ECOLOGICAL SOUNDING

Sensitivity of macroecological patterns of South American parrots to differences in data sources

Pablo V. C. Mathias†, Cláudio V. Mendonça†, Thiago F. L. V. B. Rangel‡ and José A. F. Diniz-Filho*‡§

†*Departamento de Biologia, Universidade Católica de Goiás, Goiânia, Brazil,* §*Departamento de Biologia — MCAS/VPG, Universidade Católica de Goiás, Goiânia, Brazil and* ‡*Departamento de Biologia Geral, ICB, Universidade Federal de Goiás, Goiânia, GO 74001–970, Brazil. E-mail: diniz@icb.ufg.br*

Correspondence: Departamento de Biologià Geral, ICB, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil. E-mail: diniz@icb.ufg.br

ABSTRACT

A criticism of macroecological studies has been their extensive use of secondary data sources. In this note we evaluate how different data sources affect macroecological patterns for the parrots of South America. We mapped extents of parrot occurrence based on four sources of range maps. We compared basic statistics for geographical range size distribution (mean, variance and skew) and calculated correlations between geographical range size estimates and grid cell species richness estimates. Finally, results from multiple regression analyses of species richness against six environmental variables were also compared. We found that patterns were very robust to the data source, with only relatively slight quantitative differences. Our results reinforce the notion that patterns emerging from macroecological analyses are robust to variations in data sources and cannot be merely artefacts resulting from low data quality, notably poorly defined mapping and conflicting taxonomy.

Keywords

Data resolution, data sources, geographical range size, macroecology, psittaciformes, spatial autocorrelation, species richness.

INTRODUCTION

Data quality has been one of the most criticized aspects of the macroecology research program, especially when using geographical range size estimates (Blackburn & Gaston, 1998; Gaston & Blackburn, 1999, 2000; Blackburn *et al*., 2004). With a few exceptions (e.g. Rahbek & Graves, 2000, 2001; Jetz & Rahbek, 2001), almost all taxonomically and geographically broad macroecological analyses rely on secondary sources. Thus, researchers working in macroecology must often worry not only about the overall problems associated with all meta-analytical approaches, such as publication bias (Gates, 2002), but also with deficiencies in data sources, notably coarse-grained and poorly defined maps and outdated taxonomies and phylogenies. Some have argued that many macroecological patterns may in fact be artefacts caused by these data deficiencies.

Because of these potential problems, the strategy adopted by most researchers is to find the most recent and updated dataset for the group under study, or to combine different sources. However, the data quality clearly varies among sources and even among taxonomic groups, and it is not surprising that most macroecological analyses are concentrated in mammals and birds, for which relatively well known geographical ranges and species delimitations are available (see Gaston & Blackburn, 2000). Thus, evaluating how emergent macroecological patterns are affected by differences in data sources could expand the amplitude of potential tests to less well-known groups, if it can be determined how robust the resulting patterns really are. Unfortunately, there are few studies focused on this subject.

Blackburn *et al*. (2004) recently addressed the problem of lack of data resolution by analysing macroecological patterns of parrots worldwide, mainly in terms of geographical range size and their correlates, such as body size, population size and degree of migration. Distribution maps were built using the best current available data source for the group and then the geographical range sizes were estimated using different degrees of data resolution, ranging from fine measurements made using geographical information systems up to 250,000 km^2 cells in the WORLDMAP grid. They showed that macroecological patterns of range size and the cross-species correlations were little affected by data resolution and, consequently, concluded that previous results using crude data are unlikely to change if refined measurements of geographical ranges were used instead.

On this note, we evaluated the same problem from a slightly different point of view, by asking how emergent macroecological patterns are affected by using different data sources for the same group of organisms.

VARIATION AMONG DATA SOURCES

We analysed parrots across South America, mapping their extents of occurrence based on four books (Dunning, 1987; Forshaw *et al.*, 1989; del Hoyo *et al.*, 1997; Juniper & Parr, 1998). Although it is difficult to define and measure variation in data quality among sources, these books differ in terms of how well they reflect the species real distributions. For example, Dunning (1987) is an older, general field guide for South American birds, with crudely defined and small printed maps, whereas Juniper & Parr (1998) is a more recent, specific book on parrots, with more detailed maps derived from recent primary sources. del Hoyo *et al.* (1997) and Forshaw *et al.* (1989) are considered here to be of intermediate quality, but whereas del Hoyo *et al.* (1997) possesses general maps and is part of an ongoing review of all birds of the world, Forshaw *et al.* (1989) is a more detailed, but older, book on parrots of the world.

To quantify these differences in the data sources, we evaluated variation in geographical range size (GRS) estimated for each of the 101 species that were common to the four books. The geographical range size of each species (extent of occurrence) was determined by redrawing the distribution maps of each species, based on each book, on an equal-area projection map of South America, with 374 cells of 220 km × 220 km each (see Hawkins *et al.*, 2003a). We considered in these initial specific comparisons only species that retained their taxonomic status among data sources (have the same species name) or, in a few cases, taxa for which changes in status could be clearly identified (e.g. two subspecies in one data source that were considered to be two species in other sources were included as two different species).

We then evaluated if estimated range sizes of each species varied significantly among pairs of sources using paired *t*-tests. Because of the number of tests, the critical α was conservatively adjusted by Bonferroni criterion to 0.008 (the standard 0.05 divided by the number of comparisons, equal to 6). We found that significantly different range size estimates were generated only between Dunning (1987) and both del Hoyo *et al.* (1997) and Juniper & Parr (1998) (Table 1). Differences between Forshaw *et al.* (1989) and del Hoyo *et al.* (1997) and Juniper & Parr (1998) were not significant when Bonferroni correction was applied. So, GRS estimates for each species differ among sources more or less according to what we expected, at least in the sense

of putting Dunning (1987) and Juniper & Parr (1998) at opposite ends of the spectrum. These differences may reflect a mixture of effects that could generally be referred to as data quality, including variation in mapping detail and taxonomic definitions (note that species number of each source also varies — see below). Further, since parrots include a relatively large number of threatened and endangered species (see www.redlist.org), there are probably differences in the current and 'historical' ranges of species, which could be reflected in the maps due to differences in the time periods when the data for generating the ranges maps were compiled (note that largest differences in the paired *t*-tests are between books from different decades). At the present time, it is impossible to determine how each of the sources of variation in data quality contributes to the differences in range size estimates arising among sources. Nevertheless, estimated range sizes appear to be consistent, with the probable exception of the overestimation of ranges in the very small and coarse range maps in Dunning (1987). We can now evaluate how these differences in basic dataset influence emerging macroecological patterns.

MACROECOLOGICAL ANALYSES

We used the four sets of range maps to quantify parrot species richness in each of the 374 cells of our continental grid system. Also, six environmental variables were available for this grid system: (1) Potential evapotranspiration (PET); (2) Annual evapotranspiration (AET); (3) Mean daily temperature in the coldest month (MINT); (4) Rainfall (RAIN); (5) Annual mean temperature (ANNT) and; (6) the difference between maximum and minimum elevations, or topographic heterogeneity (RELEV) (see Diniz-Filho *et al.*, 2003; Hawkins *et al.*, 2003a; for details).

We compared basic statistics (means, variances and skews) for the distribution of GRS of parrots based on different data sources, after log-transformation, and calculated pairwise Pearson correlations between GRS (across species) and between richness estimates (across grid cells). Whereas the low phylogenetic signal in GRS barely affects significance levels of cross-species correlations (Diniz-Filho & Tôrres, 2002; Webb & Gaston, 2003; but see Blackburn *et al.*, 2004), the strong spatial autocorrelation in richness across the continent requires a correction in hypothesis testing for correlations across grid cells. Following Legendre *et al.* (2002), we used the modified *t*-test of Dutilleul (1993) to correct the number of degrees of freedom used to test these correlations (see also Rangel & Diniz-Filho, 2003).

Finally, we compared stepwise-forward multiple regression models of richness against the six environmental variables (predictors), both in terms of overall explanatory power (R^2) and of the partial standardized regression coefficients. The magnitude of spatial autocorrelation in the regression residuals was also compared, using Moran's *I* coefficients, calculated for 15 distance classes, in order to evaluate the amount of spatially structured variation left in dataset after taking into account broad-scale spatial effects (see Legendre & Legendre, 1998; Diniz-Filho *et al.*, 2003).

The pairwise correlations of geographical range sizes between books (Table 2) were always quite high (all *r*-values > 0.83),

Table 1 Pairwise comparison of geographical range size estimates using paired *t*-tests across species for South American parrots based on four different books (Dunning, 1987; Forshaw *et al.*, 1989; del Hoyo *et al.*, 1997; Juniper & Parr, 1998). The critical alpha was Bonferroni-adjusted to 0.008 to correct for the number of tests

	Juniper & Parr	del Hoyo	Dunning	Forshaw
Juniper & Parr	—			
del Hoyo	0.384	—		
Dunning	3.527†	3.376†	—	
Forshaw	2.136*	2.115*	1.533	—

*$P < 0.05$; †$P < 0.008$.

although the estimates by Dunning (1987) were usually less strongly correlated with those from other data sources. Accordingly, the descriptive statistics of GRS based on the different books are also quite similar (Table 3), despite differences in the GRS estimates for each species previously described. Skew values ranged from −0.003 to −0.321, but they were not statistically different from a null expectation (distributions tend to be close to log-normal) and were not different from each other, considering their large standard errors ($t = 1.392$; $P > 0.05$, when comparing smallest and largest skew values).

The spatial patterns in species richness were also quite similar (Fig. 1), with greatest diversity in the eastern Amazon region, decreasing to the south and west. They were also similar to those reported by Fsejda & Rahbek (1998) for South American parrots,

based on very refined primary sources. Further, the correlations between richness estimated across grid cells were always quite high (all $r > 0.85$, Table 2), and all coefficients were significant at $P < 0.01$, even after Dutilleul's (1993) correction of the degrees of freedom (which were reduced from 372 to 6 or 7).

As expected from the strong positive correlations among richness measures across grid cells, the four fitted environmental models were also similar (Table 3), with R^2 values ranging from 0.73 to 0.75. A minimum adequate environmental model (see Hawkins *et al*., 2003a) obtained by a forward-stepwise multiple regression revealed that the standardized partial regression coefficients were similar. In all models, the predictor with the highest standardized coefficient was AET, followed by PET. Using data from Juniper & Parr (1998) and Dunning (1987), MINT was also retained as a significant predictor, whereas RELEV was retained as significant for both Dunning (1987) and Forshaw *et al*. (1989). After fitting AET, the coefficients of determination are close to 70% and none of the other predictors in the stepwise process added sequentially more than 4% to the R^2 (Table 3). According to the previously described tests based on autocorrelation patterns of species richness, we tested the statistical significance of these partial coefficients using a more conservative t-value, with 7 degrees of freedom (Table 3).

Patterns of spatial autocorrelation in the residuals were also similar among the models, with relatively large Moran's I at the first distance class, ranging from 0.25 to 0.44, revealing that, at these small geographical distance classes, the environmental factors analysed here do not explain entirely spatial variation in

Table 2 Pairwise correlation between geographical range sizes across species (above the diagonal) and between richness across grid cells (below the diagonal), for South American parrots based on four different books (Dunning, 1987; Forshaw *et al*., 1989; del Hoyo *et al*., 1997; Juniper & Parr, 1998)

	Juniper & Parr	del Hoyo	Dunning	Forshaw
Juniper & Parr	—	0.939	0.869	0.918
del Hoyo	0.891	—	0.835	0.886
Dunning	0.885	0.879	—	0.796
Forshaw	0.898	0.851	0.906	—

Table 3 Comparisons among four books (sources) used in this study of macroecological variation in South American parrots (Dunning, 1987; Forshaw *et al*., 1989; del Hoyo *et al*., 1997; Juniper & Parr, 1998). Results include number of species from each source, the mean and skew of log-transformed geographical range size (GRS), and results of a stepwise-forward multiple regression of richness against six environmental predictors, including the standardized partial regression coefficients and associated t-tests (in parentheses), the cumulative R^2 along the stepwise process (final R^2 in bold), and the spatial autocorrelation analyses of the residuals (Moran's I in the first distance class and patch size — see text for explanation)

	Juniper & Parr	del Hoyo	Dunning	Forshaw
Number of species	118	118	103	111
Mean GRS (± SD)	1.273 ± 0.614	1.246 ± 0.634	1.406 ± 0.507	1.356 ± 0.530
Skew GRS (± SE)	−0.209 ± 0.223	−0.233 ± 0.222	−0.403 ± 0.239	−0.003 ± 0.229
Environmental model				
Partial Coefficients				
AET	0.472 (8.80**)	0.573 (11.98**)	0.404 (7.59**)	0.445 (8.78**)
PET	0.264 (5.40**)	0.319 (6.67**)	0.326 (6.37**)	0.364 (7.69**)
RELEV	—	—	−0.098 (2.92*)	−0.207 (7.05**)
MINT	0.182 (4.07**)	—	0.150 (2.99*)	—
Cumulative R^2				
AET	0.694	0.698	0.692	0.689
PET	0.724	**0.729**	0.727	0.715
RELEV	—	—	0.743	**0.750**
MINT	**0.735**	—	**0.751**	—
Residuals				
Moran's I (< 220 km)	0.335	0.347	0.252	0.441
Patch size (km)	660	660	220	660

*$P < 0.05$; **$P < 0.01$, using t-test with 7 degrees of freedom, as suggested by Dutilleul's correction for spatial autocorrelation (see text for details).
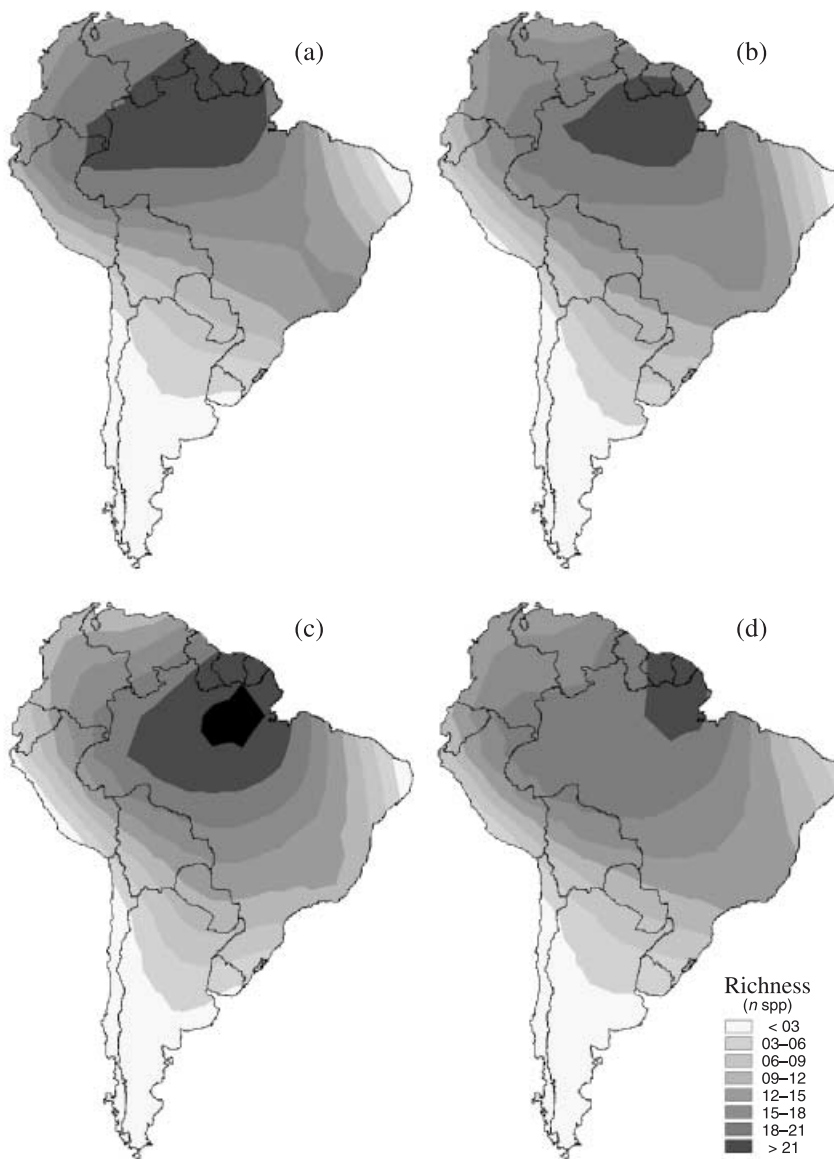
Figure 1 Interpolated patterns of species richness, using a distance-weighted least-squares algorithm (DWLS), for the South American parrots based on (a) Dunning (1987), (b) Forshaw *et al.* (1989), (c) del Hoyo *et al.* (1997), (d) Juniper & Parr (1998).

Richness
(*n* spp)

	< 03
	03–06
	06–09
	09–12
	12–15
	15–18
	18–21
	> 21

richness (see also Diniz-Filho *et al.*, 2003; Hawkins & Porter, 2003a). Whereas for data from three data sources the patch size (i.e. the distance at which autocorrelation becomes nonsignificant) in the regression residuals was around 660 km (Table 3), for data based on Dunning (1987) the patch size was the first distance class (220 km), showing that the regression model was quite effective in removing environmental effects from the data. This is expected if these data were indeed collected at a very broad scales, well described by processes acting at these scales, and with poor resolution and consequently little variation among adjacent grid cells.

CONCLUSIONS

Our results show that using different GRS measures for species in the same taxonomic group, based on different data sources, does not greatly affect macroecological patterns in this variable or in the derived richness data across the South American continent. The exception appears to be when the maps are relatively older, very small, and very coarsely resolved (i.e. Dunning, 1987), but even under these conditions results do not differ qualitatively, only quantitatively. So, similarity in macroecological patterns indicates that they are robust to variation in data sources, although there is a slight correspondence between certain similarities in these patterns (such as in the correlations among GRS estimates and in the partial coefficients of multiple regression) and differences in GRS estimates for each species (separating Dunning (1987) and perhaps Forshaw *et al.* (1989) from the other two more recent data sources).

The slight differences in the environmental models for richness patterns suggest that more detailed tests of mechanisms involved in explaining spatial patterns in species richness may be more affected by data resolution than overall macroecological patterns, such as broad-scale spatial patterns in species richness

or the geographical range size frequency distribution. However, regardless of the data sources used, AET, a variable related to productivity (or water-energy dynamics), is the most parsimonious explanation for spatial patterns in richness, as has also been found for overall bird diversity in South America (Hawkins *et al.*, 2003a; see also Hawkins *et al.*, 2003b for a recent review). Also, topographic heterogeneity (RELEV) is significant in some analyses (see Rahbek & Graves, 2001), albeit that this predictor showed more variation among data sources. Nevertheless, further investigation of this issue is desirable, because problems in model definition and multicollinearity among predictors can make regression coefficients unstable (Diniz-Filho *et al.*, 2003).

Our study leads to similar conclusions to those drawn by Blackburn *et al.* (2004) in their study of parrots of the world. They studied the effect of data resolution on macroecological patterns in this taxon, and showed that the patterns were robust to even quite dramatic differences in spatial grain size. Variation in range size estimates due to the use of different data sources is not exactly equivalent to variation in estimates due to different resolution using a single source, as it also introduces variation due to differences in taxonomy and even in range size definition. Nevertheless, both studies show that overall macroecological patterns are quite stable. This stability is also consistent with the results obtained by Hawkins & Porter (2003b), who found that analyses of the richness gradient for western Palearctic butterflies were little affected by whether the data were generated using filled-in range maps or grid-based presence/absence maps of species. However, Hawkins & Porter (2003b) varied simultaneously how the geographical range size was defined and the data sources used, in such a way that it is impossible to partition these two effects on original data.

Data quality must of course remain a concern. Ongoing mapping efforts hold the promise of better, more comprehensive and more complete data sets in the future. In the meantime, the present analysis provides some reassurance that we do currently have adequate data, at least for some taxa and regions, such that macroecological analyses can be shown to be robust to artefacts arising from variations in data quality. We suggest that further attempts to replicate macroecological analyses using alternative data sources and approaches should be encouraged.

REFERENCES

Blackburn, T.M. & Gaston, K.J. (1998) Some methodological issues in macroecology. *American Naturalist*, **151**, 68–83.

Blackburn, T.M., Jones, K.E., Cassey, P. & Losin, N. (2004) The influence of spatial resolution on macroecological patterns of range size variation: a case study using parrots (Aves: Psittaciformes) of the world. *Journal of Biogeography*, **31**, 285–294.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53–64.

Diniz-Filho, J.A.F. & Tôrres, N.M. (2002) Phylogenetic comparative methods and the geographic range size — body size relationship in New World terrestrial carnívora. *Evolutionary Ecology*, **16**, 351–367.

Dunning, J.S. (1987) *South American birds: a photographic aid to identification*. Harrowood Books, Newton Square.

Dutilleul, P. (1993) Modifying the *t* test for assessing the correlation between two spatial processes. *Biometrics*, **49**, 305–314.

Forshaw, J.T., Cooper, W.T. & Amadon, D. (1989) *Parrots of the world*. TFH Publishers, Melbourne.

Fsejda, J. & Rahbek, C. (1998) Continent-wide conservation priorities and diversification processes. *Conservation in a changing world* (ed. by G.M. Mace, A. Balmford and J.R. Ginzberg), pp. 139–160. Cambridge University Press, Cambridge.

Gaston, K.J. & Blackburn, T.M. (1999) A critique for macroecology. *Oikos*, **84**, 353–368.

Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and process in macroecology*. Blackwell, London.

Gates, S. (2002) Review of methodology of quantitative reviews using meta-analysis in ecology. *Journal of Animal Ecology*, **71**, 547–557.

Hawkins, B.A., Porter, E.E. & Diniz-Filho, J.A.F. (2003a) Productivity and history as predictors of the latitudinal diversity gradients of terrestrial birds ecology. *Ecology*, **84**, 1608–1623.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003b) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

Hawkins, B.A. & Porter, E.E. (2003a) Relative influences of current and historical factors on mammal and birds diversity patterns in deglaciated North America. *Global Ecology and Biogeography*, **12**, 475–481.

Hawkins, B.A. & Porter, E.E. (2003b) Water-energy balance and the geographic pattern of species richness of western Palearctic butterflies. *Ecological Entomology*, **28**, 678–686.

del Hoyo, J., Elliot, A. & Sargatal, J. (1997) *Handbook of the birds of the world, Vol. 4. Sandgrouse to cuckoos*. Lynx Editions, Barcelona.

Jetz, W. & Rahbek, C. (2001) Geometric constraints explain much of the species richness pattern in African birds. *Proceedings of the National Academy of Sciences USA*, **98**, 5661–5666.

Juniper, T. & Parr, M. (1998) *Parrots: a guide to the parrots of the world*. Yale University Press, London.

Legendre, P., Dale, M.R.T., Fortin, M.J., Gurevitch, J. & Hohn, M. & Myers, D. (2002) The consequences of spatial structure for design and analysis of ecological field surveys. *Ecography*, **25**, 601–615.

Legendre, P. & Legendre, L. (1998) *Numerical ecology*. Elsevier, Amsterdan.

Rahbek, C. & Graves, G.R. (2000) Detection of macro-ecological patterns in South American hummingbirds is affected by spatial scale. *Proceedings of the Royal Society of London, Series B*, **267**, 2259–2265.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences USA*, **98**, 4534–4539.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F. (2003) Spatial patterns in species richness and the geometric simulation model: a global analysis of mid-domain effect in Falconiformes. *Acta Oecologica*, **24**, 203–207.

Webb, T.J. & Gaston, K.J. (2003) On the heritability of geographic range sizes. *American Naturalist*, **161**, 553–566.

BIOSKETCHES

Pablo Vinícius Clemente Mathias is interested in ecology and biogeography of South American birds. Current projects involve the application of fractal geometry to the study of these birds.

Cláudio Veloso Mendonça is interested in ecology and biogeography of South American birds. Current projects involve using phylogenetic diversity to understand patterns in bird species richness.

Thiago Fernando Lopes Valle De Britto Rangel is interested in statistical and computational methods applied to macroecology and conservation biology. Current projects involve the development and implementation of software to simulate evolutionary dynamics of macroecological patterns.

José Alexandre Felizola Diniz-Filho is interested in statistical methods applied to macroecology and evolutionary biology. Current projects involve the application of spatial autocorrelation analysis and phylogenetic comparative methods to understand ecological processes associated with gradients in species richness and to establish priority areas for conservation.