



**HAL**  
open science

## Ostracods in databases: State of the art, mobilization and future applications

Huai-Hsuan Huang, Moriaki Yasuhara, David Horne, Vincent Perrier, Alison Smith, Simone Brandão

### ► To cite this version:

Huai-Hsuan Huang, Moriaki Yasuhara, David Horne, Vincent Perrier, Alison Smith, et al.. Ostracods in databases: State of the art, mobilization and future applications. *Marine Micropaleontology*, 2022, 174, pp.102094. 10.1016/j.marmicro.2022.102094 . hal-04823555

**HAL Id: hal-04823555**

**<https://hal.science/hal-04823555v1>**

Submitted on 7 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1

2

3 Title: Ostracods in databases: state of the art, mobilization and future applications

4

5 Authors: Huai-Hsuan M. Huang<sup>1,2,\*</sup>, Moriaki Yasuhara<sup>2,\*</sup>, David J. Horne<sup>3</sup>, Vincent Perrier<sup>4</sup>,  
6 Alison J. Smith<sup>5</sup>, Simone N. Brandão<sup>6</sup>

7

8 <sup>1</sup>Department of Paleobiology, National Museum of Natural History, Smithsonian Institution,  
9 Washington DC 20013-7012

10 <sup>2</sup>School of Biological Sciences, Division of Ecology & Biodiversity, Swire Institute of Marine  
11 Science, and State Key Laboratory of Marine Pollution, the University of Hong Kong, Kadoorie  
12 Biological Sciences Building, Pokfulam Road, Hong Kong SAR.

13 <sup>3</sup>School of Geography, Queen Mary University of London, Mile End Road, London E1 4NS, UK

14 <sup>4</sup>Université de Lyon, UCBL, ENSL, CNRS, UMR 5276 LGL-TPE, 69622 Villeurbanne, France

15 <sup>5</sup>Department of Geology, Kent State University, Kent, Ohio 44242, USA

16 <sup>6</sup>Programa de Pós-Graduação em Zoologia Departamento de Ciências Biológicas, Campus Soane  
17 Nazaré de Andrade, Rodovia Jorge Amado, km 16, Bairro Salobrinho, Ilhéus, Bahia, CEP  
18 45662-900, Brazil

19 \* authors for correspondence: [huanghuaihuan@gmail.com](mailto:huanghuaihuan@gmail.com) (HMH);

20 [moriakiyasuhara@gmail.com](mailto:moriakiyasuhara@gmail.com) (MY)

21

22 **Abstract**

23 Biodiversity databases are changing the longevity of data in the era of open science. They also  
24 represent a collaboration opportunity in analyzing large-scale (paleo)biological patterns beyond a  
25 local project or a time scale. Ostracods, microscopic crustaceans, are a component in many  
26 biodiversity databases. They live in most kinds of aquatic environments today and their fossil  
27 record spans nearly the whole of the Earth's metazoan biosphere history from Ordovician to  
28 Holocene. Thus, ostracods provide an ideal model system for understanding large-scale  
29 biodiversity patterns and dynamics in both space and time. Thanks to many contributors, current  
30 and future ostracodologists have access to databases that have gone through numerous  
31 improvements and have been populated by many datasets. However, rapid growth of databases  
32 has caused confusion among users regarding available data, technical terms and database aims.  
33 We review key databases that include ostracods, summarizing their history of development,  
34 current spatial and temporal coverage, various types of data models and the intertwined  
35 relationships between databases. We also present a quantitative summary of ostracod diversity  
36 history based on the Paleobiology Database. Our investigations show that the database field is  
37 transitioning from the traditional single focus to multipurpose, from static to dynamic data  
38 display/download and from independent systems to collaborative networks. We compare the  
39 ways several databases approach persistent challenges such as taxonomic harmonization,  
40 validation of the original sampling metadata and paleolocality uncertainties. With increasing  
41 capability of data integration, databases continue to require enormous efforts regarding high-  
42 quality data entry and careful coordination among scientists and technical teams.

43 **Keywords:** biogeography, macroecology, collaboration tool, open science, biodiversity  
44 informatics

## 45 **1. Introduction**

46 To assess the biological impacts of global changes there is an increasing demand to  
47 understand the evolving dynamics among ecological components over various spatial and  
48 temporal scales (e.g., Jablonski and Sepkoski Jr, 1996; Scholes et al., 2008; Rick and Lockwood,  
49 2013). Playing an increasingly important role in this research trend, many databases have been  
50 developed to preserve, curate, and mobilize hard-earned biodiversity data (Goddard et al., 2011),  
51 motivated by a need to increase the discoverability or accessibility of such data. Investigations  
52 into macroecology, ecosystem dynamics, and macroevolution increasingly rely on data drawn  
53 from biodiversity databases. (e.g., Liow and Stenseth, 2007; Marx and Uhen, 2010; Tittensor et  
54 al., 2010; Villéger et al., 2011; Lazarus et al., 2014; Kocsis et al., 2018b; Reddin et al., 2020;  
55 Chaudhary et al., 2021). The importance of biodiversity databases is evident, and their  
56 continuous improvement requires the support and collaboration of the scientific community.

57 Biodiversity databases are rapidly evolving. Traditionally, data have been scattered  
58 across platforms in various formats with lower interoperability. One important characteristic of  
59 improved biodiversity databases is data mobilization with higher interoperability. The term  
60 mobilization refers to making data readily operable for analysis and downloadable through a  
61 freely accessible website (Faith et al., 2013; Nelson and Ellis, 2019). It can be applied to all  
62 kinds of biodiversity data, including digitized museum collections data, field-based research  
63 datasets and laboratory-generated DNA barcode data. The implementation of data mobilization is  
64 often consolidated by research infrastructures and digital tools that are designed to (ideally)  
65 enhance the reproducibility of scientific studies, to facilitate exchange of datasets in consistently  
66 structured ways, to promote open data/science (Burgelman et al., 2019; Powers and Hampton,  
67 2019) and to facilitate integrated data analyses (e.g., Costello and Wieczorek, 2014; Peters and

68 McClennen, 2016; Williams et al., 2018). For example, metadata documentation and many data  
69 standards have been developed to allow machine readable exchange of data between databases.  
70 Depending on the database, the rules for metadata documentation vary from a free “readme”  
71 style to formal standards that define/dictate vocabularies for describing data. Although more  
72 work is required, especially for 3D digital morphological data (Davies et al., 2017),  
73 communities/working groups such as the Biodiversity Information Standards (TDWG) and the  
74 Knowledge Network for Biocomplexity (KNB) have devoted tremendous efforts to maintaining  
75 widely used metadata standards in the broad discipline of biology and paleobiology (see below).

76         A comprehensive introduction to all metadata standards is beyond the scope of this  
77 review, but here we briefly mention several widely used examples. The Dublin Core Metadata  
78 Initiative defines a set of vocabularies for basic information, such as title and creator (Weibel,  
79 1999). Darwin Core defines a set of vocabularies for biodiversity data based on taxa, such as  
80 occurrence and identification (Darwin Core Task Group, 2009; Wiczorek et al., 2012).  
81 Audubon Core defines a set of vocabularies for multimedia resources and collections  
82 (GBIF/TDWG Multimedia Resources Task Group, 2013; Morris et al., 2013). The Access to  
83 Biological Collections Data (ABCD) schema and its extensions establish a data exchange  
84 standard for mobilizing data about DNA, specimens, observations and geological samples from  
85 museums and botanical gardens (Access to Biological Collections Data Task Group, 2005;  
86 Holetschek et al., 2012). Ecological Metadata Language administers a comprehensive set of  
87 vocabularies for describing research data and associated details, such as the temporal/spatial  
88 extents of data and methods (Fegraus et al., 2005; Jones et al., 2019). These community-  
89 maintained metadata standards and their associated software development are the key to  
90 mobilizing data from all kinds of data sources.

91           In this paper, we aim to inform potential database users and contributors of the strengths  
92 and weaknesses of all mainstream biodiversity databases that include ostracod records. Ostracods  
93 are a group of microscopic crustaceans with an excellent fossil record. They cover a wide variety  
94 of ecological niches, have long evolutionary history traced back to the Ordovician (Smith and  
95 Horne, 2002; Williams et al., 2008; Yasuhara and Cronin, 2008; Rodriguez-Lazaro and Ruiz-  
96 Muñoz, 2012) and are a major component of the meiofaunal biodiversity (Brandt et al., 2007).  
97 Being ubiquitous in almost all aquatic systems and commonly preserved as fossils in  
98 sedimentary rocks, ostracods are an excellent biological proxy in contemporary environmental  
99 studies, paleoceanographic and paleoclimatic reconstructions, and deep-time paleoecological and  
100 macroevolutionary studies (e.g., Holmes and Chivas, 2002; Boomer et al., 2003; Yasuhara and  
101 Cronin, 2008; Mesquita-Joanes et al., 2012; Jöst et al., 2019; Yasuhara, 2019; Chiu et al., 2020;  
102 Martins et al., 2020). Because of their diverse applications, they have clear potential in the era of  
103 open science. Data resources of ostracods have been steadily growing since the pioneering  
104 databases, such as the Kempf Database Ostracoda (Viehberg et al., 2014) and Ellis & Messina  
105 Catalogues (Ellis and Messina, 1952). Thus, it is timely to evaluate the current role of ostracods  
106 in the growing trend of database-oriented research. The aims of this paper are:

- 107       • to give an overview of key database projects that involve ostracods, including  
108       taxonomically oriented databases (Section 3), Aphia: a multipurpose biodiversity  
109       platform (Section 4), marine occurrence-based (paleo)biodiversity databases (Section 5),  
110       non-marine occurrence-based (paleo)biodiversity databases (Section 6), and GBIF  
111       (Section 8);
- 112       • to assess the potential role of ostracod data in understanding the history of life, using  
113       based on records in the Paleobiology Database (Section 7); and

- 114       • to discuss challenges and future directions in developing ostracod databases and their  
115       applications (Section 9).

## 116 **2. Methods**

117

### 118 **2.1. Database categories**

119       To provide the reader with some orientation in the huge pool of existing databases and  
120 tools in general, we have divided them into the following categories (Table 1): (1) taxonomically  
121 oriented databases, (2) multipurpose biodiversity databases, (3) data archives/repositories, (4)  
122 data harvesters/recombiners, (5) occurrence-based (paleo)biodiversity databases, and (6)  
123 application programming interfaces (APIs). The term database can be used to describe a system  
124 where datasets are stored, or alternatively to mean a group of datasets. Taxonomically oriented  
125 databases focus on authoritative lists of taxonomic names and classification. The multipurpose  
126 biodiversity databases category refers to the Aphia platform and its associated databases. Data  
127 harvesters/recombiners, such as GBIF (Global Biodiversity Information Facility) and OBIS  
128 (Ocean Biodiversity Information System), aggregate data through the Integrated Publishing  
129 Toolkit (IPT) and currently host the largest amount of biodiversity data among all the databases  
130 discussed herein. The IPT is a free software tool developed by GBIF for registered data  
131 publishers, such as the U.S. Geological Survey and the Natural History Museum London, to  
132 share biodiversity datasets in the Darwin Core Archive standard (Robertson et al., 2014).  
133 Occurrence-based (paleo)biodiversity databases, such as the Arctic Ostracode Database and the  
134 Paleobiology Database, rely on authorized scientists or working groups to enter/compile data  
135 which mainly come from the published literature. Application Programming Interfaces are not  
136 databases, but a general software term to describe an interface system that connects programs  
137 and endpoints; they are mentioned here due to their ubiquity and usefulness in data accessibility.

138 Each database has its own unique underlying structure for hosting data. Whereas some database  
 139 designs may facilitate addressing certain questions better than others, the diversity of designs is  
 140 complementary and beneficial to almost all lines of research and education (e.g., Wright et al.,  
 141 2013; Lautenschlager and Rücklin, 2014; Hendricks et al., 2015; Lockwood et al., 2018).

142  
 143 **Table 1.** General data resources for (paleo)biology (beyond ostracods). We divide the example  
 144 databases into six categories according to their characteristics, not for a strict definition but to  
 145 help the reader navigate through the common data resources. Note that some databases can  
 146 exchange data with other databases, and thus the same datasets could exist in multiple databases.  
 147 The web interface of one database can also offer links to multiple databases.

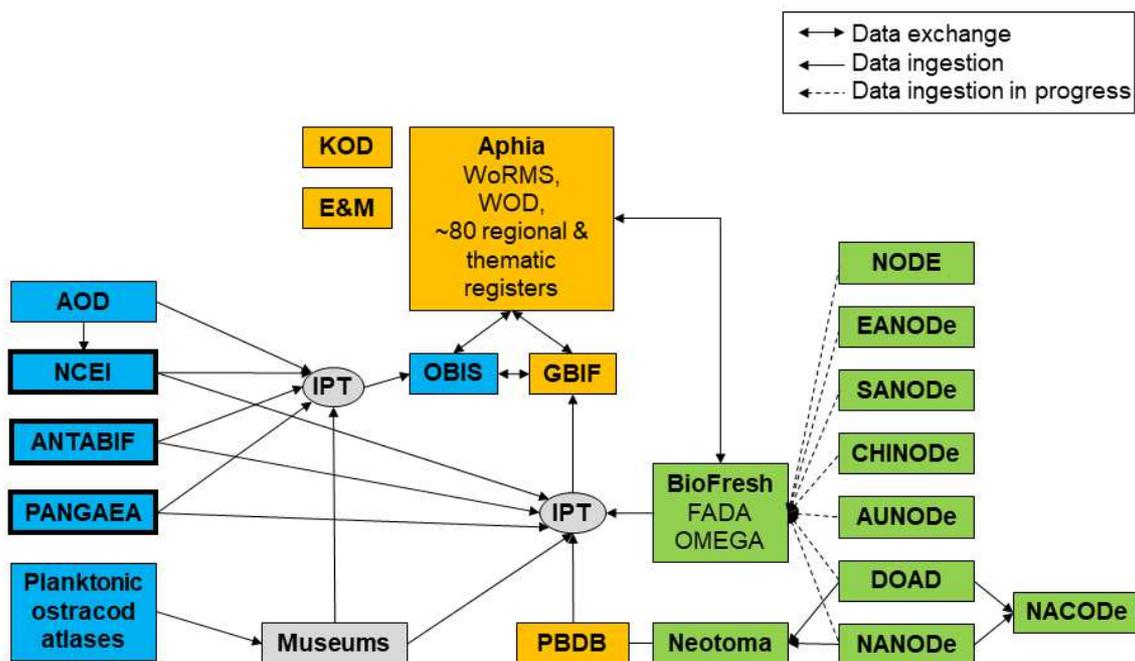
Categories	Examples	Characteristics
Taxonomically oriented databases	ITIS, Open Tree Taxonomy, Catalogue of Life	<ul style="list-style-type: none"> <li>- Authoritative lists of taxonomic names and classification</li> <li>- Contributed by taxonomic experts</li> <li>- Often as the taxonomic backbone for other databases</li> </ul>
Multi-purpose biodiversity databases	Aphia platform (WOD, WoRMS, and ~80 registers)	<ul style="list-style-type: none"> <li>- Authoritative lists of taxonomic names and classification</li> <li>- Contributed by taxonomic experts</li> <li>- Often the taxonomic backbone for other databases</li> <li>- Designed for storing ecological, morphological, biogeographical, stratigraphical and taxonomic data</li> </ul>
Data archives /repositories	Dryad, PANGAEA, NOAA's National Centers for Environmental Information, DataOne, Morpho Source (3D datasets), Morphobank (images and phylogenetic matrices), Treebase (phylogenetic matrices)	<ul style="list-style-type: none"> <li>- Populated by static data in various formats</li> <li>- Contributed by users/authors of publications</li> </ul>
Data harvesters /recombiners	GBIF, iDigBio, iDigPaleo, OBIS	<ul style="list-style-type: none"> <li>- Populated by dynamic data with the formats set by Biodiversity Information Standards (TDWG)</li> <li>- Contributed by registered publishers, such as museums and data repositories, through Integrated Publishing Toolkits</li> </ul>
Occurrence-based (paleo)biodiversity databases	Paleobiology Database, Neotoma Database, NOW fossil mammal database, Geobiodiversity Database	<ul style="list-style-type: none"> <li>- Populated by dynamic data in relational tables</li> <li>- Contributed by verified individual contributors or working groups</li> </ul>

Application programming interfaces	Earth Life Consortium, EPANDDA API	<ul style="list-style-type: none"> <li>- Not a database, but a tool for databases</li> <li>- Linkages between databases or endpoints</li> <li>- A portal for searching data in multiple databases</li> </ul>
------------------------------------	------------------------------------	--

148

149 **2.2. Reviewed databases**

150 We have reviewed both key ostracod database projects and broader databases that house  
 151 ostracod data (Figure 1 and Table 2). For each database, we have studied its stated objectives,  
 152 data sources, data accessibility, geographical and temporal coverage, and interaction with other  
 153 databases.



154

155 **Figure 1.** Diagram of the relationship among reviewed databases (in the broadest sense). This  
 156 diagram shows that many databases are in partnership rather than in isolation. Marine databases  
 157 (blue) and non-marine databases (green) are connected through databases (orange) that  
 158 accommodate both information. Some databases (bold) are independent, and some (not bold) are  
 159 a part of or adopted into Aphia or BioFresh. We simplified the connections to three kinds: (1)  
 160 data exchange (solid double arrows) between independent databases, (2) data ingestion from

161 regional databases to broader-scale databases (solid arrows), and (3) data ingestion in progress  
 162 (dashed arrows). Paleobiology Database (PBDB) and Neotoma Database are in close  
 163 collaboration (a solid line) through Earth Life Consortium. GBIF and OBIS aggregate data  
 164 through Integrated Publishing Toolkits (gray circles); NCEI, ANTABIF, and PANGAEA  
 165 (thickened box border) are contributed by authors of publications; other databases are maintained  
 166 by authorized editors/enterers/data stewards/working groups. Specimens and collections from  
 167 museums (gray) are important data sources to GBIF and OBIS. “Planktonic ostracod atlases”  
 168 refer to “Atlas of Atlantic Planktonic Ostracods” and “An Atlas of Southern Ocean Planktonic  
 169 Ostracods”. Abbreviations: see Section 2.1. for full names.

170  
 171 **Table 2.** Data resources reviewed in this paper. The table is arranged by the appearance  
 172 sequence of the databases in Sections 3–8. The network column lists some databases that link  
 173 with the focal database.

Name	Category	Marine	Non-marine	Age accommodation	Network	Reference
Kempf Database Ostracoda	Taxonomically oriented database	√	√	Extant and extinct	--	(Viehberg et al., 2014)
Ellis and Messina Catalogues: Ostracoda	Taxonomically oriented database	√	√	Extant and extinct	--	(Ellis and Messina, 1952; Micropaleontology Press, n.d.)
(Aphia) World Ostracoda Database	Multipurpose biodiversity database	√	√	Extant and extinct	EoL, ITIS, CoL, GBIF, OBIS, etc.	(Brandão & Karanovic, 2020)
Freshwater Animal Diversity Assessment (FADA): Ostracoda	Taxonomic checklists generated from an occurrence-based (paleo)biodiversity database		√	Extant	Freshwater Information Platform, BioFresh, Aphia, GBIF	(Martens et al., 2008; Martens et al., 2013)
Arctic Ostracode Database (AOD)	Occurrence-based (paleo)biodiversity database	√		Modern to the Quaternary	NOAA’s NCEI, OBIS	(Cronin et al., 1995; Gemery et al., 2017; Cronin et al., 2021)
Ocean Biogeographic Information System (OBIS)	Occurrence-based data harvester/recombiner	√		Modern to the Quaternary	GBIF, PANGAEA, ANTABIF	(OBIS, 2021)
PANGAEA	Data archive/repository	√		Modern to the Quaternary	OBIS, GBIF	(PANGAEA®)

Antarctic Biodiversity Information Facility (ANTABIF)	Data archive/repository	√	Modern/Recent	Antarctic Biodiversity Portal, OBIS	(biodiversity.aq, 2021)
Atlas of Atlantic Planktonic Ostracods	Outputs generated from an occurrence-based (paleo)biodiversity database	√	Modern/Recent	Natural History Museum London, OBIS	(Angel et al., 2008)
An Atlas of Southern Ocean planktonic ostracods	Outputs generated from an occurrence-based (paleo)biodiversity database	√	Modern/Recent	OBIS	(Blachowiak-Samolyk and Angel, 2008)
Non-marine Ostracod Distribution in Europe (NODE)	Occurrence-based (paleo)biodiversity database	√	Modern to the Quaternary	OMEGA	(Horne et al., 2011)
North American Non-marine Ostracode Database (NANODE)	Occurrence-based (paleo)biodiversity database	√	Modern/Recent	Neotoma Database, OMEGA, NACODE	(Smith et al., 2015)
Delorme Ostracode Autecological Database (DOAD)	Occurrence-based (paleo)biodiversity database	√	Modern/Recent	Neotoma Database, OMEGA, NACODE	(Curry et al., 2012)
East Asia Non-marine Ostracod Database (EANODE)	Occurrence-based (paleo)biodiversity database	√	Modern/Recent	OMEGA	Unpublished compilation
Ostracod Metadatabase of Environmental and Geographical Attributes (OMEGA)	Occurrence-based (paleo)biodiversity database	√	Modern/Recent	Freshwater Information Platform, BioFresh, GBIF	(Horne et al., 2011)
Neotoma Database	Occurrence-based (paleo)biodiversity database	√	Modern to Pliocene	Earth Life Consortium, NCEI	(Williams et al., 2018)

North American Combined Ostracode Database (NACODe)	Occurrence-based (paleo)biodiversity database	√		Modern/Recent	--	(Curry et al., 2012)
Paleobiology Database	Occurrence-based (paleo)biodiversity database	√	√	Phanerozoic	Earth Life Consortium, ePANDDA API, GBIF	(Peters and McClennen, 2016)
Global Biodiversity Information Facility (GBIF)	Data harvester/recombiner	√	√	Living and fossil	OBIS, Paleobiology Database, Biofresh, etc.	(GBIF, 2021)

174

175           In Section 3, we review the Kempf Database Ostracoda and Ellis & Messina Catalogues  
176 that are designed to preserve taxonomic information for both marine and freshwater ostracods. In  
177 Section 4, we review the Aphia platform and the World Ostracoda Database (WOD) that house  
178 multiple types of data for both marine and freshwater ostracods, including recent and fossil taxa.  
179 The Freshwater Animal Diversity Assessment (FADA) project is briefly reviewed in the Aphia  
180 platform section because they are in a close collaboration. In Section 5, we review the Arctic  
181 Ostracode Database, Ocean Biogeographic Information System, “Atlas of Atlantic Planktonic  
182 Ostracods” and “An Atlas of Southern Ocean Planktonic Ostracods”. An atlas is usually not a  
183 database but represents a static output of a database. Nonetheless, we include these two atlases  
184 here because they are valuable resources for planktonic ostracods, and because the original data  
185 have been partially published by the Natural History Museum London in OBIS as a dataset  
186 called “Personal library collection of Martin Angel of published and unpublished Halocyprid  
187 (Ostracoda) occurrences” (Angel, 2016). In Section 6, we review non-marine occurrence-based  
188 (paleo)biodiversity databases. In Section 7, we visit ostracods in the Paleobiology Database  
189 (PBDB), one of the more comprehensive spatiotemporal paleoecological databases that

190 accommodate fossil data (Ordovician to Quaternary for ostracods). In Section 8, we introduce  
191 GBIF, which aggregates data from all above databases through Integrated Publishing Toolkits.

192 We do not list static individual datasets deposited in data archives/repositories, such as  
193 PANGAEA and NOAA's National Centers for Environmental Information. This is because many  
194 mainstream data repositories have joined data harvesters, such as OBIS and GBIF. In Sections  
195 5.2. and 8, we consider how OBIS and GBIF use metadata standards and recombine datasets  
196 from numerous data repositories, respectively. Among these data repositories, PANGAEA and  
197 ANTABIF are introduced to illustrate the ostracod data in OBIS.

198 Neptune Sandbox Berlin curates microfossil occurrence and biostratigraphy data from the  
199 DSDP/ODP/IODP projects, but include only fossil microplankton (e.g., calcareous nannofossils,  
200 foraminifera, radiolarians, diatoms, and dinoflagellates) (Lazarus, 1994). Thus, it is excluded  
201 from our review since planktonic ostracods (and other planktonic arthropods) rarely become  
202 microfossils due to poorly calcified carapace (Perrier et al., 2015). However, it should be  
203 mentioned that numerous studies have generated excellent fossil records of benthic ostracods  
204 from DSDP/ODP/IODP samples (e.g., Majoran and Dingle, 2001; Bergue and Govindan, 2010;  
205 Alvarez Zarikian, 2015).

206 Abbreviations: Antarctic Biodiversity Information Facility (ANTABIF), Arctic Ostracode  
207 Database (AOD), Australian Non-marine Ostracode Database (AUNODE), Freshwater  
208 Biodiversity Data Portal (BioFresh), Catalogue of Life (CoL), Chinese Non-marine Ostracode  
209 Database (CHINODE), Delorme Ostracode Autecological Database (DOAD), East Asia Non-  
210 marine Ostracod Database (EANODE), Encyclopedia of Life (EoL), EU-FP6 project Marine  
211 Biodiversity and Ecosystem Functioning Network of Excellence (MarBEF), Ellis & Messina  
212 Catalogues (E&M), Freshwater Animal Diversity Assessment (FADA), Global Biodiversity

213 Information Facility (GBIF), Kempf Database Ostracoda (KDO), Nonmarine Ostracod  
214 Distribution in Europe (NODE), North American Combined Ostracode Database (NACODE),  
215 North American Nonmarine Ostracode Database (NANODE), Ocean Biogeographic Information  
216 System (OBIS), Ostracod Metadatabase of Environmental and Geographical Attributes  
217 (OMEGA), Paleobiology Database (PBDB), South African Non-marine Ostracode Database  
218 (SANODE), World Register of Marine Species (WoRMS), World Ostracoda Database (WOD).

219

### 220 **2.3. Data analysis**

221 We calculated database status metrics where possible, including spatial coverage,  
222 temporal coverage, resolution of taxonomic identification, and database growth over time. Table  
223 3 summarizes the download information, the requested taxa, and parameters applied on the raw  
224 data for each analysis. In Section 5.1., although a detailed biogeographical analysis is beyond the  
225 scope of this paper, we applied a network-based clustering analysis on the census data from the  
226 Arctic Ostracode Database to show a simple example of what type of analyses this database  
227 allows. This network-based clustering analysis determines biogeographical clusters or regions by  
228 calculating number of co-occurrences between pairs of species and number of shared species  
229 between pairs of sites in a bipartite occurrence network (Vilhena and Antonelli, 2015; Kocsis et  
230 al., 2018a). We used only samples with >100 specimens in the network analysis. In Section 7, we  
231 calculated the raw number of occurrences and genera in Ostracoda per geological period in the  
232 Paleobiology Database. We also calculated the number of genera per order in Ostracoda,  
233 including Podocopida, Palaeocopida, Platycopida, and Myodocopida.

234

235

236 **Table 3.** Data sources used in this paper. Data source includes the download information, request  
 237 parameters, web-link, and access date.

Figures	Data sources
Figure 2	All data downloaded from the publicly accessible Arctic Ostracode Database 2020 (Cronin et al. 2021) in National Centers for Environmental Information ( <a href="https://www.ncei.noaa.gov/">https://www.ncei.noaa.gov/</a> ) on 2021-10-13.
Figure 3	All ostracod data downloaded from the publicly accessible Ocean Biogeographic Information System ( <a href="https://obis.org/">https://obis.org/</a> ) on 2020-10-17.
Figure 4	All data provided by David J. Horne in December 2020, from Australian Non-marine Ostracode Database (AUNODE), Chinese Non-marine Ostracode Database (CHINODE), Delorme Ostracode Autecological Database (DOAD), East Asia Non-marine Ostracod Database (EANODE), North American Non-marine Ostracode Database (NANODE), Non-marine Ostracod Distribution in Europe (NODE) database, South African Non-marine Ostracode Database (SANODE)
Figure 5-7	All data downloaded from the publicly accessible Paleobiology Database ( <a href="https://paleobiodb.org/">https://paleobiodb.org/</a> ) on 2020-10-11. Requested taxa: Arthropoda. Applied filters: taxonomic resolution (all), preservation (regular taxa only), identification (latest), show accepted names only.

238

### 239 **3. Taxonomically oriented databases**

240 The Kempf Database Ostracoda (KDO) is one of the earliest initiatives to compile  
 241 taxonomic indexes and references. Kempf objectively compiled citations for more than 40,000  
 242 marine and 9,000 non-marine taxonomic names of living and fossil species from the Ordovician  
 243 to the present day (Matzke-Karasz, 2014). It is the single most complete and comprehensive  
 244 listing of ostracod taxonomic names available, referenced to the publications in which taxa were  
 245 originally described, and constituting a unique and valuable resource. It was published in CD  
 246 and book forms by the University of Cologne; the data do not include images, descriptions, or  
 247 synonymies. Since Kempf's death in 2017 there has been a necessary pause in the entry of new  
 248 data while efforts are made to establish support for the KDO and thus ensure its future  
 249 (<https://www.support-irgo.net/activities/kempf-database-ostracoda/>). It is planned to make it widely  
 250 available, for example via the German node of the Global Biodiversity Information Facility  
 251 (GBIF), provided that funds can be obtained to facilitate this.

252 Ellis & Messina Catalogues (<https://www.micropress.org/em/>) reproduce type descriptions  
 253 and illustrations of genus- and species-level taxa, together with details of stratigraphical and

254 geographical locations and the repositories of type material. New information beyond the  
255 original description, such as the repository of type material that is not shown in the original  
256 description and taxonomic note, is available for some species and genera. In addition to  
257 ostracods (>26,000 taxa) the catalogues cover foraminifera (>47,000 taxa) and diatoms (>7,000  
258 taxa), with around 300 new taxa being added annually to each catalogue (Micropaleontology  
259 Press website, viewed on 07-01-2021 at <https://www.micropress.org/em/about.php>).

260         The World Ostracoda Database (WOD) was established as a taxonomically oriented  
261 database in its first phase of database construction, compiling taxonomic information on fossil  
262 and Recent/living marine and non-marine ostracods. Using the robust infrastructure of the Aphia  
263 platform, the taxonomic information is accessible online and constantly growing. Ongoing  
264 further development phases include adding a wider range of data (e.g., biogeographical,  
265 ecological, evolutionary) (see Section 4. for more details).

266

#### 267 **4. Aphia, a multipurpose biodiversity platform, and the World Ostracoda Database**

268         The European Register of Marine Species (ERMS) is an authoritative (i.e., made by  
269 taxonomic experts) list (initially published as a printed book) of >29,000 (geologically) Recent  
270 marine and brackish-water species from European waters, of which some 700 are ostracods,  
271 together with information on their taxonomy and references for their taxonomic identification  
272 (Costello et al., 2001; Cuvelier et al., 2006). ERMS can now be accessed via the MarBEF web  
273 site (<http://www.marbef.org/data/erms.php>). Through a series of mostly EU funded programs,  
274 projects, initiatives, and institutions (e.g., MarBEF, Species 2000 Europa, Flanders Marine  
275 Institute, Species 2000), ERMS was fed into a relational digital database (Cuvelier et al., 2006),  
276 which later became the Aphia platform. ERMS was then expanded to a global register, the World

277 Register of Marine Species (WoRMS), and later the initiative expanded further to compiling  
278 information on all living and fossil species from all ecosystems (marine and non-marine), as well  
279 as starting to incorporate all aspects of taxonomy and nomenclature, also biological, ecological,  
280 evolutionary, biogeographical, genetic, bibliographical and nomenclatural information,  
281 conservation importance, economic importance, images, and notes (e.g., stratigraphy, taxonomic  
282 remarks, description). All these data are displayed to the public in the form of taxonomic,  
283 regional, and thematic registers, each one with its own webpage. Examples of these three kinds  
284 of registers are, respectively, the World Ostracoda Database (WOD), the Register of Antarctic  
285 Species (RAS) and the World Register of Introduced Marine Species (WRiMS). The different  
286 registers' web interfaces make the information contained in Aphia available to the public in a  
287 more accessible way. For example, someone looking for data on ostracods will tend to access the  
288 WOD webpage, while someone looking for Antarctic species will access RAS. However, all data  
289 are housed within Aphia, and the web interfaces provide search tools which communicate with  
290 the Aphia database.

291         Since its inception Aphia grew significantly and several tools were created; one worthy  
292 of mention is the online editing environment available since 2006, through which hundreds of  
293 accredited editors can log into the system and edit or add information only to their specific taxa  
294 (Costello, 2000; Vandepitte et al., 2015). In this way, although the information in the database is  
295 accessible to the general public, the quality of the data is assured by the fact that only accredited  
296 specialists can edit the information in the database. For example, only an editor can add, delete  
297 and correct taxonomic, ecological, evolutionary, biogeographical, genetic, bibliographical and  
298 nomenclatural information, conservation importance, economic importance, images,  
299 stratigraphical and other kinds of notes, etc. (Vandepitte et al., 2015). Furthermore, mistakes and

300 omissions can be constantly detected and passed to and corrected by the responsible editors.  
301 However, the lack of editors responsible for specific subjects or taxa will surely delay the  
302 correction of such mistakes. The low number of editors is now the main problem of WOD, with  
303 the consequence that mistakes take longer to be corrected. Nevertheless, a clear advantage of  
304 these open access, online databases (such as WOD) is to simplify or avoid taxonomic confusion,  
305 e.g., synonyms and homonyms. One example in ostracodology is *Vandenboldina* Wilson, 2010  
306 published as a new name for *Pseudoceratina* Bold, 1965, after Simone N. Brandão advertised it  
307 via Ostracon (the email list for ostracod researchers) as a junior homonym of *Pseudoceratina*  
308 Carter, 1885 (Porifera). Although the first objective of ERMS was to produce a species list, a  
309 much broader idea lay behind its initial proposals (e.g., Costello, 2000, Fig. 3), encompassing  
310 networking with other databases and scientific programs (for example, oceanographic atlases  
311 with information on water masses, bathymetry, nutrients, etc.), online tools for capturing, storing  
312 and displaying biogeographical information, and modelling of future species distribution and  
313 ocean ecosystem functions.

314         The Aphia platform contains a database, which hosts all data from WOD, WoRMS and  
315 the remaining ~80 global, regional, and thematic registers. The SQL database of Aphia with  
316 “over 400 fields, spread over 81 related tables”, which are grouped in relation to the content in  
317 “10 modules: taxonomy, distribution, traits, specimen information, vernacular names, notes,  
318 links, images, identification keys and sources” (Vandepitte et al., 2015). The EU-funded  
319 Lifewatch infrastructure for biodiversity and ecosystem research is promoting the data mining  
320 and upload of a large amount of biological, morphological, evolutionary, ecological, taxonomic  
321 and human defined characteristics of virtually all taxa, including ostracods, to the module “traits”  
322 of Aphia, e.g., body size, feeding type, life cycle, invasive species, threatened species, etc.

323 (Vandepitte et al., 2015). Aphia provides data to other databases (e.g., GBID, OBIS, CoL, ITIS,  
324 several museums, scientific institutions and individual scientists) and provides links to them in  
325 the specific taxa pages, in order to, for example, avoid duplication of work (Cuvelier et al., 2006;  
326 Vandepitte et al., 2015).

327         Within the scope of the AquaRES (Aquatic Species Register Exchange and Services)  
328 project, Aphia exchanges data with the Freshwater Animal Diversity Assessment (FADA). The  
329 AquaRES project aims at managing the data of aquatic groups occurring in both freshwater and  
330 marine ecosystems in one system (Vandepitte et al., 2015). The FADA database, led by Koen  
331 Martens from the Royal Belgian Institute of Natural Sciences, provides taxonomic checklists and  
332 distribution information of many freshwater animals, including ostracods (Balian et al., 2008;  
333 Martens et al., 2013). Georeferenced datasets can be made available via the Freshwater  
334 Information Platform which integrates FADA through the Freshwater Biodiversity Data Portal.  
335 In an output of FADA (<http://fada.biodiversity.be/group/show/18>), non-marine ostracods are  
336 grouped into eight biogeographic regions: Antarctic (3 species), Australasian (262 species),  
337 Afrotropical (453 species), Nearctic (300 species), Neotropical (290 species),  
338 Oriental/Indomalaya (222 species), Pacific/Oceania (57 species) and Palaeartic (749 species)  
339 (Martens et al., 2013). The distribution of each taxon is thus limited to its presence anywhere  
340 within a named biogeographical region, rather than single occurrences identified by coordinates  
341 and other data (e.g., latitude, longitude, altitude, date of sampling). The taxonomic classification  
342 of each taxon is provided, but not the taxonomic history or original description.

343         The World Ostracoda Database (WOD) is one of the global species registers housed in  
344 Aphia. Its aims are to compile a wide range of data types on recent and fossil Ostracoda,  
345 including taxonomy, classification, synonymy, description, bibliography, type specimens'

346 information, geographical and stratigraphical occurrences, illustrations, biological and ecological  
347 attributes (e.g., feeding type, body size, functional group, development, life cycle, habitat). WOD  
348 originated in 2008, when the chair of the WoRMS, Geoff Boxshall invited Simone N. Brandão to  
349 check a spreadsheet with approximately 8,000 ostracod taxa compiled from publications on  
350 Recent marine Ostracoda. In 2013, a grant from LifeWatch enabled a large input of data on  
351 Ostracoda, including tens of thousands of taxa and bibliographic references on marine, non-  
352 marine, extant, and extinct ostracods. WOD has kept expanding since then and has information  
353 now on more than 54,449 taxa: 45,294 species, of which 32,140 are accepted, and 3,676 genera,  
354 of which 3,446 are accepted.

355 WOD, WoRMS and the remaining registers in Aphia aim to record the entire taxonomic  
356 history of a taxon, and thus unaccepted taxa are kept in the database with an “unaccepted” label.  
357 In most cases these unaccepted taxa are invalid generic combinations at the species level, but  
358 other unaccepted taxa are subjective synonyms (i.e., taxa described with a different name and  
359 authorship, but considered later to be a synonym of a taxon described before). However, most  
360 taxa now available in WOD remain quarantined (hidden from the public). These taxa could be  
361 the ones without higher level taxonomic classification or those not yet checked by an expert.  
362 Nevertheless, 10,221 accepted species among a total of 20,548 taxa are available to everyone on  
363 the worldwide web (Brandão and Karanovic, 2020), the remaining +35,000 species in WOD will  
364 be checked before they are made accessible to the public. WOD also contains the citations of  
365 +22,000 publications on ostracods and +2,000 pdfs, which can be downloaded either directly (if  
366 no copyright applies) or upon request to the WOD (if the publication is copyrighted).  
367 Furthermore, WOD has also +2200 distribution records, +700 images and ~800 attributes (e.g.,  
368 biological, ecological, developmental, evolutionary data). In the context of Lifewatch Project,

369 ongoing steps include entering biogeographical, ecological, evolutionary data for different taxa,  
370 as well as keeping the taxonomic information and the references up to date.

371

## 372 **5. Marine occurrence-based (paleo)biodiversity databases**

### 373 **5.1. Arctic Ostracode Database**

374 The Arctic Ostracode Database (AOD) is dedicated to census data of benthic marine  
375 ostracods collected from a variety of surface and late Quaternary sediment samples from the  
376 Arctic area (Figure 2). The samples came from international sampling cruises to the Arctic since  
377 1933 (Figure 2B). The census data provide numbers of specimens per species in each sample. It  
378 also provides taxonomic information of Arctic ostracods, which all the contributors agreed on.  
379 Therefore, taxonomic names are thoroughly harmonized in the AOD. Age information at finer  
380 resolution than “modern to late Quaternary” is unavailable in the AOD but might be indicated by  
381 papers based on the same samples elsewhere.

382 The development of the AOD has been led by Thomas M. Cronin and Laura Gemery in  
383 the United States Geological Survey (USGS). The home repository of the AOD is NOAA's  
384 National Centers for Environmental Information. It started as the Modern Arctic Podocopid  
385 Ostracode Database (Cronin et al., 1991) and had several major updates, including Arctic  
386 Ostracode Database (Cronin et al., 1995), Modern Arctic Ostracode Database (Cronin et al.,  
387 2010), Arctic Ostracode Database-2015 (Gemery et al., 2017), and Arctic Ostracode Database-  
388 2020 (Cronin et al., 2021) (Figure 2B). The online USGS version of the Arctic Ostracode  
389 Database was contributed by Thomas M. Cronin, Thomas R. Holtz, Elisabeth M. Brouwers,  
390 William M. Briggs, Robin C. Whatley and Adrian Wood (Cronin et al., 1995). Some of the  
391 datasets in the AOD are also present in other databases, such as PANGAEA and OBIS. Through

392 the Integrated Publishing Toolkit, OBIS (OBIS Secretariat, 2021) also harvested the Modern  
393 Arctic Ostracode Database (Cronin et al., 2010) and the Arctic ostracod datasets in PANGAEA  
394 database (see Section 5.2.). Note that the census counts are flattened to occurrence data (i.e.,  
395 presence/absence) in the OBIS.

396 The latest version, AOD-2020, incorporates the previous versions and provides  
397 georeferenced occurrence and species census data for 96 species of benthic marine ostracods  
398 from over 1500 modern surface sediment samples (Cronin et al., 2021; Gemery et al., 2021a;  
399 Gemery et al., 2021b). This unique database grants us an opportunity to analyze and understand  
400 Arctic ostracods beyond the scale of a local project. To gain an overview, we applied a network-  
401 based clustering analysis to the AOD-2015 census data (see Gemery et al., 2017 for detailed  
402 species information and biogeography). The results suggest two biogeographical clusters (Figure  
403 2A): a deep ocean cluster is distributed in the open ocean at latitudes of 70 °N and higher, and a  
404 shallow shelf cluster in the shallow marine coastal areas at latitudes of 80 °N and lower.

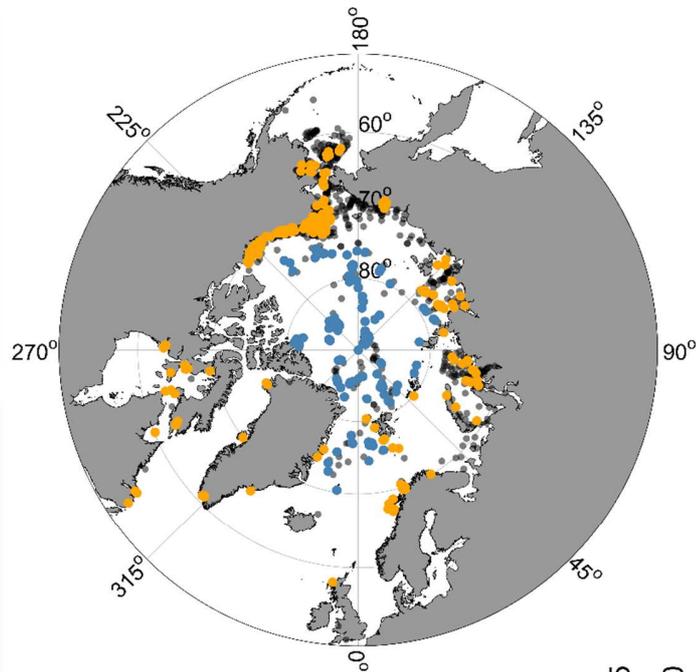
A

**Deep ocean cluster taxa**

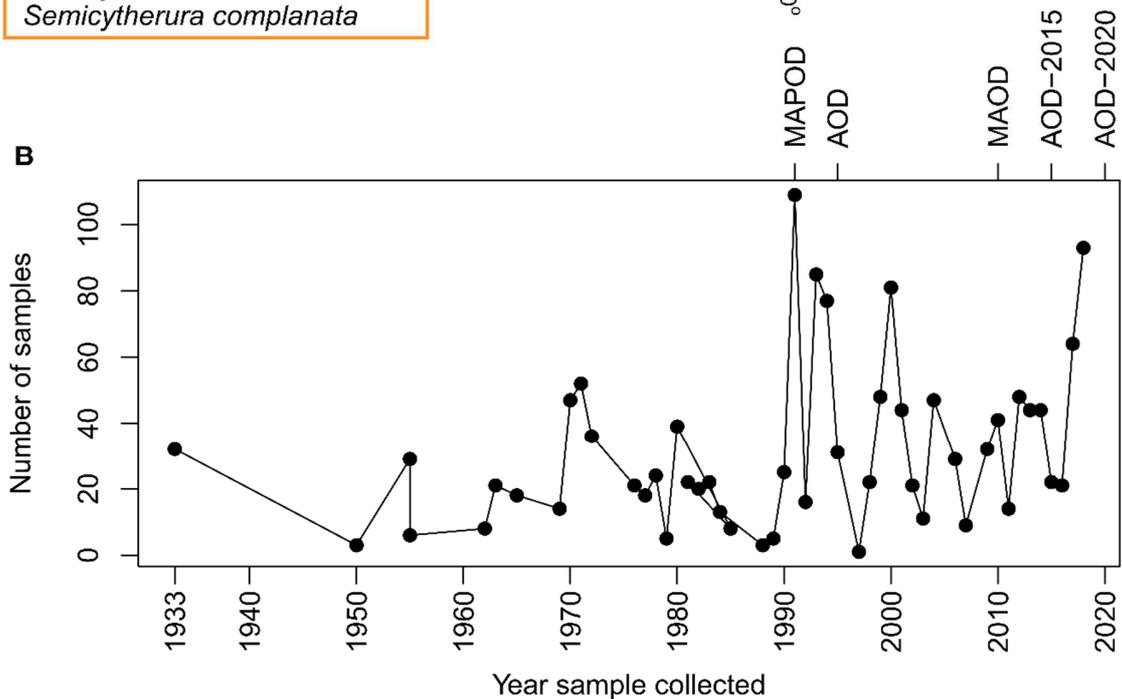
*Acetabulastoma arcticum*  
*Cytheropteron higashikawai*  
*Cytheropteron parahamatum*  
*Cytheropteron scoresbyi*  
*Cytheropteron sedovi*  
*Henryhowella asperrima*  
*Krithe hunti*  
*Paracytherois* spp.  
*Polycopse* spp.  
*Pseudocythere caudata*

**Shallow shelf cluster taxa**

*Acanthocythereis dunelmensis*  
*Finmarchinella barentzovoensis*  
*Heterocyprideis sorbyana*  
*Normanicythere leioderma*  
*Palmenella limicola*  
*Paracyprideis pseudopunctillata*  
*Sarsicytheridea bradji*  
*Semicytherura complanata*



B



405  
406  
407  
408  
409  
410  
411

**Figure 2.** Arctic Ostracode Database (AOD): Major Arctic taxa and the sampling history. (A) Two clusters determined by the network-based clustering analysis. The polar deep ocean taxa (listed in the blue box) are relatively more abundant and frequently occur in the deep ocean cluster (blue points on the map), and the same applies to the shallow shelf taxa (in the orange box) in the shallow shelf cluster (orange points on the map). Note that *Acetabulastoma arcticum* is a parasitic species living on sea-ice dwelling amphipods. Only samples with more than 100

412 specimens in the AOD-2015 version were used in the clustering analysis; the black points are the  
413 AOD samples that are not used in the analysis. **(B)** Number of samples sorted by the year of  
414 sampling. In parallel to the x-axis, the AOD versions are indicated by their published year:  
415 Modern Arctic podocopid ostracode database (MAPOD in Cronin et al., 1991); Arctic Ostracode  
416 Database (AOD in Cronin et al., 1995); Modern Arctic Ostracode Database (MAOD in Cronin et  
417 al., 2010); Arctic Ostracode Database-2015 (AOD-2015 published online in 2015 (Gemery et al.,  
418 2017)); Arctic Ostracode Database-2020 (Cronin et al. 2021).

419

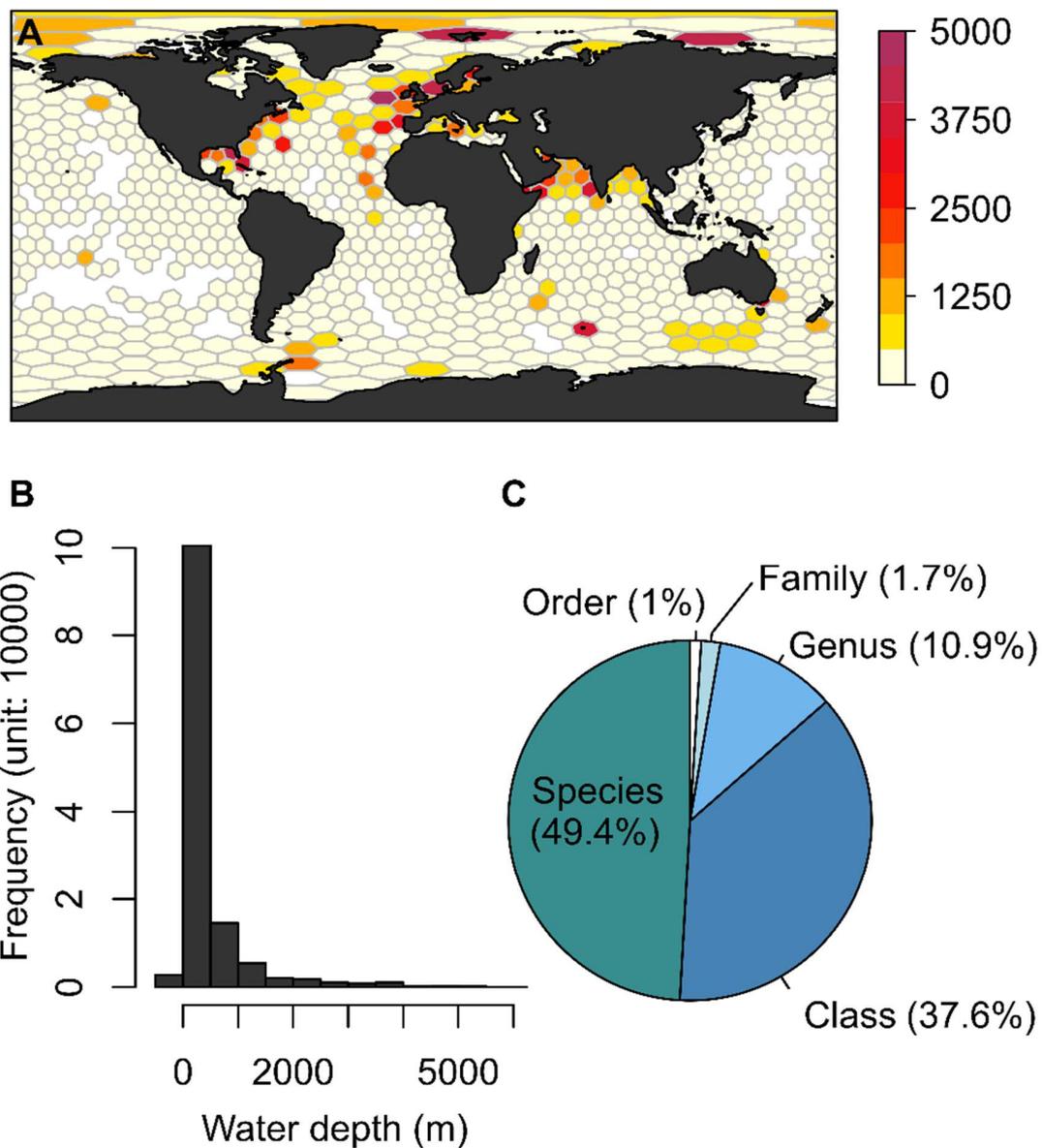
## 420 **5.2. Ocean Biogeographic Information System**

421 The Ocean Biogeographic Information System (OBIS) is dedicated to occurrence  
422 records, event records, and measurements/facts of modern marine life. It started as one initiative  
423 of the Census of Marine Life (2000-2010) and had been adopted into the UNESCO's  
424 International Oceanographic Data and Information (IODE) program since 2010. OBIS gathered  
425 data through Integrated Publishing Toolkits from more than 20 OBIS regional/country/thematic  
426 nodes (e.g., Arctic OBIS, OBIS Germany, Fish OBIS), which connect hundreds of institutional  
427 data publishers or data archives (e.g., PANGAEA, NOAA's NCEI, USGS). The Integrated  
428 Publishing Toolkit, developed by GBIF, assists with formatting the data to valid Darwin Core  
429 terms and describing dataset metadata in Ecological Metadata Language. This readily allows  
430 interoperability of datasets. OBIS reached 0.5 million records in 2002, and over 63 million  
431 occurrence records in 2020 (obis.org), together with geographic and bathymetric distribution and  
432 environmental data (e.g., salinity, temperature).

433 To ensure the utility of a system such as OBIS, which integrates datasets from multiple  
434 sources that were originally compiled in a variety of circumstances and for different purposes, a  
435 certain amount of data and metadata standardization is essential (OBIS Manual:  
436 <https://www.obis.org/manual/>). This may be achieved by having mandatory fields that must be  
437 completed for the dataset to be accepted. For example, biogeographical data in OBIS requires  
438 locations to be defined by latitude and longitude coordinates in decimal degrees to facilitate

439 mapping of records, so inclusion of datasets that use a different coordinates system such as  
440 Universal Transverse Mercator (UTM) requires conversions. Another example is taxonomic  
441 harmonization, which in the case of OBIS is achieved through the matching of names to  
442 authoritative taxonomic lists such as the WoRMS (see section 4).

443         OBIS focuses on modern marine life, but its ostracod data include both modern records  
444 from various types of samples and Quaternary records from core section sediments. At the time  
445 of writing in 2020, it houses nearly 205,000 records for 2,787 ostracod taxa (2,415 species),  
446 including both benthic and planktonic taxa. WoRMS provides the taxonomic backbone to OBIS,  
447 and about 49.4% of ostracod records were identified to species level in OBIS (Figure 3C).  
448 Ostracod data show global coverage in OBIS, but there is a strong disparity in sampling efforts  
449 between different oceans and between coastal and open oceanic regions. The areas of best  
450 coverage concentrate mostly in: (1) the northern hemisphere, and particularly the North Atlantic  
451 (versus the southern hemisphere and the Pacific Ocean; Figure 3A); (2) coastal regions (versus  
452 open ocean and deep ocean; Figure 3B).



453

454 **Figure 3.** Ocean Biogeographic Information System (OBIS): global records of marine and  
 455 brackish ostracods. **(A)** Geographical distribution of the records in a heat-map style. **(B)** Water  
 456 depth distribution of the records. **(C)** Identification resolution of the records. The map shows the  
 457 discrepancies in number of records across oceans in the OBIS. Note the intensely studied  
 458 regions: European and North American waters; and the generally understudied areas: Pacific,  
 459 Indian, and Southern Oceans. The Arctic Ocean is relatively rich in ostracod data which were  
 460 harvested from PANGAEA and the Arctic Ostracode Database (see Section 5.1.). Most records  
 461 come from 0–500 m water depths. Most records (49.4%) had identification resolution down to  
 462 the species level, but many records (37.6%) were just “Ostracoda”.

463  
464 PANGAEA is an important contributor to the ostracod data in OBIS. Hosted by the  
465 Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI) and the  
466 Center for Marine Environmental Sciences (MARUM), PANGAEA is a mainstream data  
467 archive/repository for georeferenced data from earth system research, e.g., Deep Sea Drilling  
468 Program (DSDP, 1967-1983), Ocean Drilling Program (ODP, 1984-2002), Integrated Ocean  
469 Drilling Program (IODP, 2003-2013), the International Ocean Discovery Program (IODP, 2013-  
470 2023), and many other research projects of various scales. Despite its focus on modern marine  
471 ecosystems, OBIS also harvested paleontological datasets from PANGAEA.

472 Endorsed by the Scientific Committee on Antarctic Research (SCAR), the Antarctic  
473 Biodiversity Information Facility (ANTABIF) exchanges biogeographical data of marine species  
474 with the Antarctic Thematic Node of OBIS, the AntOBIS, and provides data ultimately to GBIF.  
475 ANTABIF is a complex platform focused on continental and marine research in/off Antarctica,  
476 and it houses a diverse collection of oceanographical, ecological, biological and geological  
477 datasets (e.g., SCAR Southern Ocean Diet and Energetics Database), as well as providing online  
478 analysis tools for the scientific community (e.g., interactive identification keys, R packages  
479 relevant to Antarctic and Southern Ocean science). Concerning the biogeographical data  
480 exchanged with OBIS, currently ANTABIF hosts over 190 datasets of many taxonomic groups,  
481 including Southern Ocean Ostracoda (excluding Halocypridina) (Brandão, 2012), which focuses  
482 on benthic ostracods. These datasets can be accessed through the Antarctic Biodiversity Portal  
483 (biodiversity.aq, 2021), which integrates datasets compiled under the SCAR-MarBIN (SCAR-  
484 Marine Biodiversity Information Network) initiative with data from different sources, including  
485 the Australian Antarctic Division and other institutions. The Southern Ocean benthic Ostracoda  
486 dataset (Brandão, 2012) includes 888 occurrences of 113 taxa from 193 georeferenced locations.

487 Planktonic Southern Ocean ostracods were covered by one atlas, which offer the taxonomy,  
488 geographical distribution (static occurrence maps), illustrations and measurements of key  
489 morphological features of species (mostly Halocypridina, but all known planktonic Cypridinida  
490 are also present; see the next section).

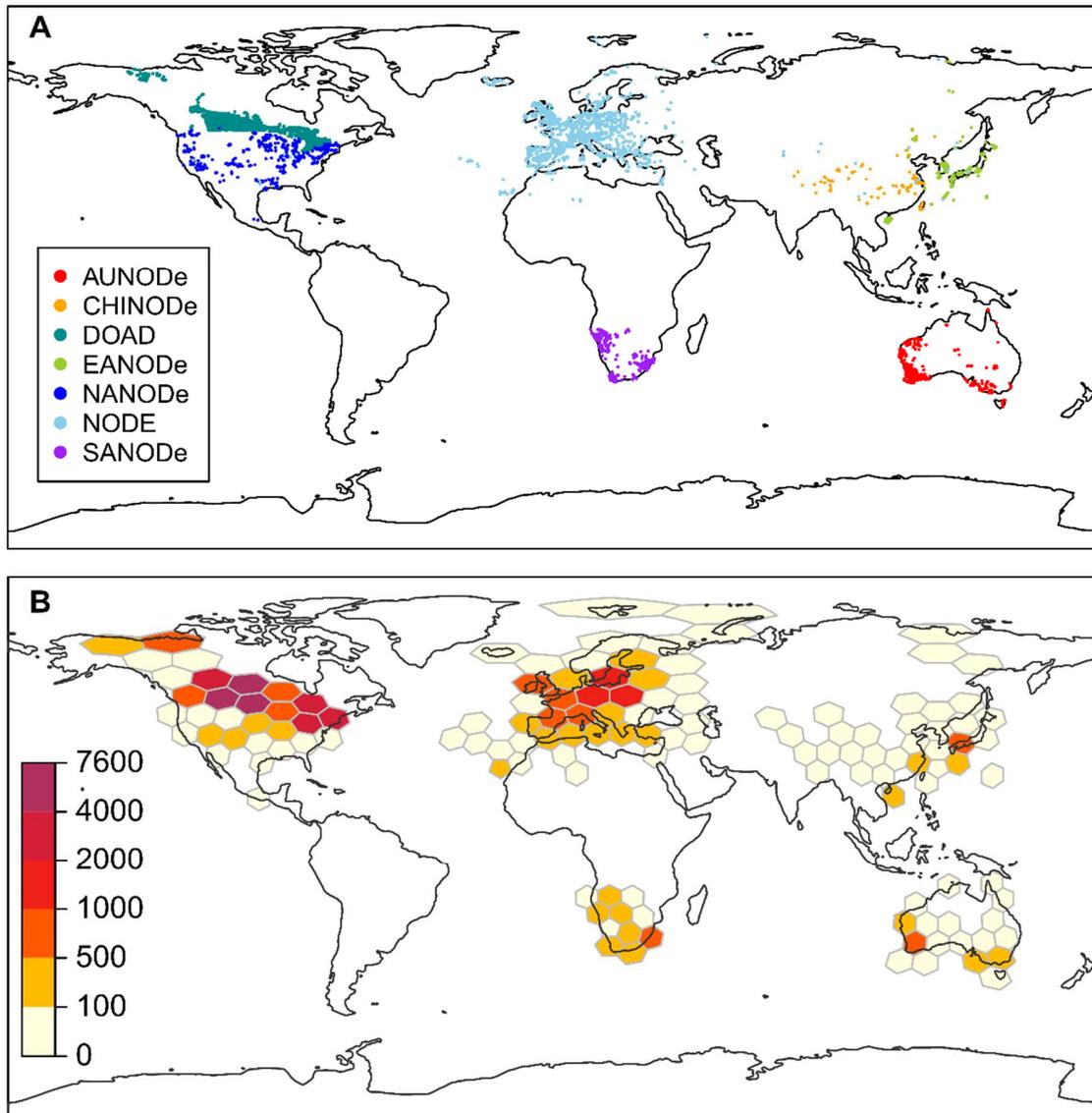
### 491 **5.3. Marine planktonic ostracods**

492 Martin Angel, Kasia Blachowiak-Samolyk, and Vladimir Chavtur compiled the available  
493 information on marine planktonic ostracods from the Atlantic and Southern Oceans, and  
494 published the “Atlas of Atlantic Planktonic Ostracods” (Angel et al., 2008) and “An Atlas of  
495 Southern Ocean Planktonic Ostracods” (Blachowiak-Samolyk and Angel, 2008). The Atlantic  
496 compilation (Angel et al., 2008) includes data from before the Challenger expedition (1870s) up  
497 to 2007, while the Southern Ocean dataset (Blachowiak-Samolyk and Angel, 2008) include data  
498 from the Discovery Investigations (1930s) up to 2007. These two compilations include detailed  
499 taxonomic and morphological (e.g., size, shape) data, static maps of geographical distribution,  
500 illustrations and measurements of important morphological characters, and list of bibliographies,  
501 all in pdf format. This means that although these are valuable resources on planktonic ostracods,  
502 specific data on the occurrences (latitude, longitude, depth, date of collection, etc.) cannot be  
503 extracted electronically from these atlases, neither it is possible to exchange data in a machine-  
504 readable format.

505

506 **6. Non-marine occurrence-based (paleo)biodiversity databases**

507         There are several regional databases focused on the ecology and distribution of living  
508 non-marine ostracods. Prominent among these are NODE (Non-marine Ostracod Distribution in  
509 Europe), NANODE (North American Non-marine Ostracode Database), DOAD (Delorme  
510 Ostracode Autecological Database), and EANODE (East Asia Non-marine Ostracod Database);  
511 key metadata and data from these and other regional databases are compiled in OMEGA  
512 (Ostracod Metadatabase of Environmental and Geographical Attributes). Datasets from DOAD  
513 and NANODE have been amalgamated into a North American Combined Ostracode Database  
514 (NACODE) as well as being made accessible through the Neotoma Paleoecology Database.  
515 Details of each of these are described briefly below. Distribution maps of all records from the  
516 above databases are in Figure 4.



517  
 518 **Figure 4.** Non-marine occurrence-based (paleo)biodiversity databases. **(A)** Geographical  
 519 distribution of all the sites in the Ostracod Metadatabase of Environmental and Geographical  
 520 Attributes (OMEGA), color-coded according to the contributing databases: Australian Non-  
 521 marine Ostracode Database (AUNODE), Chinese Non-marine Ostracode Database (CHINODE),  
 522 Delorme Ostracode Autecological Database (DOAD), East Asia Non-marine Ostracod Database  
 523 (EANODE), North American Non-marine Ostracode Database (NANODE), Non-marine  
 524 Ostracod Distribution in Europe (NODE) database, South African Non-marine Ostracod  
 525 Database (SANODE). **(B)** A heatmap of all occurrence records from (A); note the higher  
 526 concentrations of records in North America and Europe.

528 **6.1. Non-marine Ostracod Distribution in Europe**

529           The Non-marine Ostracod Distribution in Europe (NODE) database contains  
530 approximately 10,000 records of living ostracod species (plus about 2,000 Pleistocene and  
531 Holocene fossil records), representing more than 400 species and approximately 2,500 localities.  
532 It was initiated by six research teams in five countries to map the distribution of ostracod sex and  
533 parthenogenesis for the three-year (1994-1996) EU Human Capital and Mobility Programme  
534 project *Evolutionary ecology of reproductive modes in non-marine Ostracoda* (Horne et al.,  
535 1998) and subsequently developed further by David J. Horne as its Data Steward (i.e., the person  
536 responsible for all aspects of database management including quality control and accessibility).  
537 Focused on Europe, its geographical coverage extends longitudinally from the Azores to east of  
538 the Black Sea and latitudinally from the Canary Islands to Svalbard (approx. 32°W – 45°E and  
539 25–80°N). A few additional North American and Asian records are currently included in NODE  
540 for calibration purposes in connection with Quaternary paleoclimatic reconstruction (see  
541 discussion below); these may eventually be transferred to other databases. The records in  
542 NODE, compiled almost entirely from scientific literature published in the 19<sup>th</sup>, 20<sup>th</sup> and 21<sup>st</sup>  
543 centuries, are focused on occurrences of identified ostracod species but include (where possible)  
544 information such as water body type, date of collection, and whether males were found.

545 **6.2. North American Non-marine Ostracode Database**

546           The North American Non-marine Ostracode Database (NANODE) was initiated by  
547 Richard M. Forester at the U.S. Geological Survey. Field collection efforts were conducted  
548 primarily by Forester et al. (2005), and data subsequently managed by Alison J. Smith as Data  
549 Steward. The dataset ([www.kent.edu/nanode](http://www.kent.edu/nanode)) is composed of late 20<sup>th</sup> to early 21<sup>st</sup> century  
550 collections (1979 through 2005) and contains approximately 2,600 records comprising *c.* 100

551 species and *c.* 600 localities with accompanying major ion hydrochemical and field limnological  
552 data collected at the same time as the species collection (Forester et al., 2005). It provides a well-  
553 distributed coverage of the conterminous states of the USA from California in the west to New  
554 Hampshire in the east (approx. 124–70°W and 29–49°N) although the southeastern states  
555 (Florida to the Carolinas) are not yet represented. The distribution of each of the species in  
556 NANODe is shown in geographic maps and in solute graphs ([www.kent.edu/nanode](http://www.kent.edu/nanode)). The data  
557 are completely accessible through the multiproxy Neotoma Paleoecology Database (see below).  
558 The slide collection is housed in the Geology Department at Kent State University, Kent, Ohio  
559 (USA).

### 560 **6.3. Delorme Ostracode Autecological Database**

561 The Delorme Ostracode Autecological Database (DOAD), based on primary collection  
562 by Denis Delorme during the 1960s and 1970s, includes over 30,000 records representing *c.* 130  
563 species and more than 6,000 localities, together with extensive geographical, environmental, and  
564 climatic data for the waterbodies sampled. Both the database and a very substantial collection of  
565 voucher specimens (representing almost every record in the database) are now housed and  
566 curated by the Canadian Museum of Nature at its Natural Heritage Campus (Research and  
567 Collections Facility) in Gatineau, Canada, together with a collection of type and illustrated  
568 ostracod specimens (related to Delorme's publications) that was formerly kept at the Geological  
569 Survey of Canada in Ottawa. A high density of records covers Canada south of latitude 60°N  
570 from southeastern British Columbia in the west to the Great Lakes region and as far as  
571 southwestern Quebec in the East (approx. 120–70°W and 41–60°N); there are also records in  
572 Yukon Territory and Northwest Territories (approx. 140 –128°W and 65–70°N).

573

574 **6.4. East Asia Non-marine Ostracod Database**

575           The East Asia Non-marine Ostracod Database (EANODe), compiled and managed by  
576 Robin J. Smith as its Data Steward, contains approximately 1,700 records representing c. 150  
577 species and c. 650 localities, of which the majority are in Japan but including South Korea,  
578 China and the Far Eastern Federal District of Russia (approx. 108 –145°E and 18–72°N). The  
579 data are mainly from primary collections represented by specimens curated at the Lake Biwa  
580 Museum in Japan (c. 75%), the remainder being compiled from published literature (25%).

581 **6.5. Ostracod Metadatabase of Environmental and Geographical Attributes**

582           A single global database of non-marine ostracod distribution would be of immense  
583 scientific value, but creating one would be a very challenging, long-term project. As a more  
584 pragmatic approach, with short-term as well as long-term benefits, we initiated the OMEGA:  
585 Ostracod Metadatabase of Environmental and Geographical Attributes (Horne et al., 2011), with  
586 the aim of compiling and maintaining a metadatabase of regional non-marine ostracod databases.

587           It is important to understand that OMEGA is not a merging of databases to form a “super-  
588 database”. A metadatabase contains data about databases, and in this sense OMEGA might  
589 simply be a database of existing databases that contain information on ostracod ecology and  
590 distribution in different parts of the world. However, the most important data (taxonomic names  
591 and coordinates of records) are included as metadata, thus facilitating the mapping and searching  
592 of records from all of the contributing databases. Each record in OMEGA is attributed to a  
593 source database for which contact details will be supplied. It is hoped that, ultimately, links could  
594 be included to facilitate access to the regional databases; a pre-requisite for such a functionality,  
595 however, is that all contributing databases should be freely accessible online.

596           The OMEGA metadatabase is therefore more than just a convenient way of locating and  
597 accessing other databases, it constitutes a major research tool in its own right, enabling the  
598 visualization and analysis of distributions on a global scale. At the time of writing, OMEGA  
599 includes metadata (approx. 49,000 records) from NODE, NANODE, DOAD and EANODE as  
600 well as other datasets representing southern Africa (supplied by Koen Martens), Australia  
601 (supplied by Chris Gouramanis) and China (supplied by Yangmin Qin). A partial dataset of c.  
602 26,000 records compiled from NODE, DOAD and NANODE has been available for download  
603 since 2015 (via the GBIF web site:  
604 [https://data.freshwaterbiodiversity.eu/ipt/resource.do?r=bf\\_cf15](https://data.freshwaterbiodiversity.eu/ipt/resource.do?r=bf_cf15)), but data standards must be  
605 applied and very substantial taxonomic harmonization and validation of georeferenced localities  
606 remain to be completed before other data can be released (Horne, 2014).

#### 607 **6.6. Neotoma and North American Combined Ostracode Database**

608           In 2009, a public access community-curated data resource became available, an  
609 international, collaborative database named Neotoma (named for packrats of the genus *Neotoma*)  
610 (Grimm et al., 2018; Williams et al., 2018). Neotoma ([www.neotomadb.org](http://www.neotomadb.org)) was developed to  
611 house, and access independently, cohesive paleoecological datasets of Pliocene through Modern  
612 age. The need for such an international resource with data stewards linked to databases and  
613 collections was driven by the growing problem, for many researchers, of maintaining funding to  
614 sustain and grow independent regional databases. A range of biological data for multi-proxy  
615 analysis can be found in Neotoma, including data access to pollen, vertebrates, diatoms and  
616 many other proxies for past environments and climates. These independent datasets retain their  
617 cohesiveness and are managed by data stewards associated with the datasets. The entire contents  
618 of NANODE and a large subset of DOAD are accessible in Neotoma. For DOAD, 4,053

619 georeferenced sites at which living ostracods were collected were ported into Neotoma,  
620 representing a subset of the 6,719 sites in DOAD (of which many recorded empty shells only).  
621 This allows a continental-scale biogeographic view of sites in which living ostracods paired with  
622 major ion hydrochemistry and limnologic data can be studied by examining the Canadian and  
623 U.S. datasets together. Additionally, numerous paleolimnologic ostracod records of  
624 Plio/Pleistocene and Holocene age are also housed in Neotoma. Amalgamated datasets from  
625 DOAD and NANODe, forming a North American Combined Ostracode Database (NACODE),  
626 were mapped and analyzed by Curry et al. (2012).

627

## 628 **7. Paleobiology Database**

629         The PBDB is an occurrence-based paleobiodiversity database, dedicated to fossil data of  
630 all time. PBDB collaborates closely with other important databases, e.g., GBIF (see Section 8),  
631 the Neotoma Database (see Section 6.6.), iDigBio, Macrostrat, Earth Life Consortium, and  
632 ePANDDA API. The freely downloadable data types from PBDB include georeferenced  
633 occurrences, specimens and measurements, geological strata, collections, diversity over time,  
634 taxa, opinions, and bibliographic references. PBDB relies on its authorized data enterers to enter  
635 data that mainly come from the published literature. For example, the top enterers for ostracod  
636 data entries are Matthew E. Clapham and John Alroy at the time of writing. Its fossil  
637 representation is partially influenced by what taxonomic groups the enterers focus on.

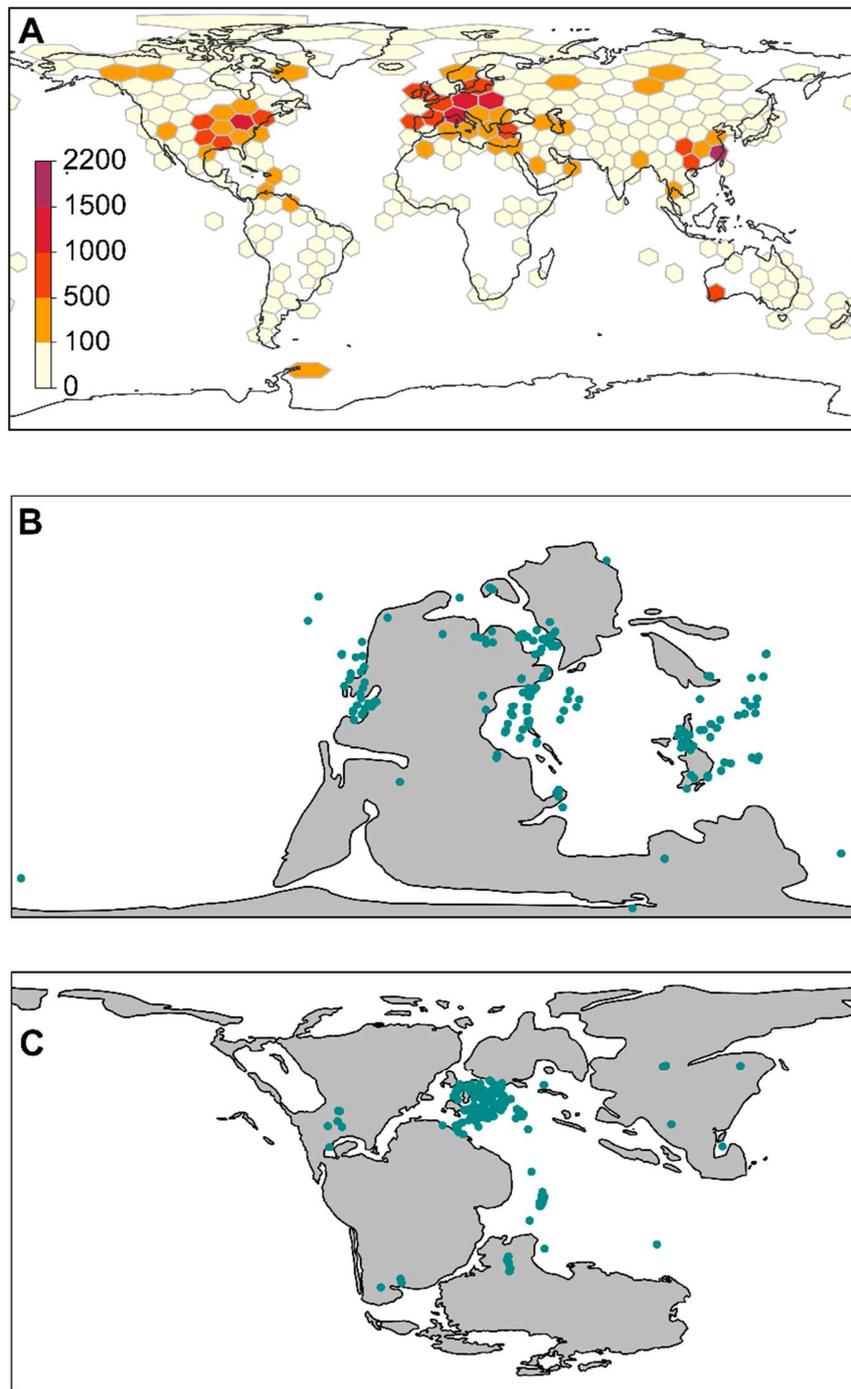
638         The PBDB has its own taxonomy system that is composed of authorities (i.e., references  
639 of taxonomic names combined with the authority) and opinions (i.e., references on the status of  
640 names and the relationships between names) (Peters and McClennen, 2016). An algorithm is  
641 used to produce the working taxonomy and to minimize influence from personal opinions (see

642 the detailed algorithm in Peters and McClennen, 2016). The users could only influence the  
643 taxonomy by changing the rank of the opinion basis. Opinion references are ranked by their basis  
644 (“stated with evidence”, “stated without evidence”, “implied”, and “second hand”) and then by  
645 published year. The algorithm then selects the highest-ranked opinions for classification and  
646 creates the working taxonomy by using the information in the selected opinions: the reason of  
647 name structure or spelling (e.g., “original spelling”, “recombination”, “rank change”,  
648 “misspelling”) and name status (e.g., “belongs to”, “synonym of”, “nomen nudum”). Thereafter,  
649 occurrence data are dynamically linked to the constantly growing taxonomy database, so the  
650 taxonomy can be instantly updated. Database users can apply various filters that deal with open  
651 nomenclature, taxonomic reidentification, and updates on the accepted names. The taxonomic  
652 treatment is fully archived and downloadable. Together with WoRMS and about 100 taxonomic  
653 databases, PBDB also supplies its checklist dataset to GBIF (Paleobiology Database, 2021).

654 Collaborating with the GPlates software team (Müller et al., 2018), PBDB has a default  
655 plate motion model for calculating the paleocoordinates for data that have present-day  
656 coordinates and geological ages. The previous default model (until 2013) was provided by  
657 Christopher Scotese, and it is still available in the PBDB. Database users can toggle between  
658 Scotese’s model and the current default model (since 2014), which is called the GPlates model  
659 (Wright et al., 2013) and is considered preferable in terms of accuracy.

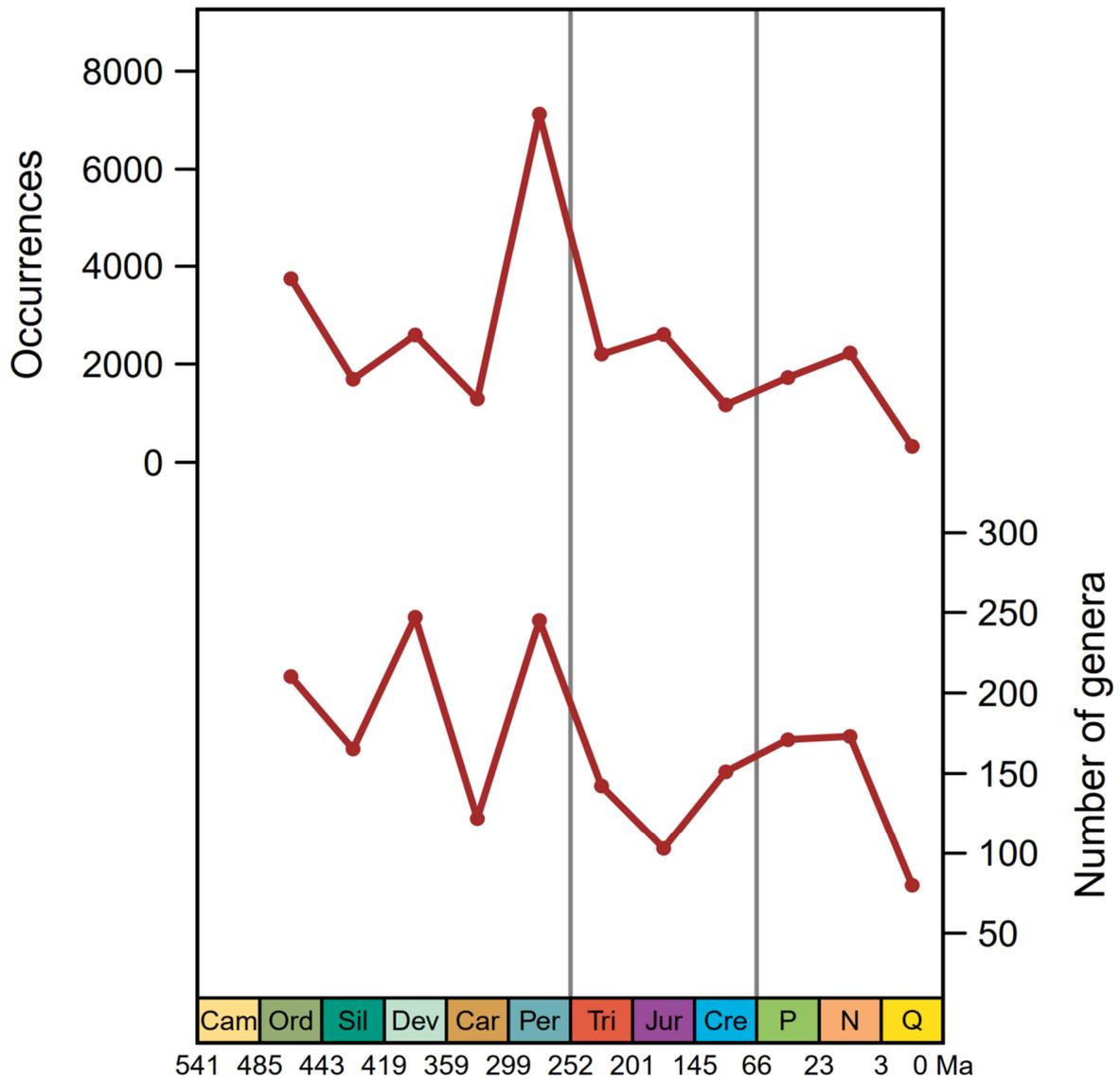
660 PBDB has accumulated over 1.4 million occurrence records since its foundation in 1998.  
661 There are in total over 26,000 records of ostracods (Figure 5A), the majority of which are marine  
662 ostracods. Ostracoda has fairly good number of occurrences with more than 100 genera in most  
663 geological periods but show decreasing numbers of data entries from the Paleozoic (63%) to the  
664 Mesozoic (20%) and the Cenozoic (17%) (Figure 6). The most common levels of taxonomic

665 resolution are genus (53%), species (39%), and class (5%). When cross checked with the WOD,  
666 about 20% of the accepted ostracod names from the PBDB are not currently registered in the  
667 WOD, and about 3% are marked as unaccepted in 2020. In the following analysis, we retain the  
668 classification in the PBDB.



669  
 670 **Figure 5.** Paleobiology Database (PBDB) ostracod records. (A) Geographical distribution of all  
 671 records shown in a heatmap style. Total number of occurrence records: 21,111; number of  
 672 locations: 2581. (B) Permian records in a paleogeographic map (275 Ma), (C) Jurassic records in  
 673 a paleogeographic map (170 Ma). The maps are based on Scotese and Wright (2018).

674



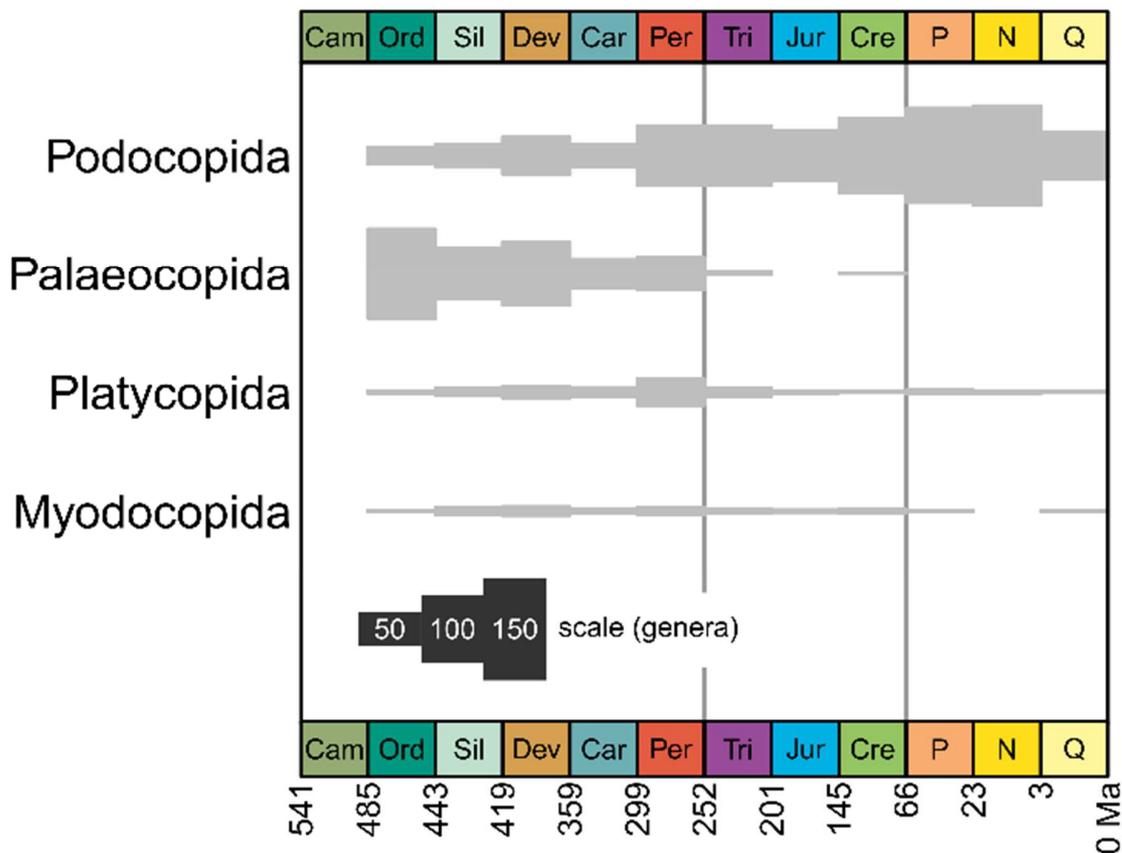
675

676 **Figure 6.** Paleobiology Database: Number of ostracod occurrences and genera in the  
 677 Phanerozoic. Abbreviations: Cambrian (Cam), Ordovician (Ord), Silurian (Sil), Devonian (Dev),  
 678 Carboniferous (Car), Permian (Per), Triassic (Tri), Jurassic (Jur), Cretaceous (Cre), Paleogene  
 679 (P), Neogene (N), Quaternary (Q), million years ago (Ma).

680

681 Within Ostracoda, Podocopida, Palaeocopida, Platycopida, and Myodocopida have fossil  
 682 records in the PBDB (Figure 7). Podocopida and Palaeocopida have more fossil records than  
 683 Platycopida, and Myodocopida. Palaeocopida is predominantly Paleozoic, whereas Podocopida  
 684 increasingly dominate the fossil records towards the Cenozoic. These general trends are  
 685 consistent with the chart presented in Figure 20.14 in Armstrong and Brasier (2004) which was  
 686 based on the data from Whatley et al. (1993).

687



688

689 **Figure 7.** Paleobiology Database: Number of genera per major orders of ostracods in the  
 690 Phanerozoic. Abbreviations follow Figure 6.

691

692 **8. GBIF**

693 GBIF is an intergovernmental collaboration between over 100 official  
694 country/economy/organization participants (represented by the “nodes” in GBIF) that coordinate  
695 their own institutions and databases (called the “publishers”) to share biodiversity data. The  
696 publishers share data through Integrated Publishing Toolkits which ensures the application of  
697 metadata standards, including Darwin Core and Ecological Metadata Language (see  
698 Introduction), on their datasets. Depending on the dataset nature, the publisher can choose from  
699 four dataset classes in GBIF: metadata-only dataset for resources that have not been digitalized,  
700 taxonomic checklist, occurrence dataset, and sampling event dataset for data collected in greater  
701 details and with some protocol. Ingesting all kinds of data from data publishers, GBIF is the  
702 world’s largest biodiversity database in terms of partnership and amount of data.

703 Both GBIF’s taxonomy and occurrence systems aggregate data from its extensive  
704 network of nodes and data publishers. The GBIF Backbone Taxonomy is assembled based on  
705 about 100 taxonomic checklist sources (GBIF Secretariat, 2021). The top providers of taxonomic  
706 names include Catalogue of Life Checklist, International Barcode of Life project (iBOL)  
707 Barcode Index Numbers (BINs), Systema Dipterorum, The World Checklist of Vascular Plants  
708 (WCVP), WoRMS, and the PBDB. GBIF is designed for modern records and the majority of the  
709 GBIF occurrences are labeled as human observation. Many of the other occurrences are labeled  
710 as preserved specimens that come from botanical gardens and museums, e.g., departments of  
711 Botany, Entomology, Invertebrate Zoology and Vertebrate Zoology in the Smithsonian  
712 Institution. Although GBIF is not designed for paleobiological data, it has ingested occurrence  
713 datasets of fossil specimens from databases and museums, e.g., the Paleobiology Department in

714 the Smithsonian Institution. These data are labeled as fossil specimens, but their geological ages  
715 have not been incorporated and their “year” column is blank in GBIF.

716 GBIF has over 376,000 occurrence records of ostracods at the time of writing. Preserved  
717 specimens account for ~43% of these records, indicating that museums are the biggest data  
718 sources of ostracods in GBIF. Fossil specimens account for ~36% of these records, and the top  
719 contributors are PANGAEA, PBDB, Smithsonian National Museum of Natural History, and  
720 Natural History Museum London. The rest are mainly labeled as human observation or material  
721 sample. Being the largest data aggregator, GBIF offers a convenient portal for checking what  
722 resources are available with which museums/institutions/databases and can potentially help  
723 ostracod researchers design new projects. On the other hand, users should be cautious about  
724 various issues that may result from aggregating a big number of datasets from different contexts.

## 725 **9. Discussion: challenges and future directions**

### 726 **9.1. Taxonomic identification and harmonization**

727 Regional datasets collected by one person or group may be internally consistent in terms  
728 of taxonomy and positional accuracy, but if they are to be utilized in a wider context in  
729 combination with other databases, then data standards must be established, and taxonomic  
730 harmonization is required. Geographically more extensive, literature-based datasets are likely to  
731 cover more of species' full distributions but, since they represent collections by many different  
732 people over many years, are more prone to inconsistencies in taxonomy as well as locational  
733 precision and accuracy. If publications include illustrations of the identified species, they can be  
734 checked and corrected where necessary, paying particular attention to synonyms and generic  
735 assignments; if unillustrated, assumptions must be made about the validity of the taxonomic

736 names listed. The use of open nomenclature, often inconsistently, needs to be dealt with by either  
737 new identifications to species level or applying uncertainty filters in the database.

738         In the marine realm, AOD, OBIS, and PBDB display different approaches for taxonomic  
739 harmonization, corresponding to their different database scopes and architectures. Our  
740 investigation showed that AOD is one unique occurrence-based database that comes with high  
741 spatial coverage and fully harmonized taxonomy (albeit for a limited geographical region). This  
742 results from its focus on the Arctic ostracods and contribution from the same group of closely  
743 collaborated researchers. Taxonomic harmonization is more challenging for OBIS because it is a  
744 data harvester that recombines a vast amount of modern and Quaternary datasets from various  
745 data publishers at the global scale. Because occurrence data themselves contain no systematics  
746 information, OBIS uses WoRMS as its taxonomic backbone. PBDB also faces a more  
747 challenging task in taxonomic quality control, hosting fossil records at a longer geological time  
748 scale. It has an internal taxonomy database that uses an objective algorithm to combine  
749 authoritative names and taxonomic opinions.

750         For non-marine ostracod databases, the subjective global checklist of extant non-marine  
751 ostracod species published by Meisch et al. (2019) provides a standard against which all database  
752 records can be checked. Taxonomic harmonization efforts are currently focused on northern  
753 hemisphere (Holarctic) databases, supported at generic level by the recent publication of keys for  
754 the Nearctic (Smith and Horne, 2016) and Palaeartic (Horne et al., 2019) regions. Simple  
755 comparisons of species lists from DOAD, NANODE, NODE and EANODE suggest that at least  
756 30 species are common to North America and Europe, and 10 of these are also common to East  
757 Asia. However, detailed investigations (in progress) show that while some are truly

758 cosmopolitan, in many cases species listed under the same name turn out to be different species  
759 in different regions, while others are listed under different names that could be synonymized.

760 Another good solution is consulting an authoritative taxonomic database for taxonomic  
761 identification, taxonomic harmonization, and resolving outdated taxonomy. There is no silver  
762 bullet for taxonomic identification, and we will always need specialist expertise to conduct high  
763 standard taxonomic identification of ostracods. Nonetheless, a free, online database can strongly  
764 accelerate the traditionally slow work on taxonomic identification, and certainly coordinate  
765 updates in taxonomic opinions and consensus. Thanks to Kempf and many database contributors,  
766 valuable taxonomic resources have become available for current and future young scientists. The  
767 WOD is arguably the most promising platform for resolving the taxonomic challenges discussed  
768 above. WOD hosts freely available taxonomic information, descriptions, and illustrations for  
769 marine/non-marine and fossil/Recent ostracods. The collaboration between WoRMS and OBIS is  
770 an excellent example of how WOD can improve taxonomic quality in occurrence-based  
771 databases. Since all kinds of data in WOD are entered by experts, WOD needs more  
772 contributions from ostracodologists.

773 If a larger number of specialists invest only a few hours per week in WOD we would in  
774 one or a few years have as much freely available information as possible with copyright issues  
775 resolved on ostracod taxonomy and geographical and stratigraphical distribution. The specialists'  
776 contributions include (1) adding newly described species, registering new combinations and new  
777 classifications; (2) synonymies accompanied by references, uploading pdfs of publications  
778 related to ostracods; (3) scanning electron microscope photographs and other illustrations of each  
779 species, especially the type taxa. Another important feature missing in WOD is the publication of  
780 a higher (suprageneric) classification for ostracods by updating Whatley et al. (1993) and Horne

781 et al. (2002). This was planned in the WOD editors' workshop in 2013, but still needs quite a lot  
782 of work. As differing opinions are common in taxonomy, all above tasks require careful  
783 coordination among the participating specialists to reach consensus opinions.

784

## 785 **9.2. Validation of locality and sediment core section**

786 Before they are made more widely available, databases that contain occurrence records  
787 should be validated by checking taxonomic and locality data for errors and making corrections  
788 where necessary. Most late 20<sup>th</sup> and 21<sup>st</sup> century publications provide adequate locality  
789 information, often including coordinates, while older literature frequently has shortcomings in  
790 this respect. Abstracting occurrence data from European literature for NODE, for example, has  
791 encountered many challenges including historical revision of political boundaries and place  
792 names, waterbodies from which ostracods were collected in the 19<sup>th</sup> century that were  
793 subsequently drained and no longer exist, and imprecise locality information (such as records  
794 from large lakes that do not specify a site within the lake, or small ponds only located by  
795 reference to the nearest town or village). Such issues can make compliance with data standards  
796 of integrative systems (e.g., OBIS, GBIF, Neotoma) very challenging. In preparation for making  
797 non-marine ostracod datasets available via the Neotoma and GBIF portals, current efforts are  
798 focused mainly on northern hemisphere databases (as with taxonomic harmonization), aided  
799 greatly by the growth of internet resources, particularly *Google Earth* and the online availability  
800 of historical maps (e.g., via *Digimap* in the UK); this work is time-consuming and for some  
801 records the confident assignment of precise coordinates may never be possible, but positional  
802 uncertainties of a few km may be considered acceptable when mapping distributions on large  
803 regional to global scales. A “traffic-light” system for indicating the validation status of species

804 records, developed for NODE and OMEGA, might usefully be adopted by other database  
805 projects: green signifies accurate coordinates, amber denotes coordinates that are acceptable (as  
806 good as possible given limitations of the available data; comments may be added in justification,  
807 e.g., “coordinates approximate for center of lake”), and red indicates uncertain or unreliable  
808 records that need further checking.

809         In the case of ocean sediment cores, there are two critical problems related to the exact  
810 position of the samples. First, the sample ID may not be completely specified, especially in the  
811 older literature. For example, in order to update the age of a sample, it is important to specify the  
812 complete DSDP/ODP/IODP sample ID, which include the Site, Hole, Core, Section, and the  
813 position in the Section. There are typically several methods for calculating the depth below  
814 seafloor for each sample because the sediments expand after being recovered from the deep sea.  
815 The depth method used in an age model could be another different method because  
816 recombination of sections taken from multiple holes could be used to create a more complete  
817 sequence. Therefore, missing any of these ID elements or only specifying the depth below  
818 seafloor is losing the exact position of the samples. This may hinder the age assignment of the  
819 sample or comparison with other studies done on the same sediment cores. Second, OBIS may  
820 not recognize the paleo nature of the Quaternary downcore data it harvested. This could be an  
821 unintended consequence of the original design of OBIS. When the paleo data are combined with  
822 the true modern observation, the problem is clear: the year of observation might be labelled by  
823 the year of sampling cruise, but the time that the observation (i.e., the fossil assemblage)  
824 represents should be the geological age of the sample. Broadly similar problems are encountered  
825 with data from cored lake sediment archives and exposed sedimentary sequences on land. Such  
826 cases may require reconsideration of relevant data standards, leading to modification of existing

827 ones or the introduction of new ones as the system evolves. Data standards are important, but  
828 they should not be allowed to exclude or confuse valuable data.

829

### 830 **9.3. Paleolocality uncertainties**

831 Fossil occurrences are georeferenced according to the present-day locations of the  
832 records, but these are not representative of their true distributions at the time when the fossils  
833 were living because tectonic plates have changed their position throughout the Phanerozoic.  
834 Uncertainty about the paleogeographic position of each occurrence is significant when mapping  
835 the distributions of fossil species. Paleo-distributions can be mapped onto plate tectonic  
836 reconstructions (examples are shown in Fig. 4) but these are typically “snapshots” representing  
837 tens of millions of years so there are still significant uncertainties in the locations. While some  
838 regions have received extensive paleontological and paleogeographical investigation, and are  
839 likely to be well constrained (e.g., European region), others (e.g., central Asia) are less well  
840 constrained due to complex tectonics. For example, the paleogeographic position of the Uzbek  
841 myodocope ostracods occurrence of Mikhailova et al. (2020) during the late Silurian is debated.  
842 Thus, the authors tentatively regard the Uzbek region/terrane in question as a small  
843 microcontinent placed between the North and South China plates based on fossil assemblages  
844 and paleomagnetic data (see Mikhailova et al. 2020). Additionally, there is very little record of  
845 Paleozoic deep-sea and open ocean planktonic faunas because the Paleozoic oceanic crust and  
846 the overlying sediment have been recycled via subduction since then (the oldest oceanic crust  
847 nowadays is less than 200 Myrs old). Thus, most of the deeper dwelling Paleozoic taxa will be  
848 preserved in shelf or slope sediments; deep basin records are generally very limited and restricted

849 to tectonically complex settings such as nappes in accretion mountain ranges (e.g., Andes,  
850 Himalaya) and accretionary prisms around subduction zones (e.g., Japan) (Isozaki, 1997).

851 Another problem is uncertainty related to the age of a species and to the stratigraphic  
852 duration of its record. The taxa recognized in paleontology are skewed towards the more  
853 abundant, widespread, and geologically long-lived species, which have the greatest total number  
854 of individuals and occur in the greatest number of localities and rock types, and so are most  
855 likely to be preserved and recorded (Chaloner and Jablonski, 1994). Thus, longer-lived taxa will  
856 tend to be overrepresented in samples because there is a higher probability that at least one  
857 individual will be preserved (Solow and Smith, 1997).

858

#### 859 **9.4. Marine ostracod database and PBDB applications**

860 Georeferenced occurrence databases are increasingly utilized in the field of  
861 macroecology and biogeography (e.g., Peters and McClennen, 2016; Williams et al., 2018).  
862 OBIS is a promising platform for global marine diversity research. However, our results showed  
863 that its ostracod coverage is concentrated in northern hemisphere and coastal areas. It currently  
864 provides little help in advancing our understanding of large-scale ecological patterns of  
865 ostracods. For example, latitudinal diversity gradients have only been recently explored in the  
866 North Atlantic, Arctic, and NW Pacific based on the AOD and individual datasets (Yasuhara et  
867 al., 2012; Jöst et al., 2019; Chiu et al., 2020). Global analysis of ostracod macroecology and  
868 biogeography remains difficult. The latest global biogeographic scheme for marine ostracods  
869 was in the 1980s based on expert knowledge without quantitative data or analyses (Whatley,  
870 1986; Titterton and Whatley, 1988). A much-needed updated global ostracod biogeography  
871 scheme would be greatly facilitated by global database developments (Yasuhara et al., 2019).

872 PBDB is the only occurrence database dedicated to integrating and curating deep-time  
873 fossil data. It is a resource for quantitatively estimating Phanerozoic-through ostracod diversity.  
874 Indeed, ostracods have a fair number of occurrences throughout the Ordovician to Present in  
875 PBDB (Figure 6). Sepkoski (2000) was the first to utilize PBDB for estimating Crustacea *s.l.*  
876 diversity, recognizing ostracods as the single dominant component of crustacean fossil records  
877 throughout the Phanerozoic. Our generic diversity curve of ostracods using the up-to-date PBDB  
878 data is consistent with Sepkoski's (2000) in general (Figure 6). Phanerozoic generic diversity  
879 trends of ostracod orders reconstructed by using the up-to-date PBDB (Figure 7) are also  
880 consistent with the ostracod experts' views (Whatley et al., 1993; Armstrong and Brasier, 2004).  
881 However, it is important to note that ostracods have received relatively little attention in the  
882 PBDB, and there is highly uneven distribution of data entries in space and time. Any estimations  
883 on the rates of origination and extinction of ostracods must be interpreted with caution.

884 PBDB is a promising platform for future collaboration in the ostracodology community.  
885 Many publications have explored macroevolutionary, paleoecological or biogeographical  
886 questions by analyzing georeferenced occurrence data, many of which have paleoenvironmental  
887 and trait information altogether in the PBDB (e.g., motility, primary skeletal mineralogy, life  
888 habitat, diet) (e.g., Marx and Uhen, 2010; Kiessling and Kocsis, 2015; Leprieur et al., 2016;  
889 Kocsis et al., 2018a; Reddin et al., 2020). In contrast, there have been very few ostracod studies  
890 utilizing PBDB (Donovan and van den Hoek Ostende, 2012; Forsey, 2016) even though  
891 Ostracoda is one of the few taxonomic groups that have abundant fossil records almost  
892 throughout the entire Phanerozoic (from Ordovician to Quaternary) (Figures 6–7). The taxonomy  
893 scheme in the PBDB is designed to mitigate the issues of outdated taxonomy, open  
894 nomenclature, and conflicting opinions. Although the current taxonomic resources of ostracods

895 are deficient, substantial future updates can presumably allow quantitative Phanerozoic-through  
896 biodiversity analyses.

897

## 898 **9.5. Non-marine ostracod database applications**

899         The combination of species occurrence records with geographical coordinates in non-  
900 marine distributional databases has high potential value as a research tool, for example in  
901 biogeographic studies of endemism and cosmopolitanism. This rich area of research holds many  
902 unanswered questions for which temporal and spatial biogeographic data are critical. Such  
903 datasets facilitate hypothesis-testing and can reveal interesting ecological as well as taxonomic  
904 insights. Calibrations of the temperature ranges of species, for use in Quaternary paleoclimatic  
905 reconstructions, are achieved by matching climate data to geographical occurrences in NODE,  
906 NANODE and DOAD (see, e.g., Horne et al., 2012; Marchegiano et al., 2020). Calibrations can  
907 be improved, moreover, by combining data from geographically separated regions in order to  
908 obtain more complete coverage of the climatic distributions of species. Molecular biological  
909 studies of geographical parthenogenesis in European non-marine ostracods have been supported  
910 by mapping sexual and asexual populations with the NODE database (e.g., Horne et al., 1998;  
911 Horne & Martens, 1999; Schmit et al., 2013).

## 912 **10. Conclusions**

913         In the rising trend of database research, it has become clear that occurrence and  
914 measurement data have more impact in science if they are digitalized, tabulated, and integrated.  
915 Even though they are far from comprehensive, ostracod data in many databases have  
916 demonstrated its unique potential in advancing our understanding of  
917         large-scale biodiversity and evolution in the future. They cover a wide range of  
918 ecosystems and are an important representative of Crustacea *s.l.* almost throughout the

919 Phanerozoic in the fossil records. Therefore, we strongly recommend ostracodologists publish  
920 georeferenced data with suitable databases after they published their studies on taxonomy,  
921 ecology, biogeography, and evolution. Contributing ostracod data to databases can not only  
922 strongly impact paleontological and biodiversity research, but also add value to personal efforts  
923 and public funds.

924 Consistent and accurate taxonomy is of key importance to large-scale biodiversity  
925 research, while global databases are integrating an increasing number of datasets. Aphia's well-  
926 developed taxonomic system is the taxonomic backbone of many distributional databases. PBDB  
927 has an algorithm-powered taxonomic classification dynamically linked to its own occurrence  
928 data. Collaborating through Aphia and PBDB can be an effective method for ostracodologists to  
929 achieve the consensus opinions and directly improve the taxonomic quality of mainstream  
930 databases in the future.

931

### 932 **Acknowledgements**

933 We would like to thank Robin J. Smith for the data of the East Asian Non-marine Ostracod  
934 Database; Renate Matzke-Karasz for the updated information on Kempf Database Ostracoda;  
935 Laura Gemery and Thomas Cronin for the updated information on Arctic Ostracode Database;  
936 two anonymous reviewers and two editors, Richard Jordan and Gene Hunt, for valuable  
937 comments. The NODE database was supported by Human Capital & Mobility Programme,  
938 European Union funding [grant number ERBCHRXCT930253]; Marie Curie Research and  
939 Training Network, European Union funding [grant number FP6-512492]; Contingency Fund  
940 award, European Union funding. NANODE and AJS were supported by the U. S. Geological  
941 Survey; Kent State University; National Science Foundation. AJS's work in Neotoma database  
942 was supported by National Science Foundation [grant numbers 0947459, 1550721, 1948297];  
943 NSF EarthCube program [grant number 1540994]. HHMH and MY are partly supported by  
944 Research Grants Council, Hong Kong Special Administrative Region, China [grant numbers  
945 HKU 17300720, HKU 17302518, HKU 17311316]; the Seed Funding Programme for Basic  
946 Research, the University of Hong Kong [grant numbers 201811159076, 201711159057,  
947 201611159053]; Faculty of Science RAE Improvement Fund, the University of Hong Kong.  
948 HHMH was supported by Ecology and Biodiversity Division Fund, the University of Hong  
949 Kong; Peter Buck Postdoc Fellowship, Smithsonian Institution. SNB's work on the World  
950 Ostracoda Database was/is funded by a Lifewatch mini grant and UESC [Projeto Biodiversidade  
951 de Ostracoda Marinho do Sul da Bahia].  
952

953 **References**

- 954 Access to Biological Collections Data Task Group, 2005. Access to Biological Collection Data  
955 (ABCD). Biodiversity Information Standards (TDWG). Permanent IRI:  
956 <http://www.tdwg.org/standards/115>.
- 957 Alvarez Zarikian, C.A., 2015. Cenozoic bathyal and abyssal ostracods beneath the oligotrophic  
958 South Pacific Gyre (IODP Expedition 329 Sites U1367, U1368 and U1370). *Palaeogeography,*  
959 *Palaeoclimatology, Palaeoecology* 419, 115–142. doi:10.1016/j.palaeo.2014.07.024
- 960 Angel, M., 2016. Personal library collection of Martin Angel of published and unpublished  
961 Halocyprid (Ostracoda) occurrences. Natural History Museum. Accessed on 2021-10-21 via  
962 [<https://obis.org/dataset/b1363fbb-afec-4dc7-85b6-39b8a5ebb55c>].
- 963 Angel, M.V., Blachowiak-Samolyk, K., Chavtur, V.G., 2008. Atlas of Atlantic Planktonic  
964 Ostracods. Natural History Museum. Accessed on 2021-10-21 via  
965 [[http://www.nhm.ac.uk/research-curation/scientific-resources/biodiversity/global-](http://www.nhm.ac.uk/research-curation/scientific-resources/biodiversity/global-biodiversity/atlantic-ostracods/atlas/browse/index.jsp)  
966 [biodiversity/atlantic-ostracods/atlas/browse/index.jsp](http://www.nhm.ac.uk/research-curation/scientific-resources/biodiversity/global-biodiversity/atlantic-ostracods/atlas/browse/index.jsp)].
- 967 Armstrong, H.A., Brasier, M.D., 2004. Ostracods, *Microfossils*, pp. 219–248.
- 968 Balian, E., Segers, H., Lévêque, C., Martens, K., 2008. An introduction to the Freshwater  
969 Animal Diversity Assessment (FADA) project. *Hydrobiologia* 595, 3–8. doi:10.1007/s10750-  
970 007-9235-6
- 971 Bergue, C., Govindan, A., 2010. Eocene-Pliocene deep sea ostracodes from ODP site 744A,  
972 Southern Indian Ocean. *Anais da Academia Brasileira de Ciências* 82, 747–760.  
973 doi:10.1590/S0001-37652010000300021
- 974 biodiversity.aq, 2021. Antarctic Biodiversity Portal. Scientific Committee for Antarctic  
975 Research. Accessed on 2021-10-21 via [[www.biodiversity.aq](http://www.biodiversity.aq)].
- 976 Blachowiak-Samolyk, K., Angel, M., 2008. An Atlas of Southern Ocean planktonic ostracods.  
977 Southampton Oceanography Centre. Accessed on 2021-10-21 via  
978 [<http://deep.iopan.gda.pl/ostracoda/index.php>].
- 979 Boomer, I., Horne, D.J., Slipper, I.J., 2003. The Use of Ostracods in Palaeoenvironmental  
980 Studies, or What can you do with an Ostracod Shell? *The Paleontological Society Papers* 9, 153–  
981 180. doi:10.1017/S1089332600002199
- 982 Brandão, S.N., 2012. Southern Ocean Ostracoda (except Halocypridina) database. Accessed on  
983 2021-10-22 via [GBIF.org]. doi:10.15468/de1pt3
- 984 Brandão, S.N., Karanovic, I., 2020. World Ostracoda Database. Accessed on 2020-12-23 via  
985 [<http://www.marinespecies.org/ostracoda>]. doi:10.14284/364
- 986 Brandt, A., Gooday, A.J., Brandão, S.N., Brix, S., Brökeland, W., Cedhagen, T., Choudhury, M.,  
987 Cornelius, N., Danis, B., De Mesel, I., Diaz, R.J., Gillan, D.C., Ebbe, B., Howe, J.A., Janussen,  
988 D., Kaiser, S., Linse, K., Malyutina, M., Pawlowski, J., Raupach, M., Vanreusel, A., 2007. First

- 989 insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature*  
990 447(7142), 307–311. doi:10.1038/nature05827
- 991 Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K.,  
992 Schoupe, M., 2019. Open science, open data, and open scholarship: European policies to make  
993 science fit for the twenty-first century. *Frontiers in Big Data* 2, 43. doi:  
994 10.3389/fdata.2019.00043
- 995 Chaloner, W.G. and Jablonski, D., 1994. Extinctions in the Fossil Record: Discussion.  
996 *Philosophical Transactions of the Royal Society of London Series B*, 344(1307), 16–17.
- 997 Chaudhary, C., Richardson, A.J., Schoeman, D.S., Costello, M.J., 2021. Global warming is  
998 causing a more pronounced dip in marine species richness around the equator. *PNAS*  
999 118(15), e2015094118. doi:10.1073/pnas.2015094118
- 1000 Chiu, W.-T.R., Yasuhara, M., Cronin, T.M., Hunt, G., Gemery, L., Wei, C.-L., 2020. Marine  
1001 latitudinal diversity gradients, niche conservatism and out of the tropics and Arctic: Climatic  
1002 sensitivity of small organisms. *Journal of Biogeography* 47(4), 817–828. doi:10.1111/jbi.13793
- 1003 Costello, M.J., 2000. Developing species information systems: the European Register of Marine  
1004 Species (ERMS). *Oceanography* 13(3), 48–55.
- 1005 Costello, M.J., Emblow, C., White, R.J., 2001. European register of marine species: a check-list  
1006 of the marine species in Europe and a bibliography of guides to their identification. Paris:  
1007 Muséum national d'histoire naturelle.
- 1008 Costello, M.J., Wicczorek, J., 2014. Best practice for biodiversity data management and  
1009 publication. *Biological Conservation* 173, 68–73. doi:10.1016/j.biocon.2013.10.018
- 1010 Cronin, T.M., Briggs Jr, W., Brouwers, E., Whatley, R., Wood, A., Cotton, M., 1991. Modern  
1011 Arctic podocopid ostracode database. US Geological Survey Open-File Report 91–385: 1–51.
- 1012 Cronin, T.M., Gemery, L., Briggs, W.M., Jakobsson, M., Polyak, L., Brouwers, E.M., 2010.  
1013 Quaternary Sea-ice history in the Arctic Ocean based on a new Ostracode sea-ice proxy.  
1014 *Quaternary Science Reviews* 29 (25-26), 3415–3429. doi:10.1016/j.quascirev.2010.05.024
- 1015 Cronin, T.M., Gemery, L.J., Briggs, W.M., Brouwers, E.M., Schornikov, E.I., Stepanova, A.,  
1016 Wood, A.M., Yasuhara, M., Siu, S., 2021. Arctic Ostracode Database 2020. NOAA's National  
1017 Centers for Environmental Information (NCEI). Accessed on 2021-10-14 via  
1018 [<https://www.ncdc.noaa.gov/paleo/study/32312>].
- 1019 Cronin, T.M., Holtz, T.R., Brouwers, E.M., Briggs, W.M., Whatley, R.C., Wood, A., 1995.  
1020 Arctic Ostracode Database. U.S. Geological Survey, Reston, VA. Accessed on 2021-10-21 via  
1021 [<https://geo-nsdi.er.usgs.gov/metadata/other/prism-arctic-ostracod/>].
- 1022 Curry, B.B., Delorme, L.D., Smith, A.J., Palmer, D.F., Stiff, B.J., 2012. The Biogeography and  
1023 Physicochemical Characteristics of Aquatic Habitats of Freshwater Ostracods in Canada and the  
1024 United States, in: Horne, D.J., Holmes, J.A., Rodriguez-Lazaro, J., Viehberg, F.A. (Eds.),

- 1025 Ostracoda as Proxies for Quaternary Climate Change. *Developments in Quaternary Sciences*,  
1026 Elsevier, pp. 85–115.
- 1027 Cuvelier, D., Claus, S., Appeltans, W., Vanhoorne, B., Vanden Berghe, E., Costello, M., 2006.  
1028 European Register of Marine Species (ERMS)-plans turning into reality. *MarBEF Newsletter* 4  
1029 14-15.
- 1030 Darwin Core Task Group, 2009. Darwin Core. Biodiversity Information Standards (TDWG).  
1031 Permanent IRI: <http://www.tdwg.org/standards/450>
- 1032 Davies, T.G., Rahman, I.A., Lautenschlager, S., Cunningham, J.A., Asher, R.J., Barrett, P.M.,  
1033 Bates, K.T., Bengtson, S., Benson, R.B., Boyer, D.M., 2017. Open data and digital morphology.  
1034 *Proceedings of the Royal Society B: Biological Sciences* 284(1852), 20170194.  
1035 doi:10.1098/rspb.2017.0194
- 1036 Donovan, S.K., van den Hoek Ostende, L.W., 2012. *Neoextrania*, a replacement name for  
1037 *Extrania* Sun & Wang, 1985 (Ostracoda), preoccupied by *Extrania* Qian in Qiu et al., 1983  
1038 (Trilobita). *Alcheringa: An Australasian Journal of Palaeontology* 36(3), 423–423.  
1039 doi:10.1080/03115518.2012.694304
- 1040 Ellis, B.F., Messina, A.R., 1952. *Catalogue of Ostracoda*. American Museum of Natural History.  
1041 Micropaleontology Press, n.d. Ellis and Messina Catalogues. Accessed on 2021-11-05 via  
1042 [<https://www.micropress.org/ellis-messina.html>].
- 1043 Faith, D., Collen, B., Ariño, A., Koleff, P.K.P., Guinotte, J., Kerr, J., Chavan, V., 2013. Bridging  
1044 the biodiversity data gaps: Recommendations to meet users' data needs. *Biodiversity Informatics*  
1045 8(2). doi:10.17161/bi.v8i2.4126
- 1046 Fegraus, E.H., Andelman, S., Jones, M.B. and Schildhauer, M., 2005. Maximizing the value of  
1047 ecological data with structured metadata: an introduction to ecological metadata language (EML)  
1048 and principles for metadata creation. *Bulletin of the Ecological Society of America*, 86 (3), 158–  
1049 168.
- 1050 Forester, R.M., Smith, A.J., Palmer, D.F., Curry, B.B., 2005. North American Nonmarine  
1051 Ostracode Database, version 1, [www.kent.edu/NANODE](http://www.kent.edu/NANODE), Kent State University, Kent, Ohio,  
1052 U.S.A.
- 1053 Forsey, G., 2016. Ostracods as proxies for past seagrass: A review. *Palaeogeography,*  
1054 *Palaeoclimatology, Palaeoecology* 447. doi:10.1016/j.palaeo.2016.01.028
- 1055 GBIF, 2021. What is GBIF? Accessed on 2021-10-14 via [<https://www.gbif.org/what-is-gbif>].
- 1056 GBIF Secretariat, 2021. GBIF Backbone Taxonomy. Checklist dataset. Accessed on 2021-10-21  
1057 via [GBIF.org]. doi:10.15468/39omei

- 1058 GBIF/TDWG Multimedia Resources Task Group, 2013. Audubon Core Multimedia Resources  
1059 Metadata Schema. Biodiversity Information Standards (TDWG). Permanent IRI:  
1060 <http://www.tdwg.org/standards/638>.
- 1061 Gemery, L., Cronin, T.M., Briggs, W.M., Brouwers, E.M., Schornikov, E.I., Stepanova, A.,  
1062 Wood, A.M., Yasuhara, M., 2017. An Arctic and Subarctic ostracode database: biogeographic  
1063 and paleoceanographic applications. *Hydrobiologia* 786(1), 59–95.
- 1064 Gemery, L., Cooper, L.W., Magen, C., Cronin, T.M. and Grebmeier, J.M., 2021a. Stable oxygen  
1065 isotopes in shallow marine ostracodes from the northern Bering and Chukchi Seas. *Marine*  
1066 *Micropaleontology*, p.102001. doi:10.1016/j.marmicro.2021.101979
- 1067 Gemery, L., Cronin, T.M., Cooper, L.W., Dowsett, H.J. and Grebmeier, J.M., 2021b.  
1068 Biogeography and ecology of Ostracoda in the US northern Bering, Chukchi, and Beaufort Seas.  
1069 *PloS one*, 16(5), p.e0251164. doi:10.1371/journal.pone.0251164
- 1070 Goddard, A., Wilson, N., Cryer, P., Yamashita, G., 2011. Data hosting infrastructure for primary  
1071 biodiversity data. *BMC bioinformatics* 12 (S15), S5.
- 1072 Grimm, E.C., Blois, J., Giesecke, T., Graham, R., Smith, A., Williams, J., 2018. Constituent  
1073 databases and data stewards in the Neotoma Paleocology Database: history, growth, and new  
1074 directions. *PAGES Magazine* 26, 64–65.
- 1075 Hendricks, J.R., Stigall, A.L., Lieberman, B.S., 2015. The Digital Atlas of Ancient Life:  
1076 Delivering Information on Paleontology and Biogeography. *Palaeontologia Electronica* 1–9.  
1077 doi:10.26879/153E
- 1078 Holetschek, J., Dröge, G., Güntsch, A., Berendsohn, W., 2012. The ABCD of primary  
1079 biodiversity data access. *Plant Biosystems-An International Journal Dealing with all Aspects of*  
1080 *Plant Biology* 146(4), 771–779. doi:10.1080/11263504.2012.740085
- 1081 Holmes, J.A., Chivas, A.R., 2002. Introduction, in: Holmes, J.A., Chivas, A.R. (Eds.), *The*  
1082 *Ostracoda: applications in Quaternary research*. American Geophysical Union, pp. 1–4.
- 1083 Horne, D., Baltanas, A., Paris, G., 1998. Geographical distribution of reproductive modes in  
1084 living non-marine ostracods, in: Martens, K. (Ed.), *Sex and parthenogenesis: evolutionary*  
1085 *ecology of reproductive modes in non-marine ostracods*. Backhuys, Leiden, pp. 77–99.
- 1086 Horne, D.J., 2014. Ostracod Metadatabase of Environmental and Geographical Attributes  
1087 (Europe, North America). *BioFresh*. Accessed on 2021-11-05 via  
1088 [<http://data.freshwaterbiodiversity.eu>]. doi:10.13148/bfcf15
- 1089 Horne, D.J., Brandon Curry, B., Denis Delorme, L., Martens, K., Smith, A.J., Smith, R.J., 2011.  
1090 OMEGA: The Ostracod metadatabase of environmental and geographical attributes. *Joannea*  
1091 *Geologie und Paläontologie* (11), 80–84.

- 1092 Horne, D.J., Cohen, A., Martens, K., 2002. Taxonomy, morphology and biology of Quaternary  
1093 and living Ostracoda, in: Holmes, J.A., Chivas, A.R. (Eds.), *The Ostracoda: Applications in*  
1094 *Quaternary Research*. American Geophysical Union, Washington, DC, pp. 5–36.
- 1095 Horne, D.J., Curry, B.B., Mesquita-Joanes, F., 2012. Mutual Climatic Range Methods for  
1096 Quaternary Ostracods, in: Horne, D.J., Holmes, J.A., Rodriguez-Lazaro, J., Viehberg, F.A.  
1097 (Eds.), *Developments in Quaternary Sciences*. Elsevier, pp. 65–84.
- 1098 Horne, D.J., Martens, K., 1998. Geographical parthenogenesis in European non-marine  
1099 ostracods: Post-glacial invasion or Holocene stability? *Hydrobiologia* 391, 1–7.  
1100 doi:10.1023/A:1003508210166
- 1101 Horne, D.J., Meisch, C., Martens, K., 2019. Arthropoda: Ostracoda, in: Rogers, D.C., Thorp, J.  
1102 (Eds.), *Thorp and Covich's Freshwater Invertebrates*, Fourth ed. Academic Press, pp. 725–760.
- 1103 Isozaki, Y., 1997. Permo-Triassic boundary superanoxia and stratified superocean: records from  
1104 lost deep sea. *Science* 276(5310), 235–238.
- 1105 Jablonski, D., Sepkoski Jr, J.J., 1996. Paleobiology, community ecology, and scales of ecological  
1106 pattern. *Ecology* 77(5), 1367–1378.
- 1107 Jones, M.B., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., Whiteaker, T.,  
1108 Earl, S., Chong, S., 2019. Ecological metadata language version 2.2. 0. KNB Data Repository.  
1109 doi:10.5063/F11834T2
- 1110 Jöst, A.B., Yasuhara, M., Wei, C.L., Okahashi, H., Ostmann, A., Martínez Arbizu, P., Mamo, B.,  
1111 Svavarsson, J., Brix, S., 2019. North Atlantic Gateway: Test bed of deep-sea macroecological  
1112 patterns. *Journal of Biogeography* 46(9), 2056–2066. doi:10.1111/jbi.13632
- 1113 Kiessling, W., Kocsis, Á.T., 2015. Biodiversity dynamics and environmental occupancy of fossil  
1114 azooxanthellate and zooxanthellate scleractinian corals. *Paleobiology* 41(3), 402–414. doi:  
1115 10.1017/pab.2015.6
- 1116 Kocsis, Á.T., Reddin, C.J., Kiessling, W., 2018a. The biogeographical imprint of mass  
1117 extinctions. *Proceedings of the Royal Society B: Biological Sciences* 285(1878), 20180232.  
1118 doi:10.1098/rspb.2018.0232
- 1119 Kocsis, Á.T., Reddin, C.J., Kiessling, W., 2018b. The stability of coastal benthic biogeography  
1120 over the last 10 million years. *Global Ecology and Biogeography* 27(9), 1106–1120.  
1121 doi:10.1111/geb.12771
- 1122 Lautenschlager, S., Rücklin, M., 2014. Beyond the print—virtual paleontology in science  
1123 publishing, outreach, and education. *Journal of Paleontology* 88(4), 727. doi: 10.1666/13-085
- 1124 Lazarus, D., 1994. Neptune: a marine micropaleontology database. *Mathematical Geology* 26  
1125 (7), 817–832.

- 1126 Lazarus, D., Barron, J., Renaudie, J., Diver, P., Türke, A., 2014. Cenozoic planktonic marine  
1127 diatom diversity and correlation to climate change. *PLOS ONE* 9(1): e84857.  
1128 doi:10.1371/journal.pone.0084857
- 1129 Leprieur, F., Descombes, P., Gaboriau, T., Cowman, P.F., Parravicini, V., Kulbicki, M., Melián,  
1130 C.J., De Santana, C.N., Heine, C., Mouillot, D., Bellwood, D.R., Pellissier, L., 2016. Plate  
1131 tectonics drive tropical reef biodiversity dynamics. *Nature Communications* 7 (1), 11461.  
1132 doi:10.1038/ncomms11461
- 1133 Liow, L.H., Stenseth, N.C., 2007. The rise and fall of species: implications for  
1134 macroevolutionary and macroecological studies. *Proceedings of the Royal Society B: Biological*  
1135 *Sciences* 274(1626), 2745–2752. doi:10.1098/rspb.2007.1006
- 1136 Lockwood, R., Cohen, P.A., Uhen, M.D., Ryker, K., 2018. Utilizing the Paleobiology Database  
1137 to provide educational opportunities for undergraduates. Cambridge University Press.
- 1138 Majoran, S., Dingle, R.V., 2001. Cenozoic deep-sea ostracods from southwestern South Atlantic  
1139 (DSDP/ODP Sites 329, 513 and 699). *Revista Espanola de Micropaleontologia* 33 (2), 205–216.
- 1140 Marchegiano, M., Horne, D.J., Gliozzi, E., Francke, A., Wagner, B., Ariztegui, D., 2020. Rapid  
1141 Late Pleistocene climate change reconstructed from a lacustrine ostracod record in central Italy  
1142 (Lake Trasimeno, Umbria). *Boreas*, 49 (4), 739–750. doi:10.1111/bor.12450
- 1143 Martens, K., Savatnalinton, S., Schön, I., Meisch, C., Horne, D.J., 2013. World checklist of  
1144 freshwater Ostracoda species. World Wide Web electronic publication. Accessed on 2021-11-05  
1145 via [<http://fada.biodiversity.be/group/show/18>].
- 1146 Martens, K., Schön, I., Meisch, C., Horne, D.J., 2008. Global diversity of ostracods (Ostracoda,  
1147 Crustacea) in freshwater. *Hydrobiologia* 595 (1), 185–193. doi:10.1007/s10750-007-9245-4
- 1148 Martins, M.J.F., Hunt, G., Thompson, C.M., Lockwood, R., Swaddle, J.P., Puckett, T.M., 2020.  
1149 Shifts in sexual dimorphism across a mass extinction in ostracods: implications for sexual  
1150 selection as a factor in extinction risk. *Proceedings of the Royal Society B* 287 (1933),  
1151 20200730. doi:10.1098/rspb.2020.0730
- 1152 Marx, F.G., Uhen, M.D., 2010. Climate, Critters, and Cetaceans: Cenozoic Drivers of the  
1153 Evolution of Modern Whales. *Science* 327 (5968), 993–996. doi:10.1126/science.1185581
- 1154 Matzke-Karasz, R., 2014. Eugen Karl Kempf, the man behind the Kempf Database Ostracoda.  
1155 *Crustaceana* 87 (8–9), 901–922.
- 1156 Meisch, C., Smith, R.J., Martens, K., 2019. A subjective global checklist of the extant non-  
1157 marine Ostracoda (Crustacea). *European Journal of Taxonomy* (492). doi:10.5852/ejt.2019.492
- 1158 Mesquita-Joanes, F., Smith, A.J., Viehberg, F.A., 2012. Chapter 2 - The Ecology of Ostracoda  
1159 Across Levels of Biological Organisation from Individual to Ecosystem: A Review of Recent  
1160 Developments and Future Potential, in: Horne, D.J., Holmes, J.A., Rodriguez-Lazaro, J.,  
1161 Viehberg, F.A. (Eds.), *Developments in Quaternary Sciences*. Elsevier, pp. 15–35.

- 1162 Mikhailova, E., Perrier, V., Williams, M., Siveter, D.J., Tarasenko, A., Salimova, F., Kim, I.A.,  
1163 2020. Cosmopolitan myodocope ostracods from the Silurian of Uzbekistan, Central Asia  
1164 ostracodes myodocopes cosmopolites du silurien Asie centrale. *Bulletin de la Société*  
1165 *Géologique de France* 191(1), 32.
- 1166 Morris, R.A., Barve, V., Carausu, M., Chavan, V., Cuadra, J., Freeland, C., Hagedorn, G., Leary,  
1167 P., Mozzherin, D., Olson, A., 2013. Discovery and publishing of primary biodiversity data  
1168 associated with multimedia resources: The Audubon Core strategies and approaches.  
1169 *Biodiversity Informatics* 8(2). doi:10.17161/bi.v8i2.4117
- 1170 Müller, R.D., Cannon, J., Qin, X., Watson, R.J., Gurnis, M., Williams, S., Pfaffelmoser, T.,  
1171 Seton, M., Russell, S.H., Zahirovic, S., 2018. GPlates: building a virtual Earth through deep  
1172 time. *Geochemistry, Geophysics, Geosystems* 19(7), 2243-2261. doi:10.1029/2018GC007584
- 1173 Nelson, G., Ellis, S., 2019. The history and impact of digitization and digital data mobilization  
1174 on biodiversity research. *Philosophical Transactions of the Royal Society B* 374(1763),  
1175 20170391. doi:10.1098/rstb.2017.0391
- 1176 OBIS, 2021. Ocean Biodiversity Information System. Intergovernmental Oceanographic  
1177 Commission of UNESCO. Accessed on 2021-11-05 via [<https://obis.org/>].
- 1178 OBIS Secretariat, 2021. OBIS Secretariat. Ocean Biogeographic Information System. Accessed  
1179 on 2021-10-14 via [<https://obis.org/node/310922b4-9d0c-4de1-92d7-9b442d34765b>]
- 1180 Paleobiology Database, 2021. The Paleobiology Database. Checklist dataset. Accessed on 2021-  
1181 10-21 via [[GBIF.org](https://gbif.org/)]. doi:10.15468/zzoyxi
- 1182 PANGAEA®, Data Publisher for Earth & Environmental Science. doi:10.1594/PANGAEA.
- 1183 Perrier, V., Williams, M., Siveter, D.J., 2015. The fossil record and palaeoenvironmental  
1184 significance of marine arthropod zooplankton. *Earth-Science Reviews* 146, 146–162.  
1185 doi:j.earscirev.2015.02.003
- 1186 Peters, S.E., McClennen, M., 2016. The Paleobiology Database application programming  
1187 interface. *Paleobiology* 42(1), 1–7. doi:10.1017/pab.2015.39
- 1188 Powers, S.M., Hampton, S.E., 2019. Open science, reproducibility, and transparency in ecology.  
1189 *Ecological applications* 29 (1), e01822. doi:10.1002/eap.1822
- 1190 Reddin, C.J., Kocsis, Á.T., Aberhan, M., Kiessling, W., 2020. Victims of ancient hyperthermal  
1191 events herald the fates of marine clades and traits under global warming. *Global Change Biology*  
1192 27 (4), 868–878. doi:10.1111/gcb.15434
- 1193 Rick, T.C., Lockwood, R., 2013. Integrating paleobiology, archeology, and history to inform  
1194 biological conservation. *Conservation Biology* 27 (1), 45–54. doi:10.1111/j.1523-  
1195 1739.2012.01920.x

- 1196 Robertson, T., Döring, M., Guralnick, R., Bloom, D., Wiczorek, J., Braak, K., Otegui, J.,  
1197 Russell, L., Desmet, P., 2014. The GBIF integrated publishing toolkit: facilitating the efficient  
1198 publishing of biodiversity data on the internet. PLOS ONE 9 (8), e102623.  
1199 doi:10.1371/journal.pone.0102623
- 1200 Rodriguez-Lazaro, J., Ruiz-Muñoz, F., 2012. Chapter 1 - A General Introduction to Ostracods:  
1201 Morphology, Distribution, Fossil Record and Applications, in: Horne, D.J., Holmes, J.A.,  
1202 Rodriguez-Lazaro, J., Viehberg, F.A. (Eds.), *Developments in Quaternary Sciences*. Elsevier, pp.  
1203 1–14.
- 1204 Schmit, O., Bode, S.N.S., Camacho, A., Horne, D.J., Lamatsch, D.K., Martens, K., Martins,  
1205 M.J.F., Namiotko, T., Rossetti, G., Rueda-Sevilla, J., Schön, I., Vandekerkhove, J., Mesquita-  
1206 Joanes, F., 2013. Linking present environment and the segregation of reproductive modes  
1207 (geographical parthenogenesis) in *Eucypris virens* (Crustacea: Ostracoda). *Journal of*  
1208 *Biogeography* 40 (12), 2396–2408. doi:10.1111/jbi.12174
- 1209 Scholes, R., Mace, G., Turner, W., Geller, G., Jürgens, N., Larigauderie, A., Muchoney, D.,  
1210 Walther, B., Mooney, H., 2008. Toward a global biodiversity observing system. *Science* 321  
1211 (5892), 1044–1045. doi:10.1126/science.1162055
- 1212 Scotese, C.R., Wright, N., 2018. PALEOMAP paleodigital elevation models (PaleoDEMS) for  
1213 the Phanerozoic. URL: <https://www.earthbyte.org/paleodem-resource-scotese-and-wright-2018>.
- 1214 Smith, A.J., Horne, D.J., 2002. Ecology of marine, marginal marine and nonmarine ostracodes,  
1215 *Geophysical Monograph Series*, pp. 37–64.
- 1216 Smith, A.J., Horne, D.J., 2016. Class Ostracoda, in: Thorp, J., Rogers, D.C. (Eds.), *Thorp and*  
1217 *Covich's Freshwater Invertebrates*, Fourth ed. Academic Press, pp. 477–514.
- 1218 Smith, A.J., Horne, D.J., Martens, K., Schön, I., 2015. Class Ostracoda, in: Thorp, J.H., Rogers,  
1219 D.C. (Eds.), *Thorp and Covich's Freshwater Invertebrates*, Fourth ed. Academic Press, Boston,  
1220 pp. 757–780.
- 1221 Solow, A.R., Smith, W., 1997. On fossil preservation and the stratigraphic ranges of taxa.  
1222 *Paleobiology* 271–277. doi:10.1017/S0094837300019680
- 1223 Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V., Worm, B., 2010.  
1224 Global patterns and predictors of marine biodiversity across taxa. *Nature* 466, 1098–1101. doi:  
1225 10.1038/nature09329
- 1226 Titterton, R., Whatley, R.C., 1988. The Provincial Distribution of Shallow Water Indo-Pacific  
1227 Marine Ostracoda: Origins, Antiquity, Dispersal Routes and Mechanisms, *Developments in*  
1228 *Palaeontology and Stratigraphy*, pp. 759–786.
- 1229 Vandepitte, L., Vanhoorne, B., Decock, W., Dekeyzer, S., Trias Verbeeck, A., Bovit, L.,  
1230 Hernandez, F., Mees, J., 2015. How Aphia—The platform behind several online and  
1231 taxonomically oriented databases—can serve both the taxonomic community and the field of

- 1232 biodiversity informatics. *Journal of Marine Science and Engineering* 3(4), 1448–1473.  
1233 doi:10.3390/jmse3041448
- 1234 Viehberg, F.A., Matzke-Karasz, R., Boush, L.P., Smith, A.J., 2014. Fossil and Recent meet  
1235 Kempf database. *Crustaceana* 87(8–9), 897–900.
- 1236 Vilhena, D.A., Antonelli, A., 2015. A network approach for identifying and delimiting  
1237 biogeographical regions. *Nature Communications* 6(1), 6848. doi:10.1038/ncomms7848
- 1238 Villéger, S., Blanchet, S., Beauchard, O., Oberdorff, T., Brosse, S., 2011. Homogenization  
1239 patterns of the world’s freshwater fish faunas. *PNAS* 108, 18003–18008.  
1240 doi:10.1073/pnas.1107614108
- 1241 Weibel, S., 1999. The State of the Dublin Core Metadata Initiative: April 1999. *Bulletin of the*  
1242 *American Society for Information Science and Technology* 25(5), 18–22. doi:10.1002/bult.127
- 1243 Whatley, R., 1986. The southern end of Tethys: An important locus for the origin and evolution  
1244 of both deep and shallow water Ostracoda, in: McKenzie, K.G. (Ed.), *Proceedings 2nd Shallow*  
1245 *Tethys Symposium, Wagga Wagga, Bolkema*, pp. 461–475.
- 1246 Whatley, R.C., Siveter, D.J., Boomer, I.D., 1993. Arthropoda (Crustacea: Ostracoda), in: Benton,  
1247 M.J. (Ed.), *The Fossil Record 2*. Chapman & Hall, London, pp. 343–356.
- 1248 Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T.,  
1249 Vieglais, D., 2012. Darwin Core: an evolving community-developed biodiversity data standard.  
1250 *PLOS ONE* 7(1), e29715. doi:10.1371/journal.pone.0029715
- 1251 Williams, J.W., Grimm, E.C., Blois, J.L., Charles, D.F., Davis, E.B., Goring, S.J., Graham,  
1252 R.W., Smith, A.J., Anderson, M., Arroyo-Cabrales, J., 2018. The Neotoma Paleocology  
1253 Database, a multiproxy, international, community-curated data resource. *Quaternary Research*  
1254 89(1), 156–177. doi:10.1017/qua.2017.105
- 1255 Williams, M., Siveter, D.J., Salas, M.J., Vannier, J., Popov, L.E., Pour, M.G., 2008. The earliest  
1256 ostracods: the geological evidence. *Senckenbergiana lethaea* 88(1), 11–21.
- 1257 Wright, N., Zahirovic, S., Müller, R., Seton, M., 2013. Towards community-driven  
1258 paleogeographic reconstructions: integrating open-access paleogeographic and paleobiology data  
1259 with plate tectonics. *Biogeosciences* 10(3), 1529–1541. doi:10.5194/bg-10-1529-2013
- 1260 Yasuhara, M., 2019. Marine biodiversity in space and time: What tiny fossils tell. *Mètode*  
1261 *Science Studies Journal-Annual Review* (9).
- 1262 Yasuhara, M., Cronin, T.M., 2008. Climatic influences on deep-sea ostracode (Crustacea)  
1263 diversity for the last three million years. *Ecology* 89 (sp11), S53–S65. doi:10.1890/07-1021.1
- 1264 Yasuhara, M., Hong, Y., Tian, S.Y., Chong, W.K., Okahashi, H., Littler, K., Cotton, L., 2019.  
1265 Eocene shallow-marine ostracods from Madagascar: southern end of the Tethys? *Journal of*  
1266 *Systematic Palaeontology* 17(9), 705–757. doi:10.1080/14772019.2018.1453555

1267 Yasuhara, M., Hunt, G., van Dijken, G., Arrigo, K.R., Cronin, T.M., Wollenburg, J.E., 2012.  
1268 Patterns and controlling factors of species diversity in the Arctic Ocean. *Journal of Biogeography*  
1269 39(11), 2081–2088. doi:10.1111/j.1365-2699.2012.02758.x