LETTER

doi:10.1038/nature12529

Integrating abundance and functional traits reveals new global hotspots of fish diversity

Rick D. Stuart-Smith[1], Amanda E. Bates[1], Jonathan S. Lefcheck[2], J. Emmett Duffy[2], Susan C. Baker[3], Russell J. Thomson[1], Jemina F. Stuart-Smith[1], Nicole A. Hill[1], Stuart J. Kininmonth[4], Laura Airoldi[5,6], Mikel A. Becerro[7], Stuart J. Campbell[8], Terence P. Dawson[9], Sergio A. Navarrete[10], German A. Soler[1], Elisabeth M. A. Strain[5], Trevor J. Willis[11] & Graham J. Edgar[1]

Species richness has dominated our view of global biodiversity patterns for centuries[1,2]. The dominance of this paradigm is reflected in the focus by ecologists and conservation managers on richness and associated occurrence-based measures for understanding drivers of broad-scale diversity patterns and as a biological basis for management[3,4]. However, this is changing rapidly, as it is now recognized that not only the number of species but the species present, their phenotypes and the number of individuals of each species are critical in determining the nature and strength of the relationships between species diversity and a range of ecological functions (such as biomass production and nutrient cycling)[5]. Integrating these measures should provide a more relevant representation of global biodiversity patterns in terms of ecological functions than that provided by simple species counts. Here we provide comparisons of a traditional global biodiversity distribution measure based on richness with metrics that incorporate species abundances and functional traits. We use data from standardized quantitative surveys of 2,473 marine reef fish species at 1,844 sites, spanning 133 degrees of latitude from all ocean basins, to identify new diversity hotspots in some temperate regions and the tropical eastern Pacific Ocean. These relate to high diversity of functional traits amongst individuals in the community (calculated using Rao's *Q*[6]), and differ from previously reported patterns in functional diversity and richness for terrestrial animals, which emphasize species-rich tropical regions only[7,8]. There is a global trend for greater evenness in the number of individuals of each species, across the reef fish species observed at sites ('community evenness'), at higher latitudes. This contributes to the distribution of functional diversity hotspots and contrasts with well-known latitudinal gradients in richness[2,4]. Our findings suggest that the contribution of species diversity to a range of ecosystem functions varies over large scales, and imply that in tropical regions, which have higher numbers of species, each species contributes proportionally less to community-level ecological processes on average than species in temperate regions. Metrics of ecological function usefully complement metrics of species diversity in conservation management, including when identifying planning priorities and when tracking changes to biodiversity values.

Species do not contribute equally to varied ecosystem processes and delivery of services[9], and thus the sustainability and resilience of these processes depend on aspects of diversity beyond the number of species present in a community. This rationale has led to recent exploration of functional diversity indices to describe variation in traits that mediate species' contributions to ecosystem processes[10]. Community-wide contributions to a range of processes will depend not only on which species and functional traits are present but also on their relative abundances; yet only a few studies of functional diversity have accounted for the uneven distribution of individuals of different species within a community[11]. So far, these studies have not used consistent quantitative sampling to account for the functional consequences of widely differing species abundances at a global scale.

Our global analysis focuses on reef fishes, using data collected from 4,357 underwater visual transects. Fishes are ideal for a global diversity analysis, as they represent the most diverse vertebrate group, encompass a range of important and diverse ecological roles, and marine fishes are relatively well known taxonomically and functionally. We build on prior analyses of coral reef fish diversity[12,13] by substantially expanding the reef fish database in the world's temperate and sub-polar regions, providing the first global maps of abundance-weighted functional diversity for any taxon in the oceans, and assessing congruence between richness and abundance-weighted functional diversity measures.

Our functional diversity index summarizes community diversity with respect to eight traits that contribute to the functional position of species in the ecosystem, encompassing body size (maximum length), feeding ecology (trophic group, trophic breadth), behaviour (water column position, diel activity pattern, gregariousness), and habitat use (preferred substratum, habitat complexity) (Extended Data Table 1). We calculate functional diversity using Rao's quadratic entropy[6], *Q*, a measure of community-level dispersion of species in functional trait space weighted by their relative abundances. Rao's *Q* is not mathematically constrained to be positively correlated with species richness[6], and thus allows unbiased tests of the relationships between species diversity and functional diversity. We use functional group richness to represent the richness of species groups sharing similar functional traits, for the purpose of describing a traditional representation of functional diversity, and for comparison with other diversity metrics.

Previous global and broad-scale studies of marine fishes based on compilations of species distribution records have shown that species richness declines with increasing latitude and distance from hotspots in southeast Asia and the Caribbean[12–14]. Here we corroborate these patterns for reef fishes using the number of species observed in a standardized area at the scale of individual reefs ('species density'; Fig. 1a). Global mapping was based on predictions using broad-scale environmental variables, from random forest models that were generated from the site-level diversity measures and corresponding environmental data, with the effect of inherent spatial autocorrelation evaluated (see Methods). As expected from studies of other taxa[7], the richness of functional groups (Fig. 1b) closely resembles the pattern in species density, with more trait combinations present in speciose tropical regions.

Functional diversity of reef fish communities, incorporating relative abundances of species, shows markedly different global patterns to

[1]Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 49, Hobart, Tasmania 7001, Australia. [2]Department of Biological Sciences, Virginia Institute of Marine Science, P.O. Box 1346, Gloucester Point, Virginia 23062-1346, USA. [3]School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia. [4]Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, Stockholm SE-114 19, Sweden. [5]Dipartimento di Scienze Biologiche, Geologiche ed Ambientali, Università di Bologna, Via San Alberto 163, Ravenna 48123, Italy. [6]Hopkins Marine Station, Stanford University, Pacific Grove, California 93950, USA. [7]Natural Products and Agrobiology Institute (IPNA-CSIC), Avda Astrofísico Francisco Sanchez 3, La Laguna 38206, Tenerife, Spain. [8]Wildlife Conservation Society, Indonesia Marine Program, Jalan Atletik no. 8, Bogor, Jawa Barat 16151, Indonesia. [9]School of the Environment, University of Dundee, Perth Road, Dundee DD1 4HN, Scotland, UK. [10]Estación Costera de Investigaciones Marinas & Center for Marine Conservation, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile. [11]School of Biological Sciences, Institute of Marine Sciences, University of Portsmouth, Ferry Road, Eastney, Portsmouth PO4 9LY, UK.

©2013 Macmillan Publishers Limited. All rights reserved

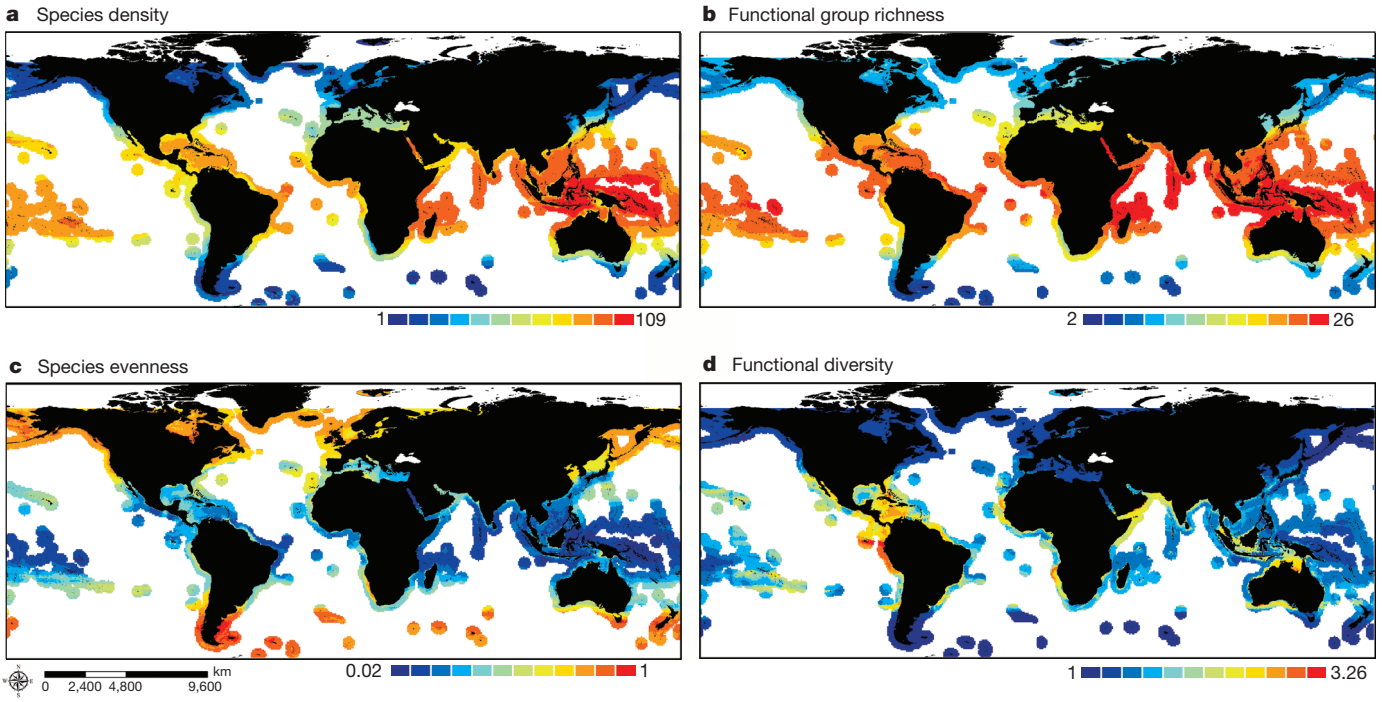


Figure 1 | Global fish diversity patterns predicted from quantitative diver censuses at 1,844 sites. **a**, Species density (a relative measure of species richness) matches the previously established gradient (decreasing with higher latitudes). **b**, Functional group richness mirrors species density. **c**, Evenness (the spread of the number of individuals of each species, across species) increases with latitude. **d**, Abundance-weighted functional diversity is highest in the tropical eastern Pacific Ocean and dispersed hotspots at a range of latitudes. Colour classifications differ between maps owing to different ranges and distributions of diversity values. Minimum and maximum observed values are provided in the key for each plot as effective numbers per 500 m^2 (see Methods).

those of species density and functional group richness (Fig. 1d), and is influenced by a previously undocumented global trend of increasing community evenness with latitude (Fig. 1c). Although current understanding of global patterns in functional diversity is based largely on richness-based measures[7], and thus highlights only tropical regions as hotspots, we found only a weak relationship between functional group richness and abundance-weighted functional diversity ($r = 0.21$; Extended Data Fig. 1; this was weaker when functional richness was expressed as the volume of trait space occupied). It is well-documented that the ecological effects of a species are generally proportional to its abundance or biomass[15]. Thus, by incorporating observed variation in abundances across species within sites, abundance-weighted functional diversity more accurately reflects community functional structure than diversity metrics based on a simple count or inventory of species present within a broader geographic region. Our abundance-weighted metric provides an improved measure of functionally relevant biodiversity and an important new perspective to our understanding of global biodiversity. In addition to some tropical regions, new hotspots of fish diversity are predicted for nutrient-rich regions, such as in cool boundary currents, where plankton and macroalgal trophic pathways are prominent. Conversely, oligotrophic regions, including some that are heavily impacted such as the Mediterranean, generally have low functional diversity (Fig. 1d), regardless of whether the total number of functional groups is high.

The reasons for high functional diversity are not consistent across hotspots. Sites that were surveyed in temperate regions were generally characterized by higher functional diversity relative to species diversity (or richness) than sites in the tropics (Fig. 2 and Extended Data Table 2) and greater evenness in abundances across species (high community evenness; Fig. 1c). Thus, temperate functional diversity hotspots result from high average variation between individual fish in terms of their functional traits and likely functional roles ('high individual complementarity'). Tropical hotspots generally occur in regions where moderate to high functional group richness coincides with regions of moderate community evenness, and thus can result from high average differentiation (complementarity) across either species or individuals (or both). The Galapagos Archipelago, which effectively comprises an isolated nutrient-rich, temperate-like region straddling the equator[16], is an exception to this pattern. This global hotspot is the result of uniquely high interspecific complementarity, where fish communities are the most functionally diverse, worldwide, for any given level of species diversity (Fig. 2), despite relatively low evenness in abundance.

Our study also provides an important perspective on the relationship between diversity and ecosystem functioning. This has typically been explored in terms of how the number of species influences various ecosystem properties in experimental systems at small scales (see ref. 17), but has not often taken advantage of large-scale field observations[18,19]. Specifically, by focusing on multiple traits and on natural abundances, rather than simply presence or absence of species, we reveal a picture of reef fish functional diversity that varies over geographic gradients in a pattern distinct from global trends in species or functional richness (Fig. 1). The main reason for this discrepancy is that in locations in which species have more even abundance distributions, such as within temperate functional diversity hotspots, a greater proportion of species reach moderate abundances at which their unique trait combinations can significantly influence ecological processes. In contrast, functionally unique tropical species tend to be rare, with a relatively weak average influence on ecosystem processes. However, this does not mean that rare species may not sometimes still have important ecological roles[20], and species with unique traits are clearly important from a conservation perspective[21].

To develop more effective biodiversity management to maximize the reliability of ecosystem processes in a changing environment requires cost-effective, broad-scale assessment and monitoring of many dimensions of biodiversity, including consideration of functional traits[22]. For example, metrics of ecological function can provide an important additional input for marine protected area (MPA) planning and management. Relatively few MPAs are located at temperate latitudes, including

©2013 Macmillan Publishers Limited. All rights reserved

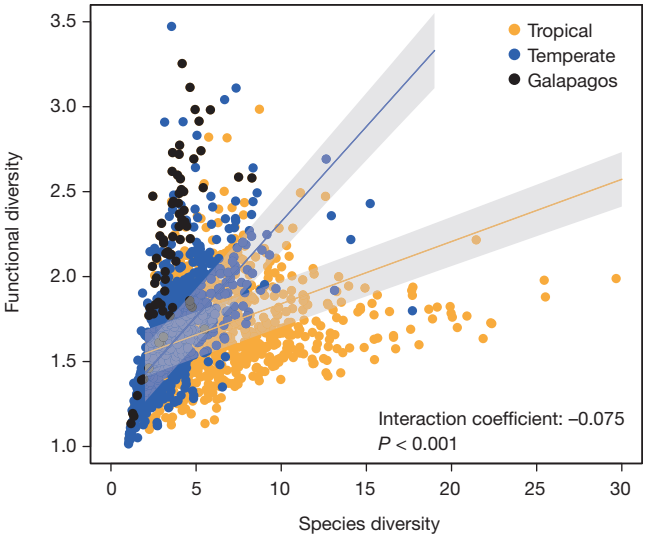


Figure 2 | The species diversity–functional diversity relationship for reef fishes differs between temperate and tropical sites. Although overlap between some temperate and tropical sites was observed, on average temperate sites have higher functional diversity relative to species diversity compared to the tropics ($P < 0.001$, $n = 1{,}817$ sites; see Methods), indicated by divergence of regression slopes predicted by a linear mixed effects model accounting for the random effects of ecological region nested in realm and heteroscedascity (Extended Data Table 2). Both functional diversity and species diversity were calculated using Rao's *Q*, incorporating abundance information and expressed as effective numbers (reducing species diversity to the inverse Simpson index), with functional diversity also incorporating data on functional traits. Grey shading represents 95% confidence intervals. The Galapagos Islands are distinct from the global pattern.

in the southern hemisphere[23], a bias that has been accentuated in recent years with global focus on declaration of large tropical MPAs[24]. Our results identify further unrecognized biodiversity value in some temperate and southern hemisphere marine regions, strengthening the argument for greater representation of these areas in global MPA protection. Nevertheless, as is also the case with species diversity, functional trait data need to be coupled with species identity information to ensure, for example, that MPAs within networks are complementary and encompass as much regional biodiversity as possible at both species and functional levels, rather than replicating ecologically similar patches of high diversity only[25]. Important next steps for building information on functional traits into further conservation applications will include quantification of relationships between metrics of ecological functional and major human impacts (including fishing)[26] and development of associated ecological indicators. Widespread use of such indicators offers enormous potential for prioritizing and informing adaptive management and in depicting global conservation targets, alongside indicators tracking decline of rare and endemic taxa[27].

METHODS SUMMARY

Field data. Standardized visual censuses along 50-m transects were undertaken at 1,844 sites in 74 of the world's marine ecological regions[28] (Extended Data Fig. 2) through the Reef Life Survey program (RLS; http://www.reeflifesurvey.com). RLS utilizes skilled volunteer SCUBA divers trained to scientific data collection standards, who undertake surveys with researchers using a consistent methodology that has been evaluated and proven effective for broad-scale studies[29]. Multiple transects were surveyed at each site (mean 2.4), and species names and abundances of all fishes observed within 5 m either side of transect lines were recorded. Details of quality-control measures are described in the Supplementary Information. Sites were classified as temperate or tropical based on Marine Ecoregions of the World categorization[28].

Diversity indices. Species density was the mean number of fish species per 500 m^2 at each site. Species and functional diversity were calculated using Rao's Q[6], incorporating relative abundance of each species and converted to effective numbers following a method from a previous study[30]. Functional diversity and functional group richness were based on eight functional traits scored for each of the 2,473 fish species in the data set (many from FishBase, http://www.fishbase.org; Extended Data Table 1) and calculated using the functional diversity package in R. Evenness was calculated by dividing species diversity by species density (richness).

Global mapping. Relationships between site diversity index values and 13 physical and environmental variables were modelled using random forest methods, where each forest was created by generating 2,000 regression trees from a bootstrap sample of the data. The same technique was used to predict values of each index across 5 arc-minute grid cells globally for mapping. A parallel reduced model was run using only data from sites spaced greater than the distance at which spatial autocorrelation occurred, which confirmed global predictions were not influenced by inherent correlation of diversity values between sites in close proximity.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 May; accepted 6 August 2013.

1. Gaston, K. J. Global patterns in biodiversity. *Nature* **405,** 220–227 (2000).
2. Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163,** 192–211 (2004).
3. Rodrigues, A. S. L. *et al.* Effectiveness of the global protected area network in representing species diversity. *Nature* **428,** 640–643 (2004).
4. Wiens, J. J. & Donoghue, M. J. Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* **19,** 639–644 (2004).
5. Wittebolle, L. *et al.* Initial community evenness favours functionality under selective stress. *Nature* **458,** 623–626 (2009).
6. Botta-Dukát, Z. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *J. Veg. Sci.* **16,** 533–540 (2005).
7. Safi, K. *et al.* Understanding global patterns of mammalian functional and phylogenetic diversity. *Phil. Trans. R. Soc. B* **366,** 2536–2544 (2011).
8. Stevens, R. D., Cox, S. B., Strauss, R. E. & Willig, M. R. Patterns of functional diversity across an extensive environmental gradient: vertebrate consumers, hidden treatments and latitudinal trends. *Ecol. Lett.* **6,** 1099–1108 (2003).
9. Luck, G. W. *et al.* Quantifying the contribution of organisms to the provision of ecosystem services. *Bioscience* **59,** 223–235 (2009).
10. Cadotte, M. W. The new diversity: management gains through insights into the functional diversity of communities. *J. Appl. Ecol.* **48,** 1067–1069 (2011).
11. Devictor, V. *et al.* Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol. Lett.* **13,** 1030–1040 (2010).
12. Allen, G. R. Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conserv.* **18,** 541–556 (2008).
13. Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295,** 1280–1284 (2002).
14. Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466,** 1098–1101 (2010).
15. Grime, J. P. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86,** 902–910 (1998).
16. Edgar, G. J., Banks, S., Fariña, J. M., Calvopiña, M. & Martínez, C. Regional biogeography of shallow reef fish and macro-invertebrate communities in the Galapagos archipelago. *J. Biogeogr.* **31,** 1107–1124 (2004).
17. Duffy, J. E. Why biodiversity is important to the functioning of real-world ecosystems. *Front. Ecol. Environ* **7,** 437–444 (2009).
18. Gamfeldt, L. *et al.* Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Commun.* **4,** 1340 (2013).
19. Mora, C. *et al.* Global human footprint on the linkage between biodiversity and ecosystem functioning in reef fishes. *PLoS Biol.* **9,** e1000606 (2011).
20. Fonseca, C. R. & Ganade, G. Species functional redundancy, random extinctions and the stability of ecosystems. *J. Ecol.* **89,** 118–125 (2001).
21. Mouillot, D. *et al.* Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol.* **11,** e1001569 (2013).
22. Duffy, J. *et al.* Envisioning a marine biodiversity observation network. *Bioscience* **63,** 350–361 (2013).
23. Wood, L. J., Fish, L., Laughren, J. & Pauly, D. Assessing progress towards global marine protection targets: shortfalls in information and action. *Oryx* **42,** 340–351 (2008).
24. Pala, C. Giant marine reserves pose vast challenges. *Science* **339,** 640–641 (2013).
25. Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405,** 243–253 (2000).
26. Mouillot, D., Graham, N. A. J., Villéger, S., Mason, N. W. H. & Bellwood, D. R. A functional approach reveals community responses to disturbances. *Trends Ecol. Evol.* **28,** 167–177 (2013).
27. Jones, J. P. G. *et al.* The why, what, and how of global biodiversity indicators beyond the 2010 target. *Conserv. Biol.* **25,** 450–457 (2011).
28. Spalding, M. D. *et al.* Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* **57,** 573–583 (2007).
29. Edgar, G. J. & Stuart-Smith, R. D. Ecological effects of marine protected areas on rocky reef communities: a continental-scale analysis. *Mar. Ecol. Prog. Ser.* **388,** 51–62 (2009).
30. Jost, L. Entropy and diversity. *Oikos* **113,** 363–375 (2006).

Supplementary Information is available in the online version of the paper.

©2013 Macmillan Publishers Limited. All rights reserved

Acknowledgements We thank the many Reef Life Survey (RLS) divers who participated in data collection and provide ongoing expertize and commitment to the program, University of Tasmania staff including T. Cooper, M. Davey, N. Barrett, J. Berkhout and E. Oh, and T. Bird for assistance running models and checking code. Development of the RLS data set was supported by the former Commonwealth Environment Research Facilities Program, and analyses were supported by the Australian Research Council, Institute for Marine and Antarctic Studies, and the Marine Biodiversity Hub, a collaborative partnership supported through the Australian Government's National Environmental Research Program. Additional funding and support for field surveys was provided by grants from the National Geographic Society, Conservation International, Wildlife Conservation Society Indonesia, The Winston Churchill Memorial Trust, and ASSEMBLE Marine.

Author Contributions R.D.S.-S., J.S.L., G.J.E., S.C.B. and A.E.B. conceived the idea, G.J.E., R.D.S.-S., J.F.S.-S., S.C.B., S.J.K., G.A.S., E.M.A.S. and many others collected the data. R.D.S.-S., A.E.B., J.E.D., G.J.E., and J.S.L. drafted the paper, with substantial input from S.C.B., R.J.T., J.F.S.-S., N.A.H., S.J.K., L.A., M.A.B., S.J.C., T.P.D., S.A.N., G.A.S., E.M.A.S. and T.J.W. A.E.B., R.J.T. and J.S.L. analysed the data and S.J.K. prepared the maps.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Ecological and trait data are available from the authors on request. Correspondence and requests for materials should be addressed to R.D.S-S. (rstuarts@utas.edu.au).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Fish survey methods. Fish abundance data were collected using standard Reef Life Survey (RLS) methods on 4,357 transects at 1,844 sites, in 74 of the world's marine ecological regions (or 'ecoregions', as defined in a previous paper[28]), with a mean of 2.4 transects per site, at a mean depth of 7.4 m (Extended Data Fig. 2). RLS uses skilled volunteer SCUBA divers trained to scientific data collection standards, who undertake surveys with researchers using a consistent methodology that has been evaluated and proven effective for broad-scale studies[19,29,31]. This methodology is based on fixed-length belt transects in shallow reef habitats, each covering an area of the seafloor 50 × 10 m, up to a height of 5 m (full details are provided in an online methods manual; http://reeflifesurvey.com/files/2008/09/NEW-Methods-Manual_15042013.pdf). All fishes sighted in the transect area (by divers swimming either side of the transect) were recorded on an underwater slate, with abundance estimates made by counting individuals of less abundant species and estimating the number of more abundant species. The majority of fishes were recorded to species level, but unidentified fishes were classified at the highest taxonomic resolution possible rather than omitting them. The use of digital photography typically allowed later identification of many unidentified species, with the assistance of taxonomic experts, as required. Appropriate transformations and diversity indices were used to reduce dominating influences of abundance estimates for the most abundant species.

A summary of the total number of transects, species and mean densities of species and individuals in each of the 11 marine realms[28] is provided in Extended Data Fig. 2b. This shows clear divisions between tropical and temperate realms in terms of densities of species and individuals, with no temperate realm averaging greater than 18 species and 870 individuals per 500 m^2, whereas all tropical realms averaged greater than 25 species and 1,430 individuals per 500 m^2.

The extent and quantity of data used for this study and in the RLS database has only been possible through volunteer efforts. Rigorous measures have been applied to ensure consistency and quality of data, with all divers involved having either substantial prior experience or detailed one-on-one training provided, and extensive data checking applied post dive and before addition to the database. More details on data management and quality control procedures are provided in the Supplementary Information. Only the most skilled subset of divers was involved in the more difficult surveys of diverse coral reefs, where >100 fish species were occasionally recorded on transects.

Functional traits. We compiled values of eight functional traits for each of the 2,473 fish species in the data set (Extended Data Table 1). These were chosen to encompass diverse attributes of fish ecology that are known to influence their functional role in an assemblage, incorporating life history, trophic position, behaviour, and habitat associations. We excluded records of fishes which were not identified to at least Class level (<0.001% of records), and allocated an average value from other members of the same genus (or in some cases, family) for records in which the species was not known, but the genus (or family) was known (0.004% of records). Many trait values were obtained from FishBase (http://www.fishbase.org/), and others were allocated based on the authors' combined knowledge of the species.

The allocation of such a comprehensive and informative suite of traits was only possible because fishes are generally so well-known. Despite this, compiling a trait database for reef fishes globally represented an enormous task, and although we attempted to minimize potential errors as much as possible, we acknowledge that a certain level of error is unavoidable. We make the assumption that there is no systematic bias related to the allocation of traits to species from different regions. Further details relating to potential biases in trait allocation are provided in the Supplementary Information.

Diversity indices. Species density was calculated as the number of species observed per 500-m^2 transect area, whereas species and functional diversity were calculated using Rao's Q[6], functional group richness (FGR) using a functional dendrogram, functional richness as the convex hull volume, and evenness by dividing species diversity by species density (species richness)[32]. Rao's Q, FGR and functional richness were calculated using the functional diversity package in R[33,34] (http://cran.r-project.org/web/packages/FD/FD.pdf), incorporating the functional trait data. We excluded transects where fewer than three species were observed before calculating Rao's Q (for functional diversity) and FGR, and transects with fewer than five species for functional richness. A similarity cut-off of 0.18 was used for FGR, based on the maximum length of basal branches in the functional dendrogram using the method 'mcquitty'. Rao's Q (for functional diversity) and functional richness were estimated using the sqrt-corrected species-by-species Gower distance matrix. Functional diversity was subsequently scaled by its maximum value across all frequency distributions to allow for comparisons across sites. Values of Rao's Q (for species and functional diversity) were converted into effective numbers using a linear transformation provided previously[30], such that the 'doubling property' was satisfied (also making Rao's Q for species diversity the same as the inverse Simpson index). Functional richness was calculated as a convex hull volume generated from 'new traits' returned following principal coordinates analysis (PCoA) on the corrected species-by-species distance matrix (R^2, indicating the quality of the reduced space representation of the traits, was 0.60 with 4 PCoA axes). Gregariousness (1 to 3) was ordered as a factor, and all other traits were coded as unordered factors.

To explore the contributions of individual traits to global patterns in functional diversity, we removed each trait from the full trait matrix and re-calculated Rao's Q (as above) to produce eight functional diversity estimates. We then compared each of the eight functional diversity estimates to functional diversity calculated using the full trait matrix using linear regression. When traits with weaker influence on functional diversity were dropped from the calculation of functional diversity, the change in functional diversity was small and R^2 was close to 1. Those traits that contributed more to functional diversity, when dropped from the trait matrix, resulted in a weaker relationship (lower R^2). This approach therefore allowed the ranking of traits based on the relative change in R^2 when each was dropped from the trait matrix, where lower R^2 values indicated a higher contribution to the overall functional diversity. This process identified water column position, substrate preference and trophic group as contributing most to global patterns in functional diversity (Extended Data Table 3).

The relationship between functional and species diversity presented in Fig. 2 was tested using a linear mixed effects (LME) model, which included the random spatial effects of the ecological region nested in Realm[28], and error structures applied to normalize variance in the residuals. We also introduced a spherical autocorrelation structure to our model, however inclusion of this term did not improve the model fit (on the basis of AIC) and this term was therefore removed. Further detail on the model structure and summary information is provided in Extended Data Table 2. The data used were from 1,817 sites (1,844 minus 27 which had fewer than three species recorded – see above) with both species and functional diversity calculated using Rao's Q and expressed as effective numbers (as above). Sites were classified as temperate or tropical based on their location in the Marine Ecoregions of the World categorizations[28].

For all methods for quantifying fish communities, visual census methods are characterized by large amounts of variation and a number of biases (for example, see ref. 35). We make the assumption that none of these biases are directional in such a way that will influence the values of diversity metrics over the global scale; for example, that species-specific behavioural traits which affect estimates of their density will be consistent throughout their range (for example, at different latitudes).

Predicting and mapping global diversity measures. Relationships between site diversity metric values and 13 environmental and geographic variables (Extended Data Table 4) were modelled using random forest methods[36], in which each forest was created by generating 2,000 regression trees from a bootstrap sample of the data. The same technique was then used to predict values of each index across 5 arc-minute grid cells globally. Environmental variables were obtained from the Bio-ORACLE environmental data set[37]. A human population index was included as a predictor variable, calculated by fitting a smoothly tapered surface to each settlement point on a year 2000 world-population density grid using a quadratic kernel function described previously[38]. Populations were screened for a density greater than 1,000 people per 0.04 degree cell, and the search radius was set at 3.959 degrees.

To estimate prediction error, observations not selected in the bootstrap sample for a tree (the 'out-of-bag' sample) were compared to their predictions, in a similar way to cross-validation. The accuracy importance (Extended Data Fig. 3), or mean decrease in accuracy when the predictor variable is randomly permuted, was measured to assess the importance of each predictor variable. Accuracy was evaluated by comparing the predictions with the actual measures of the response variable, for those sites that were out-of-bag for a given tree. Conditional importance of the predictors[39] was used, as the standard marginal procedure in the random forest methodology can suffer from bias towards correlated predictors[40]. Biodiversity indices were transformed according to the Box–Cox selection technique (Extended Data Table 5), but all predictions presented have been back-transformed. Random forest modelling was undertaken using the 'extendedForest' packages for R (https://r-forge.r-project.org/projects/gradientforest). The fit of each random forest model was measured using the Pearson correlation coefficient between the actual measures and the out-of-bag predictions, before back transformation.

From the generated random forest, predictions were made at new sites by predicting the response variables using each tree individually and then taking the average. Standard deviations of the predictions were used as a measure of prediction error. Predictions of the response variables were made at new sites across the globe and plotted on maps with a coastal buffer.

An assessment of the spatial autocorrelation (SAC) of the global diversity indices was conducted using Moran's I statistic[41], which compared the diversity values at a site with the weighted mean of neighbouring site values. Mean distance of nearest neighbour over all sites was 5.25 km, which was used as the lag distance. All diversity indices were positively spatially autocorrelated in the raw data (Moran's I from 0.23 to 0.61; Extended Data Table 6); an expected result given well-known

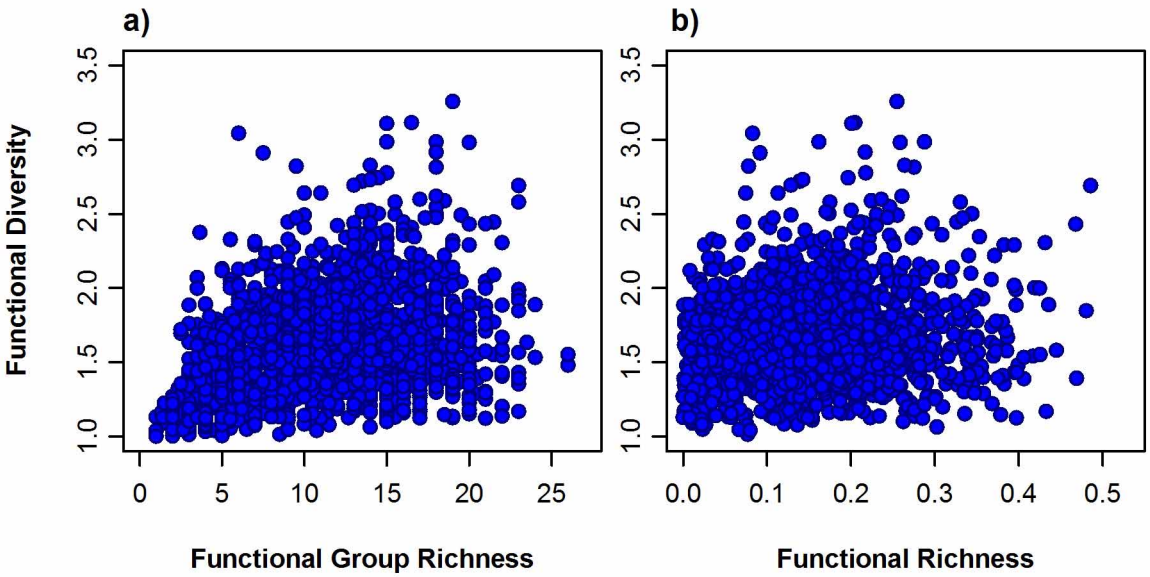
©2013 Macmillan Publishers Limited. All rights reserved

patterns in richness are strongly related to latitude and temperature[14], which are also highly spatially autocorrelated. The presence of SAC can typically result in underestimation of model fitting and incorrect hypothesis testing, but model outputs such as mean parameter estimates and predictions are less often affected[42]. There is no hypothesis testing associated with our RF models, and predictions from RF models have previously been identified to be robust to cluster-correlated data, such as in our data set[42].

Despite this, we tested for SAC in the residuals of the random forest models to be confident that SAC has not created false patterns in our diversity maps presented in Fig. 1. Although the diversity measures based on the raw data had positive SAC (noted above), the residuals of the random forest models had negative SAC (Extended Data Table 6). Thus, residuals at sites which are closer in distance are more dissimilar than residuals at more distant sites than expected. Visual assessment of variograms showed that the significant negative SAC was at scales less than 5 km for all diversity measures, possibly because the predictor values, and hence the predictions, do not vary greatly within 5 km (owing to the grid size for environmental predictor variables in the Bio-ORACLE data set). Based on this, we generated a second set of random forest models using a subset of sites, chosen with a minimum distance apart of five km (the distance at which SAC became non-significant). Comparison of the original predictions with those made from the subset of sites showed tight concordance (Extended Data Fig. 4), providing strong evidence that diversity patterns presented in our global maps are not unduly influenced by SAC.

31. Edgar, G. J., Barrett, N. S. & Stuart-Smith, R. D. Exploited reefs protected from fishing transform over decades into conservation features otherwise absent from seascapes. *Ecol. Appl.* **19,** 1967–1974 (2009).
32. Tuomisto, H. An updated consumer's guide to evenness and related indices. *Oikos* **121,** 1203–1218 (2012).
33. Laliberté, E. & Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91,** 299–305 (2010).
34. Laliberté, E. & Shipley, B. FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-11 (2011).
35. Edgar, G. J., Barrett, N. S. & Morton, A. J. Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations. *J. Exp. Mar. Biol. Ecol.* **308,** 269–290 (2004).
36. Breiman, L. Random forests. *Mach. Learn.* **45,** 5–32 (2001).
37. Tyberghein, L. *et al.* Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob. Ecol. Biogeogr.* **21,** 272–281 (2012).
38. Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. (Chapman and Hall, 1986).
39. Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinformatics* **9,** 307 (2008).
40. Ellis, N., Smith, S. & Pitcher, C. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93,** 156–168 (2012).
41. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37,** 17–23 10.2307/2332142 (1950).
42. Karpievitch, Y. V., Hill, E. G., Leclerc, A. P., Dabney, A. R. & Almeida, J. S. An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. *PLoS ONE* **4,** e7087 (2009).

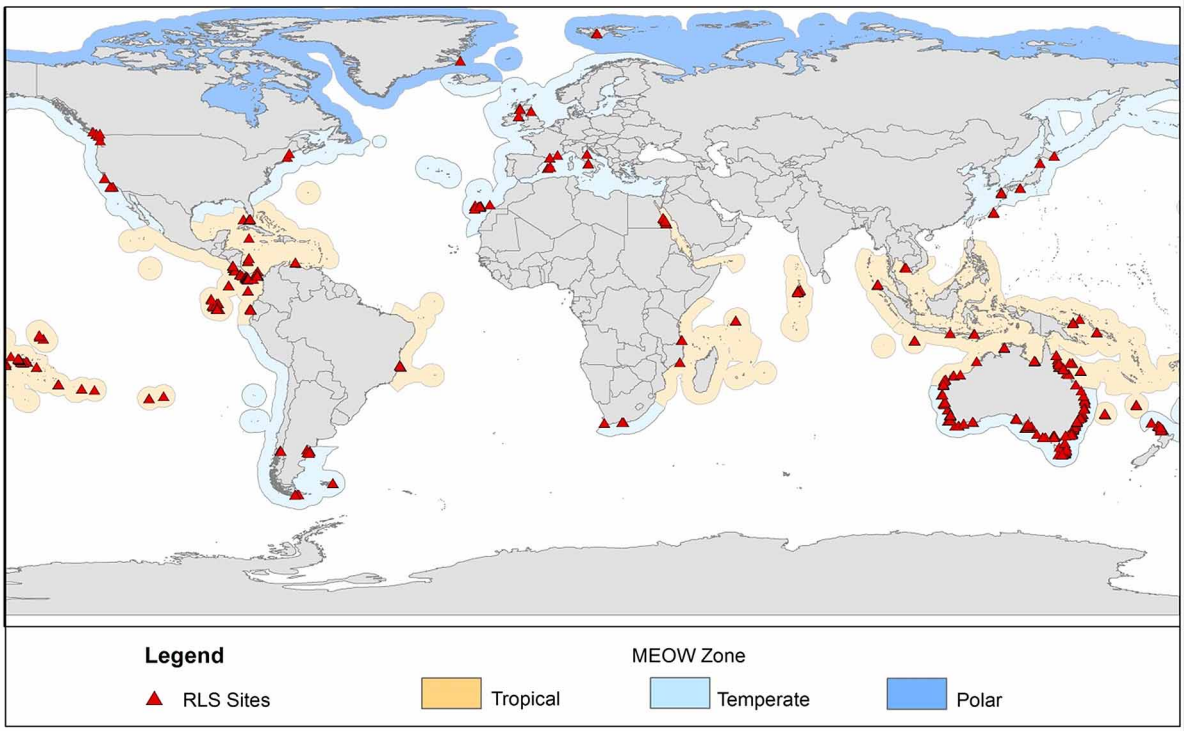
©2013 **Macmillan Publishers Limited. All rights reserved**



Extended Data Figure 1 | Global relationships between different functional diversity measures. **a**, **b**, Abundance-weighted functional diversity (Rao's *Q* expressed as effective numbers; see Methods) for reef fishes provides different information to functional richness expressed as the number of functional groups (functional group richness; FGR) (**a**) and the volume of multidimensional trait space filled by the community (convex hull volume) (**b**). Kendall's Tau correlation coefficients, $r = 0.21$ and 0.11, respectively. Points represent individual reef sites surveyed.

©2013 Macmillan Publishers Limited. All rights reserved

a)

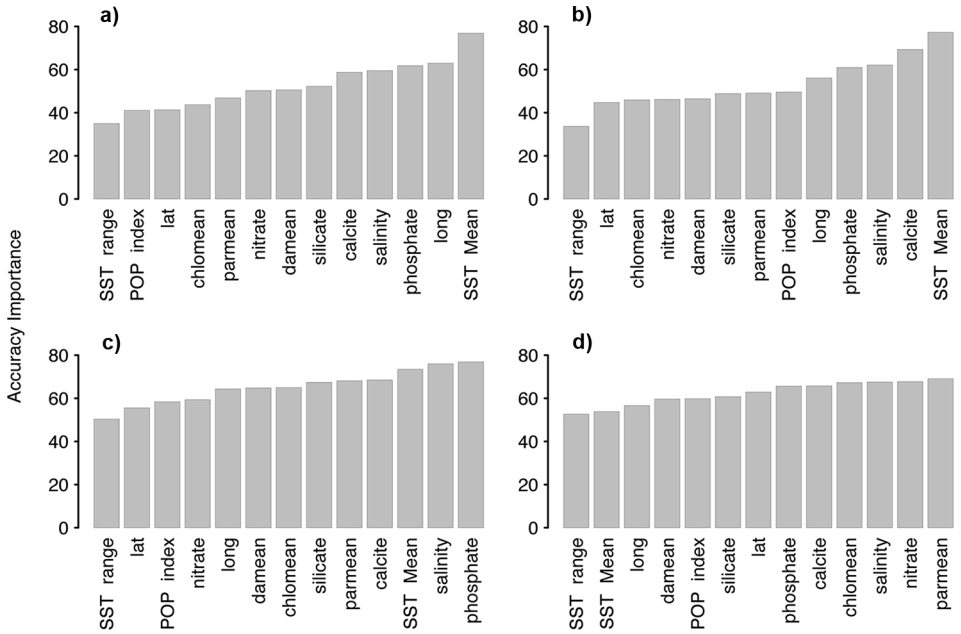


b)

Realm	Transects	Total Species	Mean species density (500 m^{-2} ±SE)	Mean total fish density (500 m^{-2} ±SE)
Arctic	6	5	1.2 (±0.2)	3.5 (±0.9)
Central Indo-Pacific	716	1162	42.6 (±0.8)	1566.3 (±94)
Eastern Indo-Pacific	238	538	43.3 (±1.1)	1669.9 (±124.9)
Temperate Australasia	2239	843	17.7 (±0.2)	865.5 (±38.3)
Temperate Northern Atlantic	200	127	9.8 (±0.3)	586.7 (±57.5)
Temperate Northern Pacific	84	166	11.2 (±0.7)	402.8 (±67.3)
Temperate South America	52	28	3.4 (±0.2)	195.5 (±33.5)
Temperate Southern Africa	24	39	8.7 (±0.7)	267.8 (±51.4)
Tropical Atlantic	132	200	32.7 (±0.7)	1435.9 (±129.2)
Tropical Eastern Pacific	562	249	25.3 (±0.3)	2489.7 (±142)
Western Indo-Pacific	104	528	62.6 (±1.5)	2193.3 (±207.7)

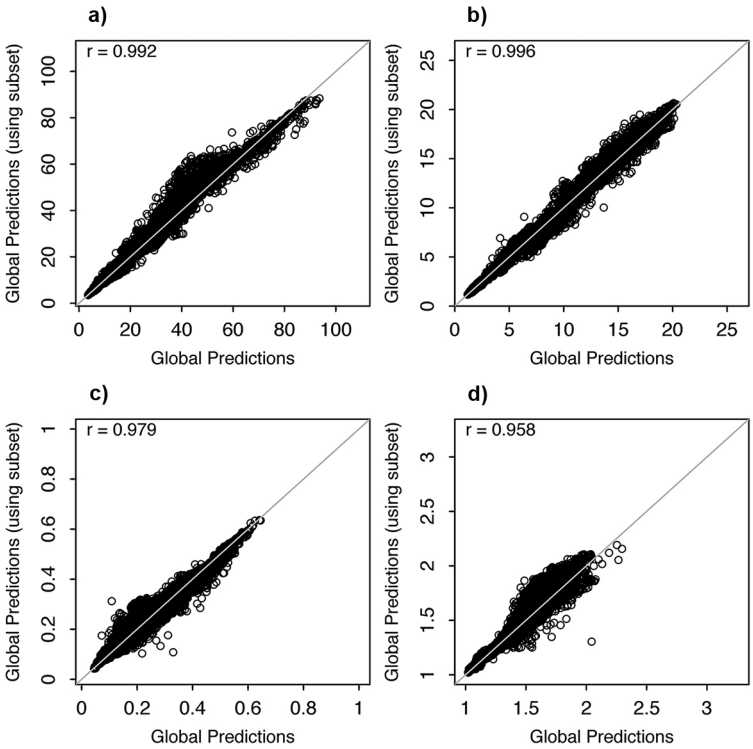
Extended Data Figure 2 | Details of global fish surveys. **a**, **b**, The fish fauna was quantitatively surveyed at 1,844 rocky and coral reef sites in 11 Marine Ecoregions of the World realms by visual census (**a**). Note that many sites are overlapping or hidden behind symbols for other sites. Tropical realms possessed much higher average fish abundance and species densities (**b**).

©2013 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 3 | The accuracy importance of the thirteen predictor variables for each of the random forest models. Models were for species density (**a**), species evenness (**b**), functional group richness (**c**) and functional diversity (**d**). Explanations and units for variables are provided in Extended Data Table 4.

©2013 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Scatter plots comparing global predictions from random forest models used for mapping in Fig. 1 with those based on a training set including only sites >5 km apart. Models were for species density (**a**), functional group richness (**b**), species evenness (**c**) and functional diversity (**d**). Predictions were compared for all global ocean grid cells where the depth was less than 20 m.

©2013 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Functional traits of reef fishes used in estimation of functional diversity.

Functional trait	Category	Type	Units
Maximum length	Body size	Numeric	Total length (cm)*
Trophic breadth	Trophic niche	Numeric	Number of prey phyla consumed (from diet studies*). Range 1-8
Trophic group	Trophic niche	Factor	Browsing herbivore, scraping herbivore, benthic invertivore, planktivore, higher carnivore*
Water column position	Behaviour	Factor	Benthic, demersal, site-attached pelagic, roaming pelagic
Gregariousness	Behaviour	Ordered Factor	Index from 1-3, representing singleton, paired to sometimes forming small schools, always schools
Diel activity pattern	Behaviour	Factor	Nocturnal, diurnal
Preferred substrate	Habitat use	Factor	Hard substrate, soft sediment
Habitat complexity	Habitat use	Factor	Typically associated with habitats characterised by low, medium, high complexity

*Obtained from www.fishbase.org/

©2013 Macmillan Publishers Limited. All rights reserved

Extended Data Table 2 | Linear mixed effects model summary table for functional diversity versus species diversity shown in Fig. 2.

Fixed-effects	**Estimate**	**SE**	**df**	**t-value**	**p-value**
Intercept	1.10	0.078	1744	14.05	<0.001
Species Diversity	0.12	0.004	1744	26.91	<0.001
Latitudinal Zone	0.38	0.108	8	3.48	0.008
Species Diversity * Latitudinal Zone	-0.08	0.005	1744	-16.90	<0.001
Random-effects	**Realm**	**Ecoregion**	**Residual**		
	35%	25%	40%		

The percentage variance explained by each random effect ('ecoregion' nested in 'realm') is reported. Error structure (within-realm standard deviation) was included to normalize variance in the residuals. The model was fitted using maximum likelihood with the nlme package in R.

©2013 **Macmillan Publishers Limited. All rights reserved**

Extended Data Table 3 | Contributions of individual traits to global patterns in functional diversity.

Trait	R^2
Water column position	0.844
Preferred substrate	0.903
Trophic group	0.904
Diel activity pattern	0.908
Habitat complexity	0.934
Gregariousness	0.952
Trophic breadth	0.984
Maximum length	0.998

Traits are listed in order of the magnitude of change in functional diversity arising from their removal, as assessed by R^2, from the relationship with functional diversity calculated using all traits. Ranks are consistent for other correlation measures.

©2013 Macmillan Publishers Limited. All rights reserved

Extended Data Table 4 | Environmental and geographic variables used in random forest models.

Code	Variable	Unit	Range (reef Sites)
nitrate	Nitrate	µmol/l	0.14 - 15.31
SST Mean	Mean Sea Surface Temperature	° C	5.35 - 31.09
damean	Diffuse attenuation coefficient at 490 nm	m^{-1}	0.02 - 0.38
phosphate	Phosphate	µmol/l	0.02 - 2.26
silicate	Silicate	µmol/l	0.31 - 37.9
calcite	Calcite	mol/m^3	5.2^{-5} - 0.45
chlomean	Mean chlorophyll A concentration	Mg/m^3	0.02 - 15.99
salinity	Salinity	PSS	27.4 - 40.1
parmean	Mean Photosynthetically available radiation	$Einstein/m^2$/ day	26.3 - 50.9
SST Range	Sea Surface Temperature Range	° C	0.96 - 17.93
POP Index	Population Index	People/5 km^2	0 - 1.75 million
lat	Latitude	° N	-55.1 - 78.5
long	Longitude	° E	-179 - 175

©2013 Macmillan Publishers Limited. All rights reserved

Extended Data Table 5 | Transformations and correlations of observed to predicted diversity values from random forest models.

Diversity metric	Transformation	*r* (full)	*r* (reduced)
Species Richness	log	0.93	0.95
Functional Group Richness	Square root	0.91	0.93
Evenness	log	0.67	0.67
Functional Diversity	log	0.58	0.62

Correlations are between recorded values at reef sites and those predicted for the maps in Fig. 1 (full model), and only for sites >5 km apart (reduced).

©2013 Macmillan Publishers Limited. All rights reserved

Extended Data Table 6 | Spatial autocorrelation measured by Moran's *I* for diversity metrics calculated from the raw data at reef sites and residuals from the four random forest models used to predict diversity values for global maps

	Species richness	Functional group richness	Evenness	Functional diversity
Raw data	0.61*	0.61*	0.31*	0.23*
RF residuals	-0.42*	-0.47*	-0.42*	-0.41*

*denotes significance for two-sided test (α=0.05)

©2013 Macmillan Publishers Limited. All rights reserved