M., Robinson, R.L. & Hewitt, G.M. (1997) Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society of London B*, **264**, 1803–1812.
- van Oppen, M.J.H., Turner, G.F., Rico, C., Robinson, R.L., Deutsch, J.C., Genner, M.J. & Hewitt, G.M. (1998) Assortative mating among rock-dwelling cichlid fishes supports high estimates of species richness from Lake Malawi. *Molecular Ecology*, **7**, 991–1001.
- Orme, C.D.L., Isaac, N.J.B. & Purvis, A. (2002) Are most species small? Not within species-level phylogenies. *Proceedings of the Royal Society of London B*, **269**, 1279–1287.

- Rees D.J., Emerson B.C., Oromi P. & Hewitt G.M. (2001) Mitochondrial DNA, ecology and morphology: interpreting the phylogeography of the *Nesotes* (Coleoptera : Tenebrionidae) of Gran Canaria (Canary Islands). *Molecular Ecology*, **10**, 427–434.
- Ribbink, A.J., Marsh, B.J., Marsh, A.C., Ribbink, A.C. & Sharp, B.J. (1983) A preliminary survey of the cichlid fishes of the rocky habitats of Lake Malawi. *South African Journal of Zoology*, **18**, 155–309.
- Rohlf, F.J. (2001) *TPS dig*. Department of Ecology and Evolution, State University of New York, Stony Brook, NY. Available from <http://life.bio.sunysb.edu/morph/>.
- 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. *Molecular Ecology*, **10**, 1207–1225.
- Salzburger, W., Baric, S. & Sturmbauer, C. (2002) Speciation via introgressive hybridization in east African cichlids? *Molecular Ecology*, **11**, 619–625.
- Schluter, D. (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford, UK.
- Seehausen, O. (1996) *Lake Victoria rock cichlids: taxonomy, ecology and distribution*. Verduijn Cichlids, Zevenhuizen, The Netherlands.
- Seehausen, O. & van Alphen, J.J.M. (1998) The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behavioural Ecology and Sociobiology*, **42**, 1–8.
- Seehausen, O. & van Alphen, J.M. (1999) Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecology Letters*, **2**, 262–271.
- Seehausen, O., Witte, F., Katunzi, E.F.B., Smits, J. & Bouton, N. (1997) Patterns of the remnant cichlid fauna in southern Lake Victoria. *Conservation Biology*, **11**, 890–905.
- Seehausen, O., Witte, F., van Alphen, J.J.M. & Bouton, N. (1998a) Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *Journal of Fish Biology*, **53A**, 37–55.
- Seehausen O., Lippitsch, E., Bouton, N & Zwennes, H. (1998b) Mpibi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyological Exploration of Freshwaters*, **9**, 129–228.
- Seehausen, O., Koetsier, E., Schneider, M.V., Chapman, L.J., Chapman, C.A., Knight, M.E., Turner, G.F., van Alphen, J.J.M. & Bills, R. (2003) Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. *Proceedings of the Royal Society of London B*, **270**, 129–137.
- Shaw, P.W., Turner, G.F., Idid, M.R., Robinson, R.L. & Carvalho, G.R. (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proceedings of the Royal Society of London B*, **267**, 2273–2280.
- Smith, P.F. & Kornfield, I. (2002) Phylogeography of Lake Malawi cichlids of the genus *Pseudotropheus*: significance of allopatric colour variation. *Proceedings of the Royal Society of London B*, **269**, 2495–2502.
- Stauffer J.R. & Hert, E. (1992) *Pseudotropheus callainos*, a new species of mbuna (Cichlidae), with analyses of changes associated with two intra-lacustrine transplantations in Lake Malawi, Africa. *Ichthyological Explorations of Freshwaters*, **3**, 253–264.
- Turner, G.F. (1996) *Offshore cichlids of Lake Malawi*. Cichlid Press, El Paso, TX.
- Turner, G.F., Seehausen, O., Knight, M.E., Allender, C.J. & Robinson, R.L. (2001a) How many species of cichlid fishes are there in African lakes? *Molecular Ecology*, **10**, 793–806.
- Turner, G.F., Robinson, R.L., Ngatunga, B.P., Shaw, P.W. & Carvalho, G.R. (2001b) Pelagic cichlid fishes of Lake Malawi/Nyasa. *Journal of Aquariology and Aquatic Sciences*, **9**, 287–302.
- Witte, F., Goldschmidt, T., Wanink, J., van Oijen, M., Goudswaard, K., Witte-Maas, E. & Bouton, N. (1992) The destruction of an endemic species flock: quantitative data on the decline of the haplochromine cichlids of Lake Victoria. *Environmental Biology of Fishes*, **34**, 1–28.
- Witte, F., Barel, K.D.N. & van Oijen, M.J.P. (1997) Intraspecific variation of haplochromine cichlids from Lake Victoria and its taxonomic implications. *South African Journal of Science*, **93**, 585–594.

BIOSKETCHES

Martin Genner is Marie Curie Postdoctoral Fellow at the University of Amsterdam with interests in the biogeographic origins of the fish and molluscs of east African lakes.

Ole Seehausen is lecturer at the University of Hull. His main interests are in the processes driving speciation and adaptive radiation.

Daniel Cleary recently completed a PhD at the University of Amsterdam. His interests focus on the mechanisms generating spatial patterns of species richness in high-diversity assemblages.

Mairi Knight is a molecular ecologist working at the University of Southampton. Her research involves using molecular tools and experimental approaches to identify behavioural barriers to gene flow.

Ellinor Michel is a lecturer at the University of Amsterdam. Her research focuses upon the speciation and biogeography of east African gastropods.

George Turner is a professor at the University of Hull. His focal interest is the behavioural and genetic basis of speciation in Lake Malawi cichlids.