The role of systematics for understanding ecosystem functions:
Proceedings of the Zoologica Scripta Symposium, Oslo, Norway, 25
August 2022
Peer-reviewed author version

The role of systematics for understanding ecosystem functions: Proceedings of the Zoologica Scripta Symposium, Oslo, Norway, 25 August 2022

Lutz Bachmann[1][†][*], Jan Beermann[2, 3], Thomas Brey[2, 3, 5], Hugo J. de Boer[1], Jennifer Dannheim[2], Bente Edvardsen[6], Per G. P. Ericson[7], Kevin C. Holston[7], Veronika A. Johansson[7], Paul Kloss[2, 3], Rebecca Konijnenberg[2], Karen J. Osborn[8], Paula Pappalardo[8], Hendrik Pehlke[2], Dieter Piepenburg[2, 3, 4], Torsten H. Struck[1], Per Sundberg[9], Stine Svalheim Markussen[10], Katharina Teschke[2, 3], Maarten P. M. Vanhove[11]

[†] Authors are listed in alphabetical order, and all authors contributed equally to the paper

[1] Natural History Museum, University of Oslo, PO Box 1172 Blindern, 0318 Oslo, Norway
[2] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Functional Ecology, Am Handelshafen 12, 27570 Bremerhaven, Germany
[3] Helmholtz Institute for Functional Marine Biodiversity at Oldenburg University, Ammerländer Heerstraße 231, 26129 Oldenburg, Germany
[4] Institute for Ecosystem Research, Kiel University, Olshausenstraße 75, 24118 Kiel, Germany
[5] University of Bremen, Bibliothekstraße 1, 28359 Bremen, Germany
[6] Section for Aquatic Biology and Toxicology, Department of Biosciences, University of Oslo, PO Box 1066 Blindern, 0316 Oslo, Norway
[7] Department of Bioinformatics and Genetics, The Swedish Museum of Natural History, SE 10405 Stockholm, Sweden
[8] Department of Invertebrate Zoology, Smithsonian National Museum of Natural History, Washington, DC, USA
[9] Department of Marine Sciences, Gothenburg University, PO Box 461, SE 40530, Gothenburg, Sweden
[10] Norwegian Biodiversity Information Centre, PO Box 1285 Torgarden, 7462 Trondheim, Norway
[11] Hasselt University, Centre for Environmental Sciences, Research Group Zoology: Biodiversity & Toxicology, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium

* coordinating author for correspondence:
Lutz Bachmann
Natural History Museum
University of Oslo
PO Box 1172 Blindern
0318 Oslo
Norway
bachmann@nhm.uio.no

ORCID-IDs:
Lutz Bachmann: 0000-0001-7451-2074
Jan Beermann: 0000-0001-5894-6817
Thomas Brey: 0000-0002-6345-2851

Jennifer Dannheim: 0000-0002-3737-5872
Hugo J. de Boer: 0000-0003-1985-7859
Bente Edvardsen: 0000-0002-6806-4807
Per G. P. Ericson: 0000-0002-4143-9998
Kevin C. Holston: 0000-0002-0786-4069
Veronika A. Johansson: 0000-0002-3028-9947
Paul Kloss: 0000-0002-5587-0850
Rebecca Konijnenberg: 0000-0002-3804-5365
Paula Pappalardo: 0000-0003-0853-7681
Hendrik Pehlke: 0000-0003-1916-7831
Dieter Piepenburg: 0000-0003-3977-2860
Karen J. Osborn: 0000-0002-4226-9257
Per Sundberg: 0000-0001-6726-0075
Torsten H. Struck: 0000-0003-3280-6239
Katharina Teschke: 0000-0001-9595-7443
Maarten P. M. Vanhove: 0000-0003-3100-7566

Bachmann, L. *et al..* (2023) The role of systematics for understanding ecosystem functions: Proceedings of the Zoologica Scripta Symposium, Oslo, Norway, 25 August 2022. *Zoologica Scripta* 00, 000-000.

Abstract

On 25 August 2022 the *Zoologica Scripta - An International Journal of Systematic Zoology* and the Norwegian Academy of Sciences and Letters arranged a symposium entitled 'The role of systematics for understanding ecosystem functions' in the Academy's premises in Oslo, Norway. The symposium aimed at offering a forum for exploring and discussing trends and future developments in the field of systematics. Eleven international experts contributed expertise on various issues related to global challenges, such as biodiversity assessments, databases, cutting-edge analysis tools, and the consequences of the taxonomic impediment. Here, we compiled a multi-author proceedings paper of the symposium contributions that is arranged in chapters and presents the content and the key conclusions of the majority of the presentations.

Keywords

1 Motivation and Introduction

Lutz Bachmann, Per G. P. Ericson, Per Sundberg

Many biologists are concerned about the taxonomic impediment recognized already in 1992 by the Convention on Biological Diversity signed in Rio de Janeiro, Brazil. The decline of expertise in taxonomy and systematics; even more so in the context of ongoing biodiversity loss and climate change. At the same time, we experience an upgrading of the portfolio of methods to highly advanced and complex 'big data' analyses, either project generated, retrieved from databases or a combination of both. A frequently heard concern is that this trend may further decrease expertise in taxonomy and systematics and increase the knowledge gap. The trends towards more complex bioinformatics analyses as well as addressing and understanding systematics in an ecological context is also reflected in the manuscripts submitted to the Zoologica Scripta - *An International Journal of Systematic Zoology*.

The taxonomic impediment is not only a scientific issue but also of importance for societal development. Roughly ten years ago the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services (IPBES) was established, aiming, among others, at advancing knowledge and data as well as at building capacities that strengthen the science-policy interface for biodiversity and ecosystem services for the conservation and sustainable use of biodiversity. In addition, national initiatives such as e.g., the Norwegian Biodiversity Information Centre (artsdatabanken.no) invest substantially in monitoring biodiversity and compiling public databases.

The Zoologica Scripta and its publisher, the Norwegian Academy of Sciences and Letters (DNVA), organized a symposium entitled 'The role of systematics for understanding ecosystem functions' in order to offer a forum for exploring and discussing trends and future developments in the field of systematics. International experts contributed a series of talks that focused on various issues related to among others global challenges, biodiversity assessments, databases, cutting-edge analysis tools. The detailed program can be found at https://dnva.no/detskjer/2022/06/role-systematics-understanding-ecosystem-functions.
Roughly 70 participants attended the symposium that was held at DNVA's premises in Oslo, Norway, on 25 August 2022 and contributed to the lively discussions.

We editors of the Zoologica Scripta think that the contributions to the symposium are of great interest to many scientists in the field that could not attend the meeting. We, therefore, suggested to the presenters to contribute to a proceedings paper. This multi-author paper covers the major topics and conclusions of the symposium and beyond. It is arranged in several chapters based on the individual talks presented at the meeting. However, the chapter titles may differ from the original titles of the symposium talks.

The symposium closed with an evening dinner to which Matthias Glaubrecht contributed an inspiring talk. The essence of his evening lecture entitled 'On the end of evolution – Humankind and the annihilation of species' will be provided as a separate contribution to this issue of the Zoologica Scripta (Glaubrecht 2023).

Finally, we wish to express our gratitude to all contributors to and participants of the symposium that turned the arrangement into an exciting and stimulating event.

2 Taxonomy is changing but where is it heading?

Hugo J. de Boer, Stine Svalheim Markussen

Taxonomy and species discovery
In order to understand the role of taxonomy, we should never be afraid to ask ourselves why taxonomy and systematics matters. Numerous studies highlight the gap between expected and described biodiversity, and how this differs from one organism group to another (Colwell & Coddington, 1994; Costello, 2015; Laurance & Edwards, 2011). To many scholars of taxonomy there is no doubt that discovering and describing all diversity on Earth is a matter of general importance (Bebber et al., 2010; Mace, 2004; Padial et al., 2010), and accurate species recognition underpins our knowledge of global biodiversity (Chapman, 2009; Pitman & Jørgensen, 2002; Stuart et al., 2010). The discipline of taxonomy has nevertheless long been neglected and underfunded (Cotterill & Foissner, 2010; Pearson et al., 2011; Wägele et al., 2011; Wheeler et al., 2004; Wilson, 2004). Furthermore, across taxonomic groups new species continue to be described at rates that suggest a backlog in both species discovery and description (Antonelli et al., 2020; Barreca et al., 2020). Not all species groups yield as many new species as others. Some groups are well known, others are poorly known. Some are easy to study, whereas others are very challenging (Jörger & Schrödl, 2013). Whether we look at terrestrial or marine ecosystems – new species continue to be discovered everywhere (Cheek et al., 2020; Glaw et al., 2012; Rouse et al., 2004).

Also, species are not discovered at equal rates across the world. In Europe and North America few species are discovered, whereas in Asia and South America many. In Africa very few new species are discovered due to a lack of research (Antonelli et al., 2020). While Norway and the other Nordic countries have relative low species diversity compared to most countries and are seen as well investigated when it comes to species richness, the Norwegian Taxonomy Initiative and related projects have shown that there still is a significant number of species to be discovered (Elven & Søli, 2021). Since its beginning in 2009, the Norwegian Taxonomy Initiative has funded 120 projects aimed to investigate poorly known species groups, and these have discovered more than 4000 species new to Norway. Elven & Søli (2021) estimate that only about 65 % of the total number of species present in Norway have been found to date. Based on species discoveries we can also predict the potential of future discoveries. An analysis of terrestrial vertebrate taxa discovery identifies those groups were the highest potential for new species discovery remains (Moura & Jetz, 2021). Taxonomy is not only important for the description of new species *per se*, but is also a great way of communicating the value of biodiversity to an audience that is increasingly disconnected from nature. The tiny chameleon *Brookesia nana* (Glaw et al., 2021), which may be the world's smallest reptile species, is an example of a species description that has generated a lot of interest in biodiversity. Another example is the discovery of a new Giant Waterlily hidden in plain sight at Kew (Smith et al., 2022). *Victoria boliviana* Magdalena & L.T.Sm. had been growing at London's Kew Gardens for 177 years and in the National Herbarium of Bolivia for 34 years, but botanists had not

realized that it was distinct from its closely related *V. amazonica* (Poepp.) J.C. Sowerby. The latter is also a good example of how not all undiscovered species are hiding in pristine rainforests, but that some are already collected in natural history museums and botanic gardens.

Species loss – a race against the clock
However, species discovery is a race against the clock, with species loss far outpacing species discovery (Humphreys et al., 2019). Around 1 million animal and plant species are now threatened with extinction, many within decades, and more than ever before in human history (IPBES, 2019). The decline in wildlife; and the number of mammal species lost are indicators of the attrition of biodiversity across the board. Thirty-two million hectares of primary or recovering forest were lost between 2010 and 2015 (IPBES, 2019). This is an area nearly the size of Norway, and Norway is not a small country. A less widely communicated phenomenon is that in some of the most economically developed countries, we are losing species and rediscovering them too (Humphreys et al., 2019). This highlights that species loss can only be accurately quantified if species diversity is known. Assessing species loss in the rest of the world can only be measured by extrapolation and approximation. Taxonomic impediment is the world-wide shortage of this important taxonomic information, the gaps in our taxonomic knowledge, and the shortage of trained taxonomists and curators to fill this need. The taxonomic impediment is also often invoked to stress the unequal distribution of taxonomic knowledge. The study by Lücking (2020) highlights this dismayingly. Nearly all of the 200 most prolific fungal taxonomists, responsible for nearly half of all approximately 360,000 species names described, lived or live in the global north.

The value of biodiversity
Communicating the value of biodiversity can be done in many ways but putting it in economic terms makes it easier to balance with other economic activities. The 2021 World Bank report, "The Economic Case for Nature: A global Earth-economy model to assess development policy pathways" (Johnson et al., 2021), estimates that protecting nature could avert global economic losses of $2.7 trillion per year by 2030. This type of doom value is alarming but doesn't help us understand in what ways biodiversity contributes to ecosystem services, in terms of livelihoods and sustainability. The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) uses 18 "Nature's Contributions to People" to analyze and value ecosystem services (Chaplin-Kramer et al., 2019; Díaz et al., 2018; Hill et al., 2021; Pascual et al., 2017). A clear framework helps stakeholders communicate the value of nature to human livelihoods and sustainability. Similarly, the UN Sustainable Development Goals (SDGs) provide a framework to communicate integrated sustainability objectives. This visualization by the Stockholm Resilience Centre shows the hierarchical organization of the SDGs. Many SDGs are related to economy and society but all rely upon a healthy biosphere. Understanding biodiversity underpins the SDGs for Life below Water and Life on Land, and its heavy dependence on biodiversity knowledge.

Is species discovery a priority?
However, can we equate biodiversity knowledge with systematics and taxonomy? How much do we need to know about biodiversity if we can already assess and quantify its value? If we know the species we need, do we need to describe the rest? Do we need to stick a label on each creature in our ecosystem? The apparent gap between taxonomic knowledge and biodiversity has been highlighted by many scholars (Giangrande, 2003; Kim & Byrne, 2006; McNeely, 2002). We have arguably entered the sixth mass extinction, with species loss being accelerated by human-induced effects like never before (Ceballos et al., 2015). We are currently in a situation where for many species we are unaware that they are at risk of extinction because we

know them so poorly (Boehm & Cronk, 2021; Reaka-Kudla, 2001; Scheffers et al., 2012; Tedesco et al., 2014). Although new species are continuously being described, at the same time, others are going extinct. Plant extinctions are shown to occur up to 500 times faster today than in pre-industrial times (Humphreys et al., 2019). Unfortunately, many more species are going extinct without us knowing about it or even having discovered them (Boehm & Cronk, 2021; Scheffers et al., 2012; Tedesco et al., 2014). If species discovery is a priority, we are in a hurry to ramp things up if we want to describe a fraction of the species being lost.

Taxonomy today – new ways in which taxonomy contributes
The CBD Global Taxonomy Initiative Forum for the post-2020 Global Biodiversity Framework (Abrahamse et al., 2021), launched a call for action stressing that "Innovative and emerging technologies provide unprecedented opportunities for generating and sharing knowledge about the biosphere, when combined with essential taxonomic knowledge, techniques and skills." Taxonomic knowledge is essential to generating and sharing knowledge about the biosphere, but in order to accelerate this work we need to harness the opportunities provided by emerging technologies. New methods empower taxonomists like never before. DNA evidence can provide a lead to reassess previously overlooked diversity (Hebert & Gregory, 2005); and molecular species delimitation can aid in separating species that are hard to recognize using morphology alone (Luo et al., 2018). The advent of DNA sequencing technologies has also helped to uncover a previously unknown world of cryptic species (Struck et al., 2018). Studies looking at environmental DNA (eDNA) have opened a Pandora's box of taxonomic needs with many studies being able to identify only a minority of operational taxonomic units (OTUs) (Arroyo et al., 2016; Carvalho et al., 2019; Gleason et al., 2022; Sinniger et al., 2016). In order to increase the resolution and use of biodiversity assessments through eDNA, we need to generate DNA barcodes from reference specimens, which in turn are based on morphologically identified specimens deposited in natural history collections, which in turn require taxonomists and species descriptions. The more we want to employ molecular biodiversity assessments for biodiversity monitoring the greater the need for taxonomic expertise. Within the Norwegian setting, these interlinked dependencies support each other. The Norwegian Environment Agency has a strong focus on the development of sampling, data and analysis standards for eDNA biodiversity monitoring. This helps the private and institute sector use validated methods for reproducible studies. The Taxonomy Initiative program of the Norwegian Biodiversity Information Centre supports species discovery and mapping of poorly known biodiversity and liaises with the distributed Norwegian Barcode of Life (NorBOL) infrastructure to ensure that all specimens are barcoded and sequences deposited in public data repositories. The Taxonomy Initiative projects stimulate transmission of taxonomic knowledge, generation of new taxonomic knowledge and help train a new generation of taxonomists within the academic and institute sector.

Taxonomy tomorrow – what does the future need?
Whereas a taxonomist in the past might have been able to focus mostly on collection, identification and description of species, a taxonomist today can benefit from a more diverse skillset to succeed within academia. The Norwegian Biodiversity Information Centre funded Scandinavian Research School in Biosystematics – ForBio, coordinated by the four Norwegian university museums, provides training in core taxonomy subjects as well as other relevant methods, including genomics, phylogenetics, phylogenomics, biogeography, grant writing, publication and presenting. It is hard to know what skillset will be helpful to have in the future but having expert taxonomic skills that can contribute to integrative and collaborative projects is certainly essential. Taxonomic skills are in increasing demand in collaborative research and being able to understand and highlight where these skills can yield synergies is important in

promoting the field. Recent publications drawing on taxonomic expertise suggest that a future taxonomist should be a taxonomic expert first and foremost, but also able to accelerate species description by complementing traditional morphology with novel data such as molecular evidence and deep learning approaches (Coleman & Radulovici, 2020; Padial et al., 2010; Valan et al., 2021; Yang et al., 2022). As a field we should increase rates of species diversity monitoring to track ecosystem health, decline and recovery, and advocate for a more equal sharing of taxonomic knowledge through Findible, Accessible, Interoperable and Reusable data.

3 The Potential and Power of Biodiversity Databases

Veronika A. Johansson, Kevin C. Holston

Biodiversity Databases – Purpose and Procedures
As the scale and scope of biological data grows and becomes available for use and reuse, examining its persistence in various aggregated forms provides insight into the range of intended purposes, or target user groups, for biodiversity databases. The category "biological data" encompasses any information derived from living organisms and their products (Dhillon, 2019; Gauthier et al., 2019), but the scope of biodiversity databases is often limited to a single data class. These can include species traits, nucleotide and protein sequences, or observation records with proprietary and additional descriptive information often associated with these individual records. In practice, these data classes usually represent transformation and refinement of raw data into datasets that are fit for use. Methods in biodiversity informatics generate these raw data via measurements from biosensors or human observations; equivalent methods generate and refine laboratory data into research datasets, for example, DNA sequence or genomic data.

Application of similar analysis methods to legacy as well as novel datasets reinforces standardization of practices in dataset compilation, which characterize scientific works subject to peer-review. In this way, community-driven practices can support the growth and maintenance of specialized databases relevant to data from particular research communities (e.g., DNA sequence and genomic data). The disadvantages of centralization often outweigh, however, the benefits from managing encyclopedic digital resources on biological or ecological units of interest, such as living cells or species. There are, however, strong, parallel interests in maintaining data warehouses to leverage the vast amounts of available biological data towards, for example, developing ecosystem and whole cell models (Karr *et al.* 2012; Gauthier 2019; Teschke *et al.* 2022).

In the domain of biological data categorized as species observations, associated vouchers, and public repositories, the Global Biodiversity Information Facility (GBIF; www.gbif.org) plays a central role in guiding data discovery and use (Heberling et al., 2021). GBIF functions as a warehouse of these classes of biological data (Lane & Edwards, 2007), a central registry and aggregator whose application of Darwin Core (DwC) data standards supports efficient transformation into accessioned datasets for research, from taxonomic revisions to ecosystem studies at various geospatial scales. The associated international, open-access biodiversity data network and infrastructure provide open access to biological data as an integrated resource, whose contributions necessitate broad application of best practices throughout the community.

Maintaining data aggregations requires technical support for unique, persistent, and actionable identifiers for data and datasets. Although data providers aim to present reliable and traceable data, this challenge is addressed by tools such as the Integrated Publishing Toolkit (IPT), which guides the interpretation of potential GBIF datasets as DwC archives while supporting publication of accessioned datasets with rich metadata (Edwards, 2000; Robertson et al., 2014). GBIF is a distributed facility, comprising a network of participant nodes and its corresponding network of systems and data managers. Strategic changes in data pipelines are intended to expand this network through coordinated deployment of new features across the biodiversity data publishing community (Martin et al., 2020).

Biodiversity Databases – Potential

Coherence and persistence of these databases represents the intention of scientists to anchor reproducibility to their analyses and establish comprehensive knowledge bases relevant to core research areas within their disciplines. Standards for data storage and exchange are meant, in this context, to increase interoperability among datasets and reduce barriers to contributions and effective use of data by stakeholders. Additionally, guidelines for applying fundamental characterizations to datasets, like taxonomic determinations and geospatial descriptors, become increasingly important elements of systems operations.

The most promising methods underscore the importance of facilitating even higher levels of interoperability across domains, such as establishing persistent identifiers (Güntsch et al., 2017), reinforcing the fundamental role of biodiversity databases as discipline-specific resources. A long history of publishing scientific research is increasingly complemented by data publication, exemplified in the FAIR (Findability, Accessibility, Interoperability, and Reuse) and CARE principles (Wilkinson et al., 2015; https://www.go-fair.org/fair-principles/; https://www.gida-global.org/care). In this way, modern biodiversity databases are managed with greater focus on the democratic responsibility of research communities to provide access to data while opening possibilities for scientific investigations spanning multiple domains. Improving deficiencies in metadata can be a major step towards establishing FAIR data practices by increasing data discovery across disciplines and between data silos, such as individual peer-reviewed journals (Contaxis et al., 2022).

This transformational perspective characterizes GBIF data, where cross-disciplinary use spans all biological sciences, well beyond its original focus on species occurrences essential to ecology (Heberling et al., 2021). Considering the unprecedented attention of GBIF in its formative years on mobilizing biodiversity informatics data from public collections, this discipline-specific data has been aggregated in a format that emphasizes transformation and subsequent data repurposing.

Guidelines for generating GBIF metadata to promote data discovery have expanded to include harmonization identifying the publishing institutions, specimen repositories or collections, and individual dataset contacts. This is a function of the continuous work within this community that reaches beyond the technical infrastructure into external development having significant impact on data available to users. In this way, GBIF exemplifies how the potential of biodiversity databases still depends on discipline-specific influences for their establishment but is positively affected by community efforts to incorporate incentives to improve interoperability.

Biodiversity Databases – Power

Biodiversity data is rapidly increasing in magnitude and scope (König et al., 2019), which allows it to be leveraged to answer increasingly difficult questions in ecology and evolution while elevating its status within disciplines as critical specialist resources. Management strategies for these discipline-specific biological databases are shifting towards enhanced interoperability of systems supporting effective data pipelines. Archives of biological data are also shifting towards repositories under various levels of federation, with best practices promoting data persistence in technical frameworks supporting accessibility and reuse.

Biodiversity databases are heterogeneous with respect to data origin and underlying motivations for compiling the data. They are, however, coherent by design with the research communities defining circumscriptions for included data, generating biodiversity data, and guiding its long-term management as its primary stakeholders. Successful community management of discipline-specific data can result in greater levels of access through, for example, a micro services-centric perspective (Sheffield et al., 2022), particularly when status as transparent, reliable resources has been established.

Due to its focus on data interoperability, GBIF is well-positioned to lead the community in a shift from domain-specific databases to data hubs for biodiversity informatics. The concept of the Digital Extended Specimen, or DES, describes the emerging paradigm in which specimen data from the natural history collections are digitally linked to ecological, environmental, and refined biological data from other domains (Hardisty et al., 2022; Macklin et al., 2022). Outside the context of GBIF, however, motivation to normalize biodiversity data and support DES development is low when the emphasis is on sharing results from analysis or data aggregations rather than sharing the underlying data.

FAIR principles lead instead towards technical implementations handling community-supported data accession procedures as well as persistent identifiers for data and data types (Stocks et al., 2016; McMurry et al., 2017). GBIF publication protocols promote data standardization so it is discoverable and interpretable in a comparative context, allowing users, for example, to apply search criteria across individual datasets. Capturing variation among datasets as structured metadata, such as environmental sampling methods, greatly improves rigor in critical post-discovery tasks, like vetting data.

Heberling et al. (2021) presented comprehensive analyses demonstrating how biological data published by GBIF enables science in research areas beyond biology. Topics of study emphasized the diversity of data use within disciplines, with interdisciplinary research successfully mediated by data standardization and integration. The authors also recognize the importance of interoperability as a guiding principle in future efforts by GBIF in support of access and reuse of biodiversity informatics data. Removing data from silos of origin improves the efficacy of attempts to identify biases and gaps in global data sets and achieve levels of taxonomic and geographical data coverage in analyses that would otherwise be impossible (König et al., 2017). Decreasing levels of heterogeneity among datasets improves the ability to find, access and combine data, securing its reuse in further scientific analysis and its long-term availability (Teschke et al., 2022).

Biodiversity Databases – Conclusions
Biodiversity databases constitute an important component of studies in ecology, evolution, and ecosystem dynamics – securing accessibility to data is considered a necessary precursor to, as well as an expected result from research programs in these fields. As a target data repository in this domain, GBIF is a central registry and data aggregator whose application of community

standards allows biodiversity data to be published as fit-for-use datasets. Assessing the impact of GBIF data has shifted from descriptive metrics, such as record totals and taxonomic scope, to measuring interoperability with other user domains and systems of biological data. The increasing magnitude of data published by GBIF will continue to make impressive gains in the upcoming years, considering the growth of citizen science platforms that report species observations. Leveraging the power of available biological data will be a major challenge for GBIF, an endeavor that will improve its role as a data publisher and data hub, focused on interoperability to improve the potential for novel data use.

4 Understanding Biodiversity Change - the Potential of Combining Genomic Data and Taxonomic Challenges like Cryptic Species

Torsten H. Struck

In the last two decades, biology is experiencing a paradigm shift driven by technological advances. Like physics before, biology is changing into big data science at different fronts. Large amounts of data are generated by remote sensing technology, high-throughput imaging technologies, large databases for environmental parameters and so forth and are accessible to big data analysis in biology (Devictor & Bensaude-Vincent, 2016; Leonelli, 2019). With high-performance computing and tools like deep learning and artificial intelligence, we have the analytical power at hand to analyze such large-scale data in reasonable times and sensible ways (e.g., Li et al., 2019). One such example is sequencing technologies, which have changed tremendously allowing the generation of sequence data of several times the human genome in just a few days or hours (e.g., Lang et al., 2020). Moreover, some of the sequencing and other molecular biological technologies such as MinION have become so small that one can take them out into the field directly (Pomerantz et al., 2022). This allows molecular biomonitoring on the spot instead of transporting the samples back to the lab for this kind of analysis. Pomerantz et al. (2022) provided a protocol and a list of necessary equipment, which allowed species identification *in situ* based on a metabarcoding approach. This included DNA isolation, amplicon sequencing of barcodes using MinION, and downstream data analysis without reliance on Internet connectivity and could be conducted in less than 10 hours. This changes the way fieldwork can be done in biology and allows a much faster reaction time to for example adjust a study design. For example, if a rare species, which is not often detected, or a new invasive species, which has not been recorded from that region before, is detected this way in a bulk collection of species (e.g., a Malaise trap) in a very remote region, one can react to this directly on site and adjust the sampling design to take more similar samples. These can also be analyzed directly on the spot or taken to the laboratory back home depending on available time. In former times, the bulk samples would have been taken back to the lab, sorted and the species identified there. If one would like to follow-up on how abundant that species actually is one might have to go back to the remote region the next year or even only a couple of years later depending on possibilities. A consequence of these new possibilities might be that scientific excursions are planned differently in the near future including time allocations also for on spot data generation and flexible slots for tailored extended sampling. Srivathsan et al. (2019) showed the potential of the combination of metabarcoding with high-throughput sequencing. In their study, they caught 7,059 specimens in one Malaise trap placed for eight weeks in Kibale

National Park, Uganda. With an accuracy of 99.99% and 0.46% of undetermined nucleotides in the barcodes, they could assign specimens to a total of 650 phorid species, which exceeded the number of phorid species currently described for the entire Afrotropical region. 90% of the detected species belong to the neglected, megadiverse genus *Megaselia* (Srivathsan et al., 2019). Hence, the approach has tremendous potential. However, metabarcoding itself has some disadvantages such as primer fit and dropout, and other amplification biases, which can be overcome by metagenomics approaches (Obiol et al., 2020). Using metagenomics would also simplify procedures further to some degree as no amplification step is necessary. On the other hand, computational demands might increase depending on the quality of the generated data. Another requirement for metagenomics approaches is the availability of a sufficient and representative amount of reference genomes for all possibly collected groups.

Fortunately, also in this respect the new sequencing technologies brought substantial progress. Nowadays, it is possible to sequence high-quality genomes for different plants, fungi, protists and animals including non-model species and species with very large and complex genomes, with very small body sizes, with high ploidy levels or with low-quality preservation status (Cerca et al., 2022; Kingan et al., 2019; Kwiatkowski et al., 2021; Martín-Durán et al., 2020; Meyer et al., 2021; Schloissnig et al., 2021; Schneider et al., 2021; Varney et al., 2022). This has led to the proposal of a biological moonshot mission, the sequencing of a reference genome for each known eukaryotic species on Earth (Lewin et al., 2018). Since its proposal, this endeavor has gained enormous momentum and support with already several large consortia being established and often funded (for example, see Table 1). The advantage of genomic sequences for our understanding in ecosystem functioning, biodiversity assessment and conservation biology does not only rest in the advantages of metagenomics, but much more importantly in more detailed understanding of the evolutionary, population genetic and demographic history of the investigated species (e.g., Brandies et al., 2019). For example, studies of the Atlantic cod revealed that besides geographic divergences there are also clear genetic differences between migratory and non-migratory cod populations at the US- and the European Atlantic coast (Berg et al., 2017). These differences are associated with major inversions in four linkage groups, possibly chromosomes. These inversions comprise hundreds of genes, which facilitate coevolution of complex behavioral traits. The knowledge of the inversions allows for an efficient and reliable ecotype identification and hence monitoring of ecological diversity within and among harvested cod populations. Another example from conversation biology is the Iberian lynx (Abascal et al., 2016). Genomic studies showed a series of severe population bottlenecks within the Iberian lynx populations. These bottlenecks were associated with drastically reduced rates of weak-to-strong substitutions, low genome-wide genetic diversity with long stretches of regions of homozygosity, multiple signatures of genetic erosion and a high frequency of potentially deleterious variants. Moreover, the genomic data allowed the development of genome-wide SNP markers for monitoring (Kleinman-Ruiz et al., 2017). These markers resulted in improvements in the identification of individuals, assignment of parentage, and estimation of relatedness, ancestry and admixture. Hence, more powerful, efficient, and flexible tools for the genetic management and non-invasive monitoring of Iberian lynx populations could be used. Finally, genome-scale data can also allow better detection of species boundaries. For example, based on only the barcoding gene COI it was assumed that the earthworm *Lumbricus rubellus* comprised four cryptic species, but RADseq data revealed the presence of only one species (Giska et al., 2015). Hence, this species is characterized by deeply divergent mitochondrial lineages within populations, which result in genetic differences comparable to interspecific distances between other species. The deep population structure is possibly due to admixture of lineages from geographically separated glacial refugia, which did not separate reproductively, large effective population sizes with low migration rate, rapid

population expansion after a single transient bottleneck on a large ancestral population or a combination of these.

Insert Table 1 here within this chapter

Genomic data might be very powerful in revealing the difference between cryptic species and deep population structure, but cryptic species in themselves are also of importance in our understanding of ecosystem functioning. With the advent of molecular species identification, it has become apparent that many species, which were only morphologically delineated, are likely to be genetically different species (Bickford et al., 2007). Several papers have shown how erroneously recognized species can affect analyses of biodiversity estimates and with that of ecosystem functioning (e.g., Bálint et al., 2011; Fišer et al., 2018; Pante et al., 2015; Poulin & Pérez-Ponce de León, 2017). This might lead to wrong conclusions concerning the threat level of species or habitats. For example, for African giraffes it has been shown that it was not one, but four species increasing the endangered species status for some (Fennessy et al., 2016). However, ongoing taxonomic research in recent years has also shown that the recognition of cryptic species can be challenging and potentially misleading. This has recently been related to the fact that the definition of cryptic species was ambiguous as it was based on taxonomic history rather than biological properties (Fišer et al., 2018; Struck et al., 2018a; Struck et al., 2018b). Amended definitions of cryptic species separated the species delineation process from the assignment if a species is a cryptic species or not (Fišer et al., 2018; Struck et al., 2018b). Only species for which it could be shown that they are phenotypically more similar than one could expect should be considered cryptic species. This allows a more accurate definition and detection of cryptic species and hence insights into evolutionary processes (Struck & Cerca, 2019, 2022; Struck et al., 2018b). With respect to ecosystem function, climate change and the biodiversity crisis, one evolutionary process is of particular interest. It is the process of morphological stasis. Several cryptic species could be shown not to change at all or much and hence remain unchanged for millions of years (Fišer et al., 2018; Struck & Cerca, 2022; Struck et al., 2018b). On the other hand, this also means that these species experienced quite substantial environmental changes already and hence can be used as study systems to understand how species can withstand strong environmental changes seemingly without evolutionary adaptation. For example, the annelid *Stygocapitella* species complex, while it is globally distributed, consists of at least 12 species, which exhibit only four morphotypes (Cerca et al., 2020). Some of these morphotypes evolved tens of millions of years ago with up to 140 million years ago (Cerca et al., 2020). Hence, while the dinosaurs went extinct the morphotype of *S. pacifica*, *S. furcata* and *S. australis* remained unchanged and identical even though *S. pacifica* and *S. furcata* occur in the Northern hemisphere and *S. australis* in the Southern one. For the Northern Atlantic species *S. subterranea*, *S. westheidei* and *S. josemariobrancoi*, it could be shown that recent gene flow does not play a role in maintenance of their identical morphologies (Cerca et al., 2021). However, shared ancestral polymorphism and standing genetic variation could potentially play a role. Hence, cryptic species in general could be ideal systems to understand stasis in evolution and the responsiveness of species to ecosystem changes (Fišer et al., 2018; Struck & Cerca, 2019, 2022).

5 Model Parasites of Model Hosts? Evolution and Ecology of Cichlid-Infecting Monogenean Flatworms

Maarten P. M. Vanhove

Parasites are the most diverse, abundant, and threatened metazoan lifeforms. Despite their critical role in ecosystem functioning, most species are scarcely documented or remain undiscovered (Carlson et al., 2020). Given the assumption all species harbour parasites, and the estimate that most species are parasites, their evolutionary influence on biodiversity is considerable (Windsor, 1998, 2021). Apart from this species richness and evolutionary and ecological importance, other features render parasites excellent target organisms for research into a range of fundamental and applied biodiversity-related topics. Host organisms present their parasites with a living and hence changing niche, which leads to ample opportunities and mechanisms for parasite speciation. However, their potential as speciation models is far from optimally harnessed (Huyse *et al.*, 2005). On a more applied note, they may serve, for example, as markers for migration and invasion of their hosts (Gagne et al., 2022).

Their often minute size and "hidden" lifestyle pose logistical challenges to collecting, identifying, and sequencing parasites. For example, genomic work on helminths is hampered by difficulties to acquire them fresh and in sufficient quantities, and by a lack of experimental protocols. This delayed advances in helminth genomics compared to progress made in free-living model organisms (McVeigh, 2020). Sequencing effort is biased towards helminths of medical, zoonotic, veterinary or conservation concern (Poulin et al., 2019). On a practical level, most parasite taxa can indeed hardly be considered as established models.

Conversely, cichlid fishes are well-known, well-studied model organisms in many fields of (evolutionary) biology, including the study of behaviour (Jordan et al., 2021) and (ecological) diversification (Burress, 2015; Salzburger, 2018). To stay within the realm of genomics, it is, therefore, unsurprising, as Parsons *et al.* (2021) put it, *"[c]ichlid biology has a continuing history of benefitting from the latest technical advances in sequencing technology"*. The latter authors even underline the potential of cichlids in clinical disease research.

Despite this (disease) model status, cichlids' parasites and other symbionts are scarcely studied, and often in a quite fragmented way. A group of cichlid parasites, the monogenean flatworms infecting West African tilapias, have however been proposed as models in the study of parasite biodiversity, by Pariselle et al. (2003). This proposal rested on the species richness of both the fish and worm taxon, the close phylogenetic relatedness of a set of ecologically comparable and often sympatrically occurring hosts, the one-host lifecycle and quite narrow host-specificity of monogeneans, the substantial variation in monogenean species richness among host species, and the idea that under natural conditions monogeneans exert little selection pressure on their hosts. Now, two decades later, what is the state-of-the-art in the parasitological approach of cichlids as a candidate model for host-parasite interactions? Moreover, how is cichlid parasitology relevant to illustrate the role of systematics in understanding ecosystem functioning?

Case studies from our team's work on cichlid parasitology mainly focus on monogeneans parasitising African tilapias (of high policy relevance in view of their economic importance and their invasiveness), and cichlids from the African Great Lakes, "natural experiments" well-

known in evolutionary biology but hardly ever studied in the context of an explicit link between host evolution and parasite diversity. While these lakes are well-known for their species flocks in various taxa, resulting from radiation events, cichlid-infecting monogeneans belonging to *Cichlidogyrus* Paperna, 1960 provided the first known case of a parasite radiation in any of the African Great Lakes (Vanhove et al., 2015). What is more: these lineages of *Cichlidogyrus*, sequenced by Vanhove et al. (2015) from the gills of tropheine cichlids in Lake Tanganyika, appeared more species-rich than their hosts. Indeed, as had already been shown for other dactylogyrid monogeneans infecting cichlids and non-cichlid freshwater fishes, co-phylogenetic analysis indicated that within-host speciation often led to several species of *Cichlidogyrus* infecting a given host species. Much less common was the proposal of co-speciation as an important diversification mechanism in these dactylogyrids. It soon became clear, however, that not all cichlid lineages in Lake Tanganyika harbour such species-rich monogenean assemblages as the littoral tropheine cichlids, whose parasites belonging to *Cichlidogyrus* exhibit quite narrow host-specificity. Indeed, Kmentová *et al.* (2016, 2021) showed that the members of Bathybatini, cichlids from the (deep) open waters of Lake Tanganyika, share a single species of *Cichlidogyrus*, *Cichlidogyrus casuarinus* Pariselle, Muterezi Bukinga & Vanhove, 2015. They proposed that the lower density of non-littoral hosts explains the reduced host-specificity of this parasite. This exemplifies how host ecology influences parasite diversification processes. Moreover, the broad host range of *C. casuarinus* was already suggested together with the species' morphology-based description (Pariselle et al., 2015), illustrating how morphological systematics may underpin hypothesis building in ecology.

Knowing that there is no single pattern in parasite speciation in a context of host radiation throughout the Tanganyika cichlids, the question arises how radiation phenomena in cichlids influence host-parasite combinations at an African scale. Applying various forms of network analysis on all available infection data for members of *Cichlidogyrus* (including the nested *Scutogyrus* Pariselle & Euzet, 1995), Cruz-Laufer et al. (2022a) point to the role of the evolutionary history of the host in determining the host repertoire of these cichlid parasites. Cichlid radiations in a more stable environment seem conducive to more specialisation in host-parasite metacommunities. The authors also show host environment to be an important predictor of these cichlid-monogenean interactions, with host-switching more likely between ecologically similar hosts. This role of ecological opportunity has implications for aquaculture-environment interactions: it can contribute to the invasive potential of parasites that are co-introduced with the countless tilapia populations that have been anthropogenically translocated. As tilapia-infecting monogeneans may be (on their way to become) the most widespread tropical fish parasites worldwide (Shinn et al., 2023), the importance of a better understanding of these possibly co-invasive parasites can hardly be overstated.

Unfortunately, parasitological aspects of the consequences of tilapia introductions are poorly investigated (Deines et al., 2016). The fact that in Africa and other tropical regions, (fish) parasite diversity is disproportionately understudied (e.g. Poulin et al., 2019, 2020) exacerbates this research gap. For want of baseline data, it is difficult to establish whether tilapia parasites in a region that underwent tilapia introductions are there as a result of co-introduction (potentially followed by transmission to native fishes) or are naturally present in the area in the first place. A potential solution is the use of historical museum collections to establish a pre-introduction baseline of parasite communities. This approach allowed Jorissen et al. (2020) to propose certain changes in the monogenean gill parasite fauna of cichlids in Central Africa to be a result of tilapia introductions. Their study showcases the promise that biodiversity infrastructure and in particular host collections hold for advancing parasite ecology (Wood &

Vanhove, 2022) and disease research in general (Thompson et al., 2021) – a notion that garnered a lot of scientific and media attention in the context of COVID-19. As collection-based parasitology largely hinges on morphological parasite identification, it proves the importance of systematics in tackling timely ecological questions (Wood et al., 2023).

Even though the work by Jorissen et al. (2020) demonstrated that morphology-based taxonomy allows species-level identification of most of these gill-infecting worms retrieved from cichlids, looking at invasion biology through a parasitological lens gives rise to questions that require genetic methods. For example, even when parasite species shared between introduced tilapia and native fishes are native to the area, there may still be "hidden transmission" or "cryptic invasion" of conspecific parasite strains that cannot be discerned morphologically. Conclusively determining whether native and introduced parasite strains differ, and whether any of these strains have been transmitted between native and introduced hosts, requires a genetic approach. Indeed, a barcoding approach using a fragment of the COI gene allowed Jorissen et al. (2022) and Geraerts et al. (2022b) to identify such transmissions of alien strains (of native monogenean species) to indigenous host species (and in the latter study, also spill-back of native parasite strains to non-native hosts).

The resolution of such genetic approaches can also be increased to the extent that parasite genetics complement knowledge on their hosts, in this case on tilapia introduction history. Indeed, the magnifying glass hypothesis predicts that parasites, because their generation time is shorter than their host's, shed additional light on evolution and biogeography of their hosts (Nieberding & Olivieri, 2007). Geraerts *et al.* (2022a) demonstrated this using the highest-resolution markers currently available for these cichlid-infecting monogeneans: mitochondrial genome sequences. In particular for *Cichlidogyrus thurstonae* Ergens, 1981, the highest intraspecific diversity was found in the same part of Africa as based on results of COI barcoding of this parasite (Jorissen et al., 2022) and RAD sequencing data of its Nile tilapia host (Geraerts et al., 2022c). Moreover, the intraspecific genomic differentiation was, between rivers in the Congo Basin, and between African countries, several times higher for this monogenean than for Nile tilapia, confirming the magnifying glass potential (Geraerts, 2022). Importantly, of four tilapia-infecting dactylogyrid species whose mitogenomes were sequenced, *C. thurstonae* was the only candidate to test the magnifying glass hypothesis, as it was the only one with a clear geographical genetic structure (Geraerts et al., 2022a). Its host-specificity probably precluded other cichlid species to act as hosts and cloud the picture. It is remarkable that patterns of genetic diversity markedly differ between closely related monogenean species parasitizing the same hosts. This should again be an argument for the importance of thorough systematic studies before embarking on ecological hypothesis testing. Indeed, as co-infections are common, sequencing without prior worm identification would render it much less straightforward to discern patterns and processes in parasite populations.

As an example of recent work in which systematic, genetic and ecological understanding of this cichlid-monogenean model fed into each other, Cruz-Laufer et al. (2022b) provided an updated multimarker phylogeny of the lineage of *Cichlidogyrus* and *Scutogyrus*. Using this reconstruction in combination with machine learning they assessed the diagnostic value, and use, of morphological characters and host repertoire data typically used for phenotyping and classifying these monogeneans. It is hoped such work provides inspiration on how to improve identification and systematics of understudied animal taxa, and shows why parasite taxonomy should not be allowed to become a "lost art". Indeed, since the discovery of new species and new host-parasite combinations in the cichlid-*Cichlidogyrus* system still seems to be in its exponential stage (Cruz-Laufer *et al.*, 2021), taxonomic and faunistic research on these animals

will remain crucial in the foreseeable future. In view of the economic and ecological importance of cichlids, especially in the Global South, (capacity building in) morphological diagnostics of their parasites is a crucial step in acquiring baseline data to assess the health of aquatic environments and how it is anthropogenically impacted (Kapepula Kasembele et al., in press).

6 Exploring marine invertebrate biodiversity: CRITTERBASE – a roadmap to the solution of the FAIR biodiversity data challenge

Dieter Piepenburg, Thomas Brey, Jennifer Dannheim, Katharina Teschke, Jan Beermann, Rebecca Konijnenberg, Hendrik Pehlke, Paul Kloss

In times of rapidly increasing climate change (Doblas-Reyes et al., 2021; Gutiérrez et al., 2021) and multiple anthropogenic impacts on marine ecosystems and biodiversity (Meredith et. al., 2019; Constable et al., 2022), sound scientific investigations to enhance understanding, sustainable-use management and protection of biotas are of major importance (Wassmann et al., 2011). Research on marine organisms, populations and communities, and their interactions with each other and the environment, is fundamental in this context, but available data are still diverse and scattered. Data on marine biota exist in many formats and sources, such as published literature, data repositories, and unpublished material (Costello et al., 2018). Because of this heterogeneity, information can be difficult to find, access and combine, severely impeding its reuse for further scientific analysis and its long-term availability for future generations. Therefore, scientists, decision makers, and the public require versatile tools to compile, synthesize and manage biodiversity data in a transparent, efficient, and comprehensible way with high-level quality assurance (Teschke et al., 2022).

To address this challenge, we developed, implemented, and utilize CRITTERBASE (https://critterbase.awi.de; Teschke et al., 2022), a publicly accessible data warehouse and interactive portal that complies with the FAIR data principles (Findability, Accessibility, Interoperability, and Reusability). The purpose of CRITTERBASE is to complement (not substitute) existing long-term data storage repositories, such as the Ocean Biodiversity Information System (OBIS) or the Data Publisher of Earth and Environmental Science (PANGAEA), through providing a versatile platform facilitating management and analysis of

verified biodiversity data across multiple spatial and temporal scales and in various contexts (research, governance, management, public information). To this end, CRITTERBASE features an operational modular data-management environment including powerful and easy-to-use data ingest, retrieval and exploration options (Collector App, Data Space, and Analyst App, respectively; Fig. 1) for handling sample-based organism-related data from marine environments on a global scale.

Insert Figure 1 here within this chapter

The Data Space contains a data model that covers a variety of biotic data types (e.g., presence/absence, abundance, biomass), sub-sample processing options and sampling methods, and can be expanded easily. The data ingestion process via the Collector App includes a thorough quality check, e.g., through the validation of taxonomic names against the current classification provided by the World Register of Marine Species (WoRMS) (Fig. 1). Through its web interface hosted by the Computing Centre of the Alfred Wegener Institute (AWI), Helmholtz Centre for Polar and Marine Research Bremerhaven, CRITTERBASE is open to the public. Direct machine-to-machine communication with the web service of CRITTERBASE is also possible through a REST interface to allow for software-based data queries.

The Analyst App, currently under development, already features a key data extraction tool, aiding users with filtering options (by region, time frame, gear, dataset) to identify and download exactly the data they are looking for. These filter options will be expanded to include, for example, the search for specific taxa. Moreover, it will also provide basic data exploration applications by means of which, for example, the distribution of data values, the presence of outliers, and relationships between different data variables can be explored to provide data users with greater insight into the raw data.

A simple use case demonstrates an example of basic data exploration options by means of CRITTERBASE in combination with R (v4.2.0; R Core Team, 2022; RStudio Team, 2022). To examine habitat preferences of six North Sea polychaete species for certain sediment grain sizes (Armonies et al., 2021), R packages DBI (v1.1.3; R-SIG-DB et al., 2022) and RPostgres (v1.4.4; Wickham et al., 2022) are used to query data from four CRITTERBASE data sets with a total of 143 grab samples, containing species abundance data and information about median sediment grain size at sampling stations. By means of R package ggplot2 (v3.4.0; Wickham, 2016), a scatterplot of the species's abundances (individuals m^{-2}) versus sampled sediment grain sizes is produced (Fig. 2), visualizing their preference for fine sand environments. In addition, R package ggOceanMaps (v.1.3.4, Vihtakari, 2022) is used to produce maps showing the geographic location and number of the stations where the species were sampled (Fig. 3). The use case exemplifies the ease and simplicity with which basic data exploration can be conducted by tying R to CRITTERBASE, to visualize with only a few R commands distributions and relationships of the queried data and identify a potentially interesting area for further investigation.

Insert Figure 2 here within this chapter

Insert Figure 3 here within this chapter

The use case also highlights the role of user choices for data query, analysis and interpretation (Teschke et al., 2022). For instance, not all CRITTERBASE data sets contain the same type of

biotic information. While some of the four data sets queried feature standardized abundance data (individuals m^{-2}), others provide only the total number of individuals alongside information about sample area (m^2). The user must choose how to address these differences and run any necessary standardizations. In this use case, we also did not discriminate amongst life stages, meaning the number of samples given in Figure 3 differs slightly from the number of actual data points in Figures 2 and 3, as some stations contained two or more entries of the same species but for a different life stage. Moreover, some stations were also sampled repeatedly. Users must carefully evaluate whether further data refinement is needed. Lastly, Figure 3 underscores that users should also consider whether they can interpret a lack of data to present true absences. In this use case, the number of grab samples differed between species, as not all species were present at each station (Fig. 3). A detailed examination of metadata provided in CRITTERBASE can help to decide whether the lack of data of a species at certain stations could be interpreted as true absences.

At the moment, CRITTERBASE contains mostly benthic biodiversity data (a total of 414,666 records of 4,550 taxa identified in approx. 30,000 samples taken at 19,900 stations by means of grabs, trawls, and underwater imaging in the Arctic, North Sea, and Antarctic). However, its comprehensive and flexible data model allows for handling geo-referenced field data on any type of marine biota (benthos, plankton, nekton, etc.) and any type of sampling technique and gear. Its connection to the Research Data Commons (RDC), a cloud-based data infrastructure that is being developed under NFDI4Biodiversity, a consortium under the umbrella of the German National Research Data Infrastructure (NFDI), will help make CRITTERBASE the core of a national information system for biodiversity data on marine organisms. In the future, CRITTERBASE will also feature interfaces for data exchange with long-term repositories such as OBIS and PANGAEA. We envision CRITTERBASE to become a valuable and continuously expanding tool for a wide range of usages, such as studies of spatio-temporal biodiversity patterns, impacts and risks of climate change or evidence-based design of marine protection measures.

7 Critical taxonomic contributions for metabarcoding studies

Paula Pappalardo, Karen J. Osborn

Assigning taxon names is essential for biodiversity research and conservation (Mace, 2004). A detailed morphological assessment is the traditional approach to identifying an organism to the lowest possible taxonomic level. Morphological study of an individual organism has some advantages, such as the potential to identify life stages, but it requires high taxonomic expertise

and is time consuming. Barcoding, a cost-effective and fast method for species identification, was developed to help identify species and cryptic diversity. Barcoding for identification uses a DNA sequence for a specific genetic marker from the sample in question, and compares it to known sequences in a DNA reference database, hoping to match the unknown sequence to a known one. A natural extension, made possible by recent sequencing technology, is metabarcoding, where multiple organisms from a mixed sample (e.g., insect trap, plankton net) can be identified simultaneously in this same manner. When the sample is environmental (e.g., water, sediment), metabarcoding techniques are known as eDNA. Metabarcoding can facilitate rapid, large-scale biodiversity sampling of diverse environments, for example, monitoring for invasive species. But the list of organisms found using metabarcoding is affected by multiple factors, and the reliability of the taxonomic identifications depends heavily on the completeness and accuracy of the reference database.

Taxonomists have two key roles in improving taxonomic assignment for metabarcoding: 1) identifying understudied organisms and contributing high quality barcode sequences for those organisms to DNA reference databases, and 2) helping to curate publicly available DNA reference databases and report taxonomic issues. Ideally, a sequence included in a DNA reference database will have an associated voucher so that it can be reexamined if there is a taxonomic conflict. BOLD (Ratnasingham & Hebert, 2007), one of the first barcode reference databases, was built with this in mind and it even requires images to complete a specimen record. Other large sequence repositories such as GenBank do not require an associated voucher to publish a genetic sequence, but do have an identifier available to provide voucher numbers when submitting sequences. In addition, for some institutions (e.g., Smithsonian National Museum of Natural History) GenBank can link the voucher number with the institution's online catalogue for additional information such as collection and handling metadata. In cases where specimens are very small and have been fully utilized for DNA extraction, storage of images and DNA extractions help with future efforts to re-examine the identification if needed. Large barcoding efforts using museum specimens (when feasible) have proven useful not only to generate DNA barcodes associated with vouchers, but also to make available the associated metadata (Hebert et al., 2013).

Regardless of how inclusive DNA reference databases are, there are large gaps in representation documented for many taxonomic groups, even bioindicator species that are used to assess aquatic ecosystem quality (Weigand et al., 2019). The gaps in representation are usually larger for invertebrates than vertebrates and are pervasive even for the most commonly used genetic markers. We used the MIDORI2 reference database (Machida et al., 2017; Leray et al., 2022) to showcase how even for COI, the genetic marker with the largest number of available sequences, there is highly variable species representation (Fig. 4). Species representation can be as low as 0.3 (for Nematoda using the maximum richness estimates) or as high as 100% (phylum Cycliophora with just two known species). This means that taxonomists can have a big impact if they contribute reference sequences, particularly those of traditionally understudied taxa.

insert Figure 4 here within this chapter

In our experience, a small group of researchers can generate a large number of novel barcodes in a short time (Pappalardo et al., 2021).The Smithsonian StreamCode project, which assessed zooplankton biodiversity in the Gulf Stream during a single field season (two weeks, 14 taxonomists, 12 staff and students) contributed 2150 sequences (Pappalardo et al., 2021). At the time of submission, 109 were novel sequences to GenBank at the species level (43 for 18S and 66 for COI), 99 were novel at the generic level (33 for 18S and 66 for COI), and 44 were

novel at the family level (11 for 18S and 33 for COI) (Pappalardo et al., 2021). These sequences were compiled to create a local reference database that outperformed other databases in the taxonomic assignment of metabarcoding samples of zooplankton to phylum (Pappalardo et al., 2021). Similarly, other initiatives involving taxonomic experts and citizen scientists to combine intense biodiversity surveys of a specific area (in a short amount of time) with DNA barcoding and specimens' vouchers (e.g. https://www.abol.ac.at/en/abol-bioblitz-2019/) are an effective way to populate reference databases.

There are multiple DNA reference databases available - they differ in the genetic marker, which taxonomic groups are included, the source of the sequences, the taxonomic backbone, curation level, submission standards, and the number of sequences (among other variables). Even for a specific marker and taxonomic group, there may be multiple reference databases available. And unfortunately, the choice of reference database does matter. For example, different family-level composition was recovered when four different reference databases were used to analyze bacterial and archaeal community composition (Robeson et al., 2021). Similarly, in our work we have seen that the percent of correct species level assignments using COI for a test set of zooplankton can vary from 69% to 100% when using a 99% percent identity cutoff for BLASTn among different reference databases (Pappalardo et al., in prep.). In general, an increase in the number of sequences included in the reference database increases the number of correct assignments. For example, we observed that the MIDORI2 "unique" database (Leray et al., 2022), which includes all the unique haplotypes for each species, performed better that the "longest" version, which only contains the single longest sequence for each species (Pappalardo et al., in prep.).

Additionally, the likelihood of a sequence being assigned correctly is also influenced by the presence of errors in the reference database. There are several studies reporting errors in sequences published in GenBank (Nilsson et al., 2006; Leray et al., 2019; van den Burg et al., 2020). Inspecting some of the papers that report errors (Leray et al., 2019; van den Burg et al., 2020), we noticed that in most of the cases the errors reported are related to taxonomy. These may be cases where similar or identical sequences have different names because of synonyms or taxonomic revisions (e.g., new genus placement). The errors related to taxonomy can be easily fixed if reported, because the NCBI taxonomic team (Schoch et al., 2020) can revise them and fix them without the need to contact the sequence author. Taxonomic issues can be flagged by writing to the NCBI address info@ncbi.nlm.nih.gov. Consider writing "taxonomy question" in the subject line to help the issue reach the taxonomic team faster. The taxonomic issues for which we or our collaborators submitted an issue were solved within a week.

If another type of problem is found with a sequence (e.g., suspected contamination) it can be reported to gb-admin@ncbi.nlm.nih.gov for them to contact the sequence author. Currently only the submitter is able to modify the identity of a submission. When alerted of a possible error, GenBank can inspect the issue, contact the sequence author, and if deemed appropriate, flag the sequence with the tag "UNVERIFIED". Sequences with the UNVERIFIED tag do not appear in BLAST searches (https://www.ncbi.nlm.nih.gov/genbank/unverified/). Depending on the response of the submitting author, these errors can also be corrected.

Another issue that sequence authors can help with is to keep their taxon names updated when the taxonomic status changes. Even after publishing a new name, Schoch et al. (2020) reported that most submitters do not update their taxon names. Updates can be reported to gb-admin@ncbi.nlm.nih.gov and in our experience revision and changes occur quickly.

To fulfill the promises of metabarcoding, we must all work together to improve DNA reference databases. Reference databases can be improved by keeping taxon names updated in GenBank, reporting errors when we find them, and publishing sequences associated with vouchers. Incorporating molecular information with species descriptions could be another way to help expand reference databases (e.g., Osborn et al., 2011; Stoev et al., 2013). Understanding the strength and limitations of the reference database is also very important, and a sensitivity analysis comparing results from two different databases is an effective way to show the robustness of the results. In terms of the choice of genetic marker, not one genetic marker is perfect for all taxonomic groups, and multiple genetic markers can increase the chances of correct species detection. Our general advice for metabarcoding projects is to have an interdisciplinary team, including expertise in taxonomy, bioinformatics, ecology, and evolution. We think that the scientific community should encourage training opportunities for the new generation of taxonomists that will use both morphological assessment and molecular approaches for species identification.

8 Metabarcoding and Systematics of Protists in the Ocean to Understand Ecosystem Functioning

Bente Edvardsen

Protist diversity
The massive loss of biodiversity due to human activities and climate change is accelerating. We need to know the organisms, who they are, their distribution in time and space, and their ecological role in order to be able to conserve biodiversity. We also need to understand the driving forces and processes that shape biodiversity and the community structure to be able to predict the consequences of human activities and mitigate losses. By combining different methods in an integrative approach, a more detailed picture of the community composition, structure, distribution and function can be obtained.

Protists are all eukaryotes that are not animals, plants or fungi, and is thus a polyphyletic group. They are mostly tiny, single-celled organisms, and are much less studied than animals, plants and fungi, despite that they dominate eukaryotic diversity and play key functional roles in all ecosystems. They span a wide range of sizes, forms and life styles and their nutrition can be phototrophic as the algae, heterotrophic consumers as the protozoans, and mixotrophic which is everything in-between, and some are symbionts or parasites living inside other organisms. Phytoplankton, consisting of different algal lineages and cyanobacteria, are primary producers and the basis for most marine food-webs and fisheries. They have therefor been studied by light microscopy and experiments for more than a century to reveal production, community composition and seasonal dynamics, and relate this to abiotic factors (eg. Hjort & Gran 1902, Braarud et al., 1958). Microscopical phytoplankton long-time series data form the basis for analysis to understand changes over time (eg. Lundsør et al., 2022). Many scientific questions in plankton ecology are the same today as 100 years ago, and we still ask: who are there?

Where, when and how do they occur, and what are the driving factors? How much is produced, what are they doing and what are their ecological roles? New technologies have steadily emerged and changed the methodologies and expanded the possibilities of what we can discover. DNA-based methods such as metabarcoding have during the last decades become a common method in microbial diversity studies as it provides high taxonomic resolution and may also detect and identify tiny, fragile and rare taxa with few morphological traits (Lopes dos Santos et al., 2022). Here I will elucidate how different methods can be combined in biodiversity studies of protists in the ocean and discuss challenges and possibilities.

Metabarcoding of protists
In metabarcoding the organisms present in an environmental sample (eg. sea water, sediment, soil) can be identified by extracting DNA from the sample, amplify a marker gene of all the targeted organisms, sequence the obtained DNA-fragments by high-throughput sequencing, and compare the sequences to a reference library with sequences of known and identified organisms (see also Burki et al., 2021). The most widely marker gene region used in protist studies is the V4 or V9 regions of the 18S rRNA gene, which is present in all eukaryotes, it has both conserved and variable regions to enable classification from kingdom to species level, and it is the DNA region with most reference sequences available in gene databases, with a curated DNA reference sequence library for taxonomic assignation, the PR2 (Guillou et al., 2013, Lopes dos Santos et al., 2022). High throughput sequencing (HTS) with the Illumina Miseq pair-end technology gives the needed sequence length (up to ca 500 base pairs, bp) and quality to separate most taxa to genus and many to species level. Other HTS technologies are emerging that enable long-read sequencing, from 1500-5000 bp, and thus higher taxonomic resolution, such as PacBio and Nanopore, that are already used as a complement to Illumina (Burki et al., 2021). Metabarcoding has during the last 15 years revealed an unprecedented diversity and provided massive data on how protist diversity is distributed in space and time and combined with metadata (environmental information) it has contributed with a better understanding of driving forces structuring protist communities (e.g. de Vargas et al., 2015, Massana et al., 2015).

Challenges with metabarcoding
Metabarcoding also has some major challenges and limitations. The laboratory step includes a polymerase chain reaction (PCR) step where the marker gene is not amplified equally well for all taxa. No primer-pair for the amplification of the V4 18S rRNA gene match 100% to the DNA of all eukaryotic taxa (Vaulot et al., 2022). This step may also introduce false diversity by combining sequences from different organisms (chimeras, Egge et al., 2013). Different bioinformatic treatments may also give different results (Egge et al., 2013). Another challenge is that the sequences with Illumina technology are rather short (about 500 bp) and some species have identical DNA sequences in the marker region (eg. some diatom species in the V4 region, Luddington et al., 2012). Also, the number of copies of the marker gene in each cell varies, not only by cell volume (a large species usually has more copies than a small species within the same taxonomic group), but also between different taxonomic groups. So, the quantification ability is disputed and at best it gives the relative abundance of a taxon in each sample. Another challenge is the lack of sufficient reference sequences for taxonomic classification.

An integrative approach
The ability for quantification by metabarcoding has been tested in several studies, e.g. by comparing with the traditional microscopical cell counts. To be able to link available quantitative protist time-series and historical data based on light microscopy to future metabarcoding data, we need to know how they compare. Santi et al. (2021) compared the

protist composition and relative abundance from marine waters by metabarcoding and microscopical cell counts. They found that some taxonomic groups were represented rather equally with the two methods, such as diatoms, whereas others did not (dinoflagellates and ciliates) and that the differences became larger at low taxonomic levels. They concluded that metabarcoding provided a better estimate of the taxonomic richness of a community while microscopy provided more accurate quantitative data of abundance and biomass. The study showed that the two methods are complementary, and by using both methods a more detailed information on taxonomic composition, richness and abundance can be obtained (Santi et al., 2021).

The lack of sufficient DNA reference sequences is presently a major limitation for species identification by metabarcoding. For most protist groups only a fraction of the described species has been isolated into culture and been sequenced in some marker gene region. For the algal division Haptophyta, 320 species have been formally described and given a name, but only 100 of these have been sequenced in the 18S rRNA gene (Edvardsen et al., 2016). Environmental sequencing has detected more than 900 different genotypes that may represent species (Edvardsen et al., 2016), and thus the majority of these cannot be identified to the species level by molecular methods. Some haptophyte taxa known only from environmental sequences may even represent novel classes. In a metabarcoding study of the haptophyte diversity in the Oslofjorden, Norway, genotypes were detected that fell into clades representing three putatively novel classes, one novel order and one genus (Egge et al., 2015).

In order to infer the phylogeny and improve the taxonomy and reference databases of haptophytes, we and collaborators have isolated 30 new haptophyte strains into culture, sequenced and determined their phylogeny and obtained new haptophyte reference sequences (e.g. Edvardsen et al., 2000, 2011). We have also described several novel haptophyte species (e.g. Eikrem & Edvardsen ,1999, Eikrem, 1996, Seoane, 2009). We steadily discover protist species new to science, and there is a strong need to formally describe more protists.

Metabarcoding studies in the Arctic have shown a large unknown protist diversity and a need for more reference sequences (e.g. Egge et al., 2021). In the project TaxMArc we have isolated more than 200 strains from Arctic waters and ice either by single cell picking under the microscope or by a dilution series. The strains have been sequenced in rRNA genes, identified under the microscope and are now maintained in and available from the Norwegian Culture Collection of Algae (NORCCA, norcca.scrol.net). With cultures we can obtain both the morphology and long sequences of many genes, which enables discovery and description of novel species and a phylogenetic placement. Supraha et al. (2022) discovered that many species in the Arctic are morphologically similar to temperate species, but are genetically different and may represent novel species or varieties. A search against the 18S rRNA gene database metaPR2 (Vaulot et al., 2022), presently containing 41 metabarcode datasets from 4000 samples from marine waters, can reveal the geographical distribution where a genotype has been found before in the datasets, and under which environmental conditions. By using the metaPR2 database Supraha et al. (2022) discovered that several diatom genotypes isolated from Arctic Svalbard seems restricted to the Arctic, whereas other diatom genotypes had either an Arctic-temperate or cosmopolitan distribution.

Linking taxonomic diversity to functional diversity
In order to link a genotype to an ecological function we need to know different traits of the organisms such as e.g. their size, shape, motility, growth rate, preferred environment for growth, pigments, trophic mode, geographical distribution etc. With a trait table containing this

information of all species and genotypes in a sample we can obtain information on the ecological functionalities from metabarcoding data (see Martini et al., 2020). Ramond et al. (2019) have compiled and tested a trait table for protists from a few marine localities. However, much work remains to obtain a more complete trait table for marine protist species. With algal cultures available, growth responses to various growth conditions should be screened. Experiments for physiological and ecological traits such as organism interactions, (host-parasite and food-grazing) are needed to be able to better understand how the ecosystems function. Increasing genomic data from protists will also add to a better understanding of physiological and ecological functioning.

Conclusion

A majority of marine protist have not yet been characterized genetically and morphologically and lack reference DNA sequences. There is a need to isolate cultures, characterize and describe more protist species and provide more reference DNA sequences and expand and curate reference libraries. To be able to link taxonomic diversity to functional diversity we need more information on different traits (eg. morphology, ecology, physiology, behavior) and more comprehensive trait tables. Further, to be able to use metabarcoding in monitoring and link available time series and historical data based on light microscopy to new molecular data, we need to know how they compare. By combining several methods in an integrated approach, a more detailed analysis of the community composition, structure, distribution, ecological function and ecosystem functioning can be obtained.

References

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M., ... Godoy, J. A. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered iberian lynx. *Genome Biology*, 17, 251. https://doi.org/10.1186/s13059-016-1090-1

Abrahamse, T., Andrade-Correa, M., Arida, C., Galsim, R., Haüser, C., Price, M., & Sommerwerk, N. (2021). The Global Taxonomy Initiative in Support of the Post-2020 Global Biodiversity Framework (No. No. 96), CBD Technical Series. Secretariat of the Convention on Biological Diversity, Montreal.

Amante, C., & Eakins, B.W. (2009) ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA.

Antonelli, A., Smith, R.J., Fry, C., & Simmonds, M.S.J. (2020). State of the World's Plants and Fungi.

Armonies, W. (2021). Who lives where? Macrobenthic species distribution over sediment types and depth classes in the eastern North Sea. *Helgoland Marine Research*, 75, https://doi.org/10.1186/s10152-021-00552-1

Arroyo, A.S., López-Escardó, D., de Vargas, C., & Ruiz-Trillo, I. (2016). Hidden diversity of Acoelomorpha revealed through metabarcoding. *Biology Letters*, 12. https://doi.org/10.1098/rsbl.2016.0674

Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., ... Nowak, C. (2011). Cryptic biodiversity loss linked to global climate change. *Nature Climate Change*, 1, 313-318. https://doi.org/10.1038/nclimate1191

Barreca, M., Spanò, V., Montalbano, A., Cueto, M., Díaz Marrero, A.R., Deniz, I., ... Bertoni, F. (2020). Marine Anticancer Agents: An Overview with a Particular Focus on Their Chemical Classes. *Marine Drugs*, 18. https://doi.org/10.3390/md18120619

Bebber, D.P., Carine, M.A., Wood, J.R.I., Wortley, A.H., Harris, D.J., Prance, G.T., ... Robson, N.K.B., Scotland, R.W. (2010). Herbaria are a major frontier for species discovery. *Proceedings* of the National Academy of Sciences *USA*, 107, 22169–22171. https://doi.org/10.1073/pnas.1011841108

Berg, P. R., Star, B., Pampoulie, C., Bradbury, I. R., Bentzen, P., Hutchings, J. A., ... Jakobsen, K. S. (2017). Trans-oceanic genomic divergence of atlantic cod ecotypes is associated with large inversions. *Heredity*, 119, 418-428. https://doi.org/10.1038/hdy.2017.54

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22, 148-155. https://doi.org/10.1016/j.tree.2006.11.004

Boehm, M.M.A., & Cronk, Q.C.B. (2021). Dark extinction: the problem of unknown historical extinctions. *Biology Letters*, 17, 20210007. https://doi.org/10.1098/rsbl.2021.0007

Braarud, T., Gaarder, K. R., & Nordli, O. (1958). Seasonal changes in the phytoplankton at various points off the Norwegian west coast Fiskeridirektoratets Skrifter Serie Havundersakelser, XII, 377 pp

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, 10, 846. https://doi.org/10.3390/genes10110846

Burki, F., Sandin, M. M., & Jamy, M. (2021). Diversity and ecology of protists revealed by metabarcoding. *Current Biology*, 31, R1267-R1280. https://doi.org/10.1016/j.cub.2021.07.066

Burress, E. D. (2015). Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. *Hydrobiologia*, 748, 7–27. https://doi.org/10.1007/s10750-014-1960-z

Carlson, C. J., Hopkins, S., Bell, K. C., Doña, J., Godfrey, S. S., Kwak, M. L., ... Wood, C. L. (2020). A global parasite conservation plan. *Biological Conservation*, 250, 108596. https://doi.org/10.1016/j.biocon.2020.108596

Carvalho, S., Aylagas, E., Villalobos, R., Kattan, Y., Berumen, M., & Pearman, J.K. (2019). Beyond the visual: using metabarcoding to characterize the hidden reef cryptobiome. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20182697. https://doi.org/10.1098/rspb.2018.2697

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., & Palmer, T.M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, 1, e1400253. https://doi.org/10.1126/sciadv.1400253

Cerca, J., Meyer, C., Purschke, G., & Struck, T. H. (2020). Delimitation of cryptic species drastically reduces the geographical ranges of marine interstitial ghost-worms (*stygocapitella*; annelida, sedentaria). *Molecular Phylogenetics and Evolution*, 143, 106663. https://doi.org/10.1016/j.ympev.2019.106663

Cerca, J., Meyer, C., Stateczny, D., Siemon, D., Wegbrod, J., Purschke, G., ... Struck, T. H. (2020). Deceleration of morphological evolution in a cryptic species complex and its link to paleontological stasis. *Evolution*, 74, 116-131. https://doi.org/10.1111/evo.13884

Cerca, J., Petersen, B., Lazaro-Guevara, J. M., Rivera-Colón, A., Birkeland, S., Vizueta, J., ... Martin, M.D. (2022). The genomic basis of the plant island syndrome in darwin's giant daisies. *Nature Communications*, 13, 3729. https://doi.org/10.1038/s41467-022-31280-w

Cerca, J., Rivera-Colón, A. G., Ferreira, M. S., Ravinet, M., Nowak, M. D., Catchen, J. M., ... Struck, T. H. (2021). Incomplete lineage sorting and ancient admixture, and speciation without morphological change in ghost-worm cryptic species. *PeerJ*, 9, e10896. https://doi.org/10.7717/peerj.10896

Chaplin-Kramer, R., Sharp, R.P., Weil, C., Bennett, E.M., Pascual, U., Arkema, K.K., ... Daily, G.C. (2019). Global modeling of nature's contributions to people. *Science*, 366, 255–258. https://doi.org/10.1126/science.aaw3372

Chapman, A.D. (2009). Numbers of living species in Australia and the world. *Australian Biodiversity Information Services*.

Cheek, M., Nic Lughadha, E., Kirk, P., Lindon, H., Carretero, J., Looney, B., ... Niskanen, T. (2020). New scientific discoveries: plants and fungi. *Plants, People, Planet*, 2, 371–388. https://doi.org/10.1002/ppp3.10148

Coleman, C.O., & Radulovici, A.E. (2020). Challenges for the future of taxonomy: talents, databases and knowledge growth. *Megataxa*, 1. https://doi.org/10.11646/megataxa.1.1.5

Colwell, R.K., & Coddington, J.A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 345, 101–118. https://doi.org/10.1098/rstb.1994.0091

Constable, A. J., Harper, S., Dawson, J., Holsman, K., Mustonen, T., Piepenburg, D., & Rost, B. (2022). Cross-Chapter Paper 6: Polar Regions. In: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 2319-2368, https://doi.org/10.1017/9781009325844.023

Contaxis, N., Clark, J., Dellureficio, A., Gonzales, S., Mannheimer, S., ... Holmes K. (2022). Ten simple rules for improving research data discovery. *PLOS Computational Biology,* 18, e1009768. https://doi.org/10.1371/journal.pcbi.1009768

Costello, M.J. (2015). Biodiversity: the known, unknown, and rates of extinction. *Current Biology*, 25, R368-71. https://doi.org/10.1016/j.cub.2015.03.051

Costello, M. J., Horton, T., & Kroh, A. (2018). Sustainable Biodiversity Databasing: International, Collaborative, Dynamic, Centralised. *Trends in Ecology & Evolution*, 33, 803–805, https://doi.org/10.1016/j.tree.2018.08.006

Cotterill, F.P.D., & Foissner, W. (2010). A pervasive denigration of natural history misconstrues how biodiversity inventories and taxonomy underpin scientific knowledge. *Biodiversity and Conservation*, 19, 291–303. https://doi.org/10.1007/s10531-009-9721-4

Cruz-Laufer, A. J., Artois, T., Pariselle, A., Smeets, K., & Vanhove, M. P. M. (2021). The cichlid-*Cichlidogyrus* network: A blueprint for a model system of parasite evolution. *Hydrobiologia*, 848, 3847–3863. https://doi.org/10.1007/s10750-020-04426-4

Cruz-Laufer, A. J., Artois, T., Koblmüller, S., Pariselle, A., Smeets, K., Van Steenberge, M., & Vanhove, M. P. M. (2022a). Explosive networking: The role of adaptive host radiations and ecological opportunity in a species-rich host–parasite assembly. *Ecology Letters*, 25(8), 1795–1812. https://doi.org/10.1111/ele.14059

Cruz-Laufer, A. J., Pariselle, A., Jorissen, M. W. P., Muterezi Bukinga, F., Al Assadi, A., Van Steenberge, M., … Vanhove, M. P. M. (2022b) Somewhere I belong: phylogeny and

morphological evolution in a species-rich lineage of ectoparasitic flatworms infecting cichlid fishes. *Cladistics*, 38(4), 465–512. https://doi.org/10.1111/cla.12506

Deines, A. M, Wittmann, M. E., Deines, J. M., & Lodge, D. M. (2016) Tradeoffs among ecosystem services associated with global tilapia introductions. *Reviews in Fisheries Science & Aquaculture*, 24(2),178–191. https://doi.org/10.1080/23308249.2015.1115466

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., ... Karsent, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348, Issue 6237. https://doi.org/10.1126/science.1261605

Devictor, V., & Bensaude-Vincent, B. (2016). From ecological records to big data: The invention of global biodiversity. *History and Philosophy of the Life Sciences*, 38, 13. https://doi.org/10.1007/s40656-016-0113-2

Dhillon, S. K. (2019). Biological Databases. In "Encyclopedia of Bioinformatics and Computational Biology". Ranganathan S, Gribskov M, Nakai K, Schönbach C. (eds.) Academic Press, pp. 96-117. https://doi.org/10.1016/B978-0-12-809633-8.20198-2.

Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R.T., Molnár, Z., ... Shirayama, Y. (2018). Assessing nature's contributions to people. *Science,* 359, 270–272. https://doi.org/10.1126/science.aap8826

Doblas-Reyes, F. J., Sörensson, A. A., Almazroui, M., Dosio, A., Gutowski, W. J., Haarsma, R., ... Z. Zuo, Z. (2021). Linking Global to Regional Climate Change. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change[Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1363–1512, https://doi.org/10.1017/9781009157896.012

Edvardsen, B., Eikrem, W., Green, J. C., Andersen, R. A., Moon-van der Staay, S. Y., & Medlin, L.K. (2000). Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia*, 39, 19-35. https://doi.org/10.2216/i0031-8884-39-1-19.1

Edvardsen, B., Egge, E. S., & Vaulot, D. (2016). Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding - a review. *Perspectives in Phycology*, 3, 77-91. https://doi.org/10.1127/pip/2016/0052

Edvardsen, B., Eikrem, W., Throndsen, J., Sáez, A.G., Probert, I., & Medlin, L.K. (2011). Ribosomal DNA phylogenies and a morphological revision set the basis for a revised taxonomy of Prymnesiales (Haptophyta). *European Journal of Phycology*, 46, 202-228. https://doi.org/10.1080/09670262.2011.594095

Egge, E. S., Bittner, L., Andersen, T., Audic, S., de Vargas, C., Edvardsen, B. (2013). 454-pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLOS One,* 8, e74371. https://doi.org/10.1371/journal.pone.0074371

Egge, E. S. Eikrem, W., & Edvardsen, B. (2015). Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. *Journal of Eukaryotic* Microbiology, 62, 121-140. https://doi.org/10.1111/jeu.12157

Egge, E. S., Elferink, S., Vaulot, D., John, U., Bratbak, G., Larsen, A., & Edvardsen, B. (2021). An 18S V4 rDNA metabarcoding dataset of protist diversity in the Atlantic inflow to the Arctic Ocean, through the year and down to 1000 m depth. *Earth System Science Data, 13,* 4913–4928. https://doi.org/10.5194/essd-13-4913-2021

Eikrem, W., & Edvardsen, B. (1999). *Chrysochromulina fragaria* sp. nov. (Prymnesiophyceae) description of a new haptophyte flagellate from Norwegian waters. *Phycologia,* 38, 149-155. https://doi.org/10.2216/i0031-8884-38-2-149.1

EMODnet Bathymetry Consortium. (2020). EMODnet Digital Bathymetry (DTM 2020). EMODnet Bathymetry Consortium. https://doi.org/10.12770/bb6a87dd-e579-4036-abe1-e649cea9881a

Elven, H., & Søli, G.E. (2021). Kunnskapsstatus for artsmangfoldet i Norge 2020, Utredning for Artsdatabanken 1/2021. Artsdatabanken.

Fennessy, J., Bidon, T., Reuss, F., Kumar, V., Elkan, P., Nilsson, Maria, A., ... Janke, A. (2016). Multi-locus analyses reveal four giraffe species instead of one. *Current Biology*, 26, 2543-2549. https://doi.org/10.1016/j.cub.2016.07.036

Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27, 613–635. https://doi.org/10.1111/mec.14486

Gagne, R. B., Crooks, K. R., Craft, M. E., Chiu, E. S., Fountain-Jones, N. M., Malmberg, J. L., ... VandeWoude, S. (2022). Parasites as conservation tools. *Conservation Biology*, 36(1), e13719. https://doi.org/10.1111/cobi.13719

Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019) A brief history of bioinformatics. *Briefings in Bioinformatics*, 20, 1981–1996. https://doi.org/10.1093/bib/bby063

GEBCO. (2021). 15-arcsecond 2021 Grid, General Bathymatric Chart of the Ocean. https://www.gebco.net/data_and_products/gridded_bathymetry_data/

Geraerts, M. (2022). Invasive Nile tilapia and their monogeneans in sub-Saharan Africa: reconstructing stocking using both host and parasite. PhD thesis. Diepenbeek: Hasselt University.

Geraerts, M., Huyse T., Barson, M., Bassirou, H., Bilong Bilong, C. F., Bitja Nyom, A. R., … Vanhove, M. P. M. (2022a). Mosaic or melting pot: the use of monogeneans as a biological tag and magnifying glass to discriminate introduced populations of Nile tilapia in sub-Saharan Africa. *Genomics*, 114(3), 110328. https://doi.org/10.1016/j.ygeno.2022.110328

Geraerts, M., Huyse, T., Barson, M., Bassirou, H., Bilong Bilong, C. F., Bitja Nyom, A. R., … Vanhove, M. P. M. (2022b). Sharing is caring? Barcoding suggests co-introduction of dactylogyrid monogeneans and transfer towards native tilapias induced by introducing Nile tilapia in sub-Saharan Africa. In M. Geraerts, *Invasive Nile tilapia and their monogeneans in sub-Saharan Africa: reconstructing stocking using both host and parasite* (pp. 163–233). PhD thesis, Hasselt University.

Geraerts, M., Vangestel, C., Artois, T., Fernandes, J. M. O., Jorissen, M. W. P., Chocha Manda, A., … Huyse T. (2022c). Population genomics of introduced Nile tilapia (*Oreochromis niloticus* (Linnaeus, 1758)) in the Democratic Republic of the Congo: repeated introductions since colonial times with multiple sources. *Molecular Ecology*, 31(12), 3304–3322. https://doi.org/10.1111/mec.16479

Giangrande, A. (2003). Biodiversity, conservation, and the ?Taxonomic impediment? *Aquatic Conservation: Marine and Freshwater Ecosystems*, 13, 451–459. https://doi.org/10.1002/aqc.584

Giska, I., Sechi, P., & Babik, W. (2015). Deeply divergent sympatric mitochondrial lineages of the earthworm *lumbricus rubellus* are not reproductively isolated. *BMC Evolutionary Biology*, 15, 217. https://doi.org/10.1186/s12862-015-0488-9

Glaubrecht, M. (2023). On the end of evolution – Humankind and the annihilation of species. *Zoologica Scripta*, 52, in press

Glaw, F., Köhler, J., Hawlitschek, O., Ratsoavina, F.M., Rakotoarison, A., Scherz, M.D., & Vences, M. (2021). Extreme miniaturization of a new amniote vertebrate and insights into the evolution of genital size in chameleons. *Scientific Reports*, 11, 2522. https://doi.org/10.1038/s41598-020-80955-1

Glaw, F., Köhler, J., Townsend, T.M., & Vences, M. (2012). Rivaling the world's smallest reptiles: discovery of miniaturized and microendemic new species of leaf chameleons (Brookesia) from northern Madagascar. *PLoS One*, 7, e31314. https://doi.org/10.1371/journal.pone.0031314

Gleason, J.E., Hanner, R.H., & Cottenie, K. (2022). Hidden diversity – DNA metabarcoding reveals hyper-diverse benthic invertebrate communities. *BioRxiv*. https://doi.org/10.1101/2022.02.28.481642

Güntsch, A., Hyam, R., Hagedorn, G., Chagnoux, S., Röpert, D., Casino, A., ... Hoffmann, J. (2017). Actionable, long-term stable and semantic web compatible identifiers for access to biological collection objects. Database, 2017. https://doi.org/10.1093/database/bax003

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, ... Christen R. (2013). The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small SubUnit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41 (Database issue), D597-D604. https://doi.org/10.1093/nar/gks1160.

Gutiérrez, J. M., Jones, R. G., Narisma, G. T., Alves, L. M., Amjad, M., Gorodetskaya, I. V. ... Yoon, J. H. (2021). Atlas. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate

Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1927–2058, https://doi.org/10.1017/9781009157896.021.

Hardisty, A. R., Ellwood, E. R-, Nelson, G., Zimkus, B., Buschbom, J., Addink, W., … Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences USA*, 118(6), e2018093118. https://doi.org/10.1073/pnas.2018093118

Hebert, P. D. N., deWaard, J. R., Zakharov, E. V., Prosser, S. W. J., Sones, J. E., McKeown, J. T. A., … La Salle, J. (2013). A DNA 'Barcode Blitz': Rapid Digitization and Sequencing of a Natural History Collection. *PLoS ONE*, 8, e68535. https://doi.org/10.1371/journal.pone.0068535

Hebert, P.D.N., & Gregory, T.R. (2005). The promise of DNA barcoding for taxonomy. Systematic Biology, 54, 852–859. https://doi.org/10.1080/10635150500354886

Hill, R., Díaz, S., Pascual, U., Stenseke, M., Molnár, Z., & Van Velden, J. (2021). Nature's contributions to people: Weaving plural perspectives. *One Earth*, 4, 910–915. https://doi.org/10.1016/j.oneear.2021.06.009

Hjort, J., & Gran, H. H. (1900). Hydrographic-Biological Investigations of The Skagerrak and the Christiania Fiord. Report on Norwegian Fishery- and Marine-lnvestigations I, 1–107.

Humphreys, A.M., Govaerts, R., Ficinski, S.Z., Nic Lughadha, E., & Vorontsova, M.S. (2019). Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nature Ecology & Evolution*, 3, 1043–1047. https://doi.org/10.1038/s41559-019-0906-2

Huyse, T., Poulin, R., & Théron, A. (2005). Speciation in parasites: a population genetics approach. *Trends in Parasitology*, 21(10), 469–475. https://doi.org/10.1016/j.pt.2005.08.009

IPBES (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat, Bonn, Germany.

Jörger, K.M., & Schrödl, M. (2013). How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology*, 10, 59. https://doi.org/10.1186/1742-9994-10-59

Johnson, J.A., Ruta, G., Baldos, U., Cervigni, R., Chonabayashi, S., Corong, E., … Polasky, S. (2021). The Economic Case for Nature: A Global Earth-Economy Model to Assess Development Policy Pathways. World Bank. https://doi.org/10.1596/35882

Jordan, A., Taborsky, B., & Taborsky, M. (2021). Cichlids as a model system for studying social behaviour and evolution. In M. E. Abate & D. L. Noakes (Eds.), *The behavior, ecology and evolution of cichlid fishes* (Fish & Fisheries Series vol 40, pp. 587–635). Springer. https://doi.org/10.1007/978-94-024-2080-7_16

Jorissen, M. W. P., Huyse, T., Pariselle, A., Wamuini Lunkayilakio, S., Muterezi Bukinga, F., Chocha Manda, A., … Vanhove, M. P. M. (2020). Historical museum collections help detect parasite species jumps after tilapia introductions in the Congo Basin. *Biological Invasions*, 22(9), 2825–2844. https://doi.org/10.1007/s10530-020-02288-4

Jorissen, M. W. P., Vanhove, M. P. M., Pariselle, A., Snoeks, J., Vreven, E., Šimková, A., … Huyse, T. (2022). Molecular footprint of parasite co-introduction with Nile tilapia in the Congo Basin. *Organisms Diversity and Evolution*, 22, 1003–1019 https://doi.org/10.1007/s13127-022-00563-x

Kapepula Kasembele, G., Chocha Manda, A., Abwe, E., Pariselle, A., Muterezi Bukinga, F., Huyse, T., … Vanhove, M. P. M. (in press). First record of monogenean fish parasites in the Upper Lufira River basin (Democratic Republic of Congo): dactylogyrids and gyrodactylids infesting *Oreochromis mweruensis*, *Coptodon rendalli* and *Serranochromis macrocephalus* (Teleostei: Cichlidae). *Parasites & Vectors*.

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., ... Covert, M. W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell*, 2012;150:389–401. https://doi.org/10.1016/j.cell.2012.05.044

Kim, K.C., & Byrne, L.B. (2006). Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science. *Ecological Research*, 21, 794–810. https://doi.org/10.1007/s11284-006-0035-7

Kingan, S. B., Heaton, H., Cudini, J., Lambert, C. C., Baybayan, P., Galvin, B. D., ... Lawniczak, M. K. N. (2019). A high-quality de novo genome assembly from a single mosquito using pacbio sequencing. *Genes*, 10, 62. 10.3390/genes10010062

Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., ... Godoy, J. A. (2017). Novel efficient genome-wide snp panels for the conservation of the highly endangered iberian lynx. *BMC Genomics*, 18, 556. https://doi.org/10.1186/s12864-017-3946-5

Kmentová, N., Gelnar, M., Mendlová, M., Van Steenberge, M., Koblmüller, S., & Vanhove, M. P. M. (2016). Reduced host-specificity in a parasite infecting non-littoral Lake Tanganyika cichlids evidenced by intraspecific morphological and genetic diversity. *Scientific Reports*, 6, 39605. https://doi.org/10.1038/srep39605

Kmentová, N., Hahn, C., Koblmüller, S., Zimmermann, H., Vorel, J., Artois, T., … Vanhove M. P. M. (2021). Contrasting host-parasite population structure: Morphology and mitogenomics of a parasitic flatworm on pelagic deepwater cichlid fishes from Lake Tanganyika. *Biology*, 10, 797. https://doi.org/10.3390/biology10080797

König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J., & Kreft, H. (2019) Biodiversity data integration—the significance of data resolution and domain. *PLOS Biology*, 17, e3000183. https://doi.org/10.1371/journal.pbio.3000183

Kwiatkowski, D., Blaxter, M., Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, & Darwin Tree of Life Consortium (2021). The genome sequence of the bootlace worm, lineus longissimus (Gunnerus, 1770). *Wellcome Open Research*, 6. 272. https://doi.org/10.12688/wellcomeopenres.17193.1

Lane, M. A., & Edwards, J. L. (2007). The global biodiversity information facility (GBIF). Systematics Association special volume. 2007;73:1.

Lang, D., Zhang, S., Ren, P., Liang, F., Sun, Z., Meng, G., ... Liu, S. (2020). Comparison of the two up-to-date sequencing technologies for genome assembly: Hifi reads of Pacific Biosciences Sequel ii system and ultralong reads of Oxford Nanopore. *GigaScience*, 9. https://doi.org/10.1093/gigascience/giaa123

Laurance, W.F., & Edwards, D.P. (2011). The search for unknown biodiversity. *Proceedings* of the National Academy of Sciences *USA,* 108, 12971–12972. https://doi.org/10.1073/pnas.1110319108

Leray, M., Knowlton, N., Ho, S.-L., Nguyen, B. N., & Machida, R. J. (2019). GenBank is a reliable resource for 21st century biodiversity research. *Proceedings of the National Academy of Sciences USA*, 116, 22651–22656. https://doi.org/10.1073/pnas.1911714116

Leray, M., Knowlton, N., & Machida, R. J. (2022). MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. *Environmental DNA*, 4, 894–907. https://doi.org/10.1002/edn3.303

Leonelli, S. (2019). The challenges of big data biology. *eLife*, 8, e47381. https://doi.org/10.7554/eLife.47381

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., ... Zhang, G. (2018). Earth biogenome project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences USA*, 115, 4325-4333. https://doi.org/10.1073/pnas.1720115115

Li, Y., Huang, C., Ding, L., Li, Z., Pan, Y., & Gao, X. (2019). Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods*, 166, 4-21. https://doi.org/10.1016/j.ymeth.2019.04.008

Lopes dos Santos, A., Gérikas, R. C., Ong, D., Garczarek, L., Li, X., Nodder S.D., ... Gutiérrez-Rodríguez, A. (2022). Chapter 11 - Phytoplankton diversity and ecology through the lens of high throughput sequencing technologies. *In:* Advances in phytoplankton ecology. Applications of Emerging Technologies, Elsevier, p. 353-393.

Luddington, I. A., Kaczmarska, I., & Lovejoy, C. (2012). Distance and character-based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae). *PLoS One*, 7, e45664. https://doi.org/10.1371/journal.pone.0045664

Lücking, R. (2020). Three challenges to contemporaneous taxonomy from a licheno-mycological perspective. *Megataxa*, 1. https://doi.org/10.11646/megataxa.1.1.16

Lundsør, E., Eikrem, W., Stige, L., Engesmo, A., Gran Stadniczenko, S., & Edvardsen, B. (2022). Changes in phytoplankton community structure over a century in relation to environmental factors. *Journal of Plankton Research,* 44, 854–871. https://doi.org/10.1093/plankt/fbac055

Luo, A., Ling, C., Ho, S.Y.W., & Zhu, C.-D. (2018). Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology*, 67, 830–846. https://doi.org/10.1093/sysbio/syy011

Mace, G.M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 359, 711–719. https://doi.org/10.1098/rstb.2003.1454

Machida, R. J., Leray, M., Ho, S.-L., & Knowlton, N. (2017). Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data*, 4, 170027. https://doi.org/10.1038/sdata.2017.27

Macklin, J. A., Mast, A. R., Miller, J. T., Monfils, A. K., Paul, D. L., Wallis, E., & Webster, M. (2022). Digital Extended Specimens: Enabling an Extensible Network of Biodiversity Data Records as Integrated Digital Objects on the Internet, *BioScience*, 72, 978–987, https://doi.org/10.1093/biosci/biac060

Martin, D,. Molina, J., dos Remedios, N., Lecoq, M.-E., Robertson, T., Ruiz Jurado, V. J. (2020) Aligning GBIF and the Atlas of Living Australia. *Biodiversity Information Science and Standards*, 4: e59274. https://doi.org/10.3897/biss.4.59274

Martín-Durán, J. M., Vellutini, B. C., Marlétaz, F., Cetrangolo, V., Cvetesic, N., Thiel, D., ... Hejnol, A. (2020). Conservative route to genome compaction in a miniature annelid. *Nature Ecology & Evolution*, 5, 231–242. https://doi.org/10.1038/s41559-020-01327-6

Martini, S., Larras, F., Boye, A., Faure, E., Aberle, N., Archambault, P., ... Ayata, S.-D. (2021). Functional trait-based approaches as a common framework for aquatic ecologists. *Limnology and Oceanography,* 66, 965-994. https://doi.org/10.1002/lno.11655

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., ... de Vargas, C .(2015). Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environmental Microbiology*, 17, 4035–4049. https://doi.org/10.1111/1462-2920.12955

McMurry, J. A., Juty, N., Blomberg, N., Burdett, T., Conlin., T, Conte, N., ... Parkinson, H. (2017). Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biology*, 15, e2001414. https://doi.org/10.1371/journal.pbio.2001414

McNeely, J.A. (2002). The role of taxonomy in conserving biodiversity. *Journal for Nature Conservation,* 10, 145–153. https://doi.org/10.1078/1617-1381-00015

McVeigh, P. (2020). Post-genomic progress in helminth parasitology. *Parasitology*, 147(8), 835–840. https://doi.org/10.1017/S0031182020000591

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., ... Schuur E. A. G. (2019). Polar Regions. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. MassonDelmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 203-320. https://doi.org/10.1017/9781009157964.005

Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J. M., Irisarri, I., ... Schartl, M. (2021). Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature*, 590, 284-289. https://doi.org/10.1038/s41586-021-03198-8

Moura, M.R., & Jetz, W. (2021). Shortfalls and opportunities in terrestrial vertebrate species discovery. *Nature Ecology & Evolution*, 5, 631–639. https://doi.org/10.1038/s41559-021-01411-5

Natural Earth Data. (2022). 1:10m Physical Vectors with the Land and Minor Island datasets combined. naturalearthdata.com.

Nieberding, C. M., & Olivieri, I. (2007). Parasites: proxies for host genealogy and ecology? *Trends in Ecology & Evolution*, 22(3), 156–165. https://doi.org/10.1016/j.tree.2006.11.012

Nilsson, R. H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.-H., & Kõljalg, U. (2006). Taxonomic Reliability of DNA Sequences in Public Sequence Databases: A Fungal Perspective. *PLoS ONE*, 1, e59. https://doi.org/10.1371/journal.pone.0000059

Obiol, A., Giner, C. R., Sánchez, P., Duarte, C. M., Acinas, S. G., & Massana, R. (2020). A metagenomic assessment of microbial eukaryotic diversity in the global ocean. *Molecular Ecology Resources*, 20, 718-731. https://doi.org/10.1111/1755-0998.13147

Osborn, K. J., Haddock, S. H. D., & Rouse, G. W. (2011). Swima (Annelida, Acrocirridae), holopelagic worms from the deep Pacific. *Zoological Journal of the Linnean Society*, 163, 663–678. https://doi.org/10.1111/j.1096-3642.2011.00727.x

Osigus, H.-J., Rolfes, S., Herzog, R., Kamm, K., & Schierwater, B. (2019). Polyplacotoma mediterranea is a new ramified placozoan species. *Current Biology*, 29, R148–R149. https://doi.org/10.1016/j.cub.2019.01.068

Padial, J.M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, 16. https://doi.org/10.1186/1742-9994-7-16

Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., et al. (2015). Species are hypotheses: Avoid connectivity assessments based on pillars of sand. *Molecular Ecology*, 24, 525-544.

Pappalardo, P., Aradhey, A., & Osborn, K. J. (in prep.). Effects of the choice of reference database for metabarcoding of marine species.

Pappalardo, P., Collins, A. G., Pagenkopp Lohan, K. M., Hanson, K. M., Truskey, S. B., Jaeckle, W., … Osborn, K. J. (2021). The role of taxonomic expertise in interpretation of metabarcoding studies. *ICES Journal of Marine Science*. https://doi.org/10.1093/icesjms/fsab082

Pariselle, A., Morand, S., Deveney, M. R., & Pouyaud, L. (2003). Parasite species richness of closely related hosts: historical scenario and "genetic" hypothesis. In C. Combes & J. Jourdane (Eds.), *Taxonomie, écologie et évolution des métazoaires parasites. Taxonomy, ecology and evolution of metazoan parasites. Livre-hommage à Louis Euzet* (pp. 147–166). Presses Universitaires de Perpignan.

Pariselle, A., Muterezi Bukinga, F., Van Steenberge, M., & Vanhove, M. P. M. (2015). Ancyrocephalidae (Monogenea) of Lake Tanganyika: IV: *Cichlidogyrus* parasitizing species of Bathybatini (Teleostei, Cichlidae): reduced host-specificity in the deepwater realm? *Hydrobiologia*, 748(1), 99–119. https://doi.org/10.1007/s10750-014-1975-5

Parsons, K. J., McWhinnie, K., & Armstrong, T. (2021). An evo-devo view of post-genomic African cichlid biology: Enhanced models for evolution and biomedicine. In M. E. Abate & D. L. Noakes (Eds.), *The behavior, ecology and evolution of cichlid fishes* (Fish & Fisheries Series vol 40, pp. 779–802). Springer. https://doi.org/10.1007/978-94-024-2080-7_21

Pascual, U., Balvanera, P., Díaz, S., Pataki, G., Roth, E., Stenseke, M., ... Yagi, N. (2017). Valuing nature's contributions to people: the IPBES approach. *Current Opinion in Environmental Sustainability*, 26–27, 7–16. https://doi.org/10.1016/j.cosust.2016.12.006

Pearson, D.L., Hamilton, A.L., & Erwin, T.L. (2011). Recovery plan for the endangered taxonomy profession. *Bioscience*, 61, 58–63. https://doi.org/10.1525/bio.2011.61.1.11

Pitman, N.C.A., & Jørgensen, P.M. (2002). Estimating the size of the world's threatened flora. *Science*, 298, 989. https://doi.org/10.1126/science.298.5595.989

Pomerantz, A., Sahlin, K., Vasiljevic, N., Seah, A., Lim, M., Humble, E., ... Prost, S. (2022). Rapid in situ identification of biological specimens via DNA amplicon sequencing using miniaturized laboratory equipment. *Nature Protocols*, 17, 1415–1443 (2022). https://doi.org/10.1038/s41596-022-00682-x

Poulin, R., Hay, E., & Jorge, F. (2019). Taxonomic and geographic bias in the genetic study of helminth parasites. *International Journal for Parasitology*, 49(6), 429–435. https://doi.org/10.1016/j.ijpara.2018.12.005

Poulin, R., & Pérez-Ponce de León, G. (2017). Global analysis reveals that cryptic diversity is linked with habitat but not mode of life. *Journal of Evolutionary Biology*, 30, 641-649. https://doi.org/10.1111/jeb.13034

Poulin, R., Presswell, B., & Jorge, F. (2020). The state of fish parasite discovery and taxonomy: a critical assessment and a look forward. *International Journal for Parasitology*, 50(10-11), 733–742. https://doi.org/10.1016/j.ijpara.2019.12.009

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

R Special Interest Group on Databases (R-SIG-DB), Wickham, H., & Müller, K. (2022). DBI: R Database Interface. R package version 1.1.3, https://CRAN.R-project.org/package=DBI

Ramond, P., Sourisseau, M., Simon, N., Romac, S., Schmitt, S., Rigaut-Jalabert, F., ... Siano, R. (2019). Coupling between taxonomic and functional diversity in protistan coastal communities. *Environmental Microbiology*, 21, 730-749. https://doi.org/10.1111/1462-2920.14537

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Reaka-Kudla, M.L. (2001). Known and unknown biodiversity, risk of extinction and conservation strategy in the sea, in: Bendell-Young, L., Gallaugher, P. (Eds.), Waters in Peril. Springer US, Boston, MA, pp. 19–33. https://doi.org/10.1007/978-1-4615-1493-0_2

Robeson, M. S., O'Rourke, D. R., Kaehler, B. D., Ziemski, M., Dillon, M. R., Foster, J. T., & Bokulich, N. A. (2021). RESCRIPt: Reproducible sequence taxonomy reference database management. *PLOS Computational Biology*, 17, e1009581. https://doi.org/10.1371/journal.pcbi.1009581

Rouse, G.W., Goffredi, S.K., & Vrijenhoek, R.C. (2004). Osedax: bone-eating marine worms with dwarf males. *Science*, 305, 668–671. https://doi.org/10.1126/science.1098650

RStudio Team. (2022). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Salzburger, W. (2018). Understanding explosive diversification through cichlid fish genomics. *Nature Reviews Genetics*, 19(11), 705–717. https://doi.org/10.1038/s41576-018-0043-9

Santi, I., Kasapidis, P., Karakassis, I., & Pitta, P. (2021). A Comparison of DNA Metabarcoding and Microscopy Methodologies for the Study of Aquatic Microbial Eukaryotes. *Diversity* 13, 180. https://doi.org/10.3390/d13050180

Scheffers, B.R., Joppa, L.N., Pimm, S.L., & Laurance, W.F. (2012). What we know and don't know about Earth's missing biodiversity. *Trends* in *Ecology & Evolution*, 27, 501–510. https://doi.org/10.1016/j.tree.2012.05.008

Schloissnig, S., Kawaguchi, A., Nowoshilow, S., Falcon, F., Otsuki, L., Tardivo, P., ... Tanaka, E. M. (2021). The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proceedings of the National Academy of Sciences USA*, 118, e2017176118. https://doi.org/10.1073/pnas.2017176118

Schneider, C., Woehle, C., Greve, C., D'Haese, C. A., Wolf, M., Hiller, M., ... Huettel, B. (2021). Two high-quality de novo genomes from single ethanol-preserved specimens of tiny metazoans (collembola). *GigaScience*, 10. https://doi.org/10.1093/gigascience/giab035

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., ... Karsch-Mizrachi, I. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database*, 2020, baaa062. https://doi.org/10.1093/database/baaa062

Stoev, P., Komerički, A., Akkari, N., Liu, S., Zhou, X., Weigand, A., ... Penev, L. (2013). Eupolybothrus cavernicolus Komerički & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae): the first eukaryotic species description combining transcriptomic, DNA barcoding and micro-CT imaging data. *Biodiversity Data Journal*, 1, e1013. https://doi.org/10.3897/BDJ.1.e1013

Seoane, S., Eikrem, W, Pienaar, R., Edvardsen, B. (2009). *Chrysochromulina palpebralis* sp. nov. (Prymnesiophyceae): a haptophyte, possessing two alternative morphologies. *Phycologia*, 48, 165-176. https://doi.org/10.2216/08-63.1

Sheffield NC, Bonazzi VR, Bourne PE, Burdett T, Clark T, Grossman RL, Spjuth O, Yates AD. (2022) From biomedical cloud platforms to microservices: next steps in FAIR data and analysis. Sci Data 9, 553. https://doi.org/10.1038/s41597-022-01619-5

Shinn, A. P., Avenant-Oldewage, A., Bondad-Reantaso, M. G., Cruz-Laufer, A., García Vásquez, A., Hernández-Orts, J. S., … Deveney, M. R. (2023). A global review of problematic and pathogenic parasites of farmed tilapia. *Reviews in Aquaculture*. https://doi.org/10.1111/raq.12742

Sinniger, F., Pawlowski, J., Harii, S., Gooday, A.J., Yamamoto, H., Chevaldonné, ... Creer, S. (2016). Worldwide Analysis of Sedimentary DNA Reveals Major Gaps in Taxonomic Knowledge of Deep-Sea Benthos. *Frontiers in Marine Science*, 3, 1–14. https://doi.org/10.3389/fmars.2016.00092

Smith, L.T., Magdalena, C., Przelomska, N.A.S., Pérez-Escobar, O.A., Melgar-Gómez, D.G., Beck, S., ... Monro, A.K. (2022). Revised species delimitation in the giant water lily genus victoria (nymphaeaceae) confirms a new species and has implications for its conservation. *Frontiers in Plant Science*, 13, 883151. https://doi.org/10.3389/fpls.2022.883151

Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W. T., Kutty, S. N., Kurina, O., & Meier, R. (2019). Rapid, large-scale species discovery in hyperdiverse taxa using 1d MinION sequencing. *BMC Biology*, 17, 96. https://doi.org/10.1186/s12915-019-0706-9

Struck, T. H., & Cerca, J. 2019. Cryptic species and their evolutionary significance*Els* pp. 1-9).

Struck, T. H., & Cerca, J. 2022. Connecting micro-and macro-evolutionary research: Extant cryptic species as systems to understand macro-evolutionary stasis. In A. K. Monro & S. J. Mayo (Eds) *Cryptic species morphological stasis, circumscription, and hidden diversity* pp. 143-188). Cambridge: Cambridge University Press.

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., ... Dimitrov, D. (2018a). Cryptic species – more than terminological chaos: A reply to Heethoff. *Trends in Ecology & Evolution*, 33, 310-312. https://doi.org/10.1016/j.tree.2018.02.008

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., ... Dimitrov, D. (2018b). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, 33, 153-163. https://doi.org/10.1016/j.tree.2017.11.007.

Stuart, S.N., Wilson, E.O., McNeely, J.A., Mittermeier, R.A., & Rodríguez, J.P. (2010). Ecology. The barometer of life. *Science,* 328, 177. https://doi.org/10.1126/science.1188606

Šupraha, L., Klemm, K., Gran Stadniczenko, S., Vaulot, D., Edvardsen, B., & John, U. (2022). Diversity and biogeography of planktonic diatoms in Svalbard fjords: the role of dispersal and Arctic endemism in phytoplankton community structuring. *Elementa:* Science *of the Anthropocene,* 10, 00117.
https://doi.org/10.1525/elementa.2021.00117

Tedesco, P.A., Bigorne, R., Bogan, A.E., Giam, X., Jézéquel, C., & Hugueny, B. (2014). Estimating how many undescribed species have gone extinct. *Conservation Biology*, 28, 1360–1370. https://doi.org/10.1111/cobi.12285

Teschke, K., Kraan, C., Kloss, P., Andresen, H., Beermann, J., Fiorentino, D., ... Dannheim, J. (2022). CRITTERBASE, a science-driven data warehouse for marine biota. *Scientific Data*, 9, 483, https://doi.org/10.1038/s41597-022-01590-1

Thompson, C. W., Phelps, K. L., Allard, M. W., Cook, J. A., Dunnum, J. L., Ferguson, A. W., … Kilpatrick, C. W. (2021). Preserve a voucher specimen! The critical need for integrating natural history collections in infectious disease studies. *mBio*, 12, e02698-20.
https://doi.org/10.1128/mBio.02698-20

Valan, M., Vondráček, D., & Ronquist, F. (2021). Awakening a taxonomist's third eye: exploring the utility of computer vision and deep learning in insect systematics. *Systematic Entomology*, 46, 757–766. https://doi.org/10.1111/syen.12492

van den Burg, M. P., Herrando-Pérez, S., & Vieites, D. R. (2020). ACDC, a global database of amphibian cytochrome-b sequences using reproducible curation for GenBank records. *Scientific Data*, 7, 268. https://doi.org/10.1038/s41597-020-00598-9

Vanhove, M. P. M., Pariselle, A., Van Steenberge, M., Raeymaekers, J. A. M., Hablützel, P. I., Gillardin, C., … Huyse, T. (2015). Hidden biodiversity in an ancient lake: phylogenetic congruence between Lake Tanganyika tropheine cichlids and their monogenean flatworm parasites. *Scientific Reports*, 5, 13669. https://doi.org/10.1038/srep13669

Varney, R., Yap-Chiongco, M., Mikkelsen, N., & Kocot, K. (2022). Genome of the lepidopleurid chiton hanleya hanleyi (Mollusca, Polyplacophora). *F1000Research*, 11, 555. https://doi.org/10.12688/f1000research.121706.1

Vaulot, D., Sim, C. W. H., Ong, D., Teo, B., Biwer, C., Jamy, M., & Lopes dos Santos, A. (2022). metaPR^2: A database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists. *Molecular Ecology Research*, 22, 3188– 3201.

https://doi.org/10.1111/1755-0998.13674

Vihtakari, M. (2022). *ggOceanMaps: Plot Data on Oceanographic Maps using 'ggplot2'*. R package version 1.3.4. https://mikkovihtakari.github.io/ggOceanMaps/.

Wägele, H., Klussmann-Kolb, A., Kuhlmann, M., Haszprunar, G., Lindberg, D., Koch, A., & Wägele, J.W. (2011). The taxonomist - an endangered race. A practical proposal for its survival. *Frontiers in Zoology*, 8, 25. https://doi.org/10.1186/1742-9994-8-25

Wassmann, P., Duarte, C. M., Agustí, S., & Sejr, M. K. (2011). Footprints of climate change in the Arctic marine ecosystem. *Global Change Biology*, 17, 1235–1249, https://doi.org/10.1111/j.1365-2486.2010.02311.x.

Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., … Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of The Total Environment*, 678, 499-524. https://doi.org/10.1016/j.scitotenv.2019.04.247

Wheeler, Q.D., Raven, P.H., & Wilson, E.O. (2004). Taxonomy: impediment or expedient? *Science*, 303, 285. https://doi.org/10.1126/science.303.5656.285

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., Ooms, J., & Müller, K. (2022). RPostgres: Rcpp Interface to PostgreSQL. R package version 1.4.4, https://CRAN.R-project.org/package=RPostgres

Wilson, E.O. (2004). Taxonomy as a fundamental discipline. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 739. https://doi.org/10.1098/rstb.2003.1440

Windsor, D. A. (1998). Most of the species on Earth are parasites. *International Journal for Parasitology*, 28(12), 1939–1941. https://doi.org/10.1016/S0020-7519(98)00153-2

Windsor, D. A. (2021). Conservation of parasites. *International Journal for Parasitology: Parasites and Wildlife*, 14, 137. https://doi.org/10.1016/j.ijppaw.2021.01.011

Wood, C. L., & Vanhove, M. P. M. (2022). Is the world wormier than it used to be? We'll never know without natural history collections. *Journal of Animal Ecology* https://doi.org/10.1111/1365-2656.13794

Wood, C. L., Leslie, K. L., Claar, D., Mastick, N., Preisser, W., Vanhove, M. P. M., & Welicky, R. (2023). How to use natural history collections to resurrect information on historical parasite abundances. *Journal of Helminthology*, 97, e6. https://doi.org/10.1017/S0022149X2200075X

Yang, B., Zhang, Z., Yang, C.-Q., Wang, Y., Orr, M.C., Wang, H., & Zhang, A.-B. (2022) Identification of species by combining molecular and morphological data using convolutional neural networks. *Systematic Biology*, 71, 690–705. https://doi.org/10.1093/sysbio/syab076

Figure Legends

Figure 1
Modular structure of the CRITTERBASE data warehouse.

Figure 2
Scatterplot showing abundances (individuals per m^{-2}) of six polychaete species in the southern North Sea in relation to the median grain size of surficial bottom sediments (μm). (a) *Chaetozone setosa*, (b) *Lagis koreni*, (c) *Lanice conchilega*, (d) *Magelona johnstoni*, (e) *Scoloplos armiger*, (f) *Spiophanes bombyx*.

Figure 3
Locations and number (n) of grab samples considered in the analysis of the distribution of six polychaete species in relation to sediment type in the southern North Sea (Figure 2). Produced using R package ggOceanMaps, the bathymetric data contained therein (Amante & Eakins, 2009; GEBCO, 2021; EMODnet Bathymetry Consortium, 2020) and the continent boundary vectors (Natural Earth, 2022).

Figure 4
Representation gaps for COI at the phylum level. Percent of species that have a COI sequence in the reference database MIDORI2 based on minimum (light blue) or maximum (orange) estimates of species richness (when a range of species richness estimates was available). Data on species richness was extracted from Table 2 in Machida et al. (2017). We revised the two phyla reporting 1 species and updated them to 2 for Cycliophora (based on the species reported in the World Register of Marine Species, https://www.marinespecies.org/aphia.php?p=taxdetails&id=22586 on 2022-10-31) and 3 for Placozoa (Osigus et al., 2019). Counts on unique species per phylum for the COI genetic marker were calculated using data from MIDORI2 version GB249 downloaded on May 13, 2022 from the MIDORI server (http://www.reference-midori.info/download.php#).

Table 1: Examples of genome project in the animal kingdom or geographically

Project name	Main focus	Webpage
Vertebrate Genome Project (VGP)	Vertebrates	https://vertebrategenomesproject.org/
10,000 Bird Genomes (B10K)	Birds	https://b10k.genomics.cn/
Fish 10,000 Genomes (Fish 10K)	Fish	https://english.cas.cn/newsroom/research_news/life/201910/t20191008_219736.shtml
Squalomix	Sharks and rays	https://github.com/Squalomix/info/
5,000 Insect Genomes (I5K)	Insects	http://i5k.github.io/
Global Ant Genomics Alliance (GAGA)	Ants	https://db.cngb.org/antbase/project
Beenome100	Bees	https://www.beenome100.org/
Global Invertebrate Genomics Alliance (GIGA)	Invertebrates without insects	http://www.gigacos.org/
InvertOmics	Lophotrochozoa	https://www.frontiersinevolutionaryzoology.com/research-1
Soil Invertebrate Genome Initiative (SIGI)	Soil invertebrates	https://tbg.senckenberg.de/sigi/
Aquatic Symbiosis Genomics Project	Freshwater and marine species	https://www.sanger.ac.uk/collaboration/aquatic-symbiosis-genomics-project/
Deep-Ocean Genomes Program	Deep Sea species	https://www.oceandecade.org/actions/deep-ocean-genomes-program/
Earth Biogenome Project (EBP)	Global	https://www.earthbiogenome.org/
Africa BioGenome Project (AfricaBP)	Africa	https://africanbiogenome.org/
Australian amphibian and reptile genomics (AusARG)	Australia – Reptiles & Amphibians	https://researchprofiles.canberra.edu.au/en/projects/ausarg-australian-amphibian-and-reptile-genomics-initiative-colla
Oz Mammals Genomics initiative	Australia - Mammals	https://ozmammalsgenomics.com/
Canada 150 Sequencing Initiative (CanSeq150)	Canada	http://www.cgen.ca/canseq150
California Conservation	California (USA)	https://www.ccgproject.org/

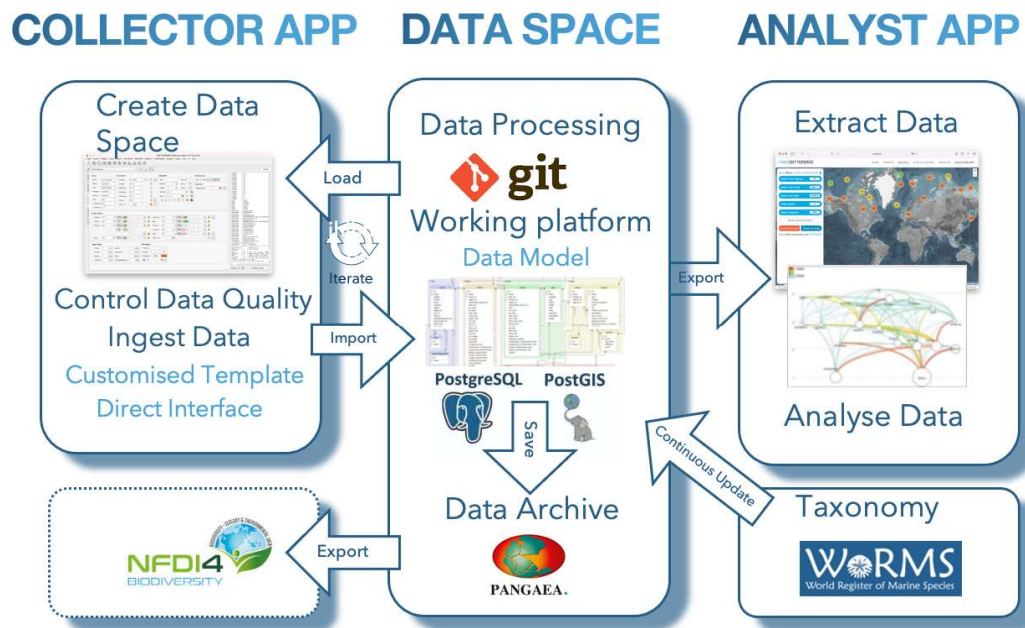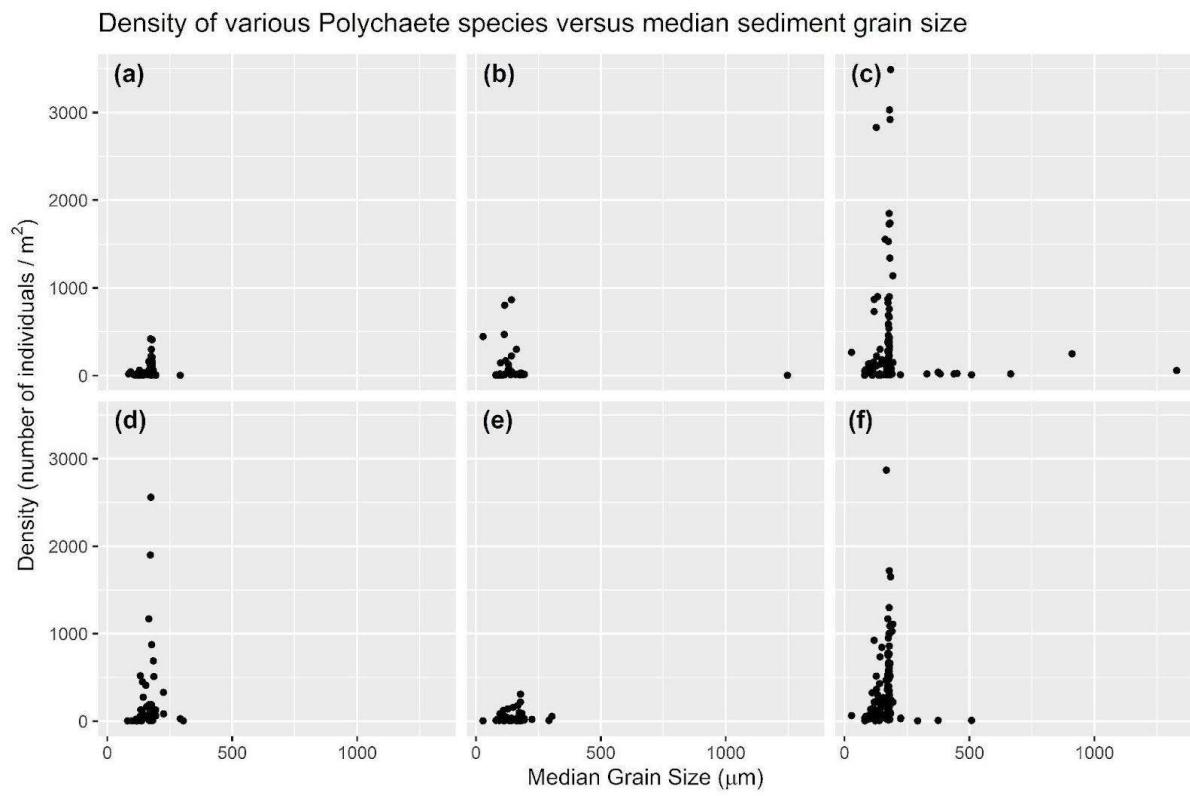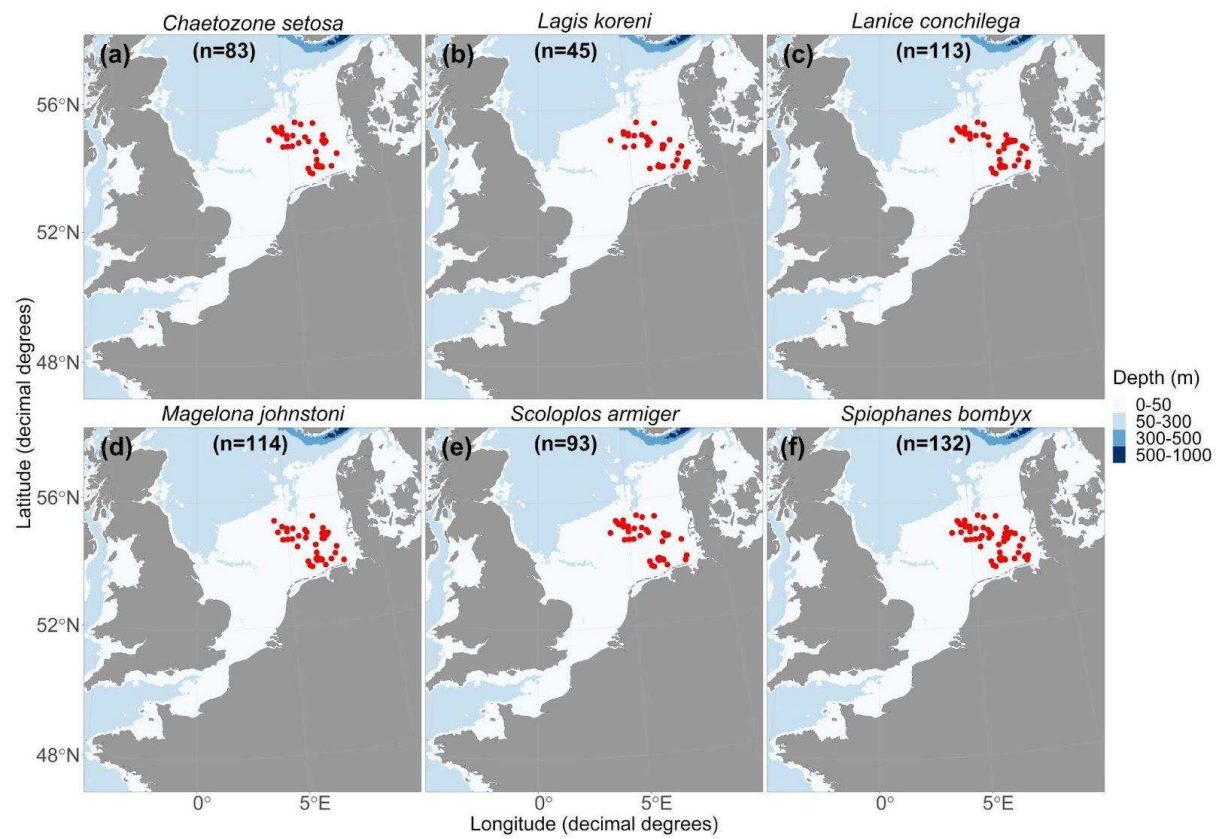|		
Genomics Project (CCGP)		
Illinois EBP Pilot	Illinois (USA)	https://blogs.illinois.edu/view/7447/790701529
European Reference Genome Atlas (ERGA)	Europe	https://www.erga-biodiversity.eu/
Biodiversity Genomics Europe (BGE)	Europe	https://biodiversitygenomics.eu/
Darwin Tree of Life (DToL)	UK & Ireland	https://www.darwintreeoflife.org/
EBP-Norway (EBP-Nor)	Norway	https://www.ebpnor.org/english/
Endemixit	Italy – Endemic species	https://endemixit.com/the-project/
Catalan Initiative for the Earth BioGenome Project (CBP)	Catalonia (Spania)	https://www.biogenoma.cat/en/home/

Figure 1

Figure 2

Density of various Polychaete species versus median sediment grain size

Figure 3

Figure 4