Diversity and Distributions, (Diversity Distrib.) (2015) **21**, 345–356

A Journal of Conservation Biogeography

Diversity and Distributions

BIODIVERSITY RESEARCH

Global diversity patterns of freshwater fishes – potential victims of their own success

Patricia Pelayo-Villamil[1], Cástor Guisande[2]*, Richard P. Vari[3], Ana Manjarrés-Hernández[4], Emilio García-Roselló[5], Jacinto González-Dacosta[5], Jürgen Heine[5], Luis González Vilas[2], Bernardo Patti[6], Enza María Quinci[6], Luz Fernanda Jiménez[1], Carlos Granado-Lorencio[7], Pablo A. Tedesco[8] and Jorge M. Lobo[9]

[1]*Grupo de Ictiología, Universidad de Antioquia, A.A. 1226, Medellín, Colombia,* [2]*Facultad de Ciencias, Universidad de Vigo, Lagoas-Marcosende, 36200 Vigo, Spain,* [3]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, MRC 159, Washington, D.C., USA,* [4]*Instituto Amazónico de Investigaciones-IMANI, Universidad Nacional de Colombia, A.A. 215, Leticia, Colombia,* [5]*Departamento de Informática, Edificio Fundición, Universidad de Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain,* [6]*Istituto per l'Ambiente Marino Costiero, U.O. Capo Granitola, Consiglio Nazionale delle Ricerche, Via del Mare 3, Campobello di Mazara, TP 91021, Italy,* [7]*Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, Sevilla, Spain,* [8]*Département Milieux et Peuplements Aquatiques, Muséum National d'Histoire Naturelle, UMR Biologie des Organismes et des Ecosystèmes Aquatiques (UMR BOREA, IRD 207-CNRS 7208-UPMC-MNHN), 43 rue Cuvier, 75231 Paris Cedex, France,* [9]*Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (CSIC), c/José Gutiérrez Abascal 2, 28006 Madrid, Spain*

*Correspondence: Cástor Guisande, Facultad de Ciencias, Universidad de Vigo, Lagoas-Marcosende, Vigo 36200, Spain.
E-mail: castor@uvigo.es

ABSTRACT

Aim To examine the pattern and cumulative curve of descriptions of freshwater fishes world-wide, the geographical biases in the available information on that fauna, the relationship between species richness and geographical rarity of such fishes, as well as to assess the relative contributions of different environmental factors on these variables.

Location Global.

Methods ModestR was used to summarize the geographical distribution of freshwater fish species using information available from data-based geographical records. The first-order jackknife richness estimator was used to estimate the completeness of such data in all terrestrial 1-degree cells world-wide. An α-shape procedure was used to build range maps capable of providing relatively accurate species richness and geographical rarity values for each grid cell. We also examined the explanatory capacity of a high number of environmental variables using multiple regressions and Support Vector Machine.

Results Cumulative species description curves show that a high number of species of freshwater fishes remain to be discovered. Completeness values indicate that only 199 one-degree grid cells, mainly located in eastern North America and Europe, could be considered as having relatively accurate inventories. Range maps provide species richness values that are positively and significantly related to those resulting from the first-order jackknife richness estimator. The relationship between species richness and geographical rarity is triangular, so that these species-rich cells are those with a higher proportion of distributionally rare species. Species richness is predicted by climatic and/or productivity variables but geographical rarity is not.

Main conclusions In general, species-rich tropical areas harbour a higher number of narrowly distributed species although comparatively species-poor subtropical cells may also contain narrowly distributed species. Historical factors may help to explain the faunistic composition of these latter areas; a supposition also supported by the low predictive capacity of climatic and productivity variables on geographical rarity values.

Keywords
Description curve, freshwater fish species richness, geographical rarity, historical factors, world distribution.

DOI: 10.1111/ddi.12271
http://wileyonlinelibrary.com/journal/ddi

INTRODUCTION

Knowledge of the patterns of species richness and rarity is fundamental for basic and applied purposes in all taxonomic groups world-wide (Gaston & Blackburn, 2000). Nonetheless, obtaining even a preliminary picture of these fundamental variables is a task plagued by uncertainties due to the provisional nature of much of the available taxonomic and distributional information, the scarcity of such data in readily accessible forms and the inherent biases in the available information (Rocchini *et al.*, 2011; Gaiji *et al.*, 2013). The prerequisite for the study of biodiversity patterns involves compiling to the maximum degree possible the pertinent data dispersed across databases, housed in natural history museums and collections and dispersed across the taxonomic literature. A common procedure is to first standardize and clean the taxonomic nomenclature to agreed-upon standards. This taxonomic information can be used to calculate the rates of species descriptions as well as provide approximate estimates of the number of remaining undescribed species (Costello *et al.*, 2012). Furthermore, the georeferencing of locations, when possible, allows us to develop point-to-grid or range maps able to represent the observed distribution of each species. As this distributional information can be plagued with uncertainties and biases, it is first necessary to discriminate regions with relatively accurate inventories from those in need of further surveys (Rocchini *et al.*, 2011). The selection of an adequate spatial resolution capable of minimizing errors as a consequence of data scarcity but, nonetheless, providing relatively reliable patterns is a key step in this process (Graham & Hijmans, 2006; Pineda & Lobo, 2012). Once obtained and evaluated, these individual distributional representations are subsequently often used to describe the basic geographical patterns of species richness and/or species rarity and the relationships, if any, between them (Gaston, 1994). Such maps of biodiversity-relevant variables can be used as a basis for studies of the most probable environmental factors capable of explaining the obtained patterns and to identify probable causal mechanisms (Wright, 1983; Wiens & Donoghue, 2004).

Although information is available for freshwater fishes at the global scale, such data are only accessible at the ecoregional (Abell *et al.*, 2008; Lévêque *et al.*, 2008) or river basin scale (Oberdorff *et al.*, 2011; Tedesco *et al.*, 2012). Further, complicating previous analyses is the very high annual pace of description of freshwater fish species with 1484 species described between 2008 and the end of 2013 (Mikšik & Schraml, 2013). In other words, circa 9.31% of the nearly 14,800 valid species of freshwater fishes recognized at the end of that period were formally described during that 6 year time span. In this study, we carry out an up-to-date global geographical representation of general biodiversity parameters for all freshwater fish species recognized by systematists through mid-2013 (Eschmeyer, 2013). We assemble and utilize a large and updated percentage of the readily available data on world freshwater fishes to (i) determine the probable degree of completeness of our current taxonomic and distributional knowledge, (ii) generate reliable world species richness and geographical rarity maps from the overlay of individual range maps, and (iii) estimate the main factors able to account for these global patterns. Thus, we first examine the pattern of species descriptions and the cumulative description of species living in freshwaters commencing with the first formal descriptions by Linnaeus to describe the magnitude of the so-called Linnaean shortfall, that is the number of fish species expected to exist in freshwaters but not yet described. Next, we tried to identify for the first time the gaps and to show the global biases in the available geographical knowledge about freshwater fishes (i.e. the so-called Wallacean shortfall). In a third step, we use an α-shape method to elaborate overlapped individual range maps to obtain a representation of the world-wide distribution in freshwater fish species richness. Our species richness projections complement those formerly provided by previous authors (Abell *et al.*, 2008; Collen *et al.*, 2014), but in our analysis, we estimated the validity of the obtained species richness representation by comparing their values with those of some areas previously determined to be relatively well-surveyed. At the same time, we also detail the world distribution of geographical rarity for freshwater fishes, examining the relationship between species richness and rarity to shed some light on the kind of processes involved in the current configuration of biodiversity patterns (Gaston & Blackburn, 2000). Lastly, we assess the relative contributions of different environmental factors on the obtained geographical variation of freshwater fishes.

METHODS

Species distributions

We used ModestR (www.ipez.es/ModestR, García-Roselló *et al.*, 2013) to develop the summary range maps for species of freshwater fishes. This was accomplished via a dataset that includes the geographical distribution of all species of freshwater fishes currently recognized as valid by systematists (Eschmeyer, 2013) and available in IPez (www.ipez.es/, Guisande *et al.*, 2010). Included species occur in freshwater, freshwater/brackish and freshwater/brackish/marine habitats (see Appendix S2 in Supporting Information). Species that are known to researchers but undescribed, only potentially present or whose presence in a given area was doubtful were not included.

Geographical records were downloaded from Global Biodiversity Information Facility (GBIF, 2013) using the menu facility of ModestR (Pelayo-Villamil *et al.*, 2012; García-Roselló *et al.*, 2013). Data were cleaned using the same application (García-Roselló *et al.*, 2014) via a process involving the removal of duplicate and erroneous georeferenced records, the correction of erroneous synonyms and false records as well as the exclusion of all the records located in marine environs. GBIF data were supplemented with records obtained from publications in which species were originally described and other taxonomic studies with geographical records, the Catalog of Fishes (Eschmeyer, 2013), Fishbase (Froese & Pauly, 2013) and the databases for fishes in the following: American Museum of Natural History – Vertebrate Zoology (http://sciweb-001.amnh.org/); Coleções Científicas do MCT-PUCRS (http://webapp.pucrs.br/); FishNet2 (http://www.fishnet2.net/); MCZBase: Museum of Comparative Zoology – Harvard University (http://mczbase.mcz.harvard.edu/); Sistema Nacional de informações sobre coleções Ictiológicas SIBIP/NEODAT III (http://www.mnrj.ufrj.br/); Smithsonian Institution, National Museum of Natural History – Division of Fishes Collections (http://collections.mnh.si.edu/); and Swedish Museum of Natural History (http://artedi.nrm.se/). Of the total of 15,939 species in the compilation from these sources, 13,120 species (82.3% of the total) have associated geographical information. On the basis of 835,502 total available geographical records, we developed range maps representing the extent of occurrence (EOO) for each of species for which georeferenced data were available.

Mapping recording effort

The number of observed species in all the terrestrial cells of 1 degree world-wide between the 80° north and −66° south was calculated using all available records (n = 14,492 grid cells). The recommended first-order jackknife richness estimator for incidence data (Hortal *et al.*, 2006) was used to estimate the completeness of these cell inventories, taking into account the number of records available for each species as a survey effort surrogate and considering species as unique in instances of only one record per cell. The *vegan* R package (Oksanen *et al.*, 2013) was used for this purpose. The percentage of observed species in each cell versus the number estimated by this predictor was used as a measure of completeness reflecting the reliability of each 1-degree cell inventory. Only those 1-degree cells with twice as many records as species and a minimum of five observed species are considered in these estimates (n = 2101).

Mapping species richness

As a point-to-grid procedure frequently results in species richness underestimates (Graham & Hijmans, 2006), we built range maps by an α-shape method both to describe the variation in species richness of freshwater fishes across the world and to study the factors associated with this variation. Convex hulls and α-shapes can be used to generate range maps from a finite set of observed occurrences (Pateiro-Lopez & Rodriguez-Casal, 2011), thus delimiting the area contained within the shortest continuous imaginary boundary encompassing all the observed occurrences (EOO; IUCN, 2013). However, the advantage of α-shape is that it minimizes EOO overestimates by incorporating discontinuities in species distributions (Burgman & Fox, 2003); depending on a parameter (α) that regulates the shape of the range size. When α approaches zero, the generated shape is near to the original point set, whereas when α increases we are able to obtain a range map similar to the typical convex hull. In this study, we use an α value of 6 in the ModestR software (García-Roselló *et al.*, 2013), because previous analyses documented that this value is adequate to obtain reliable range maps (unpublished data).

Extent of occurrence

Once EOO has been drawn for each one of the species, we calculate the EOO area (in km^2) with ModestR using the following equation to avoid the biases generated by the use of geographical coordinates:

$$1.852 * \frac{12756.2 * \pi}{21,600} \cos\left(\text{latitude} * \frac{\pi}{180}\right)$$

Subsequently, we estimate the average EOO area values for all the species present in each cell (EOO$_{\text{avg}}$). A low EOO$_{\text{avg}}$ value indicates the occurrence of a higher number of species with small geographical range sizes or geographical rarity (Gaston & Fuller, 2009).

Environmental variables

All the environmental variables used in this study are described in Appendix S1. We examined the association between anthropogenic, topographical, productivity and climatic variables and species richness or EOO$_{\text{avg}}$ values. The relationship with the degree of human impact was examined considering the spatial distribution of human populations. This variable comes from a globally consistent, spatially explicit map based on the Gridded Population of the World dataset, version 3 (GPWv3). To develop the global dataset, national population data are transformed from their native spatial units, which are usually administrative (such as state or county level) and of varying resolutions, to a global grid of quadrilateral, latitude-longitude cells at a resolution of 2.5 arc minutes, and then downscaled to 6 arc minutes. A proportional allocation gridding algorithm, utilizing more than 300,000 national and subnational administrative units, is used to assign population values to the 1-degree grid cells. Population densities show the number of humans per square kilometre based on census data available in 2000 and with estimates when necessary to fill in missing or incomplete data.

Topographical variables come from combining data from NASA's Shuttle Radar Topography Mission covering the land surface from 60 degrees south to 60 degrees north. The data for the rest of the Northern Hemisphere (60–90 degrees north) come from digital elevation models (digital versions of paper-based topographical maps) produced by the US Geological Survey. The data for the remainder of the Southern Hemisphere (60–90 degrees south) come from the 'RAMP II' project of the Radarsat Antarctic Mapping Project

Digital Elevation Model, version 2. Using all these sources, we calculate the fluctuation index of Dubois (1973) modified by Guisande *et al.* (2006) as an indicator of topographical heterogeneity using elevation, slope and slope-aspect (that is defined as the compass direction to which a slope faces measured in degrees) in cells of 5′ × 5′. Each topographical variable was scaled to range between 0 and 1 using the following equation:

$$\frac{(\text{value of the variable in the cell } 5' \times 5' - \text{variable minimum})}{(\text{variable maximum} - \text{variable minimum})}$$

Thus, topographical heterogeneity (*TH*) was finally estimated using the scaled topographical variables as:

$$TH = \sum_{i=1}^{3} p_i \log_2 \frac{p_i}{p_{im}}$$

where i are elevation, slope and aspect, p_i the relative proportion of the transformed variable i in the 5′ × 5′cell, and p_{im} the mean of the p_i considering the cell and surrounding cells. We estimated two indexes *TH8* (8 surrounding cells) and *TH24* (24 surrounding cells). *TH* increases as differences in topography between the cell and surrounding cells are greater. Both indexes are available as ASC files as a 5′ × 5′ raster in the website http://www.ipez.es/ModestR.

The area occupied by river basins within 60′ × 60′ grid cells was also used as a topographical variable following the methodology described by García-Roselló *et al.* (2013). Data were obtained from the web site http://www.openstretmap.org.

Productivity variables come from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard NASA's Terra satellite. Specifically, monthly data of terrestrial net primary productivity and vegetation index from 2001 to 2010 were obtained by averaging available information for each pixel of selected variables using the statistical software R (R Development Core Team, 2013). The net primary productivity indicates how much carbon dioxide is taken up by vegetation during photosynthesis minus how much carbon dioxide is released when plants respire. The values indicate how fast carbon was taken in, or released, for every square metre of land over the indicated time span. Values range from −1.0 g of carbon per square metre per day to 6.5 g per square metre per day. A negative value means decomposition or respiration exceeded carbon absorption; in other words, more carbon was released into the atmosphere than was absorbed by the plants. We also include the vegetation index as a productivity variable. This variable represents a measure of the greenness of Earth's landscapes.

Finally, the 19 bioclimatic variables of the WorldClim dataset from 1950 to 2000 (Hijmans *et al.*, 2005) were also used (see Appendix S1) to account for the generally well-established influence of climatic conditions on global biodiversity variations.

Statistical analyses

To estimate how far we are from relatively complete inventories, the Naperian logarithm of the accumulated number of species was related to each year using the software CurveExpert 1.4 (www.curveexpert.net) in which a large number of regression models can be simultaneously compared (linear, exponential, power, sigmoidal, growth, etc.) to find the model with the highest explanatory capacity.

Data were analysed using both classical multiple regressions and machine learning, Support Vector Machine (SVM) procedures. Dependent and independent variables were log-transformed in all models. In so far as the analysis covered the world, it encompassed areas of zero freshwater fish species richness (e.g. deserts, glaciated regions, very high elevation settings).

Stepwise multiple regressions were performed with *stats* R package (R Development Core Team, 2013). The relative contribution of each variable in the regressions was estimated with the LMG method (the r^2 contribution averaged over orderings among explanatory variables) with the R package *relaimpo* (Grömping, 2006). Kolmogorov–Smirnov test with Lilliefors correction was used to test for normality of residuals and performed with the package *nortest* (Gross, 2013). The independence of residuals was measured with the Durbin–Watson statistic and their homoscedasticity with the Breusch–Pagan test, both included in the package *lmtest* (Zeileis & Hothorn, 2002; Hothorn *et al.*, 2013). The presence of multicollinearity–linear relationship among predictors – was measured estimating the variance inflation factor (VIF; Fox & Weisberg, 2011) by means of the *car* package (Fox *et al.*, 2013).

Support Vector Machines are supervised learning models with associated learning algorithms that analyse data and recognize patterns. These are used for classification and regression analysis with a comparative good performance over other methods as a consequence of their capacity for processing nonlinear relationships (Meyer *et al.*, 2003). SVM enjoys excellent theoretical properties and good performance under very general conditions. No strict assumptions are needed. We used the function ksvm of the package *kernlab* (Karatzoglou *et al.*, 2004). SVM has been successfully applied to explain the factors affecting species richness of marine elasmobranchs (Guisande *et al.*, 2013).

RESULTS

Rate of species description

Nelson (2006) cited 11,952 strictly freshwater fish species in the world and 12,457 species using freshwater habitats at some time during their life cycle. Lévêque *et al.* (2008), in turn, reported 12,740 freshwater species. More recently, Vega & Wiens (2012) cited 15,150 species of the class Actinopterygii as inhabiting habitats including freshwaters. As of December 2013, we found 14,782 described species in IPez that occur solely in freshwater habitats, 540 species that inhabit freshwater/brackish waters and 617 species that range across freshwater/brackish/marine habitats. Thus, a total of 15,939 valid species names would be associated with freshwater environments. In the case of the strictly freshwater species, this rate of description is still very high (240.2 species per year across the last 10 years), and the cumulative description curve does not show a clear asymptotic tendency, suggesting that a high number of species remain to be described (Fig. 1). Since the first formal description of such species of fishes by Linnaeus (1758), the cumulative species description curves for freshwater/brackish and freshwater/brackish/marine habitats have similarly increased progressively, albeit at lower rates. Over the last 10 years, 2.3 species living in freshwater/brackish environs and 1.9 species inhabiting freshwater/brackish/marine habitats have been described per year (Fig. 1). The best model able to account for the yearly variation in the Naperian logarithm of the accumulated number of new species descriptions is a sigmoid model with an upper asymptote, the Morgan–Mercer–Flodin model (see Tjørve, 2003), which explains 99.73%, 99.64% and 99.08% of total variability for freshwater, freshwater/brackish and freshwater/brackish/marine fishes, respectively (Fig. 1). According to these results, around 16989, 647 and 646 species would be recognized by 2100 in these three assemblages of fishes.

Survey effort heterogeneity

Around 51% of the considered 1-degree grid cells have at least one record of a freshwater fish species, and we calculated the completeness values for approximately 29% of these cells (i.e. those cells with twice as many records as species and a minimum of five observed species). These completeness values range from 42.3% to 100% with a pooled mean (±95% CI) of 75.7 ± 0.4 (Fig. 2). Only 199 grid cells could be considered to have relatively accurate inventories as their completeness values are equal to, or higher than, 90%. The best surveyed 1-degree cells based on available data are mainly located in Europe and the eastern portions of North America (Fig. 2), while many regions of South America, Africa and Asia are poorly surveyed or simply lack any information (white areas of Fig. 2). Around 36% of total grid cells lack any records of fishes but contain river basins, while 26% of the total grid cells harbour neither river basins nor species records. Cells lacking records but including freshwater systems are obvious serious candidates for future fish surveys and as foci for efforts to capture extant but electronically unavailable data on fish occurrence for within-cell drainages (see Fig. 2). Cells harbouring neither river basins nor species records can be considered truly uninhabited areas (white areas in glaciated and desert regions; Fig. 2), although it is possible that fish could be present in special situations in such cells (e.g. oasis pools within deserts).

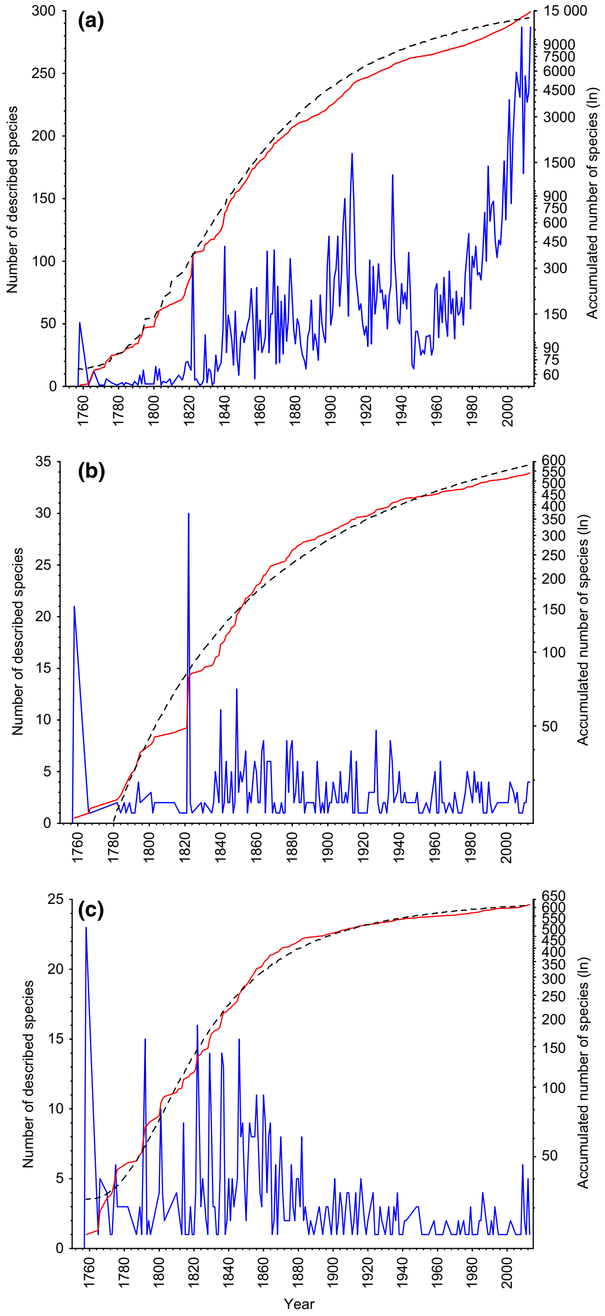

Figure 1 Number of described species (in blue) and Naperian logarithm of the accumulated number of described species (in red) by year of description for freshwater (a), freshwater/brackish (b) and freshwater/brackish/marine fishes (c). The dashed line represents the described estimated species according to the Morgan–Mercer–Flodin model.

Extent of occurrence and species richness

Range maps obtained from the application of α-shapes on point-to-grid data provide species richness values significantly and positively correlated with those generated by the first-order jackknife richness estimator (Fig. 3). The slope of the linear relationship (±95% CI) is slightly higher than one

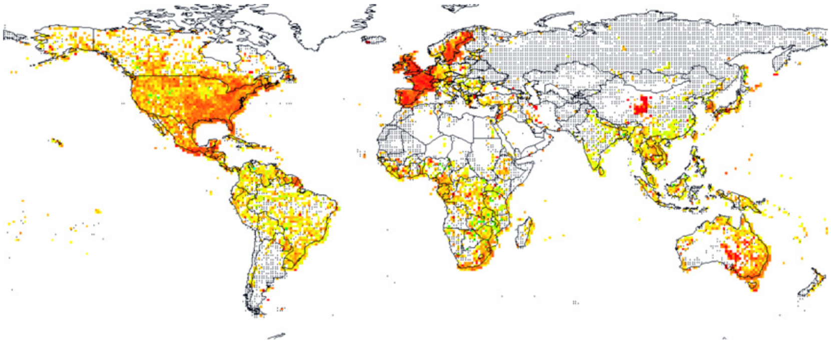

Figure 2 Completeness values (percentage of observed against predicted species richness calculated by the first-order jackknife estimator) for all the cells of 1° × 1°. Red areas are those that have completeness values higher than 75%, yellow areas have completeness values between 50% and 74% and green areas have completeness values lower than 50%. Small red spots represent cells without records but with river basins.

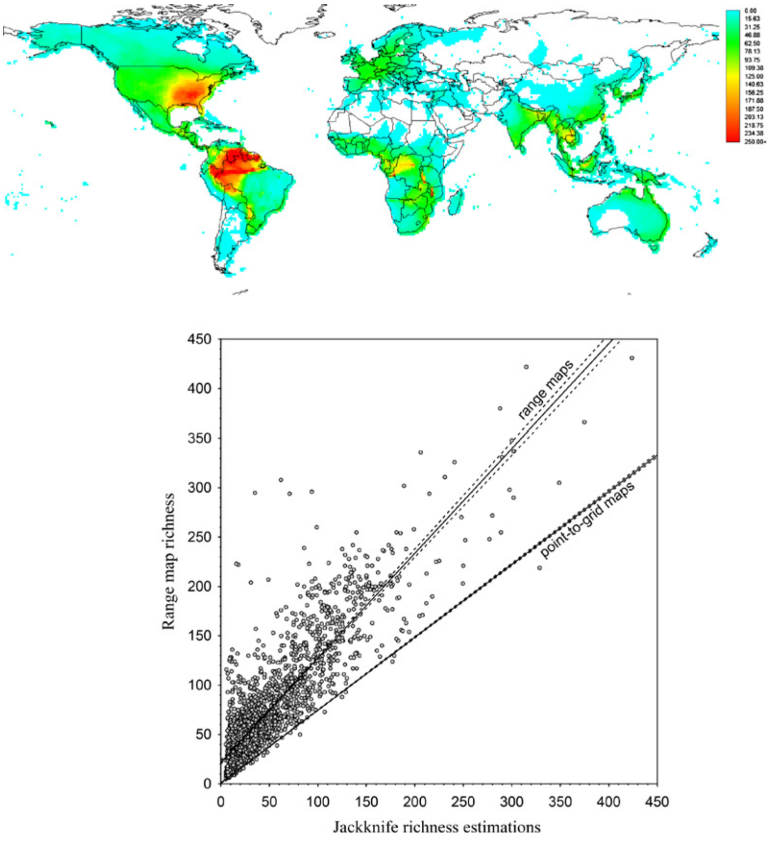

Figure 3 Variation in species richness of freshwater fishes world-wide derived from applying alpha-shape procedure (range maps) to point-to-grid data (upper panel), and relationships between values provided by first-order jackknife richness estimator and those coming from range maps and point-to-grid data (lower panel). Continuous lines (lower panel) represent linear regressions and dashed lines 95% regression bands.

(b = 1.06 ± 0.03; t = 7.78; $P < 0.001$) but much closer to unity than is the slope of the relationships between point-to-grid species data and jackknife richness values (b = 0.74 ± 0.01; t = 284.9; $P < 0.001$). Thus, although the so-obtained range maps seem to consistently overestimate richest cells, we can assume that the world-wide representation based on the available data is relatively reliable. The distribution of species richness shows that a higher richness exists in the tropical regions of Africa, Asia and South America plus to a lesser degree North America and Europe. However, richness estimates for North America and portions of Europe versus tropical regions (Fig. 3) would be influenced by the more thorough knowledge, sampling and databasing of the fish faunas in North America and Europe versus tropical regions world-wide as evidenced by the completeness values (Fig. 2).

The relationship between species richness and EOO is clearly triangular (Fig. 4) such that species-rich cells are those with a higher proportion of distributionally rare species, and as species richness decreases, the contribution of widely distributed species is higher. In fact, the distribution of $\mathrm{EOO_{avg}}$ values shows that rich freshwater fish assemblages with more narrowly distributed species are mainly distributed across tropical regions, while relatively species-poor areas containing narrowly distributed species are located in neighbouring subtropical regions and the Euromediterranean region (Fig. 4).

Associated environmental variables

The multiple regression analysis on species richness values explained 48.6% of the observed variance (Fig. 5). There was no strong collinearity among explanatory variables, with VIF values lower than 14 in all pairwise comparisons. Standardized residuals did not have a normal distribution (Kolmogorov–Smirnov, $P < 0.001$), and there was autocorrelation (Durbin–Watson, $P < 0.001$) and heteroscedasticity in the residuals (Breusch–Pagan test, $P < 0.001$). Therefore, it was impossible to test the significance of the regression. Terrestrial primary productivity (TPP), annual precipitation (Pre = BIO12), isothermality (Is = BIO3) and mean annual temperature (Tmean = BIO1) seem to be the most important of these variables, with all positively associated with the variation in species richness.

Support Vector Machine explained 81.7% of the observed variability in species richness. Most of the residuals of the model performed with SVM averaged close to zero as shown by the extensive green colouration in Fig. 5. Negative residuals identify areas with lower than predicted species richness (Fig. 5). These may be areas with undiscovered and/or undescribed species (i.e. northern South America; Vari & Malabarba, 1998), or areas with reduced species richness due to negative anthropogenic impacts (North America). As the cumulative description curve of strictly freshwater species has

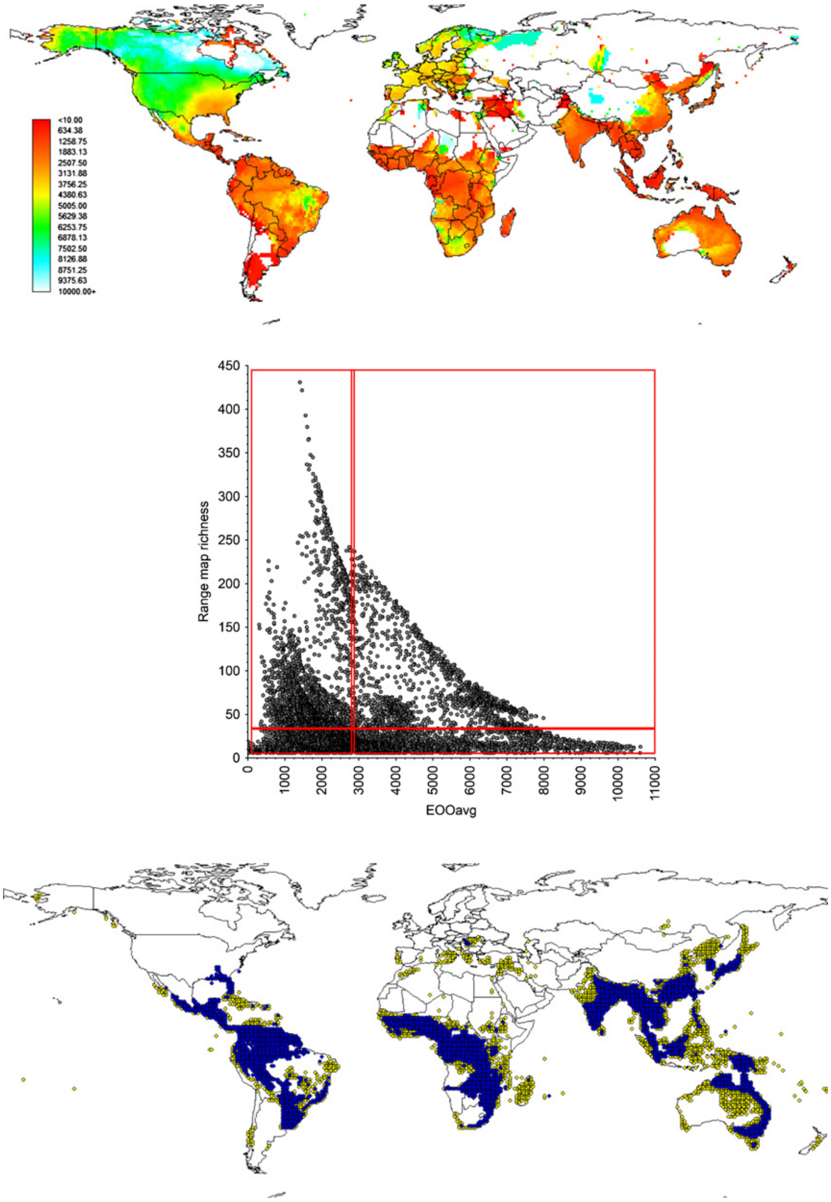

Figure 4 World variation in values of EOO_{avg} (upper panel) and triangular relationship between these values and those of species richness calculated by alpha-shape procedure (range maps; middle panel). Triangular relationship has been divided into four quadrants according to its median values. Map in lower panel presents geographical location of cells with lower than median EOO_{avg} values both in richest (blue) and in poorest cells (yellow).

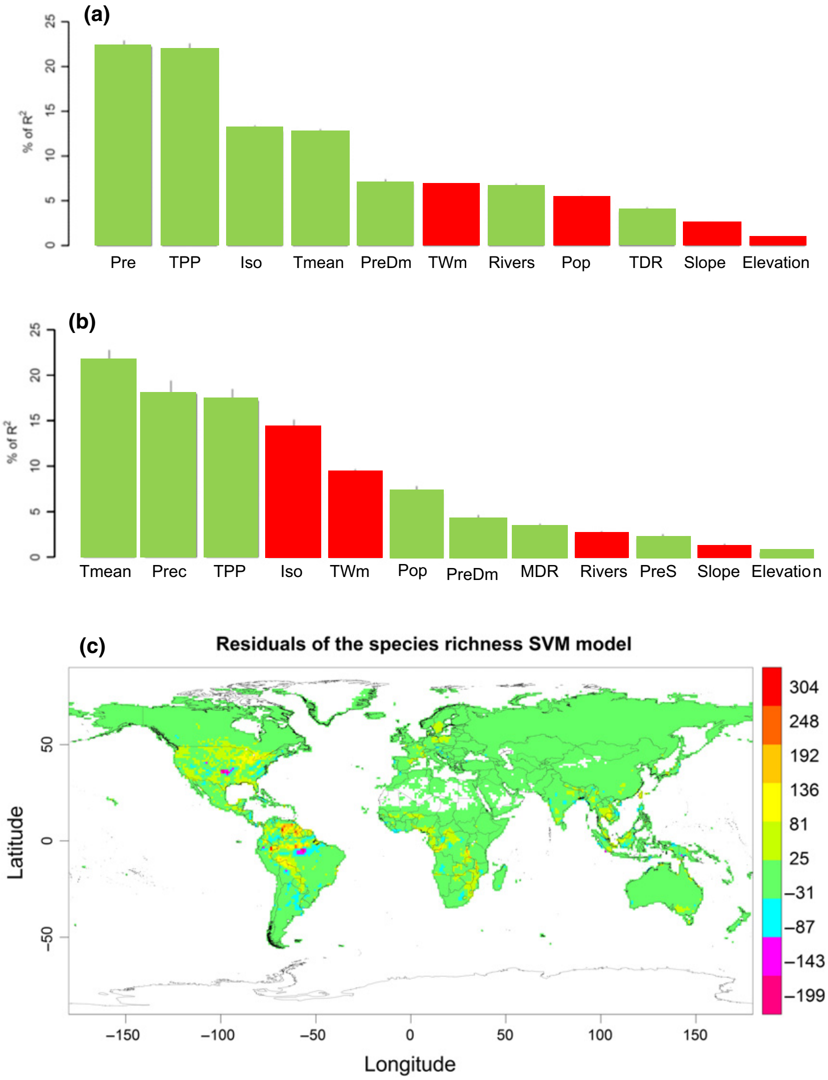

Figure 5 Relative contribution of the different considered variables in the multiple regressions obtained with the LMG method for (a) species richness and (b) EOO_{avg}. Both positively correlated (green bars) and negatively correlated (red bars) variables are shown with their 95% bootstrap confidence intervals. (c) Geographical distribution of the residuals of the SVM models in the case of species richness. Abbreviations: Elevation (metres), Tmean (BIO1, mean annual temperature in °C), MDR (BIO2, mean diurnal range, mean of monthly (max temp − min temp in °C), Iso (BIO3, isothermality (2/7) (* 100) in °C), TWm (BIO5, max temperature of warmest month), Pre (BIO12, annual precipitation in mm), PreDm (BIO14, precipitation during driest month in mm), PreS (BIO15, precipitation seasonality, coefficient of variation), Pop (human population density (number of people per km^2), Rivers (area of the river basins in km^2), Slope (topographical slope in degrees) and terrestrial primary production (TPP) (g C m^{-2} d^{-1}).

not approached its asymptote, the under-sampled regions likely contain large numbers of undiscovered/undescribed species.

The considered environmental predictors have a lower explanatory capacity on EOO_{avg} values both in the case of multiple regressions (27%) and for SVM (63%). In any case, within the considered variables, those with a higher explanatory capacity seem also to be those related to climatic or productivity parameters. Annual mean temperature (Tmean = BIO1), annual precipitation (Pre = BIO12) and TPP are the most important predictors positively related to $EOO_{avg,}$ while isothermality (Iso = BIO3) and maximum temperature of the warmest month (TWm = BIO5) are also relevant predictors but are negatively correlated with EOO_{avg} values (Fig. 5).

DISCUSSION

Information as to the rate of description and cumulative numbers of approximately 16,000 freshwater fish species clearly demonstrates that numerous species probably remain to be discovered and described in upcoming years. Furthermore, our analyses demonstrate that on a world-wide scale, proportionally few coarse one-degree grid cells can be considered well-surveyed; with these being located in regions such as North America and Europe with a long tradition of intense taxonomic and geographical studies. Although a more detailed study of the detected biases is necessary, our results suggest that more than one-third of the terrestrial area of the world is inadequately inventoried in terms of fishes, with these regions principally located in Africa, Asia and South America.

The development of range maps from point-to-grid data using an alpha-shape procedure seems to partially overcome these detected biases and data deficiencies. Thus, the observed geographical patterns of species richness and geographical rarity should be considered provisional although relatively reliable according to the available information. The distribution of species richness arrived at in this analysis is similar to the estimates provided by others authors

(Abell *et al.*, 2008; Collen *et al.*, 2014) with particular richness in tropical regions of Africa, Asia and South America plus to a lesser degree North America and Europe. Thus, some of the areas of highest species richness (portions of Africa, Asia and South America) could also be those with lower degrees of inventory completeness (compare maps of Figs 2–4).

It is rather important to point out that part of the variance in species richness between regions might be due to differences in the research effort focused on various geographical areas resulting in different degrees of success in registering and/or finding species not previously documented from a region (Guisande *et al.*, 2013). Other factors include the application of changing species criteria which, in part, account for the increased numbers of freshwater species recognized in European freshwaters in the last decade and the application of newer analytical techniques (e.g. DNA barcoding) that permit the identification of previously undetected cryptic species. It is difficult to quantify the degree to which these different factors contribute to the variance in species richness among areas; however, all are potentially important sources of bias and thus possibly responsible for a large proportion of the unexplained observed variance in species richness.

Both the relationship between species richness and geographical rarity and the distribution of $\mathrm{EOO}_{\mathrm{avg}}$ values worldwide suggest that many of the tropical and subtropical areas with a lower degree of inventory completeness may contain a high number of undetected, narrowly distributed species.

An interesting result of our study is the triangular relationship between species richness and geographical rarity measured by $\mathrm{EOO}_{\mathrm{avg}}$. Previous analyses with different groups of organisms found positive relationships between species richness and rarity (Pagel *et al.*, 1991; Gaston, 1994) even in the case of freshwater fishes (Tisseuil *et al.*, 2013), although at times yielding unrelated patterns between these two biodiversity variables (Orme *et al.*, 2005; Grenyer *et al.*, 2006). Positive relationships between species richness and rarity appear in nested assemblages (Arita *et al.*, 2012) when absences in species-rich areas are unusual. This pattern has been associated with the existence of selective colonization and/or extinction processes (Darlington, 1957; Wright *et al.*, 1998; Gaston & Blackburn, 2000). In any case, this pattern suggests that the processes that create and maintain the distribution of rare freshwater species are probably similar to the processes implied for most of the species in those studies. On the contrary, non-nested patterns tend to occur when some species-poor areas are inhabited by narrowly distributed species as a consequence of isolation and associated speciation processes (Cook & Quinn, 1995; Gaston & Blackburn, 2000). These two divergent patterns have been linked to the predominance of contemporary climate/energy factors (nested patterns and/or positive relationship between species richness and geographical rarity) or historical ones in the geographical arrangement of current biodiversity patterns (Lobo *et al.*, 2008; Baselga *et al.*, 2012).

In the case of freshwater fishes, the species-rich areas harbour a higher number of narrowly distributed species, with these areas most often located in tropical regions (see also Collen *et al.*, 2014). However, comparatively species-poor cells located in the adjacent subtropical regions may also contain narrowly distributed species (Dias *et al.*, 2013). We suggest that the faunistic composition of these latter areas have been predominantly shaped by historical factors and their role as refuge/diversification centres during past times. Necessary cautions notwithstanding, our results also suggest that a high environmental determinism exists in the case of species richness which can be accounted for by a limited number of climatic and productivity variables as happens in other groups (Hawkins *et al.*, 2012) and previously documented for some freshwater fishes (Tedesco *et al.*, 2005; Oberdorff *et al.*, 2011; Tedesco *et al.*, 2012; Griffiths *et al.*, 2014). However, geographical rarity does not seem to be as well predicted by the full spectrum of the considered variables. This supports the hypothesis that other non-climatic or non-energetic factors could have influenced the distribution of rarity among freshwater fishes, a hypothesis previously advanced for freshwater fishes (Hubert & Renno, 2006; Leprieur *et al.*, 2011; Dias *et al.*, 2014; Griffiths *et al.*, 2014).

Species richness proved to be higher in areas containing greater numbers of species confined to a specific, relatively small geographical area. This supports the hypothesis that the diversity of freshwater fishes is not only a consequence of environmental factors, but more so of the isolation resultant from river basin boundaries and the associated refuge and diversification processes promoted by such isolation. The predominance of species with small geographical range sizes in diversity hotspots may be the outcome of the role played by these contingent factors that are not directly related to climatic or energetic variables. Small geographical range sizes, however, increase extinction probabilities (Gaston, 1994; Blanchet *et al.*, 2013). Thus, the general isolation and associated smaller populations of the species in many freshwater fish communities (Toussaint *et al.*, 2014) on the one hand promote evolutionary turnover at geological time-scales as previously demonstrated for some freshwater fish groups (Bloom *et al.*, 2013). Alternatively, these factors increase the vulnerability of such species and communities to perturbations of aquatic environments and communities. Freshwater fishes can be defined as ‘potential victims of their own success’ in that the factors that promote diversity among them (limited distributional ranges and associated smaller populations) and account for their evolutionary success at the same time increase the likelihood of their possible extinction by human and other alterations of their environment – a conundrum for many species of freshwater fishes.

ACKNOWLEDGEMENT

We thank ENDESA for technical and financial support, and Juliana Herrera and Almudena Blanco for their great assistance in searching for records of species occurrences.

REFERENCES

Abell, R., Thieme, M.L., Revenga, C. *et al.* (2008) Freshwater ecoregions of the World: a new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, **58**, 403–414.

Arita, H.T., Christen, J.A., Rodríguez, P. & Soberón, J. (2012) The presence-absence matrix reloaded: the use and interpretation of range –diversity plots. *Global Ecology and Biogeography*, **21**, 282–292.

Baselga, A., Gómez-Rodríguez, C. & Lobo, J.M. (2012) Historical legacies in world amphibian diversity revealed by the turnover and nestedness components of beta diversity. *PLoS ONE*, **7**, e32341.

Blanchet, S., Reyjol, Y., April, J., Mandrak, N.E., Rodríguez, M.A., Bernatchez, L. & Magnan, P. (2013) Phenotypic and phylogenetic correlates of geographic range size in Canadian freshwater fishes. *Global Ecology and Biogeography*, **22**, 1083–1094.

Bloom, D.D., Weir, J.T., Piller, K.R. & Lovejoy, N.R. (2013) Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of New World Silversides (Atherinopsidae). *Evolution*, **67**, 2049–2057.

Burgman, M. A. & Fox, J. (2013) Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Animal Conservation*, **6**, 19–28.

Collen, B., Whitton, F., Dyer, E.E., Baillie, J.E.M., Cumberlidge, N., Darwall, W.R.T., Pollock, C., Richman, N.I., Soulsby, A.-M. & Böhm, M. (2014) Global patterns of freshwater species diversity, threat and endemism. *Global Ecology and Biogeography*, **23**, 40–51.

Cook, R.R. & Quinn, J.F. (1995) The influence of colonization in nested species subsets. *Oecologia*, **102**, 413–424.

Costello, M.J., Wilson, S.P. & Houlding, B. (2012) Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Systematic Biology*, **61**, 871–883.

Darlington, P.J. (1957) *Zoogeography: the Geographical Distribution of Animals*. Wiley, New York 675 pp.

Dias, M.S., Cornu, J.F., Oberdorff, T., Lasso, C.A. & Tedesco, P.A. (2013) Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography*, **36**, 683–689.

Dias, M.S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J.F., Brosse, S., Grenouillet, G. & Tedesco, P.A. (2014) Global imprint of historical connectivity on freshwater fish biodiversity. *Ecology Letters*, **17**, 1130–1140.

Dubois, D.M. (1973) An index of fluctuations, Do, connected with diversity and stability of ecosystems: applications in the Lotka–Volterra model and in an experimental distribution of species. Rapport de sythèse III, Programme National sur l'environment Physique et Biologique, Project Mer. Commision Interministérielle de la Politique Scientifique. Liège.

Eschmeyer, W.N. (2013) Available at: http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp (accessed 16 December 2013).

Fox, J. & Weisberg, S. (2011) *An R Companion to Applied Regression*, 2nd edn. Thousand Oaks, Sage.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Laboissiere, R., Monette, G., Murdoch, D., Nilsson, H., Ogle, D., Ripley, B., Venables, W. & Zeileis, A. (2013) Companion to Applied Regression. R package version 2.0-19. Available at: http://CRAN.R-project.org/package=lmtest (accessed 30 June 2013).

Froese, R. & Pauly, D. (2013) FishBase World Wide Web electronic publication. Available at: http://www.fishbase.org version (accessed 31 August 2013).

Gaiji, S., Chavan, V., Ariño, A.H., Otegui, J., Robles, E. & King, E. (2013) Content assessment of the primary biodiversity data through GBIF network: status, challenges and Potentials. *Biodiversity Informatics*, **8**, 94–172.

García-Roselló, E., Guisande, C., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., Vaamonde, A. & Granado-Lorencio, C. (2013) ModestR: a software tool for managing and analyzing species distribution map databases. *Ecography*, **36**, 102–1207.

García-Roselló, E., Guisande, C., Heine, J., Pelayo-Villamil, P., Manjarrés-Hernández, A., González-Vilas, L., González-Dacosta, J., Vaamonde, A. & Granado-Lorencio, C. (2014) Using ModestR to download, import and clean species distribution records. *Methods in Ecology and Evolution*, **5**, 703–713.

Gaston, K.J. (1994) *Rarity*. Chapman & Hall, London.

Gaston, K.J. & Blackburn, T.M. (2000) *Patterns and Process in Macroecology*. Blackwell, Oxford.

Gaston, K.J. & Fuller, R.A. (2009) The sizes of species' geographic ranges. *Journal of Applied Ecology*, **46**, 1–9.

GBIF (2013) GBIF data portal. Available at: http://datagbi forg (accessed 30 June 2013).

Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, **15**, 578–587.

Grenyer, R., Orme, C.D.L., Jackson, S.F., Thomas, G.H., Davies, R.G., Davies, T.J., Jones, K.E., Olson, V.A., Ridgley, R.S., Rasmusen, P.C., Ding, T.S., Bennetts, P.M., Blackburn, T.M., Gaston, K.J., Gittleman, J.L. & Owens, I.P.F. (2006) global distribution and conservation of rare and threatened vertebrates. *Nature*, **444**, 93–96.

Griffiths, D., McGonigle, C. & Quinn, R. (2014) Climate and species richness patterns of freshwater fish in North America and Europe. *Journal of Biogeography*, **41**, 452–463.

Grömping, U. (2006) Relative importance for linear regression in R: the package relaimpo. *Journal of Statistical Software*, **17**, 1–27.

Gross, J. (2013) Five omnibus tests for the composite hypothesis of normality. R package version 1.0-2. Available at: http://CRAN.R-project.org/package=nortest (accessed 30 June 2013).

Guisande, C., Barreiro, A., Maneiro, I., Riveiro, I., Vergara-Castaño, A.R. & Vaamonde, A. (2006) *Tratamiento de datos*, 367 pp. Ediciones Díaz de Santos, Madrid.

Guisande, C., Manjarrés-Hernández, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuña, A., Prieto-Piraquive, E., Janeiro, E., Matías, J.M., Patti, C., Patti, B., Mazzola, S., Jiménez, L.F., Duque, S. & Salmerón, F. (2010) IPez: an expert system for the taxonomic identification of fishes based on machine learning techniques. *Fisheries Research*, **102**, 240–247.

Guisande, C., Patti, B., Vaamonde, A., Manjarrés-Hernández, A., Pelayo-Villamil, P., García-Roselló, E., González-Dacosta, J., Heine, J. & Granado-Lorencio, C. (2013) Factors affecting species richness of marine elasmobranchs. *Biodiversity and Conservation*, **22**, 1703–1714.

Hawkins, B.A., McCain, C.M., Davies, T.J., Buckley, L.B., Anacker, B., Cornell, H.V., Damschen, E.I., Grytnes, J.-A., Harrison, S., Holt, R.D., Kraft, N.J.B. & Stephens, P.R. (2012) Different evolutionary histories underlie congruent species richness gradients of birds and mammals. *Journal of Biogeography*, **39**, 825–841.

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.

Hortal, J., Borges, P.A. & Gaspar, C. (2006) Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*, **75**, 274–287.

Hothorn, T., Zeileis, A., Farebrother, R.W., Cummins, C., Millo, G. & Mitchell, D. (2013) Testing Linear Regression Models. R package version 0.9-32. Available at: http://CRAN.R-project.org/package=lmtest (accessed 30 June 2013).

Hubert, N. & Renno, J.-F. (2006) Historical biogeography of South American freshwater fishes. *Journal of Biogeography*, **33**, 1414–1436.

IUCN (2013) *Guidelines for Using the IUCN Red List Categories and Criteria. Version 10.* IUCN Standards and Petitions Subcommittee. Available at: http://www.iucnredlist.org/documents/RedListGuidelines.pdf (accessed 16 June 2013).

Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. (2004) kernlab – an S4 package for Kernel methods in R. *Journal of Statistical Software*, **11**, 1–20.

Leprieur, F., Tedesco, P.A., Hugueny, B., Beauchard, O., Dürr, H.H., Brosse, S. & Oberdorff, T. (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. *Ecology Letters*, **14**, 325–334.

Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M.L.J. & Tedesco, P.A. (2008) Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*, **595**, 545–567.

Linnaeus, C. (1758) Systema naturæ per regna tria naturæ, secundum classses, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I, Editio Decima, Reformata 1758, Holmiæ, Laurentii Salvii (Salvius publ.), 824 pp.

Lobo, J.M., Jay-Robert, P. & Lumaret, J.P. (2008) The relationship between forecasted rarity and species richness values for Scarabaeidae and Aphodiinae species in France (Coleoptera, Scarabaeoidea). *Insect Ecology and Conservation* (ed. by S. Fattorini), pp. 299–317. Research Signpost, Trivandrum

Meyer, D., Leisch, F. & Hornik, K. (2003) The support vector machine under test. *Neurocomputing*, **55**, 169–186.

Mikšik, M. & Schraml, E. (2013) World of fishes. Available at: http://www.worldfish.de/sci.htm (accessed 30 August 2013).

Nelson, J.S. (2006) *Fishes of the World*, 4th edn, 601 pp. John Wiley & Sons, Hoboken, NJ.

Oberdorff, T., Tedesco, P.A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S. & Dürr, H.H. (2011) Global and regional patterns in riverine fish species richness: a review. *International Journal of Ecology*, Article ID 967631, 12 pages, doi:10.1155/2011/967631.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.M. & Wagner, H. (2013) Community Ecology Package. R package version 2.0-9. Available at: http://CRAN.R-project.org/package=vegan (accessed 12 September 2013).

Orme, C.D.L., Davies, R.G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V.A., Webster, A.J., Ding, T.S., Rasmussen, P.C., Ridgely, R.S., Stattersfield, A.J., Bennett, P.M., Blackburn, T.M., Gaston, K.J. & Owens, I.P.F. (2005) Global hotspots of species richness are not congruent with endemism or threat. *Nature*, **436**, 1016–1019.

Pagel, M., May, R. & Collie, A. (1991) Ecological aspects of the geographical distribution and diversity of mammalian species. *The American Naturalist*, **137**, 791–815.

Pateiro-Lopez, B. & Rodriguez-Casal, A. (2011) Alphahull: Generalization of the convex hull of a sample of points in the plane. Available at: http://cran.r484project.org/web/packages/alphahull (accessed 16 June 2013).

Pelayo-Villamil, P., Guisande, C., González-Vilas, L., Carvajal-Quintero, J.D., Jiménez-Segura, L.F., García-Roselló, E., Heine, J., González-Dacosta, J., Manjarrés-Hernández, A., Vaamonde, A. & Granado-Lorencio, C. (2012) ModestR: Una herramienta infromática para el estudio de los ecosistemas acuáticos de Colombia. *Actualidades Biológicas*, **34**, 225–239.

Pineda, E. & Lobo, J.M. (2012) The performance of range maps and species distribution models representing the geographic variation of species richness at different resolutions. *Global Ecology and Biogeography*, **21**, 935–944.

R Development Core Team (2013) *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, **35**, 211–226.

Tedesco, P.A., Oberdorff, T., Lasso, C.A., Zapata, M. & Hugueny, B. (2005) Evidence of history in explaining diversity patterns in tropical riverine fish. *Journal of Biogeography*, **32**, 1899–1907.

Tedesco, P.A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H.H., Beauchard, O., Busson, F. & Oberdorff, T. (2012) Patterns and processes of global freshwater fish endemism. *Global Ecology and Biogeography*, **21**, 977–987.

Tisseuil, C., Cornu, J.-F., Beauchard, O., Brosse, S., Darwall, W., Holland, R., Hugueny, B., Tedesco, P.A. & Oberdorff, T. (2013) Global diversity patterns and cross-taxa convergence in freshwater systems. *Journal of Animal Ecology*, **82**, 365–376.

Tjørve, E. (2003) Shapes and functions of species-area curves: a review of possible models. *Journal of Biogeography*, **30**, 827–835.

Toussaint, A., Beauchard, O., Oberdorff, T., Brosse, S. & Villéger, S. (2014) Historical assemblage distinctiveness and the introduction of widespread non-native species explain worldwide changes in freshwater fish taxonomic dissimilarity. *Global Ecology and Biogeography*, **23**, 574–584.

Vari, R.P. & Malabarba, L.R. (1998) Neotropical ichthyology: an overview. pp. 1–11. *Phylogeny and Classification of Neotropical Fishes* (ed. by L.R. Malabarba, R.E. Reis, R.P. Vari, Z.M.S. Lucena and C.S. Lucena), 603 pp. Edipucrs, Brazil.

Vega, G.C. & Wiens, J.J. (2012) Why are there so few fish in the sea? *Proceedings of the Royal Society B: Biological Sciences*, **279**, 2323–2329.

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

Wright, D.H. (1983) Species-energy theory: an extension of species-area theory. *Oikos*, **41**, 496–506.

Wright, D.H., Patterson, A.H., Mikkelson, G.M., Cutler, A. & Atmar, W. (1998) A comparative analysis of nested subset patterns of species composition. *Oecologia*, **113**, 1–20.

Zeileis, A. & Hothorn, T. (2002) Diagnostic checking in regression relationships. *R News*, **2**, 7–10. Available at: http://CRAN.R-project.org/doc/Rnews/ (accessed 3 September 2013).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 List of variables used in the manuscript in cells of $60' \times 60'$.

Appendix S2 List of species.

BIOSKETCH

Patricia Pelayo-Villamil is a PhD student at the University of Antioquia (Colombia). Her main field of interest is determining the main factors controlling the distribution and richness of freshwater fish species. To achieve that goal, she is exploring new techniques for collecting and cleaning geographical records, developing methods for estimating species distribution maps from raw geographical records, and developing statistical methods for the identification of the main factors that affect the distribution and richness of such species.

Editor: Anthony Ricciardi