Essay

Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures

Mark J. Costello,* § Bart Vanhoorne,† and Ward Appeltans‡

*Institute of Marine Science, University of Auckland, Post Bag 92019, Auckland 1142, New Zealand
†Flanders Marine Institute, Wandelaarkaai 7, Ostend 8400, Belgium
‡Ocean Biogeographic Information System, IODE, Intergovernmental Oceanographic Commission, UNESCO, Wandelaarkaai 7/61, Ostend 8400, Belgium

Abstract: *Taxonomy is the foundation of biodiversity science because it furthers discovery of new species. Globally, there have never been so many people involved in naming species new to science. The number of new marine species described per decade has never been greater. Nevertheless, it is estimated that tens of thousands of marine species, and hundreds of thousands of terrestrial species, are yet 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, new knowledge of existing species and discovery of undescribed species and their subsequent study are urgently required. To accelerate this research, we recommend, and cite examples of, more and better communication: use of collaborative online databases; easier access to knowledge and specimens; production of taxonomic revisions and species identification guides; engagement of nonspecialists; and international collaboration. "Data-sharing" should be abandoned in favor of mandated data publication by the conservation science community. Such a step requires support from peer reviewers, editors, journals, and conservation organizations. Online data publication infrastructures (e.g., Global Biodiversity Information Facility, Ocean Biogeographic Information System) illustrate gaps in biodiversity sampling and may provide common ground for long-term international collaboration between scientists and conservation organizations.*

Keywords: data access, online databases, species discovery, taxonomic revision

La Conservación de la Biodiversidad por medio de la Taxonomía, la Publicación de Datos y las Infraestructuras Colaborativas

Resumen: *La taxonomía es el fundamento de la ciencia de la biodiversidad ya que impulsa el descubrimiento de nuevas especies. Nunca ha habido tantas personas involucradas a nivel global en el nombramiento de especies nuevas para la ciencia. El número de especies marinas descritas por década nunca ha sido mayor. Sin embargo, se estima que decenas de miles de especies marinas y cientos de miles de especies terrestres no han sido descubiertas aún; muchas de las cuales podrían ser ya especímenes en colecciones. A pesar de esto, nombrar a las especies es sólo un primer paso en la documentación del conocimiento sobre su biología, biogeografía y ecología. Al considerar a las amenazas para la biodiversidad, se requiere urgentemente del conocimiento nuevo de las especies existentes y del descubrimiento de especies no descritas y su estudio subsecuente. Para acelerar esta investigación recomendamos, y citamos ejemplos de, una mayor y mejor comunicación: el uso de bases de datos colaborativas en línea, acceso más fácil al conocimiento y a los especímenes, producción de revisiones taxonómicas y guías de identificación de especies, participación de los no-especialistas, y colaboración internacional. Los "datos compartidos" deberían abandonarse en favor de la publicación de datos por encargo de la comunidad de científicos de la conservación. Un paso así requiere de apoyo por parte de los colegas revisores, editores, revistas y las organizaciones de conservación. Las infraestructuras de publicación de datos en línea (p. ej.: Global Biodiversity Information Facility, Ocean Biogeographic Information System) muestran los vacíos en el muestreo de biodiversidad y pueden proporcionar*

§Address for correspondence: Institute of Marine Science, University of Auckland, Post Bag 92019, Auckland 1142, New Zealand. email m.costello@auckland.ac.nz
Paper submitted December 23, 2014; revised manuscript accepted July 24, 2014.

DOI: 10.1111/cobi.12496

afinidades comunes para la colaboración internacional a largo plazo entre los científicos y las organizaciones de conservación.

Palabras Clave: acceso a datos, bases de datos en línea, descubrimiento de especies, revisión taxonómica

Introduction

Parsons et al. (2014) list 71 questions, which if answered could bolster efforts to conserve marine biodiversity. One of their 8 categories of questions addresses "scientific enterprise." We agree with the importance of 3 of their questions and respond and propose answers to them.

Increasing Taxonomic Expertise

Parsons et al. (2014) asked, "How can taxonomic expertise be increased to reduce uncertainty in the conservation and management of marine ecosystems?" Although we agree with these authors about the urgent need for more funding for taxonomy, there has not been a decline in taxonomic research as they state. In fact, there has been an increase in publications in the field, and the number of authors of new species descriptions has increased 7-fold since the 1950s (Fig. 1). This increase cannot be explained by the practice of naming more authors per species since the 1980s, and the relative proportions of the most and least productive authors has not changed over the last century (Costello et al. 2013*a*, 2013*b*, 2014*a*, 2014*b*). However, the number of authors from Asia and South America has increased more than the number of authors from other regions (Costello et al. 2013*c*). The number of authors is a narrow definition of a taxonomist, and reviews of taxonomy have included additional people skilled in species identification (reviewed in Costello et al. 2013*b*).

That the last decade saw more marine species named than any previous decade (Fig. 1) (Appeltans et al. 2012) indicates that the field of taxonomy has never been so productive. Nevertheless, tens of thousands of species remain unnamed. A review of 100-field studies of 33,000 marine species and statistical modeling of rates of discovery of 0.5 million species show that one-third of species are yet to be named (Appeltans et al. 2012; Costello et al. 2012). Recent reviews of marine fish, micro- and macroalgae, sea anemones, and flowering plants estimate that 61–77% have been named (Eschmeyer et al. 2010; Guiry 2012; De Clerck et al. 2013; Fautin et al. 2013; Bebber et al. 2014; Costello et al. 2014*a*, 2014*b*). Overall, it appears there are 2–3 million species on Earth, as suggested by May (1988) and Gaston (1991), but about one-third are undiscovered. That over half of all species are known indicates that the species we know, at least within the better studied places, may be good indicators of biodiversity on Earth.

We agree with Parsons et al. (2014) that increased taxonomic effort is urgently required. This could be achieved through improved communication, collaborative online infrastructures, taxonomic revisions, improved access to knowledge, improved access to specimens, engagement with nonspecialists, and international collaboration.

Increased communication and accessibility to knowledge, know-how, and publications are facilitated by email and online access to publications and author contact information. This improves awareness of current knowledge and increases exchange of expertise, both of which can lead to improved productivity.

The World Register of Marine Species (WoRMS) includes almost all marine species and access to a network of over 200 experts in marine species taxonomy (Costello et al. 2013*d*). There are over 80,000 unique users per month. The register is expanding to include more links to literature, data on species distribution, and other information about species. To date editors have published synthetic reviews of 15 taxa (e.g., Agatha 2011; Van Soest et al. 2012; Williams & Boyko 2012; Eitel et al. 2013; Mapstone 2014). The database is centralized, which aids standardization and online publication and provides cost efficiencies and a permanent host institution. This model of structured building of taxonomic knowledge merits replication in other areas of taxonomy

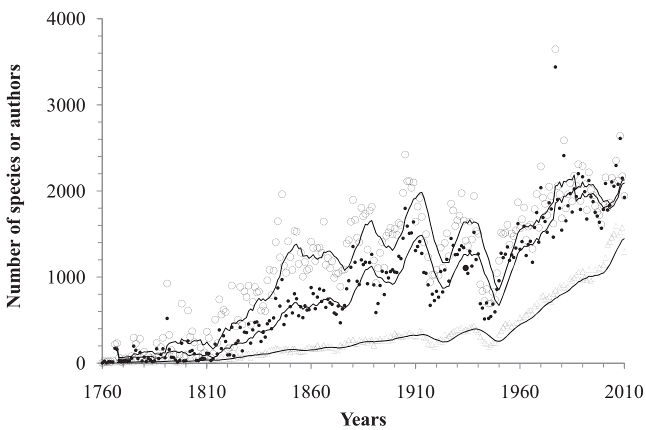

Figure 1. The number of nominal (open circle) and accepted (solid circle) marine species and distinct author surnames (lower line and triangles) per year until 2010. Lines are 10-year moving averages. The difference between nominal and accepted species names are largely synonymized species names. Data from WoRMS (11 July 2014).

and could support registration of new species in ZooBank (http://zoobank.org).

Too many taxa still lack either global or regional reviews of existing knowledge and guides to discriminate among species (Costello et al. 2006, 2010, 2013*c*, 2014*a*). 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, reviewers, 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 controlled the quality of their taxonomy) (Costello & Wieczorek 2014). Funding agencies should fund production of such taxonomic revisions, and employers should encourage and reward such benchmark publications by their scientists.

A major obstacle to engaging nontaxonomists, including conservation biologists, is the unavailability of taxonomic publications and species identification guides. Publications are increasingly easier to obtain by emailing experts, through open-access publication, and through the Biodiversity Heritage Library. A shorter time to publication of descriptions of new species would reduce the likelihood of the same species being described by different authors at the same time. However, there are no online guides, 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 (Anonymous 2014; Vanaverbeke et al. 2014).

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

That there has been more progress in taxonomy than may have been realized 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 species will be difficult to discover because they are likely to be in rarely sampled locations, low in abundance, or difficult to discriminate from other species. Filling the remaining gaps can be more cost-efficiently achieved if nonspecialists can recognize species, including unnamed species, and work with specialists to identify them (Costello et al. 2013*b*). People not employed by research organizations already play a significant role in taxonomy (reviewed in Costello et al. 2013*a*, 2013*b*; Brûlé & Touroult 2014).

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.

Changing Science and Management Cultures

Parsons et al. (2014) asked, "How can scientific and management culture be changed to promote open sharing of data in formats that are accessible (and standardized)?" A first step, in changing the culture is to stop using the term *data sharing*. This term 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 exist and then if it will be of use to their research. Instead, data should be published without restrictions on its re-use in research and education, just as with other kinds of publications. Published data sets should have a conventional citation style that indicates the persons responsible (e.g., authors, editors), its content (i.e., title), and a permanent web address for its repository (e.g., a DOI as used by the PANGAEA World Data Centre). When used, the data set should be cited in the reference list, as are other publications (Costello 2009; Costello et al. 2013*e*). When so many data sets are used that they cannot be accommodated in the main reference list of a paper, they can be cited in the appendix.

Publication is a meritorious endeavor of scientists, whereas data sharing is not. In contrast to data sharing, publication of data could include several levels of quality assurance, including peer review (Costello et al. 2013*e*). New metrics for recognizing scientific outputs include number of web views, downloads, and citations. All of these metrics, plus data use, could be applied to published data sets through the use of methods already implemented for scientific articles.

Science journals already require genetic and other molecular data used in a published article to be made publicly available. Taxonomic journals require type specimens of new species to be lodged in public specimen collections. So that results can be verified and studies reproduced, an increasing number of journals require other kinds of data to be made available when an article is published (e.g., *Nature, Science, Proceedings of the National Academy of Sciences of the United States of America*, and *Systematic Biology*). Over 31 publishers of biology journals are members of Dryad (http://datadryad.org), which archives data sets. However, *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, organizations, and journals to make supporting data publication mandatory and to cite data sets in reference lists as they would other

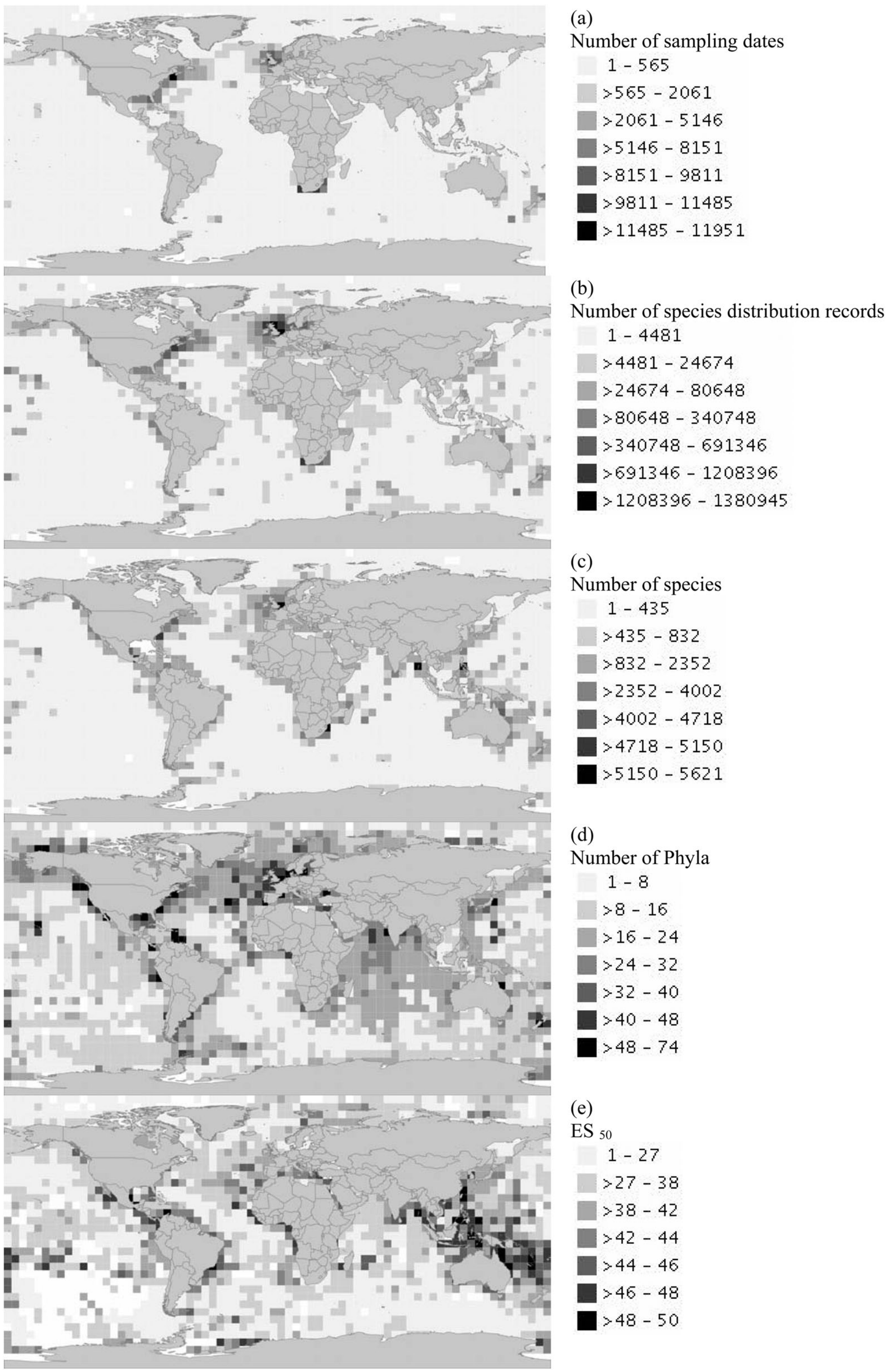

Figure 2. Number of (a) species' sampling dates (an indicator of time-series data), (b) species' distribution records (indicates sampling effort), (c) species, and (d) phyla and (e) ES_{50} (estimated species from randomized samples of 50 records) in 5 × 5 degree latitude-longitude squares. Data from Ocean Biogeographic Information System (July 2014).

publications. A recent review by Wiley (the publisher of *Conservation Biology*) found that when journals make data publication mandatory, data availability increases significantly (Ferguson 2014). It is paradoxical that the conservation community recognizes the need for more biodiversity data but has not been sufficiently disciplined to make what data already exist and have been used in publications freely available.

Strategies to Promote Collaboration

The final question of Parsons et al. (2014) we address is, "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 workshops 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 standardized open-access databases. These data are the empirical foundation of science. For marine biology, several standardized options for data publication and archiving are operational and can be expanded. For example, the 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 are available for species distribution data (Boxshall et al. 2014; Costello & Wieczorek 2014). These initiatives provide a permanent scholarly and standardized infrastructure. Every year hundreds of articles in science journals are based on these data (Costello et al. 2013*d*, 2013*e*). The OBIS and GBIF include data at local to global spatial scales, time-series data, and 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 have notable spatial gaps, particularly relative to regional, local, and temporal scales (Fig. 2). These gaps reflect 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. 2014*c*). 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 by the absence of information on what species exist, their ecology, biogeography, and trends in abundance. The measures we propose to accelerate taxonomic productivity are partly underway and demand more support from conservation scientists, managers, organizations, journals, and 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 and policy. A first step is for conservation biologists and organizations to require biodiversity data (at least that used in research publications) to be published and to recognize this as of similar merit to other kinds of publications.

Acknowledgment

The authors thank L. Vandepitte and M. Fornwall for helpful comments on this paper.

Literature Cited

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

Anonymous. 2014. Marine species identification portal. Available from http://species-identification.org (accessed July 2014).

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

Basher Z, Bowden DA, Costello MJ. 2014. Global marine environment dataset (GMED). Version 1.0 (Rev.01.2014). Available from http://gmed.auckland.ac.nz (accessed 20 October 2014).

Bebber DP, Wood JRI, Barker C, Scotland RW. 2014. Author inflation masks global capacity for species discovery in flowering plants. New Phytologist **201:**700–706.

Boxshall G, et al. 2014. World register of marine species. Available from http://www.marinespecies.org (accessed July 2014).

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

Costello MJ. 2009. Motivation of online data publication. BioScience **59:**418–427.

Costello MJ. 2015. Biodiversity: the known, unknown and rates of extinction. Current Biology: in press.

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

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

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

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

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

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

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

Costello MJ, Houlding B, Joppa L. 2014*a*. 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, Wilson S. 2014*b*. 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. 2014*c*. Strategies for the sustainability of online open-access biodiversity databases. Biological Conservation **173:**155–165.

De Clerck O, Guiry MD, Leliaert F, Samyn Y, 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:**e57131. DOI:10.1371/journal.pone.0057131.

Eschmeyer WN, Fricke R, Fong JD, Polack D. 2010. Marine fish biodiversity: a history of knowledge and discovery (*Pisces*). Zootaxa **2525:**19–50.

Fautin DG, Malarky L, 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. Available from http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/ (accessed November 2014).

Gaston KJ. 1991. The magnitude of global insect species richness. Conservation Biology **5:**283–296.

Guiry MD. 2012. How many species of algae are there? Journal of Phycology **48:**1057–1063.

Kaschner K, Rius-Barile J, Kesner-Reyes K, Garilao C, Kullander SO, Rees T, Froese R. 2013. AquaMaps: predicted range maps for aquatic species. Version 08/2013. Available from www.aquamaps.org (accessed December 2014).

Mapstone GM. 2014. Global diversity and review of Siphonophorae (Cnidaria: Hydrozoa). PLOS ONE **9:**e87737. DOI:10.1371/journal.pone.0087737.

May RM. 1988. How many species are there on Earth? Science **241:**1441–1449.

Parsons ECM, et al. 2014. Seventy-one important questions for the conservation of marine biodiversity. Conservation Biology **28:**1206–1214.

Vanaverbeke J, et al. 2014. NeMys: World database of free-living marine nematodes. Available from http://nemys.ugent.be/aphia.php?p=idkeys (accessed 20 March 2015).

Van Soest RWM, et al. 2012. Global diversity of sponges (Porifera). PLOS ONE **7:**e35105. DOI:10.1371/journal.pone.0035105.

Williams JD, Boyko CB. 2012. The global diversity of parasitic isopods associated with crustacean hosts (Isopoda: Bopyroidea and Cryptoniscoidea). PLOS ONE **7:**e35350. DOI:10.1371/journal.pone.0035350.