Insect Conservation and Diversity

Insect Conservation and Diversity (2017) **10,** 415–424 doi: 10.1111/icad.12240

Stacked species distribution and macroecological models provide incongruent predictions of species richness for Drosophilidae in the Brazilian savanna

RENATA ALVES DA MATA,[1] ROSANA TIDON,[1,2] GUILHERME DE OLIVEIRA,[3] BRUNO VILELA,[4] JOSÉ ALEXANDRE FELIZOLA DINIZ-FILHO,[4] THIAGO FERNANDO RANGEL[4] and LEVI CARINA TERRIBILE[5] [1]Programa de Pós-Graduação em Ecologia, Universidade de Brasília, UnB, Brasília, Brazil, [2]Departamento de Genética e Morfologia, Universidade de Brasília, UnB, Brasília, Brazil, [3]Laboratório de Biogeografia da Conservação, Universidade Federal do Recôncavo da Bahia, UFRB, Cruz das Almas, Brazil, [4]Departamento de Ecologia, Universidade Federal de Goiás, UFG, Goiania, Brazil and [5]Laboratório de Macroecologia, Universidade Federal de Goiás, UFG, Regional Jataí, Jataí, Brazil

Abstract. 1. We tested the adequacy of two richness-modelling approaches within the 'spatially explicit species assemblage modelling' (SESAM) framework for drosophilid flies in a tropical biome.

2. The pattern of drosophilid species richness throughout the Brazilian savanna was investigated by comparing richness estimates from macroecological models (MEM) and stacked species distribution models (S-SDM). We used occurrence records for macroecological modelling and to generate geographic ranges by modelling species' niches, which were stacked to generate SDM richness. Richness predictions were compared between models and with empirical data from well-sampled areas.

3. The spatial variation in drosophilid richness for both estimates revealed more species in the central and south-eastern regions of the biome. Nonetheless, MEM generated a more fragmented pattern than S-SDM, with scattered patches of high richness. S-SDM produced richness estimates nearer to the empirical values than MEM, which in turn strongly underestimated richness.

4. The correlation between S-SDM and observed richness suggests that climate is the major (indirect) driver of drosophilid richness in the Brazilian savanna. Richness estimates based on macroecological modelling are, however, almost certainly affected by inventory incompleteness and sampling bias. We emphasise that S-SDM can be a valuable approach to explore species richness patterns in poorly sampled regions.

Key words. Biodiversity, Cerrado, *Drosophila*, ecological niche modelling, macroecological constraints, species distribution, species richness.

Introduction

Understanding the factors that promote geographic variation in species richness is needed to preserve species in the face of the current biodiversity crisis (Whittaker *et al.*, 2005). Several decades of research and the recent availability of large datasets have provided substantial evidence of the role of climate in shaping current patterns of species diversity (e.g. Currie, 1991; Hawkins *et al.*, 2003; Wiens & Donoghue, 2004), although most of the accumulated knowledge is taxonomically and geographically biased toward vertebrates or concentrated in temperate

Correspondence: Levi Carina Terribile, Laboratório de Macroecologia, Universidade Federal de Goiás, Regional Jataí, BR 364, km 193, CEP: 75801-615, Jataí-GO, Brazil. E-mail: carina@ufg.br

regions (Beck *et al*., 2012). Many groups of invertebrates, particularly insects, remain vastly underrepresented in macroecological studies (Diniz-Filho *et al*., 2010), largely due to the incompleteness of species inventories necessary to map species richness (Ballesteros-Mejia *et al*., 2013).

One strategy to overcome this limitation is empirically modelling the number of species from environmental predictors to estimate richness. Two main approaches have been recently applied to model the geographic variation in species richness: the macroecological approach (also called top-down approach), and the stacked species distribution modelling approach (or bottom-up; Gotelli *et al*., 2009; Guisan & Rahbek, 2011; Diniz-Filho *et al*., 2012, 2013). The macroecological approach (MEM) considers richness as an emergent property of species assemblages. In this, the number of co-occurring species in a geographic unit is determined by environmental factors, predominantly energy, water, and environmental heterogeneity (e.g. Hawkins *et al*., 2003; Currie *et al*., 2004), regardless of the identity of species that compose the community (Boucher-Lalonde *et al*., 2014).

The stacked species distribution modelling approach (S-SDM) is more recent and claims that the number of species is determined by the climatic/environmental constraints on geographical ranges of individual species, thus determining which species can co-exist in an area (Guisan & Thuiller, 2005). This approach rests on the assumption that species respond individualistically to environmental variation, and richness can be predicted from the overlapping of species ranges. S- SDM is derived from the increasing use of species-niche modelling (based on niche requirements) to determine potential distributions for conservation purposes in the face of global climate and land use changes (e.g. Lemes & Loyola, 2013).

Although the approaches differ in the hierarchical level at which factors are expected to generate spatial patterns of species richness, they converge on the supposition that the environment, either through limits on the carrying capacity of the ecosystem or through individual niche requirements, is the primary determinant of the number of species across a region. Thus, it could be expected that processes driving organism distributions in a lower hierarchical level (species geographical ranges) can be directly associated with environmental factors promoting overlap of distributions at higher levels (species richness) (Terribile *et al*., 2009; Guisan & Rahbek, 2011; Diniz-Filho *et al*., 2012, 2013). In this case, patterns of richness predicted by MEM and those obtained by S-SDM would match perfectly (Guisan & Rahbek, 2011; Diniz-Filho *et al*., 2012). On the other hand, differences between the spatial patterns of species richness generated by MEM and S-SDM are expected in some situations, for example, if species are not in equilibrium with the current climate due to evolutionary and historical contingencies (e.g. dispersal barriers) and ecological factors (e.g. competition and limited dispersal abilities; Wiens & Donoghue, 2004; Araújo & Pearson, 2005). Indeed, due to the difficulty of implementing historical constraints and biotic interactions in the process of modelling the range of species, the S-SDM approach frequently overpredicts the number of species in a geographic unit, whereas MEM tends to predict values close to the observed species richness (Algar *et al*., 2009; Pineda & Lobo, 2009; Guisan & Rahbek, 2011). Guisan and Rahbek (2011) formally proposed a unifying approach, SESAM, a framework for 'spatially explicit species assemblage modelling', in which the mismatch between MEM and S-SDM richness predictions, rather than being a problem, should be compared and integrated.

It is noteworthy that, to apply the MEM approach, the geographical ranges and consequently, the species composition of the community should be known independently of the modelling itself, or should be estimated based on sampling from local assemblages using, for example, rarefaction (Diniz-Filho *et al*., 2013). On the other hand, given that in S-SDM the distributions of the species are modelled individually, richness can be estimated from incomplete information about species distributions and community composition (Algar *et al*., 2009; Guisan & Rahbek, 2011). It is expected, therefore, that incomplete species inventories and inaccurate knowledge of species distribution will result in more biased richness estimation in MEM than in S-SDM. In such cases, differences between MEM and S-SDM predictions should be interpreted cautiously, since they do not reflect processes governing species richness at different scales, but rather, an effect of sampling and biased knowledge.

Drosophilidae is a diverse family of Diptera that includes over 4000 species distributed globally (Bächli, 2016). These flies breed in a variety of decaying plant material, playing an important role in the decomposition of dead organic matter and nutrient release (Stökl *et al*., 2010), and therefore providing vital ecological services to ecosystems. Despite their importance, data describing richness patterns of drosophilids at large scales remain rare (but see Parsons & Bock, 1979). The Brazilian savanna, locally known as the Cerrado biome, is the second largest South American biome and one of the most diverse savannas in the world. Drosophilids of the Cerrado have been systematically studied for 15 years, thereby generating much information about these insects' communities (Tidon, 2006; Mata & Tidon, 2013; Roque *et al*., 2013).

In this paper, we use an extensive data compilation of drosophilid occurrence records throughout the Neotropical region, together with a set of climatic predictors, to estimate species richness from MEM, which is compared with S-SDM richness generated by stacking geographic ranges from modelling the niche of species. We also compare absolute values of species richness predicted from MEM and S-SDM with empirical values obtained from exhaustive field sampling in a set of sites in the Cerrado. We expected that, if the inventoried data currently available for drosophilids in the biome is a representative and unbiased sample of the diversity of this group, the predictions from MEM and S-SDM would match the empirical

(i.e. observed) richness from well-sampled areas. Also, given the above mentioned theoretical assumptions, if climate is the major driver of species richness, the richness values predicted by MEM should not differ from S-SDM predictions, again assuming that the field samples are reasonably accurate and unbiased. If the inventoried data, however, are incomplete or biased through the biome, we could expect for incongruences between model predictions and observed richness, especially for MEM predictions.

Materials and methods

Species and environmental data

We obtained occurrence records for the 127 drosophilid species occurring in the Cerrado biome from the online database Taxodros (The database on Taxonomy of Drosophilidae, available at http://www.taxodros.uzh.ch/) and from field samplings (e.g. Chaves & Tidon, 2008; Mata *et al*., 2008; Roque & Tidon, 2008; Roque *et al*., 2009; Valadão *et al*., 2010). All records were examined for synonymies and nomenclature errors, and the spatial locations of records were checked for correctness before modelling. Some of these species are distributed globally, whereas others are restricted to the Neotropics (Table S1). Thus, although our analyses were focused on the Cerrado, we delimitated the Neotropical region as the G (geographic) space within which niche models were built (see Peterson *et al*., 2001) to obtain the geographic distributions. All subsequent analyses (see below) were restricted to the Cerrado biome. We mapped the occurrence records over 6818 grid cells with a 0.5° grain encompassing the Neotropical region. The number of presence records in the grid cells for each species ranged from 1 to 293 (Fig. 1; see also Table S1). Species with fewer than 10 records (a total of 28 species) were excluded from the analyses (see below).

Climatic predictors used either in MEM or SDM were obtained from ecoClimate database (http://ecoclimate.org, Lima-Ribeiro *et al*., 2015), and downscaled to the 0.5° grain of our grid. Given the variety of climatic simulations (as well as SDM methods, see below) currently available, we opted to combine output predictions from

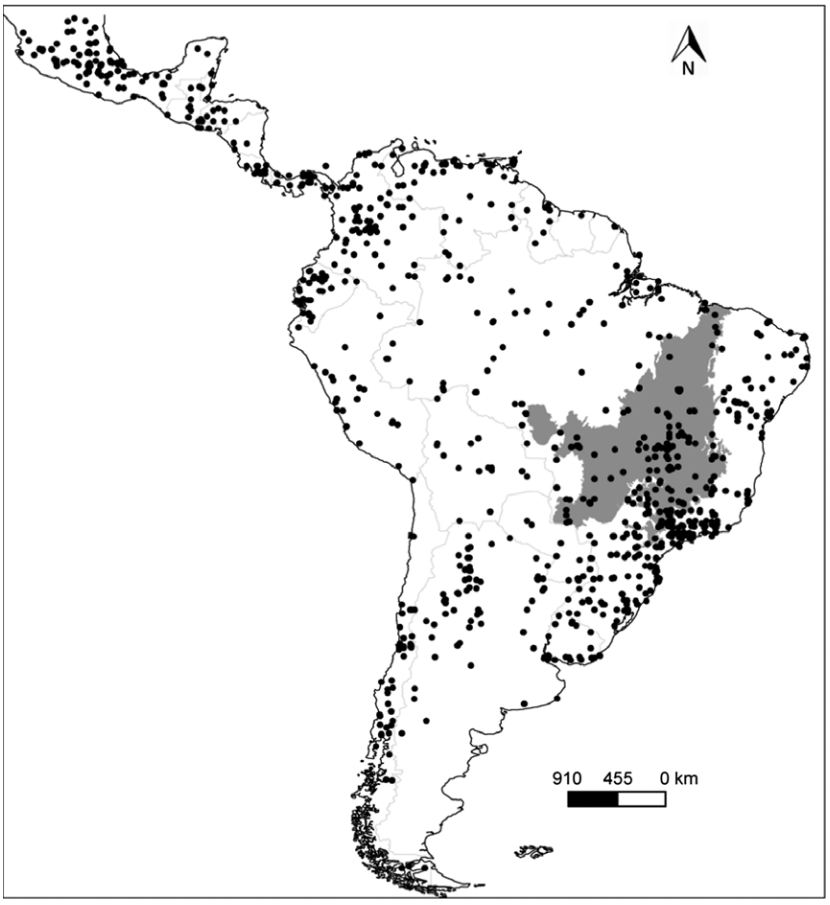

Fig. 1. Occurrence records of 127 Drosophilid species across the Neotropical region used for niche modelling and species distribution estimation. Richness analyses were performed for the Cerrado biome (grey area).

multiple climate models and SDM methods, following the 'ensemble approach' of Araújo & New, 2007 (see more details in the S-SDM richness section). We used monthly simulations of four variables (i.e. annual precipitation and mean, maximum, and minimum temperatures) from four coupled atmosphere-ocean general circulation models (AOGCM): CCSM4, GISS-E2-R, MIROC-ESM, and MRI-CGCM3 (see Table S2), to calculate 19 bioclimatic variables according to Hijmans *et al*. (2005). To minimise collinearity among bioclimatic variables, we used a varimax-rotated factor analysis to select the set of variables with highest loadings in the first five factors. Thus, the following five orthogonal variables were chosen: mean annual temperature, annual temperature range, precipitation of the wettest month, precipitation of the driest month, and precipitation of the warmest quarter. The factor analysis was performed based only on CCSM bioclimatic variables, as these variables were highly correlated across all AOGCMs (i.e. the loadings across all AOGCMs provided quite similar results, with shifts usually below 0.05 of difference in the highest loadings).

Richness estimation and macroecological models (MEM)

Biased sampling effort across the region studied poses problems for measuring the relationship between richness and environmental variables, either by creating spurious correlations (e.g. higher richness in warmer areas because they were more often visited) or masking the actual correlation (e.g. higher richness in more accessible areas). To remove the effect of sampling bias on the species richness in each grid cell, we used a modified version of the traditional individual-based rarefaction curve to estimate species richness for any given sampling effort (Gotelli & Colwell, 2001). In our modified version, all steps in the algorithm to produce a rarefaction curve are the same as the traditional method, with the exception that we used species occurrence records instead of species abundance data. Following this method, we cumulatively sampled without replacement all occurrence records in each cell to generate a species accumulation curve per cell. We repeated the procedure 1000 times to estimate a mean value of species richness per cell covering all the possible numbers of occurrence records (i.e. an occurrence record-based rarefaction curve for each cell).

A common approach to select the sampling effort (here the number of occurrence records per cell) in a rarefaction curve to estimate richness is by limiting all the values by the minimum sampling effort. By doing this, however, cells with a very small sampling effort (occurrence records ranging from 0 to 63) would force a poor richness estimation across all of the cells. Alternatively, choosing a value that maximises sampling completeness could result in very few cells to generate the MEM. So, to decide which number of occurrence records should be used for the comparison with the S-SDM predictions, we first applied an ordinary least squares regression (OLS) to relate the climatic variables and the estimated richness values over all possible numbers of occurrence records (a total of 63 models). Then, each model was used to predict the species richness across all cells in the region studied. Eventually, we considered Pearson's correlations among the predicted richness from the 63 MEMs and the S-SDM estimate. We selected the MEM that maximised the correlation between the S-SDM and MEM predictions. This model included 23 grid cells with 19 occurrence records each (Pearson's $r = 0.4$; see Fig. S1).

Niche models and S-SDM richness

There has been recent discussion of the designations of correlative approaches that explore the relationship between species occurrences and climatic predictors (Araújo & Peterson, 2012). In this paper, we used SDM (species distribution model) to refer to the output of niche modelling processes, following the duality between species-niche and geographic distributions according to Colwell and Rangel (2009). We followed the recommendation by Araújo and New (2007) to generate consensus predictions by combining outputs from different niche modelling methods, weighted by individual model accuracy (the 'ensemble forecast' approach, *sensu* Araújo & New, 2007; see also Rangel & Loyola, 2012). We used 13 ecological niche models based on a wide range of mathematical approaches, including distance and multivariate statistics (i.e. BIOCLIM, Euclidian, Gower and Mahalanobis distances, and ENFA), machine-learning (GARP, MAXENT, random forest, and artificial neural networks), and regression methods (GLM, GAM, FDA, MARS) to generate maps of potential geographic distributions for each species (see Table S3). For a general description of niche models, see Franklin (2009) and Peterson *et al*. (2011).

Prior to modelling, we excluded 28 species with fewer than 10 occurrence records. For each of the remaining 99 species, we randomly divided presence records (and pseudo-absences) into 75% for calibration and 25% for evaluation, and repeated this process 50 times. Pseudo-absences were randomly selected from the background region (excluding cells with occurrence records) with the same proportion of species records and were used both in presence-absence and presence-only ecological niche models. In the presence-only models, the pseudo-absences were used as background. Each of the 50 results for each species were converted into binary distribution maps based on thresholds established by the area under the receiver operating characteristic (ROC) curve, known as the AUC (area under the curve, Fielding & Bell, 1997). The frequency of presence of a given species in each cell in these 50 results was used to generate the species climatic suitability, varying from zero (the species was recorded as absent in the cell in all 50 results) to one (the species was recorded as present in all 50 results). Each model resulting from the 50 randomisations was weighted by true skill statistics (Allouche *et al*., 2006) to the final climatic

suitability. To generate range maps, the final climatic suitability from the combination of ecological niche models and AOGCM was converted to a consensus (binary) distribution using a threshold of 0.5. Finally, these range maps were stacked and the number of species in each grid cell was counted as the measure of richness. All models were generated using the computational platform BioEnsembles (Diniz-Filho *et al*., 2009).

Model comparisons

We evaluated the precision in richness estimates through a correlation (Pearson's r) between S-SDM and MEM richness (Algar *et al*., 2009), and between these and observed richness. Observed richness was derived from counting the number of species recorded in the grid cells based on occurrence records from online databases and field sampling (a total of 129 cells). In order to explore if S-SDM over-estimates richness values, we selected the 16 richest cells (mean richness = 44, minimum and maximum richness = 26 and 92) and compared observed richness to the values predicted by S-SDM. To select these cells, we truncated at a minimum richness of 25 species based on empirical knowledge from several years of monitoring and data collection by two of us, RAM and RT (see Fig. S2). Additional information from five thoroughly sampled areas in the Cerrado (derived from around 18 years of fieldwork, since 1997) was also used to compare observed richness with predictions (Fig. S2).

Richness estimates by rarefaction and comparative analyses were performed using the packages *raster* (Hijmans, 2016), *sp* (Pebesma & Bivand, 2005), *maptools* (Bivand & Lewin-Koh, 2016), *rgeos* (Bivand & Rundel, 2016), and *letsR* (Vilela & Villalobos, 2015) in R (v. 3.2.3).

Results

The pattern of drosophilid richness predicted by MEM revealed patches of high richness in the central, south-eastern, western, north-western, and northern regions of the biome (Fig. 2a). The predictions from S-SDM showed a north-south gradient, with higher richness in central and south-eastern areas (similar to the pattern predicted by MEM, Fig. 2b), but with fewer species in the north of the biome (in contrast to MEM).

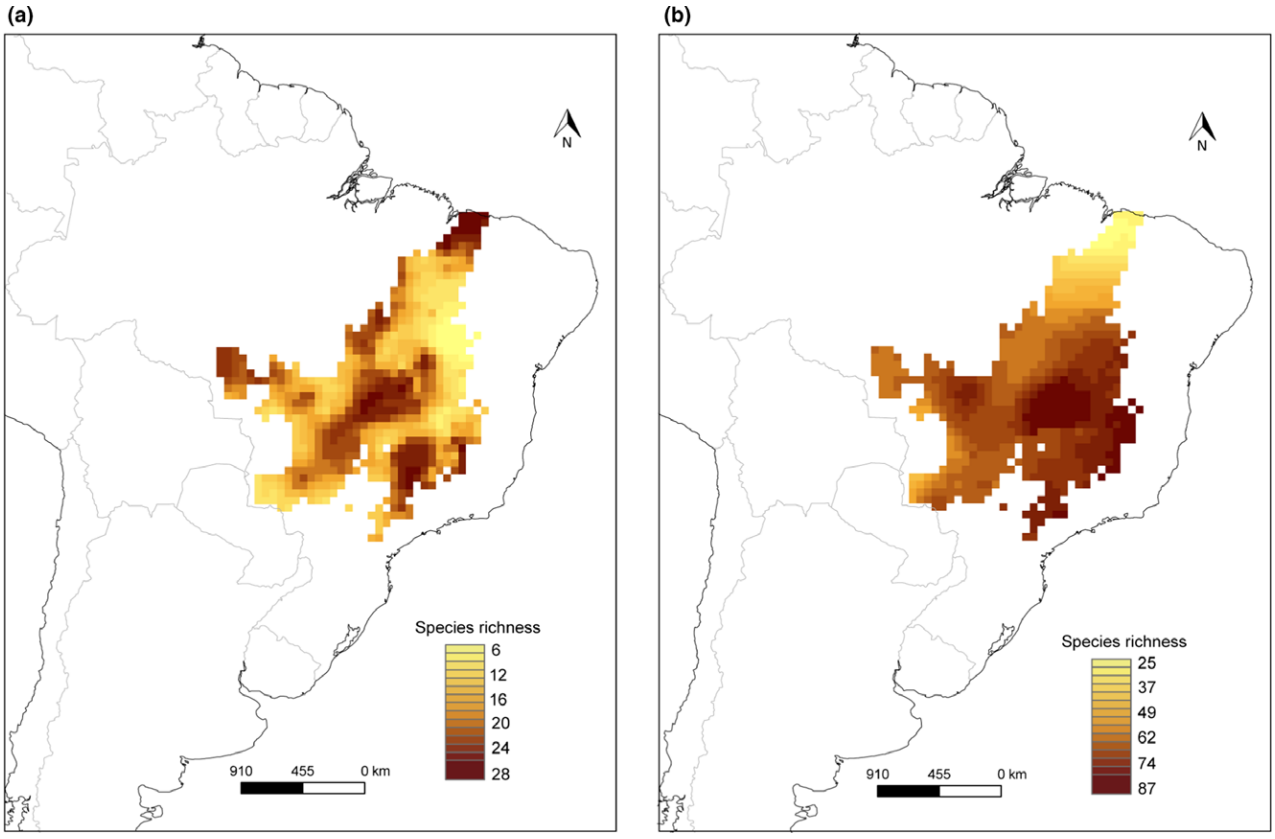

Fig. 2. Patterns of drosophilid species richness across the Cerrado biome; (a) richness from macroecological model; (b) richness from stacked species distribution models. [Colour figure can be viewed at wileyonlinelibrary.com]

The MEM prediction was moderately correlated with that from S-SDM ($r = 0.40$). Moreover, the number of species predicted by S-SDMs was markedly higher (varying from 25 to 87, mean = 69.8) than the number predicted by MEM (6–28 species, mean = 17.7; Table 1).

The observed richness of drosophilids across the selected grid cells (from 26 to 92, mean = 44) was closer to the S-SDM prediction than to MEM. For example, the maximum richness predicted by MEM in a cell was 25 species but was 87 by S-SDM, while the maximum number of species observed in the richest cells was 92. Also, for the five well-sampled cells (resulting from 18 years of monitoring and data colleting), richness predicted by S-SDM was quite similar to the observed value (mean = 69.6, Table 1; see also Fig. S2), indicating that the S-SDM predicted richness better than MEM. Moreover, correlations between S-SDM and observed richness increased as the number of species in well-sampled cells also increased, although the number of well-sampled cells was very low (Fig. 3).

Discussion

In this study, we integrated richness predictions derived from macroecological models (MEM) with stacked predictions from species distribution models (S-SDM) to investigate the geographic variation in drosophilid species richness in the Cerrado biome. The results revealed higher drosophilid richness in the central and south-eastern regions of the biome, consistent with reports for other invertebrates (e.g. young clades of termites, Eggleton *et al*., 1994; tiger moths, Ferro *et al*., 2010) and vertebrate groups (Diniz-Filho *et al*., 2006, 2008; Costa *et al*., 2007). Although MEM and S-SDM presented some similarities in their spatial patterns (i.e. high richness in the central and south-eastern areas, see Fig. 2), they strongly differed in the estimates of species richness. More importantly, S-SDM produced richness estimates that were more similar to the empirical values than the MEM predictions.

Table 1. Descriptive statistics of the richness values predicted by the macroecological model (MEM) and the stacked species distribution models (S-SDM) through the entire Cerrado (N = 679 cells), and for two sets of observed richness – one set (Observed-1) composed by the 16 richest cells whose richness values were obtained by counting the number of species based on occurrence records from online database plus field sampling, and the other (Observed-2) composed by five cells widely sampled through 18 years of fieldwork (see also Fig. S2 and Model comparisons section).

Richness	MEM	S-SDM	Observed-1	Observed-2
Mean richness	17.7	69.8	44	69.6
Minimum	6	25	26	57
Maximum	28	87	92	92
Stand. Dev.	3.3	13.5	20	13.5

These results differed from our initial expectation and from previous findings (e.g. Algar *et al*., 2009; Pineda & Lobo, 2009; Trotta-Moreu & Lobo, 2010; Dubuis *et al*., 2011; Pottier *et al*., 2013; Calabrese *et al*., 2014), in which S-SDM resulted in significant levels of overprediction error. This is expected since S-SDM do not account for biotic interactions or any constraints in the maximum number of species that can occupy the same geographic unit (Guisan & Rahbek, 2011). Nevertheless, our observation that S-SDM did not overpredict richness, but contrarily, that MEM underpredicted richness in comparison with observed values, suggests that the statements derived from comparisons between MEM and S-SDM predictions proposed by Guisan and Rahbek (2011) may not be general, but probably vary substantially across regions, and are dependent on the ecological characteristics of the assemblage and the quality of the available data, as we discuss below (see also Pellissier *et al*., 2012; Pottier *et al*., 2013; D'Amen *et al*., 2015a; Zurell *et al*., 2016).

Since species distribution models essentially represent abiotic/environmental filtering, they are expected to produce more accurate predictions of species richness for communities that are predominantly structured by climatic factors (e.g. Pottier *et al*., 2013). For instance, Pellissier *et al*. (2012) and Pottier *et al*. (2013) found that, at high altitudes, where climatic conditions represent strong environmental constraints, S-SDM produced more accurate predictions, with a significant relationship between predicted and observed species richness, than at low altitudes (see also Mateo *et al*., 2012). In such seasonal and stressful environments, species assemblages may be primarily determined by climate, and emergent patterns of species richness are the result of environmental constraints on species distributions. On the other hand, in milder climates, biotic interactions may be more important than purely abiotic effects for determining species distributions, and thus, stacking climate-based SDMs may lead to greater overprediction of richness (Pellissier *et al*., 2012; Mateo *et al*., 2012; Pottier *et al.,* 2013).

Due to high environmental seasonality (with two well-defined seasons, one dry and cold, and the other wet and warm), prolonged drought, and low nutrient availability, the Cerrado is considered a stressful environment for species growth and survival (Miranda *et al*., 1993; Hoffmann, 2000). This seasonality may influence variation in species richness because only those species sharing the required appropriate physiological, behavioural, and/or ecological attributes needed to tolerate the harsh dry season are able to survive (Costa *et al*., 2007; Pottier *et al*., 2013). Regarding drosophilids, the climatic seasonality in Cerrado plays an essential role on the dynamics of these assemblages (Tidon, 2006; Valadão *et al*., 2010). Mata *et al*. (2015) found that drosophilid assemblages retract during the dry season, decreasing to only 0.5% of their abundance during the rainy season. Such strong reduction in richness and abundance of drosophilid larva during the drought season is probably a consequence of decreasing resource availability in the environment through the dry

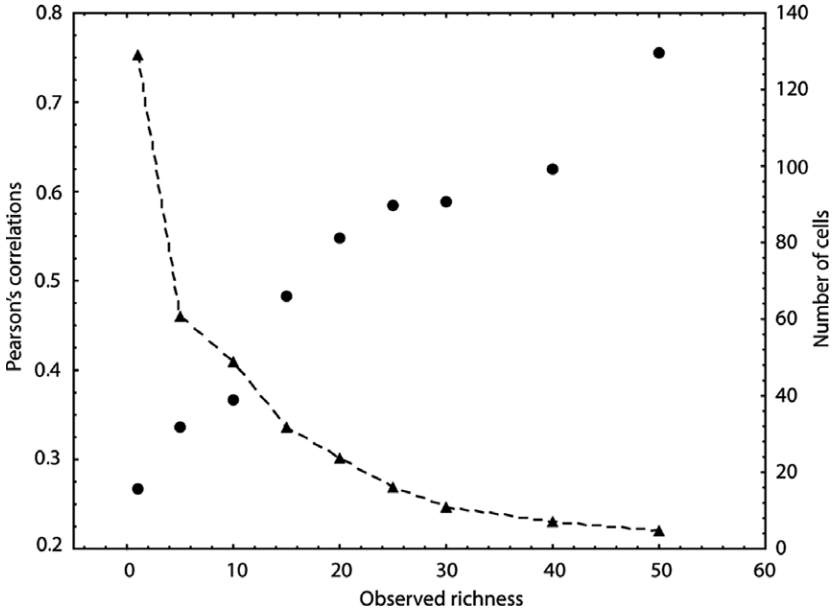

Fig. 3. Scatterplot between observed richness (x axis) and Pearson's correlations (from S-SDM and observed richness, y axis) represented by dots, and between observed richness (x axis) and the number of cells (z axis) with data for observed richness, represented by the dashed line. Triangular dots represent the observed richness datasets used to estimate Pearson's correlations (see Table S4).

season (Mata *et al*., 2015). Thus, the correlation between S-SDM and observed richness provides some support to the expectation that climate is the major (indirect) driver of drosophilid richness, and suggests that S-SDM probably captured important environmental filters that constrain drosophilid richness in the Cerrado.

Additionally, recent studies have shown that prediction accuracy of richness models may vary according to the ecological traits of the species assemblage (D'Amen *et al*., 2015b; Zurell *et al*., 2016). Zurell *et al*. (2016) showed that S-SDM overpredictions tend to be lower for habitat generalists (e.g. bird species foraging and breeding in gardens and mixed forests) and higher for habitat specialist species (bird species foraging and breeding in reeds or gravel banks). The reasoning behind this is that the communities composed predominantly of generalist species can avoid inter-specific competition using resources from different sub-regions of their fundamental niche (Colwell & Fuentes, 1975; Zurell *et al*., 2016). Thus, the role of competition in constraining species assemblages and richness patterns is probably less important in such communities. Also, some studies have suggested that the distributions of potential drosophilid competitors over the resource patches are intra-specifically aggregated, thus facilitating the coexistence of species at local scales (e.g. Shorrocks *et al*., 1979; Sevenster & Van Alphen, 1996; Krijger & Sevenster, 2001). Therefore, it is possible that competition has low influence on drosophilid community assemblage (at least at the scale of analysis of this study, see also Thuiller *et al*., 2015), thus favouring species co-occurrence and reducing S-SDM overprediction as we found here.

Finally, the strong MEM richness underprediction observed in this study differed from most previous findings (e.g. Algar *et al*., 2009; Dubuis *et al*., 2011), which frequently reported good predictive ability for MEM in comparison to S-SDM and observed richness. In our case, the cause seems to be primarily statistical instead of ecological. Firstly, the macroecological models, as originally proposed by Guisan and Rahbek (2011) to derive the SESAM assumptions, are commonly constructed based on high-resolution datasets (e.g. Dubuis *et al*., 2011; Calabrese *et al*., 2014; D'Amen *et al*., 2015a and b) or range filling maps (occurrence extent, Guisan & Rahbek, 2011) of well-sampled regions. In such cases, the number of species count in each geographical unit (i.e. the observed richness) is directly used to estimate richness by MEM. Consequently, assuming that these geographical units are sufficiently sampled, one could expect that MEM will always predict richness values very close to the observed richness; however, in regions where sampling efforts have been insufficient or uneven, the use of methods for correcting sampling bias in the estimation of species richness is needed, as we did here using rarefaction. Although rarefaction was recently recommended for richness estimation from substantial datasets (Engemann *et al*., 2015), it is also sensitive to insufficient sampling (Gotelli & Colwell, 2001), especially in communities dominated by locally rare species, as in the case of Cerrado drosophilids (Roque *et al*., 2013). In such cases, the true number of species across the geographic units tends to be underestimated (e.g. Mora *et al*., 2008; Engemann *et al*., 2015), consequently leading to an underprediction by MEM.

In summary, our study indicates that richness estimates based on macroecological modelling are almost certainly affected by inventory incompleteness and sampling bias. Unfortunately, the incompleteness and bias in diversity data are common in tropical biomes, including the Cerrado. On the other hand, SDM has proven very successful in producing reliable species distribution maps even from a limited number of specimens (e.g. Almeida *et al*., 2010) and can be a valuable approach to explore species richness patterns from georeferenced data in poorly sampled regions. We also showed that the SESAM assumptions are not always valid but may be dependent on the ecological characteristics of the modelled species, and especially, on the quality of data available to construct MEMs. Finally, the similarity between S-SDM predictions and observed richness from well surveyed areas unequivocally indicated that the pattern of drosophilid assemblage in the Cerrado is primarily driven by climatic constraints at the broad scale.

Acknowledgements

Our research programme has been continuously supported by grants to the GENPAC (Geographical Genetics and Regional Planning for natural resources in Brazilian Cerrado) research network, provided by MCTI/CNPq/CAPES, FAPEG, and FAPDF (project numbers 564717/2010-0, 564718/2010, 563727/2010-1, and 563624/2010-8). The research of LCT, RT, and JAFDF is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) productivity grants. We thank the World Climate Research Programmer's Working Group on Coupled Modeling for providing CMIP5 and the climate models available in ecoClimate. We are grateful to Markus Eichhorn and Bradford A. Hawkins for constructive comments on earlier versions of the manuscript.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: doi: 10.1111/icad.12240:

Figure S1. Relationship between the number of records in each cell and Pearson correlations from MEM*SDM used to determine the sample size (i.e. number of cells and records) from which the pattern of MEM richness was considered for further comparisons.

Figure S2. Map of drosophilid richness derived from counting the number of species recorded in each 0.5 × 0.5 degree cell based on the online database Taxodros, literature data, and from field sampling.

Table S1. List of 127 Drosophilid species and number of records used in the models.

Table S2. Details on climatic models (AOGCMs) used for Ecological Niche Modeling.

Table S3. Ecological Niche Models (ENMs) used to model the potential geographic distribution of drosophilid species.

Table S4. Pearson's correlations (r) between S-SDM and observed richness for sets of cells selected by truncating the minimum value of observed richness in the following classes: at least one species (129 cells); at least five species (61 cells), at least 10 species (49 cells), at least 15 species (32 cells), at least 20 species (24 cells), at least 25 species (16 cells), at least 30 species (11 cells), at least 40 species (7 cells), and at least 50 species (5 cells).

References

Algar, A.C., Kharouba, H.M., Young, E.R. & Kerr, J.T. (2009) Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, **32**, 22–33.

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Almeida, M.C., Côrtes, L.G. & De Marco, P. (2010) New records and a niche model for the distribution of two Neotropical damselflies: *Schistolobos boliviensis* and *Tuberculobasis inversa* (Odonata: Coenagrionidae). *Insect Conservation and Diversity*, **3**, 252–256.

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.

Araújo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527–1539.

Bächli, G. (2016) Taxodros: The Database on Taxonomy of Drosophilidae, version 2014–1. < http://taxodros.unizh.ch > 2th April 2016.

Ballesteros-Mejia, L., Kitching, I.J., Jetz, W., Nagel, P. & Beck, J. (2013) Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. *Global Ecology and Biogeography*, **22**, 586–595.

Beck, J., Ballesteros-Mejia, L., Buchmann, C.M., Dengler, J., Fritz, S.A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M. & Dormann, C.F. (2012) What's on the horizon for macroecology? *Ecography*, **35**, 001–011.

Bivand, R. & Lewin-Koh, N. (2016) *maptools: Tools for Reading and Handling Spatial Objects. R package version 0.8-39*. < https://CRAN.R-project.org/package = maptools > 30th April 2016.

Bivand, R. & Rundel, C. (2016) *rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-20*. < https://CRAN.R-project.org/package = rgeos > 18th June 2016.

Boucher-Lalonde, V., Kerr, J.T. & Currie, D.J. (2014) Does climate limit species richness by limiting individual species' ranges? *Proceedings of the Royal Society B*, **281**, 1–8.

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014) Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.

Chaves, N.B. & Tidon, R. (2008) Biogeographical aspects of drosophilids (Diptera, Drosophilidae) of the Brazilian savanna. *Revista Brasileira de Entomologia*, **52**, 340–348.

Colwell, R.K. & Fuentes, E.R. (1975) Experimental studies of the niche. *Annual Review of Ecology and systematics*, **6**, 281–310.

Colwell, R.K. & Rangel, T.F. (2009) Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences*, **106**, 19651–19658.

Costa, G.C., Nogueira, C., Machado, R.B. & Colli, G.R. (2007) Squamate richness in the Brazilian Cerrado and its environmental-climatic associations. *Diversity and Distributions*, **13**, 714–724.

Currie, D.J. (1991) Energy and large-scale patterns of animaland plant-species richness. *American Naturalist*, **137**, 27–49.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E. & Turner, J.R.G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121–1134.

D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L. & Guisan, A. (2015b) Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *Journal of biogeography*, **42**, 1255–1266.

D'Amen, M., Pradervand, J.-N. & Guisan, A. (2015a) Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography*, **24**, 1443–1453.

Diniz-Filho, J.A.F., Bini, L.M., Pinto, M.P., Rangel, T.F., Carvalho, P. & Bastos, R.P. (2006) Anuran species richness, complementarity and conservation conflicts in Brazilian Cerrado. *Acta Oecologica*, **29**, 9–15.

Diniz-Filho, J.A.F., Bini, L.M., Rangel, T.F., Loyola, R.D., Hof, C., Nogués-Bravo, D. & Araújo, M.B. (2009) Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, **32**, 897–906.

Diniz-Filho, J.A.F., Bini, L.M., Vieira, C.M., Blamires, D., Terribile, L.C., Bastos, R.P., Oliveira, G. & Barreto, B.S. (2008) Spatial patterns of terrestrial vertebrate species richness in the Brazilian Cerrado. *Zoological Studies*, **47**, 146–157.

Diniz-Filho, J.A.F., Ceccarelli, S., Hasperué, W. & Rabinovich, J. (2013) Geographical patterns of Triatominae (Heteroptera: Reduviidae) richness and distribution in the Western Hemisphere. *Insect Conservation and Diversity*, **6**, 704–714.

Diniz-Filho, J.A.D., Marco, P. & Hawkins, B.A. (2010) Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. *Insect Conservation and Diversity*, **3**, 172–179.

Diniz-Filho, J.A.F., Rangel, T.F. & Santos, M.R. (2012) Extreme deconstruction supports niche conservatism driving New World bird diversity. *Acta Oecologica*, **43**, 16–21.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.

Eggleton, P., Williams, P.H. & Gaston, K.J. (1994) Explaining global termite diversity: productivity or history? *Biodiversity and Conservation*, **3**, 318–330.

Engemann, K., Enquist, B.J., Sandel, B., Boyle, B., Jørgensen, P.M., Morueta-Holme, N. & Svenning, J.C. (2015) Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, **5**, 807–820.

Ferro, V.G., Melo, A.S. & Diniz, I.R. (2010) Richness of tiger moths (Lepidoptera: Arctiidae) in the Brazilian Cerrado: how much do we know? *Zoologia (Curitiba)*, **27**, 725–731.

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

Franklin, J. (2009) *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge, UK.

Gotelli, N.J., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R., Currie, D.J., Dunn, R.R., Graves, G.R., Green, J.L., Grytnes, J.A., Jiang, Y.H., Jetz, W., Kathleen Lyons, S., McCain, C.M., Magurran, A.E., Rahbek, C., Rangel, T.F., Soberón, J., Webb, C.O. & Willing, M.R. (2009) Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, **12**, 873–886.

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Guisan, A. & Rahbek, C. (2011) SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

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. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

Hijmans, R.J. (2016) *raster: Geographic Data Analysis and Modeling. R package version 2.5-8*. <https://CRAN.R-project.org/package=raster> 30th April 2016.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hoffmann, W.A. (2000) Post-establishment seedling success in the Brazilian Cerrado: a comparison of savanna and forest species. *Biotropica*, **32**, 62–69.

Krijger, C.L. & Sevenster, J.G. (2001) Higher species diversity explained by stronger spatial aggregation across six neotropical *Drosophila* communities. *Ecology Letters*, **4**, 106–115.

Lemes, P. & Loyola, R.D. (2013) Accommodating species climate-forced dispersal and uncertainties in spatial conservation planning. *PlosOne*, **8**, e54323.

Lima-Ribeiro, M.S., Varela, S., González-Hernández, J., Oliveira, G., Diniz-Filho, J.A.F. & Terribile, L.C. (2015) EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodiversity Informatics*, **10**, 1–21.

Mata, R.A., Roque, F. & Tidon, R. (2008) Drosophilids (Insecta: Diptera) of the Paranã valley: eight new records for the Cerrado biome. *Biota Neotropica*, **8**, 55–60.

Mata, R.A. & Tidon, R. (2013) The relative roles of habitat heterogeneity and disturbance in drosophilid assemblages (Diptera, Drosophilidae) in the Cerrado. *Insect Conservation and Diversity*, **6**, 663–670.

Mata, R.A., Valadão, H. & Tidon, R. (2015) Spatial and temporal dynamics of drosophilid larval assemblages associated to fruits. *Revista Brasileira de Entomologia*, **59**, 50–57.

Mateo, R.G., Felicísimo, Á.M., Pottier, J., Guisan, A. & Muñoz, J. (2012) Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS ONE*, **7**, e32586.

Miranda, A.C., Miranda, H.S., Dias, I.F.O. & Dias, B.F.S. (1993) Soil and air temperatures during prescribed cerrado fires in Central Brazil. *Journal of Tropical Ecology*, **9**, 313–332.

Mora, C., Tittensor, D.P. & Myers, R.A. (2008) The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proceedings of the Royal Society B*, **275**, 149–155.

Parsons, P.A. & Bock, I.R. (1979) The population biology of Australian Drosophila. *Annual Review of Ecology and Systematics*, **10**, 229–245.

Pebesma, E.J. & Bivand, R.S. (2005). *Classes and methods for spatial data in R. R News 5 (2)*. <http://cran.r-project.org/doc/Rnews/> 30th April 2016.

Pellissier, L., Pradervand, J.-N., Pottier, J., Dubuis, A., Maiorano, L. & Guisan, A. (2012) Climate-based empirical models show biased predictions of butterfly communities along environmental gradients. *Ecography*, **35**, 684–692.

Peterson, A.T., Sánchez-Cordero, V., Soberón, J., Bartley, J., Buddemeier, R.W. & Navarro-Sigüenza, A.G. (2001) Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecological modelling*, **144**, 21–30.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological niches and geographic distributions*. Monographs in Population Biology, 49. Princeton University Press, Princeton, New Jersey.

Pineda, E. & Lobo, J.M. (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182–190.

Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P. & Guisan, A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52–63.

Rangel, T.F. & Loyola, R.D. (2012) Labeling ecological niche models. *Brazilian Journal of Nature Conservation*, **10**, 119–126.

Roque, F., Hay, J.D.V. & Tidon, R. (2009) Breeding sites of drosophilids (Diptera) in the Brazilian Savanna. I. Fallen fruits of *Emmotum nitens* (Icacinaceae), *Hancornia speciosa* (Apocynaceae) and *Anacardium humile* (Anacardiaceae). *Revista Brasileira de Entomologia*, **53**, 308–313.

Roque, F., Mata, R.A. & Tidon, R. (2013) Temporal and vertical drosophilid (Insecta; Diptera) assemblage fluctuations in a neotropical gallery forest. *Biodiversity and Conservation*, **22**, 657–672.

Roque, F. & Tidon, R. (2008) Eight new records for drosophilids (Insecta: Diptera) in the Brazilian savanna. *Drosophila Information Service*, **91**, 94–98.

Sevenster, J.G. & Van Alphen, J.J. (1996) Aggregation and coexistence. II. A neotropical Drosophila community. *Journal of Animal Ecology*, **65**, 308–324.

Shorrocks, B., Atkinson, W. & Charlesworth, P. (1979) Competition on a divided and ephemeral resource. *The Journal of Animal Ecology*, **48**, 899–908.

Stökl, J., Strutz, A., Dafni, A., Svatos, A., Doubsky, J., Knaden, M., Sachse, S., Hansson, B.S. & Stensmyr, M.C. (2010) A deceptive pollination system targeting drosophilids through olfactory mimicry of yeast. *Current Biology*, **20**, 1846–1852.

Terribile, L.C., Diniz-Filho, J.A.F., Rodríguez, M.A. & Rangel, T.F. (2009) Richness patterns, species distribution models and the principle of extreme deconstruction. *Global Ecology and Biogeography*, **18**, 123–136.

Thuiller, W., Pollock, L.J., Gueguen, M. & Münkemüller, T. (2015) From species distributions to meta-communities. *Ecology letters*, **18**, 1321–1328.

Tidon, R. (2006) Relationships between drosophilids (Diptera, Drosophilidae) and the environment in two contrasting tropical vegetations. *Biological Journal of the Linean Society*, **87**, 233–247.

Trotta-Moreu, N. & Lobo, J.M. (2010) Deriving the species richness distribution of Geotrupinae (Coleoptera: Scarabaeoidea) in Mexico from the overlap of individual model predictions. *Environmental entomology*, **39**, 42–49.

Valadão, H., Hay, J.D.V. & Tidon, R. (2010) Temporal dynamics and resource availability for drosophilid fruit flies (Insecta, Diptera) in a gallery forest in the Brazilian Savanna. *International Journal of Ecology*, **2010**, 7. Article ID 152437.

Vilela, B. & Villalobos, F. (2015) letsR: a new R package for data handling and analysis in macroecology. *Methods in Ecology and Evolution*, **6**, 1229–1234. https://doi.org/10.1111/2041-210x.12401.

Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distribution*, **11**, 3–23.

Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, **19**, 639–644.

Zurell, D., Zimmermann, N.E., Sattler, T., Nobis, M.P. & Schröder, B. (2016) Effects of functional traits on the prediction accuracy of species richness models. *Diversity and Distributions*, **22**, 905–917.

Accepted 10 May 2017
First published online 14 June 2017

Editor: Yves Basset
Associate editor: Francis Gilbert