

## Synthesis

# Taxonomy 2.0: computer-aided identification tools to assist Antarctic biologists in the field and in the laboratory

THOMAS SAUCÈDE <sup>1</sup>, MARC ELÉAUME<sup>2</sup>, QUENTIN JOSSART<sup>3</sup>, CAMILLE MOREAU<sup>1,3</sup>,  
RACHEL DOWNEY<sup>4</sup>, NARISSA BAX<sup>5</sup>, CHESTER SANDS<sup>6</sup>, BORJA MERCADO<sup>7</sup>, CYRIL GALLUT<sup>2</sup> and  
RÉGINE VIGNES-LEBBE<sup>2</sup>

<sup>1</sup>Biogéosciences, UMR 6282, CNRS, EPHE, Université Bourgogne Franche-Comté, 6 boulevard Gabriel, 21000 Dijon, France

<sup>2</sup>Institut de Systématique, Evolution, Biodiversité (ISYEB UMR 7205), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier, 75005 Paris, France

<sup>3</sup>Marine Biology Lab, CP160/15 Université Libre de Bruxelles (ULB) 50, Av. F.D. Roosevelt, B-1050 Brussels, Belgium

<sup>4</sup>Fenner School of Environment & Society, Australian National University, Linnaeus Way, Acton 2601, Canberra, ACT, Australia

<sup>5</sup>Institute for Marine and Antarctic Studies, University of Tasmania & Centre for Marine Socioecology, Hobart, TAS, Australia

<sup>6</sup>British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK

<sup>7</sup>Biodiversity and Cnidarians Evolution Group, Cavanilles Institute of Biodiversity and Evolutionary Biology (ICBiBE), University of Valencia, 22085, 46071 Valencia, Spain

[thomas.saucede@u-bourgogne.fr](mailto:thomas.saucede@u-bourgogne.fr)

**Abstract:** Species inventories are essential to the implementation of conservation policies to mitigate biodiversity loss and maintain ecosystem services and their value to the society. This is particularly topical with respect to climate change and direct anthropogenic effects on Antarctic biodiversity, with the identification of the most at-risk taxa and geographical areas becoming a priority. Identification tools are often neglected and considered helpful only for taxonomists. However, the development of new online information technologies and computer-aided identification tools provides an opportunity to promote them to a wider audience, especially considering the emerging generation of scientists who apply an integrative approach to taxonomy. This paper aims to clarify essential concepts and provide convenient and accessible tools, tips and suggested systems to use and develop knowledge bases (KBs). The software *Xper3* was selected as an example of a user-friendly KB management system to give a general overview of existing tools and functionalities through two applications: the 'Antarctic Echinoids' and 'Odontasteridae Southern Ocean (Asteroids)' KBs. We highlight the advantages provided by KBs over more classical tools, and future potential uses are highlighted, including the production of field guides to aid in the compilation of species inventories for biodiversity conservation purposes.

Received 29 April 2020, accepted 29 July 2020

**Key words:** database, field guides, knowledge base, software, taxonomic key, *Xper3*

## Introduction

Species inventories are essential to the implementation of conservation policies to mitigate biodiversity loss and maintain ecosystem services and their value to society (Balmford & Gaston 1999, Gaston 2005, May 2011), to the prevention of species declines or invasions (Chapman 2005) and to forming the basis of management planning regimes (e.g. species recovery plans). This growing importance of conservation issues has increased the need for accurate identification of endangered, migratory and invasive species (Balmford & Gaston 1999).

Considering the rapid pace of climate and oceanographic changes already affecting (Turner *et al.* 2014) or predicted to affect (Allan *et al.* 2013, Gutt *et al.* 2015) the Southern

Ocean and its marine life, the identification of the most at-risk taxa and geographical areas has become a priority (Griffiths 2010) for the definition of conservation measures such as Marine Protected Areas and Vulnerable Marine Environments (CCAMLR 2009, <https://www.ccamlr.org/en/ccamlr-sm-ii/03>). Taxon inventories form an essential basis for studying the multiple abiotic and biotic stressors on species and ecosystems (Pendlebury & Barnes-Keoghan 2007, Kargel *et al.* 2014, Molinos *et al.* 2015, Byrne *et al.* 2016, Schram *et al.* 2016).

In recognition of the need for an improved understanding of species and community responses to environmental change, many key questions in the Scientific Committee for Antarctic Research (SCAR) 'Antarctic Life on the Precipice' project cover topics relating to the inclusion

of biodiversity research for the 2020 Science Horizon Scan (Kennicutt *et al.* 2015). Coordinating biodiversity collections and increasing access to taxonomic and genetic identification resources are key strategic aims.

Over the last decade, many efforts have been made to improve our knowledge of Antarctic marine life, including the use of new sampling tools for data collection (i.e. sampling gear and remote sensing and autonomous vehicles) and the development of analytical techniques (e.g. -omics approaches and isotope analyses) (Kaiser *et al.* 2013, Gutt *et al.* 2017). Following the synergistic momentum of the Census of Antarctic Marine Life (CAML) and of the International Polar Year (IPY), the inventory of Antarctic marine species has been significantly enlarged (Schiaparelli *et al.* 2013), highlighting both the uniqueness of the Antarctic biota and its high level of diversity (Clarke & Johnston 2003, Brandt *et al.* 2007, Cary *et al.* 2010, De Broyer *et al.* 2014, Kennicutt *et al.* 2015). Newly collected data are now compiled in information networks and initiatives such as the Barcode of Life Data System (BOLD, <http://www.boldsystems.org>), the Register of Antarctic Marine Species (RAMS, <http://www.marinespecies.org/rams>, Jossart *et al.* 2015) and the SCAR-Marine Biodiversity Information Network (SCAR-MarBIN, <https://www.scarmarbin.be>, Griffiths *et al.* 2011), which is now part of the SCAR Antarctic Biodiversity Portal (<https://www.biodiversity.aq>). They were also disseminated through the publication of the *Biogeographic Atlas of the Southern Ocean* (De Broyer *et al.* 2014) and a number of other open-access databases (DBs) (Danis *et al.* 2013, Gutt *et al.* 2013, <http://www.biodiversity.aq>). Altogether, these efforts have greatly advanced the documentation of Antarctic marine life (De Broyer & Danis 2011, Gutt *et al.* 2011, De Broyer *et al.* 2014). However, many biodiversity collections in museums and scientific institutes remain unsorted, unidentified and, thus far, unsearchable across these platforms, often due to a deficit of funding towards taxonomic research costs.

Naming species and using taxonomic and biodiversity information is not only a concern of taxonomists. It is of importance for many scientists and amateurs in life sciences, as it underpins any biodiversity study (Chapman 2005). It is central to countless numbers of publications and various fields such as phylogenetics, wildlife ecology, biogeography, conservation, natural resource management, biomedical research and education (Chapman 2005, Costello *et al.* 2015). These various user groups require the development of flexible, searchable, easily accessible and usable DBs following the FAIR (Findable, Accessible, Interoperable and Re-usable) guiding principles for scientific data management (Wilkinson *et al.* 2016). These DBs also need to be easily amendable as systematics and taxonomy are regularly being revised and updated, at least for a number of invertebrate taxa (Griffiths 2010).

In the scientific literature, species descriptions can be incomplete, patchy and heterogeneous in their methodology, and old-fashioned drawings or images are sometimes confusing and misleading. The interpretation of this literature and the observation of collection specimens by experts are therefore needed to build structured and comparable taxonomic descriptions. By using one or more DB, a knowledge base (KB) is an application that stores and manages the expert's knowledge for describing, classifying and identifying organisms. In KBs, the expert's knowledge is stored as formal, structured and complex information sources used by computer systems, but the information can be easily used by various categories of end users, from expert taxonomists to scientists with various fields of expertise or citizens interested in observing nature. The information summarized in KBs can be used to perform new inventories and monitoring surveys, phylogenetic analyses, as well as biogeographical and ecological studies for management and conservation purposes. Over the last 20 years, the development of new information technologies has represented an opportunity to develop such tools and to generate identification guides and species checklists that can be used in distribution information systems (Chavan & Penev 2011).

Decision criteria (i.e. diagnostic characters) are required in order to attribute a taxon name to a specimen collected in the field or accessed in academic collections. Decision criteria can be based on a wide variety of information sources and types, including morphological and ecological (e.g. sound, behaviour, etc.) features when identifying specimens in the field. Other features, such as molecular data, are of importance for post-field identifications (Teletchea 2010). It is important to clarify that formal taxonomic identification differs greatly from identification performed by non-expert taxonomists, either academic or citizen scientists with recognized expertise or people with an interest in observing nature. Unfortunately, it is often difficult to distinguish between the two in publications unless clear taxonomic methodologies, involving relevant specialist literature citations, are used. Identifications based on field guides are useful starting points, but they should not be used in ecological analyses unless verified by a taxonomic expert.

Diagnostic identification keys are the primary form of taxonomic identification tool and outline a set of characters applied in a hierarchical system for species determination. However, non-discrete and non-binary approaches also exist, such as automated identification tools. This computer-aided identification facilitates biodiversity expertise. Identification keys can be produced manually, but algorithms and programs can compute keys and have existed since the 1970s. For example, the DELTA format (DEscription Language for TAXonomy) or SDD (Structured Descriptive Data)

are two standards used for data exchange by the Biodiversity Information Standards community (formerly the Taxonomic Databases Working Group, TDWG). Several programs were based on these formats, including software producing identification keys such as *Intkey*, *Lucid*, *Xper2* or *Xper3* (Dallwitz 1980, 2000).

A workshop was organized for July 2017 in Leuven (Belgium) alongside the XIIth SCAR Biology Symposium. It was open to all scientists, either occasional or confirmed taxonomists, who have needed to identify organisms or describe species from specimens collected in the field or accessed in collections. The objectives of the workshop were to present useful tools and software that can assist biologists in a wide range of applications to build and/or use KBs. These included species identification, the formalization of diagnostic morphological characters and the production of DBs, catalogues, field guides and monographs.

The main outcomes of the workshop are synthesized in the present paper. The aim is not to provide a comprehensive survey of all existing tools and possibilities, but to clarify essential concepts and provide Antarctic biologists with convenient and accessible tools, tips and suggested systems for them to follow-up regarding their development on their own KB. The software *Xper3* (Vignes-Lebbe *et al.* 2016) was chosen as an example of a user-friendly KB management system to illustrate these concepts and tools. A general overview of this computer-aided identification software is given, along with existing tools and functionalities. Some applications and case studies are presented for the production and use of KBs, and future prospects are given, including the production of field guides.

## Computer-aided identification and decision criteria

### Identification keys

Recent advances in bioinformatics have led to the development of a large number of integrated electronic resources comprising online versions of DBs and KBs with interactive identification systems and digital imaging (Dallwitz 2000, Chapman 2005, Nimis & Vignes-Lebbe 2010). The key advantages of this approach over conventional systems include no restriction in the order in which characters can be used, character deletion and changes, error tolerance and localizing and uncertainty expressing, among many other functionalities (Dallwitz 2000). Comprehensive comparisons of the main interactive identification programs were made available online by Dallwitz (2000) and Thiele and Dallwitz (2000) ([https://www.delta-intkey.com/www/comparison.htm#Programs\\_compared](https://www.delta-intkey.com/www/comparison.htm#Programs_compared)).

Taxon identification keys are generally of two kinds (Hagedorn *et al.* 2010): 1) conventional, single-access

keys (polytomous or dichotomous) have been classically used in taxonomic books, field guides, treatises and monographs with one single possible access and decision criteria that must be followed (such keys are not very flexible and cannot be updated easily), and 2) multiple-access keys, which use any character within a set of possible descriptors and can be easily illustrated, updated and revised (however, such keys require the development of software-aided tools).

Among many different DB projects, FishBase (<http://www.fishbase.org/keys/allkeys.php>) is an example of a global biodiversity information system that proposes various finfish identification tools on the web (<http://www.fishbase.org>). Keys can be selected depending on the geographical area (following the Food and Agriculture Organization of the United Nations (FAO) zonation) and taxon (order or family) of interest. Identification keys are also available to identify species using morphometric measurements. Finally, identification by picture is also possible for selected countries, families and ecosystems. There is a wide variety of identification keys available in FishBase, but most of them are single-access polytomous systems. Once selected, they only allow one single decision direction at a time, which must be followed.

Multiple-access identification keys, using software-aided tools, can be found on the web for the identification of various taxa. Many of them are devoted to the terrestrial realm. For instance, *NaviKey* ([www.navikey.net](http://www.navikey.net), Neubacher & Rambold 2005) is implemented as a Java-embedded application for accessing descriptive data in a DELTA format, with 27 DBs available for the identification of taxa in botany, mycology and zoology. The north-west African grasshoppers and locusts website (<http://acrinwafrica.mnhn.fr/SiteAcricri/Xper.html>) proposes a set of embedded multiple-access identification keys using the software *Xper3*. It allows for the selection of any character within a set of morphological descriptors for the identification of Orthoptera (<http://acrinwafrica.mnhn.fr>). Species identification is achieved in several steps by gradual elimination of species names that do not meet the selection criteria at each step of the process. First, a morphological descriptor and associated character state are selected in no particular order, and the process is repeated until the name of a species is obtained according to the selection criteria. The e-Monograph of the Caricaceae website (<http://herbaria.plants.ox.ac.uk/bol/caricaceae/Keys>, Carvalho *et al.* 2014) or the interactive key on Hymenoptera (<http://pteromalus.identificationkey.fr/mkey.html>, Klimmek & Baur 2018) are other examples of multiple-access keys using the *Xper3* software.

Multiple-access identification keys of marine species can be found online (<http://www.marinespecies.org/aphia.php?p=idkeys>) as taxonomic facilities associated with the World Register of Marine Species (WoRMS,

<http://www.marinespecies.org>). These keys can be generated for any taxonomic group at any chosen level, and they allow the user to start identifying an organism by selecting any obvious character, then set the remaining characters relevant to the identification, thereby narrowing down to the final identification. The Mascarene corals website is another example of a marine species identification key providing computer-based assistance for the description, classification and identification of scleractinian corals from the south-west Indian Ocean. Using a collection of ~3000 specimens belonging to 185 species and 58 genera, two software options were proposed with various functionalities: *Xper2* (Ung *et al.* 2010, see below) and *IKBS* (Iterative Knowledge Base System, Conruyt *et al.* 1998), distributed via a CD and a booklet ([http://coraux.univ-reunion.fr/IMG/pdf/Plaquette-Coraux\\_des\\_Mascareignes-mai08.pdf](http://coraux.univ-reunion.fr/IMG/pdf/Plaquette-Coraux_des_Mascareignes-mai08.pdf)).

#### *Automated identification*

Recent advances in biodiversity identification tools have also led to the development of automated identification. Such tools allow experts to compile huge sets of data, which have led to significant improvements in our understanding of biodiversity. Automated identification tools are based on the computerized recognition of clusters, ordination patterns or artificial neural networks (Kumar *et al.* 2012, Hu *et al.* 2018, Joly *et al.* 2018).

Pl@ntNet (<https://plantnet.org/en>) is one of the most used systems proposing the automatic identification of plants through images. It is a citizen science project available as a smartphone application or as a web version. 'By comparing visual patterns transmitted by users via photos of plant organs', the Pl@ntNet system seeks to identify species. The images are analysed and compared to an image bank produced collaboratively and enriched daily. The system then offers a possible list of species with illustrations (Joly *et al.* 2014). iNaturalist (<https://www.inaturalist.org>) also offers a web platform for citizen science projects and collaborative identification.

Automated identification tools were also developed for marine organisms based on various descriptor types. Some are based on morphological descriptors, such as a tool implemented for shark identification based on their fin shape (<https://arxiv.org/abs/1609.06323>). Dedicated algorithms were also developed for plankton recognition and classification of images using deep machine learning methods such as deep convolutional neural networks (Kuang 2015, Bueno *et al.* 2017). However, automated identification tools are not restricted to morphological descriptors. Plankton scanning methods were also developed that rely on molecular screening tools using -omics approaches: proteomics, metabolomics, genomics and metagenomics. These methods were accompanied by important scientific outputs and received a lot of

media coverage. Several research programmes were undertaken on zooplankton automated identification (Gorsky *et al.* 2010, [http://www.obs-vlfr.fr/~gaspari/Plankton\\_Identifier/index.php](http://www.obs-vlfr.fr/~gaspari/Plankton_Identifier/index.php)). The genomic monitoring of the plankton during expeditions of the RV *Tara* in 2009 and 2013 (Pesant *et al.* 2015) is one of the most famous examples of the performance of such tools.

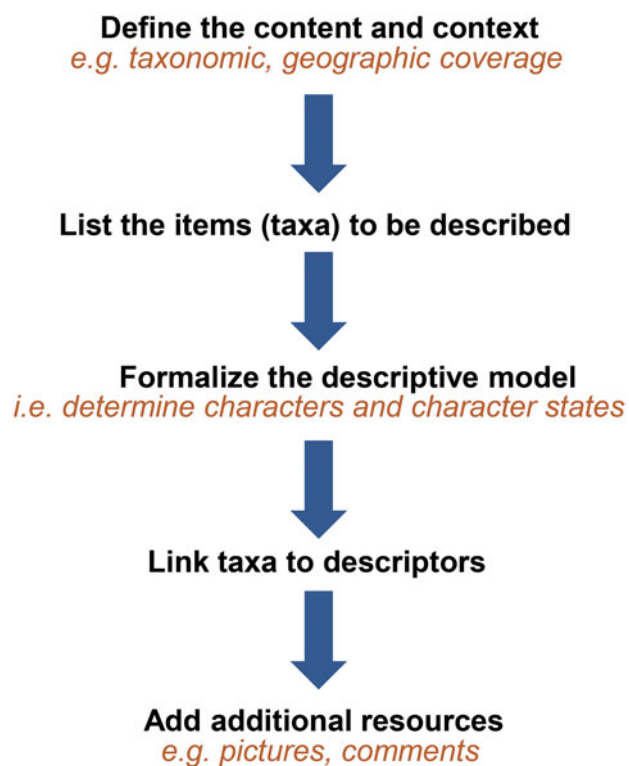
#### **Knowledge bases**

##### *The need for building knowledge bases*

Taxonomic expertise is a rare commodity that is usually not well valued and passed on from generation to generation. KBs allow for the storage, representation and preservation of the knowledge of recognized specialists in a specific field of interest as formal, structured and complex information sources. They allow for the management, manipulation, comparison, processing and transmission of knowledge. Based on a descriptive model to standardize the terminology and a DB to store the data such as comparable taxonomic descriptions, KBs can use machine learning techniques and generate additional knowledge, such as diagnoses (by computing best discriminant descriptors) and decision trees for identification.

Multiple-access keys using computers as proposed by *Xper3* (Vignes-Lebbe *et al.* 2016), *NaviKey* (Neubacher & Rambold 2005), *Lucid* (<https://www.lucidcentral.org>) and other programs (Dallwitz 2000) require an algorithm to automatically analyse and compare taxonomic descriptions. New descriptions produced by taxon specialists result in new observations and measurements that can then be easily added to previous descriptions. In practical terms, a KB can be stored in DB management systems (*Xper3* used MySQL) or XML files (data exchange standards such as SDD) or text files (DELTA files).

Building and using KBs can provide many uses for addressing taxonomic issues compared to more classical tools such as books, as exemplified below, but this also applies to traditional electronic resources. For instance, identification books occasionally need to be updated because taxonomic descriptions change with taxonomy, new collections and our knowledge of species biogeography. Revising, editing, producing and publishing new book versions is time- and energy-consuming, while updates can be regularly and readily done in KBs. Some identification books may not be 'user friendly', leading to species misidentifications. Using KBs, complicated identification keys can be substantially and regularly improved based on users' feedback. Furthermore, software such as *Xper3* can explain identification results and display the discordant character states of eliminated taxa. The spreading of taxonomic information in the formal scientific literature can be patchy and data can be



**Fig. 1.** Processing flowchart for building a knowledge base.

organized either geographically or taxonomically, which are serious limitations to species identification, especially as clear identification keys are still needed for most taxa, and especially those from the deep sea and polar regions. Scattered taxonomic data can be easily merged and combined in KBs. Lastly, there have been paradigm shifts in the ways in which we undertake taxonomic research. These are related to recent advances in new technologies and DNA sequencing techniques that make us question traditional morphology-based taxonomy and look for better ways by which to identify specimens. This process is part of the general shift in taxonomy to a data-intensive, expert-led, logically constructed, unambiguous taxonomic identification process, meaning that non-experts can be involved in the global categorizing of the living world. Near-global access to computers, the Internet and collaborative scientific platforms enables taxonomists and data scientists to easily and regularly update KBs.

#### *Building a knowledge base*

Building a KB includes several steps that are synthesized as a flow chart in Fig. 1.

1) The metadata. The first step consists in defining the metadata; that is, the content, limits and usages, which means the taxonomic coverage, geographical area of interest, phenological states considered by the

key, the public (end users) and the needed prerequisites (basic knowledge).

- 2) Items. The objects to be described must be listed as items. Most often these items are taxa, but sometimes they can correspond to groups (e.g. for the crowdsourcing project on pollinators, the *Spipoll* key includes morphological groups when it is too difficult for identification at the species level; see <http://spipoll.snv.jussieu.fr/mkey/mkey-spipoll.html>).
- 3) The descriptive model. Character description must be formalized along with the corresponding terminology (descriptors/characters) using a glossary to describe the taxa. This means that the characters, the character states in case of categorical characters and the unit for numerical characters must be defined. Producing descriptive models and rules is a crucial point in KB development in order to obtain comparable and intelligible descriptions and keys that are easy to use.
- 4) Character dependency and weights. Consistency rules must be edited, which are the experts' knowledge of character relevance. Characters can be weighted to take into account observational difficulties or to influence the key algorithm by suggesting an order in the identification process. In this way, the KB is not limited to a simple matrix and translates more of the taxonomic expertise.
- 5) Descriptions. This step corresponds to the item descriptions linking taxa to descriptors, including polymorphism. A single non-hierarchical identification system can be generated for all species of a genus or family using a restricted number of characters, most of which are declined into many character states. However, generating an identification key for a higher taxonomic level (i.e. order, class) is more ambitious considering the incompatibility (specificity) of some characters. In such a case, a hierarchical structure of identification keys is possible by using internal URL links to connect identification keys from various taxonomic levels.
- 6) Additional resources. Additional resources such as comments or images can be provided. Both overview pictures of specimens and close-ups of some characters can be added to help the user in the identification process or to illustrate the species with living specimens exhibiting vivid colours taken in the field.

#### *Xper3 software as an application tool*

*Xper3* is a free and portable KB software that provides an easy-to-use interface. It allows us to check the consistency of existing descriptions, compare descriptions or sets of descriptions, create and use either single- or multiple-access keys and publish identification keys and taxonomic forms. There have been three different

**Table I.** Comparison between *Xper2* and *Xper3*.

Function	<i>Xper2</i>	<i>Xper3</i>
Installation	Local. Windows, macOS, Linux	Not required. Remote access via Internet browser
User	Single user	Multiple users
Language interface	French, English, German, Spanish and Chinese	French and English
Items	List, hierarchy, group and scope	List and hierarchy
Descriptors	Categorical, numerical	Categorical, numerical and calculated. Weight
Illustrations	Local or remote files	Only remote files accessible via URL
Check base	Yes	Yes
Comparison of descriptions	Yes: two, several or comparisons of sets of descriptions	Yes: two, several or comparisons of sets of descriptions
Diagnose	Special features	Special features
Editor	Add, modify, copy, rename, merge, delete, order	Add, modify, copy, rename, merge, delete, order
Detailed sheet	In editor and in identification	In identification only
Publishing	KB content and descriptors with their discriminator power, identification	KB content, identification
Statistics	Yes	Yes
Import formats	<i>Xper</i> , CSV, SDD	CSV, SDD
Export formats	<i>Xper</i> , CSV, SDD, Nexus, HTML, wiki	CSV, SDD, Nexus, HTML
Identification	Multiple (free) access keys	Both single and multiple (free) access keys

KB = knowledge base.

versions of *Xper* (Forget *et al.* 1986, Ung *et al.* 2010, Vignes-Lebbe *et al.* 2016). Two of them were made available at [www.infosyslab.fr](http://www.infosyslab.fr). *Xper2* was developed in years 2003–07 and is a one-user local system. It is available for Windows, macOS and Linux environments, with a user interface in five languages (French, English, German, Spanish and Chinese). *Xper3* is the most recent version and has been available since 2014. It offers a collaborative, multi-user interface in French and English without any installation step. The system is directly opened in an Internet browser, and the KB is stored in a remote MySQL DB. An exchange format in XML (SDD, <http://www.tdwg.org/standards/116>) ensures data exchange with the local version of *Xper2* (Ung *et al.* 2010). The main differences between the two versions are listed below (see Table I).

Currently, *Xper3* requires an Internet connection; however, when there is no connection available, which is often the case when at sea in remote field locations, the easiest solution is to export the KB in SDD format to use it with *Xper2*, the local version of *Xper*. As soon as the Internet is reconnected, it is possible to come back to *Xper3*, as switching between the two versions is easy. Another solution is to export the SDD file for use with *BiodiversiClés*, a software that was developed as a free access key for android mobile phones (<https://apkpure.com/fr/biodiversiclés/mnhn.inpn.biodiversiclés>).

Future developments are in progress that will modify the *Xper* architecture in order to maintain a single core code and various user interfaces. This will allow for the same functionalities in both local and remote access situations. *Xper3* currently offers an approachable way to define terminology for a taxonomic context. A prototype to export descriptive models in ontological language,

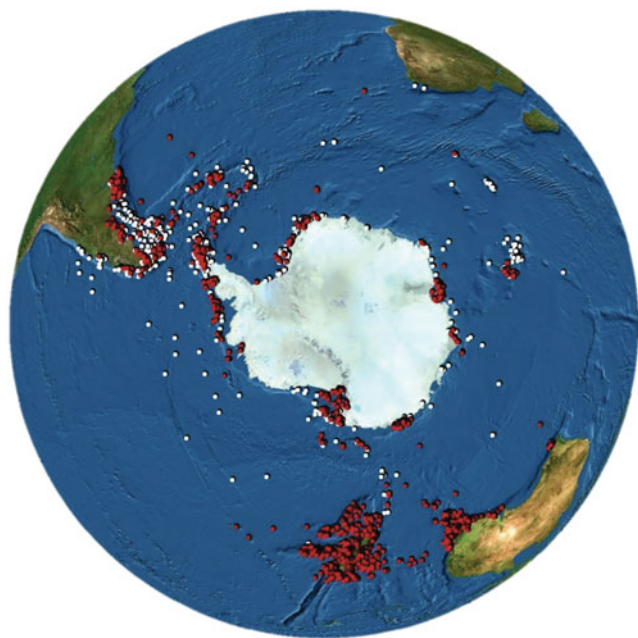
*OWL*, was developed in 2018 (Pellen *et al.* 2018). At the French national level, a project to link between *Xper3* and collections (e-Recolnat) is in progress, and it will offer the possibility to tag collection images with controlled vocabularies structured in the *Xper3* KB (see ANNOTATE on <https://www.recolnat.org/fr>).

#### Two case studies

The following two case studies illustrate an attempt to create a KB and address common practical issues encountered when identifying species in the field or in the laboratory. Firstly, for most taxa in the Southern Ocean, taxonomic and distribution data from the literature are often patchy, which stresses the need for identification keys. In addition, when identification books are finally released, they do not include our current knowledge on systematics and biogeography, and these data need to be updated in a better format. Lastly, experience shows that traditional books are not always 'user friendly' and that species can be misidentified. The two following examples are new KBs that were built with various objectives. The first one, the 'Antarctic Echinoidea' KB, was created to update book content and facilitate access to taxonomic and biogeographical data. The second one, the 'Odontasteridae Southern Ocean (Asteroids)' KB, digitized taxonomic information that had been previously published.

#### Transferring book content to a knowledge base: an example among Antarctic echinoids

In 2005, David *et al.* (2005a) published a book entitled *Antarctic Echinoidea* as a special volume of the book series 'Synopses of the Antarctic Benthos'. The book is a



**Fig. 2.** Recent advances in our knowledge of echinoid distribution in the Southern Ocean, comparing the databases from David *et al.* (2005a, white dots, 2000 records) and Fabri-Ruiz *et al.* (2017, red dots, 7100 records).

synthesis and partial revision of the systematics and distribution of the Antarctic echinoid species south of 45°S. The book content was also complemented by a DB entitled 'Antarctic Echinoids: An Interactive Database' (David *et al.* 2005b). The book started with a general introduction to the Southern Ocean and to the phylogeny and biogeography of Antarctic echinoids, then continued with practical information on the morphology and anatomy of echinoids, including a useful section devoted to anatomical terminology for echinoids. The systematics of Antarctic echinoids constituted the main part of the book. This section proposes a system of three embedded identification keys for the identification of taxa at the family, genus and species level, respectively. The systematic description of genera and species includes a synonymy list, a diagnosis, a morphological description for each species and species distribution data provided on a map. The book has proven useful for non-expert taxonomists and was cited in several biological studies for the identification of Antarctic echinoids (Linse *et al.* 2008, Gutt *et al.* 2011, Hardy *et al.* 2011, Moya *et al.* 2012, Lecointre *et al.* 2013, among others).

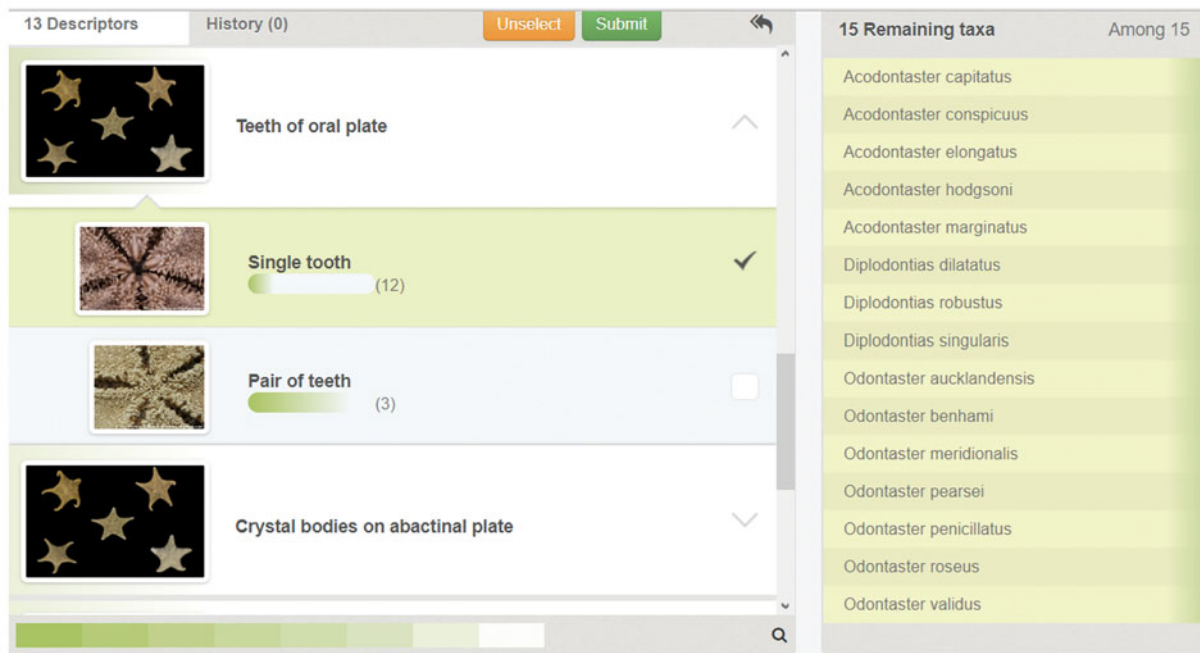
During the last decade, occurrence data on Antarctic marine species have been significantly improved following the increased number of field campaigns undertaken in the Southern Ocean, with many of them being achieved under the CAML and IPY initiatives (Schiaparelli *et al.* 2013). However, our knowledge of

Antarctic echinoid distribution has been improved substantially since the *Antarctic Echinoidea* book was released (Fig. 2) (Saucède *et al.* 2015b; Fabri-Ruiz *et al.* 2017). Currently, echinoid systematics are being revised based on new integrated molecular and morphological research (Díaz *et al.* 2011, Saucède *et al.* 2014, 2015a, David *et al.* 2016). An example of this update is seen in the echinoid genus *Sterechinus* (Díaz *et al.* 2011, Saucède *et al.* 2015a), with now only three entities distinguished, compared to the six nominal species initially attributed to the genus (David *et al.* 2005a).

The content of the *Antarctic Echinoidea* book was transferred to an *Xper3* KB, which included echinoid systematics and occurrence data, which has been updated following recent findings and revisions. Seventy-seven species are described using a descriptive model based on morphological descriptors that are commonly used for echinoid systematics (David *et al.* 2005a). In addition, for each species, a synonymy list, pictures and anatomical drawings, a distribution map and references are given. This *Xper3* KB can be accessed at <http://echinoidea-so.identificationkey.org>. Transferring the book content to the *Xper3* KB proved to be simple and relatively immediate. Adding pictures, specimen plates and figures (occurrence maps) using URL links was easy as soon as the files were ready and available online. Finally, the *Xper3* 'View Description Matrix' tool proved particularly useful for checking the assigned character states and the consistency of the KB.

#### *Building a knowledge base from various sources of information: an example of Antarctic asteroids*

As for many Antarctic taxa, there is no single reference that merges all taxonomic information in order to easily identify Antarctic asteroid specimens collected in the field by non-expert taxonomists. Clark & Downey's (1992) *Starfishes of the Atlantic* represents the most complete attempt to date for the identification of asteroids; however, it is restricted to Atlantic waters (including its southernmost parts, which are adjacent to the Southern Ocean). To find a more comprehensive work focusing on Antarctic asteroids, one must mine through many different sources, including field guides dedicated to particular areas (e.g. Brueggeman 1998 (McMurdo), Schories & Kohlberg 2016 (King George Island)), legacy works of original descriptions (e.g. Koehler 1906, Fisher 1940) and contemporary scientific literature (Janosik & Halanych 2010, Mah & Foltz 2014, Mah *et al.* 2014). In addition to this patchy bibliography, most of the aforementioned references do not include actual identification keys. Therefore, the objective of the Antarctic asteroids KB was to produce an *Xper3*-based identification tool as a unique and useful reference for



**Fig. 3.** *Xper3* identification key of the Odontasteridae (Asteroidea). Descriptors can be selected in the left panel, while remaining species (considering the descriptors already selected) are listed in the right panel. Several macro-pictures were uploaded to show the whole-specimen morphology or close-ups of particular characters.

those who need to identify common species of sea stars from the Southern Ocean.

In the present case, a prototype of an identification key was built for the family Odontasteridae (Fig. 3). Odontasteridae represents one of the most commonly encountered families in the Southern Ocean. In particular, the species *Odontaster validus* is one of the most studied invertebrates of the entire Antarctic continent and plays a major role in shallow-water food webs (McClintock 1994). The family was also a good choice for a prototype key, as diagnostic characters were recently reinvestigated and validated (Janosik & Halanych 2010, 2013; Janosik *et al.* 2011). We focused on 15 species (3 genera: *Acodontaster*, *Diplodontias* and *Odontaster*) and 13 descriptors (e.g. general morphology, number and organization of skeletal plates, morphology of spines, number of teeth and type of pedicellariae). We also took macro-photographs to illustrate whole specimens and descriptors, allowing for a realistic identification. This *Xper3* identification key can be accessed at <http://odontasteridae-so.identificationkey.org>.

Transferring the experts' knowledge from recent scientific papers to an interactive identification key and building the KB through digitization was straightforward and easier to implement than initially assumed. Supplementary tools provided in *Xper3*, such as matrices comparing the diagnostic characters among a set of selected taxa ('Compare Items' tool), proved particularly convenient.

Following the completion of this first key, a medium-term objective is to build an enlarged identification key including the ~120 extant sea star genera in the Southern Ocean (Moreau *et al.* 2015, Jossart *et al.* 2020).

### Knowledge bases as a development framework

#### *Data papers*

Building an *Xper3* KB, as well as any DB, is time-consuming for taxon and data experts, as terminology must be defined unambiguously and all credible data sources must be merged and organized. Such a work can be valuable for data paper publications. A data paper is a relatively new publication type designed to describe a dataset, such as how data were collected, processed and verified, and to make other researchers aware of these data (Pasquetto *et al.* 2017, Kim 2020). Data papers were initiated in the journal *Biodiversity* by a Global Biodiversity Information Facility (GBIF) consortium in discussion with the publisher Pensoft in order to add value to the process of collecting, identifying and building a dataset that can be freely shared amongst all researchers. GBIF now lists 30 journals with such data publications. Other domains now publish data papers, such as physics, chemistry, environmental science and climatology (e.g. *Nature's Scientific Data* journal). The core of a data paper is built on the metadata of a dataset, which precisely describe



dataset coverage, methodological protocols, data validation, field and museum collection information, the relevant usage, etc. Examples are available on the GBIF France website ([www.gbif.fr](http://www.gbif.fr)), including very recent examples (Dillen *et al.* 2019, Simier *et al.* 2019). Data papers on *Xper3* identification keys and KBs also appear in Engel *et al.* (2016), Jouveau *et al.* (2018) and Klimmek & Baur (2018), among others.

### Field guides

The production of taxonomic field guides is another application that can be greatly facilitated by using an established KB. In practice, field guides can prove highly popular and useful to a wide range of users in the field (Féral *et al.* 2019). Field guides are used for naming morphospecies from images and samples taken in the field. They can provide an important baseline resource to assist biologists in monitoring species, communities and environments. They represent a useful first resource for non-specialist taxonomists during collection expeditions. However, they should not be treated as taxonomic references or considered complete, and firm identifications should be avoided without strong evidence. Using field guides must be considered an initial step for the identification of specimens collected in the field before a more precise identification is performed using complementary tools. For instance, identification of species complexes or assessing cryptic diversity requires the combined use of molecular tools and morphological approaches (Sands *et al.* 2008, Tan *et al.* 2010, Martin-Ledo *et al.* 2012, Jossart *et al.* 2019, Moreau *et al.* 2019).

The production of photographic field guides for non-experts is complementary to the production of KBs. The production of field guides and their updates can be expensive, and therefore irregular and incomplete for certain regions, and the taxonomies of some field guides may not be checked by experts for major biological groups. Usually, field guides do not include detailed taxonomic information such as the specimen depository in academic collections and the barcoding associated with specimen depositories, and identification keys are not always provided. Barriers to making marine imagery discoverable and accessible also constitute a major issue, and specific recommendations were recently refined (<https://www.nesp.marine.edu.au/document/data-discoverability-and-accessibility-report-july-2019-workshop-marine-imagery>). Such issues or potential limitations could be partly addressed using KBs. Field guides and updates can be easily generated from a KB such as *Xper3* and, conversely, images from field guides can be integrated into existing KBs. Therefore, field guides and KBs can build upon each other to become more informative and better illustrated guides.

The tools provided by KBs to produce field guides could be particularly helpful for Antarctic biologists, as there is a

real need for generating new Antarctic field guides or updating existing ones. A relatively low number of field guides to the flora and fauna of Antarctica have been made available in comparison to temperate and tropical regions (Brueggeman 1998, Hibberd & Moore 2009, Rauschert & Arntz 2015, Schories & Kohlberg 2016, <https://niwa.co.nz/coasts-and-oceans/marine-identification-guides-and-fact-sheets/amazing-antarctic-asteroids>). In addition, regarding the marine fauna and flora of the Southern Ocean, it appears that the biodiversity of the Antarctic zone has been assessed more often (Brueggeman 1998, De Broyer *et al.* 2014, Rauschert & Arntz 2015, Schories & Kohlberg 2016) than the species of the sub-Antarctic islands (however, see Fischer & Hureau 1985, Hibberd & Moore 2009), in nearshore habitats in particular (Féral *et al.* 2019). Available Antarctic guides mainly focus on epifaunal species (Barnes 2007, Hibberd & Moore 2009, Schories & Kohlberg 2016) or Vulnerable Marine Ecosystems taxa (<https://www.ccamlr.org/en/system/files/VME-guide.pdf>), or include only some benthic fauna (Fischer & Hureau 1985). Coverage of infaunal communities is minimal or only available for larger visible infaunal taxa at the sediment surface (i.e. burrowing bivalves and echinoderms). Many organisms remain undescribed, with either little to no published literature or no other available information.

Initiated under the aegis of SCAR, the SCAR-MarBIN Antarctic Field Guides (AFG) initiative was launched to overcome such a need. This is a collaborative project aimed at establishing and supporting a system of interoperable DBs to provide free access to data to identify Antarctic organisms (<http://afg.biodiversity.aq>). Field guides can be easily built upon, updated and customized from the content of controlled data sources, such as the SCAR Antarctic Biodiversity Portal (free and open access to Antarctic marine biodiversity data), ANTABIF (Antarctic biodiversity data), RAMS (access to taxonomic lists of species occurring in the Southern Ocean), GBIF and Panoramio (a photo-sharing community). The generated field guides can be browsed online or downloaded into a PDF format, published or printed to be taken into the field. Some examples are provided, but the taxonomic coverage remains limited to common or emblematic organisms (<http://afg.biodiversity.aq/guides>).

### Conclusions and future prospects

The objectives of the Leuven taxonomic workshop were to present useful computer-aided identification tools and software to assist Antarctic biologists in areas from species identification to the production of KBs and field guides. The software *Xper3* was chosen as an example of

a KB management system and development framework for the various applications. In general, the *Xper3* software would greatly benefit from a wider audience. More specifically, the tools provided by KB could be particularly helpful for generating and disseminating new Antarctic field guides and updating existing ones based on existing and controlled data on organisms' morphology and taxonomy (either from personal data or free, open-access DBs). Additional field guides and identification keys for a wide range of organisms are needed by the Antarctic community for ecological studies, the monitoring of biodiversity dynamics and the application of conservation measures.

There is, however, a need to increase the accessibility and interoperability of *Xper3* with other taxonomic informatics tools and field guides such as the SCAR Antarctic Biodiversity Portal, ANTABIF and RAMS, and to encourage the use of such tools in the scientific community. Moreover, another welcome addition would be the inclusion of taxonomic, genetic and bibliographical backbones from reference DBs (e.g. WoRMS, Catalogue of Life, bioRxiv, Entrez, Ensembl) directly into the KBs in the form of dynamic links.

Identification tools are often neglected and considered helpful only for taxonomists. KBs could present them to a wider audience, making a better link between taxonomists and non-taxonomists and encouraging the creation of preliminary KBs from their observations. Broadly, we highlight several advantages provided by KBs over more classical tools, such as improved portability, update frequency and visibility, even if renewed efforts are still required for the continuous advancement of KBs and a successful application of electronic field guides and computer-aided identification tools.

In recognition of the need for an improved understanding of species and community responses to environmental change, many key questions in the SCAR 'Antarctic Life on the Precipice' project cover topics relating to the inclusion of biodiversity research for the 2020 Science Horizon Scan (Kennicutt *et al.* 2015). Coordinating biodiversity collections and increasing access to taxonomic and genetic identification resources are key strategic aims. The risk of non-native species introductions into Antarctica is increasing (Bartlett *et al.* 2020). Recent evidence suggests that physical barriers to biological dispersal, previously assumed to isolate Southern Ocean fauna, can be crossed, a phenomenon enhanced by changing climate conditions (Fraser *et al.* 2018). Biofouling organisms can survive sea-ice scraping of vessel hulls (Hughes & Ashton 2017) and anti-fouling treatment (Lee & Chown 2007). Finally, as shipping visits to Antarctic coastal waters increase, particularly in the tourist industry (Verbitsky 2013, Kruczek *et al.* 2018), the risk of non-native introduction increases (McCarthy *et al.* 2019). Facilitating species inventories

using KBs can help us to monitor native species and the community state and track non-native species. Field guides also provide an important baseline resource to assist in monitoring Antarctic marine environments for invasive species by non-expert observers and conservation managers.

### Acknowledgements

We are grateful to the organizing committee of the SCAR XIIth Biology Symposium for the excellent facilities provided for the holding of the taxonomic workshop. Two anonymous reviewers provided helpful comments on an earlier draft of this manuscript.

### Author contributions

TS and ME conceived and designed this work. TS, QJ and CM conceived and built the two example KBs. TS, RV-L, QJ and CM prepared the figures and the table. TS wrote the manuscript with contributions from all authors.

### Financial support

TS benefited from the support of the 'Refugia and Ecosystem Tolerance in the Southern Ocean' project (RECTO; BR/154/A1/RECTO) funded by the 'Belgian Science Policy Office' (BELSPO). The participation of NB was made possible thanks to support from David Barnes from the British Antarctic Survey and the Antarctic Seabed Carbon Capture Change project.

### References

- ALLAN, E.L., FRONEMAN, P.W., DURGADO, J.V., MCQUAID, C.D., ANSORGE, I.J. & RICHOUX, N.B. 2013. Critical indirect effects of climate change on sub-Antarctic ecosystem functioning. *Ecology and Evolution*, **3**, 10.1002/ece3.678.
- BALMFORD, A. & GASTON, K.J. 1999. Why biodiversity surveys are good value. *Nature*, **398**, 204–205.
- BARNES, D.K.A. 2007. Benthic communities in the Southern Ocean. In RIFFENBURGH, B., ed. *Encyclopedia of the Antarctic*. New York: Routledge, 142–144.
- BARTLETT, J.C., CONVEY, P., PERTIERRA, L.R. & HAYWARD, S.A. 2020. An insect invasion of Antarctica: the past, present and future distribution of *Eretmoptera murphyi* (Diptera, Chironomidae) on Signy Island. *Insect Conservation and Diversity*, **13**, 10.1111/icad.12389.
- BRANDT, A., GOODAY, A.J., BRANDAO, S.N., BRIX, S., BROKELAND, W., CEDHAGEN, T., *et al.* 2007. First insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature*, **447**, 307–311.
- BRUEGGEMAN, P. 1998. *Underwater field guide to Ross Island & McMurdo Sound, Antarctica*. San Diego, CA: The National Science Foundation's Office of Polar Programs. Available at <http://www.peterbrueggeman.com/nsf/fguide/>
- BUENO, G., DENIZ, O., PEDRAZA, A., RUIZ-SANTAQUITERIA, J., SALIDO, J., CRISTÓBAL, G., *et al.* 2017. Automated diatom classification (part A): handcrafted feature approaches. *Applied Sciences*, **7**, 10.3390/app7080753.
- BYRNE, M., GALL, M., WOLFE, K. & AGÜERA, A. 2016. From pole to pole: the potential for the Arctic seastar *Asterias amurensis* to invade

- a warming Southern Ocean. *Global Change Biology*, **22**, 10.1111/gcb.13304.
- CARVALHO, F.A., FILER, D. & RENNER, S.S. 2014. Taxonomy in the electronic age and an e-monograph of the papaya family (Caricaceae) as an example. *Cladistics*, **31**, 10.1111/cla.12095.
- CARY, S.C., McDONALD, I.R., BARRETT, J.E. & COWAN, D.A. 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. *Nature Reviews Microbiology*, **8**, 129–138.
- CCAMLR. 2009. *Conservation measure 91-03. Protection of the South Orkney Islands southern shelf*. Hobart: CCAMLR.
- CHAPMAN, A.D. 2005. Uses of Primary Species-Occurrence Data, version 1.0. *Report for the Global Biodiversity Information Facility*. Available at <http://www.gbif.org/document/80545>.
- CHAWAN, V. & PENEV, L. 2011. The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, **12**, S2.
- CLARK, A. M. & DOWNEY, M.E. 1992. *Starfishes of the Atlantic*. London: Chapman & Hall, 820 pp.
- CLARKE, A. & JOHNSTON, N.M. 2003. Antarctic marine benthic diversity. *Oceanography and Marine Biology*, **41**, 47–114.
- CONRUYT, N., GROSSER, D., FAURE, G., PICHON, M., GEYNET, Y., GUILLAUME, M., *et al.* 1998. *IKBS: an iterative knowledge base system for managing knowledge in remote systematics. Application to corals of the Mascarene Archipelago*. Perpignan: International Society for Reef Studies (ISRS) European Meeting.
- COSTELLO, M.J., VANHOORNE, B. & APPELTANS, W. 2015. Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conservation Biology*, **29**, 10.1111/cobi.12496.
- DALLWITZ, M.J. 1980. A general system for coding taxonomic descriptions. *Taxon*, **29**, 41–46.
- DALLWITZ, M.J. 2000. *A comparison of interactive identification programs*. Available at <https://www.delta-intkey.com>.
- DANIS, B., VAN DE PUTTE, A., RENAUDIER, S. & GRIFFITHS, H. 2013. Connecting biodiversity data during the IPY: the path towards e-polar science. In VERDE, C. & DI PRISCO, G., eds. *Adaptation and evolution in marine environments 2*. Berlin: Springer, 21–32.
- DAVID, B., CHONÉ, T., MOOI, R. & DE RIDDER, C. 2005a. *Antarctic Echinoidea. Synopses of the Antarctic benthos*. Königstein: Koeltz Scientific Books, 274 pp.
- DAVID, B., CHONÉ, T., FESTEUAU, A., MOOI, R. & DE RIDDER, C. 2005b. Biodiversity of Antarctic echinoids: a comprehensive and interactive database. *Scientia Marina*, **69**, 201–203.
- DAVID, B., SAUCÈDE, T., CHENUIL, A., STEIMETZ, A. & DE RIDDER, C. 2016. The taxonomic challenge posed by the Antarctic echinoids *Abatus bidens* and *Abatus cavernosus* (Schizasteridae, Echinoidea). *Polar Biology*, **39**, 10.1007/s00300-015-1842-5.
- DE BROYER, C. & DANIS, B. 2011. How many species in the Southern Ocean? Towards a dynamic inventory of the Antarctic marine species. *Deep-Sea Research II*, **58**, 5–17.
- DE BROYER, C., KOUUBI, P., GRIFFITHS, H.J., RAYMOND, B., UDEKEM D'ACCOZ, C. D., VAN DE PUTTE, A.P., *et al.* 2014. *Biogeographic atlas of the Southern Ocean*. Cambridge: Scientific Committee on Antarctic Research, xii + 498 pp.
- DÍAZ, A., FÉRAL, J.-P., DAVID, B., SAUCÈDE, T. & POULIN, E. 2011. Evolutionary pathways among shallow and deep-sea echinoids of the genus *Sterechinus* in the Southern Ocean. *Deep-Sea Research II*, **58**, 10.1016/j.dsr2.2010.10.012.
- DILLEN, M., GROOM, Q., CHAGNOUX, S., GÜNTSCH, A., HARDISTY, A., HASTON, E., *et al.* 2019. A benchmark dataset of herbarium specimen images with label data. *Biodiversity Data Journal*, **7**, 10.3897/BDJ.7.e31817.
- ENGEL, J., BROUSSEAU, L. & BARALOTO, C. 2016. *GuiaTreeKey*, a multi-access electronic key to identify tree genera in French Guiana. *PhytoKeys*, **68**, 10.3897/phytokeys.68.8707.
- FABRI-RUIZ, S., SAUCÈDE, T., DANIS, B. & DAVID, B. 2017. Southern Ocean echinoids database - an updated version of Antarctic, sub-Antarctic and cold temperate echinoid database. *ZooKeys*, **697**, 1–20.
- FÉRAL, J.-P., POULIN, E., DE RIDDER, C. & SAUCÈDE, T. 2019. A field guide to coastal echinoderms of the Kerguelen Islands. *Zoosymposia*, **15**, 10.11646/zoosymposia.15.1.6.
- FISCHER, W.K. & HUREAU, J.-C. 1985. *FAO species identification sheets for fishery purposes Southern Ocean: fishing Areas 48, 58 and 88 (CCAMLR Convention Area)*. Rome: FAO, 470 pp.
- FISHER, W.K. 1940. Asteroidea. *Discovery Reports*, **20**, 69–306.
- FORGET, P.-M., LEBBE, J., PUIG, H., VIGNES, R., HIDEUX, F.L.S. & HIBEUX, M. 1986. Microcomputer aided identification: an application to trees from French Guiana. *Botanical Journal of the Linnean Society*, **93**, 10.1111/j.1095-8339.1986.tb01021.x.
- FRASER, C.I., MORRISON, A.K., HOGG, A.M., MACAYA, E.C., VAN SEBILLE, E., RYAN, P.G., *et al.* 2018. Antarctica's ecological isolation will be broken by storm-driven dispersal and warming. *Nature Climate Change*, **8**, 704.
- GASTON, K.J. 2005. Biodiversity and extinction: species and people. *Progress in Physical Geography*, **29**, 239–247.
- GORSKY, G., OHMAN, M.D., PIPHERAL, M., GASPARINI, S., STEMMANN, L., ROMAGNAN, J.B., *et al.* 2010. Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research March*, **32**, 285–303.
- GRIFFITHS, H.J. 2010. Antarctic marine biodiversity - what do we know about the distribution of life in the Southern Ocean? *PLoS One*, **5**, 10.1371/journal.pone.0011683.
- GRIFFITHS, H.J., DANIS, B. & CLARKE, A. 2011. Quantifying Antarctic marine biodiversity: the SCAR-MarBIN data portal. *Deep-Sea Research II*, **58**, 10.1016/j.dsr2.2010.10.008.
- GUTT, J., BARNES, D.K.A., LOCKHART, S. & VAN DE PUTTE, A. 2013. Antarctic macrobenthic communities: a compilation of circumpolar information. *Nature Conservation*, **4**, 10.3897/natureconservation.4.4499.
- GUTT, J., BARRATT, I., DOMACK, E., D'UDEKEM D'ACCOZ, C., DIMMLER, W., GRÉMARE, A., *et al.* 2011. Biodiversity change after climate-induced ice-shelf collapse in the Antarctic. *Deep-Sea Research II*, **58**, 74–83.
- GUTT, J., BERTLER, N., BRACEGIRDLE, T.J., BUSCHMANN, A., COMISO, J., HOSIE, G., *et al.* 2015. The Southern Ocean ecosystem under multiple climate change stresses? An integrated circumpolar assessment. *Global Change Biology*, **21**, 1434–1453.
- GUTT, J., ISLA, E., BERTLER, A.N., BODEKER, G.E., BRACEGIRDLE, T., CAVANAGH, R.D., *et al.* 2017. Cross-disciplinary in the advance of Antarctic ecosystem research. *Marine Genomics*, **37**, 1–17.
- HAGEDORN, G., RAMBOLD, G. & MARTELLI, S. 2010. Types of identification keys. In NIMIS, P.L. & VIGNES-LEBBE, R., eds. *Tools for identifying biodiversity: progress and problems. Proceedings of the International Congress, Paris, September 20–22, 2010*. Trieste: EUT Edizioni Università di Trieste, 59–64.
- HARDY, C., DAVID, B., RIGAUD, T., DE RIDDER, C. & SAUCÈDE, T. 2011. Ectosymbiosis associated with cidaroids (Echinodermata: Echinoidea) promotes benthic colonization of the seafloor in the Larsen Embayments, Western Antarctica. *Deep-Sea Research II*, **58**, 84–90.
- HIBBERD, T. & MOORE, K. 2009. *Field identification guide to Heard Island and McDonald Islands benthic invertebrates, a guide for scientific observers on board fishing vessels in that area*. Kingston: The Department of Environment, Water, Heritage, and the Arts, Australian Antarctic Division and the Fisheries Research and Development Corporation, Australia, 158 pp.
- HU, J., CHEN, Z., YANG, M., ZHANG, R. & CUI, Y. 2018. A multiscale fusion convolutional neural network for plant leaf recognition. *IEEE Signal Processing Letters*, **25**, 853–857.
- HUGHES, K.A. & ASHTON, G.V. 2017. Breaking the ice: the introduction of biofouling organisms to Antarctica on vessel hulls. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **27**, 10.1002/aqc.2625.
- JANOSIK, A.M. & HALANYCH, K.M. 2010. Unrecognized Antarctic biodiversity: a case study of the genus *Odontaster* (Odontasteridae; Asteroidea). *Integrative and Comparative Biology*, **50**, 981–992.
- JANOSIK, A.M. & HALANYCH, K.M. 2013. Seeing stars: a molecular and morphological investigation into the evolutionary history of

- Odontasteridae (Asteroidea) with description of a new species from the Galapagos Islands. *Marine Biology*, **160**, 821–841.
- JANOSIK, A.M., MAHON, A.R. & HALANYCH, K.M. 2011. Evolutionary history of Southern Ocean *Odontaster* sea star species (Odontasteridae; Asteroidea). *Polar Biology*, **34**, 575–586.
- JOLY, A., VROCHIDIS, S., KARATZAS, K., KARPPINEN, A. & BONNET, P. 2018. *Multimedia tools and applications for environmental & biodiversity informatics*. Cham: Springer International Publishing, 221 pp.
- JOLY, A., GOËAU, H., BONNET, P., BAKIĆ, V., BARBE, J., SELMI, S., *et al.* 2014. Interactive plant identification based on social image data. *Ecological Informatics*, **23**, 22–34.
- JOSSART, Q., SANDS, C.J. & SEWELL, M.A. 2019. Dwarf brooder *versus* giant broadcaster: combining genetic and reproductive data to unravel cryptic diversity in an Antarctic brittle star. *Heredity*, **123**, 622–633.
- JOSSART, Q., KOCHZIUS, M., DANID, B., SAUCÈDE, T. & MOREAU, C. 2020. Diversity of the Pterasteridae (Asteroidea) in the Southern Ocean: a molecular and morphological approach. *Zoological Journal of the Linnean Society*, 10.1093/zoolinnean/zlaa097/5907661.
- JOSSART, Q., MOREAU, C., AGÜERA, A., DE BROYER, C. & DANIS, B. 2015. The Register of Antarctic Marine Species (RAMS): a ten-year appraisal. *ZooKeys*, **524**, 137.
- JOUVEAU, S., DELAUNAY, M., VIGNES-LEBBE, R. & NATTIER, R. 2018. A multi-access identification key based on colour patterns in ladybirds (Coleoptera, Coccinellidae). *ZooKeys*, **758**, 10.3897/zookeys.758.22171.
- KAISER, S., BRANDÃO, S.N., BRIX, S., BARNES, D.K.A., BOWDEN, D.A., INGELS, J., *et al.* 2013. Patterns, processes and vulnerability of Southern Ocean benthos: a decadal leap in knowledge and understanding. *Marine Biology*, **160**, 10.1007/s00227-013-2232-6.
- KARGEL, J.S., BUSH, A.B.G., COGLEY, J.G., LEONARD, G.J., RAUP, B.H., SMIRAGLIA, C.M., *et al.* 2014. A world of changing glaciers: summary and climatic context. In KARGEL, J.S., LEONARD, G.J., BISHOP, M.P., KÄÄB, A. & RAUP, B.H., eds. *Global land ice measurements from space*. Berlin: Springer, 781–840.
- KENNICUTT, M.C., CHOWN, S.L., CASSANO, J.J., LIGGETT, D., PECK, L.S., MASSOM, R., *et al.* 2015. A roadmap for Antarctic and Southern Ocean science for the next two decades and beyond. *Antarctic Science*, **27**, 10.1017/S0954102014000674.
- KIM, J. 2020. An analysis of data paper templates and guidelines: types of contextual information described by data journals. *Science Editing*, **7**, 16–23.
- KLIMMEK, F. & BAUR, H. 2018. An interactive key to Central European species of the *Pteromalus albipennis* species group and other species of the genus (Hymenoptera: Chalcidoidea: Pteromalidae), with the description of a new species. *Biodiversity Data Journal*, **6**, 10.3897/BDJ.6.e27722.
- KOEHLE, R. 1906. Note préliminaire sur les Échinodermes recueillis par l'expédition, antarctique fran Yaise du Dr. Charcot (Échinides, Astéries et Ophiures). *Bulletin du Museum Paris*, **1905**, 464–470.
- KRUCZEK, Z., KRUCZEK, M. & SZROMEK, A. 2018. Possibilities of using the tourism area life cycle model to understand and provide sustainable solution for tourism development in the Antarctic Region. *Sustainability*, **10**, 89.
- KUANG, Y. 2015. *Deep neural network for deep sea plankton classification. Technical report*. Stanford, CA: Stanford University Press, 4321–4327.
- KUMAR, N., BELHUMEUR, P., BISWAS, A., JACOBS, D., KRESS, W., LOPEZ I., *et al.* 2012. *Leafsnap*: a computer vision system for automatic plant species identification. In FITZGIBBON, A., LAZEBNIK, S., PERONA, P., SATO, Y., & SCHMID, C., eds. *Computer Vision - ECCV 2012*. Berlin: Springer, 502–516.
- LECOINTRE, G., AMÉZIANE, N., BOISSELIER, M.-C., BONILLO, C., BUSSON, F., CAUSSE R., *et al.* 2013. How operational is the species flock concept? The Antarctic shelf case. *PLoS One*, **8**, 10.1371/journal.pone.0068787.
- LEE, J.E. & CHOWN, S.L. 2007. *Mytilus* on the move: transport of an invasive bivalve to the Antarctic. *Marine Ecology Progress Series*, **339**, 307–310.
- LINSE, K., WALKER, L.J. & BARNES, D.K. 2008. Biodiversity of echinoids and their epibionts around the Scottia Arc, Antarctica. *Antarctic Science*, **20**, 227–244.
- MAH, C.L. & FOLTZ, D.W. 2014. New taxa and taxonomic revisions to the Poraniidae (Valvatacea; Asteroidea) with comments on feeding biology. *Zootaxa*, **3795**, 327–372.
- MAH, C.L., NEILL, K., ELÉAUME, M. & FOLTZ, D. 2014. New species and global revision of Hippasteria (Hippasterinae: Goniasteridae; Asteroidea; Echinodermata). *Zoological Journal of the Linnean Society*, **171**, 422–456.
- MARTÍN-LEDO, R., LÓPEZ-GONZÁLEZ, P. & SANDS, C.J. 2012. A new brooding species of brittle star (Echinodermata: Ophiuroidea) from Antarctic waters. *Polar Biology*, **36**, 10.1007/s00300-012-1242-z.
- MAY, R.M. 2011. Why worry about how many species and their loss? *PLoS Biology*, **9**, 10.1371/journal.pbio.1001130.
- MCCARTHY, A.H., PECK, L.S., HUGHES, K.A. & ALDRIDGE, D.C. 2019. Antarctica: the final frontier for marine biological invasions. *Global Change Biology*, **25**, 2221–2241.
- MCCLINTOCK, J.B. 1994. Trophic biology of Antarctic shallow-water echinoderms. *Marine Ecology Progress Series*, **111**, 191–202.
- MOLINOS, J.G., HALPERN, B.S., SCHOEMAN, D.S., BROWN, C.J., KIESSLING, W., MOORE, P.J., *et al.* 2015. Climate velocity and the future global redistribution of marine biodiversity. *Nature Climate Change*, **6**, 10.1038/nclimate2769.
- MOREAU, C., AGÜERA, A., JOSSART, Q. & DANIS, B. 2015. Southern Ocean Asteroidea: a proposed update for the Register of Antarctic Marine Species. *Biodiversity Data Journal*, **3**, 10.3897/BDJ.3.e7062.
- MOREAU, C., DANIS, B., JOSSART, Q., ELÉAUME, M., SANDS, C., ACHAZ, G., *et al.* 2019. Is reproductive strategy a key factor in understanding the evolutionary history of Southern Ocean Asteroidea (Echinodermata)? *Ecology and Evolution*, **9**, 8465–8478.
- MOYA, F., SAUCÈDE, T. & MANJÓN-CABEZA, M.E. 2012. Environmental control on the structure of echinoid assemblages in the Bellingshausen Sea (Antarctica). *Polar Biology*, **35**, 10.1007/s00300-012-1176-5.
- NEUBACHER, D. & RAMBOLD, G. 2005. *NaviKey* - a Java applet and application for accessing descriptive data coded in DELTA format. Available at <http://www.navikey.net>.
- NIMIS, P.L. & VIGNES-LEBBE, R. 2010. *Tools for identifying biodiversity: progress and problems. Proceedings of the International Congress, Paris, September 20–22*. Trieste: EUT Edizioni Università di Trieste, 445 pp.
- PASQUETTO, I.V., RANDLES, B.M. & BORGMAN, C.L. 2017. On the reuse of scientific data. *Data Scientific Journal*, **16**, 10.5334/dsj-2017-008.
- PELLEN, F., BOUQUIN, S., MOUGENOT, I. & VIGNES-LEBBE, R. 2018. Building an OWL ontology with Xper3. *Biodiversity Information Science and Standards*, **2**, 10.3897/biss.2.25614.
- PENDLEBURY, S.F. & BARNES-KEOGHAN, I.P. 2007. Climate and climate change in the sub-Antarctic. *Papers and Proceedings of the Royal Society of Tasmania*, **141**, 10.26749/rstpp.141.1.67.
- PESANT, S., NOT, F., PICALER, M., KANDELS-LEWIS, S., LE BESCOT, N., GORSKY, G., *et al.* 2015. Open science resources for the discovery and analysis of Tara Oceans data. *Sciences Data*, **2**, 10.1038/sdata.2015.23.
- RAUSCHERT M. & ARNTZ W.E. 2015. *Antarctic macrobenthos: a field guide of the invertebrates living at the Antarctic seafloor*. Wurster Nordseekueste: Arntz & Rauschert Selbstverlag, 143 pp.
- SANDS, C.J., MCINNES, S.J., MARLEY, N.J., GOODALL-COPESTAKE, W.P., CONVEY, P. & LINSE K. 2008. Phylum Tardigrada: an 'individual' approach. *Cladistics*, **24**, 861–871.
- SAUCÈDE, T., PIERRAT, B. & DAVID, B. 2014. Echinoids. In DE BROYER, C., KOUUBI, P., GRIFFITHS, H.J., RAYMOND, B., UDEKEM D'ACOS D', C., VAN DE PUTTE, A.P. *et al.*, eds. *Biogeographic atlas of the Southern Ocean*. Cambridge: Scientific Committee on Antarctic Research, 213–220.
- SAUCÈDE, T., DÍAZ, A., PIERRAT, B., SELLANES, J., DAVID, B., FÉRAL, J.-P., *et al.* 2015a. The phylogenetic position and taxonomic status of *Sterechinus bernasconiae* Larrain, 1975 (Echinodermata, Echinoidea), an enigmatic Chilean sea urchin. *Polar Biology*, **38**, 1223–1237.

- SAUCÈDE, T., GRIFFITHS, H., MOREAU, C., JACKSON, J., SANDS, C., DOWNEY, R., *et al.* 2015b. East Weddell Sea echinoids from the JR275 expedition. *ZooKeys*, **504**, 10.3897/zookeys.504.8860.
- SCHIAPARELLI, S., DANIS, B., WADLEY, V. & STODDART, D.M. 2013. The census of Antarctic marine life: the first available baseline for Antarctic marine biodiversity. In VERDE, V. & DI PRISCO, G., eds. *Adaptation and evolution in marine environments*, vol. 2. Berlin: Springer, 3–19.
- SCHORIES, D. & KOHLBERG, G. 2016. *Marine wildlife King George Island Antarctica: identification guide*. Rostock: Dirk Schories Publications, 348 pp.
- SCHRAM, J.B., SCHOENROCK, K.M., MCCLINTOCK, J.B., AMSLER, C.D. & ANGUS, R.A. 2016. Testing Antarctic resilience: The effects of elevated seawater temperature and decreased pH on two gastropod species. *ICES Journal of Marine Science*, **73**, 10.1093/icesjms/fsv233.
- SIMIER, M., ECOUTIN, J. & TITO DE MORAIS, L. 2019. The PPEAO experimental fishing dataset: fish from West African estuaries, lagoons and reservoirs. *Biodiversity Data Journal*, **7**, 10.3897/BDJ.7.e31374.
- TAN, D.S.H., ANG, Y., LIM, G.S., ISMAIL, M.R.B. & MEIER, R. 2010. From 'cryptic species' to integrative taxonomy: an iterative process involving DNA sequences, morphology, and behaviour leads to the resurrection of *Sepsis pyrrhosoma* (Sepsidae: Diptera). *Zoologica Scripta*, **39**, 10.1111/j.1463-6409.2009.00408.x.
- TELETCHÉA, F. 2010. After 7 years and 1000 citations: comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA*, **21**, 206–226.
- THIELE, K. & DALLWITZ, M.J. 2000. A critique of Dallwitz's 'A comparison of interactive identification programs'. Available at <https://www.delta-intkey.com/www/thiele.htm>.
- TURNER, J., BARRAND, N., BRACEGIRDLE, T., CONVEY, P., HODGSON, D.A., JARVIS, M., *et al.* 2014. Antarctic climate change and the environment: an update. *Polar Record*, **50**, 10.1017/S0032247413000296.
- UNG, V., DUBUS, G., ZARAGÜETA-BAGILS, R. & VIGNES-LEBBE, R. 2010. *Xper2*: introducing e-taxonomy. *Bioinformatics*, **26**(5), 10.1093/bioinformatics/btp715.
- VERBITSKY, J. 2013. Antarctic tourism management and regulation: the need for change. *Polar Record*, **49**, 10.1017/S003224741200071X.
- VIGNES-LEBBE, R., CHESSELET, P. & DIEP THI, M.-H. 2016. *Xper3*: new tools for collaborating, training and transmitting knowledge on botanical phenotypes. In RAKOTOARISOA, N.R., BLACKMORE, S. & RIERA, B., eds. *Botanists of the twenty-first century: roles, challenges and opportunities*. Paris: UNESCO, 228–239.
- WILKINSON, M.D., DUMONTIER, M., AALBERSBERG, I.J., APPLETON, G., AXTON, M., BAAK, A., *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, **3**, 10.1038/sdata.2016.18.