Progressing conservation of biodiversity through taxonomy, data publication and collaborative infrastructures

Keywords: species, database, marine, discovery, global.

ABSTRACT

Taxonomy is the foundation of biodiversity science through discovering its primary metric, namely 'species'. Globally, there has never been so many people involved in naming species new to science. The number of new marine (but not terrestrial) species described per decade has never been greater. Nevertheless, it is estimated that tens of thousands of marine, and hundreds of thousands of terrestrial, species remain to be discovered; many of which may already be in specimen collections. However, naming species is only a first step in documenting knowledge about their biology, biogeography and ecology. Considering the threats to biodiversity, the discovery of knowledge of existing and undescribed species is urgently required for their conservation. To accelerate this research we recommend, and cite examples of, increased communication: use of collaborative online databases; easier access to knowledge and specimens; production of taxonomic revisions and species identification guides; engagement of non-specialists; and international collaboration. The paradigm of 'data-sharing' should be abandoned in favour of mandating 'data publication' by the conservation biology community, including by peer-reviewers, editors, and journal and organisation policies. Examples of online data publication infrastructures illustrate gaps in sampling biodiversity and may also provide a common ground for long-term international collaboration between scientists and conservation initiatives.

INTRODUCTION

Parsons et al. (2014) listed 71 questions which if answered would bolster efforts to conserve marine biodiversity. One of their eight categories of questions addressed "scientific enterprise". We agree with the importance of their questions 69 to 71, and both respond and propose answers to them here.

69. How can taxonomic expertise be increased to reduce uncertainty in the conservation and management of marine ecosystems?

While we agree with the urgent need for more funding for taxonomy, there has not been a decline in taxonomic research as stated by Parsons et al. (2014). In fact, there has been an increase in publications in the field, and the number of authors of new species descriptions has increased seven-fold since the 1950's (Figure 1). This increase cannot be explained by the practice of naming more authors per species since the 1980's, and the relative proportions of the most and least productive authors has not changed over the last century (Costello et al. 2013a, b, 2014a, b). However, the number of authors from Asia and South America has been increasing more than those from other regions (Costello et al. 2013c). This is a narrow definition of a taxonomist and reviews of taxonomy have included people skilled in species identification (reviewed in Costello et al. 2013b). That the last decade saw more marine species named than any previous decade (Figure 1) (Appeltans et al. 2012) indicates that the field of taxonomy has never been so productive. Nevertheless, tens of thousands of species remain to be named. A review of 100 field studies of 33,000 marine species, and statistical modelling of 0.5 million species' rates of discovery both found one-third remained to be named (Appeltans et al. 2012, Costello et al. 2012). Recent reviews of marine fish, micro and macro-algae, sea anemones, and flowering plants, estimated that 61-77%, have been named (Eschmeyer et al. 2010, Guiry 2012, Fautin et al. 2013, De Clerck et al. 2013, Bebber et al. 2014, Costello et al. 2014a, b). Overall it appears that there are 2 to 3 million species on Earth, as suggested by May (1988) and Gaston (1991), but about one-third remain to be discovered. That over half of all species are known indicates that the species we know, at least within the better studied regions and habitats, may be a good indicator of biodiversity on Earth.

We agree with Parsons et al. (2014) that increased taxonomic effort is urgently required. This could be achieved through the following:

Communication: Increased communication and accessibility to knowledge, know-how, and publications is facilitated by email and online access to publications and authors contact details. This improves awareness of current knowledge and exchange of expertise that can lead to improved productivity.

Collaborative online infrastructures: The World Register of Marine Species (WoRMS) includes almost all marine species and a network of over 200 experts to marine species taxonomy (Costello et al. 2013d), and over 80,000 unique users per month. It is expanding to include more links to literature, species distribution, and other information about species. To date editors have published synthetic reviews of 15 taxa (Agatha 2011, Cairns 2011, Neiber et al. 2011, Rasmussen et al. 2011, Suárez-Morales 2011, Williams 2011, Blazewicz-Paszkowycz et al. 2012, Mah & Blake 2012, Poore & Bruce 2012, Stöhr et al. 2012, Van Soest et al. 2012, Williams & Boyko 2012, Eitel et al. 2013, Mapstone 2014). The database is centralised which aids standardisation and online publication, provides cost efficiencies, and has a permanent host institution. This model of structured building of taxonomic knowledge merits replication in other areas of taxonomy. It could also support registration of new species (e.g. Zoobank, http://zoobank.org).

Taxonomic revisions: Too many taxa still lack either global or regional reviews of existing knowledge and guides to discriminate their species (Costello et al. 2006, 2010, 2013c, 2014a). Such taxonomic revisions should resolve synonyms, identify early species' descriptions that were inadequate and thus species' names that are doubtful, and provide guides to species identification. Journal editors, referees and authors should support recognition of such publications by requiring authors to cite what guides they used to identify and name species, i.e. how they quality controlled their taxonomy (Costello & Wieczorek 2014). Funding agencies should announce calls for proposals to fund production of such taxonomic revisions. Employers should encourage and reward such benchmark publications by their scientists.

Access to knowledge: A major obstacle to engaging non-taxonomists, including conservation biologists, is the unavailability of taxonomic publications and species identification guides. Publications are increasingly easier to obtain by emailing experts, being published open-access, and through the Biodiversity Heritage Library. The shorter time to publication of descriptions of new species will reduce the likelihood of the same species being described by different authors at the same time. However, there is no online guide, or portal to guides, for the identification of all species or even higher taxa. Efforts to create such an online identification key to life are rudimentary, although there are some to marine species (Anon. 2014, Vanaverbeke et al. 2014).

Access to specimens: Undescribed marine species in collections of museums and other organisations may number 65,000 (Appeltans et al. 2012). However, too many collections still lack online registers of what taxa they may include. Access to this information would accelerate the planning of research to study these specimens and making best use of already archived specimens.

Engaging non-specialists: The fact that there has been more progress in taxonomy than may have been realised until recently does not mean enough has been done. It has taken over 250 years to get the most basic information, often only a species description, of about two-thirds of species on

Earth. The remaining number will be harder to discover as they are likely to be in rarely sampled locations, low in abundance, and/or difficult to discriminate from other species. Filling the remaining gaps can be more cost-efficiently achieved if non-specialists can recognise species, including unnamed species, and work with specialists to identify them (Costello et al. 2013b). People not employed by research organisations already provide a significant role in taxonomy (reviewed in Costello et al. 2013a, b, Brûlé & Touroult 2014).

International cooperation: Most research funding is still nationally based and many countries lack funding targeted at taxonomy. If countries cooperate by sharing taxonomic expertise, including access of non-nationals to sampling sites and specimens, this will provide cost-efficiencies. It is not realistic or necessary that every country has specialists in every taxon, especially when some taxa may be rare in their country.

70. How can scientific and management culture be changed to promote open sharing of data in formats that are accessible (and standardized)?

A first step is to stop using the term 'data sharing'. This implies some type of reciprocation, such as authorship on another paper or payment. This kind of data sharing requires potential users to know in advance if the data exists and then if it will be of use to their research. Instead, data should be published without any restrictions on its use just as with other kinds of publications. Such data publications should provide a conventional citation indicating the persons responsible (e.g. authors, editors), its content (i.e. title) and repository (e.g. using a DOI as used by the PANGAEA World Data Centre). When used, the dataset should be cited in the reference list as are other publications (Costello 2009, Costello et al. 2013e). When so many datasets are used that they cannot be accommodated in the main reference list of a paper, they can be cited in an Appendix.

Publication is a meritorious expectation of scientists whereas data sharing is not. In contrast to data sharing, it can include several levels of quality assurance, including peer-review (Costello et al. 2013e). New metrics for recognising scientific outputs include number of web views, downloads, and citations. All of these, plus data use, could be applied to published datasets using methods already implemented for papers.

Science journals already require genetic and other molecular data to be made publicly available upon publication of papers that used it. Taxonomic journals require type specimens of new species to be lodged in public specimen collections. An increasing number of journals require other kinds of data to be published once a paper has been published in their journal. These include *Nature*, *Science*, *Proceedings of the National Academy of Sciences of the United States of America*, and *Systematic Biology*, so the results of analyses can be verified and reproducible. Over 31 publishers of biology journals are members of Dryad (http://datadryad.org) which archives datasets. However, the journal *Conservation Biology* has no policy on data availability and *Biological Conservation* only encourages it. An overdue action to encourage data availability would be for conservation biologists, organisations and journals to make supporting data publication mandatory and to cite datasets in reference lists as they would other publications. A recent review by Wiley (the publisher of *Conservation Biology*) found that when journals made data publication mandatory it significantly increased data availability (Ferguson 2014). It is paradoxical that the conservation community recognises the need for more biodiversity data but has not taken the discipline to make what data already exists and has been used in journal publications freely available.

71. What strategies can be used to promote long-term integrated multidisciplinary collaborations?

The 'long-term' component of this question is the most challenging. Short term funding, conferences and workshop regularly foster collaboration. Similar strategies outlined above to improve taxonomic productivity could be applied to research to support other aspects of marine biodiversity conservation. We propose that the 'long-term' component of this research can best be served by publication of primary data in standardised open-access databases. These data are the empirical foundation of science. For marine biology, several standardised options for data publication and archiving are operational and can be expanded. For example, the World Register of Marine Species (WoRMS) is available for taxonomic and related biological data, and the Ocean Biogeographic Information System (OBIS) and Global Biodiversity Information Facility (GBIF) and associated databases for species distribution data (Boxshall et al. 2014, Costello & Wieczorek 2014). These initiatives provide a permanent scholarly standardised infrastructure. Hundreds of papers in science journals use this data every year (Costello et al. 2013d, 2013e). OBIS and GBIF include data at local to global spatial scales and time-series data; data from ecological and fishery surveys, citizen scientists and museum collections. Additional data fields and linking with other databases (e.g. WoRMS) may provide wider ecological (e.g. which species are Introductions), and environmental (e.g. AquaMaps, GMED) context (Kaschner et al. 2013, Basher et al. 2014). However, the data show notable spatial gaps, particularly when scrutinised at regional, local and temporal scales (Figure 2). These reflect the limited sampling in some geographic areas, including greater depths, and the need to publish historic data from the literature and specimen collections. These databases are now part of the international scientific infrastructure but are not yet within the mainstream of conservation science and management. In addition to their need for infrastructure support, these databases need mechanisms to ensure continued engagement of scientists in their development and quality assurance (Costello et al. 2014c). With such integration they can provide the pivot point for long-term international collaboration.

DISCUSSION

Despite the productivity and health of taxonomic research, it has never been so urgently needed because of the threat of species extinctions (Costello 2015). Conservation is compromised in the absence of information on what species exist, their ecology, biogeography and trends in abundance. The measures proposed here to accelerate taxonomic productivity are partly underway and demand more support from conservation scientists, managers, organisations, journals and their funding agencies. The publication of data in existing open access databases needs to become a mainstream activity that will provide the data necessary to inform conservation management and policy. A first step is for conservation biologists and organisations to require biodiversity data to be published and recognise this as of similar merit to other kinds of publications.

References

Agatha S 2011. Global diversity of aloricate Oligotrichea (Protista, Ciliophora, Spirotricha) in marine and brackish sea water. PLoS ONE 6(8): e22466. doi:10.1371/journal.pone.0022466.

Anon. 2014. Marine Species Identification Portal. Accessed at http://species-identification.org 11th July 2014.

Appeltans W, et al. 2012. The magnitude of global marine species diversity. *Current Biology* **22**, 1-14.

Basher, Z., Bowden, D. A. and Costello, M. J., 2014. Global Marine Environment Dataset (GMED). World Wide Web electronic publication. Version 1.0 (Rev.01.2014). Accessed at http://gmed.auckland.ac.nz on <Access DATE>

Bebber, D. P., Wood, J. R. I., Barker, C., and Scotland, R. W. 2014. Author inflation masks global capacity for species discovery in flowering plants. *New Phytologist*, **201**: 700–706.

Blazewicz-Paszkowycz M, Bamber R, and Anderson G 2012) Diversity of Tanaidacea (Crustacea: Peracarida) in the world's oceans – how far have we come? PLoS ONE 7(4): e33068. doi:10.1371/journal.pone.0033068.

Boxshall, G., et al. 2013. World Register of Marine Species. Accessed 2014-07-08 from http://www.marinespecies.org.

Brûlé S, and Touroult J 2014. Insects of French Guiana: a baseline for diversity and taxonomic effort. ZooKeys 434: 111–130. doi: 10.3897/zookeys.434.7582

Cairns SD 2011. Global diversity of the Stylasteridae (Cnidaria: Hydrozoa: Athecatae). PLoS ONE 6(7): e21670. doi:10.1371/journal.pone.0021670.

Costello M.J. 2009. Motivation of online data publication. *BioScience* **59**, 418-427.

Costello M.J. 2015. Biodiversity: the known, unknown and rates of extinction. *Curr. Biol*., in press.

Costello MJ, and Wieczorek J. 2014. Best practice for biodiversity data management and publication. *Biological Conservation* **173**, 68-73.

Costello, M.J., Emblow C.S., Bouchet P. and Legakis A. 2006. European marine biodiversity inventory and taxonomic resources: state of the art and gaps in knowledge. *Marine Ecology Progress Series* **316**, 257-268.

Costello MJ, Wilson SP, and Houlding B. 2012. Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Systematic Biology* **61**: 871-883.

Costello MJ, Coll M, Danovaro R, Halpin P, Ojaveer H, and Miloslavich P. 2010. A census of marine biodiversity knowledge, resources and future challenges. *PLoS ONE* (8): e12110.

Costello MJ. Wilson S, and Houlding B. 2013a. More taxonomists but a declining catch of species discovered per unit effort. *Systematic Biology* **62**, 616–624.

Costello MJ, May RM, and Stork NE 2013b. Can we name Earth’s species before they go extinct? *Science* **339**, 413-416.

Costello MJ, May RM, and Stork NE 2013c. Response to Comments on “Can we name Earth’s species before they go extinct?” *Science* **341**, 237.

Costello MJ, et al. 2013d. Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLoS ONE* 8(1): e51629.

Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, and Bourne P. 2013e. Data should be published, cited and peer-reviewed. *Trends in Ecology and Evolution* **28**, 454-461.

Costello MJ. Houlding B, and Joppa, L. 2014a. Further evidence of more taxonomists discovering new species, and that most species have been named: response to Bebber *et al.* (2014). *New Phytologist* **202**, 739–740.

Costello MJ., Houlding B., and Wilson S. 2014b. As in other taxa, relatively fewer beetles are being described by an increasing number of authors: Response to Löbl and Leschen. *Systematic Entomology* **39**, 395–399.

Costello MJ, et al. 2014c. Strategies for the sustainability of online open-access biodiversity databases. *Biological Conservation* **173**, 155-165.

De Clerck O., Guiry M.D., Leliaert F., Samyn Y. and Verbruggen H. 2013. Algal taxonomy: a road to nowhere? *Journal of Phycology* 49(2):215–225.

Eitel M, Osigus H-J, DeSalle R, Schierwater B 2013. Global Diversity of the Placozoa. PLoS ONE 8(4): e57131. doi:10.1371/journal.pone.0057131.

Eschmeyer, W.N., Fricke R., Fong J.D. and Polack D. 2010. Marine fish biodiversity: A history of knowledge and discovery (Pisces). *Zootaxa* 2525, 19-50.

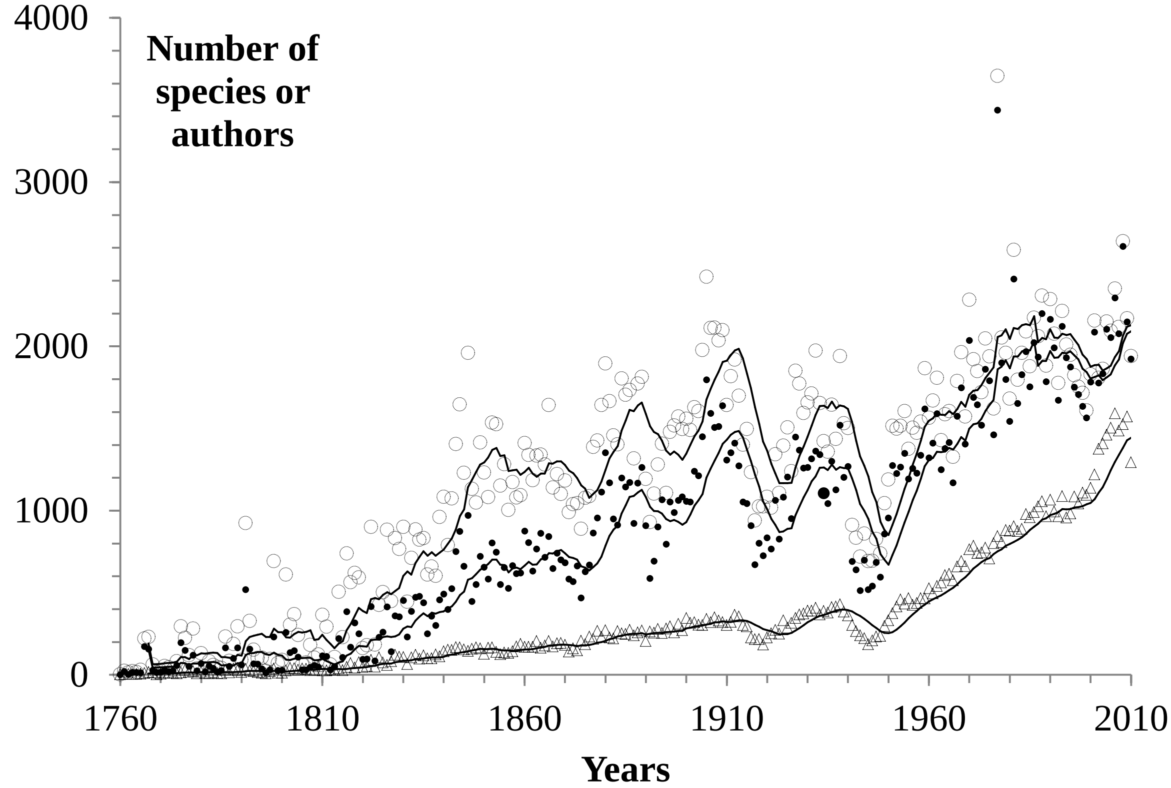
Fautin D.G., Malarky L., and Soberon J. 2013. Latitudinal diversity of sea anemones (Cnidaria: Actinaria). *Biological Bulletin* 224(2): 89–98.

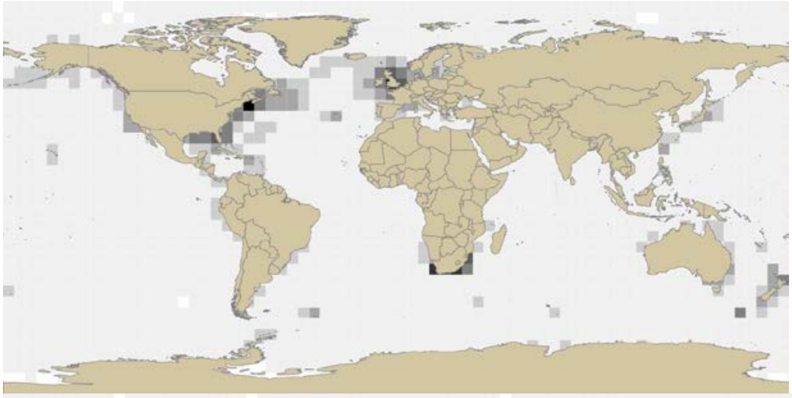
Ferguson L. 2014. How and why researchers share data (and why they don’t). Wiley Exchanges. Accessed at http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/ on 20 November 2014.

Gaston, K.J. 1991. The magnitude of global insect species richness. *Conservation Biology* **5**, 283-296.
Guiry M.D. 2012 How many species of algae are there? *Journal of Phycology* 48, 1057-1063.
Kaschner, K., J. Rius-Barile, K. Kesner-Reyes, C. Garilao, S.O. Kullander, T. Rees, and R. Froese. 2013. AquaMaps: Predicted range maps for aquatic species. World wide web electronic publication, www.aquamaps.org, Version 08/2013.
Parsons, E.C.M., et al. 2014. Seventy-one important questions for the conservation of marine biodiversity. *Conservation Biology*, in press.Mah CL and Blake DB 2012. Global diversity and phylogeny of the Asteroidea (Echinodermata). PLoS ONE 7(4): e35644. doi:10.1371/journal.pone.0035644.
Mapstone GM 2014. Global diversity and review of Siphonophorae (Cnidaria: Hydrozoa). PLoS ONE 9(2): e87737. doi:10.1371/journal.pone.0087737.
May, R.M. 1988. How many species are there on Earth? *Science* **241**, 1441-1449.
Neiber MT, et al. 2011. Global biodiversity and phylogenetic evaluation of Remipedia (Crustacea). PLoS ONE 6(5): e19627. doi:10.1371/journal.pone.0019627.
Poore GCB and Bruce NL 2012. Global diversity of marine isopods (except Asellota and Crustacean Symbionts). PLoS ONE 7(8): e43529. doi:10.1371/journal.pone.0043529.
Rasmussen AR, Murphy JC, Ompi M, Gibbons JW, Uetz P 2011. Marine reptiles. PLoS ONE 6(11): e27373. doi:10.1371/journal.pone.0027373.
Stöhr S, O'Hara TD and Thuy B 2012. Global diversity of brittle stars (Echinodermata: Ophiuroidea). PLoS ONE 7(3): e31940. doi:10.1371/journal.pone.0031940.
Suárez-Morales E 2011. Diversity of the Monstrilloida (Crustacea: Copepoda). PLoS ONE 6(8): e22915. doi:10.1371/journal.pone.0022915.
Vanaverbeke J, et al. 2014. NeMys: World database of free-living marine nematodes. Accessed at http://nemys.ugent.be on 2014-07-16.
Van Soest RWM, et al. 2012. Global diversity of sponges (Porifera). PLoS ONE 7(4): e35105. doi:10.1371/journal.pone.0035105.
Williams GC 2011. The global diversity of sea pens (Cnidaria: Octocorallia: Pennatulacea). PLoS ONE 6(7): e22747. doi:10.1371/journal.pone.0022747.
Williams JD and Boyko CB 2012. The global diversity of parasitic isopods associated with crustacean hosts (Isopoda: Bopyroidea and Cryptoniscoidea). PLoS ONE 7(4): e35350. doi:10.1371/journal.pone.0035350.

Figure 1. The number of nominal (hollow circle) and accepted (solid dot) marine species, and distinct author surnames per year until 2010. Lines are 10-year moving averages. The difference between nominal and accepted species names are largely synonymised species names. Data from WoRMS 11th July 2014.

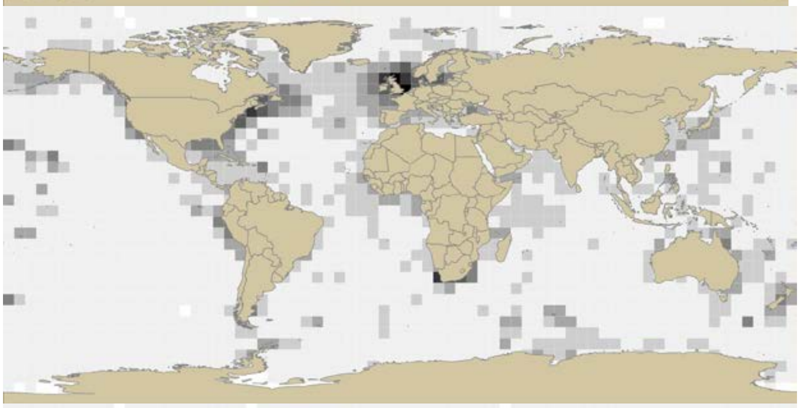
Figure 2. Maps of the number of (from top) (a) sampling dates (an indicator of time-series data), (b) species distribution records (indicates sampling effort), (c) species, (d) phyla, and (e) ES_{50} (estimated species from randomised samples of 50 records); in 5 by 5 degree latitude-longitude squares. Data from OBIS July 2014.


Number of species or authors
4000
3000
2000
1000
0
1760
1810
1860
1910
1960
2010
Years



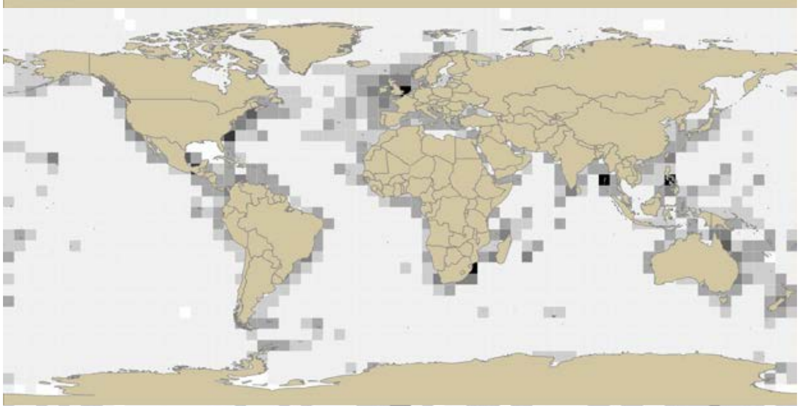
(a)
Number of sampling dates

- 1 - 565
- >565 - 2061
- >2061 - 5146
- >5146 - 8151
- >8151 - 9811
- >9811 - 11485
- >11485 - 11951



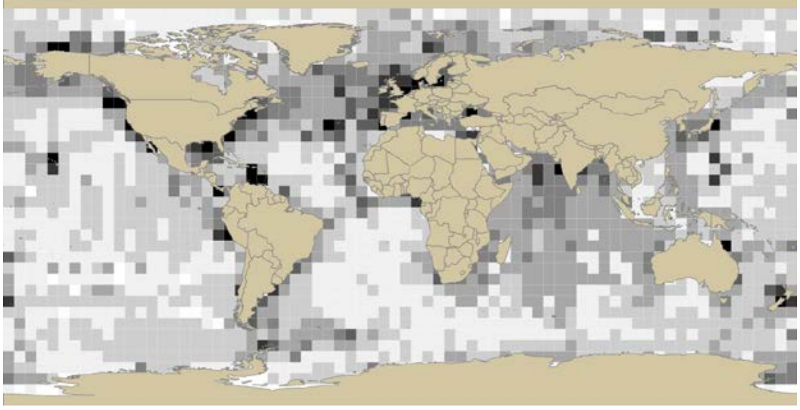
(b)
Number of species distribution records

- 1 - 4481
- >4481 - 24674
- >24674 - 80648
- >80648 - 340748
- >340748 - 691346
- >691346 - 1208396
- >1208396 - 1380945



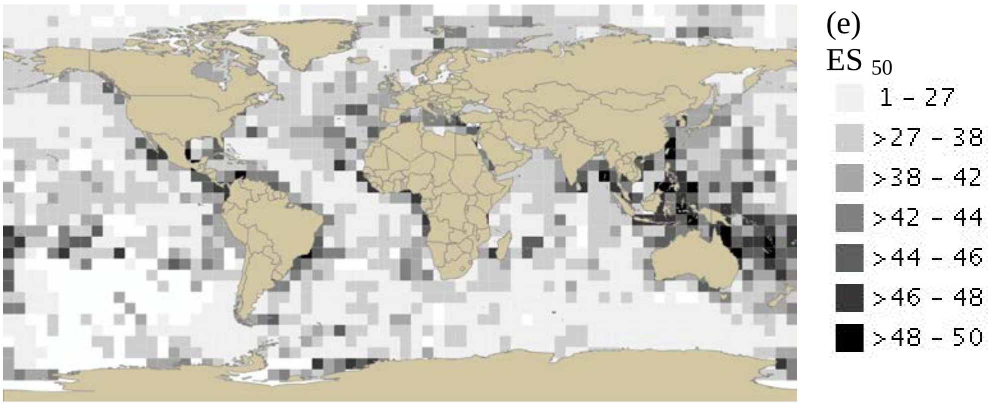
(c)
Number of species

- 1 - 435
- >435 - 832
- >832 - 2352
- >2352 - 4002
- >4002 - 4718
- >4718 - 5150
- >5150 - 5621



(d)
Number of Phyla

- 1 - 8
- >8 - 16
- >16 - 24
- >24 - 32
- >32 - 40
- >40 - 48
- >48 - 74


(e)
ES 50
1 – 27
>27 – 38
>38 – 42
>42 – 44
>44 – 46
>46 – 48
>48 – 50