LETTER

https://doi.org/10.1038/s41586-018-0273-1

An inverse latitudinal gradient in speciation rate for marine fishes

Daniel L. Rabosky[1,10]*, Jonathan Chang[2,10], Pascal O. Title[1,10], Peter F. Cowman[3,4], Lauren Sallan[5], Matt Friedman[6], Kristin Kaschner[7], Cristina Garilao[8], Thomas J. Near[3], Marta Coll[9] & Michael E. Alfaro[2,10]

Far more species of organisms are found in the tropics than in temperate and polar regions, but the evolutionary and ecological causes of this pattern remain controversial[1,2]. Tropical marine fish communities are much more diverse than cold-water fish communities found at higher latitudes[3,4], and several explanations for this latitudinal diversity gradient propose that warm reef environments serve as evolutionary 'hotspots' for species formation[5–8]. Here we test the relationship between latitude, species richness and speciation rate across marine fishes. We assembled a time-calibrated phylogeny of all ray-finned fishes (31,526 tips, of which 11,638 had genetic data) and used this framework to describe the spatial dynamics of speciation in the marine realm. We show that the fastest rates of speciation occur in species-poor regions outside the tropics, and that high-latitude fish lineages form new species at much faster rates than their tropical counterparts. High rates of speciation occur in geographical regions that are characterized by low surface temperatures and high endemism. Our results reject a broad class of mechanisms under which the tropics serve as an evolutionary cradle for marine fish diversity and raise new questions about why the coldest oceans on Earth are present-day hotspots of species formation.

The steep decline in species richness from the equator to the poles is one of the most general large-scale patterns in biology[9,10] and has existed in its general form for more than 30 million years[11]. Many proposed mechanisms for this latitudinal diversity gradient (LDG) explain high tropical diversity as the outcome of faster rates of species origination: the tropics are an evolutionary cradle for new species, and the gradient reflects—at least in part—lower rates of species formation in regions outside the tropics[1,12]. Studies on fossil mollusks[12], plankton[13] and corals[5] support the hypothesis that rates of marine species formation are faster in the tropics than at higher latitudes.

We tested whether latitudinal variation in the rate of speciation can explain the LDG in marine fish diversity by reconstructing speciation rates across fishes and analysing them in a geographical context. We focused explicitly on recent speciation rates[2,14,15], because extinction reduces our ability to infer rates deep in the past[16]. We also ignored phylogenetic estimates of extinction rates, given the unreliable nature of these parameters in phylogenetic diversification models[17]. If speciation rates are controlled by energy—perhaps owing to accelerated chemical reactions, life histories or mutation rates[18,19]—then we should observe a footprint of rapid speciation in the distribution of recent speciation times for tropical taxa.

We assembled a distribution of all-taxon assembled (ATA) time-calibrated phylogenetic trees of ray-finned fishes (31,526 species). The ATA phylogenies include 11,638 species with genetic data (5,231 marine species); the remaining 19,888 species that did not have genetic data were placed using stochastic polytomy resolution (Methods) to generate taxonomically consistent resolutions of all taxa without genetic data under a conservative constant-rate birth–death process. The ATA trees were time-calibrated using a database of 139 fossil taxa (Extended Data Fig. 1 and Supplementary Information). We estimated or compiled geographic ranges for the majority of known marine species, including all species with genetic data. We estimated speciation rates across the phylogenies using BAMM[20], a Bayesian framework for reconstructing complex evolutionary dynamics from phylogenetic trees, and DR, a summary statistic that infers recent speciation rates for all tips in the phylogeny without requiring a formal parametric inference model[21]. We denote these two analyses of speciation rates as λ_{BAMM} and λ_{DR}, respectively. The λ_{BAMM} and λ_{DR} rates include substantial historical information and are best interpreted as the rate of lineage splitting averaged across the past 10–20 million years (Myr)[2]; units for speciation presented here are per-lineage rates in units of lineages per Myr. We also computed a simple interval-based measure of speciation rate for a series of path intervals from 0.25 Myr to 50 Myr before present[22], providing a window of reliability for λ_{BAMM} and λ_{DR}. Estimates of λ_{DR} were computed across the distribution of ATA phylogenies, thus generating rate estimates conditional on the uncertainty in placements of taxa without genetic data. λ_{BAMM} was estimated from the primary dated tree including all taxa with genetic data ($n = 11,638$), and incomplete sampling was incorporated by using family-specific sampling fractions.

Consistent with previous studies[3,4], we find a strong LDG in marine fish diversity, with an extreme richness peak in the Coral Triangle of the tropical Indo-Pacific Ocean (Fig. 1a). However, analysis of per-cell mean speciation rates reveals a notable inverse relationship between the rate of species formation and latitude (Fig. 1b–e). Mean speciation rate for cell assemblages from tropical regions (<23.5°; $n = 6,698$ cells) was $\lambda_{BAMM} = 0.08$ ($\lambda_{DR} = 0.11$) and the corresponding rate for high-latitude (>45°; $n = 4,347$ cells) assemblages was $\lambda_{BAMM} = 0.14$ ($\lambda_{DR} = 0.16$). These rate differences are substantially greater when comparing more species-rich assemblages from continental shelf and slope regions: shallow (mean depth <2,000 m) tropical cells have $\lambda_{BAMM} = 0.08$ ($\lambda_{DR} = 0.11$; $n = 833$), whereas corresponding high-latitude cells have $\lambda_{BAMM} = 0.18$ ($\lambda_{DR} = 0.22$; $n = 1,182$). We computed means for 232 marine biogeographic ecoregions—encompassing the Earth's shallow and coastal regions—and used spatial simultaneous autoregressive (SAR) models with breakpoints to assess the relationship between latitudinal position and speciation rate. Regardless of how regional mean rates are computed, all SAR models have highly significant effects of latitude on speciation rate ($P < 0.001$; $n = 232$ regions; Extended Data Fig. 2). In general, for latitudes greater than 25° north or south, each ten-degree increase in latitude increases the assemblage-wide speciation rate by approximately 0.025 lineages per Myr. However, speciation rate is effectively decoupled from latitude

[1]Museum of Zoology, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA. [3]Peabody Museum of Natural History and Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [4]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland, Australia. [5]Department of Earth and Environmental Science, University of Pennsylvania, Philadelphia, PA, USA. [6]Museum of Paleontology and Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA. [7]Department of Biometry and Environmental System Analysis, Albert-Ludwigs-University of Freiburg, Freiburg, Germany. [8]GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany. [9]Instituto de Ciencias de Mar, Spanish National Research Council (ICM-CSIC), Barcelona, Spain. [10]These authors contributed equally: Daniel L. Rabosky, Jonathan Chang, Pascal O. Title, Michael E. Alfaro. *e-mail: drabosky@umich.edu

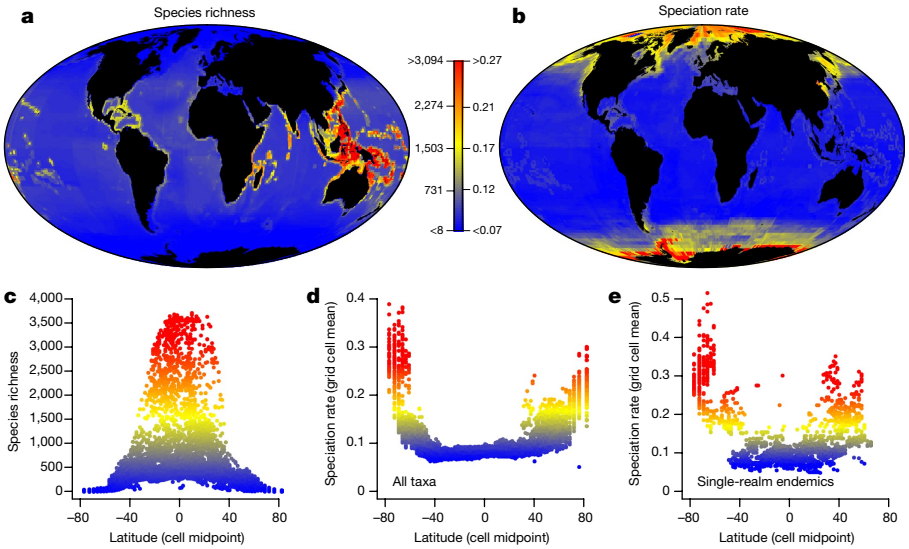


Fig. 1 | Latitudinal gradient in species diversity and speciation rate in marine fishes. a, **b**, Mean species richness (**a**) and speciation rate (**b**) for marine fish assemblages at the global scale. **c**, **d**, Marginal distributions of richness (**c**) and speciation rate (**d**) with respect to latitude ($n = 16{,}150$), with cell colours corresponding to scale bars in **a**, **b**. **e**, Mean speciation rates for endemic taxa only ($n = 2{,}698$). Results shown for λ_{BAMM} but similar results are obtained for λ_{DR} (Extended Data Figs. 2, 3). Grid cell size is 150 × 150 km for all panels.

for tropical and subtropical regions (Extended Data Fig. 2h). General results reported here are robust across all of the measures of speciation rate and associated weighting schemes that we considered (Extended Data Figs. 2, 3).

Speciation rate is strongly and negatively associated with both species richness (Fig. 2a) and annual sea surface temperature (Fig. 2b), although sea surface temperature is highly correlated with latitude ($r = -0.95$ across 16,150 grid cells). Regional assemblages of fishes with the fastest rates of speciation occur at the highest latitudes and are characterized by cold surface temperatures (Fig. 2 and Extended Data Fig. 4). The south polar seas, dominated by the in situ radiation of highly specialized and geographically restricted icefishes and their relatives[23], are characterized by the fastest overall rates of species formation of any marine region on Earth. Continental shelf and slope assemblages from the Southern Ocean surrounding Antarctica have mean speciation rates of $\lambda_{BAMM} = 0.27$ and $\lambda_{DR} = 0.26$ ($n = 179$ cells); these rates substantially exceed those observed for the Coral Triangle ($\lambda_{BAMM} = 0.08$ and $\lambda_{DR} = 0.11$; $n = 220$ cells), despite a mean 62-fold difference in per-cell species richness for these regions. Assemblages from the Arctic also have high speciation rates ($\lambda_{BAMM} = 0.17$ and $\lambda_{DR} = 0.24$; $n = 511$ cells), despite little overlap between the clades that comprise the northern and southern polar faunas[24]. There is a strong positive relationship between several analyses of regional endemism and assemblage-wide speciation rate (Fig. 2c and Extended Data Fig. 4e; $n = 60$ regions). The correlation between λ and endemism is high overall (λ_{BAMM}, $r = 0.81$; λ_{DR}, $r = 0.79$). The Mediterranean Sea is a clear outlier with respect to this overall pattern, combining high endemism with relatively low speciation (Fig. 2c). This suggests that the factors contributing to endemism per se are not necessarily those that promote fast speciation.

As an alternative to the analysis of mean speciation rates by grid cell and biogeographical region (Figs. 1, 2), we analysed λ_{BAMM} and λ_{DR} for individual fish species with respect to their latitudinal midpoint. High-latitude fish clades are characterized by rapid speciation relative to low-latitude and reef-associated clades, and there is a strong relationship between the centroid midpoint of the geographic range for each species and its estimated rate of species formation (Fig. 3 (inset) and Extended Data Fig. 5). We formally tested the relationship between latitudinal midpoint and speciation rate using several methods that are robust to model misspecification and phylogenetic pseudoreplication[25,26]. The correlation between absolute latitudinal midpoint and λ_{DR} is 0.27 ($P < 0.001$); similar results are obtained for λ_{BAMM} and latitude ($r = 0.3$; $P = 0.006$). Across a range of latitudinal thresholds, we find a highly significant difference in speciation rate for high- and low-latitude fishes ($P < 0.001$ across all thresholds), and cold–temperate and polar lineages speciating approximately twice as fast as the average low-latitude lineage (Extended Data Table 1).

Species with latitudinal midpoints in the tropics (23.5° S to 23.5° N; $n = 3{,}461$) have mean speciation rates of $\lambda_{BAMM} = 0.09$ and $\lambda_{DR} = 0.12$. By contrast, species with latitudinal midpoints greater than 45° N or 45° S ($n = 574$) have $\lambda_{BAMM} = 0.20$ and $\lambda_{DR} = 0.25$. These rates are even more extreme for subpolar and polar taxa: across fishes in our dataset with latitudinal midpoints greater than 60° ($n = 122$), mean speciation rates were $\lambda_{BAMM} = 0.29$ and $\lambda_{DR} = 0.35$. Interval-based estimates of speciation rate[22] indicate that the overall tropical–temperate–polar gradient that we report here has been present for millions of years, extending back in time at least the Miocene/Pliocene boundary (Extended Data Fig. 6).

Reef-associated clades, which comprise a substantial fraction of the tropical diversity peak, are not characterized by exceptional rates of species formation. Three of the largest such clades—the wrasses, damselfishes and gobies—collectively account for approximately 3,000 species, yet have low to moderate rates of speciation estimated using BAMM (wrasses: $\lambda_{BAMM} = 0.10$; gobies: $\lambda_{BAMM} = 0.07$; damselfishes: $\lambda_{BAMM} = 0.12$) and DR (wrasses: $\lambda_{DR} = 0.12$; gobies: $\lambda_{DR} = 0.10$; damselfishes: $\lambda_{DR} = 0.14$). By contrast, temperate and polar fish faunas are dominated by members of multiple clades that have exceptionally high rates of species formation (Fig. 3), including snailfishes, eelpouts, *Sebastes* rockfishes and Antarctic notothens (icefishes and allied species). These coldwater taxa are characterized by speciation rates that exceed 0.26 (λ_{BAMM}) and 0.34 (λ_{DR}). With the possible exception of gobies, we find little evidence for early bursts of speciation during the radiations of major tropical and reef-associated clades across the past 20–60 Myr (Extended Data Fig. 6). We note that 79.7% of marine speciation events in our ATA phylogenies are inferred to have occurred after the Oligocene/Miocene boundary, suggesting that the timescales over which we have estimated speciation rates are relevant to the origin and maintenance of modern LDG in marine fishes.

An alternative explanation for the global gradient in speciation rates that we report involves environmental or biogeographical filtering on traits that are also associated with rapid speciation. For example, perhaps speciation rates are most rapid for fishes that inhabit cold and dark bathyal or abyssal regions; physiological adaptations for life in those environments might predispose these lineages towards disproportionate representation in high-latitude communities. This hypothesis predicts that deep-sea lineages should speciate more rapidly than shallow lineages, regardless of latitude. However, mean rates for high-latitude (>45°) deep-sea fishes are much faster than for low-latitude (<45°) deep-sea species (high latitude: $\lambda_{BAMM} = 0.29$, $\lambda_{DR} = 0.37$, $n = 75$; low latitude: $\lambda_{BAMM} = 0.15$, $\lambda_{DR} = 0.15$, $n = 218$). Across all deep-sea fishes represented in our dataset ($n = 293$), there is a strong positive correlation between absolute latitudinal midpoint and speciation rate ($r = 0.50$; $P < 0.001$). There is no effect of depth classification on speciation rate

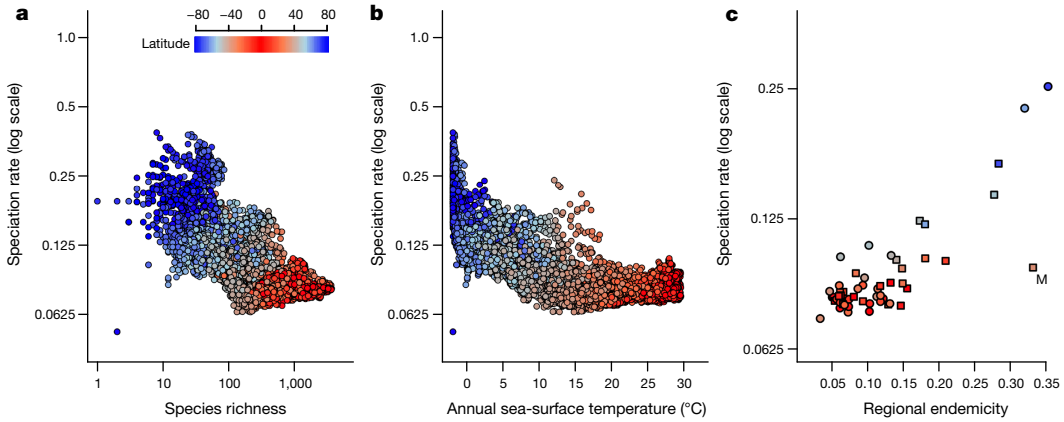


Fig. 2 | Species richness, temperature and speciation rate in marine fishes for individual grid cells. a, Negative relationship between species richness and mean speciation rate (λ_{BAMM}) for individual grid cells ($n = 16{,}150$). **b**, Negative relationship between mean annual sea-surface temperature and mean speciation rate for cells. **c**, Positive relationship between regional endemism and mean speciation rate for all species occurring in a particular biogeographical province ($n = 60$ biogeographical provinces). Squares and circles denote provinces with latitudinal midpoints north and south of the equator, respectively; cell colours denote latitude. Point labelled 'M' in the lower right of **c** is the Mediterranean Sea, which is characterized by high endemism and low speciation rate. Nearly identical results are obtained for λ_{DR} and for BAMM analyses that assume time constancy within rate regimes (Extended Data Fig. 4).

for tropical fishes ($P > 0.25$ across all classification schemes; Extended Data Fig. 7a). A secondary prediction of the filtering hypothesis is that high-rate, high-latitude clades should be nested within high-rate tropical or deepwater clades. We tested this hypothesis for perciform fishes, which account for 66.3% of high-latitude fishes (Extended Data Fig. 7b). Perciformes include four of the most-rapidly speciating

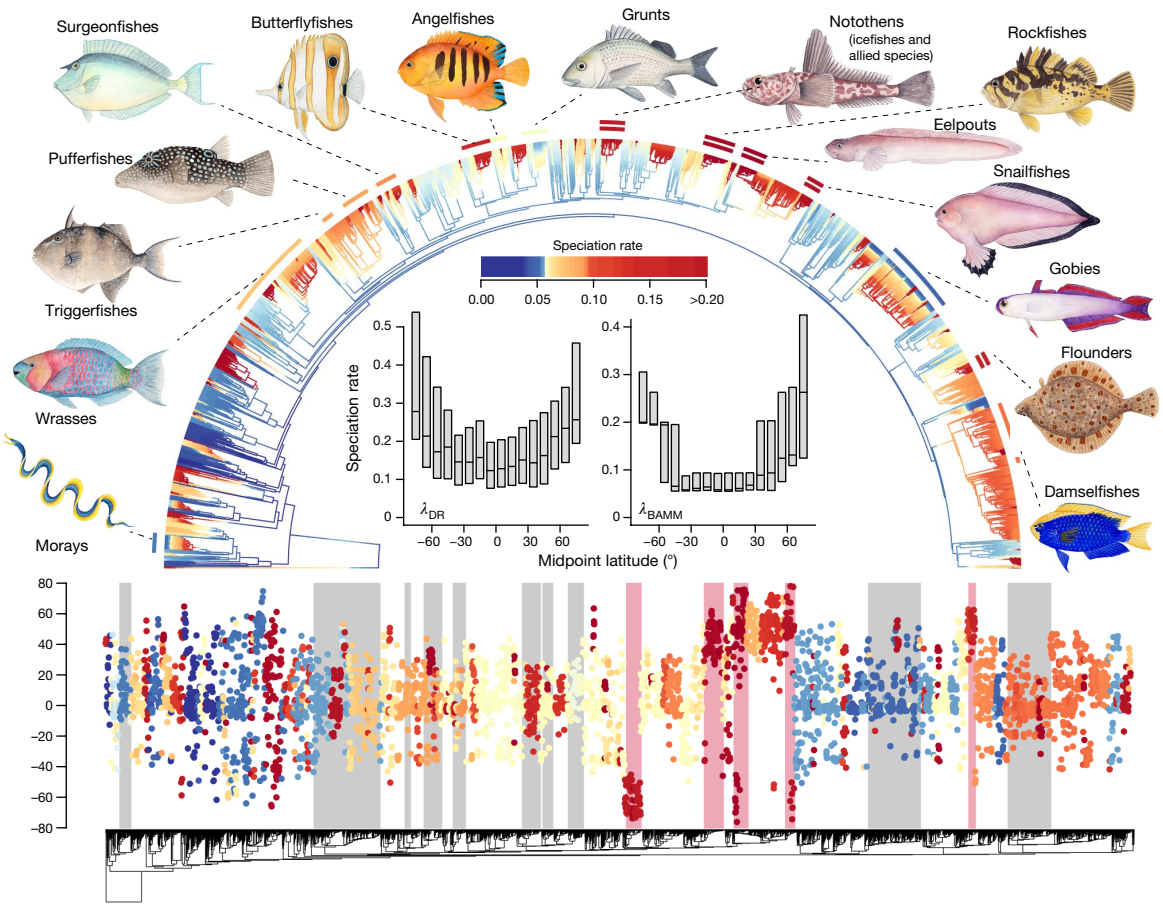


Fig. 3 | Latitudinal gradient in per-taxon speciation rate for marine fishes. Top, BAMM-estimated speciation rates across phylogenetic tree of 5,223 marine fishes for which genetic and geographic range data were available. Iconic coral reef clades are indicated with single arc segments; double segments denote high-latitude lineages that drive the overall fast speciation rate for temperate and polar fishes. Inset box plots show the median and interquartile range in distribution of rates (λ_{DR} and λ_{BAMM}) for individual taxa with respect to the centroid midpoint of their latitudinal distribution, with species values binned in 10° increments. Bottom, phylogenetic niche conservatism in marine fish lineages as reflected by the geographical distribution of latitudinal midpoints; each point is the centroid midpoint of an individual species, and colours reflect corresponding λ_{BAMM} estimates. Clades denoted with pink polygons are dominant high-latitude fish clades; grey polygons are predominantly reef-associated clades. The fish images were created by J. Johnson.

major clades of marine fishes (Notothenioids, Sebastidae, Zoarcidae and Liparidae), but these high-latitude clades are either nested within other high-latitude clades or within largely tropical clades that have low speciation rates (Extended Data Fig. 7c). The overall latitudinal gradient in speciation rate is thus unlikely to be explained by filtering on deepwater clades with rapid speciation rates into high-latitude biogeographical provinces.

We performed a complementary set of analyses based only on primary occurrence records from museum databases and other sources (see Methods). These estimates of species ranges yield highly congruent results (Extended Data Fig. 8). Our results are not conditional on a specific parametric model for inference; the terminal branch lengths themselves are strongly associated with latitude (Extended Data Fig. 9), indicating that few assumptions are required to obtain the results presented here. Furthermore, these results cannot be explained by variation in the completeness of taxonomic sampling with respect to latitude or by alternative reconstructions of geographic range (Extended Data Figs. 8, 9).

Our results reject the hypothesis that rapid speciation explains the spectacular diversity of tropical marine shallow-water fishes and reveal that, paradoxically, speciation rates are fastest in the geographical regions with the lowest species richness. Several evolutionary explanations for the LDG propose that fundamental relationships between energy and speciation rate control the accumulation of biodiversity over time[18,19], and it has been said that the tropics are more diverse because 'the Red Queen runs faster when she is hot'[27]. For the marine fish species that were studied here—and for many terrestrial vertebrates[2,21]—there is no evidence to support these biophysical linkages between energy, metabolism and speciation. Faster speciation contributes to total species richness in some island and freshwater lacustrine systems[2,28], but for larger biogeographical provinces—including the marine realms considered in the present study—it is increasingly unlikely that speciation rate variation is the primary cause of diversity gradients[2,29]. Whether the rapid speciation that we have documented in Earth's cold oceanic provinces reflects a recent and ongoing expansion of marine diversity is a key frontier for future research on the LDG in marine organisms.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0273-1.

Received: 2 July 2017; Accepted: 29 May 2018;
Published online 4 July 2018.

1. Mittelbach, G. G. et al. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* **10**, 315–331 (2007).
2. Schluter, D. & Pennell, M. W. Speciation gradients and the distribution of biodiversity. *Nature* **546**, 48–55 (2017).
3. Tittensor, D. P. et al. Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).
4. Stuart-Smith, R. D. et al. Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature* **501**, 539–542 (2013).
5. Kiessling, W., Simpson, C. & Foote, M. Reefs as cradles of evolution and sources of biodiversity in the Phanerozoic. *Science* **327**, 196–198 (2010).
6. Alfaro, M. E., Santini, F. & Brock, C. D. Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). *Evolution* **61**, 2104–2126 (2007).
7. Cowman, P. F. & Bellwood, D. R. Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *J. Evol. Biol.* **24**, 2543–2562 (2011).
8. Siqueira, A. C., Oliveira-Santos, L. G. R., Cowman, P. F. & Floeter, S. R. Evolutionary processes underlying latitudinal differences in reef fish biodiversity. *Glob. Ecol. Biogeogr.* **25**, 1466–1476 (2016).
9. Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163**, 192–211 (2004).
10. MacArthur, R. H. *Geographical Ecology* (Princeton Univ. Press, Princeton, 1972).
11. Mannion, P. D., Upchurch, P., Benson, R. B. J. & Goswami, A. The latitudinal biodiversity gradient through deep time. *Trends Ecol. Evol.* **29**, 42–50 (2014).
12. Jablonski, D., Roy, K. & Valentine, J. W. Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science* **314**, 102–106 (2006).
13. Allen, A. P. & Gillooly, J. F. Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Ecol. Lett.* **9**, 947–954 (2006).
14. Weir, J. T. & Schluter, D. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315**, 1574–1576 (2007).
15. Rabosky, D. L., Title, P. O. & Huang, H. Minimal effects of latitude on present-day speciation rates in New World birds. *Proc. R. Soc. Lond. B* **282**, 20142889 (2015).
16. Liow, L. H., Quental, T. B. & Marshall, C. R. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Syst. Biol.* **59**, 646–659 (2010).
17. Davis, M. P., Midford, P. E. & Maddison, W. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol.* **13**, 38 (2013).
18. Rohde, K. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* **65**, 514–527 (1992).
19. Allen, A. P., Brown, J. H. & Gillooly, J. F. Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**, 1545–1548 (2002).
20. Rabosky, D. L., Mitchell, J. S. & Chang, J. Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Syst. Biol.* **66**, 477–498 (2017).
21. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
22. Freckleton, R. P., Phillimore, A. B. & Pagel, M. Relating traits to diversification: a simple test. *Am. Nat.* **172**, 102–115 (2008).
23. Near, T. J. et al. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proc. Natl Acad. Sci. USA* **109**, 3434–3439 (2012).
24. Eastman, J. T. Comparison of the Antarctic and Arctic fish faunas. *Cybium* **21**, 335–352 (1997).
25. Harvey, M. G. & Rabosky, D. L. Continuous traits and speciation rates: alternatives to state-dependent diversification models. *Methods Ecol. Evol.* **9**, 984–993 (2018).
26. Rabosky, D. L. & Huang, H. A robust semi-parametric test for detecting trait-dependent diversification. *Syst. Biol.* **65**, 181–193 (2016).
27. Brown, J. H. Why are there so many species in the tropics? *J. Biogeogr.* **41**, 8–22 (2014).
28. Wagner, C. E., Harmon, L. J. & Seehausen, O. Cichlid species–area relationships are shaped by adaptive radiations that scale with area. *Ecol. Lett.* **17**, 583–592 (2014).
29. Quintero, I. & Jetz, W. Global elevational diversity and diversification of birds. *Nature* **555**, 246–250 (2018).

Acknowledgements We thank M. Grundler for statistical and coding advice, and M. Venzon and A. Noonan for assistance with dataset assembly. We are grateful to the many institutions that curate the primary biodiversity data that underlie several of our analyses (see Supplementary Table 6). This research was carried out using computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor. This work was supported in part by NSF grant DEB-1256330 (D.L.R.), an NSF DDIG grant to J.C. (DEB-1601830), an Encyclopedia of Life Rubenstein Fellowship to J.C. (EOL-33066-13) and by a Fellowship from the David and Lucile Packard Foundation (D.L.R.). P.F.C. was funded by a Gaylord Donnelley Postdoctoral Environment Fellowship (Yale) and through the ARC Centre of Excellence for Coral Reef Studies. We thank J. Johnson for creating the fish images in Fig. 3 and Extended Data Fig. 7.

Reviewer information *Nature* thanks O. Bininda-Emonds, O. Seehausen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.L.R. and M.E.A. designed the study. D.L.R. drafted the paper with substantial input from P.O.T., M.E.A. and J.C. J.C., P.O.T., M.E.A., P.F.C., L.S., M.F., K.K., C.G., T.J.N., M.C. and D.L.R. contributed data. J.C., P.O.T. and D.L.R. developed methods, and P.O.T. and J.C. developed pipelines for data processing and analysis. D.L.R., P.O.T., J.C. and M.E.A. analysed data. All authors contributed to interpretation and discussion of results. Authorship order for P.O.T. and J.C. was determined by coin toss.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0273-1.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0273-1.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.L.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Matrix assembly. We used the PHLAWD pipeline[30] to generate a 27-gene multilocus alignment for ray-finned fishes (Supplementary Information). Guide alignments were constructed using data from recently published studies of higher-level actinopterygian relationships[31,32]. Guide alignments also included new sequences for 442 species of actinopterygians (Supplementary Table 2; see 'Data availability') generated using a standardized phylogenetic workflow for fishes[32]. PHLAWD produced a preliminary alignment of 15,606 species. We performed a series of curation steps including BLAST searches of each sequence back to GenBank to identify taxonomically misassigned species, taxonomic name reconciliation against the California Academy of Sciences taxonomy, duplicate species detection and visual identification of poorly aligned sequences (Supplementary Information). We removed rogue sequences using the RogueNaRok searches[33] and performed preliminary tree searches in RAxML to identify and remove sequences with pathologically long branches due to misalignment. After curation of the raw alignment, our final alignment contained 11,638 species. We used PartitionFinder[34] to identify a model of sequence evolution for multigene alignment and RAxML to find the maximum likelihood topology and calculate Shimodaira–Hasegawa-like support values[35] (Supplementary Information).

Divergence time analysis. We surveyed the palaeontological literature and museum catalogues to assemble our actinopterygian fossil calibration set (139 early occurrences for 130 nodes; see Supplementary Information and 'Data availability'). Fossil assignment to nodes was based upon synapomorphies for that node from published phylogenetic studies, diagnostic characters for taxonomic ranks and/or detailed surveys of clade fossil records by experts. Fossil ages were used as minimum age constraints; maximum ages were derived for all nodes following the Whole Tree Extension of the Hedman algorithm[36], a probabilistic method that incorporates outgroup ages and that has recently been implemented in R[37]. We identified 130 nodes that could be assigned fossil constraints (Supplementary Information) and time-calibrated the phylogeny using treePL[38]. A graphical summary of the distribution of calibrations across the phylogeny is shown in Extended Data Fig. 1.

Phylogenetic placement of unsampled species by taxonomy. Using the time-calibrated phylogeny as a backbone, we generated a distribution of trees in which missing taxa were placed according to their taxonomy. For each of the unsampled species of ray-finned fish, we assigned the most restrictive taxonomic rank (for example, genus, family, order) that was recovered as monophyletic in our maximum likelihood phylogeny. To determine divergence times for unsampled species in the phylogeny, we sampled from a distribution of waiting times conditioned on rank-specific estimates of the speciation rate and sampling fraction using a custom Python script implementing functions from TreePar and SimTree[39–41], and added unsampled species based on the assigned taxonomic rank and inferred waiting time. This procedure was repeated 100 times to generate a distribution of fully sampled ray-finned fish phylogenies (Supplementary Information). This procedure is similar to stochastic polytomy resolution as implemented in PASTIS[42], but permits construction of extremely large phylogenies using all molecular data in a single analysis, rather than a two-stage process that begins with a reduced backbone dataset followed by separate tree searches for each crown lineage. Additionally, our procedure generates a local estimate of diversification rates at each taxonomic node, rather than using a global diversification rate, permitting more accurate placements of unsampled taxa when diversification rate heterogeneity may exist.

Estimation of geographic ranges and species richness. We used the AquaMaps algorithm[43,44] to estimate geographic ranges for marine fishes. These maps were generated using an environmental envelope approach that predicts species distributions based on available species-specific occurrence records at the 0.5°-grid-cell scale in conjunction with the following environmental predictors: depth, sea surface temperature, salinity, proportional ice cover and primary productivity[45]. The predictive algorithm also incorporated geographical bounding boxes to limit occurrences to known ocean-scale distributions for each taxon. We transformed the AquaMaps distributions to a Behrmann equal area projection, and upscaled the resulting grids to 150 × 150-km resolution. We then converted the AquaMaps suitability scores for each cell to binary presence or absence by applying a fixed threshold of 0.5. This threshold was selected based on manual inspection of a number of individual species ranges. Expert opinion was then used when available to further refine the projected distributions, typically by truncating the AquaMaps predictions in light of museum occurrence data, known biogeographical barriers and specialist literature on particular taxa. Where available, we incorporated more accurate distributional maps produced by taxonomic experts in particular groups[46–48]. The final dataset included maps for 12,050 marine species out of an estimated total of 15,500 described marine species[49]. Our conclusions should be unaffected by these missing and uncommon taxa, given that we were able to reconstruct the previously hypothesized pattern of marine fish richness on a global scale[3,4].

Occurrence-based analyses. As an alternative to range predictions from AquaMaps suitability scores, we performed a parallel set of occurrence-based analyses in which we reconstructed cell-based species assemblages as well as species latitudinal midpoints. We obtained all actinopterygian records from four major biodiversity occurrence aggregators (Global Biodiversity Information Facility (GBIF), Ocean Biogeographic Information System (OBIS), Fishnet2 and VertNet) between February 2014 and January 2015 and removed redundancies, resulting in a total of 13,322,575 marine fish occurrences. We downloaded all actinopterygian data from GBIF (https://www.gbif.org/) using their download API version 1; FishNet2 data (http://www.fishnet2.net/) were acquired using a custom Python script to download KML files for each species. VertNet (http://www.vertnet.org/) and OBIS (http://www.iobis.org/) data were retrieved by contacting the administrators of these databases, who then provided us with the relevant data. To reconcile and combine the four datasets, we used museum accession numbers to deduplicate identical records contained in multiple databases. Where accession identifiers were inconsistent within a single museum, we unified these accessions onto a common scheme using a custom Python script. To reconcile species names by resolving synonyms and other sources of error, we used the same procedure described in the Supplementary Information. Institutions contributing substantially to the occurrence dataset are listed in Supplementary Table 6.

The occurrence dataset was filtered to exclude records that fell on land, and records with zero–zero or other nonsensical coordinates. Species richness counts were then calculated across a global grid at 300 × 300 km resolution, using the Behrmann equal area projection. We further excluded isolated grid cells with recorded species, and removed cells that were greater than two standard deviations from the residuals of a thin plate spline interpolation that was fit to the species richness grid. These filters allowed us to remove cells that were probably unrealistic representations of the species diversity at those locations. For all analyses presented, the same richness and bathymetry filters were applied that were used with the primary map data.

Estimates of speciation rate. We reconstructed speciation rates using (i) an inverse equal-splits measure of speciation rate (λ_{DR}), also known as the 'DR statistic'[21,50,51], (ii) BAMM estimates of speciation rate allowing time-varying rate regimes (λ_{BAMM})[20,52,53], and (iii) BAMM estimates of speciation rate assuming time constancy of speciation rates within rate regimes ($\lambda_{BAMM\text{-}TC}$). For the λ_{DR} analyses, we accounted for missing taxa by computing λ_{DR} for each tip in the ATA 31,526 taxon phylogenies; we then computed the average value for each taxon across the set of 100 trees generated with stochastic polytomy resolution. Stochastic polytomy resolution generates taxonomic placements that may compromise inferences of trait-dependent diversification because taxa are placed on trees in a manner that is inconsistent with the underlying process of trait evolution[54], and we excluded all taxa lacking genetic data from formal statistical analysis of the relationship between latitude and speciation. However, including these taxa during estimates of λ_{DR} reduces bias due to incomplete taxon sampling and our calculations of λ_{DR} effectively integrate over the number and location of unsampled species.

BAMM analyses were performed on the time-calibrated phylogeny containing 11,638 tips for which genetic data were available. For each of the two classes of BAMM models (λ_{BAMM} and $\lambda_{BAMM\text{-}TC}$), we performed three BAMM runs for 50 million generations using default MCMC operators and a prior expectation of 500 shifts to facilitate convergence[55]. Raw output and control files to repeat these analyses are available through the Dryad data repository (see 'Data availability'). We were unable to achieve satisfactory convergence when running BAMM on the all-taxon (31,526 tip) phylogenies; we therefore used sampling fractions to account for the effects of incomplete sampling. We corrected for incomplete sampling at the family level. We computed the mean of the marginal posterior distribution of speciation rates for each tip in the phylogeny for both λ_{BAMM} and $\lambda_{BAMM\text{-}TC}$. As an alternative to λ_{BAMM} and λ_{DR}, we computed a simple node density estimate of speciation rate[22]. For each taxon, these estimates are computed simply as n_T/T, in which n_T is the number of nodes on a path of length T, traversing the tree backwards from the tips towards the root. An estimate for an interval of 5 Myr would represent the average speciation rate for a given tip during the past 5 Myr. We computed node density estimates of speciation rate for a sequence of intervals between 0.25 and 50 Myr (Extended Data Fig. 6 and Supplementary Information). As for λ_{DR}, the node density estimates of speciation rate were computed over the full set of ATA phylogenies.

Grid-based analyses of speciation rate. We computed mean speciation rates (λ_{DR}, λ_{BAMM} and $\lambda_{BAMM\text{-}TC}$) for regional assemblages of fishes, focusing on sets of species that are presumed to occur together at the scale of the 150 × 150-km grid cell. We computed the mean rate for individual grid cells four different ways, to reduce spatial and taxonomic pseudoreplication across cells. The simplest approach involved computing the arithmetic mean λ for all species inferred to be present in a particular cell (Fig. 1). Following Jetz et al.[21], we computed weighted arithmetic and geometric means of speciation rates to reduce the contribution of geographically widespread taxa to the overall mean. For the arithmetic mean, the rate for the kth

grid cell is computed as $\lambda_k = (\Sigma w_i \lambda_i)/\Sigma w_i$ in which Σ denotes a summation over all N species ($i = 1$ to $i = N$) present in the kth cell, and w_i is the weight assigned to the ith species. We computed weights for each species as the inverse of the number of grid cells in which the species was found[21]. Therefore, geographically widespread taxa contribute less to a cell mean than a taxon with a highly restricted geographical distribution. Finally, we computed cell means for 'realm endemics'—species uniquely found in one of 12 biogeographical realms under the MEOW marine bioregionalization scheme[56] ($n = 1{,}053$ and 3,100 realm endemics with and without genetic data). The analysis of endemic taxa is particularly informative as such taxa provide more localized information about speciation rates in particular geographical regions relative to widespread taxa that may be found in multiple regions[57].

To formally assess the relationship between latitude and assemblage-level speciation rate, we first computed the mean speciation rate for all cells within a particular ecoregion ($n = 262$) under the MEOW biogeographical regionalization[56]. We modelled the speciation–latitude relationship at the scale of ecoregions rather than individual cells because of the high autocorrelation between adjacent cells, which was reduced at the ecoregion scale, and to reduce the computational burden associated with analysing the full (16,150 grid cell) dataset. To account for spatial autocorrelation between ecoregions, we implemented simultaneous autoregressive error (SAR) models using the spdep package in R[58–60]. These and other statistical tests are two-tailed. We defined neighbours for SAR models as those ecoregions with contiguous boundaries; we then selected the appropriate weighting scheme using Akaike information criterion (AIC) model selection. Simple visual inspection of our data (Fig. 1 and Extended Data Fig. 2) and ordinary least squares (OLS) breakpoint regressions reveal a clear biphasic signal in the relationship between speciation rate and latitude, with a linear increase for higher latitude cells (approximately 30° N and 30° S) and a much weaker relationship for low (tropical) latitudes. We therefore considered an expanded set of breakpoint SAR models with no relationship between absolute latitude and speciation for cells below a particular threshold value, and a linear effect of absolute latitude on speciation above the threshold. We used maximum likelihood analyses to estimate the threshold location and we compared the fit of the breakpoint model to a simple no-breakpoint SAR model using AIC (Extended Data Fig. 2). We used Moran's I to test for spatial autocorrelation in the residuals of OLS and SAR regressions to determine whether the SAR model successfully accounted for spatial non-independence in the data. We tested the relationship between assemblage speciation rate and latitude for ecoregions with absolute latitude less than the previously identified breakpoints (for example, tropical and other low-latitude regions). In general, there is at most a marginal effect of latitude on speciation rate for tropical and subtropical regions (Extended Data Fig. 2h). Finally, we estimated endemism for each MEOW marine biogeographical province using two analyses of occupancy. These two analyses of regional endemism, E, are given by $E = (1/N)\Sigma(1/O_k)$, in which N is the number of species occurring in the focal region, O_k is the estimated global occupancy of the kth species from that region, and Σ denotes a summation from $k = 1$ to $k = N$. Occupancy is computed as either the total number of biogeographical provinces or as the total number of equal-area grid cells in which the taxon is found.

Trait-dependent speciation. We treated the absolute value of the latitudinal midpoint of each species as a 'trait' and tested its relationship to speciation rates using formal statistical methods for analysing trait-dependent diversification[26,51]. The latitudinal midpoint for each species was computed as the centroid midpoint of the geographical range of the species. We used three recently developed methods for testing the effects of species traits on lineage speciation rates that are robust to phylogenetic pseudoreplication and model misspecification[61,62]. Using ES-SIM[25], we tested whether λ_{DR} was correlated with absolute latitudinal midpoint for individual species. Using FiSSE[51], we then tested whether two discrete classifications of species by latitude ('low latitude/tropical' versus 'high latitude/temperate') differed in their rate of speciation as measured using λ_{DR}. We performed the FiSSE test across a range of thresholds for classifying lineages as tropical and temperate (23.5°, 25°, 30°, 35°, 40°, 45°, 50°, 55° and 60°). Regardless of the threshold, all FiSSE results indicated a highly significant effect of latitude on speciation rate (Extended Data Table 1). As an alternative method for continuous traits, we used STRAPP[26] to test whether latitude was correlated with the two BAMM-based measures of speciation rate (λ_{BAMM} and $\lambda_{BAMM\text{-}TC}$). Results for λ_{BAMM} and $\lambda_{BAMM\text{-}TC}$ were almost identical and identified a strong effect of latitude on speciation rate (Pearson's $r = 0.30$–0.31; $P \leq 0.006$). One possible explanation for our results is that high-latitude assemblages are enriched for deep-sea taxa, and that faster speciation is actually a property of deep-sea environments and not high latitudes. To test this hypothesis, we obtained depth classifications for marine fishes from FishBase (http://www.fishbase.org); minimum and maximum depths were available for 4,089 species (of 5,231 total species). We used ES-SIM to test the relationship between latitude and speciation λ_{DR} for fishes with minimum depth >200 m. Using FiSSE, we tested whether speciation rates were faster for low-latitude deep-sea fishes relative to low-latitude shallow or reef-associated species (Extended Data Fig. 7a).

Additional checks on statistical robustness. We performed several additional checks on the robustness of the latitude–speciation correlation. We visualized latitudinal trends in terminal branch length, which is expected to correlate inversely with underlying speciation rate. We obtained estimates of the mean terminal branch length of each species from the distribution of ATA (31,526 taxon) phylogenies. The inverse of these branch lengths is the simplest possible estimate of the instantaneous rate of speciation, although it is an extremely noisy metric; λ_{DR} is similar but includes the weighted contribution of earlier branches to increase the signal-to-noise ratio. Despite the overall noisiness of the terminal branch length metric, there is a clear trend towards shorter terminal branches for high-latitude taxa (Extended Data Fig. 9).

We also tested whether our speciation rate estimates could have been driven by latitudinal gradients in genetic taxon sampling, as might be the case if a higher percentage of high-latitude taxa had DNA sequences with which to infer their phylogenetic position without relying on stochastic polytomy resolution. To formally address this potential confounding variable, we fitted multiple regression models to the relationship between lineage-specific speciation rate and latitudinal midpoint, but including the sampling fraction for each species as a covariate. The sampling fraction for each taxon was simply the percentage of total species from the corresponding family-level clade that contained genetic data (for example, the percentage of total species from each family that were represented in the genetic supermatrix). These sampling fractions were identical to those used to correct for incomplete sampling in the BAMM analyses. Visual inspection and formal analysis shows minimal effect of sampling fraction on the patterns reported here (Extended Data Fig. 9).

Finally, we tested whether variation in the rate of molecular evolution could drive spurious variation in the rate of diversification. A systematic bias towards low rates of molecular evolution can lead to apparent fast rates of diversification in slowly evolving lineages as an artefact of the algorithms used for time-scaling the raw (uncalibrated) phylogeny. If our results are affected by this bias, we expect to observe (i) a general trend towards slower rates of molecular evolution at high latitudes, and (ii) a negative correlation between speciation rate and the rate of molecular evolution. To estimate rates of molecular evolution, we computed the relative root-to-tip distances for each taxon in the phylogeny and estimated their correlation with both latitude and λ_{DR} (Extended Data Fig. 9). There is no evidence that higher latitudes are associated with slower rates of molecular evolution, or that rates of molecular evolution are negatively correlated with λ_{DR} (see Extended Data Fig. 9).

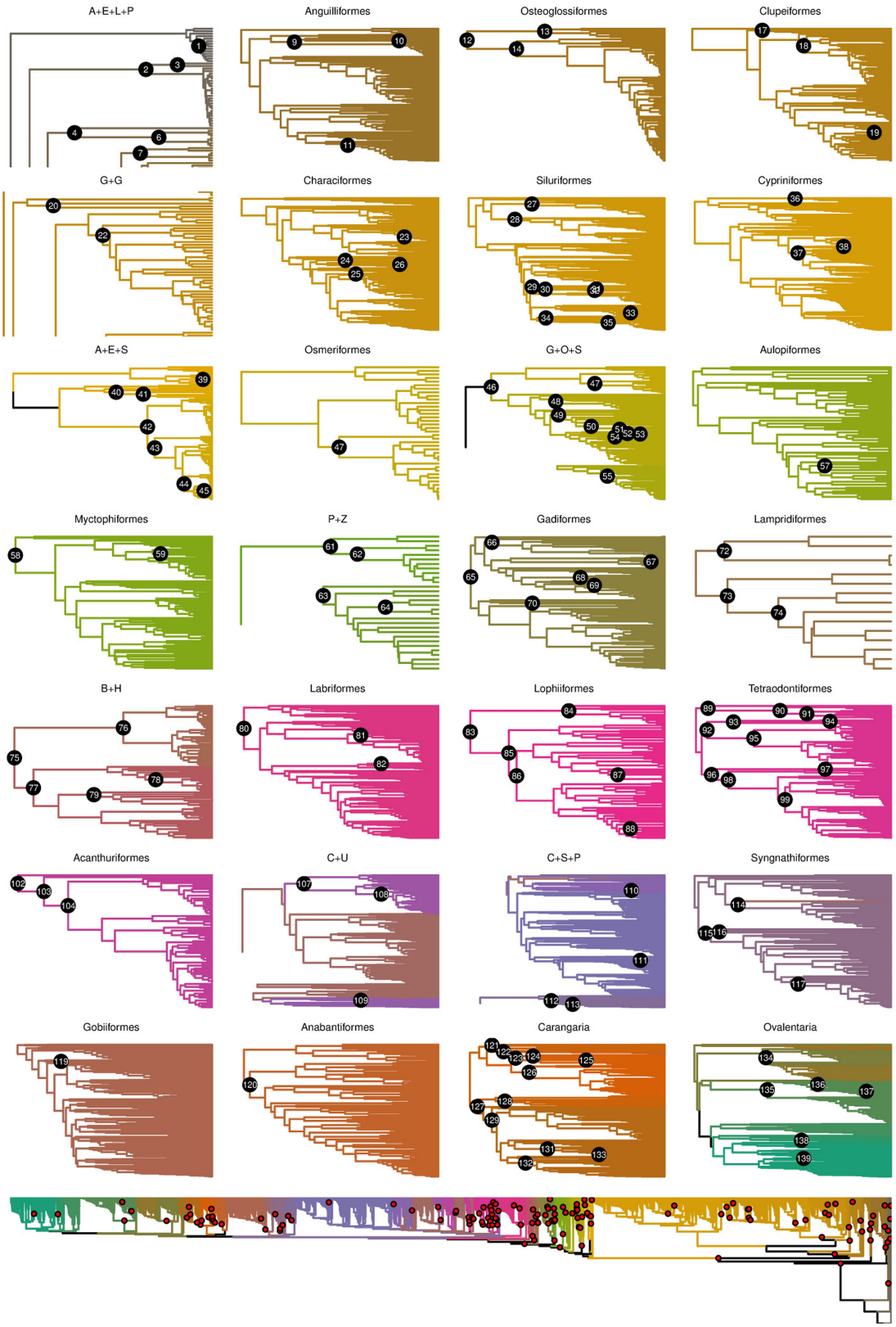
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. All scripts and code necessary to repeat the analyses described here have been made available through the Dryad digital data repository (https://doi.org/10.5061/dryad.fc71cp4).

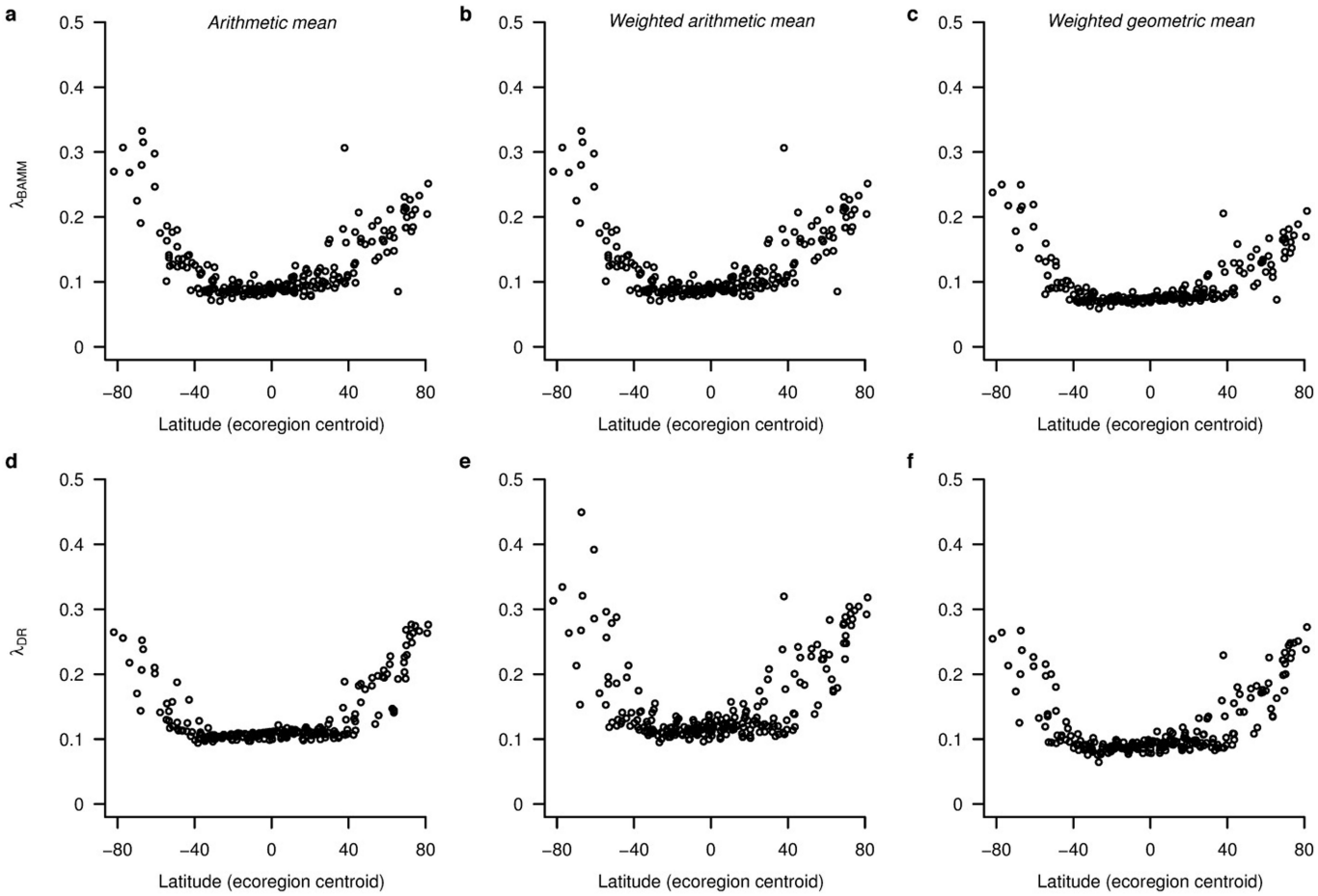
Data availability. All data necessary to repeat the analyses described here have been made available through the Dryad digital data repository (https://doi.org/10.5061/dryad.fc71cp4). Phylogenetic tree distributions are also available through http://fishtreeoflife.org.

30. Smith, S. A., Beaulieu, J. M. & Donoghue, M. J. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evol. Biol.* **9**, 37 (2009).
31. Betancur-R, R. et al. The tree of life and a new classification of bony fishes. *PLoS Curr. Tree of Life* https://doi.org/10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288 (2013).
32. Near, T. J. et al. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl Acad. Sci. USA* **110**, 12738–12743 (2013).
33. Aberer, A. J., Krompass, D. & Stamatakis, A. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst. Biol.* **62**, 162–166 (2013).
34. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
35. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
36. Hedman, M. M. Constraints on clade ages from fossil outgroups. *Paleobiology* **36**, 16–31 (2010).
37. Lloyd, G. T., Bapst, D. W., Friedman, M. & Davis, K. E. Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. *Biol. Lett.* **12**, 20160609 (2016).
38. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
39. Stadler, T. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *J. Theor. Biol.* **261**, 58–66 (2009).
40. Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl Acad. Sci. USA* **108**, 6187–6192 (2011).
41. Stadler, T. Simulating trees with a fixed number of extant species. *Syst. Biol.* **60**, 676–684 (2011).

42. Thomas, G. H. et al. PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences. *Methods Ecol. Evol.* **4**, 1011–1017 (2013).
43. Ready, J. et al. Predicting the distribution of marine organisms at the global scale. *Ecol. Modell.* **221**, 467–478 (2010).
44. Kaschner, K. et al. AquaMaps: predicted range maps for aquatic species. version 08/2016 http://www.aquamaps.org (2016).
45. Kaschner, K. et al. AquaMaps environmental dataset: half-degree cells authority file (HCAF). version 6, 08/2016 https://www.aquamaps.org/main/envt_data.php (2016).
46. Mecklenburg, C. W., Mecklenburg, T. A., Sheiko, B. A. & Steinke, D. *Pacific Arctic Marine Fishes* (CAFF, Akureyri, 2016).
47. Coll, M. et al. The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. *PLoS ONE* **5**, e11842 (2010).
48. IUCN. The IUCN Red List of Threatened Species. version 2016-1 http://www.iucnredlist.org, downloaded on 8 February 2018 (2016).
49. 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. B* **275**, 149–155 (2008).
50. Belmaker, J. & Jetz, W. Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients. *Ecol. Lett.* **18**, 563–571 (2015).
51. Rabosky, D. L. & Goldberg, E. E. FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution* **71**, 1432–1442 (2017).
52. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, e89543 (2014).
53. Rabosky, D. L., Donnellan, S. C., Grundler, M. & Lovette, I. J. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Syst. Biol.* **63**, 610–627 (2014).
54. Rabosky, D. L. No substitute for real data: a cautionary note on the use of phylogenies from birth–death polytomy resolvers for downstream comparative analyses. *Evolution* **69**, 3207–3216 (2015).
55. Mitchell, J. S. & Rabosky, D. L. Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. *Methods Ecol. Evol.* **8**, 37–46 (2017).
56. Spalding, M. D. et al. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* **57**, 573–583 (2007).
57. Cowman, P. F., Parravicini, V., Kulbicki, M. & Floeter, S. R. The biogeography of tropical reef fishes: endemism and provinciality through time. *Biol. Rev. Camb. Philos. Soc.* **92**, 2112–2130 (2017).
58. Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. *R News* **5**, 9–13 (2005).
59. Bivand, R., Hauke, J. & Kossowski, T. Computing the Jacobian in Gaussian spatial autoregressive models: an illustrated comparison of available methods. *Geogr. Anal.* **45**, 150–179 (2013).
60. Bivand, R. & Piras, G. Comparing implementations of estimation methods for spatial econometrics. *J. Stat. Softw.* **63**, 1–36 (2015).
61. Rabosky, D. L. & Goldberg, E. E. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* **64**, 340–355 (2015).
62. Maddison, W. P. & FitzJohn, R. G. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol.* **64**, 127–136 (2015).



Extended Data Fig. 1 | Phylogenetic placement of fossil calibrations in major fish lineages. Major lineages are broken into subclades (top) to visualize fossil calibrations and are coloured by taxonomic order. Numbered nodes are described in the calibration report in the Dryad data repository. The same calibrations are red circles in the full phylogeny (bottom). A + E + L + P: Acipenseriformes, Elopiformes, Lepisosteiformes, Polypteriformes; A + E + S: Argentiniformes, Esociformes, Salmoniformes; B + H: Beryciformes, Holocentriformes; C + S + P: Centrarchiformes, Scombriformes, Perciformes; C + U: Chaetodontiformes, Uranoscopiformes; G + G: Gonorynchiformes, Gymnotiformes; G + O + S: Galaxiiformes, Osmeriformes, Stomiatiformes; P + Z: Percopsiformes, Zeiformes.



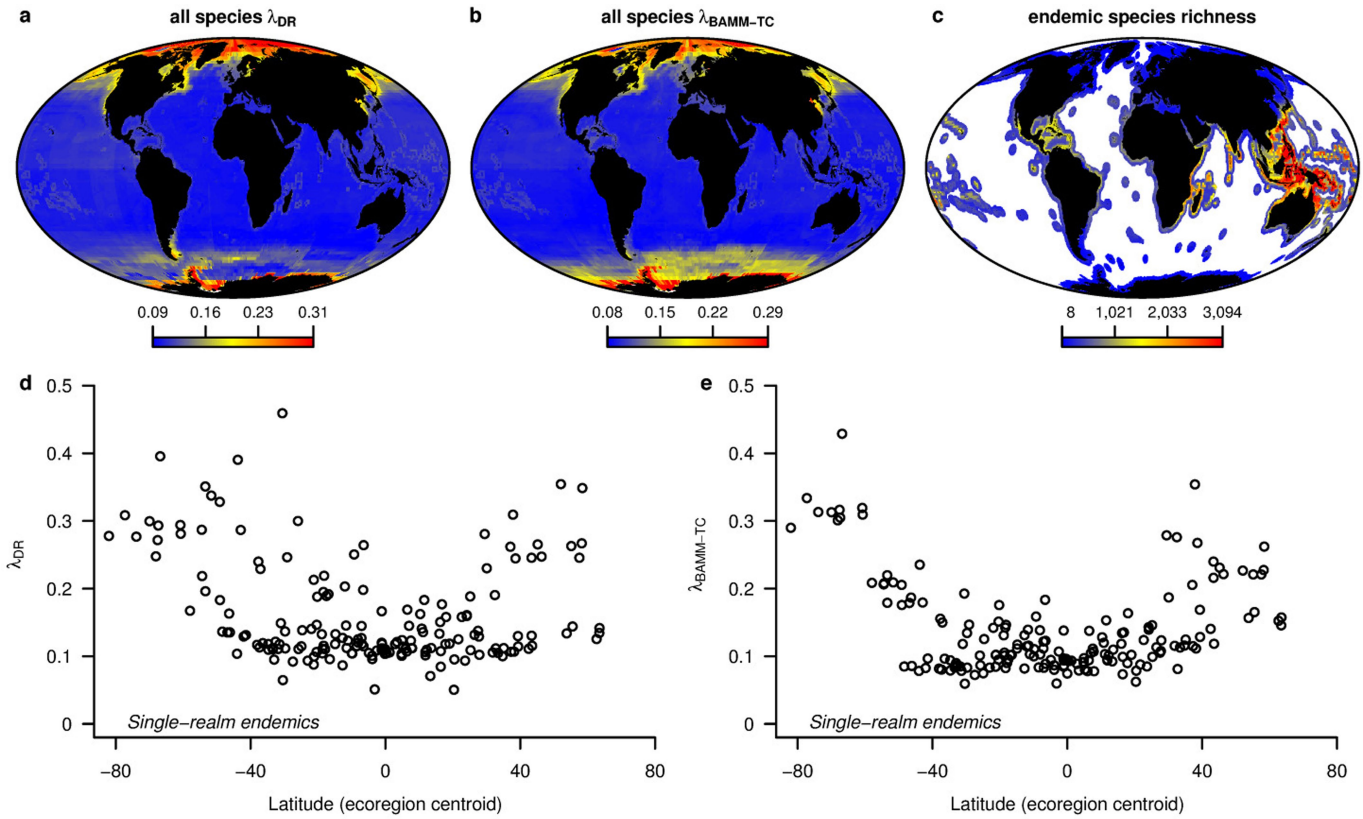
g

metric	AIC1	AIC2 (breakpoint)	breakpoint	slope	p	SAR.I	SAR.I_p	OLS.I	OLS.I_p
λ_{BAMM} : unweighted arithmetic mean	−1225.502	−1306.396	40	0.003	< 0.001	−0.011	0.546	0.55	< 0.001
λ_{BAMM} : weighted mean, arithmetic	−1045.137	−1072.144	27	0.003	< 0.001	0.015	0.357	0.525	< 0.001
λ_{BAMM} : weighted mean, geometric	−1196.958	−1250.247	34.5	0.003	< 0.001	0.011	0.385	0.544	< 0.001
λ_{DR} : unweighted arithmetic mean	−1166.941	−1275.119	41.5	0.004	< 0.001	−0.048	0.778	0.525	< 0.001
λ_{DR} : weighted mean, arithmetic	−929.573	−967.732	33.5	0.004	< 0.001	0.045	0.187	0.521	< 0.001
λ_{DR} : weighted mean, geometric	−1091.167	−1164.07	35	0.003	< 0.001	0.021	0.322	0.442	< 0.001

h

metric	AIC (OLS)	AIC (SAR)	OLS.slope	SAR.slope	p	slope ratio	SAR.I	SAR.I_p	OLS.I	OLS.I_p
λ_{BAMM} : unweighted arithmetic mean	−1029.162	−1060.477	0.0002	0.0002	0.0125	0.076	0.052	0.115	0.318	< 0.001
λ_{BAMM} : weighted mean, arithmetic	−736.036	−778.44	0.0002	< 0.0001	0.7473	0.014	0.032	0.285	0.495	< 0.001
λ_{BAMM} : weighted mean, geometric	−991.157	−1014.502	0.0002	0.0001	0.0856	0.044	0.017	0.365	0.334	< 0.001
λ_{DR} : unweighted arithmetic mean	−1039.202	−1083.7	< 0.0001	0.0001	0.2089	0.028	0.028	0.267	0.386	< 0.001
λ_{DR} : weighted mean, arithmetic	−751.094	−805.78	0.0003	< 0.0001	0.6231	0.022	0.001	0.453	0.499	< 0.001
λ_{DR} : weighted mean, geometric	−902.099	−950.599	< 0.0001	< 0.0001	0.3601	−0.034	−0.012	0.527	0.465	< 0.001

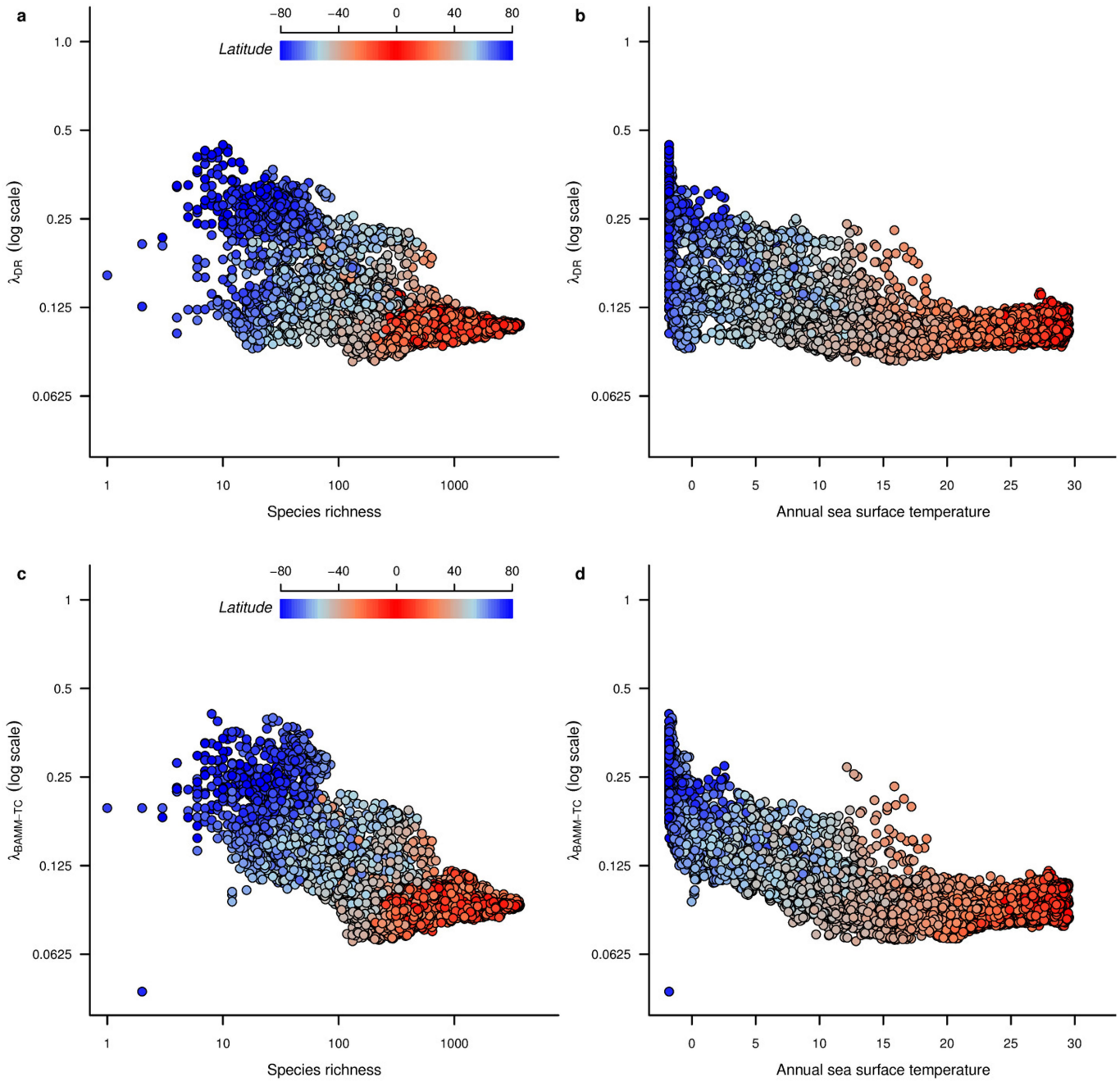
Extended Data Fig. 2 | Relationships between mean speciation rates and latitude for 262 marine ecoregions using alternative methods for the computation of the cell rates. a–c, λ_{BAMM} versus latitude. **d–f**, λ_{DR} versus latitude. Ecoregion rates are mean rates across all cells assigned to each biogeographical region. Arithmetic mean is the mean rate across all taxa inferred to occur in the cell; weighted arithmetic and weighted geometric means assign proportionately greater weight to species with small geographical ranges. Weighting schemes for speciation metrics are described in the Methods. **g**, Simultaneous autoregressive (SAR) spatial error models for the effects of absolute latitude on mean speciation rates for ecoregions. AIC1 gives the Akaike information criterion (AIC) for a linear model with a single slope and intercept term; AIC2 is the corresponding AIC for a breakpoint model that assumes no relationship (slope = 0) between absolute latitude and speciation rate for all values below some threshold, and a linear relationship for latitudes that exceed the threshold. SAR.*I* and SAR.I_P are global Moran's *I* estimates and associated *P* values for assessing the presence of residual spatial autocorrelation in the model residuals; OLS.*I* and OLS.I_P are the corresponding values for ordinary least squares (OLS) regression that ignores spatial autocorrelation. All SAR models show highly significant effects of latitude on speciation rate, and breakpoint models provided a consistently better fit than models without a breakpoint. **h**, OLS and SAR models for the effects of absolute latitude on speciation rate for low-latitude grid ecoregions only. The slope ratio term gives the ratio of slopes for low-latitude ecoregions (below the corresponding breakpoint; **g**) to the slope for ecoregions with latitude above the breakpoint. Overall, there is a marginal effect of latitude on speciation rate for low-latitude ecoregions.



f

metric	AIC1	AIC2 (breakpoint)	breakpoint	slope	p	SAR.I	SAR.I_p	OLS.I	OLS.I_p
λ_{BAMM} : unweighted arithmetic mean	−692.771	−703.745	27.5	0.003	< 0.001	−0.001	0.47	0.566	< 0.001
λ_{BAMM} : weighted mean, arithmetic	−598.557	−609.271	27.5	0.003	< 0.001	0.013	0.389	0.493	< 0.001
λ_{BAMM} : weighted mean, geometric	−692.316	−707.397	32	0.003	< 0.001	0.012	0.392	0.462	< 0.001
λ_{DR} : unweighted arithmetic mean	−575.146	−579.98	37	0.003	< 0.001	−0.09	0.898	0.622	< 0.001
λ_{DR} : weighted mean, arithmetic	−471.397	−479.062	27.5	0.004	< 0.001	−0.012	0.537	0.46	< 0.001
λ_{DR} : weighted mean, geometric	−556.135	−568.005	31.5	0.004	< 0.001	−0.002	0.475	0.327	< 0.001

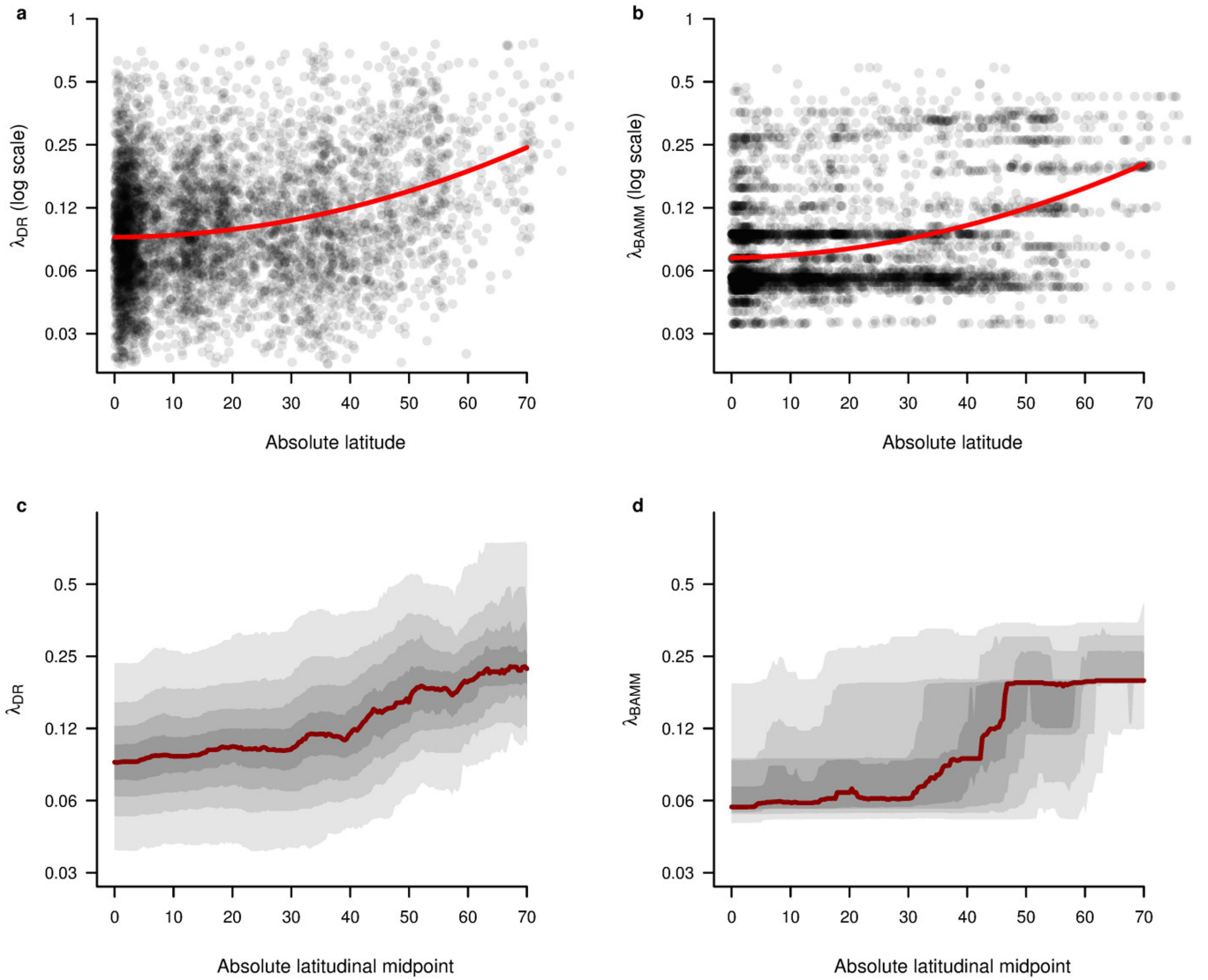
Extended Data Fig. 3 | Relationships between speciation rate and latitude for alternative speciation rate metrics and for endemic taxa only. a, **b**, Global maps of λ_{DR} and $\lambda_{BAMM\text{-}TC}$, as in Fig. 1. **c**, Global map of endemic species richness, by grid cell. 'Endemic' taxa are those that are restricted to a single MEOW realm; an endemic taxon can occur in multiple grid cells provided all grid cells are contained within a single realm. **d**, Relationship between speciation rates (λ_{DR}) and latitude for ecoregions ($n = 232$), computed using realm endemics only. **e**, Relationship between speciation rates ($\lambda_{BAMM\text{-}TC}$) and latitude for ecoregions, computed using realm endemics only. **f**, SAR spatial error models for the relationship between ecoregion speciation rates and absolute latitude, for which ecoregion means are computed from single-realm endemics only. Weighting schemes for assemblages are described in the Methods. SAR modelling results are presented as in Extended Data Fig. 2g and show a strong correlation between latitude and speciation rate.



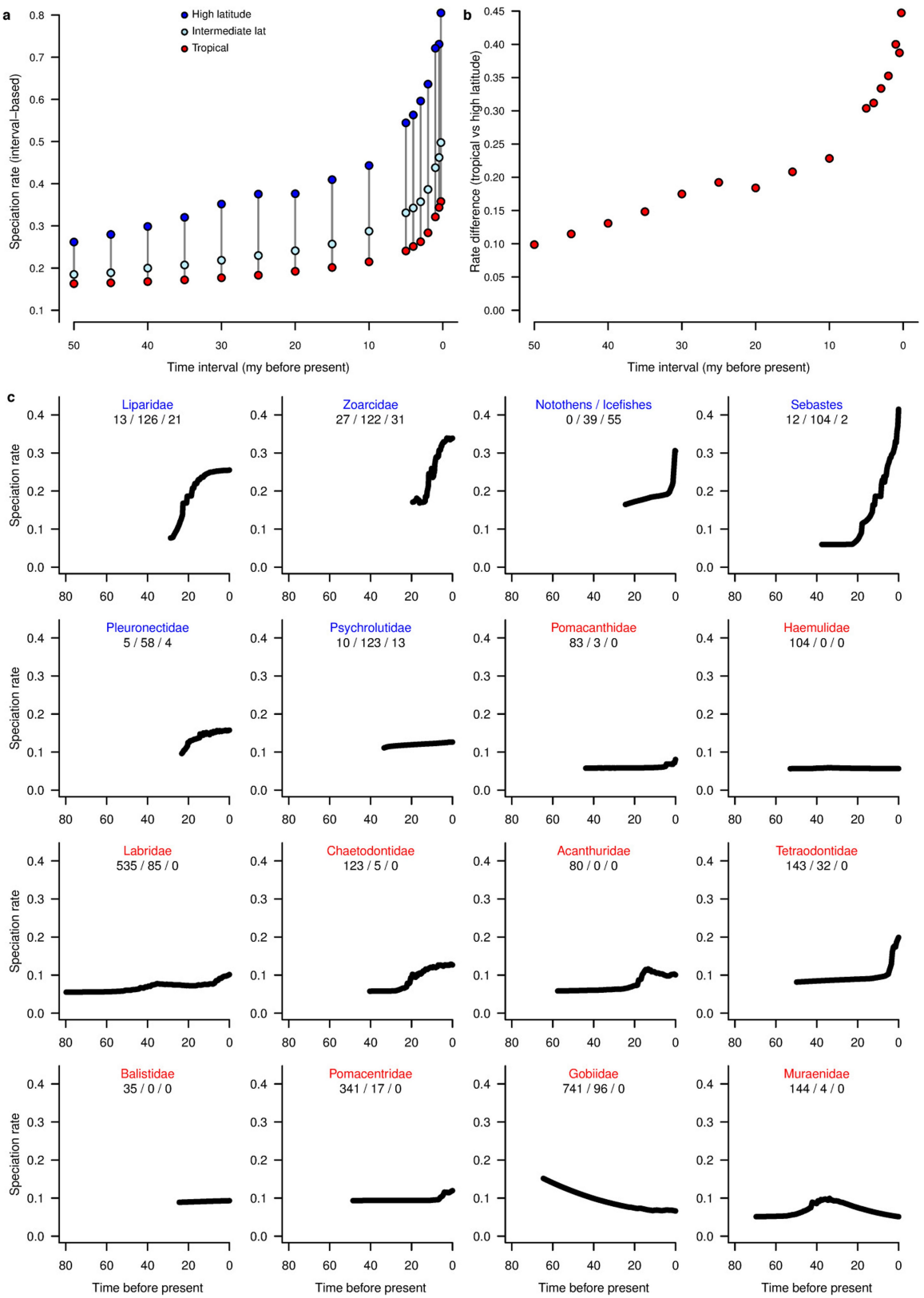
e

Description	Endemicity measure	Speciation measure	r	p
Speciation rate by endemism	Occupancy (provinces)	λ_{BAMM}	0.79	< 0.001
Speciation rate by endemism	Occupancy (provinces)	λ_{DR}	0.79	< 0.001
Speciation rate by endemism	Occupancy (provinces)	$\lambda_{BAMM-TC}$	0.79	< 0.001
Speciation rate by endemism	Range size	λ_{BAMM}	0.57	< 0.001
Speciation rate by endemism	Range size	λ_{DR}	0.61	< 0.001
Speciation rate by endemism	Range size	$\lambda_{BAMM-TC}$	0.56	< 0.001
Endemism by absolute latitude	Occupancy (provinces)	—	0.58	< 0.001
Endemism by absolute latitude	Range size	—	0.46	< 0.001

Extended Data Fig. 4 | Speciation rate, species richness, temperature and endemism. a, Negative relationship between species richness and mean speciation rate (λ_{DR}) for individual grid cells. **b**, Negative relationship between mean annual sea surface temperature and mean speciation rate. **c**, **d**, Same as **a**, **b**, but for BAMM with time-constant rate regimes ($\lambda_{BAMM\text{-}TC}$). Grid cells as in Fig. 1 ($n = 16{,}150$). See Fig. 2 for comparison. **e**, Correlation between mean speciation rate for MEOW biogeographical provinces and two measurements of regional endemism. 'Occupancy (provinces)' measures endemism as the inverse of the mean number of provinces occupied by each taxon that occurs in a particular province. 'Range size' is the inverse mean range size across all taxa occurring in a given province. High values of endemism indicate that a given region consists of species that are found in fewer additional provinces, or of species with smaller geographical ranges. The bottom two rows show the correlations between the endemism parameters and latitude.



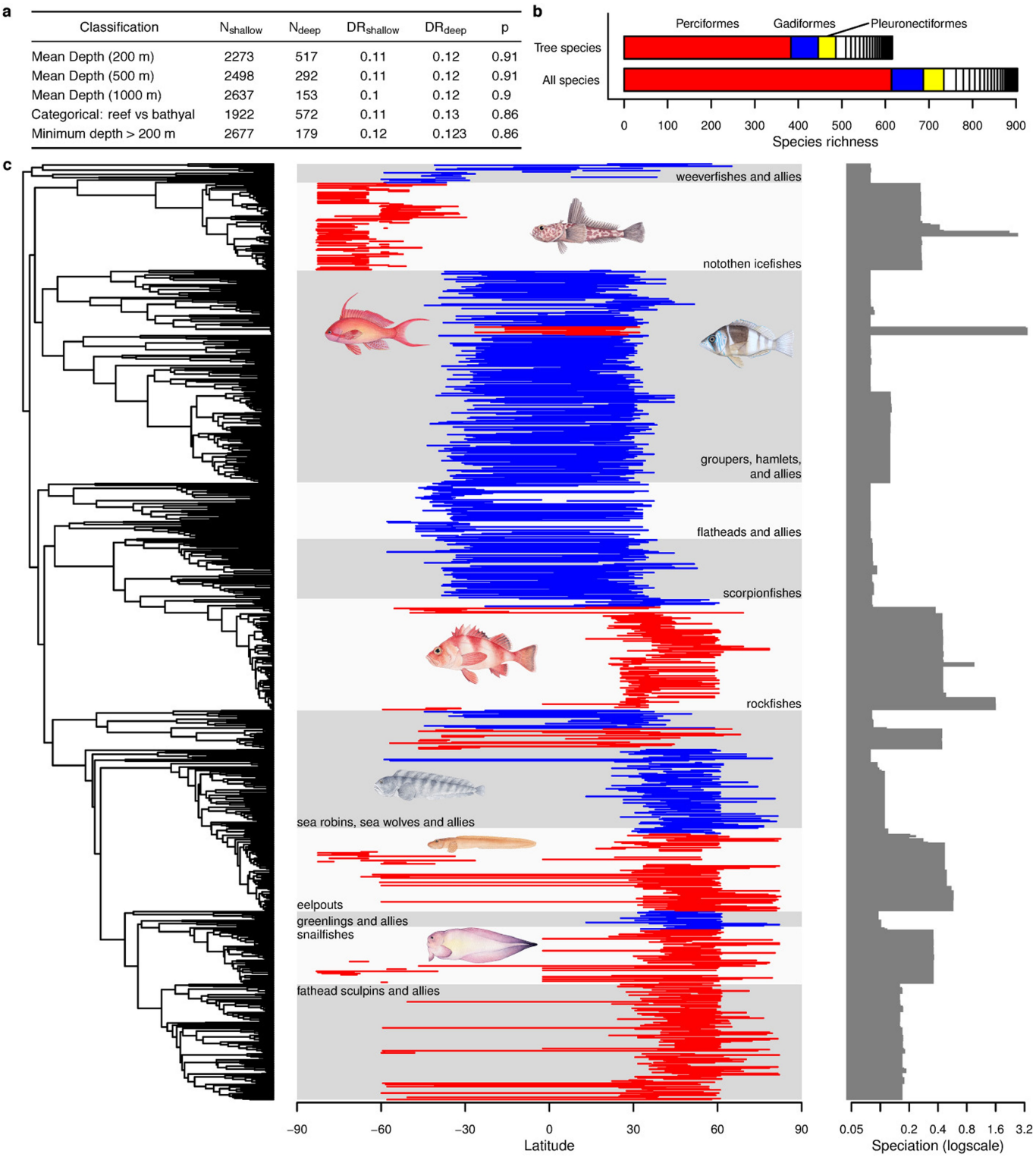
Extended Data Fig. 5 | Speciation rates for individual taxa as a function of latitudinal midpoint. a, λ_{DR} for all marine species with genetic data ($n = 5{,}229$) as a function of the latitudinal (centroid) midpoint of their geographical range. Non-phylogenetic OLS regression with quadratic term is overlaid on points to denote trend in mean rates. **b**, λ_{BAMM} for the same taxon set. **c**, Sliding window analysis of λ_{DR} distributional quantiles in speciation rates by individual taxa with respect to latitudinal midpoint. Contours denote quantiles from 0.10 to 0.90, in 0.10 increments, with a sliding window size of 6°. Dark red line is the median rate. **d**, Distributional quantiles of λ_{BAMM} for all species with respect to latitudinal midpoint; dark red line is median rate.



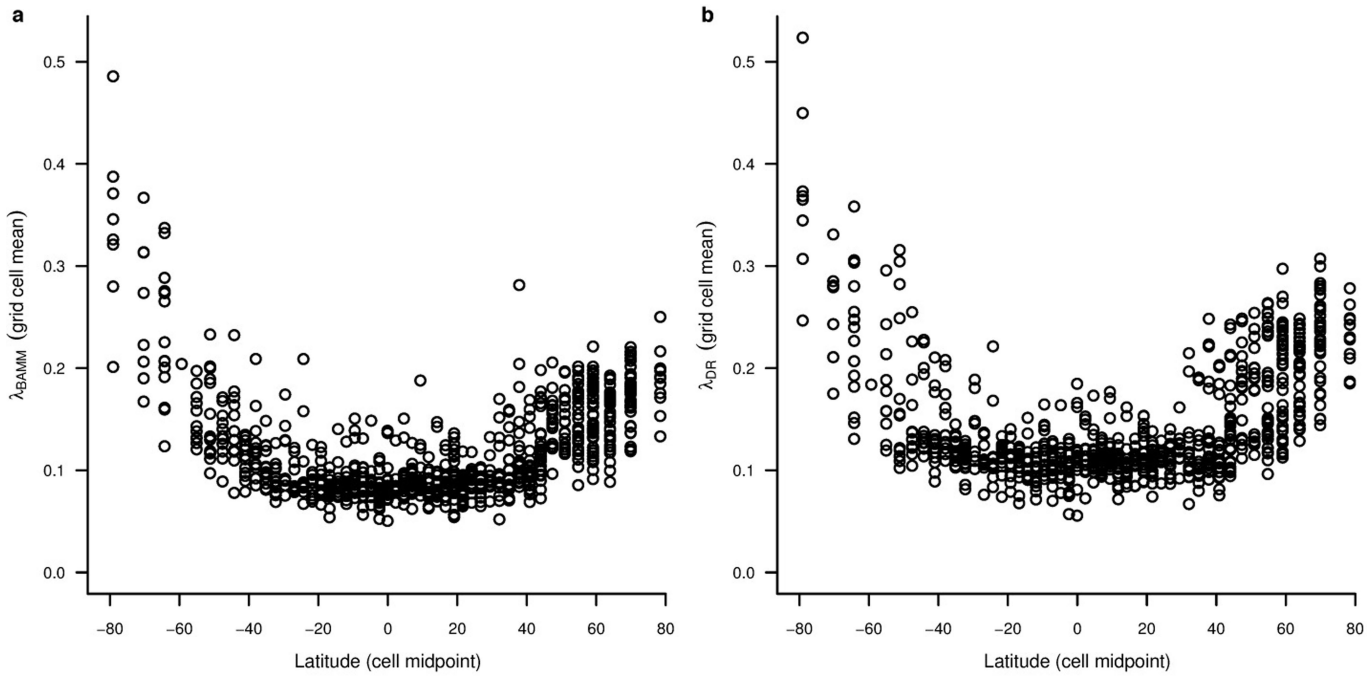
Extended Data Fig. 6 | Temporal dimension of speciation rate variation as a function of latitude. a, Mean speciation rates for taxa from low latitudes (<30°), intermediate latitudes (30–60°) and high latitudes (>60°) computed using the interval method. Per-taxon interval-based rates were computed for time intervals between 0.25 and 50 million years before present. Time-averaged speciation rates for high-latitude fishes are much higher than those inferred for low-latitude fishes, even across timescales that exceed 20 million years. **b**, Rate differential between high-latitude and low-latitude taxa as a function of interval duration. **c**, Speciation-over-time curves reconstructed using the time-varying rates model in BAMM for 14 clades of high-latitude (blue) and low-latitude (red) fishes. Inset numbers for each panel give the numbers of low-, intermediate- and high-latitude (from left to right) taxa from each clade for which geographical range data are available. Low-latitude clades were selected to represent high-diversity and iconic reef-associated clades that contribute substantially to the tropical diversity peak in marine fishes. With the possible exception of gobies, there is no signal of early, rapid speciation in low-latitude or tropical shallow-water clades.

a

Classification	$N_{shallow}$	N_{deep}	$DR_{shallow}$	DR_{deep}	p
Mean Depth (200 m)	2273	517	0.11	0.12	0.91
Mean Depth (500 m)	2498	292	0.11	0.12	0.91
Mean Depth (1000 m)	2637	153	0.1	0.12	0.9
Categorical: reef vs bathyal	1922	572	0.11	0.13	0.86
Minimum depth > 200 m	2677	179	0.12	0.123	0.86



Extended Data Figure 7 | Speciation rates in deep-sea fishes and the phylogenetic structure of high-latitude fish diversity. a, Formal test of the relationship between speciation rate and depth classification for tropical fishes. 'Classification' is the criterion used to define fishes as deep sea versus shallow water; mean depth (200 m) thus classifies all fishes with mean depth greater than 200 m as deep sea. Among tropical fishes, there is no effect of depth state on speciation rate. **b**, Phylogenetic composition of high-latitude fish diversity by taxonomic order, across all marine fishes (top) and for the subset of species with genetic data (bottom). High latitude is defined as having a centroid midpoint greater than 45° north or south. Only the three most species-rich high-latitude orders are labelled. Most high-latitude marine fishes are Perciformes. **c**, Phylogenetic and geographical structure of the diversity of Perciformes. The latitudinal range of each perciform species in the phylogenetic dataset is shown, along with the corresponding speciation rate (λ_{BAMM}). Latitudinal ranges from species with speciation rates that are faster and slower than the median rate are shown in red and blue, respectively. High-latitude and rapidly speciating clades are nested within slowly speciating tropical lineages, and speciation rates for high-latitude taxa of Perciformes are higher than those observed in tropical lineages. Mean speciation rates for high-latitude species (>45°, $n = 376$) are faster than those observed for tropical (<25°, $n = 287$) species (tropical: $\lambda_{DR} = 0.16$, $\lambda_{BAMM} = 0.15$; high latitude: $\lambda_{DR} = 0.30$, $\lambda_{BAMM} = 0.23$). For polar species (>60°, $n = 105$), these rate differentials are even more extreme, with mean $\lambda_{DR} = 0.38$ and $\lambda_{BAMM} = 0.31$.



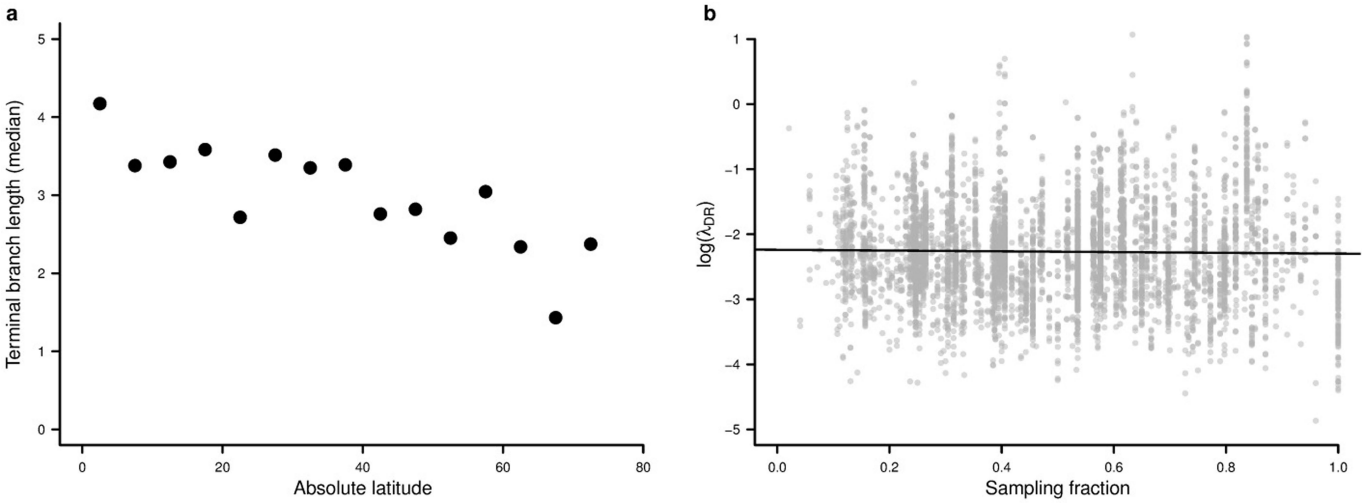
c

metric	AIC1	AIC2 (breakpoint)	breakpoint	slope	p	SAR.I	SAR.I_p	OLS.I	OLS.I_p
λ_{BAMM}	−861.634	−892.706	32.5	0.0031	0	−0.0143	0.5550	0.5094	1.37E-16
λ_{DR}	−805.025	−848.045	34	0.0038	0	−0.0308	0.6545	0.5623	2.76E-19

d

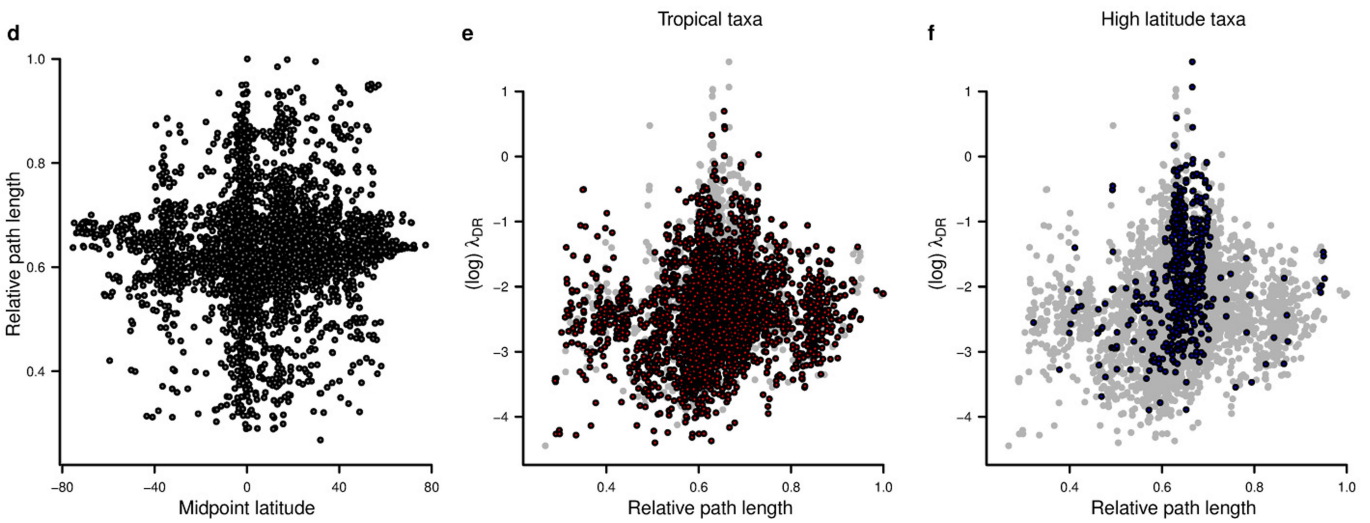
Threshold	λ_1	λ_0	p	Null $\Delta\mu$	Null σ	$n_{parsimony}$	$q_{parsimony}$
30	0.1194	0.2057	< 0.001	0.0001	0.0254	398	0.0044
35	0.1232	0.2194	< 0.001	0.0009	0.0276	303	0.0031
40	0.1270	0.2306	< 0.001	0.0007	0.0310	218	0.0021
45	0.1285	0.2528	< 0.001	−0.0004	0.0316	178	0.0017
50	0.1310	0.2682	< 0.001	0.0004	0.0323	171	0.0017
55	0.1352	0.3019	< 0.001	0.0006	0.0356	108	0.0010
60	0.1370	0.3789	< 0.001	−0.0010	0.0419	65	0.0005

Extended Data Figure 8 | Latitudinal gradient in speciation rate for cell assemblages inferred from occurrence data. Cell assemblages ($n = 843$) and species latitudinal midpoints were inferred from a non-redundant merge of four primary occurrence-based biodiversity databases (GBIF, OBIS, Fishnet2 and VertNet). **a**, λ_{BAMM} for cell assemblages as a function of latitude. **b**, λ_{DR} as a function of latitude. **c**, SAR spatial error models for the effects of absolute latitude on mean speciation rates for grid cells. AIC1 is a linear model with a single slope and intercept term; AIC2 is the corresponding AIC for a breakpoint model that assumes no relationship (slope = 0) between absolute latitude and speciation rate for all values below some threshold, and a linear relationship for latitudes that exceed the threshold. All other column headings as in Extended Data Fig. 2g. Results indicate a strong effect of latitude on speciation rate and are nearly identical to results obtained using the dataset of the primary map. **d**, Effects of absolute latitudinal midpoint for individual taxa on corresponding tip speciation rates, as assessed using FiSSE. Each row gives the results of FiSSE using a different threshold for classifying lineages as tropical and temperate. λ_0 and λ_1 denote estimated speciation rates (similar to λ_{DR}) for tropical and temperate lineages, respectively. All column headings are identical to those shown in Extended Data Table 1. Results are nearly identical to those obtained using explicit range reconstructions and reveal a pervasive effect of latitude on lineage-level speciation rates, regardless of the threshold used to classify species.



c

	Variable	Slope	Sum of Squares	P–value
λ_{BAMM}	latitude	0.012	255.96	< 0.0001
	sampling frac.	0.075	1.45	0.048
λ_{DR}	latitude	0.011	206.71	< 0.0001
	sampling frac.	–0.103	2.74	0.021



Extended Data Fig. 9 | Additional checks of statistical robustness. **a**, Relationship between terminal branch lengths and absolute latitudinal midpoint; means are shown for all species falling into a given bin (±2.5° from the focal value, $n = 15$). Mean branch lengths decrease with increasing latitude, reflecting faster speciation at high latitudes. **b**, Relationship between the estimated speciation rate for each taxon (λ_{DR}, $n = 5{,}155$) and the sampling fraction for the corresponding family-level clade to which the taxon belongs; the sampling fraction is simply the percentage of known taxa from the family that were represented in the phylogenetic dataset with genetic data. There is no clear relationship between the sampling fraction and the estimated speciation rates. **c**, Multiple regression analysis (OLS) of the relationship between taxon-specific speciation rate (λ_{BAMM} or λ_{DR}) and two predictors (latitude and family-level sampling fraction) in a multiple regression framework ($n = 5{,}155$). If the relationship between speciation rate and latitude is driven by progressively greater (or lower) genetic taxon sampling as a function of latitude, the sampling fraction term should explain a large fraction of the overall sums of squares. Even when sampling fraction is included as a covariate, the overwhelming fraction of variance is explained by latitude. For both λ_{DR} and λ_{BAMM}, more than 98% of the total sums of squares is explained by latitude and not sampling. **d**–**f**, Test for the effects of molecular evolutionary rate variation and latitudinal bias in speciation rate. **d**, Relationship between root-to-tip branch length sum for uncalibrated (non-ultrametric) RAxML phylogeny and midpoint latitude for each marine taxon ($n = 5{,}149$). **e**, **f**, Relationship between root-to-tip distance and λ_{DR}. There is effectively no relationship between the total path length for individual tips and their absolute latitudinal midpoint (Pearson $r = 0.020$). Plots in **e** and **f** emphasize tropical (midpoint latitude <25°; $n = 3{,}481$; red) and temperate–polar (midpoint latitude >45°; $n = 567$; blue) taxa, respectively, all other taxa are shown in grey. Overall relationship between (log)λ_{DR} and the rate of molecular evolution (root-to-tip sum) is weak but positive (Pearson $r = 0.130$) and inconsistent with the hypothesis that slow rates of molecular evolution at high latitudes results in fast but spurious estimates of speciation rate.

Extended Data Table 1 | Effects of absolute latitudinal midpoint on speciation rates

Threshold	λ_1	λ_0	p	Null $\Delta\mu$	Null σ	$n_{parsimony}$	$q_{parsimony}$
23.5	0.119	0.187	< 0.001	0.00122	0.022	600	0.007
25	0.12	0.19	< 0.001	−0.00098	0.022	576	0.007
30	0.122	0.2	< 0.001	−0.00035	0.023	468	0.005
35	0.126	0.205	< 0.001	−0.00007	0.025	388	0.004
40	0.128	0.223	< 0.001	0.00043	0.028	263	0.002
45	0.129	0.246	< 0.001	0.00039	0.03	201	0.002
50	0.132	0.267	< 0.001	0.00034	0.031	185	0.002
55	0.136	0.275	< 0.001	0.00174	0.034	132	0.001
60	0.137	0.353	< 0.001	0.00061	0.041	74	0.001

The effect of latitude on diversification was assessed using FiSSE, a method for inferring the effects of a binary character on lineage diversification rates. Each row gives the results of FiSSE using a different threshold for classifying lineages as tropical and temperate. λ_0 and λ_1 denote estimated speciation rates (similar to λ_{DR}) for tropical and temperate lineages, respectively. *P* values indicate the proportion of simulations with a rate difference ($\lambda_1 - \lambda_0$) that is greater than the observed difference ($\Delta\mu_{null}$). The number of parsimony-reconstructed changes between states 0 and 1 is given by $n_{parsimony}$; *q* denotes the empirically estimated transition rate used to generate the null distribution. Results are based on 2,000 simulations; the observed difference in rates exceeded all simulated values, regardless of threshold.

nature research

Corresponding author(s): Daniel L. Rabosky

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Custom scripts written in the R programming language were used to extract and process data; all scripts are provided in the Dryad data package that accompanies our article.
Data analysis	Analysis: we used numerous statistical routines developed for the R programming language (many specific packages: ape, spdep, BAMMtools, and others), including many custom approaches for this study. Many standard phylogenetics software programs including RAxML, treePL, BAMM, PHLAWD, and PartitionFinder. All computer code and associated workflows are archived in the Dryad data repository (see Data and Code Availability statement).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data, scripts, and code necessary to repeat the analyses described here have been made available through the Dryad digital data repository (doi:10.5061/dryad.fc71cp4). Phylogenetic tree distributions are also available through http://fishtreeoflife.org. There are no restrictions on data availability.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We tested the relationship between assemblage-wide rates of speciation and latitude, and between species-specific rates of speciation and latitude. The study focused on marine fishes.
Research sample	The research sample included all species of marine fishes for which we could obtain DNA sequence data such that they could be included in a comprehensive phylogenetic tree of marine fishes. The sample was global in scope, and there were no data exclusions.
Sampling strategy	Sample size was determined by availability of genetic and geographic data for marine fishes. We used all available data in our study. The sample size used in our study exceeds the vast majority of published trait-dependent diversification studies. Moreover, the extent of speciation rate heterogeneity across fishes as inferred using BAMM is sufficiently great as to enable us to detect effects (see Rabosky and Huang, 2016, DOI:10.1093/sysbio/syv066 for relevant discussion of statistical power).
Data collection	DNA sequence data were extracted from public DNA sequence databases (GenBank) or sequenced de novo for this study. Spatial data were extracted from the AquaMaps spatial data repository. Ecological data for fishes (e.g., depth classification) were extracted from the publicly available FishBase database.
Timing and spatial scale	Spatial scale (scope) of the study is global. However, this question is generally not applicable to our study.
Data exclusions	No data were excluded.
Reproducibility	Our data are not experimental and experiments were thus not replicated. However, we used multiple distinct statistical tests (e.g., FiSSE, STRAPP, generalized linear models) and all approaches yielded concordant results. We also tested extensively for artifacts that might explain the pattern (see Extended Data 10).
Randomization	We did not perform an experiment and there was thus no group allocation. We used all species of marine fishes for which (1) DNA sequence data were available, and (2) spatial data were available. There was no further group partitioning of data beyond the natural groupings associated with geography and clade membership.
Blinding	Blinding was not relevant to our study, because all available data were used (our study did not perform an experiment).

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging