Global patterns and predictors of tropical reef fish species richness

V. Parravicini, M. Kulbicki, D. R. Bellwood, A. M. Friedlander, J. E. Arias-Gonzalez, P. Chabanet, S. R. Floeter, R. Myers, L. Vigliola, S. D'Agata and D. Mouillot

V. Parravicini, (valeriano.parravicini@ird.fr) and M. Kulbicki, IRD, Inst. de Recherche pour le Développement, UR 227 CoReUs, LABEX Corail, Laboratoire Arago, BP 44, FR-66651 Banyuls/mer, France. VP also at: CESAB-FRB, Immeuble Henri Poincaré, Domaine du Petit Arbois, FR-13857 Aix-en-Provence cedex 3, France. – D. R. Bellwood, School of Marine and Tropical Biology, James Cook Univ., Townsville, QLD 4811, Australia. DRB also at: Australian Research Council, Center of Excellence for Coral Reef Studies, James Cook Univ., Townsville, QLD 4811, Australia. – A. M. Friedlander, Hawaii Cooperative Fishery Research Unit and Univ. of Hawaii at Manoa, Honolulu, HI 096822, USA. – J. E. Arias-Gonzalez, Laboratorio de Ecología de Ecosistemas de Arrecifes Coralinos, Depto de Recursos del Mar, Centro de Investigación y de Estudios Avanzados del Inst. Politécnico Nacional, Unidad Mérida, Antigua Carretera a Progreso Km. 73 Cordemex, Mérida, Yucatán, 97310 Mexico. – P. Chabanet, IRD, Inst. de Recherche pour le Développement, UR 227 CoReUs, LABEX Corail, BP 50172, CS 41095, FR-97495 Ste Clotilde cedex, Reunion Island, France. – S. R. Floeter, Depto de Ecologia e Zoologia, Univ. Federal de Santa Catarina, Florianópolis, Santa Catarina 88010-970, Brazil. – R. Myers, Seaclicks – Coral Graphics, 9273 Silent Oak Circle, Wellington, FL 33411, USA. – L. Vigliola and S. D'Agata, IRD, Inst. de Recherche pour le Développement, UR 227 CoReUs, LABEX Corail, BP A5 – 98848 Nouméa Cedex, New Caledonia. SD also at: Laboratoire Ecologie des Systèmes Marins Côtiers UMR 5119, CNRS, IRD, IFREMER, UM2, UM1, cc 093, Place E. Bataillon, FR-34095 Montpellier Cedex 5, France. – D. Mouillot, Australian Research Council, Center of Excellence for Coral Reef Studies, James Cook Univ., Townsville, QLD 4811, Australia. DM also at: Laboratoire Ecologie des Systèmes Marins Côtiers UMR 5119, CNRS, IRD, IFREMER, UM2, UM1, cc 093, Place E. Bataillon, FR-34095 Montpellier Cedex 5, France.

In the marine realm, the tropics host an extraordinary diversity of taxa but the drivers underlying the global distribution of marine organisms are still under scrutiny and we still lack an accurate global predictive model. Using a spatial database for 6336 tropical reef fishes, we attempted to predict species richness according to geometric, biogeographical and environmental explanatory variables. In particular, we aimed to evaluate and disentangle the predictive performances of temperature, habitat area, connectivity, mid-domain effect and biogeographical region on reef fish species richness. We used boosted regression trees, a flexible machine-learning technique, to build our predictive model and structural equation modeling to test for potential 'mediation effects' among predictors. Our model proved to be accurate, explaining 80% of the total deviance in fish richness using a cross-validated procedure. Coral reef area and biogeographical region were the primary predictors of reef fish species richness, followed by coast length, connectivity, mid-domain effect and sea surface temperature, with interactions between the region and other predictors. Important indirect effects of water temperature on reef fish richness, mediated by coral reef area, were also identified. The relationship between environmental predictors and species richness varied markedly among biogeographical regions. Our analysis revealed that a few easily accessible variables can accurately predict reef fish species richness. They also highlight concerns regarding ongoing environmental declines, with region-specific responses to variation in environmental conditions predicting a variable response to anthropogenic impacts.

Understanding the mechanisms generating and maintaining the global distribution of species richness is one of the major goals of biogeography and ecology (Gaston 2000), but we still lack conclusive evidence to separate processes from patterns (Hawkins et al. 2003). Although mechanisms are not well understood, strong relationships between species richness and environmental or geographical variables have been identified at large scales. Indeed, in the marine realm, recurrent predictors of species richness are consistently sea surface temperature, habitat area, and the mid-domain effect (Bellwood et al. 2005, Tittensor et al. 2010). This suggests that, even if understanding mechanisms producing

vs. maintaining variation in species richness remains challenging, it is possible to identify, contrast and disentangle those factors that best explain current-day large-scale patterns (Hawkins et al. 2003, Jetz and Fine 2012).

In the marine realm, the tropics host an extraordinary diversity of taxa upon which ecosystem goods and services depend (Worm et al. 2006). Tropical coastal zones, however, are presently in severe decline (Bellwood et al. 2004, Knowlton and Jackson 2008). Human pressures and global environmental changes are inducing local biodiversity erosion which may imperil ecosystem functioning (Bellwood et al. 2012). Identifying and ranking environmental or

geographical variables maintaining species richness is thus critical to addressing the worldwide biodiversity crisis.

Here we built a predictive model of reef fish species richness at a global scale and we disentangled the relative contribution of environmental (i.e. temperature, habitat area, connectivity) and geographical variables (i.e. mid-domain effect, biogeographical region) in explaining observed global diversity patterns. Previous works on reef fishes focused on the Indo-Pacific (Bellwood et al. 2005) or Tropical Eastern Pacific (Mora and Robertson 2005). These studies identified habitat area (Bellwood and Hughes 2001), the mid-domain effect (Connolly et al. 2003), connectivity to the Indo-Australian Archipelago (IAA; Mora et al. 2003), and a combination of mid-domain effect and habitat area (Bellwood et al. 2005) as the main predictors of reef fish richness. The mid-domain effect considers the spatial variation of species richness due to bounding constraints (Colwell and Hurtt 1994). If the geographical ranges of species are randomly placed within a bounded domain, the resultant richness gradient will form a peak in the middle of the domain (i.e. the mid-domain). Following Stehli, Wells, Potts and Rosen (reviewed by Bellwood et al. 2012) we propose a connectivity hypothesis that postulates that areas with low geographical distances between habitat patches will sustain comparatively higher species richness as it allows for potentially more interconnected meta-communities. At smaller scales, this hypothesis is supported by experimental evidences documenting higher fish species richness along continuous reefs than in isolated reef patches (Belmaker et al. 2011). The area hypothesis predicts that the greater the habitat area, the higher the number of species that can be maintained (Gray 2001). Finally, the energy hypothesis predicts that higher temperature supports high species richness through its effects on the biochemical kinetics of metabolism (Allen et al. 2002).

Although major predictors of reef fish richness have been identified, a more complex scenario emerges when attempting to disentangle their relative influence at a global scale. Different oceanic basins have markedly different evolutionary histories (Floeter et al. 2008, Cowman and Bellwood 2013). The Indo-Pacific region, where most previous research was focused, shows a unique pattern of variation in species richness which is markedly different from that observed in the Atlantic Ocean. In the Indo-Pacific,

species richness declines nearly uniformly with increasing distance from the IAA, which is close to the mid-domain of the region (Bellwood et al. 2012). In contrast, the biodiversity hotspot in the Atlantic Ocean (i.e. the Caribbean) is distant from the mid-domain and comparatively less connected to surrounding areas because of several biogeographical barriers (Luiz et al. 2012).

Contrary to the pattern observed in the tropical Indo-Pacific (Bellwood et al. 2005), at a broader latitudinal scale in the Atlantic Ocean, Macpherson (2002) found a major role of temperature and a negligible effect of area in explaining the distribution of species richness for several marine taxa, including fishes. Complex interactions between environmental, geographic and historical variables may thus blur the patterns and weaken the predictive power of models at a global scale.

To overcome these constraints and to provide a robust predictive model of fish species richness in the tropics, we accounted for the biogeographical region in our analysis. Our goal is to test whether an accurate predictive model can be proposed for reef fishes at global scale based on geometric, biogeographical and environmental factors and accounting for their interactions. After construction of the predictive model, we then used structural equation models to further explain the complex pathways between explanatory factors and fish richness.

Methods

Fish species lists

We obtained estimates of reef fish species richness at 163 locations worldwide (Fig. 1). We examined almost 500 references and extracted information from published works, regional checklists, monographs on specific families or genera, and reports (Supplementary material Appendix 1). We restricted data collection to the oceans portion showing a minimum monthly sea surface temperature (hereafter SST) of 17°C. Although the present study focused on the tropics, whose SST limit is usually set at 20°C, we decided to use a broader definition of 'tropical' oceans, thereby including locations where species of tropical affinity are present. The inclusion of those locations between 17°C and 20°C broadens

Figure 1. Geographical position of the 163 locations (black circles) for which information on the presence–absence of tropical reef-associated fish was available. Dotted lines indicates known physical or physiological barriers considered as boundaries of the domains. Stars represents the position of the mid-domains. EPB: Eastern Pacific Barrier; MAB: Mid-Atlantic Barrier; AOB: Amazon-Orinoco Barrier.

the range of variables, potentially increasing the accuracy of our predictive model. Overall, we obtained information on the distribution of 6336 reef fishes from 37 locations in the Atlantic Ocean, 39 in the Indian Ocean, 70 in the Pacific Ocean, and 17 in the Tropical Eastern Pacific.

Explanatory variables

Explanatory variables were selected according to several hypotheses explaining the variation in species richness based on previous works and general ecological theories. In particular, we considered the following hypotheses: 1) energy; 2) area; 3) mid-domain (hereafter MD) effect and 4) connectivity.

Estimates for each explanatory variable were obtained for each location using two geographical definitions of location. First, according to maps and descriptions in the original literature used to estimate species richness, we defined for each location, the area of the shelf (sea-bottom down to 200 m depth) to which the species list pertained. The shelf was defined using SRTM30_PLUS bathymetry (Shuttle Radar Topography Mission) available at <http://topex.ucsd.edu/WWW_html/srtm30_plus.html>. In this way, we obtained polygons of different size according to the locations, which were later used for the extraction of explanatory variables from raster and vector spatial files. Second, we estimated explanatory variables within a standard radius of 600 km around each location. The 600 km radius was chosen because it was already successfully employed by previous research conducted in the Indo-Pacific (Bellwood et al. 2005) and encompasses the probable larval dispersal envelope of reef fishes (Swearer et al. 2002). All spatial analyses were conducted using a global equal area Behrmann projection with the prime meridian set at 145° longitude.

For the energy hypothesis, we evaluated SST (i.e. minimum, mean, maximum and the range of annual values of sea surface temperature). This information was obtained from the Bio-ORACLE database at a resolution of 5 arcmin (Tyberghein et al. 2012) available at <www.oracle.ugent.be>. SST estimates were based on monthly mean values from 2002 to 2009.

For the area hypothesis we considered the total coral reef area (km²), the shelf area (km²) and the coast length (km). Shelf area was considered as a relevant proxy of the historical habitat availability as it provides an approximate estimate of the coastal waters during the Pleistocene low sea level stands (Bellwood et al. 2005). Coast length was used as an estimate of the present extent of coastline used by shallow water reef fish. Both coast length and shelf area were obtained using SRTM30_PLUS bathymetry (Shuttle Radar Topography Mission). Estimates of coral reef area were obtained from data based on the Coral Reef Millennium Census project (Andréfouët et al. 2006) and available at: <http://data.unep-wcmc.org/>.

The potential effect of MD on species richness was also considered. Typically it is assessed using null-models based on resampling of species with defined geographical ranges (Currie and Kerr 2008). However, there is a substantial lack of agreement on the most appropriate null-model and this approach does not easily account for multiple predictors

in the same analysis. Thus, we implemented the approach developed by Bellwood et al. (2005) calculating the distance from the MD relative to the domain size (i.e. the normalised distance to the MD) for each location. We operatively defined the domains as those areas within well-known physical or physiological barriers that are rarely crossed by reef fishes. We defined five such domains in all (Fig. 1). In the Indo-Pacific we considered the area from the western Indian Ocean to Easter Island and Salas y Gómez as separated from the Tropical Eastern Pacific by the East Pacific Barrier (Robertson et al. 2004, Robertson and Cramer 2009). In the Atlantic Ocean, we considered two distinct domains along the west coast, respectively separated by the Amazon River barrier and one domain in the east coast, east and west coasts of the Atlantic being separated by the Mid-Atlantic Barrier (Floeter et al. 2008, Luiz et al. 2012).

In order to account for the connectivity hypothesis, connectivity was defined as the relative proximity of each location to patches of reef fish habitat. Connectivity estimates were calculated using a nearest neighbour approach, which is typically employed in ecological research (Moilanen and Nieminen 2002). In particular, using a Behrmann projection, the world was divided into cells (200 × 200 km at the equator) from which only cells containing reef fish habitat (i.e. coast or coral reef) were retained. These cells represent habitat patches and for each location we computed the inverse of the mean distance from the location to the 10 nearest patches. We arbitrarily used 10 nearest neighbours instead of 1 in order to limit the strong dependency of the variable to a single neighbour, which may not always be highly connected to the most proximate reefs because of current flows (Mora et al. 2012). Finally, we considered the different evolutionary histories characterising each biogeographical region (Cowman and Bellwood 2013). We therefore included a 'region' factor as a predictor variable. We used four biogeographical regions: the Atlantic Ocean, Indian Ocean, and Pacific Ocean and the Tropical Eastern Pacific.

Data analysis

In order to build a predictive model for global patterns of species richness for tropical reef fishes we implemented boosted regression trees (hereafter BRT, Friedman 2001). BRT differ from traditional regression methods that produce a single 'best' model in that they combine large numbers of simple trees to optimize predictive performance (Elith et al. 2008). This approach was preferred over linear or additive models for two main reasons: 1) BRT can handle complex relationships and interactions among predictor variables that were expected to emerge at a global scale; 2) BRT are considered as a robust technique which can handle outliers and nonlinearity (Friedman 2001, Elith et al. 2008). Although BRT were also considered relatively robust to collinearity among predictor variables, it has been recently shown that removing collinearity from explanatory variables showing a Pearson correlation coefficient higher than $|0.7|$ is a desirable practice also with techniques based on boosting (Dormann et al. 2012).

After testing for collinearity between explanatory variables, we decided to remove maximum SST, minimum SST and the extent of the continental shelf (Supplementary

material Appendix 2). Regarding SST estimates, high collinearity was detected between mean SST and minimum or maximum SST, thus suggesting that mean SST alone contains most of information about temperature extremes. In addition, we detected collinearity between shelf area and coast length and we decided to remove the shelf area from further analyses because previous research evidenced a negligible role of this variable on reef fish richness (Bellwood et al. 2005).

BRT were fitted using the gbm library in R (Ridgeway 2007) with additional functions provided by Elith et al. (2008). This technique allows for the specification of four main parameters: bag fraction (bf), learning rate (lr), tree complexity (tc) and the number of trees (nt). bf is the proportion of samples used at each step, lr is the contribution of each fitted tree to the final model, tc is the number of nodes of each fitted tree determining the extent to which statistical interactions were fitted, and nt represents the number of trees corresponding to the number of boosting iterations. Optimal setting of the parameters was chosen using 5-fold cross validation (CV). The procedure provides a parsimonious estimate, CV–D^2 (i.e. the cross validated proportion of the deviance explained), representing the expected performance of the model when fitted to new data (Elith et al. 2008). Using CV, we explored different combinations of the parameters to be set and retained the model showing the highest CV–D^2 (Supplementary material Appendix 3).

We also explored the possibility of eliminating non-informative predictor variables to select the most parsimonious model. This simplification process uses CV procedure, progressively simplifying the model fitted to each fold, and using the average CV errors to decide which and how many variables can be removed from the original model without decreasing its performance (see Appendix 2 in Elith et al. 2008 for details on the simplification procedure).

The relative influence of predictor variables in BRT models was evaluated with the contribution to model fit attributable to each predictor, averaged across all the trees fitted (Friedman 2001). In order to account for uncertainty around these estimates, the relative influence of each explanatory variable was calculated on 1000 bootstrap replicates of the original dataset in order to compute 95% confidence intervals. Similarly, we evaluated the accuracy of our predictive model calculating and mapping 95% confidence intervals of model predictions at each location.

We also identified the most important interactions between each pair of predictors (Elith et al. 2008). The effect of the most important predictors on fish richness was visualized by means of partial dependence plots, obtained by plotting predicted values in relation to major variables, while maintaining other variables constant.

As BRT cannot provide a simple way to depict model structure (i.e. no equation) and outputs, we illustrated BRT by plotting the first regression tree fitted by the technique.

In addition to BRT, which was used with the main goal of building a predictive model, structural equation models (SEM) were employed to disentangle direct vs indirect effects of explanatory variables on fish richness by evaluating many causal paths. SEM thus quantifies the relative direct and indirect contributions of predictors by incorporating partial regression models in order to disentangle pure variable effects (Mitchell 1992). This analysis was conducted because ecological relationships between our predictor variables may blur direct relationships. Thus, disentangling the existence of 'mediation effects' on fish richness may allow for a better understanding of their interactions.

The modeling process in SEM requires the a priori specification of pathways considering the relationships between variables. In this study, we specified the conceptual model based on hypothesized mechanisms expected to operate in tropical reef systems. First, we included all the direct relationships between explanatory variables and reef fish richness. Second, we introduced hypothesized relationships between pairs of explanatory variables, largely based on previous knowledge and general ecological theory. A potential effect of temperature-related variables on coral reef area was included in the model as temperature is known to be a limiting factor for coral reef development and also relates to the aragonite saturation (Kleypas et al. 1999).

Moreover we hypothesized an effect of the region on several variables. In particular, we hypothesized a region effect on coral reef area, because coral reef area is asymmetrically distributed across regions (Bellwood et al. 2005). Similarly, we considered a regional effect on connectivity, as the Pacific, Indian and Atlantic Oceans differ markedly in their history of connectivity (Cowman and Bellwood 2013). Other potential effects of regions, namely on temperature and MD effect were excluded because we forcedly considered the same temperature range in each region and the metric used to measure the MD influence was a priori standardized according to the domain size.

SEM was fitted using the 'lavaan' library in R using the Bollen–Stine bootstrapping technique which is adequate when fitting SEMs on variables not showing normal distribution (Rosseel 2012).

Spatial autocorrelation is a fundamental property of biogeographical data, with nearby locations are more related than distant ones. However, the autocorrelation of model residuals is a direct evidence that some unmeasured variables are missing to explain the spatial structure in the data (Hawkins et al. 2007, Hawkins 2012). We thus tested for spatial autocorrelation in the residuals of the final models (both BRT and SEM) by computing Moran's I index for different spatial lags. Moran's I correlograms were computed using 5 classes of distance-based neighbors defined by increasing the spatial lag by 500 km, from 2500 km (i.e. the minimum distance needed for having at least one neighbor for each location) to 5000 km which represents the largest spatial gap between islands (Victor and Wellington 2000). The analysis was conducted using the 'spdep' package in R (Bivand et al. 2013). Spatial patterns of the residuals were also mapped in order to highlight locations where overestimation or underestimation of reef fish richness occurred.

Results

The highest values of reef fish richness were disproportionally found in the Indo-Pacific, and especially in the IAA (Table 1; Fig. 2) with 2016 species in the Philippines. From the IAA, richness tends to decrease almost regularly both eastward and westward. An exception to this regular pattern occurred in the western Indian Ocean where species richness between

Table 1. Species and genera richness in each of the four regions (right part of the table) and species and genera in common for each pair of regions (left part of the table). Grey cells represent information at genus level while white ones are referred to species level.

	Atlantic	Indian	Pacific	TEP	Species richness	Genera richness
Atlantic		159	148	161	1186	372
Indian	84		678	137	3236	809
Pacific	74	2383		142	3919	815
TEP	65	103	120		634	281

Madagascar and Mauritius rises to approximately half the IAA with more than 950 species. The Atlantic Ocean and the Tropical Eastern Pacific (hereafter TEP) showed a comparatively lower richness in reef fishes. The richest area in the TEP was Panama with 330 species, while the richest area of the Atlantic Ocean was located between Cuba and the Virgin Islands with 461 species.

The final models obtained using either locations defined as polygons of varying size or standard areas within a 600 km radius yielded almost identical results. Therefore, we report only the results obtained using polygons of varying sizes for each location because they better reflect the basis of our fish data, while the results obtained using standard areas are reported in Supplementary material Appendix 4.

The final model accounted for all seven predictor variables, none of which was removed by the simplification procedure. Their relative influences are in descending order: coral reef area, region, coast length, connectivity, normalized

distance to the MD, mean SST and SST range (Fig. 3). Area predictors (i.e. coral reef area and coast length) showed a more or less complex positive asymptotic relationship with species richness (Fig. 4). A positive relationship with fish richness was found for the mean SST and connectivity as well. The relationship between fish richness and the distance from MD was instead negative and the SST range was linked to fish richness by a complex hump-shaped relationship.

The final BRT model showed a high predictive performance, explaining 80% of the total deviance using a cross-validated procedure. The relationship between observed and predicted species richness was remarkably high ($R^2 = 0.95$ for a regression of slope 1 and 0 intercept, Fig. 5). In addition, uncertainty around predictions was extremely low, generally being $< 2\%$ of the observed richness at a given location (see maps in Supplementary material Appendix 5).

Interaction between predictors played an important role. BRTs considering interactions between predictors performed much better than simpler BRTs (Supplementary material Appendix 6). Region, in particular, was the most 'interacting' variable and had relatively strong interactions especially with coral reef area, distance from MD, and coast length, thus highlighting that these factors have different effects in each region (Supplementary material Appendix 6, Fig. 4). In general individual effects in the Atlantic Ocean and TEP appeared to be less pronounced than in the Pacific and Indian oceans. This is partially due to an 'intercept' effect, since the diversity of the Indian and Pacific oceans is much higher than that of the other regions.
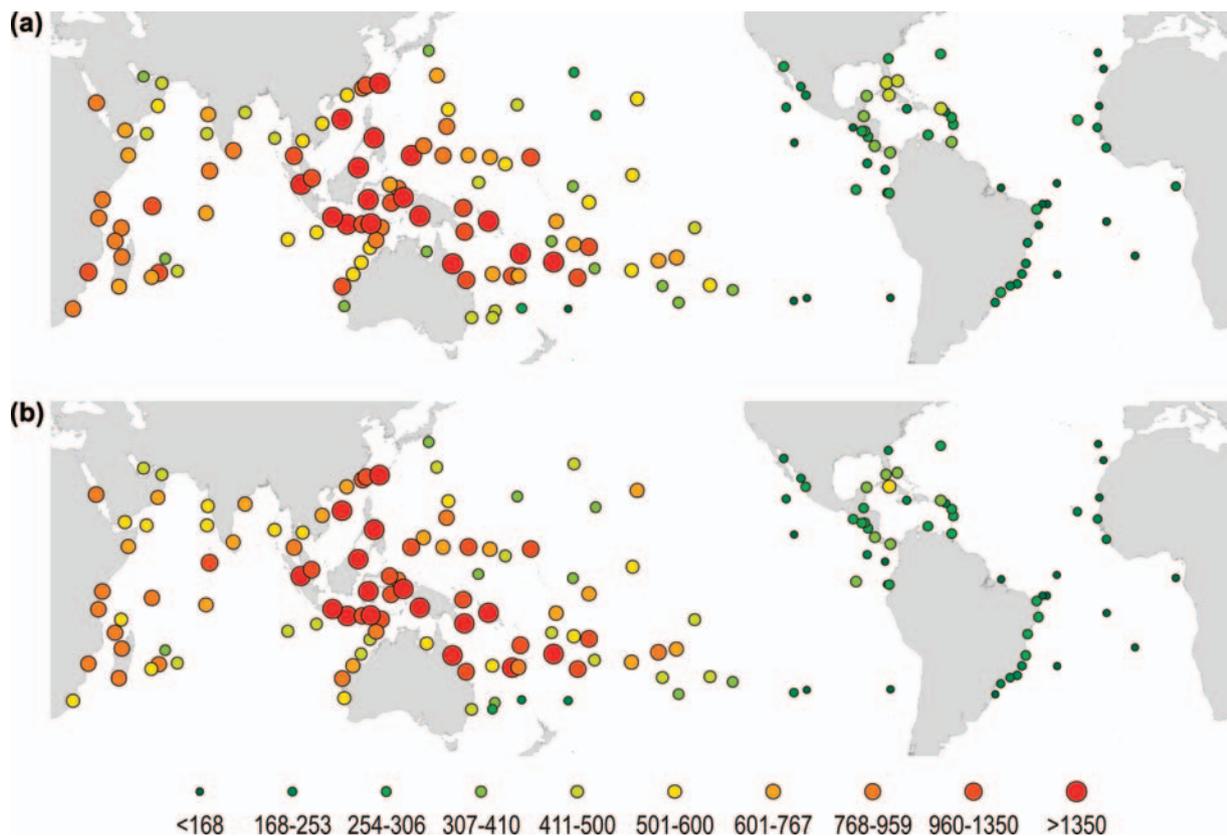
Figure 2. (a) Map of the observed richness of reef-associated fishes; (b) map of the richness of reef-associated fishes predicted by the BRT model. Colors and circle size represent 10 richness classes indentified based on the observed richness. The same scheme was later applied on the map of predicted richness.
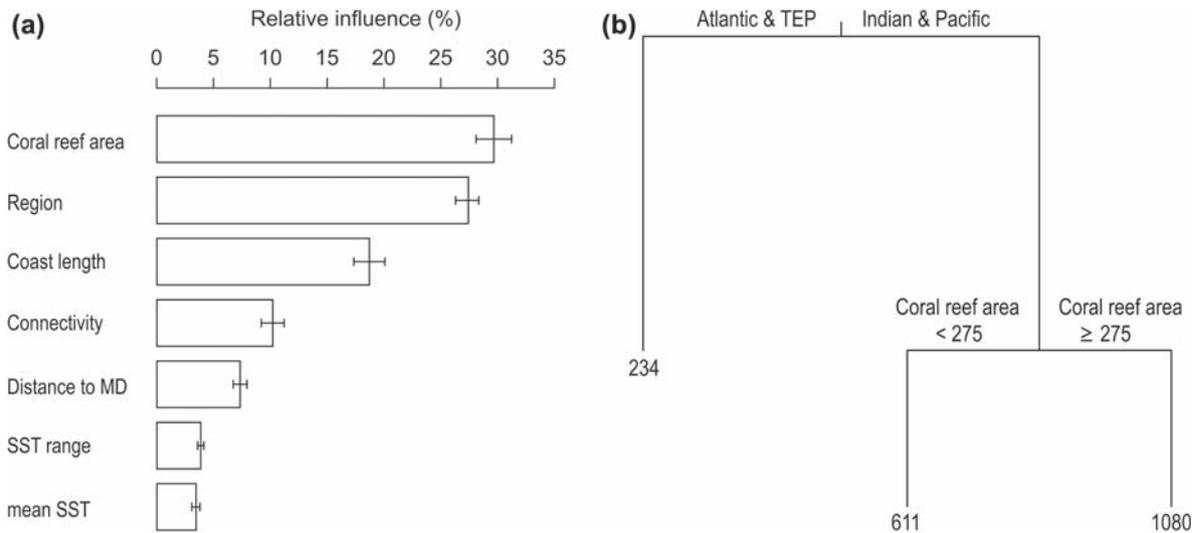
Figure 3. (a) Relative influence of the predictor variables as ranked by the final BRT model. Error bars represent 95% confidence intervals obtained from 1000 bootstrap samples of the original dataset. (b) First tree fitted by the BRT. The end points of the tree (nodes) represent mutually exclusive combinations of independent variables. At each node, a mean estimate of the dependent variable is calculated from the data points falling within that node. The final BRT model contains 1700 trees.

The tests for detecting spatial autocorrelation in the residuals were not significant for all the spatial lags considered (at 1st lag: Moran's I = 0.0528, p = 0.068, Supplementary material Appendix 7), indicating that a spatial structure in overestimation or underestimation of species richness was not found.

The structural equation model was highly significant (χ^2 = 214, DF = 12, p < 0.0001) and explained 54% of the total variation in fish richness. However, we detected significant spatial autocorrelation in SEM residuals (at 1st lag: Moran's I = 0.1708, p < 0.0001, Supplementary

material Appendix 7). Since the violation of spatial independence tends to inflate type I errors (Legendre 1993), we limited the interpretation of the SEM model to the path coefficients, without reporting information on the significance of the effects (Kissling et al. 2008). SEM results were generally consistent with the results obtained with the BRT, particularly the ranking of direct predictor effects (Fig. 6). Path coefficients highlighted that region played mainly a direct role on species richness; its effect on coral reef and connectivity was comparatively lower.
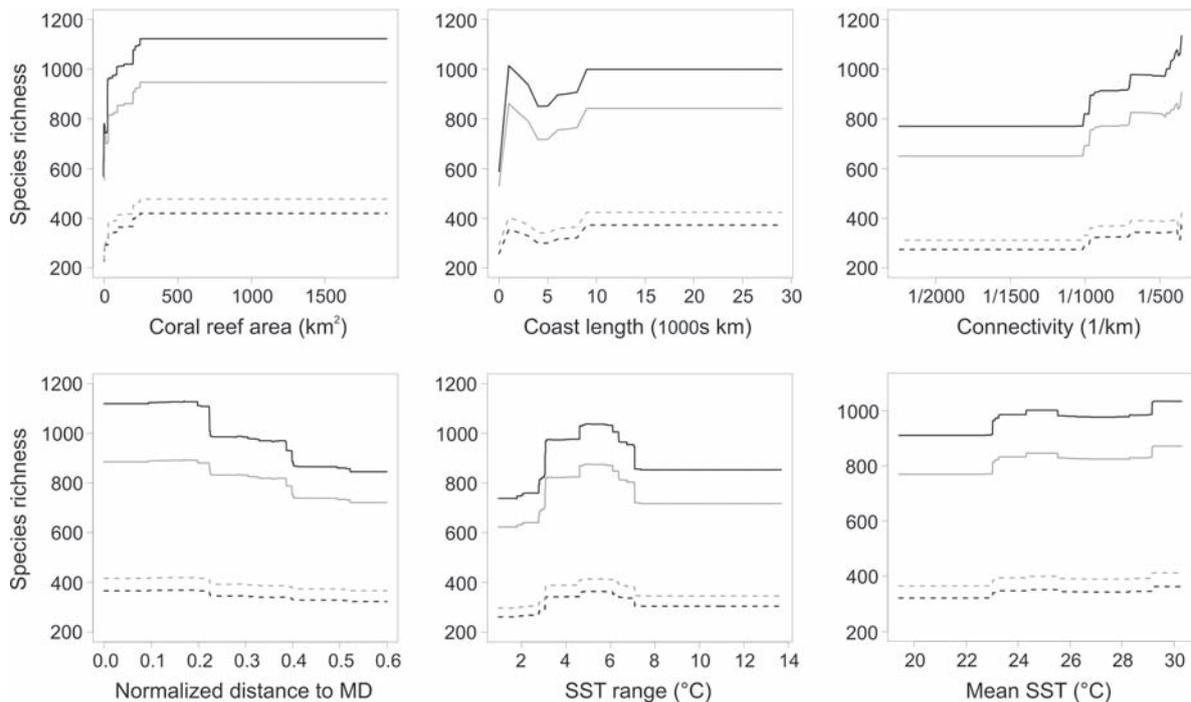
Figure 4. BRT model predictions of reef-associated fish richness according to several explanatory variables for the different regions. Values for predictors other than those on the x-axis are held constant at their mean. Black lines: Pacific Ocean; grey lines: Indian Ocean; dotted black lines: Atlantic Ocean; dotted grey lines: Tropical Eastern Pacific.
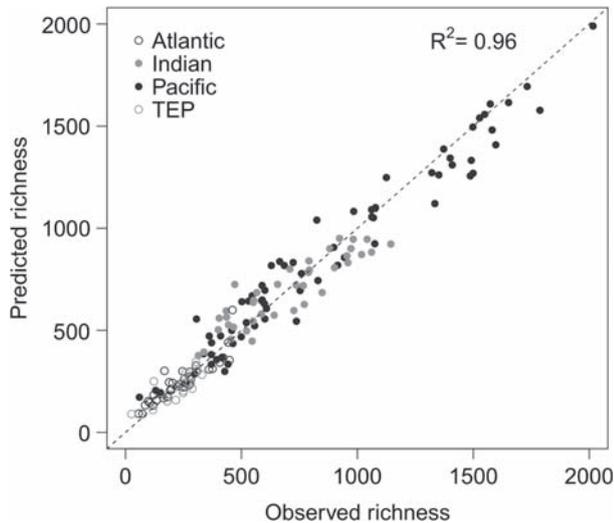
Figure 5. Relationship between observed species richness at each location and species richness predicted by the BRT model. Empty black circles: Atlantic Ocean; empty grey circles: Tropical Eastern Pacific (TEP); filled black circles: Pacific Ocean; filled grey circles: Indian Ocean.

In contrast, mean SST showed a strong positive indirect effect on species richness, mediated by coral reef area.

Discussion

We analyzed the relative contribution of a range of variables in explaining current-day tropical reef fish richness at a global scale. Our analysis revealed the potential for biogeographical history to interact with present environmental variables to shape the distribution of reef fish richness. Environmental predictors, in fact, showed different effects according to the biogeographical region. This has been previously documented for terrestrial environments, where the role of present environmental variables has been shown to differentially shape amphibians richness in different biogeographical realms (Buckley and Jetz 2007). Likewise, with deeper insights into evolutionary
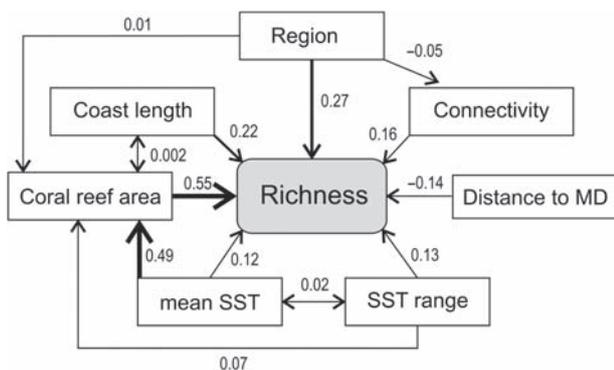
Figure 6. Results of the structural equation model performed to assess direct and indirect effects of explanatory variables on reef fish species richness. Numbers represent standardized path coefficients. Arrows thickness is proportional to path coefficients. Since we detected spatial autocorrelation for SEM residuals potentially inflating type I error, we reported all the path coefficients without considering p-values.

history, Jetz and Fine (2012) demonstrate how jointly considering historical and contemporary environmental predictors accounts for more than 80% of the variation in the diversity of terrestrial vertebrates at a global scale.

The analysis of interaction strength between explanatory variables highlighted that the region (used here as a proxy of evolutionary history) established strong interactions, especially with the MD effect, coral reef area, coast length and SST range. The effect of evolutionary and biogeographical history on reef fish distribution patterns has been documented worldwide (Floeter et al. 2008, Bellwood et al. 2012, Cowman and Bellwood 2013) and has underpinned major shifts in species richness distribution during the past 50 million yr (Renema et al. 2008). Our results and the interactions detected between region and other predictors are consistent with the hypothesis that a combined effect of phylogenetic and biogeographical histories, and present environmental characteristics is likely to shape the present geographical pattern of reef fish richness. In this regard, it is worth mentioning that the relative importance of variables obtained with the BRT model includes most of their interactions. This also explains the higher influence of the region factor recorded by BRT compared to SEM models. While BRT accounts for interactions between predictors but not for mediation effects, SEM does the opposite and their results should be seen as complementary. The spatial autocorrelation detected in SEM residuals but not in BRT residuals is likely due to the higher predictive performance of the latter models which show a low proportion of unexplained variation.

Besides history, our analyses evidence a primary role of area variables and, in particular, of coral reef area. This is in agreement with the findings of Bellwood et al. (2005), who identified coral reef area and MD effect as the key variables explaining most variation in reef fish diversity patterns across the Indo-Pacific. However, contrary to their analyses, we detect only a minor role of the MD effect. This is, at least in part, due to the global scale of our work, embracing five separate domains, whereas Bellwood et al. (2005) focused on the Indo-Pacific domain. Partial dependence plots confirm an important role of the MD in the Pacific Ocean while its effect is less influential in the Indian Ocean and especially in the TEP and the Atlantic Ocean (Fig. 4). The comparatively lower influence of the MD effect in the Indian Ocean is likely due to the high richness of the south-west coast of the basin reaching a fish species richness almost half that of the IAA. This secondary hotspot has already been documented to be the cause of a main deviation from the null expectation depicted by the sole MD effect (Connolly et al. 2003). Despite the high number of species that the Indian Ocean shares with the Pacific, our findings suggest that the Indian Ocean may act, at least in part, as a separate domain. The negligible role of the MD effect in the Atlantic Ocean, is probably due to an asymmetric distribution of shallow water or reefal habitats, low diversity faunas, limited geographical extents, and extensive extinction (O'Dea et al. 2007, Floeter et al. 2008).

Our analyses also evidence an important role of connectivity, especially in the Indian and Pacific Oceans. This result is highly consistent with Cowman and Bellwood (2013) who highlight the evolutionary role of dispersal from the IAA

in the development of Indian and Pacific Ocean reef fish faunas. Our results are also in agreement with the findings of Mora et al.(2003) who explained the low richness characterizing isolated Pacific islands as the result of a filtering from the IAA (cf. the center of origin hypothesis; Bellwood et al. 2012). Similarly, although working at smaller scales, Sandin et al. (2008) report an important role of connectivity on species richness in the Caribbean.

Our results highlighted a limited influence of SST related variables. This is in agreement with previous research conducted in the Indo-Pacific (Bellwood et al. 2005), but in contrast with other studies conducted in the Atlantic Ocean or at a global scale (Macpherson 2002, Tittensor et al. 2010). This is probably due to the extent of the latter works which explored a broader latitudinal gradient, including multiple taxa in temperate and polar environments. From this perspective, the relationship between species richness and temperature may be not linear, thereby exerting a significant role at the interface between cold-temperate and tropical waters, but a minor effect once tropical temperatures (i.e. higher than 17°C in our case) are reached.

Structural equation models provided further insight into the way the predictors of species richness may interact. In this regard, we expected history to play mainly indirectly on reef fish richness, most isolated islands and coral reefs being asymmetrically located in the Pacific Ocean. Our results, however, suggest that the region factor has a major direct effect on species richness and help to disentangle the relationship between fish richness and the other predictors for different regions. SST, instead, has a major indirect role on reef fish richness mediated by coral reef area. This partially clarifies the minor role of temperature on reef species richness documented by previous research (Bellwood et al. 2005) and from the BRT models in the present work. Coral reef development is known to be severely limited by SST (Kleypas et al. 1999) and, according to the results obtained by SEM, its effect on fish richness may be interpreted as linked to habitat provision more than to the energy hypothesis itself.

One of the most striking results was the among-region variation in the response of fish diversity to environmental variables. Previous work has reported variation among major marine groups in their response to environmental variables. Macroalgae, for example, have been reported to be more susceptible than fishes to SST (Kerswell 2006). Variation within a taxon is, in some ways, counterintuitive. One may expect fishes to react to temperature in a similar manner, regardless of the region. The observed variation among regions, however, may have both methodological and historical explanations. In terms of methodology, the regions vary greatly in the magnitude of the gradients observed. For example, the area of reef habitat in the Pacific (within 600 km radius areas) varies from 0 km^2 (e.g. Sala Y Gomez) to $>$ 10 000 km^2 (Great Barrier Reef), while in the TEP it ranges from 0 to about 100 km^2 (Costa Rica), as the entire TEP has $<$ 200 km^2 of reef area. Thus the gradients, and the capacity of fishes to respond, are region-specific.

Historically, the various regions have experienced markedly different evolutionary histories. The Indian and Pacific Oceans have experienced extended periods of connectivity, and within these oceans reef fish lineages have exhibited expansion and dispersal (Cowman and Bellwood 2013). In contrast, the geographically restricted Caribbean, Atlantic and TEP have had a long history of isolation, with the resultant extinction of marine taxa and limited diversification within the remaining fish lineages (Bellwood and Wainwright 2002, O'Dea et al. 2007, Cowman and Bellwood 2013). This evolutionary filtering of the fish faunas may, at least in part, help explain why Caribbean fish assemblages have much stronger associations with non-reef habitats (rocky reefs, mangroves, seagrasses etc) than their Indo-Pacific counterparts (Bellwood and Wainwright 2002, Barnes et al. 2012); because they had to survive during periods with limited reef area. It may also explain why the composition of Caribbean and TEP reef fish faunas have a stronger resemblance to those of non-tropical coastal faunas, such as that of New Zealand, rather than low-latitude coral reef assemblages (Bellwood and Wainwright 2002). It is this history that may explain, at least in part, the limited response of TEP and Atlantic fish faunas to reef area, temperature or other reef-associated variables.

Our results have important implications in terms of reef fish conservation. Coral reef area was the most influential variable according to both BRT and SEM analyses. However coral reefs are presently at risk and under multiple human pressures, including climate change, habitat destruction, pollution, over-fishing, and ocean acidification (Pandolfi et al. 2011) leading to the reduction of live coral cover and suitable habitat. In addition, our findings show an important positive effect of mean SST on reef fish richness, mediated by coral reef area. Anthropogenic climate change has already triggered profound responses of coral reefs due to increasing temperature and positive thermal peaks that have resulted in massive die-off of coral species and a significant decrease of coral cover worldwide (Goreau et al. 2000) and their associated fishes (Wilson et al. 2006). Human pressure is thus acting on the main predictors of reef fish species richness both directly by local impacts, and indirectly by global warming. Given the strong interactions identified, our results suggest that different regions will respond differently to ongoing anthropogenic pressures.

Our work demonstrates that a few and readily available variables are sufficient to accurately predict the pattern of variation in reef fish species richness at global scale. We found a very strong relationship between observed and predicted fish species richness and low uncertainty around prediction estimates. The high proportion of cross-validated deviance highlights the applicability of our model to new locations, thereby representing an important tool for attempting filling the gaps in those areas where information on fish species richness is still unavailable or inaccessible.

References

Allen, A. P. et al. 2002. global biodiversity, biochemical kinetics, and the energetic-equivalence rule. – Science 297: 1545–1548.

Andréfouët, S. et al. 2006. Global assessment of modern coral reef extent and diversity for regional science and management applications: a view from space. – Proceedings of the International Coral Reef Symposium.

Barnes, L. et al. 2012. The use of clear-water non-estuarine mangroves by reef fishes on the Great Barrier Reef. – Mar. Biol. 159: 211–220.

Bellwood, D. R. and Hughes, T. P. 2001. Regional-scale assembly rules and biodiversity of coral reefs. – Science 292: 1532–1534.

Bellwood, D. R. and Wainwright, P. C. 2002. The history and biogeography of fishes on coral reefs. – In: Sale, P. F. (ed.), Coral reef fishes. Dynamics and diversity on a complex ecosystem. Academic Press, pp. 5–32.

Bellwood, D. R. et al. 2004. Confronting the coral reef crisis. – Nature 429: 827–833.

Bellwood, D. R. et al. 2005. Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. – Ecol. Lett. 8: 643–651.

Bellwood, D. R. et al. 2012. Biodiversity hotspots, evolution and coral reef biogeography: a review. – In: Gower, D. et al. (eds), Biotic evolution and environmental change in southeast Asia. Cambridge Univ. Press, pp. 2–32.

Belmaker, J. et al. 2011. The influence of connectivity on richness and temporal variation of reef fishes. – Landscape Ecol. 76: 113–133.

Bivand, R. et al. 2013. spdep: spatial dependence: weighting schemes, statistics and models. – R package ver. 0.5-56, <http://CRAN.R-project.org/package=spdep>.

Buckley, L. B. and Jetz, W. 2007. Environmental and historical constraints on global patterns of amphibian richness. – Proc. Biol. Sci. 274: 1167–1173.

Colwell, R. K. and Hurtt, G. C. 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. – Am. Nat. 144: 570–595.

Connolly, S. R. et al. 2003. Indo-Pacific biodiversity of coral reefs: deviations from a mid-domain model. – Ecology 84: 2178–2190.

Cowman, P. F. and Bellwood, D. R. 2013. The historical biogeography of coral reef fishes: global patterns of origination and dispersal. – J. Biogeogr. 40: 209–224.

Currie, D. J. and Kerr, J. T. 2008. Tests of the mid-domain hypothesis: a review of the evidence of the mid-domain. – Ecol. Monogr. 78: 3–18.

Dormann, C. F. et al. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. – Ecography 35: 1–20.

Elith, J. et al. 2008. A working guide to boosted regression trees. – J. Anim. Ecol. 77: 802–813.

Floeter, S. R. et al. 2008. Atlantic reef fish biogeography and evolution. – J. Biogeogr. 35: 22–47.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. – Ann. Stat. 29: 1189–1232.

Gaston, K. J. 2000. Global patterns in biodiversity. – Nature 405: 220–227.

Goreau, T. et al. 2000. Conservation of coral reefs after the 1998 global bleaching event. – Conserv. Biol. 14: 5–15.

Gray, J. S. 2001. Marine diversity: the paradigms in patterns of species richness examined. – Sci. Mar. 65: 41–56.

Hawkins, B. A. 2012. Eight (and a half) deadly sins of spatial analysis. – J. Biogeogr. 39: 1–9.

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. Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. – Ecography 30: 375–384.

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.

Kerswell, A.P. 2006. Global biodiversity patterns of benthic marine algae. – Ecology 87: 2479–2488.

Kissling, W. D. et al. 2008. Spatial patterns of woody plant and bird diversity: functional relationships or environmental effects? – Global Ecol. Biogeogr. 17: 327–339.

Kleypas, J. A. et al. 1999. Environmental limits to coral reef development: where do we draw. – Am. Zool. 159: 146–159.

Knowlton, N. and Jackson, J. B. C. 2008. Shifting baselines, local impacts, and global change on coral reefs. – PLoS Biol. 6: e54.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? – Ecology 74: 1659–1673.

Luiz, O. J. et al. 2012. Ecological traits influencing range expansion across large oceanic dispersal barriers: insights from tropical Atlantic reef fishes. – Proc. Biol. Sci. 279: 1033–1040.

Macpherson, E. 2002. Large-scale species-richness gradients in the Atlantic Ocean. – Proc. Biol. Sci. 269: 1715–1720.

Mitchell, R. J. 1992. Testing evolutionary equation modelling path analysis and structural equation modelling. – Funct. Ecol. 6: 123–129.

Moilanen, A. and Nieminen, M. 2002. Simple connectivity measures in spatial ecology. – Ecology 83: 1131–1145.

Mora, C. and Robertson, D. R. 2005. Causes of latitudinal gradient in species richness: a test with fishes of the Tropical Eastern Pacific. – Ecology 86: 1771–1782.

Mora, C. et al. 2003. Patterns and processes in reef fish diversity. – Nature 421: 933–936.

Mora, C. et al. 2012. High connectivity among habitats precludes the relationship between dispersal and range size in tropical reef fishes. – Ecography 35: 89–96.

O'Dea, A. et al. 2007. Environmental change preceded Caribbean extinction by 2 million years. – Proc. Natl Acad. Sci. USA 104: 5501–5506.

Pandolfi, J. M. et al. 2011. Projecting coral reef futures under global warming and ocean acidification. – Science 333: 418–422.

Renema, W. et al. 2008. Hopping hotspots: global shifts in marine biodiversity. – Science 321: 654–657.

Ridgeway, G. 2007. gbm: generalized boosted regression models. – R package ver. 1.6-3, <http://cran.r-project.org/web/packages/gbm.pdf>.

Robertson, D. R. and Cramer, K. L. 2009. Shore fishes and biogeographic subdivisions of the Tropical Eastern Pacific. – Mar. Ecol. Prog. Ser. 380: 1–17.

Robertson, D. R. et al. 2004. Tropical transpacific shore fishes. – Pac. Sci. 58: 507–565.

Rosseel, Y. 2012. lavaan: an R package for structural equation modeling. – J. Stat. Softw. 48: 1–36.

Sandin, S. et al. 2008. Island biogeography of Caribbean coral reef fish. – Global Ecol. Biogeogr. 17: 770–777.

Swearer, S. E. et al. 2002. Evidence of self-recruitment in demersal marine populations. – Bull. Mar. Sci. 70: 251–271.

Tittensor, D. P. et al. 2010. Global patterns and predictors of marine biodiversity across taxa. – Nature 466: 1098–1101.

Tyberghein, L. et al. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. – Global Ecol. Biogeogr. 21: 272–281.

Victor, B. C. and Wellington, G. M. 2000. Endemism and the pelagic larval duration of reef fishes in the eastern Pacific Ocean. – Mar. Ecol. Prog. Ser. 205: 241–248.

Wilson, S. K. et al. 2006. Multiple disturbances and the global degradation of coral reefs: are reef fishes at risk or resilient? – Global Change Biol. 12: 2220–2234.

Worm, B. et al. 2006. Impacts of biodiversity loss on ocean ecosystem services. – Science 314: 787–790.