REPORT

The coincidence of rarity and richness and the potential signature of history in centres of endemism

Walter Jetz,[1,2]* Carsten Rahbek[3] and Robert K. Colwell[4]

[1]Ecology and Evolutionary Biology Department, Princeton University, Princeton, NJ, USA
[2]Biology Department, University of New Mexico, Albuquerque, NM, USA
[3]Zoological Museum, University of Copenhagen, Copenhagen, Denmark
[4]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA
*Correspondence and present address as of December 2004: Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA.
E-mail: wjetz@princeton.edu

Abstract

We investigate the relative importance of stochastic and environmental/topographic effects on the occurrence of avian centres of endemism, evaluating their potential historical importance for broad-scale patterns in species richness across Sub-Saharan Africa. Because species-rich areas are more likely to be centres of endemism by chance alone, we test two null models: Model 1 calculates expected patterns of endemism using a random draw from the occurrence records of the continental assemblage, whereas Model 2 additionally implements the potential role of geometric constraints. Since Model 1 yields better quantitative predictions we use it to identify centres of endemism controlled for richness. Altitudinal range and low seasonality emerge as core environmental predictors for these areas, which contain unusually high species richness compared to other parts of sub-Saharan Africa, even when controlled for environmental differences. This result supports the idea that centres of endemism may represent areas of special evolutionary history, probably as centres of diversification.

INTRODUCTION

A growing number of regional and continental analyses suggest that a large proportion of the geographic variation in species richness can be explained by contemporary factors, such as productivity and habitat heterogeneity (Brown 1995; Rosenzweig 1995; Currie *et al.* 1999; Rahbek & Graves 2001; Jetz & Rahbek 2002; Francis & Currie 2003). However, there is a history behind species richness patterns, and despite the prominent role of present-day factors there is no doubt that history is reflected, in one form or another, in the distribution of contemporary assemblages (Ricklefs & Schluter 1993; Cracraft 1994; Patterson 1999). Advocates of the role of regional history, biogeographic barriers and the large-scale processes of allopatric speciation and extinction urge caution over the neglect of historical mechanisms such as past isolation dynamics, which tend to be ignored in analyses that focus on contemporary patterns of species richness (Latham & Ricklefs 1993; Ricklefs *et al.* 1999; Ricklefs 2004). Meanwhile, other authors hold that the consistently strong explanatory power of contemporary climate suggests only a minor role for historical processes for which direct evidence is notoriously

difficult to establish (e.g. Francis & Currie 2003). It follows that geographic analyses of species richness often exemplify a divide between ecological (MacArthur 1972; Endler 1982a; Brown 1995) and historical (Rosen 1978; Nelson & Platnick 1981; Haffer 1982) approaches to the analyses of species distributions that has marked the past forty years of research in the interface of biogeography, ecology and evolution.

While palaeoecological evidence allows increasingly more accurate prediction of past vegetation patterns, exact spatio-temporal habitat dynamics and their specific effect on animal distributions and gene flow remain obscure. However, coarse indices of climatic stability can be attained and yield interesting first insights about the direct effect of past climate on species distributions (Dynesius & Jansson 2000). Another promising analytical angle is given by the ever more accurate and comprehensive phylogenies that allow phylogeographic analyses at an increasingly large spatial and phylogenetic scale (Fjeldsa & Lovett 1997; Schneider *et al.* 1999; Moritz *et al.* 2000).

Across taxa, regions and scales, contemporary environment models are repeatedly found to explain a very large proportion of overall species richness, usually between 60 and 85% (Schall & Pianka 1978; Currie 1991; Rahbek &

Graves 2001; Jetz & Rahbek 2002; Francis & Currie 2003). However, these statistically strong relationships mask two important phenomena. First, outliers with much higher richness than predicted by contemporary environment models occur (Rahbek & Graves 2001; Jetz & Rahbek 2002) and often tend to be spatially clustered in areas that have been pinpointed in the past to contain phylogenetically or biogeographically unique species (Fjeldsa & Lovett 1997; Stattersfield *et al.* 1998). Second, species with very small geographic ranges tend to show a pattern in richness that is very different to that of all species together, they are affected by different variables, and they are less predictable by contemporary environment (Jetz & Rahbek 2002). Both findings highlight locations where contemporary environmental models fail, despite their strong explanatory power for overall species richness, and thus where historical processes may prevail. We suggest that the occurrence of narrow-ranged species in so-called 'centres of endemism' and underprediction of richness by environmental models may be interrelated, each pointing to the geographic occurrence of historical processes that affect both endemicity and richness.

In the debate about historical interpretations of species distributions, 'centres of endemism' have repeatedly been regarded as exemplifying the role of history in contemporary patterns of species distributions (Rosen 1978; Nelson & Platnick 1981; Haffer 1982; Prance 1982). One suggestion is that these regions, if marked by primary endemics, are centres of clade origin and speciation and should still testify to this special historical role by a very high overlap of contemporary geographical ranges (Croizat *et al.* 1974; Terborgh 1992; Ricklefs & Schluter 1993). That is, overall species richness in such places should be higher than elsewhere, regardless of the particular endemic species within them.

This view, and support for any specific historical mechanism could potentially be challenged if the geographic distribution of centres of endemism largely follows contemporary factors (Endler 1982b; see also Francis & Currie 2003). Yet, even the latter interpretation may be challenged if one could show that chance alone was enough to explain the observed pattern. The apparent local excess of narrow-ranged species that define centres of endemism might simply represent the expected number of such species, given locally higher richness, and such a simple 'sampling effect' would call any further historical or ecological inferences into question (Connor & Simberloff 1979; Gotelli & Graves 1996; Maurer 1999). It follows that any special (e.g. evolutionary historical) role of centres of endemism can be evaluated only if the effect of species richness, *per se*, is properly accounted for. This requirement has so far left unresolved the issue of whether centres of endemism do indeed contain an unexpectedly greater number of species than other regions. Historical biogeography has so far focused on the importance of areas of endemism in quantifying vicariance, but has only begun to specifically address the confounding issue of random effects on area selection at a large scale (Mast & Nyffeler 2003).

Beyond their role as indicators for testing biogeographic hypotheses, centres of endemism represent regions of high conservation concern (Stattersfield *et al.* 1998). With an ever-increasing rate of extinction and lack of distributional information, knowledge about the potential large-scale predictability of centres of endemism from environmental factors reaches beyond traditional hypothesis testing. Whether centres of endemism – besides containing species that are threatened according to currently accepted criteria – have a special role as areas of high past and possibly future evolutionary potential is a matter of particular importance for large-scale conservation priority setting (Fjeldsa *et al.* 1999; Crandall *et al.* 2000).

Here we set out to address these issues by creating a methodological bridge between those studies of species richness that focus exclusively on contemporary environmental correlates and those studies that attempt to infer historical process based on variance left conditionally unexplained by contemporary and stochastic models. Using null models to control for the confounding effect of species richness, we identify areas of endemism that cannot be conditionally explained by contemporary environmental or stochastic effects, suggesting a potential signature of historical processes on species richness. Specifically, we ask (1) How many narrow-ranged species would be expected in an assemblage based solely on the overall species richness of that assemblage, and how many and which centres of endemism remain when this effect is controlled for? (2) To what extent can the occurrence of these centres of endemism be explained from environment and topography alone? and (3) Do centres of endemism contain more species than other regions, even after controlling for any sampling effects and accounting for potential differences in environment?

DATA AND METHODS

Distribution data

The distributional data and grid used here are identical to that in Jetz & Rahbek (2002), as compiled by the Zoological Museum, University of Copenhagen (Burgess *et al.* 1998). This database consists of breeding distribution data for all 1599 birds endemic to Sub-Saharan Africa across a 1° latitudinal–longitudinal grid of 1738 quadrats and contains 366 853 species presence records (quadrats containing ≤ 50% dry land were excluded). We defined species with geographic range sizes ≤ 10 quadrats as 'narrow-ranged

species' ($n = 190$ species, representing 0.27% of all quadrat records) and the quadrats in which they occur as 'Centres of Endemism' ($n = 423$). 'Areas of endemism' are traditionally defined as regions with at least two overlapping species restricted in range (Harold & Mooi 1994; Stattersfield *et al.* 1998; Hausdorf 2002). Our approach here is somewhat different, in that we are interested in the potential special signature of the occurrence of narrow-ranged species as such, disregarding their immediate relevance for vicariance biogeography. For conservation studies a range size cut-off of 50 000 km^2 is often used (e.g. Stattersfield *et al.* 1998), but here we chose 10 quadrats (approximately 110 000 km^2) as a compromise between statistical power and the ability to identify unique regions (see also Fjeldsa 2003).

Null model predictions

Some quadrats may be more likely than others to include many narrow-ranging species simply because of sample-size effects (Connor & Simberloff 1979; Gotelli & Graves 1996; Maurer 1999; Mast & Nyffeler 2003) or geometric constraints (Colwell & Lees 2000; Jetz & Rahbek 2001; Colwell *et al.* 2004; Pimm & Brown 2004). With regard to sample size, all other things being equal, one would expect species-rich quadrats to contain more species of all range size categories, including more narrow-ranged species, than species-poor quadrats. Researchers have attempted to address this issue by deriving indices that down-weight the occurrence of wide ranging species and/or divide by overall richness (assuming a proportional relationship, e.g. Linder 2001), but none of these corrections control for richness satisfactorily. With regard to geometric constraints, mid-domain models (Colwell & Lees 2000) predict that wide-ranging species are more likely to overlap in the interior of a bounded domain (such as sub-Saharan Africa, bounded by sea and desert) than nearer its edges for wholly non-biological reasons, thus producing a different pattern of range-size frequencies in different quadrats (Lees *et al.* 1999; Jetz & Rahbek 2002).

These issues have so far thwarted rigorous tests as to whether assemblages do in fact contain an unusual number of endemics, given their overall richness, or conversely, whether putative centres of endemism are in fact more species rich than other areas. Here we develop two null models that are intended to remove the 'noise' (quadrats that have high numbers of narrow-ranged species simply because they have high richness) from the 'signal' (areas with more narrow-ranged species than expected by chance given their level of richness). To select between the two models, we here assume that the null model that better fits the observed distribution quantitatively, and therefore removes the most 'noise,' is to be preferred.

Model 1

All else being equal, a given quadrat is more likely to contain species with a large than a small geographic range size (i.e. number of quadrat records). In order to allow differences in range size to bear consequence on species representation in communities, the probability of a species' inclusion should be proportional to its geographic range size in relation to the sum of all range sizes and the species richness N_A of the assemblage. This representation in a local quadrat assemblage can be simulated by a random draw of species from the observed list of species' quadrat-records. Wide-ranged species contribute more quadrats than narrow-ranged species, and are thus more likely to be sampled. Only sampling without replacement achieves an unbiased representation of species, and ensures that each species can be represented in an assemblage at most once.

We performed Model 1 simulations using a custom-written C program. For each hypothetical quadrat species richness value N_A (between 1 and 1000), quadrat records were drawn at random from the observed list of quadrat records (366 853 overall) and the species represented by that record, if not already present in the quadrat, was added to the list of species until N_A distinct species had been drawn. This procedure was repeated 1000 times for each value of N_A (yielding 1 million species lists, in all) and the average number (and 95% percentiles) of narrow-ranged species was calculated. This yielded a general relationship between quadrat species richness N_A and expected number of narrow-ranged species per quadrat, which we then applied to the empirical patterns.

Model 2

Geometric constraints imposed by hard boundaries (such as continental edge for terrestrial species) on species richness are expected to 'force' the occurrence of wide-ranged species towards the middle, while narrow-ranged species should be unaffected (Lees *et al.* 1999; Colwell & Lees 2000; Jetz & Rahbek 2001; Colwell *et al.* 2004; Pimm & Brown 2004). One potential prediction of this effect is a higher overall richness of species in the middle of a continent, driven by the higher number of wide-ranged species that tend to overlap there (e.g. Jetz & Rahbek 2002). However, a mixed scenario is possible, in which overall quadrat species richness is mostly driven by historical and environmental factors, but geometric constraints may affect the range size composition of assemblages. If the middle of a continent is more likely to contain wide-ranged species than the edge, quadrats of equal overall species richness should contain more narrow-ranged species near the edge than the middle. Thus, compared with the assumption of Model 1 – that the random draw from quadrat records applies equally to all quadrats – consideration of geometric constraints

predicts the presence of a relative 'excess' of wide-ranging species in interior quadrats and a relative 'deficiency' of wide-ranging species in quadrats near the continental edge.

We modelled the location-specific predicted number of narrow-ranged species, given geometric constraints and observed overall species richness, as follows: first we used the two-dimensional 'spreading dye' model presented by Jetz & Rahbek (2001) as implemented in GEOSPOD (Jetz 2001) and the observed list of range sizes to simulate for each quadrat a list of species occurrences (with number of species per quadrat given by the geometric constraints predictions), performing 100 runs. This resulted in a list of 366 853 000 species occurrences across the 1738 1° quadrats. For each quadrat we then performed a random draw (without replacement, i.e. each species was only sampled once) from the quadrat-specific list of species occurrences until the actually observed species richness, N_A, of that quadrat was reached, and repeated this procedure 100 times. We then used this species list to calculate the average number of narrow-ranged species predicted for this quadrat.

Predictor variables

Overall, we used 14 predictor variables to evaluate the effect of environmental and topographic conditions. These include eleven variables related to contemporary climate and derived features (e.g., net primary productivity, i.e. NPP and NPP^2, habitat heterogeneity, eight climatic variables, three of which reflect seasonality) and three variables associated with altitude (mean and range) and area. Levels of overall richness may be affected by geometric constraints (mid-domain effect, Colwell & Lees 2000; Jetz & Rahbek 2001; Colwell *et al.* 2004; Pimm & Brown 2004). Thus, for the analyses on overall species richness inside and outside centres of endemism we included predictions (expected species richness values) from null model simulations using the observed range sizes and the assumption of fully continuous ranges (using the program GEOSPOD, Jetz 2001). Details on sources and calculations of all 15 predictor variables are given in Jetz & Rahbek (2002).

Statistics

We tested the performance of null models and environmental variables in predicting presence and absence of centres of endemism using sensitivity and specificity, with a cut-off of ≥ 1 narrow-ranged species for null model predictions, and 50% presence–absence probability for environmental logistic regression predictions. Additionally we calculated the accuracy measure 'area under the curve'

(AUC) of receiver–operating characteristic (ROC) plots, a threshold-independent measure of goodness-of-fit (Fielding & Bell 1997). Following Swets (1988) AUC values < 0.7 are considered as poor, those > 0.7 as reasonable, and those > 0.9 as good. We use logistic regression to further examine null model predictions and to investigate the role of environmental predictors. In order to pre-select core environmental variables for multiple regression within each type of predictor (Table 1) we identified pairs of variables that showed high collinearity [abs(r_S) > 0.5)] and only retained the variable with the higher explanatory power in one-predictor regressions. We used model simplification in order to find minimum adequate models for endemism status (logistic regression). Initially, all core variables were included in the model. Subsequently, the predictor with the lowest log-likelihood was excluded until all variables remaining in the model were significant at the 0.05 level. This analysis is confounded by spatial autocorrelation, which can affect parameter estimates (Cliff & Ord 1981; Lennon 2000; Jetz & Rahbek 2002). Owing to a lack of readily available methods for spatial logistic regression we did not control for spatial autocorrelation in this endemism prediction analysis. The resulting likely spatial non-independence of model residuals affects parameter estimates, measures of fitness and thus ranking of predictor variables. The strength of this effect depends on the specific relationship between spatial autocorrelations of response and predictor variable. Strong differences in the fit of predictor variables with similar spatial autocorrelation are likely to be robust to this issue.

We performed linear regression analysis on species richness of all 1599 bird species endemic to Africa over all 1738 quadrats of sub-Saharan Africa, similar to Jetz & Rahbek (2002), with an *ad hoc* model of all 15 environmental/topographic/geometric constraints predictor variables. We compared species richness predictions between Center of Endemism quadrats and all others using a traditional regression model. For statistical evaluation, we entered Center of Endemism status (yes/no) as a categorical variable with the aforementioned 15-predictor variables and compared its significance relative to other important predictors. We repeated the richness prediction analysis employing spatial regression techniques to evaluate and control for spatial autocorrelation (Cliff & Ord 1981). Spatial regression separates the variation across a lattice into large-scale changes due to predictor effects and small-scale variation due to interactions with neighbours, which can be modelled by iterative fitting of an autoregressive covariance model to the dispersion matrix. Here, we use a simultaneous autoregressive model (SAR) and a King's case (eight immediate neighbours) neighbourhood structure and performed the calculations in the S+ SPATIAL STATS Module (Kaluzny *et al.* 1998).

Category	Predictor	One-predictor models			Minimum adequate model		
		Dir	Chi-square	P	Dir	Chi-square	P
Altitude	Mean altitude	+	18.80	***			
	Altitudinal range	+	122.60	***	+	108.89	***
Productivity	NPP	+	0.12				
	NPP^2	+	0.11				
Habitat	NDVI classes	+	58.80	***			
heterogeneity	TempRange	−	53.76	***	−	52.43	***
Seasonality	NDVI seasonality	−	17.57	***	−	14.08	***
	RainRange	+	0.01				
	MaxTemp	−	62.39	***			
	Precipitation	+	0.08				
Climatic factors	AET	+	0.14				
	PET	−	4.03	*			
	Radiation	−	2.04		+	7.05	**
Area	Area dry land	+	11.91	**			

Table 1 Environmental/topographic logistic regression one-predictor and minimum adequate model results for the occurrence of Model 1 Centres of Endemism

Habitat heterogeneity was estimated by counting the annual average of monthly classes of NDVI (normalized difference vegetation index) per quadrat (0.05° resolution).
Dir, direction of the relationship; Chi-square, change in −2 log-likelihood compared with a statistical model without that predictor; NPP, net primary productivity.
TempRange and RainRange refer to average annual range in monthly mean temperature and precipitation, respectively. NDVI seasonality refers to the intra-annual coefficient of variation in monthly mean quadrat NDVI. AET refers to actual evapotranspiration; PET to potential evapotranspiration.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

RESULTS

Model 1 predicts a non-linear increase of expected narrow-ranged species with increasing quadrat species richness (Fig. 1a). Any quadrat with over 285 species is expected to contain at least one narrow-ranged species. Model 2 similarly predicts a concave upward increase of narrow-ranged species richness with overall quadrat species richness, but with predicted richness levels modulated by the distance of quadrats to the continental edge (Fig. 1b). In both models the maximum number of narrow-ranged species predicted per quadrat is under four, which is in stark contrast to the observed data with values as high as 28 (Fig. 1c).

Figure 2 illustrates the performance of null models in predicting the spatial occurrence of Centres of Endemism (quadrats with at least one observed/predicted narrow-ranged species). Both null models perform better in correctly predicting absences than presences (high specificity, Model 1: 0.86, Model 2: 0.85; low sensitivity, Model 1: 0.49, Model 2: 0.47) and have reasonable accuracy (AUC, Model 1: 0.77, Model 2: 0.75). They both show a highly significant logistic regression fit (likelihood ratio test, Model 1: $\chi^2 = 323.88$, Model 2: $\chi^2 = 271.59$, $P < 0.001$ in both cases, $n = 1738$). Model 2 appears to better mimic

the geographic pattern of Center of Endemism occurrence (Fig. 2), but quadrat-by-quadrat Model 1 achieves a significantly better fit than Model 2 (logistic regression, likelihood ratio test; $\chi^2 = 52.30$, $P < 0.001$). When model performance is tested on the number of narrow-ranged species among quadrats that contain them, both Model 1 and 2 are statistically significant predictors (Poisson regression, likelihood ratio tests; Model 1: $\chi^2 = 294.12$, $P < 0.001$, Model 2: $\chi^2 = 147.55$, $P < 0.001$, $n = 423$), but Model 1 yields a significantly stronger fit ($\chi^2 = 146.57$, $P < 0.001$). Comparing the logistic regression fits achieved by the null models to those of 14 environmental/topographic variables we find that even the most significant predictor variable (habitat heterogeneity) has much smaller predictive power than the null model predictions (likelihood ratio test, Model 1: $\chi^2 = 99.74$, $P < 0.001$; Model 2: $\chi^2 = 78.34$, $P < 0.001$).

As Model 1 shows the better fit, we choose it over Model 2 as our null model in subsequent analysis. Thus, we set out to select those quadrats among the defined Centres of Endemism that contain more narrow-ranged species than expected by the chance effects of overall quadrat richness (as modelled by Model 1) alone. We use the upper 95% percentile predictions of Model 1 as lower cut-off to identify these areas (Fig. 3a). Of the original 423 quadrats
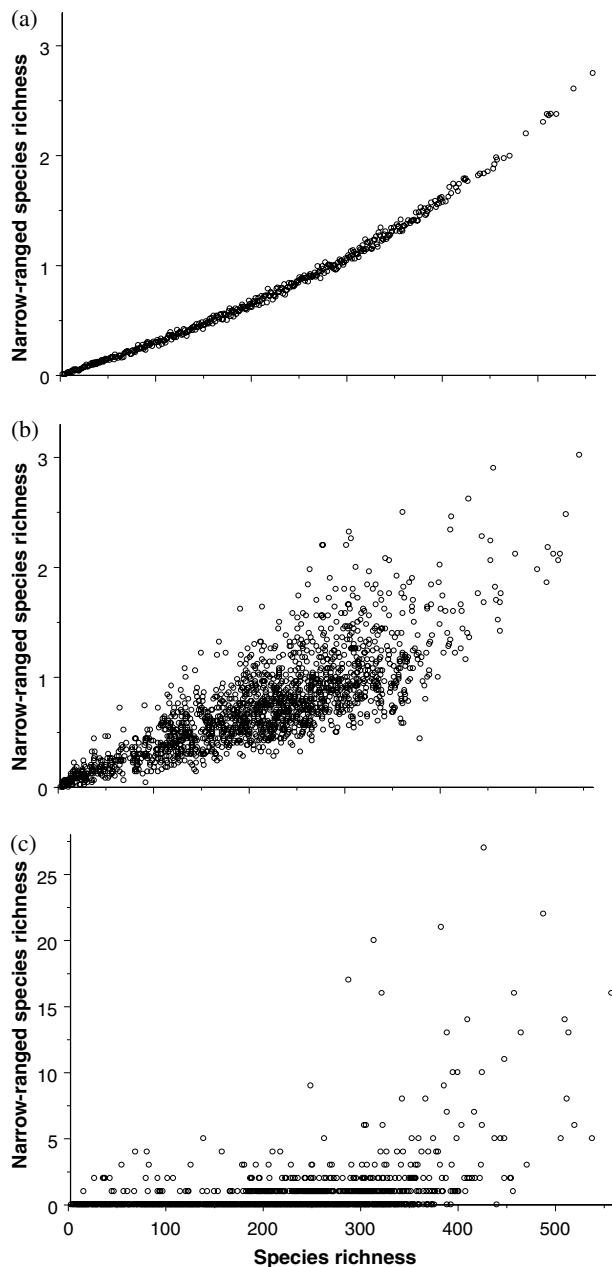
Figure 1 Relationship between the observed overall species richness and the predicted [(a) and (b)] or observed (c) narrow-ranged species richness across 1° quadrats of sub-Saharan Africa. Narrow-ranged species are those with a geographic range size ≤ 10 quadrats. (a) Model 1 predictions (random draw of species from continental list of quadrat occurrences). (b) Model 2 predictions (random draw of species from GEOSPOD simulated quadrat specific list of quadrat occurrences). (c) Observed data (note different scale on *y*-axis).

with narrow-ranged species (Fig. 3b), 79 qualify under this criterion (Fig. 3c). These include species rich regions (Cameroon Highlands, Albertine Rift Mountains, Kenya

Highlands, Eastern Zimbabwe mountains), as well as less speciose areas such as the Central Somali Coast and North Somali Highlands, Western Angola, Lesotho Highlands and Southeast Namibia.

We proceed to evaluate the contemporary environmental/ topographic predictability of the occurrence of Model 1 Centres of Endemism. We first perform single-predictor logistic regressions with all 14 environmental/topographic predictor variables to identify important variables (Table 1). We find habitat heterogeneity, altitudinal range, maximum temperature and annual range of temperature to be the statistically most powerful predictors of Model 1 Centres of Endemism. In a minimum adequate model that accounts for collinearity and selects core predictors, a positive effect of topographic heterogeneity (i.e. altitudinal range), a negative effect of two seasonality variables (seasonal temperature range and variation in productivity) and, much less significant, a positive effect of solar radiation emerge as important (Table 1). Of these, altitudinal range is by far the most significant. Using a 0.5 probability cut-off, this logistic regression model performs well in predicting absence, but not well for predicting presence (Fig. 3d, sensitivity: 0.10, specificity: 0.99). However, the cut-off independent goodness-of-fit is reasonable (AUC: 0.92). Well predicted are the mountainous Centres of Endemism in East Africa and Cameroon and coastal areas in Angola and North Somali, but not northern Nigeria, Somali East Coast and Highlands, and the Lesotho highlands. A high concentration of predicted but not (yet?) observed presences along the West Coast of Central Africa down to Southern Namibia is noticeable.

Observed (unadjusted) Centres of Endemism are expected to be more species rich than other regions due to the effect that species richness has on the probability of a quadrat being a Center of Endemism (see above). Observed Centres of Endemism contain on average 100 more species than other quadrats ($U = 129.09$, $P < 0.001$). Yet, even richness-controlled Model 1 Centres of Endemism harbour on average 69 species more species ($U = 45.02$, $P < 0.001$). This strong difference remains when the two to three (on average, per quadrat) narrow-ranged species are taken out of the analysis (reduction in U marginal; $P < 0.001$ remains in all tests). This higher overall species richness inside Centres of Endemism could simply be because of differences in the environmental conditions that correlate with species richness. That is, areas of endemism could simply be areas that environmentally favour high species richness. To test this hypothesis, we entered 'Endemism Status' (whether a quadrat is a Center of Endemism or not) as a binary variable in a full regression model with overall quadrat richness minus narrow-ranged species richness as the dependent variable and all 14 environmental/topographic predictor variables plus geometric constraints predictions as independent variables (Table 2). It emerges that Centres of
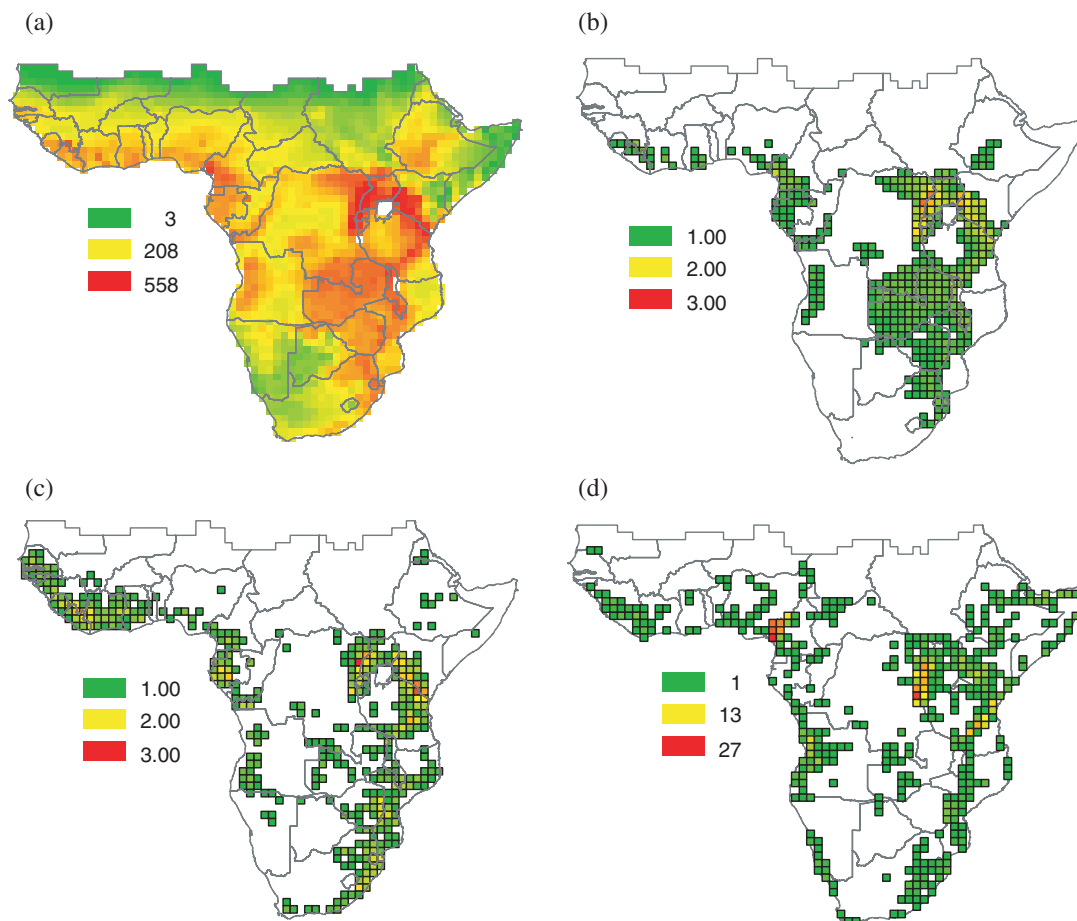
Figure 2 Null model base data and predictions. (a) Observed richness of all 1599 species endemic to Africa. (b) Model 1 predictions for the number of narrow-ranged species (range size ≤ 10, 1° quadrats) expected across Africa, given the observed richness pattern. Only quadrats for which at least one narrow-ranged species is predicted are plotted. (c) same for Model 2. (d) Observed Centres of Endemism of Africa and their narrow-ranged species richness. All are equal interval classification.

Endemism are expected to contain relatively high numbers of species because of their environment alone, in particular because of their tendency to be located in productive areas (NPP is consistently the top predictor of quadrat richness). However, both observed and Model 1 Centres of Endemism, consistently contain even more species than the environmental model predicts. In both cases, in the multi-predictor model, Endemism Status came out as a highly significant additional variable and as an important predictor (Table 2). We repeated the analysis using spatial regression, which supported this result.

DISCUSSION

Our study attempts a synthetic continental analysis of centres of endemism, seeking to investigate the interrelated effects of species richness, environment and history. In methodology and approach it sets out to provide a

link between traditional environmental-correlate based analyses of species richness, null model focused investigations and historical approaches that emphasize regional context.

The notion that simple chance effects or constraints should be controlled for, or at least evaluated, is now an established concept in community and broad-scale ecology (Connor & Simberloff 1983; Gotelli & Graves 1996; Gotelli 2001), although still debated by some. However, null models still appear to play only a minor (and contentious, e.g. Hawkins & Diniz 2002; Zapata *et al.* 2003; Colwell *et al.* 2004) role in biogeographic and historical analyses. In analyses of local communities it has become clear that, beyond environmental and historical explanation on the local scale, broader consideration is required of the surrounding or source region in which the local assemblage is embedded, as well as the degree to which local biotic composition actually differs from chance expectation. The
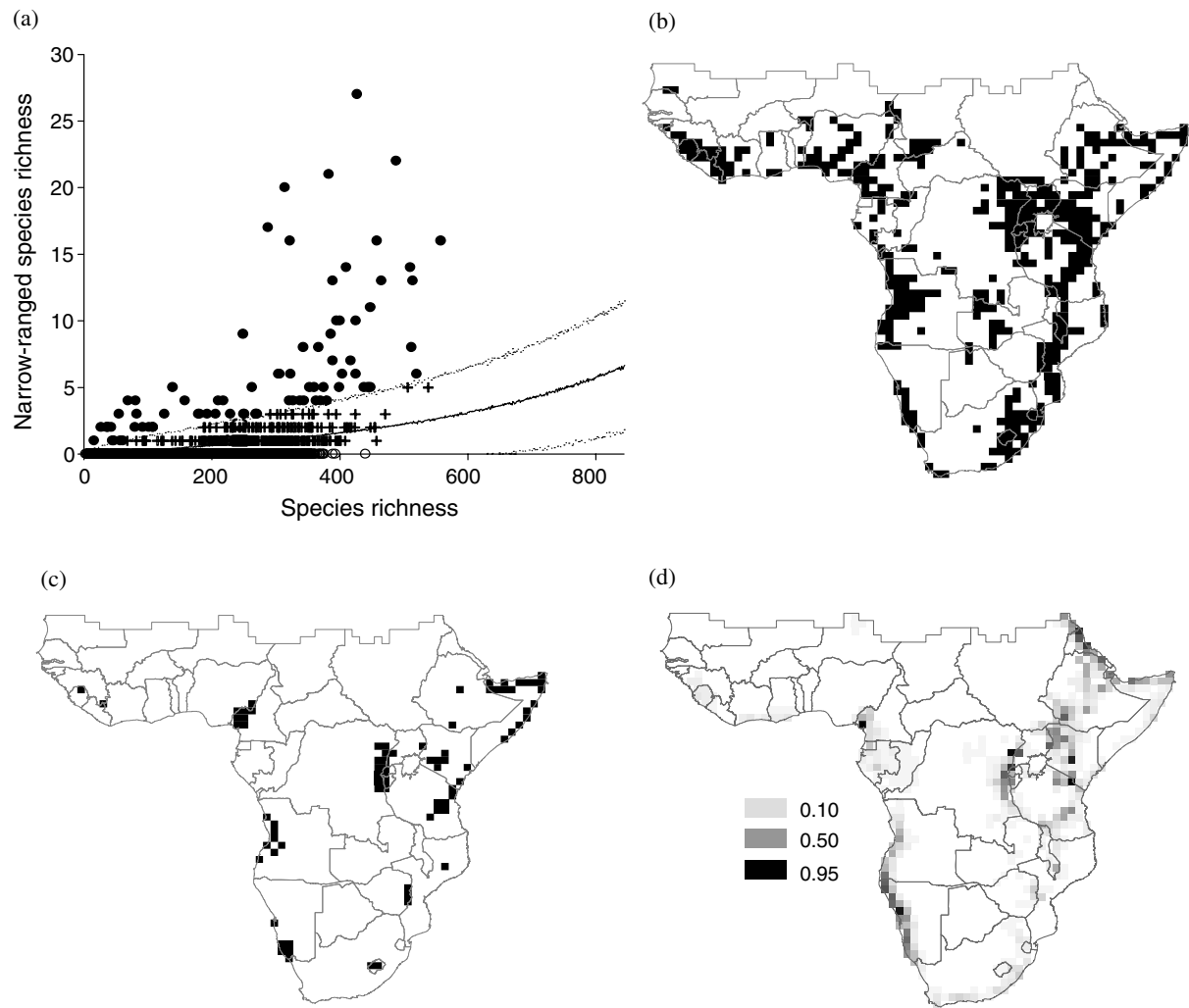
Figure 3 Selection, distribution and environmental predictability of Model 1 Centres of Endemism. (a) Richness of narrow-ranged species in a quadrat in relation to its overall species richness as predicted by Model 1. Solid thick line: predictions [± 95% confidence intervals (CI), thin lines] of Model 1. Open circles: quadrats with no narrow-ranged species. Crosses: quadrats with at least one narrow-ranged species, but less than predicted by Model 1 (≤ 95% CI) . Solid circles: Model 1 Centres of Endemism, i.e. quadrats with more narrow-ranged species than predicted by Model 1 (> 95% CI). (b) Observed Centres of Endemism before selection [all symbols in (a)]. (c) Model 1 Centres of Endemism [i.e. the solid circles in (a)]. (d) Geographic predictions of probability of Model 1 Center of Endemism occurrence according to the minimum adequate environmental logistic regression model (Table 2). Equal interval classification.

extent to which ecological characteristics of assemblages deviate from random draw models has already provided invaluable insights for many studies on island communities (see Gotelli & Graves 1996 for review). The application of this approach to local assemblages as part of regional pools on the mainland may be similarly valuable (Maurer 1999; Blackburn & Gaston 2001), and this is the first study to use it on a continental scale.

As typical for most taxa on broad scales (Gaston 2003), the range size distribution of African birds is highly right-skewed: there are many more narrow- than wide-ranged species (see also Hall & Moreau 1962; Pomeroy & Ssekabiira

1990). Therefore, the chance relationship between overall and narrow-ranged species richness is not linear (Fig. 1). As demonstrated by our simulation models, relatively more narrow-ranged species are expected in species rich assemblages. Of course, by allocating species from the full continental pool and thus sampling from the composite range-size frequency distribution across the continent, one may neglect essential local processes or factors that contribute to this frequency distribution. In the context of this analysis, the evolution and existence of extremely narrow ranges within the overall range size distribution may rest on an explanation or a constraint that does not warrant

Observed Centres of Endemism ($n = 423$)				Model 1 Centres of Endemism ($n = 79$)			
Rank	Variable	t	P	Rank	Variable	t	P
GLM							
1	NPP^2	−12.72	***	1	NPP^2	−13.69	***
2	RainRange	12.04	***	2	NPP	11.99	***
3	NPP	10.65	***	3	RainRange	11.50	***
4	Endemism Status	9.99	***	10	Endemism Status	4.10	***
	Moran's I	0.67	***		Moran's I	0.69	***
SAR							
1	NPP	9.75	***	1	NPP	9.96	***
2	Altitudinal range	8.87	***	2	NDVI classes	8.64	***
3	NDVI classes	8.63	***	3	NPP^2	−8.10	***
4	Endemism status	7.57	***	6	Endemism status	4.86	***
	Moran's I	0.06	***		Moran's I	−0.01	n.s.

Table 2 Select results of 15 predictors regression on overall minus narrow-ranged species richness of birds across sub-Saharan Africa

Independent variables include the 14 topographic/environmental predictors and geometric constraints predictions for overall species richness (see Data and Methods). Endemism status is a binary variable that indicates whether a quadrat is a Centre of Endemism or not. Only results of top three predictors are shown, together with the rank and results for the focal variable, endemism status.

Moran's I measures the spatial autocorrelation of the model residuals. For predictor terms and symbols see Table 1.

GLM, traditional linear regression model; SAR, spatial autoregressive model.

random allocation across the continent. This critique can only be addressed by careful interpretation. Here we propose that the close and highly confounding inter-relationship between overall and narrow-ranged species richness both justifies and requires a null model approach. It confirms a potentially prominent role for random draw models in the study of continental biota. Our simplistic criterion – best logistic regression fit – selected Model 1 as best null model, but did not take into account spatial autocorrelation effects and similarity in geographic pattern (Fig. 3) which may have favoured Model 2. Model 2 is logically valid and may well have higher explanatory power in other datasets.

Because of its restriction to a standardized 1° grid, our study is limited in the extent to which it allows interpretation of exact distributional boundaries of narrow-ranged birds (see e.g. Hall & Moreau 1962; Terborgh & Winter 1982; De Klerk et al. 2002). Generally, we find that species with narrow ranges show a very distinct pattern of occurrence that can only partly be predicted from contemporary factors. One of the two strong predictors of the geographic location of centres of endemism is large altitudinal range within a quadrat (not altitude per se). Although altitudinal range has sometimes been used as an estimate of habitat heterogeneity, topographic heterogeneity measured as altitudinal range might better be viewed as a rough surrogate variable reflecting historical opportunities for allopatric speciation (e.g. Rahbek & Graves 2001; Jetz & Rahbek 2002). Altitudinal separation within a quadrat measures topogra-

phical variation and occurrence of narrow homothermous elevational bands and thus indicates the potential existence of past and present barriers that facilitate speciation and beta-diversity (see also Janzen 1967; Vuilleumier 1969; Graves 1988; Rahbek 1997). Montane areas, in particular montane forests, have repeatedly been demonstrated to be key areas for narrow-ranged species (Terborgh & Winter 1982; Stattersfield et al. 1998), also in Africa (Diamond & Hamilton 1980; Collar & Stuart 1988; Johnson et al. 1998; Linder 2001).

Contemporary climate conditions are usually seen as important for processes maintaining species richness (e.g. Currie et al. 1999), however they may also convey a strong historical signature (e.g. Latham & Ricklefs 1993). In our study we identify low seasonality, best captured through annual temperature range and obviously a measure of contemporary climate, as the second important predictor of centres of endemism. The role of low seasonality, as such, in promoting occurrence of centres of endemism has as yet not been quantitatively demonstrated. From an ecological perspective the connection between very small range size and low seasonality may not be surprising. Seasonal environments are likely to limit tight habitat specializations, which in turn may thwart small ranges (Janzen 1967; Huey 1978; Stevens 1989). The significance of seasonality may be related to the idea of ecoclimatic stability furthering the persistence of relictual endemics (Fjeldsa et al. 1997; Dynesius & Jansson 2000). One interpretation of the role of eco-climatic stability was

triggered by the observation that in African montane forests the distributions of both phylogenetically old and young species coincide (Fjeldsa & Lovett 1997), suggesting a causal link between speciation, endemism and long-term stability (Fjeldsa *et al*. 1997).

We may conclude that that centres of endemism are concentrated in regions that offered unusually many opportunities for past speciation, combined with stable climates that allowed survival of narrow endemics despite their small geographic ranges. It has been argued that if centres of endemism really have acted as centres of speciation, as 'speciation pumps', in the past (Terborgh 1992), they are likely to contain more species than other regions today (Endler 1982b; Haffer 1982; Prance 1982). Here we are able confirm this prediction and show that, in Africa, centres of endemism indeed do contain more species than expected by chance or environment and topography alone. However, we are unable to separate to what degree this pattern arises due to past differences in rates of speciation and extinction, or immigration and emigration, and can therefore not directly test the 'speciation pump' hypothesis. Not only speciation but species persistence (locally low extinction) may determine the occurrence of centres of endemism (Mayr 1963). Knowledge of which centres of endemism are primary (based on relictual, autochthonous endemics) would yield stronger insights about historical mechanisms. The fact that clear patterns appear even in an analysis without that knowledge underlines the strength of a historical explanation and supports the likely role of centres of endemism as past centres of cladogenesis (Croizat *et al*. 1974; Ricklefs & Schluter 1993).

We believe that our results on the distribution of centres of endemism help to elucidate the role of history in shaping avian distributions in Africa. We have illustrated how an analysis with contemporary environmental predictors can help to support historical interpretations, although better palaeoclimatic and phylogenetic information is still badly needed. Further, our results demonstrate that not even advanced information on environmental variables and modelling techniques are likely to be sufficient to delineate areas and species of prime conservation concern. Greater recognition of the value of primary ecological surveys is needed.

ACKNOWLEDGEMENTS

REFERENCES

Blackburn, T.M. & Gaston, K.J. (2001). Local avian assemblages as random draws from regional pools. *Ecography*, 24, 50–58.

Brown, J.H. (1995). *Macroecology*. University of Chicago Press, Chicago, IL.

Burgess, N., Fjeldså, J. & Rahbek, R. (1998). Mapping the distributions of Afrotropical vertebrate groups. *Species*, 30, 16–17.

Cliff, A.D. & Ord, J.K. (1981). *Spatial Processes: Models & Applications*. Pion, London.

Collar, N.J. & Stuart, S.N. (1988). *Key Forests for Threatened Birds in Africa*. International Council for Bird Preservation, Cambridge, UK.

Colwell, R.K. & Lees, D.C. (2000). The mid-domain effect: geometric constraints on the geography of species richness. *Trends Ecol. Evol.*, 15, 70–76.

Colwell, R.K., Rahbek, C. & Gotelli, N.J. (2004). The mid-domain effect and species richness patterns: what have we learned so far? *Am. Nat.*, 163, E1-E23.

Connor, E.F. & Simberloff, D.S. (1979). The assembly of species communities: chance or competition. *Ecology*, 60, 1132–1140.

Connor, E.F. & Simberloff, D. (1983). Neutral models of species' co-occurrence patterns. In: *Ecological Communities: Conceptual Issues and the Evidence* (eds Strong, D.R., Simberloff, D. & Abele, L.G.). Princeton University Press, Princeton, NJ.

Cracraft, J. (1994). Species-diversity, biogeography, and the evolution of biotas. *Am. Zool.*, 34, 33–47.

Crandall, K.A., Bininda-Emonds, O.R., Mace, G.M. & Wayne, R.K. (2000). Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.*, 15, 290–295.

Croizat, L., Nelson, G. & Rosen, D.E. (1974). Centers of origin and related concepts. *Syst. Zool.*, 23, 265–287.

Currie, D.J. (1991). Energy and large-scale patterns of animal species and plant species richness. *Am. Nat.*, 137, 27–49.

Currie, D.J., Francis, A.P. & Kerr, J.T. (1999). Some general propositions about the study of spatial patterns of species richness. *Ecoscience*, 6, 392–399.

De Klerk, H.M., Crowe, T.M., Fjeldsa, J. & Burgess, N.D. (2002). Biogeographical patterns of endemic terrestrial Afrotropical birds. *Diversity Distributions*, 8, 147–162.

Diamond, A.W. & Hamilton, A.C. (1980). The distribution of forest passerine birds and Quaternary climatic change in Africa. *J. Zool.*, 191, 379–402.

Dynesius, M. & Jansson, R. (2000). Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proc. Natl Acad. Sci. USA*, 97, 9115–9120.

Endler, J.A. (1982a). Pleistocene forest refuges: fact or fancy? In: *Biological Diversification in the Tropics* (ed. Prance, T.G.). Columbia University Press, New York, NY, pp. 641–657.

Endler, J.A. (1982b). Problems in distinguishing historical from ecological factors in biogeography. *Am. Zool.*, 22, 441–452.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24, 38–49.

Fjeldsa, J. (2003). Patterns of endemism in African birds: how much does taxonomy matter? *Ostrich*, 74, 30–38.

Fjeldsa, J. & Lovett, J.C. (1997). Geographical patterns of old and young species in African forest biota: the significance of specific montane areas as evolutionary centres. *Biodiversity Conserv.*, 6, 325–346.

Fjeldsa, J., Ehrlich, D., Lambin, E. & Prins, E. (1997). Are biodiversity 'hotspots' correlated with current ecoclimatic stability? A pilot study using the NOAA-AVHRR remote sensing data. *Biodiversity Conserv.*, 6, 401–422.

Fjeldsa, J., Bayes, M.K., Bruford, M.W. & Roy, M.S. (in press). Biogeography and diversification of African forest faunas: implications for conservation. In: *Tropical Rainforests – Past, Present and Future* (eds Bermingham, E., Dick, C., Moritz, C.) Chicago University Press, Chicago, IL.

Francis, A.P. & Currie, D.J. (2003). A globally consistent richness-climate relationship for angiosperms. *Am. Nat.*, 161, 523–536.

Gaston, K.J. (2003). *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, Oxford, UK.

Gotelli, N.J. (2001). Research frontiers in null model analysis. *Global Ecol. Biogeogr.*, 10, 337–343.

Gotelli, N.J. & Graves, G.R. (1996). *Null Models in Ecology*. Smithsonian Institution, Washington, DC.

Graves, G.R. (1988). Linearity of geographic range and its possible effect on the population-structure of Andean birds. *Auk*, 105, 47–52.

Haffer, J. (1982). General aspects of the refuge theory. In: *Biological Diversification in the Tropics* (ed. Prance, T.G.). Columbia University Press, New York, NY, pp. 6–24.

Hall, B.P. & Moreau, R.M. (1962). A study of rare birds of Africa. *Bull. Br. Museum (Nat. Hist.) Zool.*, 8, 313–378.

Harold, A.S. & Mooi, R.D. (1994). Areas of endemism – definition and recognition criteria. *Syst. Biol.*, 43, 261–266.

Hausdorf, B. (2002). Units in biogeography. *Syst. Biol.*, 51, 648–652.

Hawkins, B.A. & Diniz, J.A.F. (2002). The mid-domain effect cannot explain the diversity gradient of Nearctic birds. *Global Ecol. Biogeogr.*, 11, 419–426.

Huey, R.B. (1978). Latitudinal pattern of between-altitude faunal similarity: mountains might be 'higher' in the tropics. *Am. Nat.*, 112, 225–229.

Janzen, D.H. (1967). Why mountain passes are higher in the tropics. *Am. Nat.*, 101, 233–249.

Jetz, W. (2001). GEOSPOD – a program to model the effect of geometric constraints. Oxford, UK, URL: http://evolve.zoo.ox.ac.uk/.

Jetz, W. & Rahbek, C. (2001). Geometric constraints explain much of the species richness pattern in African birds. *Proc. Natl Acad. Sci. USA*, 98, 5661–5666.

Jetz, W. & Rahbek, C. (2002). Geographic range size and determinants of avian species richness. *Science*, 297, 1548–1551.

Johnson, D.D.P., Hay, S.I. & Rogers, D.J. (1998). Contemporary environmental correlates of endemic bird areas derived from meteorological satellite sensors. *Proc. R. Soc. Lond. B Biol. Sci.*, 265, 951–959.

Kaluzny, S.P., Vega, S.C., Cardoso, T.P. & Shelly, A.A. (1998). S+ SPATIAL STATS. Mathsoft Inc., Seattle, WA.

Latham, R.E. & Ricklefs, R.E. (1993). Global patterns of tree species richness in moist forests: energy–diversity theory does not account for variation in species richness. *Oikos*, 67, 325–333.

Lees, D.C., Kremen, C. & Andriamampianina, L. (1999). A null model for species richness gradients: bounded range overlap of butterflies and other rainforest endemics in Madagascar. *Biol. J. Linn. Soc.*, 67, 529–584.

Lennon, J.J. (2000). Red-shifts and red herrings in geographical ecology. *Ecography*, 23, 101–113.

Linder, H.P. (2001). Plant diversity and endemism in sub-Saharan tropical Africa. *J. Biogeogr.*, 28, 169–182.

MacArthur, R.H. (1972). *Geographical Ecology*. Harper & Row, New York, NY.

Mast, A.R. & Nyffeler, R. (2003). Using a null model to recognize significant co-occurrence prior to identifying candidate areas of endemism. *Syst. Biol.*, 52, 271–280.

Maurer, B.A. (1999). *Untangling Ecological Complexity: The Macroscopic Perspective*. University of Chicago Press, Chicago, IL.

Mayr, E. (1963). *Animal Species and Evolution*. Belknap Press, Cambridge, MA.

Moritz, C., Patton, J.L., Schneider, C.J. & Smith, T.B. (2000). Diversification of rainforest faunas: an integrated molecular approach. *Annu. Rev. Ecol. Syst.*, 31, 533–563.

Nelson, G. & Platnick, N.I. (1981). *Systematics and Biogeography*. Columbia University Press, New York, NY.

Patterson, B.D. (1999). Contingency and determinism in mammalian biogeography: the role of history. *J. Mammal.*, 80, 345–360.

Pimm, S.L. & Brown, J.H. (2004). Domains of diversity. *Science*, 304, 831–833.

Pomeroy, D. & Ssekabiira, D. (1990). An analysis of the distributions of terrestrial birds in Africa. *Afr. J. Ecol.*, 28, 1–14.

Prance, T.G. (1982). *Biological Diversification in the Tropics*. Columbia University Press, New York, NY.

Rahbek, C. (1997). The relationship among area, elevation, and regional species richness in neotropical birds. *Am. Nat.*, 149, 875–902.

Rahbek, C. & Graves, G.R. (2001). Multiscale assessment of patterns of avian species richness. *Proc. Natl Acad. Sci. USA*, 98, 4534–4539.

Ricklefs, R.E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.*, 7, 1–15.

Ricklefs, R.E. & Schluter, D. (1993). Species diversity: regional and historical influences. In: *Species Diversity in Ecological Communities* (eds Ricklefs, R.E. & Schluter, D.). Chicago University Press, Chicago, IL.

Ricklefs, R.E., Latham, R.E. & Qian, H. (1999). Global patterns of tree species richness in moist forests: distinguishing ecological influences and historical contingency. *Oikos*, 86, 369–373.

Rosen, D.E. (1978). Vicariant patterns and historical explanations in biogeography. *Syst. Zool.*, 27, 159–188.

Rosenzweig, M.L. (1995). *Species Diversity in Space and Time*. Cambridge University Press, Cambridge, UK.

Schall, J.J. & Pianka, E.R. (1978). Geographical trends in number of species. *Science*, 201, 679–686.

Schneider, C.J., Smith, T.B., Larison, B. & Moritz, C. (1999). A test of alternative models of diversification in tropical rainforests: ecological gradients vs. rainforest refugia. *Proc. Natl Acad. Sci. USA*, 96, 13869–13873.

Stattersfield, A.J., Crosby, M.J., Long, A.J. & Wege, D.C. (1998). *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. Bird Life International, Cambridge, UK.

Stevens, G.C. (1989). The latitudinal gradient in geographical range: how so many species coexist in the tropics. *Am. Nat.*, 133, 240–256.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.

Terborgh, J. (1992). *Diversity and the Tropical Rain Forest*. Freeman, New York, NY.

Terborgh, J. & Winter, B. (1982). Evolutionary circumstances of species with small ranges. In: *Biological Diversification in the Tropics* (ed. Prance, T.G.). Columbia University Press, New York, NY, pp. 587–600.

Vuilleumier, F. (1969). Pleistocene speciation of birds living in the high Andes. *Nature*, 223, 1179–1180.

Zapata, F.A., Gaston, K.J. & Chown, S.L. (2003). Mid-domain models of species richness gradients: assumptions, methods and evidence. *J. Anim. Ecol.*, 72, 677–690.