

Biogeographic Distribution Patterns of South American Amphibians: A Regionalization Based on Cluster Analysis

Tiago da Silveira Vasconcelos^{1,*}, Miguel Ángel Rodríguez² & Bradford Alan Hawkins¹

¹ Department of Ecology & Evolutionary Biology, University of California, USA

² Department of Ecology, University of Alcalá, Alcalá de Henares, Spain

Abstract

We devise a regionalization system based on the geographical distribution of 2,265 amphibian species in South America. We used range maps of amphibians to obtain a presence/absence data grid resolved to 9310 km². Biogeographical regions were generated by submitting the dataset to *k*-means clustering combined with *v*-fold cross-validation. The boundaries of the three clusters generated by multiple runs of the analysis are congruent with broadly defined biome structure in South America: 1) the Andes, the Atacama desert, Patagonia, and subtropics including grassland in southern Brazil, Argentina and Uruguay; 2) Amazon forest; 3) Atlantic forest and the Cerrado-Caatinga-Chaco complex. A few runs further distinguished Atlantic forest and Caatinga biome from other drier, more open biomes. The variable most strongly associated with the distribution of clusters was species richness, but climate also had moderately strong explanatory power. The regionalization scheme based on clustering is less finely resolved than previous schemes generated by expert opinion and rates of endemism but provides a general overview of the biogeographic signal contained in the current distribution patterns of amphibian species.

Key words: Amphibians, Biogeographic Regions, *k*-means Clustering, Neotropics, Regionalization, Range Maps.

Introduction

An interest in the geographical distributions of species has long led researchers to divide parts of the planet (or the entire planet) into floristic kingdoms, zoogeographical regions, and ecoregions (for a brief history and examples, see Krefl & Jetz 2010). These regionalization schemes provide explicit frameworks for a range of basic and applied questions in historical and ecological biogeography, evolutionary biology, systematics and conservation (Morrone 2009). Historically, regionalizations were generated intuitively by specialists based on their knowledge of species distributions of a given group. As computers have become more powerful, and knowledge of species distributions has increased and become accessible in digital databases, biogeographical regionalizations are now being performed and/or revised for animal and plant groups in different regions of the world (e.g. Krefl & Jetz 2010; Rueda *et al.* 2010).

The Neotropics, particularly South America, have been less studied in many biological fields, including macroecology and biogeography, when compared to the northern temperate zone. The South American situation is even worse when we consider that the continent supports much of the biodiversity

of the Earth but has also suffered from some of the most severe deforestation in the world (Whitmore 1997). To attract the attention of research supporting agencies, conservation biologists and political decision-makers, Myers (2003) identified, among 29 regions throughout the world, five areas in South America distinguished by their high levels of biodiversity and endemism but also highly threatened by human activities. These regions, called hotspots of biodiversity, were determined based on species richness and endemism rates of vascular plants and vertebrates. On the other hand, hotspots may not be effective if the designation of conservation areas for a specific taxonomic group is needed, since diversity patterns need not be congruent across all groups. For instance, Atlantic forest and Cerrado are two South American hotspots mostly found in Brazil. However, if the hotspot criterion is considered for establishing amphibian conservation actions in Brazil, decision-makers would neglect important areas in the Amazon forest, whose amphibian richness (~600 species) (*sensu* Lima *et al.* 2006) is substantially higher than the Cerrado (133 species) (*sensu* Colli *et al.* 2002; Pádua *et al.* 2008).

In this study we explore how amphibians are distributed throughout South America to generate a map of amphibian diversity that focuses on the composition of regional faunas within the continent rather than species richness *per se*.

*Send correspondence to: Tiago da Silveira Vasconcelos
Department of Ecology & Evolutionary Biology,
University of California, Irvine, CA 92697, USA
E-mail: zoologia@ig.com.br

Amphibians are known to have their current distributions strongly influenced by both environmental variables and evolutionary history (e.g. Hawkins *et al.* 2003; Buckley & Jetz 2007), and due to their global decline are the object of much conservation concern. Our goal is to identify biogeographic regions based on the geographical distribution of species using a quantitative methodology. Specifically, we determine the number and positions of these regions within South America using a quantitative clustering method (*k*-means clustering with *v*-fold cross validation). Further, following Rueda *et al.* (2010), we use general linear modeling and variation partitioning to examine abiotic and biotic predictors of the distribution of the faunal regions identified by our clustering method.

Material and Methods

Species distribution maps

Of the 2,461 species of amphibians currently recorded in South America (AmphibiaWeb 2010), we obtained digitalized species distribution (range) maps for 2,265, available at the International Union for Conservation of Nature (IUCN) portal (<http://www.iucnredlist.org/technical-documents/spatial-data>). The range maps were overlaid in a continental grid of 9310 km² grain in a Behrmann projection to generate a presence/absence matrix, which was then submitted to cluster analysis.

Cluster analysis

The data matrix was analyzed using *k*-means clustering (Bishop 2005). The classical *k*-means clustering algorithm requires the number of clusters (*k*) to be established in advance, which utilizes a subset of *k* random initialization cells that are treated as the initial cluster centres, and then proceeds as a two-step iterative procedure in which cluster centres and clusters are successively recalculated. The process begins an iterative adjustment of cells so that each of them is assigned to one of the *k* clusters, and each of the *k* clusters is the mean of its assigned cells (Bishop 2005). First, each cell is assigned to its nearest cluster centre in terms of species compositional distance, herein considered by the Hellinger distances (Legendre & Gallagher 2001). Then, each cluster center is updated by making it equal to the mean of the cells assigned to it. The process is repeated (we used 50 iterations) so that the clusters and cluster centres change in each replicate, converging to a locally optimal position in the data space.

To obtain the optimal number of clusters based on species composition without regard to the spatial proximity of grids, *k*-means clustering was combined with *v*-fold cross-validation (see Rueda *et al.* 2010). In this approach, the range of potential *k* groups is established in advance (we considered from two to 25 clusters), but the algorithm determines the 'best' number of clusters within this range. To check the

robustness of the results we ran the analysis multiple times (ca. 10). The *k*-means clustering and *v*-fold cross-validation were performed using Statistica 8.0 (StatSoft 2007).

Correlates of cluster patterns

The distribution of clusters for any taxonomic group will be to some extent a reflection of species richness patterns, since both are manifestations of species distributions. However, other biotic and abiotic factors would also be expected to play a role via their influences on vegetation structure (e.g., Rueda *et al.* 2010), or as barriers to dispersal. To explore these relationships we gridded the data in a 55 × 55 km grid system, to which we added the cluster results, three climatic variables (annual temperature, annual precipitation, and annual evapotranspiration [AET]), and topography (range of elevation in the cell). The temperature and precipitation data were extracted from the 10 arc-minute database in WorldClim (available at <http://www.worldclim.org/>), and AET data were from the 30 arc-minute GNV183 database (available at <http://www.grid.unep.ch/data/summary.php?dataid>). The elevation data were from the 30 arc-second gtopo30 digital elevation model (available at http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info).

We generated GLM models using a generalized multinomial logit-link to identify the amounts of deviance in the cluster assignments that can be explained statistically by our predictor variables. We first generated a single-factor model for each predictor, followed by a three-factor model in which temperature, precipitation and AET were combined to generate an overall climate model. Finally, we used variation partitioning (see Rueda *et al.* 2010) to partition explained deviances into independent and shared components of climate, species richness and topography. The analyses were conducted on both potential cluster solutions (see below).

Results and Discussion

The *k*-means clustering with *v*-fold cross-validation identified three compositionally distinct regions in South America in most runs. The boundaries of the clusters are broadly congruent with aspects of biome structure in South America (Figure 1). One cluster encompasses high altitude and/or colder regions in the Andes, the Atacama Desert, the savannah of Venezuela, Patagonia, and subtropical areas including the grasslands of southern Brazil, Argentina and Uruguay (the Pampas biome *sensu* IBGE 2004) (Figure 1a). A second cluster comprises the Amazon forest, while the third includes areas of Atlantic forest and the Cerrado-Caatinga-Chaco complex (*sensu* Duellman 1999), which is predominantly composed of savanna-like vegetation, but also with semi-arid vegetation (Caatinga in northeastern Brazil, *sensu* Veloso *et al.* 1991) and semideciduous forest areas in the southeastern and southern Brazil (*sensu* Veloso *et al.* 1991) (Figure 1a). The four-cluster solution, which occurred in two runs, further subdivided the third cluster into two pieces, distinguishing Chaco, Cerrado, and

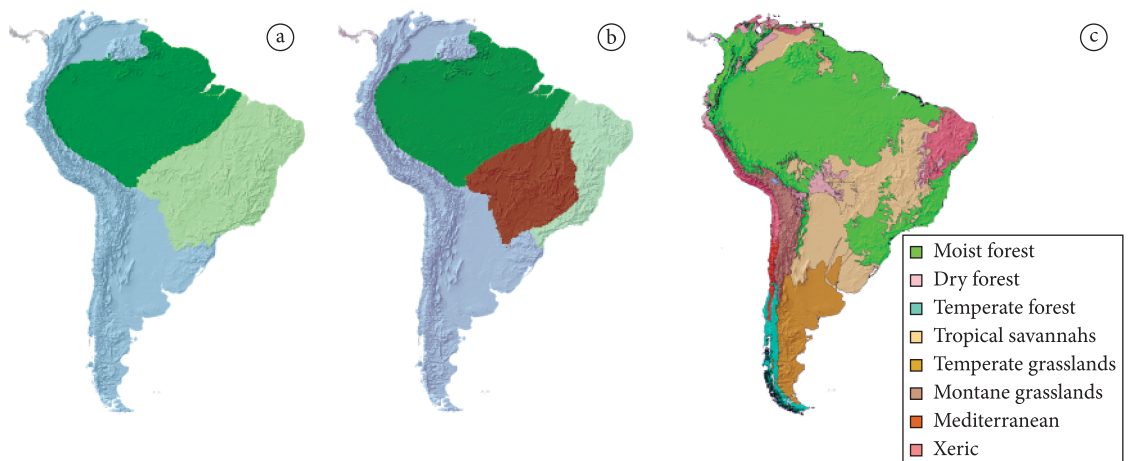


Figure 1. Biogeographic regions based on South American amphibians generated by k -means clustering with v -fold cross-validation; a) the three-cluster solution found in most runs; and b) a four-cluster solution found in two runs only. c) Major South American biomes, modified slightly from World Wildlife Fund designations.

most of tropical dry forest to the west of coastal Atlantic forest and Caatinga (Figure 2b).

The strongest correlate of cluster configuration was species richness, explaining approximately half of the deviance in the 3-cluster solution and slightly less in the 4-cluster solution (Table 1). Among the climatic variables, annual temperature was the strongest predictor; however, combining all three climatic variables explained substantially more deviance than the variables considered in isolation. The general climate model explained slightly less deviance than species richness (Table 1). Topography explained substantially less deviance in both the 3- and 4-cluster solutions.

Partitioning the deviances into independent and overlapping components also identified species richness as having the strongest relationship with cluster assignments (Figure 2). However, substantial covariation between richness and climate makes it difficult to determine whether richness itself has the strongest influence on the cluster assignments or if much of the relationship is actually due to climate. Indeed, a cell based multiple regression of the climatic variables and species richness across South America identified a fairly strong link between the richness pattern and climatic gradients ($R^2 = 0.627$). Irrespective, the relationships between species richness and cluster configurations in South America appear stronger than that found for a regionalization scheme for European amphibians generated using an identical approach (Rueda *et al.* 2010). We are unsure why this should be, but it may be related to differences in range sizes of amphibians within South America and Europe or to the very strong richness gradient in South America.

In a previous amphibian regionalization in South America, Duellman (1999) established 12 biogeographic subregions based on morphoclimatic domains (*sensu* Ab'Saber 1977), adapting geographical boundaries or combining multiple domains based on amphibian distributions and endemism

at each region. It is important to emphasize that the regionalization method we use, k -means clustering, which is based on differences in species composition, is not designed to generate biotic regions based on high diversity or the distributions of endemic taxa, but rather based on differences in species assemblages, and some clusters may include endemics and/or high richness, but some may not (Procheş 2005). Therefore, the biogeographic subregions based on the amphibian distributions defined by Duellman (1999) are partially a more finely resolved version of the general patterns recorded here. For instance, the blue cluster recorded here (Figure 1a) includes at least seven regions (*e.g.* Andes, Atacama Desert, Caribbean Coastal Forest, Patagonia, among others) considered by Duellman (1999). These areas are known to have either many species or a high proportion of endemic species, mainly in the Andes, Atacama and Patagonia (Duellman 1999).

The Amazon forest cluster (Figure 1) is reasonably congruent with the Amazonia-Guiana region defined by Duellman (1999), but he considered the northern boundary further north than the region delimited by our analysis. The light-green cluster (Figure 1a) represents two regions defined by Duellman (1999), the Atlantic forest and the Cerrado-Caatinga-Chaco complex. Although the Atlantic forest represents one of the richest areas in the Neotropics with a high level of amphibian endemism (Duellman 1999; Haddad & Prado 2005), we found that this region, together with the Cerrado-Caatinga-Chaco complex, was combined a single biogeographic unit based on overall distribution patterns of species. This probably reflects the large number of species shared between these areas, since amphibian species in semideciduous forest (which is found in transition zones with Cerrado and Chaco areas) are a mix of typical Cerrado and Atlantic forest species as well as widely distributed species (Santos *et al.* 2009).

Comparison of the regions we found for amphibians with earlier regionalizations for mammals shows that amphibian patterns agree with mammals only for colder/high-altitude areas (the central-southern Andes, the Pampas and Patagonia) (Procheş 2005; Kreft & Jetz 2010). Mammals (Kreft & Jetz 2010), and specifically bats (Procheş 2005), have distinct clusters in high-altitude and/or colder regions in South America, but no clusters were found in

the northern Andes nor in central tropical South America. These studies are similar to the biogeographic patterns considered by Morrone (2002), which separated South America in two major biogeographic subregions: the Andean and the Neotropical. On the other hand, methodological differences and the scale of analysis might limit general comparisons of the numbers and positions of clusters, since Procheş (2005) and Kreft & Jetz (2010) conducted their biogeographic regionalization at a global scale and used different clustering methods (*e.g.* UPGMA). Further, climatic variables are known to influence species distributions (*e.g.* Hawkins *et al.* 2003; Buckley & Jetz 2007), and high-altitude areas (the Andes) and the subtropics (*e.g.* grasslands, deserts, among others) represent areas with the greatest spatial climatic turnover rates in South America (Buckley & Jetz 2008). Consequently, these areas also have the highest rates of species turnover of both birds and amphibians in South America (Buckley & Jetz 2008), thereby generating a separate biogeographic subregion in at least some South American vertebrates, including amphibians (present study) and mammals (Procheş 2005; Kreft & Jetz 2010).

Of the major biomes and/or morphoclimatic domains found in South America (*e.g.* Ab'Saber 1977; Veloso *et al.* 1991; Morrone 2002; IBGE 2004), only the Amazon forest can be recognized as a biogeographic unit for amphibians, while the remaining biomes (*e.g.* temperate forest, savanna, grasslands, desert, among others, Figure 1c) were always combined in various combinations to form biogeographic subregions for the amphibians. Even when the marginally supported fourth cluster was distinguished, the new subregion combines Cerrado and Chaco (Figure 1b), which are characterized by seasonal climates with open formations or dry, semideciduous forest. These areas are strongly influenced by their seasonal climates (Pennington *et al.* 2000), which in turn may have led to similar amphibian species composition throughout this area (see further discussion in Santos *et al.* 2009; Vasconcelos *et al.* 2010). Interestingly, the fourth cluster (brown in Figure 1b) does not include the Caatinga, which remains associated with Atlantic forest despite its semi-arid climate and vegetation physiognomy being more similar to Cerrado-Chaco than to moist forest. Indeed, Caatinga had been previously considered a sub-region of the Neotropics (Morrone 2002), as well as an amphibian biogeographic unit (Duellman 1999) together with other open-areas formations (Cerrado and Chaco). However, its closer association with Atlantic biome in the present study might be related to the large number of widespread amphibians that also occur in

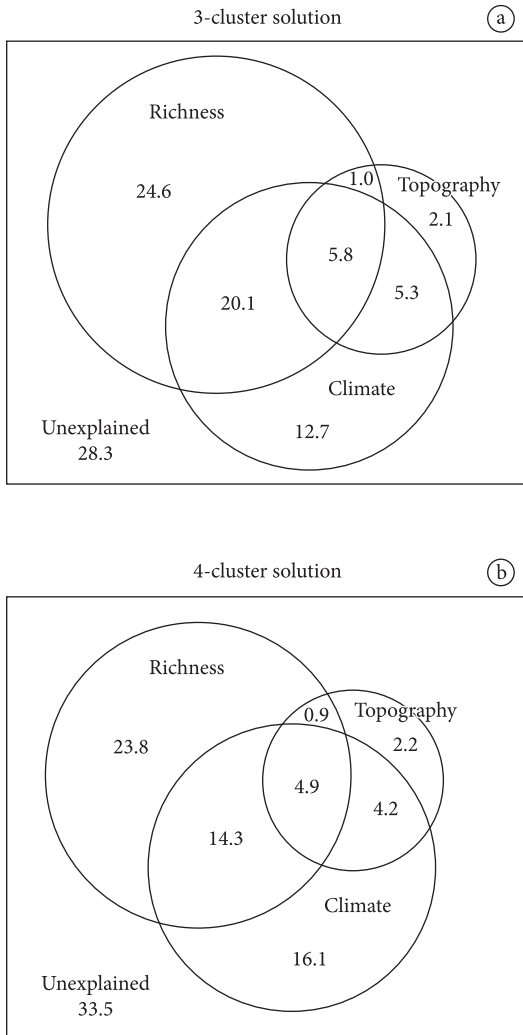


Figure 2. Deviance partitioning analysis for the a) 3-cluster; and b) 4-cluster solutions, representing the deviance in cluster configurations explained by climate (mean annual temperature + annual precipitation + annual evapotranspiration), topography (range in elevation) and amphibian species richness.

Table 1. Percentage of deviance in *k*-means solutions as explained by mean annual temperature [T], annual precipitation [P], annual actual evapotranspiration [AET], these three climatic variables combined, topography (range in elevation), and species richness.

K-means solution	T	P	AET	Climate (T+P+AET)	Topography	Richness
3-clusters	32.7	24.9	22.2	44.0	14.3	51.6
4-clusters	28.3	22.4	19.3	39.6	12.3	43.9

Cerrado, Amazon, and Atlantic forest, although Caatinga contains some endemic amphibians as well (Loebmann & Haddad 2010).

Finally, from a conservation point of view, we found that only two regions (Figure 1a, the blue and light-green clusters) encompass places that had been attracted special attention from conservation biologists over the last decade. Myers (2003) recognized five ecoregions in South America that represent hyper-rich places with high rates of endemism but have been simultaneously hardly threatened by human activities. Among these hotspots, our blue cluster contains three of them (Chilean Valdivian Forests, Tropical Andes, and Tumbes-Chocó-Magdalena), whose amphibian fauna has the highest richness and/or endemism rates in South America (Duellman 1999). The other two hotspots, Atlantic forest and Cerrado, are located in a single cluster. Although both of them are considered hotspots, amphibian diversity and endemism are more extensive in Atlantic forest (Colli *et al.* 2002; Haddad & Prado 2005), which is known to have one of the highest richness and amphibian endemism rates in South America (Duellman 1999; Haddad & Prado 2005). On the other hand, the Amazon forest would not be considered a hotspot according to Myers (2003), but the present study identifies this region as a distinctive biogeographic amphibian unit in South America, indicating that the extension of Amazon habitat structure was basically conserved with respect to amphibian distributions. On the other hand, other taxonomic groups may have different patterns, and future work to generate a consensus of biogeographic regions across multiple biological groups is needed to generalize the usual metrics (*e.g.* species richness, endemism rates, and percentage of endangered species) used to generate priority areas in conservation plans.

Acknowledgements

TSV was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Brazil: grant nº 2009/17195-3) and MAR by the Ministry of Science and Innovation of Spain (grant: CGL2010-22119).

References

- Ab'Saber NA, 1977. Os domínios morfoclimáticos na América do Sul. *Geomorfologia*, 52: 1-159.
- AmphibiaWeb, 2010. *Information on amphibian biology and conservation*. Berkeley, California: AmphibiaWeb. Available from: <www.amphibiaweb.org/>.
- Bishop CM, 2005. *Neural networks for pattern recognition*. New York: Oxford University Press.
- Buckley LB & Jetz W, 2007. Environmental and historical constraints on global patterns of amphibian richness. *Proceedings of the Royal Society B*, 274:1167-1173. <http://dx.doi.org/10.1098/rspb.2006.0436>
- Buckley LB & Jetz W, 2008. Linking global turnover of species and environment. *Proceedings of the National Academy of Sciences of the United States of America*, 105:17836-17841. <http://dx.doi.org/10.1073/pnas.0803524105>
- Colli GR., Bastos RP & Araújo AFB, 2002. The character and dynamics of the Cerrado herpetofauna. In: Oliveira PS & Marques RJ (eds.). *The Cerrados of Brazil: ecology and natural history of a neotropical savanna*. New York: Columbia University Press. p. 223-241.
- Duellman WE, 1999. Distribution Patterns of Amphibians in South America. In: Duellman WE (ed.). *Patterns of Distribution of Amphibians*. Baltimore and London: The Johns Hopkins University Press. p. 255-327.
- Haddad CFB & Prado CPA, 2005. Reproductive modes in frogs and their unexpected diversity in the Atlantic Forest of Brazil. *BioScience*, 55:207-217. [http://dx.doi.org/10.1641/0006-3568\(2005\)055\[0207:RMIFAT\]2.0.CO;2](http://dx.doi.org/10.1641/0006-3568(2005)055[0207:RMIFAT]2.0.CO;2)
- Hawkins BA *et al.*, 2003. Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84:3105-3117. <http://dx.doi.org/10.1890/03-8006>
- Instituto Brasileiro de Geografia e Estatística - IBGE, 2004. *Mapa de Biomas do Brasil: primeira aproximação*. Rio de Janeiro: IBGE. Available form: <http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=169>.
- Kreft H & Jetz W, 2010. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37:2029-2053. <http://dx.doi.org/10.1111/j.1365-2699.2010.02375.x>
- Legendre P & Gallagher ED, 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129:271 -280. <http://dx.doi.org/10.1007/s004420100716>
- Lima AP *et al.*, 2006. *Guia de sapos da Reserva Adolpho Ducke, Amazônia Central*. Manaus: Attema Design Editorial.
- Loebmann D & Haddad CFB, 2010. Amphibians and reptiles from a highly diverse area of the Caatinga domain: composition and conservation implications. *Biota Neotropica*, 10:227-256. <http://dx.doi.org/10.1590/S1676-06032010000300026>
- Morrone JJ, 2002. Presentación sintética de un nuevo esquema biogeográfico de América Latina y el Caribe. *Monografías Tercer Milenio*, 2:267-275.
- Morrone JJ, 2009. *Evolutionary biogeography: an integrative approach with case studies*. New York: Columbia University Press.
- Myers N, 2003. Biodiversity hotspots revisited. *BioScience*, 53:916-917. [http://dx.doi.org/10.1641/0006-3568\(2003\)053\[0916:BHR\]2.0.CO;2](http://dx.doi.org/10.1641/0006-3568(2003)053[0916:BHR]2.0.CO;2)
- Pádua GCC, Pinto MP & Diniz-Filho JAF, 2008. Escolha de áreas prioritárias de conservação de anfíbios anuros do Cerrado através de um modelo de populações centrais-periféricas. *Iheringia (Série Zoologia)*, 98:200-204.
- Pennington RT, Prado DE & Pendry CA, 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, 27:261-273. <http://dx.doi.org/10.1046/j.1365-2699.2000.00397.x>
- Procheş Ş, 2005. The world's biogeographical regions: cluster analysis based on bat distributions. *Journal of Biogeography*, 32:607-614. <http://dx.doi.org/10.1111/j.1365-2699.2004.01186.x>

- Rueda M, Rodríguez MÁ & Hawkins BA, 2010. Towards a biogeographic regionalization of the European biota. *Journal of Biogeography*, 37:2067-2076. <http://dx.doi.org/10.1111/j.1365-2699.2010.02388.x>
- Santos TG *et al.*, 2009. Anurans of a seasonally dry tropical forest: the Morro do Diabo State Park, São Paulo State, Brazil. *Journal of Natural History*, 43:973-993.
- StatSoft Inc., 2007. *STATISTICA (data analysis software system)*. Version 8.0. Available from: <www.statsoft.com>.
- Vasconcelos TS *et al.*, 2010. Climatic variables and altitude as predictors of anuran species richness and number of reproductive modes in Brazil. *Journal of Tropical Ecology*, 26:423-432. <http://dx.doi.org/10.1017/S0266467410000167>
- Veloso HP, Rangel-Filho AL & Lima JCA, 1991. *Classificação da vegetação brasileira adaptada a um sistema universal*. Rio de Janeiro (Brazil) (1st ed.): IBGE.
- Whitmore TC, 1997. Tropical forest disturbance, disappearance, and species loss. In: Laurance WF & Bierregaard Jr RO (eds.). *Tropical forest remnants: ecology, management, and conservation of fragmented communities*. Chicago: The University of Chicago Press. p. 3-12.

Received: May 2011

First Decision: May 2011

Accepted: June 2011