

Putting insects on the map: near-global variation in sphingid moth richness along spatial and environmental gradients

Liliana Ballesteros-Mejia, Ian J. Kitching, Walter Jetz and Jan Beck

L. Ballesteros-Mejia (ballesteros.liliana@gmail.com), Laboratorio de Genética e Biodiversidade, Inst. de Ciências Biológicas, Univ. Federal de Goiás (UFG), Goiania, Goiás, Brazil. – I. J. Kitching, Natural History Museum, Dept of Life Sciences, London, UK. – W. Jetz, Yale Univ., Dept of Ecology and Evolution, New Haven, USA, and Dept of Life Sciences, Imperial College London, Ascot, Berks, UK. – J. Beck, Univ. of Colorado Boulder, Museum of Natural History, Boulder, CO, USA. LB-M and JB also at: Univ. of Basel, Inst. of Biogeography, Basel, Switzerland.

Despite their vast diversity and vital ecological role, insects are notoriously underrepresented in biogeography and conservation, and key broad-scale ecological hypotheses about them remain untested – largely due to generally incomplete and very coarse spatial distribution knowledge. Integrating records from publications, field work and natural history collections, we used a mixture of species distribution models and expert estimates to provide geographic distributions and emergent richness patterns for all ca 1000 sphingid moth species found outside the Americas in high spatial detail. Total sphingid moth richness, the first for a higher insect group to be documented at this scale, shows distinct maxima in the wet tropics of Africa and the Oriental with notable decay toward Australasia. Using multivariate models controlling for spatial autocorrelation, we found that primary productivity is the dominant environmental variable associated with moth richness, while temperature, contrary to our predictions, is an unexpectedly weak predictor. This is in stark contrast to the importance we identify for temperature as a niche variable of individual species. Despite divergent life histories, both main sub-groups of moths exhibit these relationships. Tribal-level deconstruction of richness and climatic niche patterns indicate idiosyncratic effects of biogeographic history for some of the less species-rich tribes, which in some cases exhibit distinct richness peaks away from the tropics. The study confirms, for a diverse insect group, overall richness associations of remarkable similarity to those documented for vertebrates and highlights the significant within-taxon structure that underpins emergent macroecological patterns. Results do not, however, meet predictions from vertebrate-derived hypotheses on how thermoregulation affects the strength of temperature–richness effects. Our study thus broadens the taxonomic focus in this data-deficient discourse. Our procedures of processing incomplete, scattered distribution data are a template for application to other taxa and regions.

The relative roles of evolutionary and ecological mechanisms driving species distributions and the spatial biodiversity pattern they form, as well as their susceptibility to changing environmental gradients, are at the core of biodiversity science (Mittelbach et al. 2001, Willig et al. 2003, Currie et al. 2004). However, most insights to date are based on a limited set of taxa, i.e. vertebrates and vascular plants, with much of the remaining biodiversity remaining heavily underrepresented (Stork 1988, May 1994, Beck et al. 2012). The vast majority of species on Earth are invertebrates, and among them herbivorous insects make a sizable contribution in terrestrial systems (Godfray et al. 1999, Hamilton et al. 2013). Apart from their overwhelming diversity, insects show life history trait combinations that are not represented among vertebrates. This raises the question how representative vertebrate biogeographic patterns are for insects, and whether the mechanisms shaping them may differ. Unfortunately, for most invertebrate species only vague ideas on their geographic ranges exist (i.e. at continental resolution for many species). This currently

prevents a broad-scale spatial biodiversity assessment of insect groups and testing of key hypotheses on large-scale biodiversity patterns (Willig et al. 2003) for their generality. While data for some taxa exist for North America and Europe, earlier attempts on larger extents (incl. tropical regions) were severely constrained by coarse resolution data (e.g. based on national checklists; Pearson and Cassola 1992). A recent push of new initiatives and methodologies on which our study builds has begun to address this (Basset et al. 2012, Jetz et al. 2012, Keil et al. 2013).

We provided and use a new dataset on sphingid moth species distributions to investigate hypothesized environmental associations of species richness hitherto only globally assessed for plants and vertebrates (Currie et al. 2004, Kreft and Jetz 2007, Jetz and Fine 2012). Sphingids are a family of large, fast-flying and, in some cases, extremely dispersive Lepidoptera (Kitching and Cadiou 2000). Caterpillars are folivorous with a moderate to low degree of hostplant specialization (typically, plant family or order). The group consists of seven well-supported taxonomic tribes (Kawahara

et al. 2009) that differ in species richness, distribution and ecological traits (such as host plant use, mobility and adult feeding). The two most species-rich tribes, in particular, are most divergent in their life histories. In ‘income-breeding’ *Macroglossini*, adult feeding allows a prolonged reproductive phase and morphology suggests higher mobility (Janzen 1984). In contrast, ‘capital breeding’ *Smerinthini* do not feed as adults, utilize a different larval host plant spectrum, morphology suggests lower mobility, and local studies indicate a preference for stable, forested habitats (Beck et al. 2006a, Beck and Kitching 2007). Most of the smaller tribes probably share characteristics with the *Macroglossini* rather than with the *Smerinthini*, with a possibly intermediate set of traits in *Ambulycini* (A. Kawahara pers. comm., Beck et al. 2006b).

Among hypothesized environment-diversity links at large spatial scales, effects of ambient energy are of particular interest (Willig et al. 2003, Evans et al. 2005). Available food resources (i.e. potential energy) positively affect population sizes, thereby diminishing extinction probabilities and ultimately increasing species richness (Evans et al. 2005, Allen et al. 2007, Hurlbert and Jetz 2010). Consistent with this, a positive relationship between richness and net primary productivity has repeatedly been reported (Currie et al. 2004, Jetz and Fine 2012) but tests using insects have mostly been restricted to very small spatial extents (Classen et al. 2015). We predict a positive correlation between actual evapotranspiration (AET), a proxy for net primary productivity, and moth richness patterns.

Temperature (i.e. kinetic energy) has been hypothesized to affect richness via its effects on metabolism, growth rate, generation time and speciation rate, especially in ectotherms, which are most directly dependent on ambient thermal conditions (Allen et al. 2007, Buckley et al. 2012). Furthermore, low temperature may restrict ectotherms’ ability to forage, leading to limitations in potential energy despite food being available (Classen et al. 2015). In contrast, in endotherms key biological rates are only weakly connected to ambient temperatures, at least within thermoneutral zones. Ideas on the link between thermoregulation and the relative roles of potential and kinetic energy on richness were shaped by, and tested on, comparison of the four main terrestrial vertebrate groups (Buckley et al. 2012). Broadening the taxonomic database by testing predicted relationships for non-vertebrate ectotherms is paramount to put this hypothesis on firmer ground. Leaving aside the ability of select sphingid species to undertake pre-flight heating of thoracic muscles (Heinrich 1993), these moths are a strictly ectothermic group. Low temperatures should therefore strongly constrain their species richness, and we expect an important role for direct temperature effects relative to those of productivity.

Temperature may also act as an important niche dimension for individual species, apart from affecting richness at the assemblage level. Productivity, however, can only be assumed to constrain the richness of an assemblage (McArthur 1972), whereas it seems less plausible that individual species adapt physiologically to particular productivity levels. Such links of environmental effects across aggregation levels (i.e. species’ niches vs assemblage richness) are poorly understood (Terribile et al. 2009). We expect that climates that fit the group’s climatic niche preferences will have higher richness.

Environmental correlations with species richness can only serve as general explanations if they are found across taxa irrespective of their particular life histories and adaptations (Allen et al. 2007). Sufficiently inclusive taxonomic groups may permit rigorous testing of environment–richness associations, although historical idiosyncrasies may dilute or bias patterns in particular groups. By deconstructing patterns of richness and species’ climatic niche placement at the tribal level, we investigate whether they follow the same spatial and environmental gradients, or whether they differ from one to another with total richness a largely emergent property. Such among-clade discrepancy may arise from disparate geographic centres of diversification and dissimilar, strongly conserved environmental niche preferences (Wiens and Donoghue 2004, Buckley et al. 2010). We also investigate the link between aggregation levels (assemblage richness vs species-level niche) replicated across tribes.

Currently, gaps in knowledge of invertebrate distributions are only partly due to limited incentive for field research (Basset et al. 2012). Many data on species occurrences are already available in the form of natural history collections, but only their digitization, mobilization, quality control and cautious use will fully empower their use in biodiversity research (Beck et al. 2012, Jetz et al. 2012). At the same time, species distribution modelling (SDM; Elith and Leathwick 2009) is a methodological toolbox that builds correlative models by combining recorded species presences with climatic and remotely sensed landcover data to infer likely regions of occurrence for taxa. These methods make many unwarranted assumptions, have clear limitations regarding mechanistic inference, and manifold options for wrong implementation and erroneous interpretation (Beale and Lennon 2012, Joppa et al. 2013). They nevertheless provide a sound operational approach to attain reproducible baseline distribution estimates for many of the more poorly known taxa. Here we showcase this opportunity for the limited distribution data typical for insects and demonstrate the use and transparent validation of resulting range predictions.

Specifically, we develop and further analyse predicted distributional information for all ca 1000 species of sphingid moths occurring outside the Americas (see Supplementary material Appendix 1 for detail). We compiled records of species’ presences from a multitude of sources and generated distribution estimates with expert-edited SDMs. Resulting range estimates are the first of their kind (i.e. high resolution, near-global extent) for a higher insect taxon and now enable testing the generality of key environmental predictors of large-extent biodiversity patterns on an insect group. To confirm the reliability of resulting species richness assessments, and to evaluate the potential trade-off that fine-grained data may be associated with higher uncertainty, we validated richness estimates across grain sizes.

Thus, the aims of our study are twofold. First, we provide details of our compilation, processing, mapping and validating of novel species range data and resulting species richness patterns. Second, we utilize these data to test general hypotheses on the links of richness and environmental variation, focussing in particular on the roles of energy in the form of primary productivity and temperature.

Material and methods

Here we first describe in detail the methodology applied to retain comprehensive, gridded maps of species' geographic ranges and richness patterns from a compilation of raw distribution records (see Fig. 1 for general approach). A more detailed version of this section is presented in Supplementary material Appendix 1. Below, we explain the methods for addressing the environment–richness relationship with these data.

Compiling and processing the distribution record database

We extracted distribution records for sphingid moths from published literature (>1300 references, Supplementary material Appendix 2), supplemented by own field sampling, and we used distribution records from online data bases such as GBIF (<www.gbif.org>, Nov. 2009) and BOLD (<www.barcodinglife.org>, Aug. 2010). Furthermore, we retrieved specimen label data from >400 natural history collections (Supplementary material Appendix 2), from amateur collectors, and from online searches. We restricted analyses of data to species ranges west of 180°E (i.e. excluding data for islands of the Eastern Pacific, and species endemic to that region); remaining data refer to 972 species. Details on data processing, including treatment of migrants, vagrants and likely errors, are presented in Supplementary material Appendix 1.

For the majority of species, we followed the nomenclature of Kitching and Cadiou (2000) and more recent taxonomic publications. However, taxonomic findings can have a considerable time-lag until reaching official status according to the International Code of Zoological Nomenclature (ICZN 1999), so we allowed deviations for the purposes of this data compilation. Supplementary material Appendix 1 provides more detail and Supplementary material Appendix 3 lists species' names as utilized in this dataset. For higher-taxon associations we followed the molecular phylogeny of Kawahara et al. (2009). Generally, in order to avoid errors due to misidentifications, identification of difficult taxa was checked by the taxonomist in our team (IJK; either on the specimen or by photograph) unless data stemmed from a renowned expert on sphingid identification.

Georeferencing

Based on locality data given in publications or associated to specimen labels, we assigned geographic coordinates to distribution records. If not given from GPS measurement, we found coordinates of localities in online gazetteers, Google Earth, Atlases, and local maps. For difficult localities we also made use of historical atlases and travel itineraries of the collectors in questions.

We generally aimed at georeferencing with a resolution of 0.01° (ca 1 km) or higher. However, sometimes this was not feasible either because we could not find sites precisely enough, or because no detailed locality was given in original sources. We estimated the spatial precision of records (i.e. $\leq 0.01^\circ$, $\leq 0.1^\circ$, $\leq 1^\circ$ or 'unspecific' (e.g. 'southern

India'; Wieczorek et al. 2004) to facilitate filtering our data depending on the resolution required. If data could not be localized precisely but altitude information was given, we set coordinates to a region of similar altitude (using Google Earth, which incorporates a 90 m resolution digital elevation model) to minimize environmental deviation from the true site.

We spent great lengths to identify georeferencing errors by mapping data and by checking consistency between records processed by different people (see Acknowledgements). We also subjected data already containing coordinates to this procedure (e.g. collectors' GPS-data or downloads from databases), and we found many errors (often apparently due to mistyping). We prioritized databasing and georeferencing towards regions and taxa with relatively few records, hereby adding more information per work effort (Beck et al. 2013).

Distribution modelling and its validation

We based all SDMs on climatic data provided by Worldclim and on vegetation cover from remote sensing (MODIS vegetation continuous fields, 2000–2001 data; Supplementary material Appendix 1, Fig. A1). Modelling was carried out at a spatial resolution of 2.5 arcmin (ca 5 km). SDM was based on implementing the maximum entropy algorithm (Maxent; Phillips et al. 2006).

For species with a large number of available records we used only data that were georeferenced with high precision (i.e. if more than 50 spatially independent records, at modelling resolution, where available, records with a precision $> 0.1^\circ$ were excluded). However, we included records up to a precision of 1° for data-deficient species. Unspecific (regional) records were not used for modelling, but we considered them when validating and editing range estimates for dispersal barriers (see below). For species known from fewer than five spatially independent localities we reviewed SDM data (if models could be run at all) rigorously as if they were tentative, expert-drawn range estimates.

We used Maxent software (ver. 3.3.3e) with default settings, except that we used 10 000 background points chosen from a bias file to account for undersampled regions (Phillips et al. 2009). The bias file was produced from kernel densities of raw records (100 km radius). As a result, background sampling of environmental conditions is down-weighted in regions where no distinction at all is possible between a species' absence and lack of sampling. Modelling regions were restricted a priori according to the known biogeographic association of species (e.g. excluding Europe and Asia when modelling a species restricted to sub-Saharan Africa). However, if in doubt on the potential spread of a species, we erred on the side of modelling more inclusive extents.

We evaluated raw SDMs according to two criteria: 1) we followed standard SDM procedures by randomly splitting available data and using 75% for model fitting or 'training', and the remaining 25% for testing. This cross-validation procedure retrieved a metric akin to the area under the receiver-operating characteristic (AUC), based on averaging five replicate model runs. Because Maxent software replaces commission error with relative predicted area, we denote

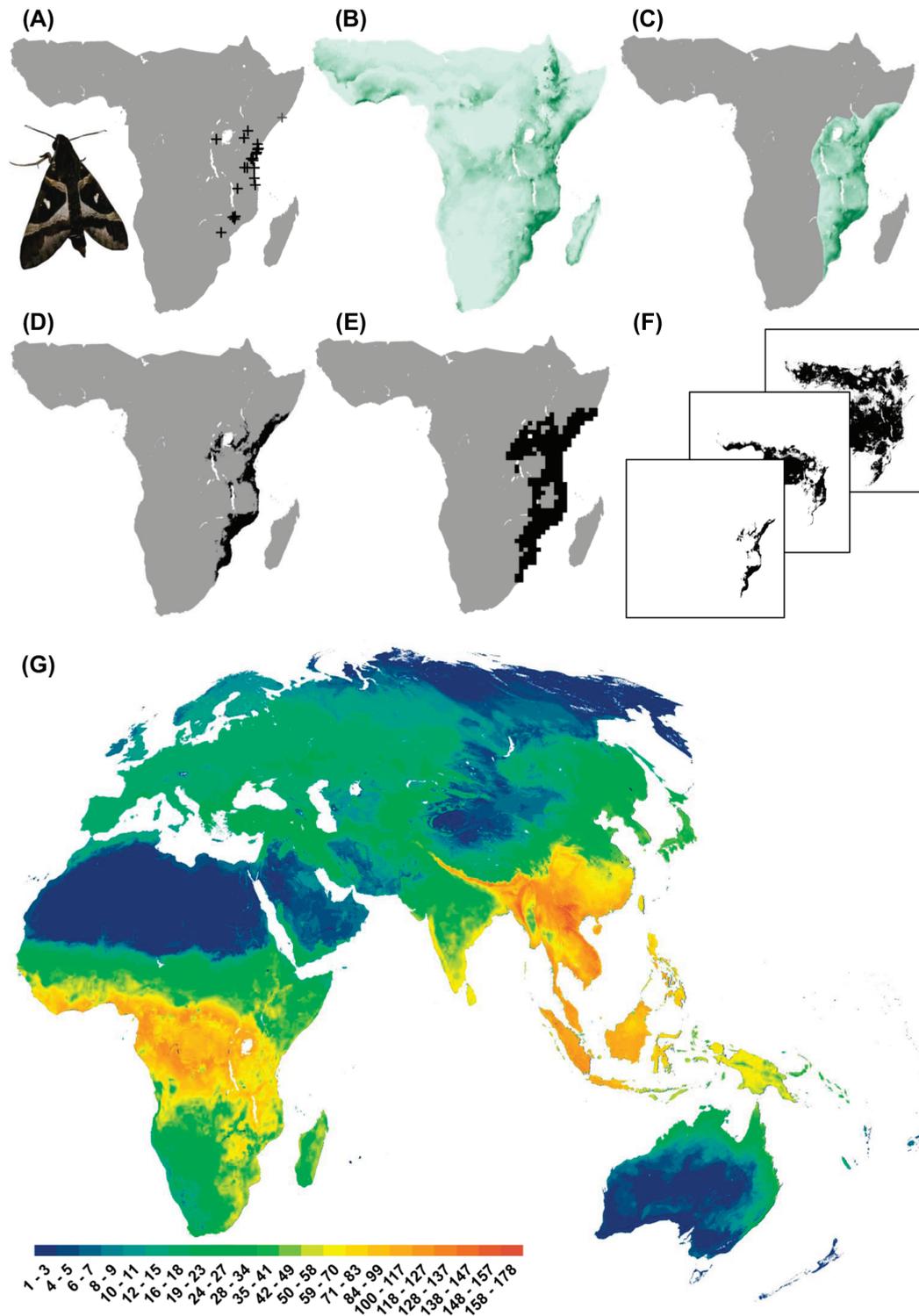


Figure 1. SpHINGid moth data processing and richness predictions. (A) The example species *Nephela argentifera* is known from 39 records. Modeling extent was a priori restricted to sub-saharan Africa. (B) SDM output highlights suitable regions (darker green = better habitat; cross-validation $AUC_{Mx} = 0.95$). (C) The model was evaluated and post-edited for dispersal limitation (considerations: the species seems genuinely absent from well-sampled South Africa and Madagascar, there is no evidence for its presence in Ethiopia or at some of the well-sampled locations in the Congo region). (D) Thresholding leads to a presence-prediction at modelling resolution. (E) Range estimates were aggregated to different grain sizes (shown here: 100 km). (F) Ranges of all species were stacked; the number of species predicted within each cell was counted. (G) Spatial variation in sphingid moth richness, mapped at a grain size of 5 km (Mollweide World equal area projection). See Supplementary material Appendix 4, Fig. A4 for maps at all available grain sizes.

these as AUC_{Mx} to avoid confusion with true AUC based on commission assessments. We do not rest our quality assessment on AUC_{Mx} alone because the lack of commission

data, as well as other issues, complicates its interpretation (Merckx et al. 2011, Jiménez-Valverde 2012). 2) Range predictions were checked for plausibility based on record

data (including those not used for modelling), inventories at very well-sampled sites (providing information on true absences of a species) and our knowledge of the species' biology (e.g. host plant associations). We tried to find and fix sources of error for obviously bad models (e.g. input data biases; Beck et al. 2014). If this did not lead to improvement, we excluded species from SDM and provided expert-drawn range estimates instead.

Range editing, thresholds, and expert estimates

Current SDM methods are unable to account for dispersal limitation. Rather, the output is a measure of environmental suitability, irrespective of whether the species has reached a region or not. Beyond the first step of limiting modelling regions if the broad extent of occurrence of a species was certain, we applied expert-opinion edits of modelled distributions based on known biogeographic barriers for sphingids (Beck et al. 2006a) or for other taxa, as well as large gaps in suitable habitat inferred from model output (see Supplementary material Appendix 1 for details). We assumed that species did not cross such potential barriers unless demonstrated by records. The extent of model predictions was limited by these conclusions.

For transforming continuous suitability into binary presence-absence predictions we set a threshold so that at least 90% of recorded presences were predicted correctly (Engler et al. 2004, Bean et al. 2012, Radosavljevic and Anderson 2014). Other, recently recommended thresholding rules (Bean et al. 2012) require confirmed absences, which are not available in any presence-only dataset. However, by taking absences from 'well-sampled' cells (see below) we could investigate model quality and threshold selection for a small test set of widespread species. For these, we calculated alternative thresholds, resulting range estimates, and 'true' AUC for comparison.

For some species we were unable to model distributions, e.g. due to lack of sufficient occurrence data. In these cases we created 'expert-opinion' range estimates by plotting records on maps and drawing estimated extents of occurrence. We intersected these with our assessment of habitat restrictions (elevation, temperature, precipitation, tree cover; Hurlbert and Jetz 2007) and converted them to a presence-absence grid of the same resolution as SDM-based maps.

Grain size, stacking, and validation of richness patterns

For further analysis we projected range data into Mollweide World equal area projection. To investigate effects of grain size we up-scaled thresholded range estimates (5 km grain) to grid cell sizes of 15, 25, 50, 100, 200 and 400 km. For each grain size, we estimated species richness by summing predicted species for each cell.

Conceptual problems and a trend towards overprediction from stacking thresholded range data have been noted (Calabrese et al. 2014), leading to the recommendation that raw SDM data be used. To investigate this, we estimated species richness as the sum of (logistic) output from Maxent models (after dispersal editing, 5 km grain), adding

presence-predictions of the non-modelled species to produce richness data consistent with thresholded data.

While independent evaluation data on species' presence and absence can be used for selected, common species in methodological studies, they do not exist for a data set that contains many rare taxa. Instead, we provide a quasi-independent validation of emergent species richness patterns by defining 'well-sampled' locations, which were then used to test model-derived richness predictions.

We defined 52 'well-sampled' 5 km cells (Fig. 2) and correlated observed species richness (disregarding records for non-permanent populations) against predictions from stacked range maps. We used observed species richness of larger cells containing those 'well-sampled' 5 km cells to test predictions at larger grains. Record density showed strong spatial clumping, so at larger grain size these regions were also thoroughly sampled (see Supplementary material Appendix 1 for detail).

Environment-richness correlations and niche assessments

We carried out analyses of species richness and climatic niche space at a grain size of 1° equal-area cells (ca 110 km, based on our 100 km grain data). We restricted analyses to cells containing only land (to avoid area effects), but repeated relevant analyses with all cells containing ≥ 20 percent land to avoid missing effects due to the exclusion of coastal areas. We used actual evapotranspiration (AET), topographic range, and mean annual temperature as predictors. For multivariate modelling, data were log-transformed and standardized to a mean of zero and unit SD. We also acknowledged possible effects of evolutionary history by including a categorization of biogeographic realms (as binary dummy variables; Table 1, see Supplementary material Appendix 1 for detail). We used general linear models (ordinary least squares, OLS, no interactions) as well as spatially explicit simultaneous

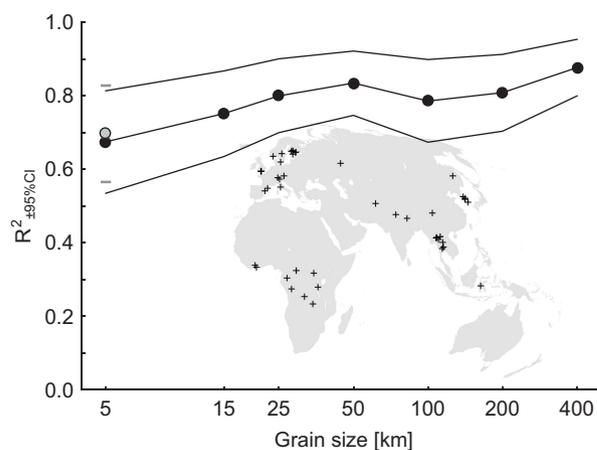


Figure 2. Validation of species richness estimates. Richness predicted by stacked distribution maps of different grain is related to observed values in 52 well-sampled cells (black crosses in map; see Supplementary material Appendix 1 for definition). Black symbols refer to thresholded data; grey refers to summed SDM output (at 5 km grain). Supplementary material Appendix 1, Table A1 shows details and extended test results.

Table 1. Multivariate models, restricted to land-only cells (110 km grain size, standardized data; see Fig. 3 for graphics, Supplementary material Appendix 4 for inclusion of all cells $\geq 20\%$ land). Ordinary least squares (OLS) as well as spatial explicit simultaneous autoregressive (SAR) models were fitted.

n = 6312 cells	OLS: $R^2_{adj} = 0.778$; AIC = 8237.1			SAR: $R^2_{pseudo} = 0.923$; AIC = 1863.1		
	Coefficient \pm SE	t	p	Coefficient \pm SE	z	p
Intercept	0.014 \pm 0.010	1.38	0.167	-2.470 \pm 2.411	-1.03	0.306
AET	0.665 \pm 0.008	87.05	<0.001	0.399 \pm 0.007	55.87	<0.001
Temperature	0.174 \pm 0.008	20.83	<0.001	0.096 \pm 0.015	6.49	<0.001
TopoRange	0.198 \pm 0.006	31.25	<0.001	0.125 \pm 0.005	25.49	<0.001
Palaearctic	0*			0*		
Afrotropics	0.262 \pm 0.021	12.20	<0.001	0.660 \pm 0.027	24.23	<0.001
Indomalaya	0.353 \pm 0.028	12.55	<0.001	-0.009 \pm 0.030	-0.30	0.763
Australasia	-0.803 \pm 0.025	-32.70	<0.001	16.720 \pm 8.105	2.06	0.039

*by default.

autoregressive models (SAR, 1500 km neighborhood; R package ‘spdep’) to investigate relationships between richness and environmental predictors. To assess SAR model fit, we calculated pseudo- R^2 s from correlations of model prediction and observed data. We repeated environment–richness modelling separately for the two largest tribes.

As a broad characterization of the climatic niche of species we extracted data on mean annual temperature and annual precipitation in the cells where a species is predicted to occur. We aggregated these as median and the range from 5 to 95% percentiles to characterize niche position and niche breadth. We used these to compare trends in climatic niche space of taxonomic tribes, juxtaposing them with maps of tribal species richness. Additionally, we analysed the niche overlap of tribes using a randomization method that allows assessing overlap in multivariate niche space and testing the similarity and equivalence of niches (on tribal level). We used kernel densities of species’ climatic niche placement to visualize a clade’s preference. Using linear models of temperature and precipitation we projected these densities onto geographic space and correlated them with species richness patterns.

Results

Creating geographic range estimates for all Old World spingid species

Properties of raw data and geographic range predictions

We used 109 880 distribution records for the analysis of 981 species, of which almost half were mobilized de novo from specimen label information (Supplementary material Appendix 2). A ‘record’ indicates a unique combination of locality, year and collector or source, but it may contain one or many specimens.

Almost 80 percent of our data were georeferenced with an estimated precision $\leq 0.01^\circ$ latitude and longitude, while < 3 percent were ‘unspecific’. Notably, 39 percent of records represented spatial replicates (i.e. the species was already known from the same locality yet from a different year or a different collector or source). When considered at the grain size used for modelling (2.5 arcmin $\approx 5 \times 5$ km), the number of spatially unique records shrank to < 50 000 (ca 45% of the original records). Information on collection year was available for ca 76 200 records ($\approx 69\%$)

with most (ca 63 600 $\approx 83\%$) from after 1950, and roughly half from 1989 or later. The number of records per species closely followed a lognormal distribution, indicating the dominance of rare and poorly-known taxa (Supplementary material Appendix 1, Fig. A1).

We computed and edited SDMs for 789 species and provided expert range estimates for the remaining 192 species (Fig. 1). Annual temperature range and precipitation variables were the most important predictors for range models, whereas landcover variables had only little additional effect on predictions (Supplementary material Appendix 1, Fig. A1).

Evaluating range models and richness patterns

Standard SDM evaluation metrics indicated successful modeling (i.e. median $AUC_{Mx} = 0.94$; 90 percent of models had $AUC_{Mx} \geq 0.8$; Fig. 2). See Supplementary material Appendix 1, Fig. A1 for AUC and threshold distribution, and Supplementary material Appendix 6 for comparison to alternative thresholds.

We validated species richness estimates from stacking thresholded SDMs with data from ‘well-sampled’ regions. The fits were very strong (Fig. 2; Supplementary material Appendix 1, Table A1). Adjusted correlation coefficients increased from $r^2 = 0.68$ for a grain size of 5 km to $r^2 = 0.88$ for 400 km cells, and this trend was unaffected by spatial autocorrelation. Species richness from thresholded SDMs overpredicted observed data slightly and consistently across all grains (slopes < 0.8). Fits of non-thresholded SDM output were similar, but provided an underprediction of observed data to a similar degree as thresholded data overpredicted (slope = 1.3). Patterns of species richness estimates from both methods were very similar (5 km grain; linear regression: $n = 9706$ random points, $r^2 = 0.97$).

Species richness and its relationship with environmental energy availability

Richness patterns

Overall species richness patterns, irrespective of grain size, show a concentration of diversity in the moist tropics (Fig. 1G and 3; correlation of absolute latitude and richness, Spearman’s $\rho = -0.495$). Noteworthy geographic patterns are a ‘finger’ of high richness stretching from southeast-Asia

along the southern Himalayan foothills, a drop in species numbers from the continental plate of tropical Asia towards Australasia, and outstanding richness in montane east Africa at coarse but not fine grains (indicating effects of species turnover).

When deconstructed to the tribal level (Fig. 3), richness patterns of Macroglossini and Smerinthini exhibit almost identical patterns as data for all sphingids (502 species; Spearman $\rho \geq 0.9$), and those of Sphingini were very similar (82 species; $\rho = 0.81$). However, some of the less species-rich groups differ from pooled data. Ambulycini are mostly restricted to the Oriental region (61 species; $\rho = 0.59$), Acherontiini richness peaks on isolated Madagascar and Sulawesi (13 species; $\rho = 0.78$), and Dilophonotini (25 species; $\rho = 0.49$) and Sphingulini (26 species; $\rho = 0.22$) have distinctive richness peaks in temperate Asia. As a result from the declines of other tribes Macroglossini increasingly dominate assemblages toward desert and arctic biomes and in isolated regions. Sphingulini make a relatively high contribution to assemblages in Australia.

Environmental associations

Individual species' positions in climatic niche space (annual mean temperature vs annual precipitation, Fig. 3D) indicate that most species ranges are centred in warm and wet conditions. This matches the observation that cells in moist-tropical regions also featured highest species richness (Fig. 3A; see Supplementary material Appendix 5 for quantitative exploration). In fact, AET was by far the strongest univariate predictor of species richness (Fig. 3B; see Supplementary material Appendix 4 for separate analyses of Macroglossini and Smerinthini). It was also consistently the strongest predictor in multivariate models (Fig. 3C; Table 1), which accounted for $> 75\%$ of variation in richness. Effects of temperature and topographic heterogeneity on species richness were clearly weaker.

The seven tribes show distinct differences in their climatic niches (Fig. 3E), supporting the importance of niche conservatism in the dimensions analyzed. For median temperatures, across-species means of Dilophonotini and Sphingulini are consistently lower (15°C) than for other tribes ($> 20^\circ\text{C}$; Anova: $n = 972$, $F_{\text{DF}=6} = 12.2$, $p < 0.001$), whereas no differences in temperature ranges were found ($p \geq 0.40$) for any tribe group. For median precipitations, Ambulycini and Macroglossini have higher across-species means (> 1.6 m) than other tribes (< 1.4 m; $F_{\text{DF}=6} = 8.4$, $p < 0.001$), and Acherontiini occur under a much wider range of precipitations (> 1.9 m) than other tribes (< 1.5 m; $F_{\text{DF}=6} = 5.9$, $p < 0.001$). A more rigorous randomization test of pairwise niche similarity (in temperature–precipitation space; Supplementary material Appendix 5) showed that tribes are significantly less similar in pairwise comparisons than expected by chance. This suggests divergent niche evolution in these clades.

Among-clade niche differences notwithstanding, we found that, consistently for most tribes, AET was a substantially stronger predictor of richness than temperature (Fig. 3E, Supplementary material Appendix 4), in line with the results for all sphingids. Exceptions to this were Dilophonotini and Sphingulini, which exhibit distinctively non-tropical peaks of richness. These species-poor groups generally cor-

responded poorly to climatic predictors of richness or niche preferences (Fig. 3E, Supplementary material Appendix 5). Biogeographic regions had different effects on the models of tribal species richness (Supplementary material Appendix 4), reflecting a signal of different phylogeographic history of the groups.

Richness data are available for 1° equal-area polygons (for use in a Geographic Information System GIS; Supplementary material Appendix 7) while individual species' range estimates can be browsed at Map of Life (www.mol.org).

Discussion

Our expert compilation and model predictions of sphingid moth ranges enabled a macroecological assessment of an insect group at unprecedented detail and extent, spanning four continents. We uncovered, across multiple spatial grains, substantial variation in geographic variation in richness. Regions of highest richness were generally located in the moist tropics, with tribal-level deconstruction highlighting important divergence to this for some sub-groups. We identified productivity as the strongest environmental predictor of richness, while a link with ambient temperature (expected for an ectotherm taxon according to ideas on metabolic effects) was only weakly supported. These findings underscore the importance of broadening the taxonomic spectrum beyond the four main terrestrial vertebrate taxa when evaluating energy effects on richness. In particular, reptiles and amphibians alone may not well represent all possible ectothermic adaptations of organisms.

Data quality and distribution modelling

We developed range predictions using a combined strategy of expert knowledge and species distribution modelling (SDM). SDMs have seen explosive development and application over the last decade (Elith and Leathwick 2009, Radosavljevic and Anderson 2014), and scientists' acceptance ranges from suggestions of automated (i.e. unsupervised, possibly uncritical) applications to substantial scepticism (Beale and Lennon 2012, Joppa et al. 2013). New methods are continually being developed and old approaches challenged (e.g. thresholding, Supplementary material Appendix 6; Calabrese et al. 2014). Cautious use that includes expert-based quality control and feedback, however, may offer a pragmatic way to close distribution data gaps for poorly known taxa and provide first baseline biodiversity estimates that help ameliorate the ongoing knowledge bias toward vertebrate taxa. Expert edits would ideally be captured and made available for downstream uses as well as further review and updating, to transparently advance species distribution knowledge (Jetz et al. 2012).

Environment and the causes of richness patterns

The proxy for net primary productivity (AET) was by far the strongest predictor of species richness patterns in our data. An effect of productivity on population sizes has been widely discussed as a plausible mechanism affecting extinction rates

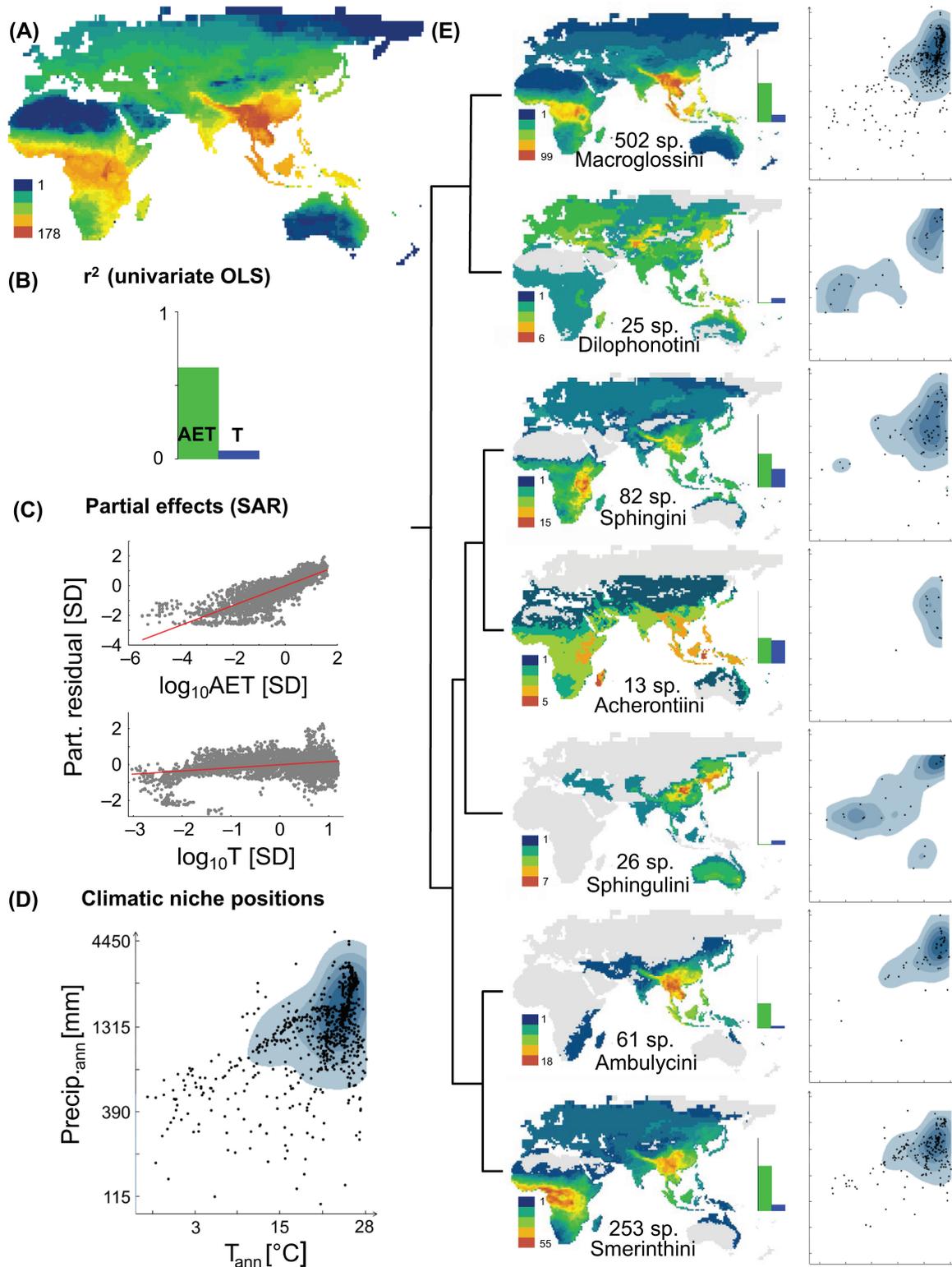


Figure 3. Total richness, environmental associations and niche positions of 972 species, for all species and partitioned by subclade. (A) Richness in 1°-cells (cells < 1/3 land area excluded). The color scheme is the same as in Fig. 3. (B) Strengths of univariate regressions of AET (green) and temperature (blue) with species richness (OLS, land-only cells). (C) Partial effects of AET and temperature in multivariate SAR model (Table 1; each dot is one grid cell). (D) Climatic niche positions (temperature, precipitation). Each dot in the plots represents a species, defined by the median of conditions where it occurs; density kernels visualize the climates most commonly occupied by the group (search radius 1 SD; only above-average densities are shown). (E) Deconstruction of richness patterns (grey cells: absence of a tribe), univariate richness correlations and niche positions for the seven tribes (with topology of phylogeny). All axis scaling is identical to those of data for all sphingids (D).

and species richness (Mittelbach et al. 2001, Willig et al. 2003, Evans et al. 2005), albeit with scale-dependent changes in shape. Strong additional effects of temperature on assemblage richness due to metabolic effects have been hypothesized for ectotherms (Allen et al. 2007), but we found only weak support for this idea (Hawkins et al. 2007). Current understanding of how thermoregulation affects the relative importance of environmental drivers of species richness and other phenomena is based almost exclusively on vertebrate studies (Buckley et al. 2012). Unification of ideas about the role of physiology in governing these patterns rests on a taxonomic broadening of the empirical support.

For example, particularly strong effects of air temperature data can be expected in amphibians as their water-permeable skin prevents sunbasking, an efficient means of behavioural thermoregulation in reptiles as well as many terrestrial invertebrates (Shine 2005). Strong, rather than weaker, temperature–richness links in reptiles compared to amphibians (Buckley et al. 2012) contradicts this expectation, possibly due to taxon-specific idiosyncrasies (Powney et al. 2010). In sphingids, behavioural thermoregulation is unlikely (most adults are nocturnal, and caterpillar sunbasking appears restricted to few Lepidoptera species with very efficient anti-predation adaptations; Ruf and Fiedler 2002, McClure et al. 2011). The above reasoning would therefore predict rather strong temperature effects on sphingid metabolism and, by purported extension, richness – but this was not supported empirically. As the number of exceptions to the theoretical prediction increases, a re-evaluation of ideas around the temperature–metabolism–richness connection may be warranted.

However, at spatial scales smaller than treated here (e.g. elevational transect samples), some studies on terrestrial insects (in particular, hymenopterans) supported strong temperature effects on richness patterns (Sanders et al. 2007, Dunn et al. 2009, Classen et al. 2015), whereas others reported links with productivity (Kaspari et al. 2000). Our preliminary analyses of sphingid elevational richness patterns across the study region are consistent with strong statistical effects of temperature as well as productivity (Beck et al. unpubl.). Hawkins et al. (2003) reported correlations of water availability with richness in a number of insect datasets, while effects of temperature were restricted to northern regions. They suggested this may be due to water availability affecting plant productivity, which was not directly measured in many datasets used in Hawkins et al. (2003). Alternatively, they hypothesized on heightened desiccation risk for small-bodied insects (in comparison to, e.g. larger-bodied lizards), which may explain links with water in hot regions.

In sphingids, a particular role for primary productivity on individual growth, hence generation times and diversification rates, may be related to their larval growth adaptations. Broadly speaking, sphingid caterpillars follow a ‘grow fast, use resources wastefully’-syndrome. Typically, they feed on fresh, nutritious leaves of succession plants, whereas related ‘big moth’ taxa such as saturniids generally feed on less nutritious old leaves and grow more slowly (Janzen 1984). It would be premature to claim that these life history adaptations are part of the cause for strong productivity effects on richness patterns, but these ideas highlight the need for

broader taxonomic comparisons (e.g. with saturniids) and also for establishing trait databases for insect groups to link macroecological patterns with physiological mechanisms (Kingsolver et al. 2011).

The explanatory fits we find for environmental predictors of sphingid richness are comparable to those found for other taxa at similar extents and resolutions (Kreft and Jetz 2007), as were the directions of modelled effects. This, as well as the similar results for the constituent tribes *Macroglossini* and *Smerinthini*, supports the view that large-scale environment–richness relationships are not strongly dependent on the studied taxon as long as a group is sufficiently inclusive not to be affected by idiosyncrasies of its evolutionary history (McArthur 1972, Jetz et al. 2009).

However, it becomes increasingly clear that understanding the causes of richness patterns requires a historical perspective (Wiens and Donoghue 2004, Fritz et al. 2013). Recent studies (Pyron and Wiens 2013) have begun to link phylogenies and environmental data to disentangle effects of speciation, extinction and dispersal rates, but such approaches require even more extensive support with data (e.g. distribution, phylogeny, present and past environment). To date, only scattered knowledge on the historical biogeography of sphingid moths is available. Based on a phylogenetic reconstruction of tribes and subtribes, multiple separate origins of the radiations in the Old World and the Americas have been postulated (Kawahara et al. 2009), whereas the sensitivity of Oriental *Smerinthini* to dispersal barriers, leading to a rapid decline of the group’s species richness towards Melanesia and Australia, has also been noted (Beck et al. 2006a). Data presented in Fig. 3, although not phylogenetically explicit beyond tribal level, clearly map differences in species richness ‘hot spots’ of the different clades. These data indicate deviating niche space for some tribes (implying effects of niche conservatism), but notably there are no large differences between the species-rich tribes despite substantially different life histories. This is consistent with finding no major differences in environmental effects on richness pattern of the two largest groups (Supplementary material Appendix 4).

Distribution modelling rendered annual temperature range (i.e. seasonality) as the most important predictor variable across taxa (Supplementary material Appendix 1, Fig. A1). This invites interpretation towards limited spread of tropical taxa into seasonal climates (due to a lack of adequate adaptations, such as diapause). However, because 1) climate variables used for SDM were collinear, and 2) Maxent-models are designed and optimized for prediction, not for hypothesis testing, these observations should not be overinterpreted. Further, hypothesis-driven testing (ideally within a phylogenetic framework) will be needed to substantiate our speculations.

Conclusions

We show that a careful compilation of sphingid distribution records, in combination with expert-edited ‘supervised’ SDMs, can provide reliable first baseline estimates of global-scale species distribution data at a spatial resolution comparable to that of terrestrial vertebrates. The methodological recommendations and findings developed here may be applicable

to other invertebrates and have the potential to support a taxonomically more inclusive approach to biodiversity science. Investigating a broader set of taxa with different adaptations is necessary to explore and test general theories, e.g. by linking thermoregulation with the strength of the temperature–richness relationships as discussed above. Putting insects on the map is therefore a timely step to make more use of Earth’s organismic diversity to understand the spatial patterns of species richness.

We were able to document remarkably strong geographic variation in species richness, with centers of richness in the wet tropics of the African and Oriental region. Data at tribal level, however, showed idiosyncratic richness peaks in other regions for some groups, indicating important effects of variation in climatic niches, dispersal limitation and different evolutionary centres of origin. Furthermore, we documented an intuitive link between the densities of species-level climatic niches and richness patterns. Deconstructing this on a tribal level illustrated how phylogeny, adaptations and richness patterns are intrinsically linked to each other.

Richness patterns and their environmental correlates revealed strong links with energy availability (i.e. AET), whereas the predicted effect of temperature was, at the scale of analysis, weaker and less robust to analytical approaches. We found the environmental associations of total richness to be similar among many of the investigated clades, notably between the two largest sphingid subgroups despite their strikingly divergent life histories. This robustly supports the importance of potential energy availability in particular in shaping sphingid species richness variation at this scale.

Acknowledgements – We are very grateful with all those amateurs and professional collectors who have made available their collections to use within this project. M. Curran, M. Kopp, R. Hagmann, S. Lang and S. Widler helped with georeferencing, G. Amatulli and P. Vogt with GIS and IT. We thank K. Fiedler for valuable input. The study received financial support from the Swiss National Science Foundation (SNF grant 3100AO_119879 & ISV program), US National Science Foundation grants DEB 1026764, and DEB 1441737 and DBI 1262600, NASA Biodiversity grant NNX11AP72G, the Synthesys program and the Freiwillige Akademische Gesellschaft (FAG) Basel.

References

- Allen, A. et al. 2007. Recasting the species–energy hypothesis: the different roles of kinetic and potential energy in regulating biodiversity. – In: Storch, D. et al. (eds), *Scaling biodiversity*. Cambridge Univ. Press, pp. 1–29.
- Basset, Y. et al. 2012. Arthropod diversity in a tropical forest. – *Science* 338: 1481–1484.
- Beale, C. M. and Lennon, J. J. 2012. Incorporating uncertainty in predictive species distribution modelling. – *Phil. Trans. R. Soc. B* 367: 247–258.
- Bean, W. T. et al. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. – *Ecography* 35: 250–258.
- Beck, J. and Kitching, I. J. 2007. Correlates of range size and dispersal ability: a comparative analysis of sphingid moths from the Indo-Australian tropics. – *Global Ecol. Biogeogr.* 16: 341–349.
- Beck, J. et al. 2006a. Wallace’s line revisited: has vicariance or dispersal shaped the distribution of Malaysian hawkmoths (Lepidoptera: Sphingidae)? – *Biol. J. Linn. Soc.* 89: 455–468.
- Beck, J. et al. 2006b. Diet breadth and host plant relationships of southeast-Asian sphingid caterpillars. – *Ecotropica* 12: 1–13.
- Beck, J. et al. 2012. What’s on the horizon for macroecology? – *Ecography* 35: 1–11.
- Beck, J. et al. 2013. Online solutions and the “Wallacean shortfall”: what does GBIF contribute to our knowledge of species’ ranges? – *Divers. Distrib.* 19: 1–8.
- Beck, J. et al. 2014. Spatial bias in the GBIF database and its effect on modeling species’ geographic distributions. – *Ecol. Inform.* 19: 10–15.
- Buckley, L. B. et al. 2010. Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. – *Proc. Natl Acad. Sci. USA* 277: 2131–2138.
- Buckley, L. B. et al. 2012. Broad-scale ecological implications of ectothermy and endothermy in changing environments. – *Global Ecol. Biogeogr.* 21: 873–885.
- Calabrese, J. M. et al. 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. – *Global Ecol. Biogeogr.* 23: 99–112.
- Classen, A. et al. 2015. Temperature versus resource constraints: which factors determine bee diversity on Mount Kilimanjaro, Tanzania? – *Global Ecol. Biogeogr.* 24: 642–652.
- Currie, D. J. et al. 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. – *Ecol. Lett.* 7: 1121–1134.
- Dunn, R. R. et al. 2009. Climatic drivers of hemispheric asymmetry in global patterns of ant species richness. – *Ecol. Lett.* 12: 324–333.
- Elith, J. and Leathwick, J. 2009. Species distribution models: ecological explanation and prediction across space and time. – *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Engler, R. et al. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. – *J. Appl. Ecol.* 41: 263–274.
- Evans, K. L. et al. 2005. Species–energy relationships at the macroecological scale: a review of the mechanisms. – *Biol. Rev.* 80: 1–25.
- Fritz, S. A. et al. 2013. Diversity in time and space: wanted dead and alive. – *Trends Ecol. Evol.* 28: 509–516.
- Godfray, H. C. et al. 1999. Studying insect diversity in the tropics. – *Phil. Trans. R. Soc. B* 354: 1811–1824.
- Hamilton, A. J. et al. 2013. Estimating global arthropod species richness: refining probabilistic models using probability bounds analysis. – *Oecologia* 171: 357–365.
- Hawkins, B. A. et al. 2003. Energy, water, and broad-scale geographic patterns of species richness. – *Ecology* 84: 3105–3117.
- Hawkins, B. A. et al. 2007. A global explanation of metabolic theory as an explanation for terrestrial species richness gradients. – *Ecology* 88: 1877–1888.
- Heinrich, B. 1993. The hot-blooded insects: strategies and mechanisms of thermoregulation. – Harvard Univ. Press.
- Hurlbert, A. H. and Jetz, W. 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. – *Proc. Natl Acad. Sci. USA* 104: 13384–13389.
- Hurlbert, A. H. and Jetz, W. 2010. More than “more individuals”: the nonequivalence of area and energy in the scaling of species richness. – *Am. Nat.* 176: E50–E65.
- ICZN: International Code of Zoological Nomenclature 1999. International Trust for Zoological Nomenclature, 4th ed. – London.
- Janzen, D. H. 1984. Two ways to be a tropical big moth: Santa Rosa saturniids and sphingids. – In: Dawkins, R. and Ridley, M. (eds), *Oxford surveys in evolutionary biology*. Oxford Univ. Press, pp. 85–140.

- Jetz, W. and Fine, P. V. A. 2012. Global gradients in vertebrate diversity predicted by historical area-productivity dynamics and contemporary environment. – *PLoS Biol.* 10: e1001292.
- Jetz, W. et al. 2009. Global associations between terrestrial producer and vertebrate consumer diversity. – *Proc. Natl Acad. Sci. USA* 276: 269–278.
- Jetz, W. et al. 2012. Integrating biodiversity distribution knowledge: toward a global map of life. – *Trends Ecol. Evol.* 27: 151–159.
- Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. – *Global Ecol. Biogeogr.* 21: 498–507.
- Joppa, L. N. et al. 2013. Troubling trends in scientific software use. – *Science* 340: 814–815.
- Kaspari, M. et al. 2000. Energy, density, and constraints to species richness: ant assemblages along a productivity gradient. – *Am. Nat.* 155: 280–293.
- Kawahara, A. Y. et al. 2009. Phylogeny and biogeography of hawkmoths (Lepidoptera: Sphingidae): evidence from five nuclear genes. – *PLoS One* 4: e5719.
- Keil, P. et al. 2013. Downscaling of species distribution models: a hierarchical approach. – *Methods Ecol. Evol.* 4: 82–94.
- Kingsolver, J. G. et al. 2011. Complex life cycles and the responses of insects to climate change. – *Integr. Comp. Biol.* 51: 719–732.
- Kitching, I. and Cadiou, J. M. 2000. Hawkmoths of the world. – The Natural History Museum London and Cornell Univ. Press.
- Krefl, H. and Jetz, W. 2007. Global patterns and determinants of vascular plant diversity. – *Proc. Natl Acad. Sci. USA* 104: 5925–5930.
- May, R. 1994. Conceptual aspects of the quantification of the extent of biological diversity. – *Phil. Trans. R. Soc. B* 345: 13–20.
- McArthur, R. H. 1972. Geographical ecology: patterns in the distribution of species. – Harper and Row.
- McClure, M. et al. 2011. Thermal ecology and behaviour of the nomadic social forager *Malacosoma disstria*. – *Physiol. Entomol.* 36: 120–127.
- Merckx, B. et al. 2011. Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. – *Ecol. Model.* 222: 588–597.
- Mittelbach, G. G. et al. 2001. What is the observed relationship between species richness and productivity? – *Ecology* 82: 2381–2396.
- Pearson, D. L. and Cassola, F. 1992. World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): indicator taxon for biodiversity and conservation studies. – *Conserv. Biol.* 6: 376–391.
- Phillips, S. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Phillips, S. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – *Ecol. Appl.* 19: 181–197.
- Powney, G. D. et al. 2010. Hot, dry and different: Australian lizard richness is unlike that of mammals, amphibians and birds. – *Global Ecol. Biogeogr.* 19: 386–396.
- Pyron, R. A. and Wiens, J. J. 2013. Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. – *Proc. Natl Acad. Sci. USA* 280: 20131622.
- Radosavljevic, A. and Anderson, R. P. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. – *J. Biogeogr.* 41: 629–643.
- Ruf, C. and Fiedler, K. 2002. Tent-based thermoregulation in social caterpillars of *Eriogaster lanestris* (Lepidoptera: Lasiocampidae): behavioral mechanisms and physical features of the tent. – *J. Therm. Biol.* 27: 493–501.
- Sanders, N. J. et al. 2007. Temperature, but not productivity or geometry, predicts elevational diversity gradients in ants across spatial grains. – *Global Ecol. Biogeogr.* 16: 640–649.
- Shine, R. 2005. Life-history evolution in reptiles. – *Annu. Rev. Ecol. Evol. Syst.* 36: 23–46.
- Stork, N. E. 1988. Insect diversity: facts, fiction and speculation. – *Biol. J. Linn. Soc.* 35: 321–337.
- Terribile, L. C. et al. 2009. Richness patterns, species distributions and the principle of extreme deconstruction. – *Global Ecol. Biogeogr.* 18: 123–136.
- Wieczorek, J. et al. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. – *Int. J. Geogr. Inform. Sci.* 18: 745–761.
- Wiens, J. J. and Donoghue, M. J. 2004. Historical biogeography, ecology and species richness. – *Trends Ecol. Evol.* 19: 639–644.
- Willig, M. R. et al. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. – *Annu. Rev. Ecol. Evol. Syst.* 34: 273–309.

Supplementary material Appendix ECOG-02438 at <www.ecography.org/appendix/ecog-02438>. Appendix 1–7.