

## LETTERS

# Global patterns and predictors of marine biodiversity across taxa

Derek P. Tittensor<sup>1</sup>, Camilo Mora<sup>1</sup>, Walter Jetz<sup>2</sup>, Heike K. Lotze<sup>1</sup>, Daniel Ricard<sup>1</sup>, Edward Vanden Berghe<sup>3</sup> & Boris Worm<sup>1</sup>

Global patterns of species richness and their structuring forces have fascinated biologists since Darwin<sup>1,2</sup> and provide critical context for contemporary studies in ecology, evolution and conservation. Anthropogenic impacts and the need for systematic conservation planning have further motivated the analysis of diversity patterns and processes at regional to global scales<sup>3</sup>. Whereas land diversity patterns and their predictors are known for numerous taxa<sup>4,5</sup>, our understanding of global marine diversity has been more limited, with recent findings revealing some striking contrasts to widely held terrestrial paradigms<sup>6–8</sup>. Here we examine global patterns and predictors of species richness across 13 major species groups ranging from zooplankton to marine mammals. **Two major patterns emerged: coastal species showed maximum diversity in the Western Pacific, whereas oceanic groups consistently peaked across broad mid-latitude bands in all oceans. Spatial regression analyses revealed sea surface temperature as the only environmental predictor highly related to diversity across all 13 taxa. Habitat availability and historical factors were also important for coastal species, whereas other predictors had less significance. Areas of high species richness were disproportionately concentrated in regions with medium or higher human impacts. Our findings indicate a fundamental role of temperature or kinetic energy in structuring cross-taxon marine biodiversity, and indicate that changes in ocean temperature, in conjunction with other human impacts, may ultimately rearrange the global distribution of life in the ocean.**

We compiled data on the global distribution of 11,567 species across 13 different taxonomic groups (Table 1 and Supplementary Table 1) and spanning 10 orders of magnitude in body mass. In

contrast to previous syntheses on latitudinal gradients<sup>9</sup>, our focus was on groups where we could analyse two-dimensional spatial patterns on a global scale. These included marine zooplankton (foraminifera and euphausiids), plants (mangroves and seagrasses), invertebrates (stony corals, squids and other cephalopods), fishes (coastal fishes, tunas and billfishes, oceanic and non-oceanic sharks), and mammals (cetaceans and pinnipeds). For each group, we mapped the global distribution of species richness and assessed the extent to which it varied across taxa. We investigated further whether observed patterns were consistent with mechanisms proposed to structure global-scale species diversity patterns, and how marine richness overlapped with recently mapped cumulative human impacts across the world ocean<sup>10</sup>.

We found two distinct spatial patterns of marine species richness. Primarily coastal taxa had peaks of diversity in the western Pacific and showed clear latitudinal gradients along the coasts of continents (Fig. 1a–g). Pinnipeds were an exception, peaking at higher latitudes (Fig. 1d). They depicted an interesting contrast to all other groups owing to their low tropical diversity. Primarily oceanic taxa tended to show pantropical or circumglobal distributions with diversity peaking at latitudes between 20° and 40° in all oceans (Fig. 1h–m). Correlations among diversity patterns supported this separation of oceanic and coastal taxa (Supplementary Table 3).

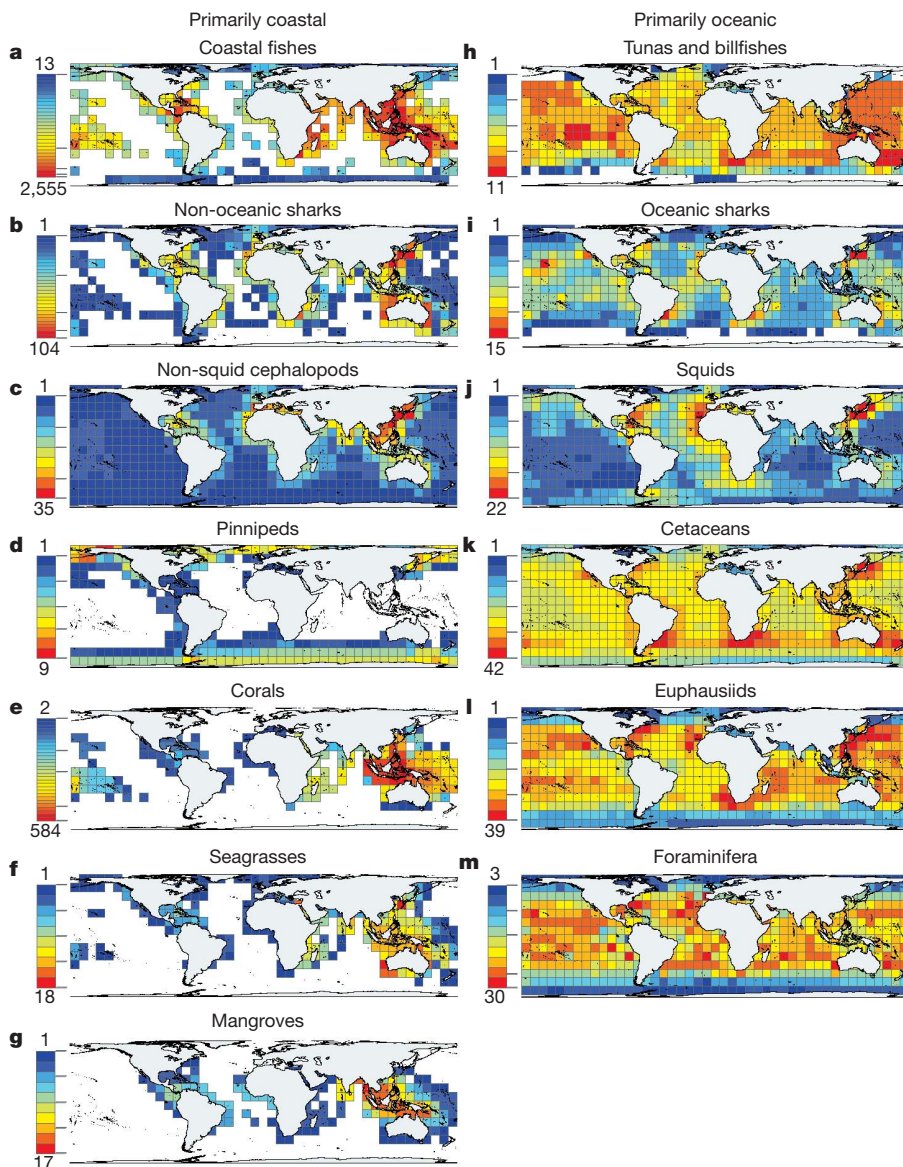
Total species richness across taxa was mostly driven by fishes, the most diverse group we examined, with regional peaks in Southeast Asia (Pacific), Southeast Africa (Indian) and the Caribbean (Atlantic) (Fig. 2a). By normalizing diversity for each taxon, then averaging across all taxa present in each cell, we derived a synthetic pattern of mean diversity (Fig. 2b). The highest mean diversity occurred in

**Table 1 | Spatial modelling of species richness**

Species group	Species richness	Per cent of known	SST	SST slope	Coastline length	Primary productivity	SST range	Oxygen stress	Indian Ocean	Pacific Ocean	Pseudo R <sup>2</sup>
<b>Primarily coastal species</b>											
Coastal fishes	9,713	79	10.7***		7.9***				3.7**	4.5***	0.71
Non-oceanic sharks	480	100	7.1***	2.4*	13.0***	3.6**	−2.5*				0.75
Non-squid cephalopods	122	25	7.1***		6.5***			−1.8*	−2.8**		0.89
Pinnipeds	36	100	−10.0***	4.3**	4.5***	5.5***	−3.2**				0.88
Corals	794	95	7.7***		3.1**			3.8**	3.5**		0.73
Seagrasses	60	100	4.4***		4.3**			2.6*	2.0*		0.75
Mangroves	32	60	9.3***		2.0*						0.85
<b>Primarily oceanic species</b>											
Tunas and billfishes	12	63	7.0***	3.1**			−3.6**				0.76
Oceanic sharks	27	100	11.8***		5.8***						0.81
Squids	85	25	4.0**	2.7**							0.88
Cetaceans	81	96	6.6***			12.1***					0.89
Euphausiids	86	100	6.6***	3.9**		3.4**	−7.8***				0.85
Foraminifera	39	88	16.6***	3.3**	−2.8**			−2.3*			0.79

Number and completeness (per cent of known) of species by taxa and minimal-adequate SLM results for environmental correlates are given. Numbers are z-values; stars represent significance levels at  $P < 0.05$  (\*),  $0.01$  (\*\*), or  $0.00001$  (\*\*\*). Ocean column z-values represent contrast against the Atlantic Ocean.

<sup>1</sup>Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax B3H 4J1, Canada. <sup>2</sup>Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, Connecticut 06520-8106, USA. <sup>3</sup>Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08901-8521, USA.



**Figure 1 | Patterns of species richness for individual taxa.** a–m, Species groups that primarily occur in coastal (a–g) or oceanic habitats (h–m) are shown. Empirically gathered point data (coastal fishes, foraminifera) were co-kriged using GLM predictions as the co-variable<sup>30</sup> to extrapolate to undersampled regions for display purposes. Horizontal tick marks on colour-bars indicate richness quartiles; colour-scaling is adjusted by taxon to optimize contrast.

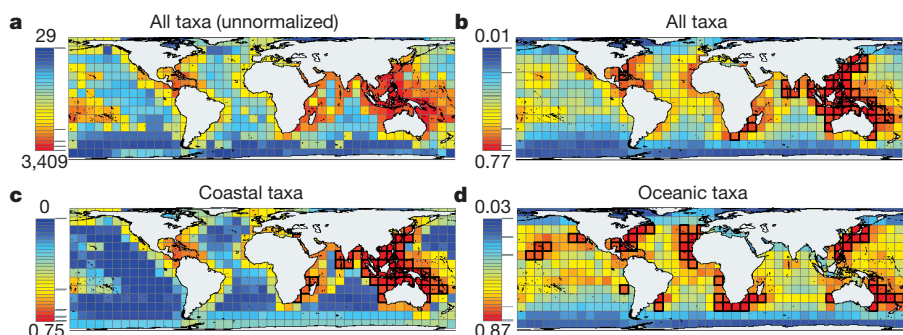
hotspots around the Philippines, Japan, China, Indonesia, Australia, India and Sri Lanka, South Africa, and the Caribbean and southeast USA. Coastal species groups tended to be disproportionately concentrated in Southeast Asia (Fig. 2c), whereas cross-taxon oceanic diversity showed consistent bands of high average richness at ~30° latitude North or South in all oceans (Fig. 2d). Within these latitudes, oceanic diversity peaked closer to the continents and along boundary currents such as the Gulf Stream and Kuroshio Current, probably owing to the availability of favourable habitat (discussed below).

We used multivariate spatial linear models (SLMs) to evaluate the support in our data for six prominent hypotheses that may explain

observed diversity patterns through relating richness to mean environmental conditions per grid cell.

(1) The kinetic energy or temperature hypothesis posits that at higher temperatures increased metabolic rates may promote higher rates of speciation<sup>11,12</sup> leading to greater diversity<sup>11–14</sup>, or that range limits are set by thermal tolerance, with more species tolerant of warm conditions<sup>13</sup>; in both cases we would expect a positive correlation with temperature, particularly for ectotherms.

(2) The potential energy or ‘productivity–richness’ hypothesis predicts a positive effect of primary productivity on richness at coarse grain sizes, such as facilitating larger population sizes that avert



**Figure 2 | Global species richness and hotspots across taxa.** (a) Global marine species richness for all taxa from Table 1 combined. Richness values for each taxon were then normalized by rescaling from zero to one, and averaged across taxa by cell for all taxa (b), primarily coastal taxa (c) and primarily oceanic taxa (d). Cells with a bold outline are hotspots (defined as the 10% of cells with highest mean richness). Horizontal tick marks on colour-bars indicate quartiles.

extinctions or support niche specialists<sup>5,13,15</sup>. Peaks of temperature and productivity are spatially separated in the ocean, allowing better discrimination of these first two hypotheses than on land.

(3) The stress hypothesis predicts a negative relationship of richness with environmental stress<sup>16</sup>, tested by quantifying the extent of oxygen depletion, a unique and increasingly important stressor in the marine environment<sup>17</sup>.

(4) The climate stability hypothesis assumes higher diversity in more environmentally stable regions<sup>16</sup>, tested by using a measure of temporal variance in sea surface temperature (SST).

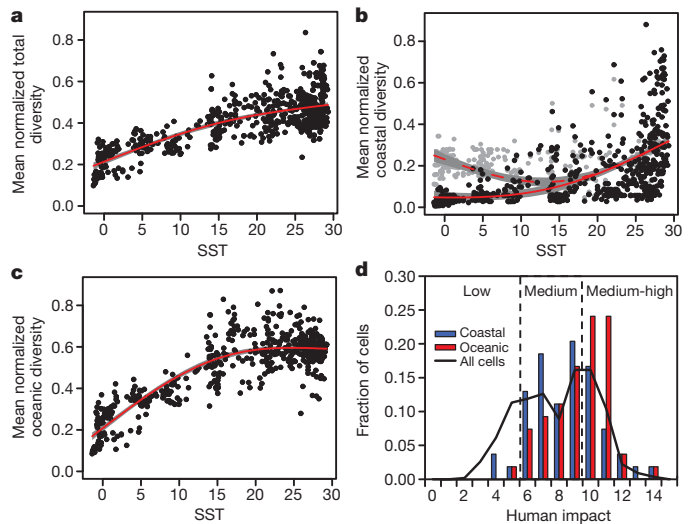
(5) The availability of important habitat features is expected to influence positively both abundance and richness<sup>18</sup>, specifically coastline length for coastal species and frontal systems (detected as SST slopes) for oceanic species<sup>7,19</sup>.

(6) We also sought to account for potentially different signatures of evolutionary history among ocean basins; hence an 'oceans' term was included in our models.

SST was the only predictor of species richness identified as statistically significant across all species groups in the SLMs (Table 1). This lends general support to the kinetic energy or temperature hypothesis, that is, that higher metabolic rates or relaxed thermal constraints promote diversity. This overriding importance of SST was supported further by minimal-adequate generalized-linear models (GLMs), as well as single-predictor SLMs and GLMs, indicating that results are robust to collinearity and independent of methods (see Methods and Supplementary Information for details). The relationship between SST and species richness was generally positive (Fig. 3a–c), except for pinnipeds. This is probably due to their endothermy, which represents a selective advantage in cold waters by decoupling metabolic rates from ambient temperatures<sup>20</sup>. Whereas mean richness of coastal groups other than pinnipeds tended to increase monotonically with temperature (Fig. 3b and Supplementary Fig. 3), oceanic groups on average showed an asymptotic relationship with SST (Fig. 3c and Supplementary Fig. 3), with some taxa declining in diversity at high temperatures (25–30 °C, Supplementary Fig. 3).

In addition to SST, all coastal groups showed significant relationships with coastline length (Table 1), supporting the hypothesis that habitat area also influences species richness for these taxa. Similarly, SST slope was significant for most oceanic taxa, indicating that oceanographic habitat features associated with steep SST gradients (fronts and eddies) mediate species richness patterns in the open ocean<sup>7,19</sup>. However, neither coastline length nor SST slope generally explained more variation in richness than SST (Table 1). Significant differences between ocean basins were detected for four coastal but no oceanic taxa, indicating a larger signature of historical geographic factors on coastal species diversity. Whereas coastal fish species are more likely to remain close to their place of origin<sup>21</sup>, oceanic species are highly mobile and live in a continuous habitat with high connectivity. We thus propose that species' life-history factors relating to mobility, dispersal and range size may affect the relative signature of evolutionary factors on diversity patterns in the sea. Other hypotheses involving measures of primary productivity, oxygen stress and temporal stability were not consistently supported (Table 1). It is interesting to note that both endothermic groups (cetaceans and pinnipeds) showed stronger positive relationships with primary productivity than SST, potentially indicative of smaller constraints on activity rates but larger demands on thermoregulatory activity. In sum, these results indicate that temperature or kinetic energy has a consistent and dominant role in structuring broad-scale marine diversity patterns, particularly for ectothermic species, with habitat area and historical factors important for coastal taxa, and support for other factors varying by taxon.

One important motivation for understanding geographic diversity patterns is concern over biodiversity threats being concentrated in areas of high richness and thus affecting large numbers of species<sup>22</sup>. For the oceans, multiple threats such as exploitation, habitat destruction, pollution and climate change have been overlaid into a cumulative



**Figure 3 | Diversity, SST and human impact overlap.** a–c, Relationship between mean normalized diversity and SST for (a) all taxa, (b) coastal taxa without pinnipeds (solid line and black points) and coastal taxa with pinnipeds (dashed line and grey points) and (c) oceanic taxa. Trends (red lines with grey 95% confidence limits) indicated by generalized additive model fit with basis dimension 3. Taxon-specific plots provided in Supplementary Fig. 3. d, Histogram of diversity hotspots (10% of cells with highest mean richness from Fig. 2c, d) by human impact. For each cell, average human impact score calculated from ref. 10. The distribution of human impacts across all marine cells (black line) is provided for reference.

impact map, showing large human impacts in coastal areas of East Asia, Europe, North America and the Caribbean<sup>10</sup>. We found weak but statistically significant correlations between mean anthropogenic impacts and total species richness ( $r = 0.19$ ,  $P < 0.01$ ) and mean normalized richness ( $r = 0.35$  for all species,  $r = 0.15$  for coastal species,  $r = 0.43$  for oceanic species,  $P < 0.01$  in all cases) per cell. Hotspots for both oceanic and coastal species (defined as 90th percentile of mean normalized richness) occurred in areas with medium or higher human impact more frequently than expected by chance alone (Fig. 3d; null permutation model  $P < 0.05$ ). Such overlapping hotspots of species richness and human impact should be further assessed across finer grain sizes, and may be an important focus for marine management and conservation efforts across taxa.

The global diversity patterns presented here need to be interpreted in context. Our database is limited to taxa for which sufficient records were accessible to determine global distribution. A large gap is the distribution of deep-sea diversity, where data remain scarce<sup>23</sup>. Likewise, we did not consider microbes or viruses, and have limited marine invertebrate data. The cephalopod pattern has a higher level of uncertainty, as available data only account for ~25% of known diversity (Table 1), and are biased towards commercial species. Furthermore, our results are based on a single, relatively coarse grain size necessary to maximize sampling effort and minimize errors in extrapolation and record accuracy. Finally, we analysed only a subset of mechanisms that may shape biodiversity<sup>11</sup>, and necessarily did so in a correlative fashion that cannot establish causation in the same way as experimental approaches. Although a macroecological approach can evaluate support for hypotheses given available data<sup>13</sup>, phylogenetic and paleoecological data will prove critical to separate ecological and evolutionary processes further<sup>24,25</sup>.

In conclusion, we found two distinct patterns of global marine biodiversity, one for species primarily associated with coastal habitats, the other for open ocean taxa. Although spatial patterns of individuals groups such as corals<sup>26</sup> and foraminifera<sup>6</sup> have been well studied before, our analyses provide a new level of generality by assembling and synthesizing global distributions across 13 key taxa. In accordance with the idea of high kinetic energy facilitating greater species richness over evolutionary and ecological timescales,

temperature emerged as the primary environmental correlate of diversity at the large geographic scale tested here. Furthermore, factors related to available habitat (coastline length and SST slope) were also important, whereas other mechanisms were inconsistently supported across taxa. Based on these findings, changes in the temperature of the global ocean<sup>27</sup> may have strong consequences for the distribution of marine biodiversity<sup>28</sup>. Other disturbances could further modify these patterns, as cumulative human impacts of medium to high magnitude were present in most areas of high mean richness described here. Limiting the extent of ocean warming, and mitigating multiple human impacts, may be of particular importance to securing marine biodiversity in the future.

## METHODS SUMMARY

**Data.** Our analysis builds on the decade-long effort by the Census of Marine Life to compile occurrence records for marine species in an Ocean Biogeographic Information System ([www.iobis.org](http://www.iobis.org)). This database has been particularly successful for fishes (>6.5 million records), which data we used to generate global maps of coastal fish species richness by extrapolating from species discovery curves<sup>29</sup> (to correct for highly variable sampling effort) and projecting onto an 880-km resolution equal-area grid (Supplementary Fig. 4). We derived richness of other taxa from multiple sources of both expert-verified range maps and empirical occurrence data (Supplementary Table 1). For extrapolating fish richness, we compared four established extrapolation methods on both simulated and real data. We validated extrapolated coastal fish richness first against independent checklist data, and second by comparing co-kriged estimates against further checklist data (see Methods). Environmental data (Supplementary Table 2) were compiled from published sources and projected on the same grid as species richness.

**Analysis.** We modelled relationships between species richness and environmental predictors using both GLMs and SLMs. Both dependent and independent variables were log-transformed to linearize and normalize data. We excluded cells with zero diversity and those containing <10% ocean area from the analyses. GLMs produced model residuals that were spatially non-independent for all taxa (Supplementary Table 8); thus SLMs, which account for spatial autocorrelation patterns, were used for inference. Analyses were conducted using maximum-likelihood spatial autoregressive (SAR) models. We used an error-SAR model and tested neighbourhood thresholds between 1,000 and 10,000 km in 200-km intervals. Final selection of neighbourhood size was based on minimising Akaike Information Criterion values for the spatial null model independently for each taxon. Backwards stepwise elimination of insignificant predictors was used to determine the minimal adequate model.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 March; accepted 8 July 2010.

Published online 28 July 2010.

1. Darwin, C. *The Origin of Species* (John Murray, 1859).
2. Briggs, J. C. *Marine Zoogeography* (McGraw-Hill, 1974).
3. Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405**, 243–253 (2000).
4. Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
5. Jetz, W., Kreft, H., Ceballos, G. & Mutke, J. Global associations between terrestrial producer and vertebrate consumer diversity. *Proc. R. Soc. Lond. B* **276**, 269–278 (2009).
6. Rutherford, S., D'Hondt, S. & Prell, W. Environmental controls on the geographic distribution of zooplankton diversity. *Nature* **400**, 749–753 (1999).
7. Worm, B. *et al.* Global patterns of predator diversity in the open oceans. *Science* **309**, 1365–1369 (2005).
8. Whitehead, H., McGill, B. & Worm, B. Diversity of deep-water cetaceans in relation to temperature: implications for ocean warming. *Ecol. Lett.* **11**, 1198–1207 (2008).
9. Hillebrand, H. Strength, slope and variability of marine latitudinal gradients. *Mar. Ecol. Prog. Ser.* **273**, 251–267 (2004).

10. Halpern, B. S. *et al.* A global map of human impact on marine ecosystems. *Science* **319**, 948–952 (2008).
11. Rohde, K. Latitudinal gradients in species-diversity: the search for the primary cause. *Oikos* **65**, 514–527 (1992).
12. Allen, A. P., Gillooly, J. F. & Brown, J. H. in *Scaling Biodiversity* (eds Storch, D., Marquet, P. A. and Brown, J. H.) (Cambridge Univ. Press, 2007).
13. Currie, D. J. *et al.* Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* **7**, 1121–1134 (2004).
14. Clarke, A. & Gaston, K. J. Climate, energy and diversity. *Proc. R. Soc. Lond. B* **273**, 2257–2266 (2006).
15. Evans, K. L., Warren, P. H. & Gaston, K. J. Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biol. Rev. Camb. Philos. Soc.* **80**, 1–25 (2005).
16. Fraser, R. H. & Currie, D. J. The species richness–energy hypothesis in a system where historical factors are thought to prevail: coral reefs. *Am. Nat.* **148**, 138–159 (1996).
17. Keeling, R. F., Körtzinger, A. & Gruber, N. Ocean deoxygenation in a warming world. *Annu. Rev. Mar. Sci.* **2**, 199–229 (2010).
18. Rosenzweig, M. L. *Species diversity in space and time*. (Cambridge Univ. Press, 1995).
19. Etnoyer, P., Canny, D., Mate, B. & Morgan, L. Persistent pelagic habitats in the Baja California to Bering Sea (B2B) ecoregion. *Oceanography* **17**, 90–101 (2004).
20. Cairns, D. K., Gaston, A. J. & Huettmann, F. Endothermy, ectothermy and the global structure of marine vertebrate communities. *Mar. Ecol. Prog. Ser.* **356**, 239–250 (2008).
21. Mora, C. & Robertson, D. R. Factors shaping the range-size frequency distribution of the endemic fish fauna of the Tropical Eastern Pacific. *J. Biogeogr.* **32**, 277–286 (2005).
22. Myers, N. *et al.* Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
23. Webb, T. J., Vanden Berghe, E. & O'Dor, R. K. Biodiversity's big wet secret: The global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. *PLoS ONE* doi:10.1371/journal.pone.0010223 (2010).
24. Jablonski, D., Roy, K. & Valentine, J. W. Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science* **314**, 102–106 (2006).
25. Ricklefs, R. E. History and diversity: explorations at the intersection of ecology and evolution. *Am. Nat.* **170**, S56–S70 (2007).
26. Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284 (2002).
27. Levitus, S. *et al.* Anthropogenic warming of the Earth's climate system. *Science* **292**, 267–270 (2001).
28. Worm, B. & Lotze, H. K. in *Climate and Global Change: Observed Impacts on Planet Earth* (ed. Letcher, T.) 263–279 (Elsevier, 2009).
29. Mora, C., Tittensor, D. P. & Myers, R. A. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proc. R. Soc. Lond. B* **275**, 149–155 (2008).
30. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *Proc. Natl Acad. Sci. USA* **104**, 5925–5930 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge contributions and sharing of data from L. Lucifora, V. Garcia, M. Kulbicki and P. Hull. We thank all the sources in Supplementary Tables 1 and 2 for making their data available. We are grateful to all OBIS data providers for making this study possible; see [www.iobis.org](http://www.iobis.org) for a full list. A. Rollo, G. Britten and D. Boyce provided technical help; W. Blanchard offered statistical advice. This paper builds on the efforts of all Census of Marine Life contributors, and long-term support from the Sloan Foundation. W.J. acknowledges support from NSF grants DBI-0960550 and BCS-0648733.

**Author Contributions** B.W., H.K.L., D.P.T., W.J. and C.M. conceived the study, D.P.T., C.M., E.V.B., D.R., B.W. and H.K.L. compiled the data, D.P.T., W.J. and C.M. conducted the analyses, and all authors contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.P.T. ([derek@mathstat.dal.ca](mailto:derek@mathstat.dal.ca)).

## METHODS

**Biological data.** Biological data (Supplementary Table 1) were compiled from empirical sampling data (fishes, foraminifera and tunas and billfishes) or expert-verified range maps (cephalopods, corals, euphausiids, marine mammals, mangroves, seagrasses and sharks). Marine mammals, fishes, sharks and cephalopods were subdivided into groups of primarily oceanic and primarily coastal species, because they contain numerous species of both habitat associations. Other taxa were mostly associated with either coastal (corals, mangroves, seagrasses) or oceanic (tunas and billfishes, planktonic foraminifera and euphausiids) habitats. Cells with zero values (missing values or zero species) were excluded from the analyses. Fish data were gridded to global equal-area grids of varying grain sizes (110, 220, 440 and 880 km) to test the trade-off between grain size and having sufficient data to extrapolate accumulation curves<sup>29</sup> (to correct for globally varying sampling effort) and conduct co-kriging (see 'Statistical analyses' later). The 880-km resolution grid was optimal for retaining sufficient cells to conduct co-kriging and extrapolate global diversity patterns. All other biological and environmental data were compiled on the same grid. Grid cells of <10% ocean area were removed from the analysis.

Fish data were retrieved from the Ocean Biogeographical Information System (OBIS) database. OBIS was created as the data integration component of the Census of Marine Life<sup>31</sup>, and, at present, makes available through its portal over 700 individual data sets, with more than 22 million records. It is the largest primary provider of marine biogeographical information, and one of the main providers of data to the Global Biodiversity Information Facility (GBIF). As part of the data management routine of OBIS, species names as provided by the data custodians are checked against a number of standard lists: Nomenclator Zoologicus (<http://uiom.mbl.edu/NomenclatorZoologicus/>), the Interim Register for Marine- and Non-marine Genera (IRMNG, <http://www.obis.org.au/irmng/>), the Catalogue of Life (<http://www.catalogueoflife.org/search.php>) and the World Register of Marine Species (WoRMS, <http://www.marinespecies.org>). The latter is used as the preferred source for species names, as this system is at present the most complete for marine biota; it is also used to find the currently accepted synonyms, and to add taxonomic classification information. Data were retrieved on 23 October 2009. We extracted ~6.5 million fish records from OBIS and binned them to the 880-km grid. To determine the richness of coastal species, we extracted habitat association data for each of the 13,049 species from Fishbase<sup>32</sup> and excluded species listed as pelagic, pelagic-oceanic, bathypelagic or bathydemersal, leaving 9,713 species. There were insufficient data on primarily oceanic fishes to extrapolate sufficient cells to construct a representative co-kringed map of richness; we therefore used the independently derived data on tunas and billfishes and oceanic sharks to represent the richness patterns of pelagic fishes. Sampling effort varied substantially between grid cells. We constructed accumulation curves and extrapolated total coastal fish richness (see 'Statistical analyses' below) to make cells comparable.

Cephalopod data were digitized and regridded from range maps<sup>33,34</sup>. Two synonymous species were removed and two subspecies combined with their parent species. We partitioned this group into primarily oceanic squids and primarily coastal non-squid cephalopods. Stony warm-water corals (Scleractinia) were regridded from range maps<sup>35</sup>. Euphausiids were digitized and regridded from range maps<sup>36</sup>. Foraminifera data were point data from the Brown University foraminifera database<sup>37</sup>. To minimize the effects of calcium carbonate dissolution on assemblage composition, sites from >3,500 m in the Pacific and Indian oceans and from >4,500 m elsewhere were excluded. Foraminiferan richness per cell was correlated with the number of samples in that cell ( $r = 0.35$ ). We standardized richness by taking mean richness per sample for cells with multiple samples. Marine mammals were regridded from range maps<sup>38</sup>, partitioned into cetaceans as primarily oceanic and pinnipeds as primarily coastal species, and their taxonomy checked. Mangroves were regridded from range maps<sup>39,40</sup>. Seagrasses were regridded from range maps<sup>41–44</sup>; species labelled as unknown were removed. Cells containing mangrove, seagrass or coral ranges but no coastline were removed. Shark geographic distributions were gathered from the scientific literature and collection specimens<sup>45</sup>. Shark data were partitioned into 'Oceanic' and 'Non-oceanic' following the IUCN Shark Specialist Group habitat classifications ([www.iucnredlist.org](http://www.iucnredlist.org)). Tuna (Thunnini), billfish (Istiophoridae), and swordfish (Xiphiidae) ranges were compiled from species occurrence records from Japanese longline fisheries operating worldwide between 1990 and 1999 (ref. 7), and further combined with global catch data taken from the United Nations Food and Agriculture Organization (FAO) Atlas of Tuna and Billfish Statistics (<http://www.fao.org/fishery/statistics/tuna-atlas/4/en>; longlining, pole and line, and purse seine) from the same period to ensure optimal range coverage. Both sources used 5° cells that we combined, converted to ranges, and mapped to the 880-km grid. Data were filtered by removing cells with, on average, less than one individual or 1 ton of catch per year, to eliminate isolated stragglers and possible misidentifications.

Sailfish (*Istiophorus albicans*, *Istiophorus platypterus*) and shortbill spearfish (*Tetrapturus angustirostris*), although present in the Japanese longline data, were excluded from analysis owing to the lack of FAO catch data.

**Environmental data.** Environmental predictors (Supplementary Table 2) were compiled to assess the support in our data for prominent hypotheses structuring large-scale marine diversity patterns. Data were averaged on the 880-km grid from their original resolution, unless otherwise indicated. Predictors that were strongly correlated ( $r > 0.5$ ) with other environmental data were removed (surface-variability index as in ref. 46, number of islands, depth) to prevent problems with model identifiability. Oxygen at 100 m (ref. 7) from the World Ocean Atlas 2005 (ref. 47) was highly correlated with temperature ( $r = -0.83$ ); we thus converted it to the proportion of grid cell that was <2 millilitres per litre O<sub>2</sub>, a critical physiological stress threshold for many species<sup>48</sup>. SST was calculated as long-term (1961–1990) annual means from NASA optimum interpolation data v.2 (ref. 49). SST slope was calculated as the maximum slope between each one-degree grid cell and its eight neighbours using the same data<sup>49</sup>, then averaged for each 880-km grid cell. Mean standard deviation of SST was calculated using the AVHRR climatology<sup>50</sup> by taking the grand mean annual standard deviation from 1985 to 2008 and averaging within each 880-km grid cell. Net primary productivity (gram C m<sup>-2</sup> year<sup>-1</sup>) was derived from the standard parameterization of the vertically generalized production model<sup>51</sup> (VGPM). We did not include the mid-domain effect<sup>32</sup> as a potential predictor of species richness because interconnections among oceans and the fact that boundaries are unlikely to be equally limiting for all taxa make the delineation of geographic domains very subjective at the scale of the global ocean.

**Statistical analyses.** Sampling effort for OBIS data varied markedly between cells. We corrected for this by extrapolating species accumulation curves of combined yearly species descriptions within each grid cell. Extrapolation of asymptotic values can be conducted using parametric or non-parametric approaches<sup>53</sup>; we tested both categories by using four separate approaches (multimodel-averaged parametric asymptote<sup>27</sup>; Chao 2 estimator<sup>54,55</sup>; Jackknife<sup>56</sup>; Bootstrap<sup>56</sup>) against both simulated and real data to optimize performance and assess the effect of varying sampling effort on predicted richness. Cells with less than ten years of sampling data were removed from the analysis. As is typical for this type of biodiversity data, sampling effort varied within and between years. We addressed this by summarizing the data by year and performing randomized ordering of samples or rarefaction when calculating richness estimators, which smoothes out temporal variability in the accumulation curves. We compared estimator performance across four simulated communities with Poisson-lognormally distributed abundances of increasing unevenness and varying temporal sampling effort. The Chao estimator had the highest precision of all estimators tested in the two simulated communities that most closely reflect biological reality (uneven to moderately uneven abundance structure). The multi-model averaged parametric approach frequently gave similarly good predictions, but for simplicity we focused on the Chao estimator for analyses of the OBIS database. We then validated this approach by comparing estimates of extrapolated richness against independently gathered coastal fish checklists (expanded from refs 57–61) of spatial extent smaller than our grid cells. We found that when rarefied accumulation curves were near-linear, estimates of richness were unreliable, potentially leading to substantial under- or over-predictions owing to insufficient sampling of the community. This was not reflected in the ratio of the mean to the variance of the Chao estimator. We used a heuristic approach to identify and exclude cells with such insufficient sampling. We measured the curvilinearity of extrapolated fish richness as  $dy/dx$  in linear space for each sample, and excluded cells with a maximum slope ratio between samples of <2. Given the larger spatial extent of the grid cells and the fact that we were extrapolating total predicted coastal fish species (rather than just discovered species), predicted species richness should be correlated with but equal to or higher than the checklist richness. Although some cells did underpredict, underpredictions were relatively small (Supplementary Fig. 5), and predictions were highly correlated with the checklist data ( $r = 0.78$ ;  $n = 29$ ). This comparison is obviously limited by the different shapes and spatial extent of checklist census regions and may be affected by a geographically non-uniform spatial scaling of species richness. Finally, we cross-validated our approach by comparing co-kringed estimates (see below) for cells in which we did not have OBIS sampling data to further checklists; the correlation was  $r = 0.76$  ( $n = 44$ ).

When modelling the relationship between environmental predictors and response variables, the presence of spatial autocorrelation violates assumptions of traditional statistical approaches, resulting in deflated estimates of variance and corresponding impacts on inference, among other issues. Hence we modelled variables and conducted inference using both GLMs and SLMs. We log-transformed both dependent and independent variables to linearize data, and to homogenize variances. GLMs resulted in model residuals that were spatially non-independent for all taxa (measured using Moran's  $I$  tests; Supplementary Table 8); we therefore used spatial linear models for final inference (no Moran's  $I$  test

significant). We conducted spatial analyses using maximum-likelihood spatial autoregressive (SAR) models. We used an error-SAR model<sup>62</sup> and tested neighbourhood thresholds of between 1,000 and 10,000 km in 200-km intervals. Final selection of neighbourhood size was based on minimizing AIC values for the spatial null model (that is, only retaining a spatial autocorrelation term) and was determined independently for each taxon, thus reflecting potential differences in spatial autocorrelation due to life history or other evolutionary effects.

We used backwards stepwise elimination of insignificant predictors to determine the minimal adequate model. The importance of individual predictors was ascertained through *t*-tests (GLMs), *z*-tests (SLMs), and the marginal importance of individual predictors ascertained through single-factor models. We used AIC and pseudo-*R*<sup>2</sup> to assess model fit, to compare GLM and SLM results, and to assess the importance of spatial autocorrelation on model results. Although our models can fit functional forms in arithmetic space including asymptoting, linear and exponential, several previous studies have suggested a potentially unimodal relationship between SST and diversity for some oceanic taxa. There was little evidence of ill-fitting in our minimal adequate SLMs, as judged by examining residual plots, indicating that unimodal relationships with SST, if present, are likely to be captured adequately through incorporating multiple predictors. We tested models further by including a quadratic SST term; although quadratic terms were retained in minimal models for most taxa, our main results (SST important for all taxa; coastline and historical effects for coastal taxa) were robust to this difference (SST and historical effects unchanged in terms of taxa; one less group (mangrove) associated with coastline). We chose to present linear models in the main text for ease of interpretation, and owing to uncertainties in how quadratic and spatial autocorrelation terms interact in multiple predictor models.

Empirically gathered point data (fishes, foraminifera) were then co-kriged using minimal-adequate GLM predictions as the co-variable<sup>30</sup> to extrapolate to unsampled or undersampled regions. Co-kriged data were used to validate the extrapolation approach (see above), for visual purposes (Fig. 1), and for constructing the hotspot maps (Fig. 3). Co-kriged coastal fish predictions were constrained to cells with non-zero coastline length.

We used null models to test for statistical significance of the observed relationship between hotspots and human impact. The null expectation was that the hotspots could occur anywhere within the potential domain; we distributed them randomly, calculated the cumulative distribution relative to human impacts, and repeated this permutation 999 times. We then used the ordered set of permutations to determine 2.5% and 97.5% confidence limits.

All analyses were conducted in the open-source language R<sup>63</sup>. Non-parametric estimates were calculated using the R package *vegan*<sup>64</sup>. Multimodel averaging approaches were conducted using a bespoke R package. Spatial model analyses and co-kriging were conducted using the *spdep* and *gstat* R packages, respectively<sup>65,66</sup>.

31. Fornwall, M. Planning for OBIS: examining relationships with existing national and international biodiversity information systems. *Oceanography* **13**, 31–38 (2000).
32. Froese, R. & Pauly, D. *FishBase* (<http://www.fishbase.org>) (2010).
33. FAO. *FAO Species Catalogue* (eds Roper, C. F. E., Sweeney, M. J. & Nauen, C. E.) Vol. 3 (FAO, 1984).
34. FAO. *Cephalopods of the World: An Annotated and Illustrated Catalogue of Cephalopod Species Known to Date* (eds Jereb, P. & Roper, C. F. E.) Vol. 1 (FAO, 2005).
35. Veron, J. E. N. *Corals of the World* Vols 1–3 (Australian Institute of Marine Science, 2000).
36. Brinton, E. *et al.* *Euphausiids of the World Ocean* CD-ROM (ETI Bioinformatics, 2000).
37. Prell, W., Martin, A., Cullen, J. & Trend, M. *The Brown University Foraminiferal Data Base, IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series # 1999-027*. (NOAA/NGDC Paleoclimatology Program, 1999).
38. Schipper, J. *et al.* The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* **322**, 225–230 (2008).
39. UNEP-WCMC. *Global Distribution of Mangroves* (UNEP World Conservation Monitoring Centre/ISME) In *World Mangrove Atlas* (eds Spalding, M. D., Blasco, F. & Field, C. D.) (The International Society for Mangrove Ecosystems, 1997).
40. Spalding, M. D., Blasco, F. & Field, C. D. (eds) *World Mangrove Atlas* (The International Society for Mangrove Ecosystems, 1997).
41. UNEP-WCMC & Short, F. T. *Global Seagrass Diversity* (v1.0) in *World Atlas of Seagrasses* (Green, E. P. & Short, F. T.) (Univ. of California Press, 2003).
42. UNEP-WCMC & Short, F. T. *Global Seagrass Species Ranges* (v1.0) in *World Atlas of Seagrasses* (Green, E. P. & Short, F. T.) (Univ. of California Press, 2003).
43. UNEP-WCMC & Short, F. T. *Global Distribution of Seagrasses* (v2.0) in *World Atlas of Seagrasses* (Green, E. P. & Short, F. T.) (Univ. of California Press, 2003/2005).
44. Green, E. P. & Short, F. T. *World Atlas of Seagrasses* (Univ. of California Press, 2003).
45. Lucifora, L., Garcia, V. & Worm, B. Global diversity hotspots and conservation priorities for sharks. *PLoS ONE* (in the press).
46. Lutz, M. J., Calderia, K., Dunbar, R. B. & Behrenfeld, M. J. Seasonal rhythms of net primary production and particulate organic carbon flux describe biological pump efficiency in the global ocean. *J. Geophys. Res.* **112**, C10011 (2007).
47. Garcia, H. E., Locarnini, R. A., Boyer, T. P. & Antonov, J. I. *World Ocean Atlas 2005* (ed. Levitus, S.) Vol. 3 (US Government Printing Office, 2006).
48. Vaquer-Sunyer, R. & Duarte, C. M. Thresholds of hypoxia for marine biodiversity. *Proc. Natl Acad. Sci. USA* **105**, 15452–15457 (2008).
49. Reynolds, R. W. *et al.* An improved in situ and satellite SST analysis for climate. *J. Clim.* **15**, 1609–1625 (2002).
50. Casey, K. S., Brandon, T. B., Cornillon, P. & Evans, R. *Oceanography from Space, Revisited* (eds Barale, V., Gower, J. F. R. & Alberotanza, L.) Ch. 16 (Springer, 2010).
51. Behrenfeld, M. J. & Falkowski, P. G. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol. Oceanogr.* **42**, 1–20 (1997).
52. Colwell, R. K. & Lees, D. C. The mid-domain effect: geometric constraints on the geography of species richness. *Trends Ecol. Evol.* **15**, 70–76 (2000).
53. Colwell, R. K. & Coddington, J. A. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. B.* **345**, 101–118 (1994).
54. Chao, A. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* **11**, 256–270 (1984).
55. Chao, A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791 (1987).
56. Smith, E. P. & van Belle, G. Nonparametric estimation of species richness. *Biometrics* **40**, 119–129 (1984).
57. Robertson, D. R. & Allen, G. R. *Shorefishes of the Tropical Eastern Pacific: an Information System* CD-ROM (Smithsonian Tropical Research Institute, 2002).
58. Kulbicki, M., Labrosse, P. & Ferraris, J. *Challenging Coasts: Transdisciplinary Excursions into Integrated Coastal Zone Development* (ed. Visser, L. E.) (Amsterdam Univ. Press, 2004).
59. Floeter, S. R. *et al.* Atlantic reef fish biogeography and evolution. *J. Biogeogr.* **35**, 22–47 (2007).
60. Kulbicki, M. Biogeography of reef fishes of the French territories in the South Pacific. *Cybius* **31**, 275–288 (2007).
61. Kulbicki, M. *Du Macrocosme au Microcosme: Les Poissons de Récif du Pacifique Comme Modèle* HDR report (Univ. de Perpignan, 2007).
62. Dormann, C. F. *et al.* Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609–628 (2007).
63. R Core Development Team. *R: A Language and Environment for Statistical Computing* (<http://www.R-project.org>) (2009).
64. Oksanen, J. *et al.* *vegan: Community Ecology Package*. R package version 1.15-4. (<http://CRAN.R-project.org/package=vegan>) (2009).
65. Bivand, R. *spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. R package version 0.4-56. (<http://CRAN.R-project.org/package=spdep>) (2009).
66. Pebesma, E. J. Multivariable geostatistics in S: the *gstat* package. *Comput. Geosci.* **30**, 683–691 (2004).