

Are marine biodiversity hotspots still blackspots for barcoding?

Authors: Francesco Mugnai¹, Emese MeglécZ², CoMBoMed group^{**}, Federica Costantini^{1,3}

****** Marco Abbiati^{4,3}, Giorgio Bavestrello^{5,3}, Fabio Bertasi⁶, Marzia Bo^{5,3}, María Capa⁷, Anne Chenuil⁸, Marina Antonia Colangelo^{1,3}, Olivier De Clerck⁸, José Miguel Gutiérrez⁹, Loretta Lattanzi⁶, Michèle Leduc¹⁰, Daniel Martin¹¹, Kenan Oguz Matterson¹, Barbara Mikac¹, Laetitia Plaisance^{12,13}, Massimo Ponti^{1,3}, Ana Riesgo¹⁴, Vincent Rossi¹⁵, Eva Turicchia^{3,4,16}, Andrea Waeschenbach¹⁷, Owen S Wangensteen¹⁸

Affiliation:

¹ University of Bologna, Dipartimento Di Scienze Biologiche, Geologiche e Ambientali, Ravenna, Italy

² Aix Marseille Univ, Avignon Université, CNRS, IRD, IMBE, Marseille, France

³ Consorzio Nazionale Interuniversitario per le Scienze del Mare (CoNISMa), Rome, Italy

⁴ University of Bologna, Dipartimento di Beni Culturali, Ravenna, Italy

⁵ Università degli Studi di Genova, Dipartimento di Scienze della Terra, dell'Ambiente e della Vita, Italy

⁶ ISPRA Istituto per la Protezione e la Ricerca Ambientale, Roma, Italy

⁷ Universitat de les Illes Balears, Dpto Biologia, Palma, Spain

⁸ Phycology Research Group, Biology Department, University of Ghent, Ghent, Belgium

⁹ TAXON Estudios Ambientales S.L., Alcantarilla (Murcia), Spain

¹⁰ STARESO marine station, Corse, France

¹¹ Centre d'Estudis Avançats de Blanes (CEAB - CSIC), Blanes, Spain

¹² Laboratoire Evolution et Diversité Biologique, CNRS/UPS, Toulouse, France

¹³ National Museum of Natural History, Smithsonian Institution, Washington, D.C., USA

¹⁴ Department of Life Sciences, Natural History Museum, Cromwell Road, SW7 5BD, London, UK

¹⁵ Aix Marseille Univ., Université de Toulon, CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France

¹⁶ CIRI FRAME, Università di Bologna, Ravenna, Italy

¹⁷ School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, UK

¹⁸ Norwegian College of Fishery Science, UiT the Arctic University of Norway

Keywords: metabarcoding, biological diversity, genetic diversity, marine animals, cryptic species, integrative taxonomy

Article impact statement: Lack of COI sequences of marine species in public databases hamper comprehensive biodiversity census for effective conservation strategies.

Acknowledgements: This paper is a result of the CoMBoMed workshop (Census of Marine Mediterranean Benthic Biodiversity: an integrative metabarcoding approach), held in Ravenna (Italy) on February 11-14, 2019 and organized by the Department of Biological, Geological and Environmental Sciences (BiGeA) of the University of Bologna (UniBo). The CoMBoMed initiative was supported by the European Marine Research Network (EUROMARINE Network), the Inter-Departmental Research Centre for Environmental Sciences (CIRSA – UniBo), the Cultural Heritage Department (DBC - UniBo, <https://beniculturali.unibo.it/it>), the Fondazione Flaminia and the ERANet Mar-Tera Project SEAMoBB (Solutions for sEmi-Automated Monitoring of Benthic Biodiversity).

Abstract:

Marine biodiversity underpins ecosystem health and societal well-being. Preservation of biodiversity hotspots is a global challenge. Molecular tools, like DNA barcoding and metabarcoding, hold great potential for biodiversity monitoring, possibly outperforming more traditional taxonomic methods. However, metabarcoding-based biodiversity assessments are limited by the availability of sequences in barcoding reference databases; a lack thereof results in high percentages of unassigned sequences. In this study we (i) present the current status of known vs. barcoded marine species at a global scale based on online taxonomic and genetic databases; and (ii) compare the current status with data from ten years ago. Then we analyzed occurrence data of marine animal species from five Large Marine Ecosystems (LMEs) classified as biodiversity hotspots, to identify any consistent disparities in COI barcoding coverage between geographic regions and at phylum level. Barcoding coverage varied among LMEs (from 36.8% to 62.4% COI-barcoded species) and phyla (from 4.8% to 74.7% COI-barcoded species), with Porifera, Bryozoa and Platyhelminthes being highly underrepresented, compared to Chordata, Arthropoda and Mollusca. We demonstrate that although barcoded marine species increased from 9.5% to 14.2% since the last assessment in 2011, about 15,000 (corresponding to 7.8% increase) new species were described from 2011 to 2021. The next ten years will thus be crucial to enroll concrete collaborative measures and long term initiatives (e.g., Horizon 2030, Ocean Decade) to populate barcoding libraries for the marine realm.

1 Healthy and well-functioning ecosystems are crucial for providing essential goods and services (e.g., primary
2 production, climate regulation, CO₂ sequestration). A critical ecohealth indicator (*sensu* Charron 2012) is biodiversity
3 (Myers et al. 2000) as its loss disrupts ecosystem processes and services. Unfortunately, global biodiversity is predicted
4 to decline over the 21st century due to anthropogenic stressors such as commercial species overexploitation, habitat
5 artificialization and destruction, eutrophication, pollution, and introduction of non-indigenous species (NIS) (Pereira et
6 al. 2010). These pressures intensify the effects of climate change by accentuating associated alterations in oceanic
7 biogeochemistry (Harley 2011) and cause local and global biodiversity losses and spatial reshuffling of species (Pecl et
8 al. 2017; Blowes et al. 2019). Marine biodiversity loss and redistribution are recognized as planetary challenges (Worm
9 et al. 2006; Pimm et al. 2014), but the effective implementation of conservation initiatives depends on having an accurate
10 knowledge of species diversity and their geographic distributions. Therefore, comprehensive biodiversity censuses,
11 particularly of vulnerable marine ecosystems, are imperative to inform policy-making.

12 Biodiversity studies typically focus on charismatic or commercially-valuable species (Troudet et al. 2017) or
13 well-known biodiversity hotspots (Myers et al. 2000; Jenkins and Van Houtan 2016), where high species richness,
14 endemism and human pressures coincide, leaving many oceanic regions largely unexplored (Snelgrove 2016).
15 While there is a great need to characterize marine biodiversity more widely, few resources are dedicated to taxonomic
16 research and training, which are fundamental for identifying and describing new species (de Carvalho et al. 2007; Boero
17 2010), particularly of small and cryptic organisms. Furthermore, morphology based biodiversity assessments are time
18 consuming and require highly specialized taxonomic knowledge, which is fast disappearing in the absence of funding.
19 Although genetic characterization (barcoding) of species is highly effective in identifying known taxonomic entities
20 (DeSalle and Goldstein 2019), an entirely molecular approach applied to lesser known taxa or geographic regions goes
21 little beyond identifying molecular operational taxonomic units (MOTUs). Thus, an integrative approach (combining
22 barcoding and morphology) is needed for reliable species identification and high-quality biodiversity assessments.

23 The rapid generation of high volumes of data resulting from the recent advances in “Omics” technologies and
24 high-throughput sequencing methods poses great opportunities for metabarcoding, metagenomics and
25 metatranscriptomics studies (Raupach et al. 2016; Wangensteen et al. 2018; Holman et al. 2019). Yet, substantial gaps
26 remain in public DNA barcoding databases, leading to a high proportion of unassigned sequences in metabarcoding-based
27 studies (e.g., Martin et al. 2021), and substantial uncertainty in the estimated number of species living in a surveyed area
28 (Valentini et al. 2016). Furthermore, many national and regional initiatives aimed at expanding DNA barcode databases
29 are not providing complete species documentation, including morphology, ecology, and biology. Moreover, databases
30 are usually not fully curated and harmonized by the time funds are exhausted after project completions (Collins et al.
31 2020). Consequently, some databases may have limited accessibility or are taken offline only a few years after their

32 publication, and the obtained sequences are not always uploaded on public barcoding databases (see Online Resource S1,
33 based on Radulovici et al. 2010 and Trivedi et al. 2016). In addition, most DNA barcoding initiatives were, and still are,
34 focusing on terrestrial or freshwater environments (see Online Resource S1) or on marine organisms that can serve as
35 bioindicators of environmental impacts (Weigand et al., 2019). However, more substantial contributions are needed to
36 improve the genetic knowledge for underrepresented marine taxa at a global scale. Previous attempts of quantifying the
37 proportion of barcoded species across all marine animal species (Bucklin et al. 2011; Bergsten et al. 2012; Vargas et al.
38 2012; Ratnasingham and Hebert 2013; Aylagas et al. 2018; Weigand et al. 2019; Ramirez et al. 2020) found that only
39 9.5% of the 192,702 described species had been assigned a molecular cytochrome *c* oxidase subunit I (COI) barcode
40 (Bucklin et al. 2011).

41 The number of marine species discovered has significantly progressed on a global scale during the last decade
42 (Mora et al. 2011; Costello et al. 2012; Appeltans et al. 2012). Currently, around 238,000 marine species have been
43 described (Horton et al. 2021), with about 2,000 new species added each year (Costello et al. 2012). Here, we provide an
44 updated estimate of COI-barcoded marine animal species at a global scale. Moreover, we selected five Large Marine
45 Ecosystems (LMEs, <https://www.lmehub.net>) considered as biodiversity hotspots (Mediterranean Sea, Caribbean Sea,
46 North Sea, Indonesian Sea, and Red Sea) to detect possible disparities in COI barcoding coverage at regional scale and
47 to underline where more efforts should be enrolled from now on.

48 Data for all marine animal species worldwide were gathered from the World Register of Marine Species
49 (WoRMS, Horton et al. 2021), while their presence within the five LMEs was retrieved from the Ocean Biodiversity
50 Information System (OBIS, Intergovernmental Oceanographic Commission of UNESCO; <http://www.iobis.org> accessed
51 on May 2021). Correspondent availability of COI barcode records at species level were retrieved from the National Center
52 for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov>) and Barcode of Life Data System (BOLD,
53 <http://www.boldsystems.org>) repositories using custom Perl scripts (<https://osf.io/qsn5e/>).

54 Globally, from 207,686 known marine animal species, only 14.5% (NCBI) and 13.8% (BOLD) possessed COI
55 barcodes. This represents 4.3% and 5% increases (NCBI and BOLD respectively) (see Online Resource S2) since the
56 estimate by Bucklin et al. (2011). The five LMEs encompass a total of 34,286 nominal marine animal species, with 14,472
57 (42.2%, NCBI) and 14,351 (41.9%, BOLD) having at least one COI sequence (Fig. 1). Evidently, the proportion of
58 barcoded species in these hotspots largely exceeds that at global level. Nevertheless, the current LMEs barcoding data
59 are still not sufficient for comprehensive biodiversity assessments. Furthermore, this barcoding coverage and species
60 richness varies across LMEs (Fig. 1). The North and Indonesian Seas have comparable species richness, but the number
61 of barcoded species is much higher (+15.8% for both NCBI and BOLD) in the latter. Variable research intensity and
62 disparity in sampling efforts and access to marine biological collections may explain the differences in number and

63 relative occurrences (i.e., frequency) of detected species, as well as the variable barcoding coverage among the LMEs
64 (Collins et al. 2020) (Fig. 2A). In particular, the North Sea is the most intensively studied LME for which species
65 occurrence records started in 1753, while research efforts for the other four LMEs started as late as the 1970s. Moreover,
66 the magnitude of occurrence records does not follow that of species richness. The North Sea has a species richness
67 comparable to the Indonesian Sea (8255 species), but it is considerably lower than that of the Mediterranean (9977
68 species) and Caribbean (12161 species) Seas. However, it has several millions more species occurrences (9.57 millions)
69 than the other LMEs (Fig. 2B).

70 Regarding taxon-specific statistics, the phyla with the highest numbers of nominal and COI-sequenced species
71 in the five LMEs were Chordata, Arthropoda and Mollusca, followed by Annelida and Cnidaria (Fig. 3) (see Online
72 Resource S3 for LME-specific data). COI barcoding coverage for Porifera, Bryozoa, and Platyhelminthes was very low,
73 which is likely due to (1) high crypticity (Trontelj and Fišer 2009; Fehlaue-Ale et al. 2014), (2) variability in the annealing
74 position of traditional COI primers leading to failed PCR amplification, (3) co-amplification of epibionts and
75 endosymbionts (Vargas et al. 2012). Chordata clearly shows the highest number of nominal species and barcoding
76 coverage (68.6% NCBI, 71.5% BOLD). Most of these contributions originate from Vertebrata (70.1% NCBI, 73.1%
77 BOLD), whereas non-vertebrate Chordata (Ascidians) are highly underrepresented (36.3% NCBI, 35.0% BOLD) (see
78 Online Resources S4).

79 Our results evidence a global lack of COI barcode availability for a large proportion of marine invertebrates,
80 with the North, Caribbean, and Mediterranean Seas being highly underrepresented in barcode databases, thus requiring
81 further efforts. The ongoing EU's key funding programme for research and innovation "Horizon 2030" (2021-2027)
82 allocates 8,953 Bi€ for Pillar II, Cluster 6 "Food, Bioeconomy, Natural Resources, Agriculture & Environment", offering
83 co-funded initiatives such as "Rescuing Biodiversity to Safeguard Life on Earth" (https://ec.europa.eu/info/horizon-europe_en) that could be pivotal for increasing the biodiversity knowledge at European level, including the marine realm.

84 To bridge the gaps in biodiversity assessment and monitoring worldwide, national and international
85 collaborations among multidisciplinary teams (including taxonomists, molecular biologists, bioinformaticians,
86 evolutionary biologists, ecologists, and environmental scientists) must be fostered. Such an integrated approach enables
87 physical specimens, accessioned in permanent natural history museum collections, to be linked to molecular,
88 morphological and ecological data, creating meaningful and long-lived reference data repositories. These efforts need to
89 be accompanied by creating an interactive virtual platform able to interface with different database sources, showing up-
90 to-date species-specific ecological, morphological and molecular data. Configuring these public repository following a
91 top-down layout linking other subset databases based on FAIR principles (Findable, Accessible, Interoperable and
92 Reusable, <http://wilkinsonlab.info/node/FAIR>) is mandatory to allow efficient data sharing.

94 European initiatives such as DNAquaNet (<https://dnaqua.net>), MBON (<https://marinebon.org>), SEAMoBB
95 (<https://seamobb.osupytheas.fr>) are working to reduce morpho-genetic gaps and to expand our knowledge on marine
96 biodiversity. Such initiatives should be further promoted worldwide. The BIOSCAN initiative (2019-2027,
97 <https://ibol.org/programs/bioscan/>) from the iBOL Consortium aimed at "Revealing species, their interactions, and
98 dynamics" is expected to promote new species descriptions and molecular characterization of terrestrial, freshwater, and
99 marine species. Last but not least, in the present era of unprecedented marine biodiversity loss, society as a whole must
100 be enthused about marine life in view to support measures that preserve marine biodiversity at a global scale. To attain
101 this goal, citizen science initiatives (e.g., marine BioBlitz - <https://www.abol.ac.at/en/abol-bioblitz-2020/>, LifeWatch -
102 <https://www.lifewatch.eu>, educational field activities etc.) will be pivotal, as emphasized by the UN Decade of Ocean
103 Science for Sustainable Development (<https://www.oceandecade.org>).

Literature cited:

- 104 Appeltans W, Ah Yong ST, Anderson G, et al (2012) The magnitude of global marine species diversity. *Curr Biol*
105 22:2189–2202. <https://doi.org/10.1016/j.cub.2012.09.036>
- 106 Aylagas E, Borja Á, Muxika I, Rodríguez-Ezpeleta N (2018) Adapting metabarcoding-based benthic biomonitoring into
107 routine marine ecological status assessment networks. *Ecol Indic* 95:194–202.
108 <https://doi.org/10.1016/j.ecolind.2018.07.044>
- 109 Bergsten J, Bilton DT, Fujisawa T, et al (2012) The effect of geographical scale of sampling on DNA barcoding. *Syst*
110 *Biol* 61:851–869. <https://doi.org/10.1093/sysbio/sys037>
- 111 Blowes SA, Supp SR, Antão LH, et al (2019) The geography of biodiversity change in marine and terrestrial
112 assemblages. *Science* (80-) 366:339–345. <https://doi.org/10.1126/science.aaw1620>
- 113 Boero F (2010) The study of species in the era of biodiversity: a tale of stupidity. *Diversity* 2:115–126.
114 <https://doi.org/10.3390/d2010115>
- 115 Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine metazoa. *Ann Rev Mar Sci* 3:471–508.
116 <https://doi.org/10.1146/annurev-marine-120308-080950>
- 117 Charron DF (2012) Ecohealth: origins and approach. In: Charron DF (ed) *Ecohealth Research in Practice: Innovative*
118 *Applications of an Ecosystem Approach to Health*. Springer New York, New York, NY, pp 1–30
- 119 Collins JE, Rabone M, Vanagt T, et al (2020) Strengthening the global network for sharing of marine biological
120 collections: recommendations for a new agreement for biodiversity beyond national jurisdiction. *ICES J Mar Sci*.
121 <https://doi.org/10.1093/icesjms/fsaa227>
- 122 Costello MJ, Wilson S, Houlding B (2012) Predicting total global species richness using rates of species description and
123 estimates of taxonomic effort. *Syst Biol* 61:871. <https://doi.org/10.1093/sysbio/syr080>
- 124 de Carvalho MR, Bockmann FA, Amorim DS, et al (2007) Taxonomic Impediment or Impediment to Taxonomy? A
125 Commentary on Systematics and the Cybertaxonomic-Automation Paradigm. *Evol Biol* 34:140–143.
126 <https://doi.org/10.1007/s11692-007-9011-6>
- 127 DeSalle R, Goldstein P (2019) Review and Interpretation of Trends in DNA Barcoding. *Front Ecol Evol* 7:1–11.
128 <https://doi.org/10.3389/fevo.2019.00302>
- 129 Fehlaue-Ale KH, Mackie JA, Lim-Fong GE, et al (2014) Cryptic species in the cosmopolitan *Bugula neritina* complex
130 (Bryozoa, Cheilostomata). *Zool Scr* 43:193–205. <https://doi.org/10.1111/zsc.12042>
- 131 Harley CDG (2011) Climate Change, Keystone Predation, and Biodiversity Loss. *Science* (80-) 334:1124–1127.
132 <https://doi.org/10.1126/science.1210199>
- 133 Holman LE, de Bruyn M, Creer S, et al (2019) Detection of introduced and resident marine species using environmental
134 DNA metabarcoding of sediment and water. *Sci Rep* 9:11559. <https://doi.org/10.1038/s41598-019-47899-7>
- 135 Horton T, Kroh A, Ah Yong S, et al (2021) World Register of Marine Species (WoRMS)
- 136 Jenkins CN, Van Houtan KS (2016) Global and regional priorities for marine biodiversity protection. *Biol Conserv*
137 204:333–339. <https://doi.org/10.1016/j.biocon.2016.10.005>
- 138 Martin D, Aguado MT, Álamo MAF, et al (2021) On the diversity of phyllodocida (annelida: Errantia), with a focus on
139 glyceridae, goniadidae, nephtyidae, polynoidae, sphaerodoridae, syllidae, and the holoplanktonic families.
140 *Diversity* 13:. <https://doi.org/10.3390/d13030131>
- 141 Mora C, Tittensor DP, Adl S, et al (2011) How many species are there on earth and in the ocean? *PLoS Biol*
142 9:e1001127. <https://doi.org/10.1371/journal.pbio.1001127>
- 143 Myers N, Mittermeier RA, Mittermeier CG, et al (2000) Biodiversity hotspots for conservation priorities. *Nature*
144 403:853–858. <https://doi.org/10.1038/35002501>
- 145 Pecl GT, Araújo MB, Bell JD, et al (2017) Biodiversity redistribution under climate change: Impacts on ecosystems and
146 human well-being. *Science* (80-) 355:. <https://doi.org/10.1126/science.aai9214>
- 147 Pereira HM, Leadley PW, Proença V, et al (2010) Scenarios for global biodiversity in the 21st century. *Science* (80-)
148 330:1496–1501. <https://doi.org/10.1126/science.1196624>
- 149 Pimm SL, Jenkins CN, Abell R, et al (2014) The biodiversity of species and their rates of extinction, distribution, and
150 protection. *Science* (80-) 344:1246752–1246752. <https://doi.org/10.1126/science.1246752>
- 151 Radulović AE, Archambault P, Dufresne F (2010) DNA barcodes for marine biodiversity: Moving fast forward?
152 *Diversity* 2:450–472. <https://doi.org/10.3390/d2040450>
- 153 Ramirez JL, Rosas-Puchuri U, Cañedo RM, et al (2020) DNA barcoding in the Southeast Pacific marine realm: Low
154 coverage and geographic representation despite high diversity. *PLoS One* 15:1–13.
155 <https://doi.org/10.1371/journal.pone.0244323>
- 156 Ratnasingham S, Hebert PDN (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN)
157 system. *PLoS One* 8:. <https://doi.org/10.1371/journal.pone.0066213>
- 158 Raupach MJ, Amann R, Wheeler QD, Roos C (2016) The application of “-omics” technologies for the classification
159 and identification of animals. *Org Divers Evol* 16:1–12. <https://doi.org/10.1007/s13127-015-0234-6>
- 160 Trivedi S, Aloufi AA, Ansari AA, Ghosh SK (2016) Role of DNA barcoding in marine biodiversity assessment and
161 conservation: An update. *Saudi J Biol Sci* 23:161–171. <https://doi.org/10.1016/j.sjbs.2015.01.001>
- 162 Trontelj P, Fišer C (2009) Perspectives: Cryptic species diversity should not be trivialised. *Syst Biodivers* 7:1–3.
163 <https://doi.org/10.1017/S1477200008002909>

- 164 Troudet J, Grandcolas P, Blin A, et al (2017) Taxonomic bias in biodiversity data and societal preferences. *Sci Rep* 7:1–
165 14. <https://doi.org/10.1038/s41598-017-09084-6>
- 166 Valentini A, Taberlet P, Miaud C, et al (2016) Next-generation monitoring of aquatic biodiversity using environmental
167 DNA metabarcoding. *Mol Ecol* 25:929–942. <https://doi.org/10.1111/mec.13428>
- 168 Vargas S, Schuster A, Sacher K, et al (2012) Barcoding Sponges: An Overview Based on Comprehensive Sampling.
169 *PLoS One* 7:e39345. <https://doi.org/10.1371/journal.pone.0039345>
- 170 Wangenstein OS, Palacín C, Guardiola M, Turon X (2018) DNA metabarcoding of littoral hard-bottom communities:
171 high diversity and database gaps revealed by two molecular markers. *PeerJ* 6:e4705.
172 <https://doi.org/10.7717/peerj.4705>
- 173 Weigand H, Beermann AJ, Čiampor F, et al (2019) DNA barcode reference libraries for the monitoring of aquatic biota
174 in Europe: Gap-analysis and recommendations for future work. *Sci Total Environ* 678:499–524.
175 <https://doi.org/10.1016/j.scitotenv.2019.04.247>
- 176 Worm B, Barbier EB, Beaumont N, et al (2006) Impacts of Biodiversity Loss on Ocean Ecosystem Services. *Science*
177 (80-) 314:787–790. <https://doi.org/10.1126/science.1132294>

Figure legends

Fig. 1 Number of nominal marine and COI-sequenced Animalia species for the five target Large Marine Ecosystems (data accessed in May 2021)

Fig. 2 (a) Species occurrences over time for each target Large Marine Ecosystem and **(b)** total species occurrences relative to total number of described species per each LME (data accessed in May 2021)

Fig. 3 Cumulative number of nominal marine Animalia and COI-sequenced species per phylum in the five target Large Marine Ecosystems (data accessed in May 2021)

Fig. 1

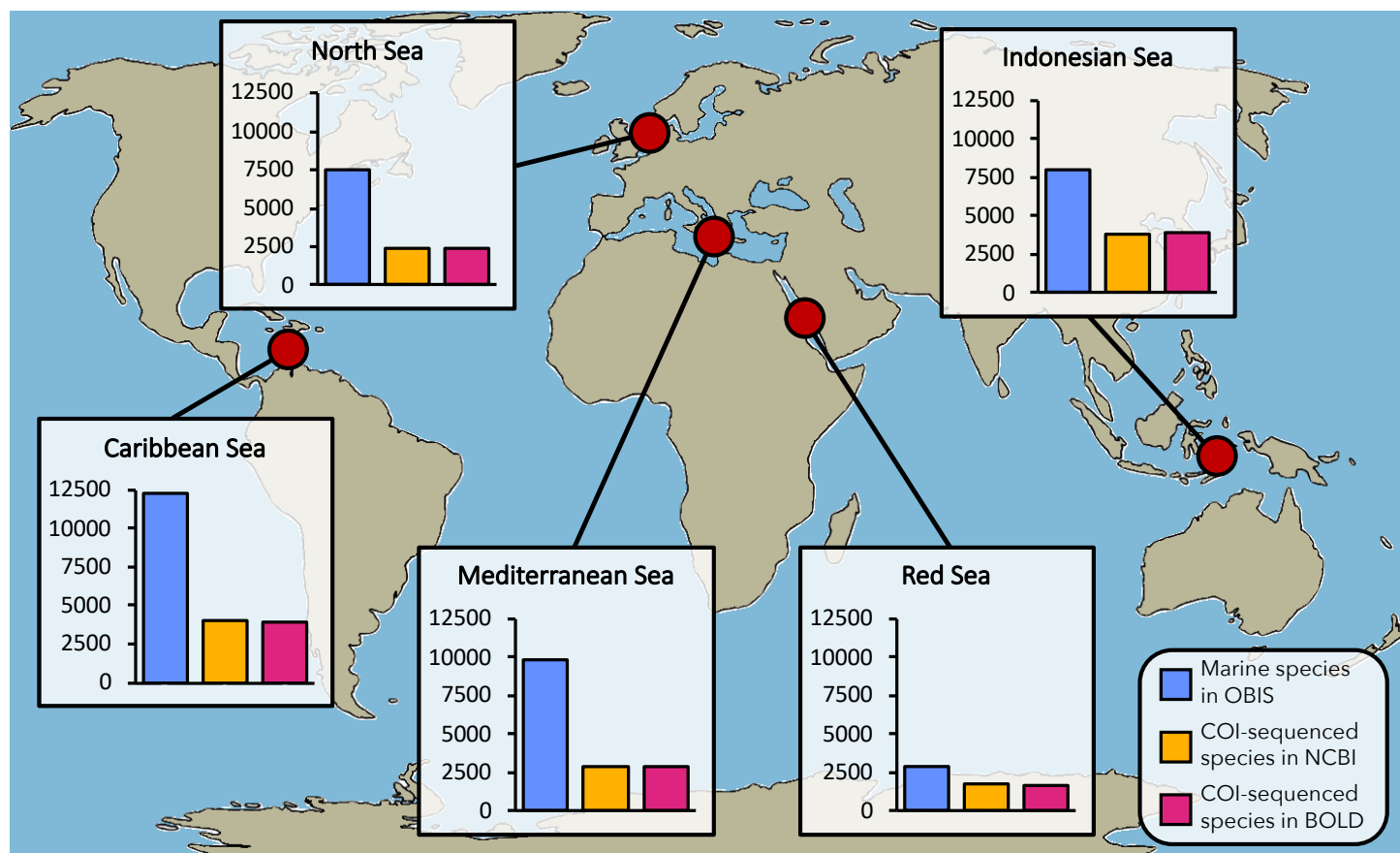


Fig. 2

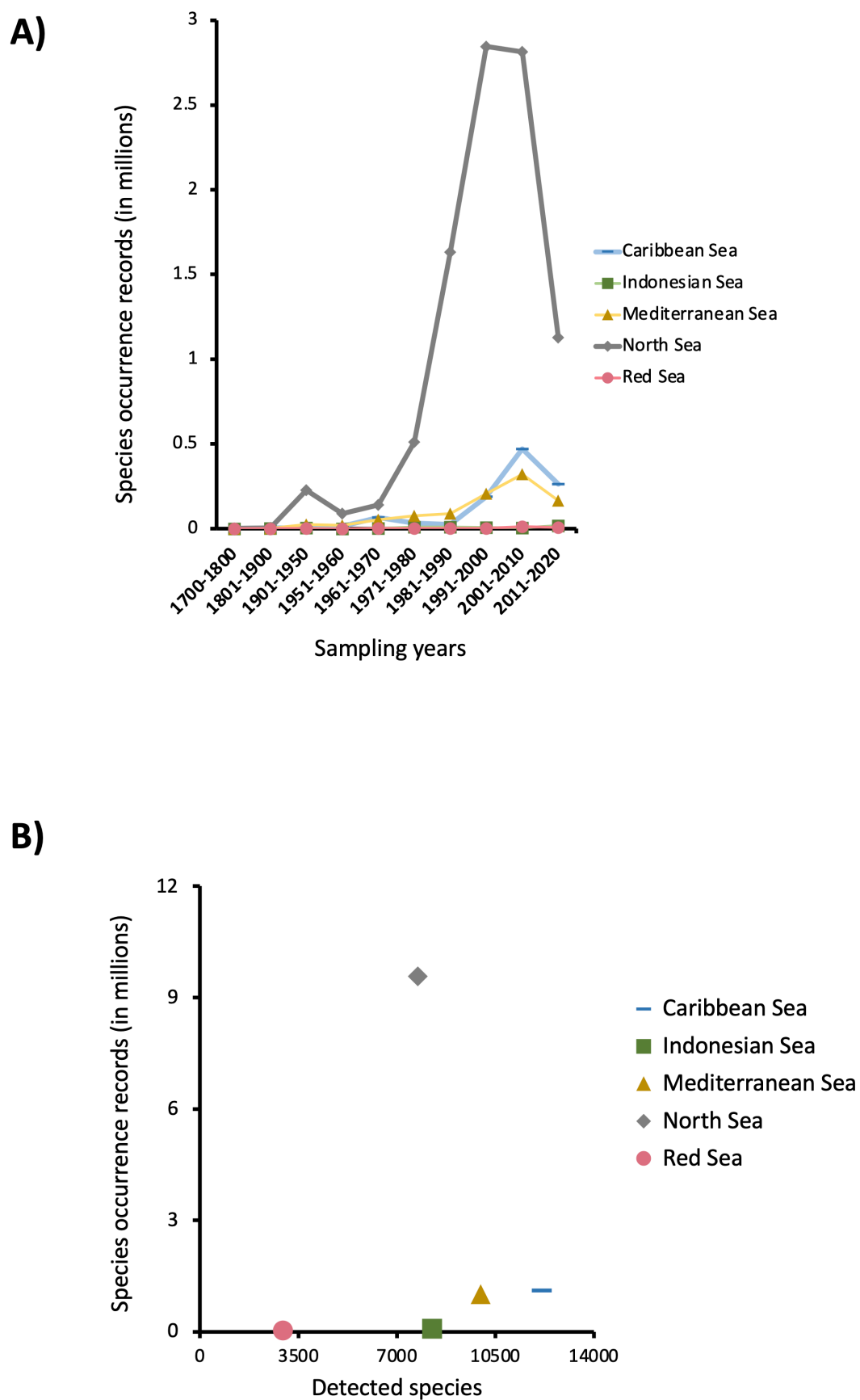


Fig. 3

