



OPEN

DATA DESCRIPTOR

Myctobase, a circumpolar database of mesopelagic fishes for new insights into deep pelagic prey fields

Briannyn Woods^{1,11}✉, Rowan Trebilco^{1,2}, Andrea Walters¹, Mark Hindell¹, Guy Duhamel³, Hauke Flores⁴, Masato Moteki^{5,6}, Patrice Pruvost³, Christian Reiss⁷, Ryan A. Saunders⁸, Caroline Sutton², Yi-Ming Gan⁹ & Anton Van de Putte^{9,10,11}✉

The global importance of mesopelagic fish is increasingly recognised, but they remain poorly studied. This is particularly true in the Southern Ocean, where mesopelagic fishes are both key predators and prey, but where the remote environment makes sampling challenging. Despite this, multiple national Antarctic research programs have undertaken regional sampling of mesopelagic fish over several decades. However, data are dispersed, and sampling methodologies often differ precluding comparisons and limiting synthetic analyses. We identified potential data holders by compiling a metadata catalogue of existing survey data for Southern Ocean mesopelagic fishes. Data holders contributed 17,491 occurrence and 11,190 abundance records from 4780 net hauls from 72 different research cruises. Data span across 37 years from 1991 to 2019 and include trait-based information (length, weight, maturity). The final dataset underwent quality control processes and detailed metadata was provided for each sampling event. This dataset can be accessed through Zenodo. *Myctobase* will enhance research capacity by providing the broadscale baseline data necessary for observing and modelling mesopelagic fishes.

Background & Summary

Open-ocean pelagic ecosystems are under-represented in databases of marine biodiversity, despite holding the largest biomass of organisms on Earth¹. The open-ocean pelagic community predominantly consists of fish, crustaceans and cephalopods that inhabit mesopelagic (200–1000 m) and bathypelagic (1000 m to >4000 m) depths². Many pelagic species undertake diel vertical migration (DVM), moving from the depths to shallower waters at dusk and in the reverse direction at dawn, possibly constituting the largest animal migration on the planet³. This has implications for carbon sequestration and climate regulation, as organisms actively transport organic carbon from the surface to the deep ocean^{3–6}.

Mesopelagic fishes are a central component of open-ocean pelagic communities dominating global vertebrate biomass with estimates of up to 10 billion tons^{4,7,8}. They are thought to represent a key link to coupling physical-biogeochemical models to the population dynamics of top-predators^{6,9}. Thus, the open-ocean pelagic environment and its inhabitants are critical components for the provision of globally important ecosystem services^{1,2}.

¹Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, 7004, Australia. ²CSIRO, Oceans and Atmosphere, Hobart, 7004, Australia. ³Laboratoire BORA (UMR 8067), Muséum National d'Histoire Naturelle, Paris, 75005, France. ⁴Alfred-Wegener-Institut Helmholtz-Zentrum für Polar und Meeresforschung (AWI), Bremerhaven, 27570, Germany. ⁵Tokyo University of Marine Science and Technology, Tokyo, 108-8477, Japan. ⁶National Institute of Polar Science, Tokyo, 190-8518, Japan. ⁷Antarctic Ecosystem Research Division, Southwest Fisheries Science Center, La Jolla, California, 92037, USA. ⁸British Antarctic Survey, Cambridge, CB3 0ET, UK. ⁹Royal Belgian Institute for Natural Sciences, Brussel, B-1000, Belgium. ¹⁰Université Libre de Bruxelles, Brussel, B-1000, Belgium. ¹¹These authors contributed equally: Briannyn Woods, Anton Van de Putte. ✉e-mail: bree.woods@utas.edu.au; avandeputte@naturalsciences.be

In the Southern Ocean, mesopelagic fishes are key prey for sentinel species such as seals and seabirds^{10–12}. As major consumers of secondary productivity (zooplankton) they also exert control on lower trophic levels of oceanic food webs^{13,14}. Despite their ecologically important role, there are gaps in our knowledge of their biodiversity, abundance, biomass, and the processes that shape their distribution, life cycles and behaviour^{4,8,15–19}. Mesopelagic fishes are difficult to sample due to their patchy distribution and ability to avoid and escape pelagic nets^{8,20,21}. This is confounded by the differences in catch efficiency between gear types, leading to biased estimates of abundance and biomass^{4,8}.

Mesopelagic fishes are likely to be impacted by ocean-warming with evidence indicating future range shifts of temperate species, range reductions of Antarctic species^{16,22} and possible biogeographic shifts in body size patterns²³. This will have implications for the predators that feed upon them¹⁴ and potentially more broadly for global biogeochemical cycles²⁴. Tracking the future distribution and abundance of species requires the development of reliable population baseline estimates and an understanding of the biases associated with different sampling strategies^{25–28}.

The advent of biodiversity informatics has seen the development of open-access biodiversity data repositories, such as the Global Biodiversity Information Facility (GBIF) and the Ocean Biogeographic Information System (OBIS), to enhance research output in the context of a global biological monitoring system²⁹. However, much of the data for mesopelagic fishes archived in these repositories are for species occurrence only, without information on abundance or biomass. Often, the methodological metadata (such as net-type or depth of trawl) are also lacking, further limiting the scope of the analyses^{1,30}.

Decades of data from net sampling of the pelagic environment exist for localised regions of the Southern Ocean as part of multiple national Antarctic research programs³¹. Here, we take the important step of integrating and standardising the format of published and unpublished survey data on the abundance, biomass, biodiversity, and methodological metadata for mesopelagic fishes of the Southern Ocean. We collected and synthesised data from the Scotia Sea^{14,32–34}, the Antarctic Peninsula^{35,36}, the Lazarev Sea^{37,38}, the Kerguelen Plateau^{39–43}, East Antarctica⁴⁴, Macquarie Island^{45,46} and unpublished data from the Indian Ocean from Japan's national Antarctic program into a single dataset called *Myctobase* (Fig. 1).

The long-term aims of *Myctobase* are (1) to provide a single findable, standardised, and citable resource for multiple combined datasets; (2) to provide the baseline data upon which to measure the status and future trends of the mesopelagic fish assemblage; (3) to support further research such as investigating the patterns and biophysical drivers of diversity; and (4) to provide the broadscale spatiotemporal perspective necessary for the holistic management and conservation of the open-ocean pelagic ecosystem in the Southern Ocean. The dataset is available on GBIF, OBIS and Zenodo to support the management of the open-ocean pelagic ecosystem. Researchers wishing to contribute to this project should contact the corresponding authors. We hope that this database will continue to grow, providing a resource for ongoing monitoring of the Southern Ocean pelagic ecosystem.

Methods

Sampling design. Samples were collected onboard multiple national Antarctic research cruises between the years 1991–2016 and for the year 2019 and jointly covered all calendar months (Online-only Table 1). Net sampling occurred predominantly in the Indian and Atlantic Sectors of the Southern Ocean (Fig. 1).

Mesopelagic fishes were sampled using a range of gear types as a “standard” sampling gear does not currently exist²¹. Samples were collected from pelagic trawls using opening-closing net systems with a range of Rectangular Midwater Trawl (RMT) and International Young Gadoid Pelagic Trawl (IYGPT) nets, an Isaacs-Kidd Midwater Trawl (IKMT) net and a Matsuda-Oozeki-Hu Trawl (MOHT) net (Table 1, Online-only Table 2). Opening-closing net systems allowed for the sampling of discrete depth intervals and included stratified oblique and horizontal trawls. Survey designs (for trawl locations) included randomised design, designated stations, and target trawls on acoustically detected aggregations of fish (see Online-only Table 1 for more information on sampling methodology).

Once onboard, samples were sorted to the lowest taxonomic level possible using published guides^{41,47} and personal/institutional reference collections. Taxonomic identities were verified in home laboratories. Standard length (mm) and wet weight (g) measurements were taken onboard with a motion compensated balance or in home laboratories. Samples were preserved in either ethanol, formalin or frozen for further analyses.

Detailed information on methodologies utilised for each research cruise can be found in the relevant citations listed in Online-only Table 1.

Data collection and processing. Potential data holders were identified by compiling a metadata catalogue of scientific publications and existing survey data for Southern Ocean mesopelagic fishes. Data holders were invited to contribute published and unpublished data to the *Myctobase* project. A standardised template for the collation of multiple datasets was created using Darwin Core terms⁴⁸ where possible. The terms used in the template alongside definitions for each term can be found in Table 1. Data were collated with R statistical software, version 4.0.4⁴⁹.

Units of measurement were standardised across datasets. Length and weight measurements were converted into millimetres and grams, respectively. The taxonomy of each observation was verified and associated alpha IDs (globally unique and stable identifiers for each taxonomic name) were retrieved from the World Register of Marine Species using the R package, WoRMS⁵⁰ (Fig. 2).

Abundance was calculated for each net haul by species per filtered volume of water. This was performed by dividing the species count (n) by the volume filtered (m³). This was not possible for all net hauls and species as count and volume filtered information were not always recorded or technical difficulties were encountered during the research cruise. Methodology for calculating volume filtered varied between datasets. Several cruises

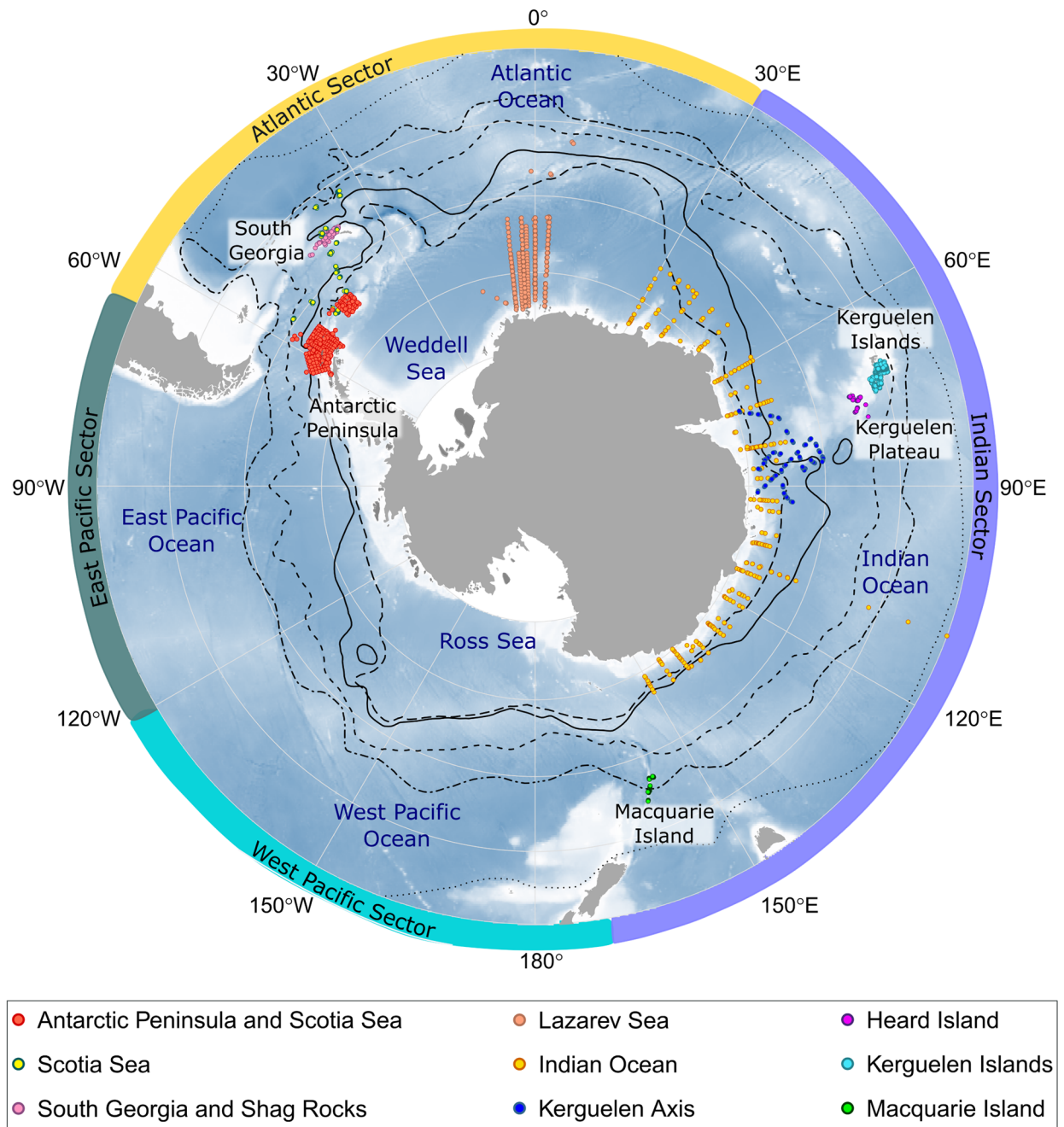


Fig. 1 The sampling locations of mesopelagic fish trawl stations that are currently held in *Myctobase*. Trawl stations are colour coded to correspond to the location listed in Online-only Table 1 (colour key is indicated within the black box). Black lines indicate mean frontal positions⁵². Starting from the outer line, frontal features are as follows: Subtropical Front (dotted line); Subantarctic Front (dash-dot line); Polar Front (dashed line); southern Antarctic Circumpolar Current Front (solid line); southern boundary of the ACC (long dashed line). Numbered labels around the outside of the map indicate the longitude. Each latitude line represents 10° of latitude where 75° S is at the Antarctic continent and 40° S is at the external line.

used mechanical flow meters while others used a calculation where the nets filtering area (m^2) was multiplied by the distance it was towed (m). The filtering area of the MOHT net was calculated by multiplying the net mouth area by a coefficient estimated from calibration tows in calm seas. The coefficient was estimated under a calm sea condition using a net frame with a flowmeter which was vertically deployed up to 100 m wire out. The net frame was retrieved slowly to the sea surface and a count from the flowmeter was recorded. This procedure was repeated five times and an average value was calculated to obtain the coefficient. For some datasets, volume filtered was not previously calculated, but the information needed to calculate the volume filtered had been recorded. In these instances, we retrospectively calculated the volume filtered using this information. This was shown in a separate column to the values that were obtained at the time of a cruise, this is labelled

Net Type	Acronym
Rectangular Midwater Trawl net	RMT
International Young Gadoid Pelagic Trawl net	IYGPT
International Young Gadoid Pelagic Trawl with Mid-water Open Close net	IYGPT with MIDOC net
Matsuda-Oozeki-Hu Trawl net	MOHT
Isaacs-Kid Midwater Trawl net	IKMT

Table 1. Data in *Myctobase* were collected with the following net types which are commonly referred to by their associated acronyms.

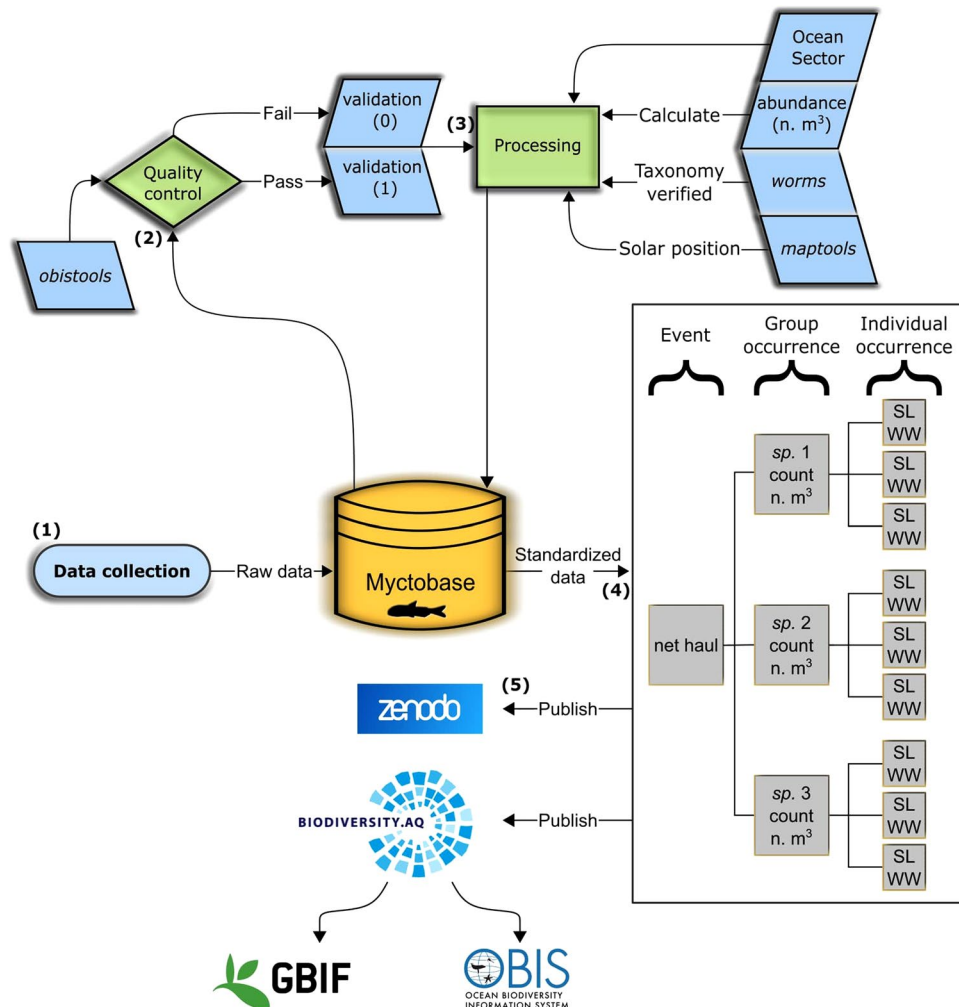


Fig. 2 Schematic illustrating the quality control and processing steps leading to the standardised data output of *Myctobase*. The standardised data are made available through Zenodo, the Antarctic Biodiversity Portal, the Global Biodiversity Information Facility (GBIF) and the Ocean Biogeographic Information System (OBIS). Abbreviations under ‘Individual occurrence’ are standard length (SL) and wet weight (WW).

volumeFiltered2 in the database. Detailed information on the methodologies utilised to obtain volume filtered, including our retrospective calculations, for each cruise can be found in Online-only Table 2.

The solar position and day/night information was added for each observation based on the date, time and the start latitude and longitude of each net haul using the R package, *maptools*⁵¹ (Fig. 2). Solar position is the angle of the sun in relation to the horizon (0°), thus dawn was defined as a solar position of -12° to 12° before midday, and dusk was defined as a solar position of 12° to -12° after midday. Day was defined as the period after dawn and before dusk and night was defined as the period after dusk and before dawn. We added the zone (eg. Northern, Subantarctic and Antarctic) in which net hauls were undertaken using the definitions in⁵². The sector in which the net haul was undertaken was additionally added following definitions in⁵³, which define four major sectors: Atlantic, Indian, West Pacific and East Pacific (Fig. 2).

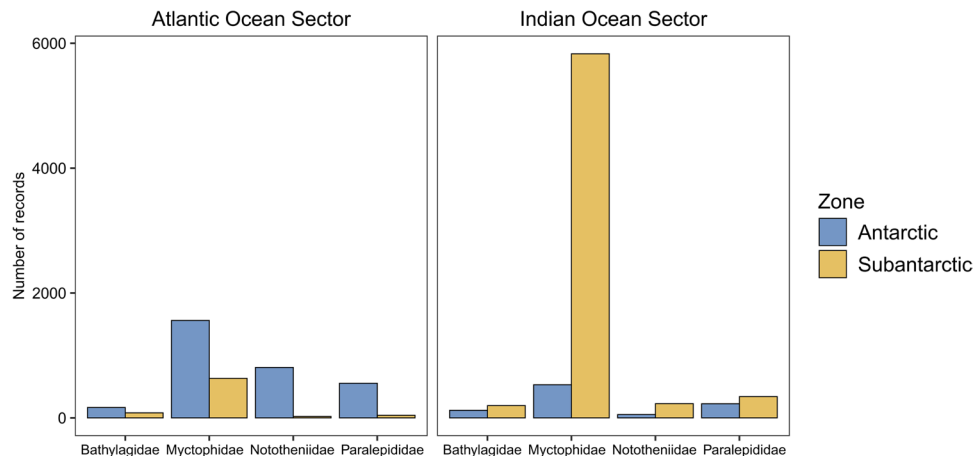


Fig. 3 The number of records of occurrence for the four most abundant fish families in the groupOccurrence data record of *Myctobase*. Data are divided into Ocean Sectors (Atlantic and Indian) and zones (Antarctic and Subantarctic) of the Southern Ocean. Zones of the Southern Ocean are defined by⁵².

Data Records

The dataset is comprised of three comma-separated files which are freely available at Zenodo⁵⁴. Filenames adhere as closely as possible to the naming convention set out by the Darwin Core Standard⁴⁸. The first file (event.csv) describes the survey methodology. Each row has its own unique event ID, which consists of the institute, cruise, event number (as recorded in the voyage logbook) and the net number (institute_cruise_event_net). An event ID represents the sampling event or net haul and contains the details of the event including date, time, position (latitude, longitude, and depth), sampling protocol, net type, net mesh size, tow speed, volume filtered and haul type (station, routine, target, test, surface or failed). The second file (groupOccurrence.csv) contains the catch data linked to the survey methodology by an event ID. Each row has its own unique occurrence ID, which is the event ID and aphia ID (retrieved from WoRMS) combined (eventID_aphiaID). An occurrence ID contains taxonomic information (eg. phylum, class, order, and family), the number of individuals (n) and estimated abundance (n_m⁻³) for the associated sampling event. The final file (individualOccurrence.csv) contains measurements of individuals. Each row contains the event and occurrence ID, which links each measurement to the first and second file, and a 'catalogNumber', linking the data to the original dataset for traceability. Rows also contain taxonomic information, standard length (mm), weight (g), life stage and reproductive maturity of the individual (following definitions in⁵⁵) and sex, where available. Additional notes on the preservation method and whether the specimen was measured before or after preservation are included. The presence of NA values in *Myctobase* are indicative of missing data. See associated metadata record for definitions and units for each variable (Supplementary Table 1). Citations for all datasets held in *Myctobase* can be found in Online-only Table 1.

Spatial and temporal coverage. *Myctobase* currently holds 4780 net hauls from 3775 sampling stations and 72 different research cruises from across the Southern Ocean. There are currently 17,491 occurrence records in *Myctobase*.

The highest concentration of data points is within the Atlantic Sector (45% of net hauls) spanning from the Antarctic continent to the Polar Front at 50° S (Fig. 1). The East and West Pacific Sector stands as a major gap in the spatial coverage of *Myctobase* (21% and 1% of net hauls in the East and West, respectively).

Vertically, data extend from the surface down to a maximum depth of 2000 m. More than half of the trawls took place in the epipelagic layer at 0–200 m (n = 3713). Data from the mesopelagic (200–1000 m) and bathypelagic (>1000 m) layers were also recorded from stratified oblique trawls.

Data span from 1991–2016 and 2019, with no data for 2017 and 2018 (Online-only Table 1). The year 2006 contains the highest number of net hauls (11%). Cumulatively, the datasets within *Myctobase* cover data for every month (NB. not every month of every year), with 76% of net hauls occurring in the summer months (November to March).

Records of occurrence. The four most abundant fish families in *Myctobase* are Paralepididae, Nototheniidae, Myctophidae and Bathylagidae. Myctophidae make up the highest number of records in *Myctobase* for both the Atlantic and Indian sectors (Fig. 3). In the Indian Ocean sector, the highest number of records for Myctophidae are in the Subantarctic zone. Conversely, in the Atlantic Ocean sector the highest number of records for Myctophidae are in the Antarctic zone. There are a similar number of records for Bathylagidae, Nototheniidae and Paralepididae in the Indian sector for both the Antarctic and Subantarctic zones. In the Atlantic Ocean sector, the highest number of records for Bathylagidae, Nototheniidae and Paralepididae are in the Antarctic zone.

Technical Validation

The final dataset was subject to quality control and validation processes. Rather than removing ambiguous or incomplete records altogether, two extra columns were added to the event information (event.csv). The first column labelled *validation* is an index used to indicate whether the record passed quality control measures (0 = fail, 1 = pass). The second column labelled *validationDescription* details the reasons for a failure. Indexing the data in this way prevents the loss of potentially valuable data. For example, some events are missing latitude and longitude information, however a more general sampling location is often available and may be enough for regional analyses (Fig. 2).

The R package *obistools*⁵⁶ was used to validate sampling locations and dates of events. Sampling locations on land or at depth values higher than the bathymetry raster used in *obistools* did not pass the quality control step and thus were given a '0' in the *validation* column. Further, records missing latitude, longitude and date/time data were similarly identified. Records where technical difficulties were recorded, such as net failing to open or close were also allocated a '0' in the *validation* column (Fig. 2).

Abundance was standardised to number of individuals per m³ (n_m³) for all datasets. In some instances, abundance could not be calculated due to missing count or volume filtered data. Additionally, abundances were not calculated for *US AMLR Program* as this data were considered bycatch and not the sole focus of the sampling program. Subsequently, this data is important for documenting the species observed and overall distribution in the region of the Antarctic Peninsula. Important information regarding the use of abundance values is available in 'Usage Notes'.

Although the purpose of the dataset is to document the occurrences and relevant metadata of mesopelagic fish species, we also retained records of cephalopods. Occurrence records of cephalopods alongside those of fish were included in some of the datasets contributed to *Myctobase*. Cephalopods are similarly important species of the open-ocean pelagic community and are a key group supporting the flow of energy from primary producers to higher order predators in Southern Ocean food webs^{57–59}. Data on squid are limited due to their low catchability with scientific nets⁵⁹. As such, we retained records to maximise availability of valuable data. The data are structured to enable users to easily filter out cephalopod occurrences if they are not of interest.

Taxonomic fields within the final dataset were checked for spelling errors and to verify the usage of valid/accepted names according to WoRMS. Where appropriate, records were corrected. All original data records were archived for future data checking and validation. Users can provide feedback for any data record to the corresponding authors.

Usage Notes

Myctobase will enhance research capacity by facilitating international effort toward observing and modelling mesopelagic fish taxa. The data held in *Myctobase* are suitable for a number of applications, for example, investigating the biophysical determinants shaping patterns of occurrence and biodiversity. Further examples of data use can be found in the citations listed in Online-only Table 1. Data can be analysed using a variety of statistical software such as R⁴⁹ or Matlab⁶⁰.

Myctobase contains data from 72 different research cruises across the Southern Ocean. Each cruise employed a specific sampling strategy and provided varying levels of detail on methodology and samples collected. The different sampling methodologies used between datasets necessitates caution when comparing abundance estimates, as well as length and weight measurements across research cruises. The biases associated with net type and mesh size of the net have previously been described^{8,21}, and are likely to influence the final output of analyses that use data from different projects. For example, the IYGPT has a larger mesh size at the front of the net limiting the size of fish that may be caught due to smaller fish escaping through the mesh. The minimum size limit has been suggested to be approximately 25–35 mm⁶¹. This suggests that nets with larger mesh sizes may not be appropriate for calculating densities of species that can escape the net. However this is a complex issue which requires further study as the net will also exert a herding effect on fish³². Biases should be considered and acknowledged. *Myctobase* provides detailed information for each sampling event to ensure that researchers can account for these differences in their analyses or to enable comparison of data using similar methods.

Further, abundance values should be treated as relative rather than absolute values due to the patchy distribution of species and the limited spatial and temporal coverage of sampling creating uncertainty around estimates^{4,8}. Data are best used to demonstrate the community composition, distribution, and occurrence of fish species, life history and relative abundance within the Southern Ocean. For example, there are a multitude of available modelling techniques which may be used to predict a species geographic distribution in relation to environmental variables^{62,63}. Furthermore, this information provides the data necessary for ground-truthing acoustic data⁶⁴.

We anticipate that *Myctobase* will continue to grow into a fully circumpolar database with continued collaboration from across the scientific marine community. We invite researchers to contact the corresponding author(s) to contribute data (particularly from under-represented regions, taxa, and data types such as abundance data), including data from unpublished research which can be embargoed until publication.

Terms of use. The database is released under a CC-BY license. Users are encouraged to formally cite the data record used according to the standards and format of the journal in which they are published.

Code availability

We used freely available code from the following packages: *maptools*⁵¹, *obistools*⁵⁶ and *WoRMS*⁵⁰

Received: 5 November 2021; Accepted: 28 June 2022;

Published online: 13 July 2022

References

- Webb, T. J., vanden Berghe, E. & O'Dor, R. Biodiversity's big wet secret: The global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. *PLoS ONE* **5**, <https://doi.org/10.1371/journal.pone.0010223> (2010).
- Drazen, J. C. & Sutton, T. T. Dining in the Deep: The Feeding Ecology of Deep-Sea Fishes. *Annual Review of Marine Science* **9**, 337–366, <https://doi.org/10.1146/annurev-marine-010816-060543> (2017).
- Brierley, A. S. Diel vertical migration. *Current Biology* **24**, R1074–R1076, <https://doi.org/10.1016/j.cub.2014.08.054> (2014).
- Irigoin, X. *et al.* Large mesopelagic fishes biomass and trophic efficiency in the open ocean. *Nature Communications* **5**, 10, <https://doi.org/10.1038/ncomms4271> (2014).
- Anderson, T. R. *et al.* Quantifying carbon fluxes from primary production to mesopelagic fish using a simple food web model. *ICES Journal of Marine Science* **76**, 690–701, <https://doi.org/10.1093/icesjms/fsx234> (2018).
- Saba, G. K. *et al.* Toward a better understanding of fish-based contribution to ocean carbon flux. *Limnology and Oceanography* **66**, 1639–1664, <https://doi.org/10.1002/lno.11709> (2021).
- Koslow, J. A., Kloser, R. J. & Williams, A. Pelagic biomass and community structure over the mid-continental slope off southeastern Australia based upon acoustic and midwater trawl sampling. *Marine Ecology Progress Series* **146**, 21–35, <https://doi.org/10.3354/meps146021> (1997).
- Kaartvedt, S., Staby, A. & Aksnes, D. L. Efficient trawl avoidance by mesopelagic fishes causes large underestimation of their biomass. *Marine Ecology Progress Series* **456**, 1–6, <https://doi.org/10.3354/meps09785> (2012).
- Lehodey, P., Murtugudde, R. & Senina, I. Bridging the gap from ocean models to population dynamics of large marine predators: A model of mid-trophic functional groups. *Progress in Oceanography* **84**, 69–84, <https://doi.org/10.1016/j.pocean.2009.09.008> (2010).
- Van de Putte, A., Flores, H., Volckaert, F. & van Franeker, J. A. Energy content of Antarctic mesopelagic fishes: Implications for the marine food web. *Polar Biology* **29**, 1045–1051, <https://doi.org/10.1007/s00300-006-0148-z> (2006).
- Stowasser, G. *et al.* Food web dynamics in the Scotia Sea in summer: A stable isotope study. *Deep-Sea Research Part II-Topical Studies in Oceanography* **59**, 208–221, <https://doi.org/10.1016/j.dsr2.2011.08.004> (2012).
- McCormack, S. A. *et al.* Decades of dietary data demonstrate regional food web structures in the Southern Ocean. *Ecology and Evolution* **11**, 227–241, <https://doi.org/10.1002/ece3.7017> (2021).
- Griffiths, S. P., Olson, R. J. & Watters, G. M. Complex wasp-waist regulation of pelagic ecosystems in the Pacific Ocean. *Reviews in Fish Biology and Fisheries* **23**, 459–475, <https://doi.org/10.1007/s1160-012-9301-7> (2013).
- Saunders, R. A., Hill, S. L., Tarling, G. A. & Murphy, E. J. Myctophid Fish (Family Myctophidae) Are Central Consumers in the Food Web of the Scotia Sea (Southern Ocean). *Frontiers in Marine Science* **6**, <https://doi.org/10.3389/fmars.2019.00530> (2019).
- Dornan, T., Fielding, S., Saunders, R. A. & Genner, M. J. Swimbladder morphology masks Southern Ocean mesopelagic fish biomass. *Proceedings of the Royal Society B-Biological Sciences* **286**, 8, <https://doi.org/10.1098/rspb.2019.0353> (2019).
- Freer, J. J., Tarling, G. A., Collins, M. A., Partridge, J. C. & Genner, M. J. Predicting future distributions of lanternfish, a significant ecological resource within the Southern Ocean. *Diversity and Distributions* **25**, 1259–1272, <https://doi.org/10.1111/ddi.12934> (2019).
- Hidalgo, M. & Browman, H. I. Developing the knowledge base needed to sustainably manage mesopelagic resources Introduction. *ICES Journal of Marine Science* **76**, 609–615, <https://doi.org/10.1093/icesjms/fsz067> (2019).
- Proud, R. *et al.* From siphonophores to deep scattering layers: Uncertainty ranges for the estimation of global mesopelagic fish biomass. *ICES Journal of Marine Science* **76**, 718–733, <https://doi.org/10.1093/icesjms/fsy037> (2019).
- Caccavo, J. A. *et al.* Productivity and Change in Fish and Squid in the Southern Ocean. *Frontiers in Ecology and Evolution* **9**, <https://doi.org/10.3389/fevo.2021.624918> (2021).
- Davison, P., Lara-Lopez, A. & Anthony Koslow, J. Mesopelagic fish biomass in the southern California current ecosystem. *Deep-Sea Research Part II: Topical Studies in Oceanography* **112**, 129–142, <https://doi.org/10.1016/j.dsr2.2014.10.007> (2015).
- Pakhomov, E. & Yamamura, O. Report of the Advisory Panel on Micronekton Sampling Inter-calibration Experiment. *Tech. Rep., PICES* (2010).
- Cheung, W. W. L. *et al.* Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries* **10**, 235–251, <https://doi.org/10.1111/j.1467-2979.2008.00315.x> (2009).
- Saunders, R. A. & Tarling, G. A. Southern Ocean Mesopelagic Fish Comply with Bergmann's Rule. *American Naturalist* **191**, 343–351, <https://doi.org/10.1086/695767> (2018).
- Proud, R., Cox, M. J. & Brierley, A. S. Biogeography of the Global Ocean's Mesopelagic Zone. *Current Biology* **27**, 113–119, <https://doi.org/10.1016/j.cub.2016.11.003> (2017).
- Robison, B. H. Conservation of Deep Pelagic Biodiversity. *Conservation Biology* **23**, 847–858, <https://doi.org/10.1111/j.1523-1739.2009.01219.x> (2009).
- Constable, A. J. *et al.* Developing priority variables (“ecosystem Essential Ocean Variables” - eEOVs) for observing dynamics and change in Southern Ocean ecosystems. *Journal of Marine Systems* **161**, 26–41, <https://doi.org/10.1016/j.jmarsys.2016.05.003> (2016).
- St John, M. A. *et al.* A Dark Hole in Our Understanding of Marine Ecosystems and Their Services: Perspectives from the Mesopelagic Community. *Frontiers in Marine Science* **3**, 6, <https://doi.org/10.3389/fmars.2016.00031> (2016).
- Newman, L. *et al.* Delivering Sustained, Coordinated, and Integrated Observations of the Southern Ocean for Global Impact. *Frontiers in Marine Science* **6**, <https://doi.org/10.3389/fmars.2019.00433> (2019).
- Costello, M. J. & Vanden Berghe, E. ‘Ocean biodiversity informatics’: a new era in marine biology research and management. *Marine Ecology Progress Series* **316**, 203–214, <https://doi.org/10.3354/meps316203> (2006).
- Van de Putte, A. *et al.* From data to marine ecosystem assessments of the Southern Ocean, achievements, challenges, and lessons for the future. *Frontiers in Marine Science* **8**, <https://doi.org/10.3389/fmars.2021.637063> (2021).
- Duhamel, G. *et al.* Biogeographic Patterns of Fish. In *Biogeographic Atlas of the Southern Ocean*, 328–362 (Scientific Committee of Antarctic Research, Cambridge, UK, 2014).
- Piatkowski, U., Rodhouse, P. G., White, M. G., Bone, D. G. & Symon, C. Nekton community of the Scotia Sea as sampled by the RMT-25 during the austral summer. *Marine Ecology Progress Series* **112**, 13–28, <https://doi.org/10.3354/meps112013> (1994).
- Collins, M. A. *et al.* Patterns in the distribution of myctophid fish in the northern Scotia Sea ecosystem. *Polar Biology* **31**, 837–851, <https://doi.org/10.1007/s00300-008-0423-2> (2008).
- Collins, M. A. *et al.* Latitudinal and bathymetric patterns in the distribution and abundance of mesopelagic fish in the Scotia Sea. *Deep-Sea Research Part II-Topical Studies in Oceanography* **59**, 189–198, <https://doi.org/10.1016/j.dsr2.2011.07.003> (2012).
- Loeb, V. J., Hofmann, E. E., Klinck, J. M., Holm-Hansen, O. & White, W. B. ENSO and variability of the Antarctic Peninsula pelagic marine ecosystem. *Antarctic Science* **21**, 135–148, <https://doi.org/10.1017/s0954102008001636> (2009).
- Reiss, C. S. *et al.* Overwinter habitat selection by Antarctic krill under varying sea-ice conditions: implications for top predators and fishery management. *Marine Ecology Progress Series* **568**, 1–16, <https://doi.org/10.3354/meps12099> (2017).
- Flores, H. *et al.* Distribution, abundance and ecological relevance of pelagic fishes in the Lazarev Sea, Southern Ocean. *Marine Ecology Progress Series* **367**, 271–282, <https://doi.org/10.3354/meps07530> (2008).
- Flores, H. *et al.* Seasonal changes in the vertical distribution and community structure of Antarctic macrozooplankton and micronekton. *Deep-Sea Research Part I-Oceanographic Research Papers* **84**, 127–141, <https://doi.org/10.1016/j.dsr.2013.11.001> (2014).
- Duhamel, G. The Pelagic Fish Community of the Polar Frontal Zone off the Kerguelen Islands. In *Fishes of Antarctica*, 63–74, https://doi.org/10.1007/978-88-470-2157-0_5 (Springer, Milano, 1998).

40. Duhamel, G., Koubbi, P. & Ravier, C. Day and night mesopelagic fish assemblages off the Kerguelen Islands (Southern Ocean). *Polar Biology* **23**, 106–112, <https://doi.org/10.1007/s003000050015> (2000).
41. Duhamel, G., Gasco, N. & Davaine, P. *Poissons des îles Kerguelen et Crozet: Guide régional de l'océan Austral* (Muséum national d'Histoire naturelle, Paris, 2005).
42. Trebilco, R. *et al.* Mesopelagic community structure on the southern Kerguelen Axis. In *The Kerguelen Plateau: Marine Ecosystem and Fisheries*, 49–54 (Australian Antarctic Division, Kingston, Tasmania, 2019).
43. Constable, A. J. & Swadling, K. M. Ecosystem drivers of food webs on the Kerguelen Axis of the Southern Ocean. *Deep-Sea Research Part II-Topical Studies in Oceanography* **174**, 6, <https://doi.org/10.1016/j.dsr2.2020.104790> (2020).
44. Van de Putte, A. P., Jackson, G. D., Pakhomov, E., Flores, H. & Volckaert, F. A. M. Distribution of squid and fish in the pelagic zone of the Cosmonaut Sea and Prydz Bay region during the BROKE-West campaign. *Deep-Sea Research Part II-Topical Studies in Oceanography* **57**, 956–967, <https://doi.org/10.1016/j.dsr2.2008.02.015> (2010).
45. Flynn, A. J. & Williams, A. Lanternfish (Pisces: Myctophidae) biomass distribution and oceanographic-topographic associations at Macquarie Island, Southern Ocean. *Marine and Freshwater Research* **63**, 251–263, <https://doi.org/10.1071/mf11163> (2012).
46. Sutton, C. A., Kloser, R. J. & Gershwin, L. A. Micronekton in southeastern Australian and the Southern Ocean; a collation of the biomass, abundance, biodiversity and distribution data from CSIRO's historical mesopelagic depth stratified new samples. CSIRO, Aust. <http://hdl.handle.net/102.100.100/365479?index=1> (2018).
47. Gon, O. & Heemstra, P. C. *Fishes of the Southern Ocean* (J.L.B. Smith Institute of Ichthyology, Grahamstown, South Africa, 1990).
48. Darwin Core Maintenance Group. List of Darwin Core terms (2021).
49. R Core Team. R: A language and environment for statistical computing (2021).
50. Holstein, J. worms: Retrieving Aphia Information from World Register of Marine Species (2018).
51. Bivand, R. *et al.* maptools: Tools for handling spatial objects. R package version 1.1-1 (2021).
52. Orsi, A. H., Whitworth, T. & Nowlin, W. D. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Research Part I-Oceanographic Research Papers* **42**, 641–673, [https://doi.org/10.1016/0967-0637\(95\)00021-w](https://doi.org/10.1016/0967-0637(95)00021-w) (1995).
53. Constable, A. J. *et al.* Climate change and Southern Ocean ecosystems I: how changes in physical habitats directly affect marine biota. *Global Change Biology* **20**, 3004–3025, <https://doi.org/10.1111/gcb.12623> (2014).
54. Woods, B. *et al.* Myctobase. Zenodo <https://doi.org/10.5281/zenodo.5590999> (2021).
55. Saunders, R. A., Collins, M. A., Stowasser, G. & Tarling, G. A. Southern Ocean mesopelagic fish communities in the Scotia Sea are sustained by mass immigration. *Marine Ecology Progress Series* **569**, 173–185, <https://doi.org/10.3354/meps12093> (2017).
56. Provoost, P. & Bosch, S. obistools: Tools for data enhancement and quality control (2021).
57. Murphy, E. J. *et al.* Understanding the structure and functioning of polar pelagic ecosystems to predict the impacts of change, <https://doi.org/10.1098/rspb.2016.1646> (2016).
58. McCormack, S. A., Melbourne-Thomas, J., Trebilco, R., Blanchard, J. L. & Constable, A. Alternative energy pathways in Southern Ocean food webs: Insights from a balanced model of Prydz Bay, Antarctica. *Deep-Sea Research Part II-Topical Studies in Oceanography* **174**, <https://doi.org/10.1016/j.dsr2.2019.07.001> (2020).
59. Rodhouse, P. G. K. Role of squid in the Southern Ocean pelagic ecosystem and the possible consequences of climate change. *Deep-Sea Research Part II-Topical Studies in Oceanography* **95**, 129–138, <https://doi.org/10.1016/j.dsr2.2012.07.001> (2013).
60. The MathWorks Inc., V. MATLAB (2019).
61. Potter, D. C., Lough, R. G., Perry, R. I. & Neilson, J. D. Comparison of the mocness and iygt pelagic samplers for the capture of 0-group cod (*Gadus morhua*) on Georges Bank. *ICES Journal of Marine Science* **46**, <https://doi.org/10.1093/icesjms/46.2.121> (1990).
62. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *Journal of Animal Ecology* **77**, 802–813, <https://doi.org/10.1111/j.1365-2656.2008.01390.x> (2008).
63. Oppel, S. *et al.* Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation* **156**, <https://doi.org/10.1016/j.biocon.2011.11.013> (2012).
64. McClatchie, S., Thorne, R. E., Grimes, P. & Hanchet, S. Ground truth and target identification for fisheries acoustics. *Fisheries Research* **47**, 173–191, [https://doi.org/10.1016/s0165-7836\(00\)00168-5](https://doi.org/10.1016/s0165-7836(00)00168-5) (2000).
65. Collins, M., Piatkowski, U. & Saunders, R. A. Distribution of mesopelagic fish in the Scotia Sea from RMT25 and pelagic trawls deployed from RRS James Clark Ross and RRS John Biscoe, UK Polar Data Centre <https://doi.org/10.5285/f4dfc0ee-4f61-47c5-a5a8-238e02ff2fdd> (2021).
66. Hoddell, R. J., Crossley, C., Hosie, G. & Williams, D. Fish and zooplankton from RMT-8 net hauls on the BROKE voyage. *Australian Antarctic Data Centre* <https://doi.org/10.4225/15/57BA97EA8A22D> (2016).
67. Constable, A., Williams, D. & Lamb, T. Heard Island and McDonald Islands (HIMI) Marine Ecosystem. *Australian Antarctic Data Centre* <https://doi.org/10.4225/15/5b31be45e8977> (2018).
68. Van de Putte, A. Fish catches from Rectangular Midwater Trawl - data collected from the BROKE-West voyage of the Aurora Australis, 2006. *Australian Antarctic Data Centre* <https://doi.org/10.4225/15/598d453109182> (2010).
69. Flynn, A. J., Kloser, R. J. & Sutton, C. Micronekton assemblages and bioregional setting of the Great Australian Bight: A temperate northern boundary current system. *Deep-Sea Research Part II: Topical Studies in Oceanography* **157**–158, <https://doi.org/10.1016/j.dsr2.2018.08.006> (2018).
70. Oozeki, Y., Hu, F., Tomatsu, C. & Kubota, H. Development of a new multiple sampling trawl with autonomous opening/closing net control system for sampling juvenile pelagic fish. *Deep-Sea Research Part I-Oceanographic Research Papers* **61**, <https://doi.org/10.1016/j.dsr.2011.12.001> (2012).

Acknowledgements

The authors would like to thank and acknowledge the contributions of all researchers and their institutions. We also thank Dave Connell and Petra Ten Hoopen for their assistance and support with data acquisition, Sophie Bestley for providing custom code for calculating day/night periods. The relevant project funders and acknowledgements for each dataset provided to *Myctobase* can be accessed through associated citations listed in Online-only Table 1. For those without citations, the authors acknowledge Institut Polaire Français Paul-Emile Victor (IPEV) for the support of the programme “IPEKER”. The Japanese survey was supported by JSPS KAKENHI Grant No. JP17H06319 to M. Moteki and Japanese Antarctic Research Expedition (No. AP-0939 to M. Moteki). This Publication was funded by the Belgian Science Policy Office (BELSPO, contract n°FR/36/AN1/AntaBIS) in the Framework of EU-Lifewatch.

Author contributions

B.W. and A.V. were responsible for the conceptualisation of the project and the creation of the database. R.T., A.W., G.D., H.F., M.M., P.P., C.R., R.A.S., C.S., M.H., and A.V. contributed data and provided advice on database design. Y.G. contributed to data preparation and quality control processes. B.W. wrote the manuscript. A.V., R.T. and A.W. contributed to the manuscript. All authors read, reviewed and approved the manuscript. All authors verified the data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01496-y>.

Correspondence and requests for materials should be addressed to B.W. or A.V.d.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022