*CORRESPONDENCE
Jennifer J. Freer
✉ jenfree@bas.ac.uk

Assessing key influences on the distribution and life-history of Arctic and boreal *Calanus*: are online databases up to the challenge?

Jennifer J. Freer* and Geraint A. Tarling

Ecosystems, British Antarctic Survey, Cambridge, United Kingdom

Despite the importance of calanoid copepods to healthy ecosystem functioning of the Arctic Ocean and Subarctic Seas, many aspects of their biogeography, particularly in winter months, remain unresolved. At the same time, online databases that digitize species distribution records are growing in popularity as a tool to investigate ecological patterns at macro scales. The value of such databases for *Calanus* research requires investigation - the long history of *Calanus* sampling holds promise for such databases, while conditions at high latitudes may impose limits through spatial and temporal biases. We collated records of three *Calanus* species (*C. finmarchicus*, *C. glacialis*, and *C. hyperboreus*) from the Ocean Biodiversity Information System (OBIS) and the Global Biodiversity Information Facility (GBIF) providing over 230,000 unique records spanning 150 years and over 100 individual datasets. After quality control and cleaning, the latitudinal and vertical distribution of occurrences were explored, as well as the completeness of informative metadata fields. *Calanus* sampling was found to be temporally and spatially biased towards surfacemost layers (<10m) in spring and summer. Only 3.5% of records had an average collection depth ≥400m, approximately half of these in months important for diapause. Just over 40% of records lacked associated information on sampling protocol while 11% of records lacked life-stage information. OBIS data contained fields for maximum and minimum collection depth and so were subset into discrete "shallow summer" and "deep winter" life cycle phases and matched to sea-ice and temperature conditions. 23% of OBIS records north of 66° latitude were located in regions of seasonal sea-ice presence and occurrences show species-specific thermal optima during the shallow summer period. The collection depth of *C. finmarchicus* was significantly different to *C. hyperboreus* during the deep winter. Overall, online databases contain a vast number of *Calanus* records but sampling biases should be acknowledged when they are used to investigate patterns of biogeography. We advocate efforts to integrate additional data sources within online portals. Particular gaps to be filled by existing or future collections are (i) widening the spatial extent of sampling during spring/summer months, (ii) increasing the frequency of sampling during winter, particularly at depths below 400m, and (iii) improving the quality, quantity and consistency of metadata reporting.

1 Introduction

Calanoid copepods of the genus *Calanus* are the dominant zooplankton in the sub-Arctic and Arctic seas (Conover and Huntley, 1991; Mauchline, 1998; Falk-Petersen et al., 2009). The Arctic *Calanus* complex comprises three species; *Calanus hyperboreus, C. glacialis*, and *C. finmarchicus*, which are thought to have core distributions in the Arctic continental slope, Arctic shelf and marginal seas, and Atlantic Ocean respectively (Falk-Petersen et al., 2009; Choquet et al., 2017), though sea-ice dynamics may be more important in shaping species core distributions in the Arctic Ocean (Ershova et al., 2021). From eggs to adult life-stages, they are vital in transferring energy between primary producers and higher trophic levels including stocks of herring, mackerel and cod, which in turn are prey for larger fish, seabirds, and mammals (Wassmann et al., 2006; Heath and Lough, 2007; Darnis et al., 2012; Langoy et al., 2012). In doing so, they contribute to the health and productivity of these polar ecosystems and to the goods and services they provide (Murphy et al., 2016).

Decades of basin-scale sampling campaigns and long term monitoring programmes have produced numerous observations of *Calanus* during spring and summer months. This research effort has revealed their complex physiological (Maps et al., 2014), behavioral (Sainmont et al., 2014; Banas et al., 2016), and morphological adaptations across latitudinal clines and varying seasonality. Key to their life cycle is an overwintering strategy known as diapause, where *Calanus* typically descend to depths of 400-2500m and enter a state of inactivity and reduced metabolism (Hirche, 1996; Maps et al., 2014). Overwintering behavior, onset and vertical distribution are dynamic, and thought to be influenced by multiple physical (e.g. bottom depth) and environmental (e.g. photoperiod and water mass depth) factors (Krumhansl et al., 2018). During diapause a store of high-energy wax ester lipids is used to fuel respiration, maturation, and reproduction (Jónasdóttir, 1999; Rey-Rassat et al., 2002; Saumweber and Durbin, 2006).

However, knowledge gaps in the overwintering ecology and distribution of *Calanus* have persisted due to the logistical and economic constraints of ship-borne sampling during the Arctic winter (Hop et al., 2021). Such data gaps have severe implications for the understanding of these key species because overwintering conditions account for a significant proportion of their yearly habitat - at least eight months for *C. hyperboreus* (Hirche, 1997). While modelling studies are able to utilize the limited empirical data to investigate overwintering strategies and processes, rigorous testing of their hypotheses will only be possible with increased effort to observe the timing, depth, behavior and physiology of diapausing copepods at finer spatial and temporal scales (Pierson et al., 2013).

Meanwhile, the last few decades have witnessed a transformation in the access of biodiversity data due to the concerted effort to digitize natural history collections and biological surveys on online databases and data portals (Soberon and Peterson, 2004). For example, the data available through the Global Biodiversity Information Facility (GBIF) reached 1.6 billion occurrence records in 2020, representing a growth of

1150% since 2012 (Heberling et al., 2021). The data held within these portals offer an unparalleled archive of life on Earth over the last few centuries, and have played a pivotal role in biodiversity research, most notably on fields relating to macroecology (Heberling et al., 2021). With a long and international history of *Calanus* research in northern seas (Barnard et al., 2004; Bandara et al., 2021b), the quantity of data and ease of access makes online databases an attractive source of information for undertaking novel *Calanus* studies, and has already been utilized to investigate long term shifts in the surface habitat of *C. finmarchicus* (Freer et al., 2022).

An in-depth analysis of the data available for these three Arctic and boreal *Calanus* species remains lacking. This is important to overcome, as while occurrences of Crustacea dominate data within databases covering northern European seas (Vandepitte et al., 2011; Ramirez et al., 2022), the quality of data remains equivocal for the few marine species investigated (Moudry and Devillers, 2020). Moreover, georeferencing uncertainty (Marcer et al., 2022), data quality (Maldonado et al., 2015) and uneven spatial and temporal coverage (Yesson et al., 2007; Ramirez et al., 2022) are inherent issues that must be understood before data are used or analyzed. If left unaccounted for, such issues have been shown to have unintended consequences when used to model species distributions (Beck et al., 2014; Marcer et al., 2022), predict species richness-environment relationships (Yang et al., 2013) or measure biodiversity change (Boakes et al., 2010).

As well as gaining a better understanding of available data, compiling current observations can also highlight gaps that can guide future sampling, as has been achieved for deep sea biodiversity data (Webb et al., 2010). Moreover, Kvile et al. (2019) compiled abundances of overwintering *C. hyperboreus* and *C. glacialis*, finding a lack of sampling in slope and central basin regions, as well as challenges associated with data collected *via* depth integrated net hauls. Assessing whether data stored on online portals also show these patterns will provide further evidence of under-sampled regions and depths that may help prioritize future research of the Arctic *Calanus* complex. Here we focus on GBIF and OBIS databases which are the two largest global scientific knowledge bases for marine species, although we acknowledge further datasets may exist that are yet to be integrated within these larger compilations. While processing these occurrence records, we also have an opportunity to add value to existing data. This includes associating the records to vertically and seasonally resolved environmental parameters, providing a unique insight into interspecific differences in these variables.

In this study we aimed to investigate the value and completeness of online databases for *Calanus* research in the high north. Using available *Calanus* occurrence records from OBIS and GBIF, we explore: i.) latitudinal, vertical and temporal biases in currently available data; ii.) regional temperature affinities of occurrences collected during typical "shallow summer" and "deep winter" life-cycle phases; and iii.) gaps in *Calanus* observations which should be targeted through the mobilization of existing data sources or future sampling efforts.

2 Materials and methods

2.1 Occurrence records and quality control

Georeferenced occurrence records of *Calanus finmarchicus, C. glacialis* and *C. hyperboreus* were compiled from the Oceanographic Biodiversity Information System (OBIS; www.obis.org) and the Global Biodiversity Information Facility (GBIF; www.gbif.org). To maintain consistency, all encounters were treated as presence only, i.e. no weighting was given to records with information on abundance, biomass or number of individuals and no absence records were included. Thus, the use of the term "records" hereafter relates to occurrence data that record the location and time of an encounter with a species, most commonly in the form of human observations or preserved specimens (Vandepitte et al., 2011). Citations for contributing datasets from OBIS and GBIF are available at See data availability statement to access citations of all contributing datasets from OBIS and GBIF. GBIF download information is also available at https://doi.org/10.15468/dl.fjggtx. All subsequent data processing and analysis were carried out in R v4.2.1 (R Core Team 2022).

The following cleaning and filtering steps were applied to OBIS and GBIF data separately. Records were excluded if they were erroneously located on land, in the Pacific sector or southern; hemisphere were not identified to species level; had missing identifiers for month or year of collection. Duplicate records were removed as defined by records that had the following identical identifiers: species, longitude, latitude, year, month, day, minimum and maximum depth (OBIS), average depth (GBIF), life-stage. Lastly, cleaned OBIS and GBIF data were combined into a single dataset and duplicate occurrences between databases were removed using the same identical identifiers listed above.

We analyzed how the remaining records were distributed by collection year, and quantified the number of records that contained metadata within the fields "life-stage" and "sampling protocol" (i.e. gear type used to collect the record). This was repeated for GBIF and OBIS data separately and combined.

2.2 Spatial and temporal sampling effort

To create a three dimensional picture of *Calanus* records, we incorporated information on each record's depth (average vertical depth of sample collection) and latitudinal position. The study region was then divided in to 1 degree latitudinal bins (30°-90°) and 50m depth bins (0-4000m). Using the combined OBIS+GBIF dataset, we populated this 60x80 matrix with the total number of records occurring within each matrix bin. Records that contained inaccurate depth information (i.e. average depth of collection was greater than bathymetric depth) were excluded and the matrix was log transformed before plotting.

Hereafter, analyses use records originating from the OBIS database only due to these data containing fields for maximum and minimum sampling depth, which are necessary to subset the data into discrete "shallow" and "deep" groupings.

To assess the proportion of records within different Arctic and boreal regions, records were assigned to one of eleven different geographic subareas based upon the International Hydrographic Organization (IHO) Sea Areas of the Arctic Ocean (IHO, 1953). For each species and season (Jan-Feb-Mar; Apr-May-Jun; Jul-Aug-Sep; Oct-Nov-Dec), the number of records within each region was calculated and visualized using spatial heatmaps. This was repeated for occurrences collected in the upper "shallow" water column (maximum net depth ≤ 200m), and occurrences collected from "deep" sampling strata (minimum net depth ≥ 400m).

2.3 Temperature and sea-ice associations

Occurrence records were matched to vertically resolved temperature data from the 2018 World Ocean Atlas (Locarnini et al., 2018). Raster grids of mean temperature were collated for nine vertical layers (0m, 50m, 100m, 200m, 400m, 600m, 800m, 1000m, 1500m), for each season (Jan-Feb-Mar; Apr-May-Jun; Jul-Aug-Sep; Oct-Nov-Dec) within two multi-decadal eras (1955-1984 and 1985-2017). These eras were chosen as they represent two different (cool and warm) oceanographic regimes in the northern North Atlantic, which are known to have affected zooplankton community dynamics and distribution (Beaugrand, 2009; Freer et al., 2022). For each occurrence record, temperature values were extracted from the most appropriate temperature grid in relation to the season, depth and year in which the record was collected. For example, a record collected in May 2002 with a mean net depth of 170m was assigned the environmental conditions from the 200m, Apr-May-Jun, 1985-2017 climatology. Similarly, occurrences were matched to seasonal sea-ice concentration data from the same two multi-decadal eras using gridded sea-ice concentration data from the National Snow and Ice Data Centre (Walsh et al., 2019). Any records collected outside of climatology years (i.e. before 1955) and containing inaccurate depth information (i.e. average collection depth greater than bathymetric depth) were excluded.

To explore how temperature associations changed with seasonal use of habitat, we further subset the "shallow" and "deep" groupings into "shallow summer" (representing records with a maximum depth ≤ 200m and between April-July only) and "deep winter" (representing records with a minimum depth ≥400m and between August and March only). These month/depth assignments follow the general seasonal vertical migration patterns known from *Calanus* observations (Hirche, 1996; Heath et al., 2000; Heath, 2003) and simulations (Bandara et al., 2021a), though we acknowledge that the timing of diapause entry and exit varies with latitude and local factors such as sea-ice extent, spring bloom timing etc. (Ji et al., 2012).

2.4 Overwinter depth associations

The bathymetric depth (i.e. total water column depth) and collection depth (i.e. average net depth) of records within the "deep winter" subset were compared between species using Kruskal-Wallis rank sum tests, after testing for normality (Shapiro-Wilk

test) and equal variance (Brown-Forsythe test). *Post-hoc* pairwise comparisons were analyzed using Wilcoxon rank sum tests with the default "holm" adjusted p-values.

3 Results

3.1 Occurrence records and quality control

The OBIS and GBIF databases contained 243,680 and 208,857 records respectively for the three *Calanus* species combined (both databases consisted of approx. 85% *C. finmarchicus*, 8% *C. glacialis*, 7% *C. hyperboreus*). After cleaning and quality control steps, 86% of OBIS records and 55% of GBIF records remained (See Table 1 for a summary of records removed per cleaning step per database). After merging these two cleaned datasets and removing duplications, 230,416 (71% of combined records) remained. The majority of these records were present in the OBIS database, with GBIF providing an additional 9% of unique records (Table 1).

From the combined GBIF+OBIS dataset, the greatest number of records overall was from 1995 (Figure 1A), but this pattern differed when analyzing databases separately (Figures S1–2A). GBIF records peaked in 1966 (*C. finmarchicus* = 1966, *C. hyperboreus* and *C. glacialis* = 1995) whilst OBIS records peaked in 1998 (*C. finmarchicus* = 1997, *C. glacialis* = 2001 and *C. hyperboreus* = 2005).

Just 6% of GBIF data had no associated information on sampling protocol, compared to 43% of OBIS data which is carried over into the combined dataset (Figure 1B). The most popular sampling method for all three species within GBIF was the continuous plankton recorder (86% of records, Figure S1B)

while WPII netting was the most popular method for *C. hyperboreus* within the OBIS data (Figure S2B). Overall, 11% of records in the combined dataset were not assigned to a life-stage category and the majority of records with life-stage information belonged to copepodite stages CIV-CVI (Figure 1C). This was similar across OBIS and GBIF databases however OBIS contained a greater spread of life-stage categories than GBIF (Figures S1–2C). Adult male and naupilus stages represented 2.5 and 1.5% of OBIS records compared to 0.5% and 0.1% of GBIF records respectively.
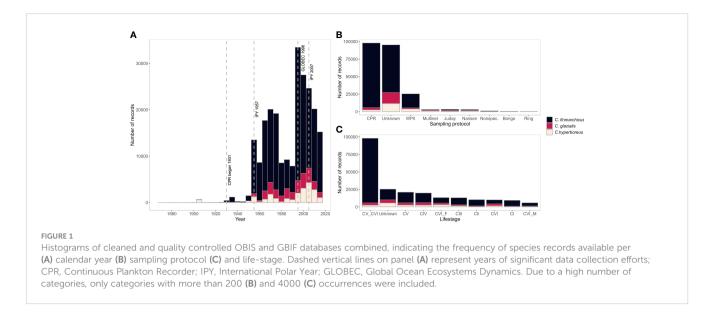
3.2 Spatial and temporal sampling effort

Across all three species, *Calanus* sampling was found to have strong vertical bias toward near-surface collections, with the number of records falling exponentially with depth (Figure 2A). In total, 68% of combined GBIF+OBIS records had an average collection depth <50m, 91% of records <200m and 60% of all records were collected between 40 and 60°N (Figure 2A). Records that were flagged as having depth inaccuracies were mostly located at latitudes >70°N (Figures 2B, C), further decreasing the number of records at northern latitudes.

Calanus sampling at shallow depths (≤200m) is widespread throughout the Arctic and boreal region (Figure 3). The majority of samples are located within the North Atlantic, Barents Sea and Norwegian Sea with variation across species and seasons. *Calanus* sampling at deeper depths (≥400m) is restricted to the basins of the Arctic and North Atlantic oceans and the Norwegian Sea, with greater variability within these basins across species and seasons (Figure 4). According to these data, *C. hyperboreus* has been sampled

TABLE 1 Summary table of (a) the number of occurrence records removed after each cleaning and quality control step within each database and (b) the number of occurrence records (absolute number and % of cleaned dataset) falling within defined categories for each database separately and combined.

(a) Data cleaning steps	OBIS	GBIF	GBIF+OBIS combined
Original download	243680	208857	
Missing species name	4321	0	
Wrong hemisphere	184	232	
Located on land	1496	2276	
Located in Pacific	331	102	
Duplicate record	28211	92391	
Total after cleaned + duplicates removed	209137 (86%)	113856 (55%)	230416 (71%)**
(b) Summary within cleaned dataset	OBIS	GBIF	GBIF+OBIS combined
Collection depth inaccuracy	6802 (3%)	5087 (4%)	11202 (5%)
Missing life-stage	15687 (8%)	10377 (9%)	25361 (11%)
Missing sampling protocol	88986 (43%)	7187 (6%)	95321 (41%)
Missing collection depth	1641 (1%)	1175 (1%)	2706 (1%)
Collection depth ≤ 200m*	185829 (89%)	106035 (93%)	209172 (91%)
Collection depth ≥ 400m*	1981 (1%)	3979 (3%)	8163 (4%)

**OBIS contribution = 209049 (91%), GBIF contribution = 21367 (9%); *GBIF and OBIS+GBIF combined = depth based on average depth field, OBIS = depth based on minimum and maximum depth fields. All values represent all three Calanus species combined.

FIGURE 1
Histograms of cleaned and quality controlled OBIS and GBIF databases combined, indicating the frequency of species records available per
(A) calendar year **(B)** sampling protocol **(C)** and life-stage. Dashed vertical lines on panel **(A)** represent years of significant data collection efforts;
CPR, Continuous Plankton Recorder; IPY, International Polar Year; GLOBEC, Global Ocean Ecosystems Dynamics. Due to a high number of
categories, only categories with more than 200 **(B)** and 4000 **(C)** occurrences were included.

within the Arctic Ocean all year round, whilst samples of *C. glacialis* are present January-September and *C. finmarchicus* is absent at these depths except between July and September (Figure 4).

Assessing *Calanus* records based on month of collection, records collected at shallow depths (≤200m) were most numerous between April and June for all species (Figure 5A). Records collected at deeper depths (≥400m) were most numerous in April for *C. finmarchicus* and in August for *C. glacialis* and *C. hyperboreus* January and February contain almost no records (Figure 5B).

3.3 Temperature and sea-ice associations

During the summer shallow period, records show species-specific associations with temperature (Figure 6A). Densities of *C. finmarchicus* occurrences peak at ~10°C, whilst *C. glacialis* and *C. hyperboreus* densities are bi-modal, peaking at ~2 and 12°C for *C. glacialis* and ~1 and 8°C for *C. hyperboreus* (Figure 6A). During the deep winter period, *C. glacialis* and *C. hyperboreus* records exhibit a clear peak in density between -1 and 1°C (Figure 6B). A small proportion of *C. finmarchicus* occurrences also associate with this temperature range, but higher densities are found at ~8°C (Figure 6B). Focusing on records collected in Arctic waters (those from 66°N and above), 23% of records were located in regions of seasonal sea-ice concentrations ≥ 15%. The majority of records are in areas with absent or low average sea ice concentrations (mean sea ice concentration across all species records = 14%; Figure 6C).

3.4 Overwinter depth associations

During the deep winter period, sampling was carried out at bathymetries between 600 and 4000m (Figure 7A). The bathymetric depth of sampling locations differed between species (Kruskal-Wallis: Chi^2 = 7.05, df = 2, p = 0.03), however *post-hoc* analysis found only marginal significance between *C. finmarchicus* and *C. hyperboreus* (p = 0.05). Collection depth was also significantly
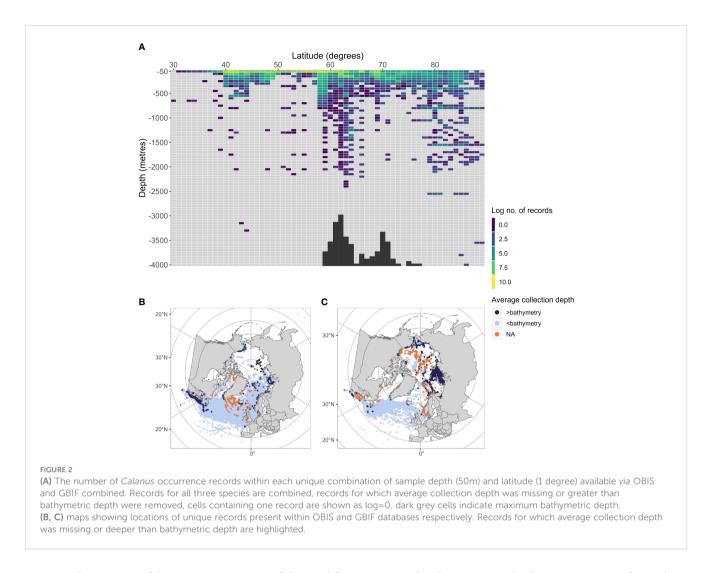
different between species (Kruskal-Wallis: Chi^2 = 245.9, df = 2, p = <0.001; Figure 7B), with *post-hoc* analysis indicating that the pairwise comparison between *C. finmarchicus* (median = 481.5 m) and *C. hyperboreus* (median = 1407.2 m) was significant (p = <0.001). All other pairwise comparisons were insignificant.

4 Discussion

Calanus sampling efforts have historically favored shallow summer events, potentially resulting in spatial, temporal and vertical sampling biases that impede a full understanding of their life history and climate change response. Combining the GBIF and OBIS catalogues, we have assessed the extent of these biases and will discuss the extent and suitability of these data for understanding of *Calanus* biogeography across both summer and winter environments.

4.1 Occurrence quality control

Manual data quality control and cleaning procedures are known to be an important part of using online biodiversity databases (Maldonado et al., 2015; Veiga et al., 2017). 14% of OBIS data and 45% of GBIF *Calanus* records had to be removed due to missing, duplicate or poor quality information, which is largely in line with the 35-55% reported for marine mammal data (Moudry and Devillers, 2020). Nevertheless, even records that have essential fields such as geographic coordinates, year and month of collection, often lack auxiliary metadata that can be vital for their interpretation or inclusion in subsequent analyses. As an example, just under half of OBIS occurrence records lacked information on sampling protocol. This precludes knowledge on other aspects of the sample such as tow direction and mesh size of the net, which has consequences for interpreting other fields such as sample effort or depth of occurrence. Another notable factor was the lack of consistent reporting within the metadata fields themselves. Life-stage, for example, had 135 different entries, which had to be

FIGURE 2
(A) The number of *Calanus* occurrence records within each unique combination of sample depth (50m) and latitude (1 degree) available *via* OBIS and GBIF combined. Records for all three species are combined, records for which average collection depth was missing or greater than bathymetric depth were removed, cells containing one record are shown as log=0, dark grey cells indicate maximum bathymetric depth.
(B, C) maps showing locations of unique records present within OBIS and GBIF databases respectively. Records for which average collection depth was missing or deeper than bathymetric depth are highlighted.

compressed in to 15 useful categories. Reporting of discrete life-stages is important as it would allow a more detailed understanding of the full overwintering stock of *Calanus*, and give new insights into the ecology of lesser known life-stages (Daase et al., 2018).
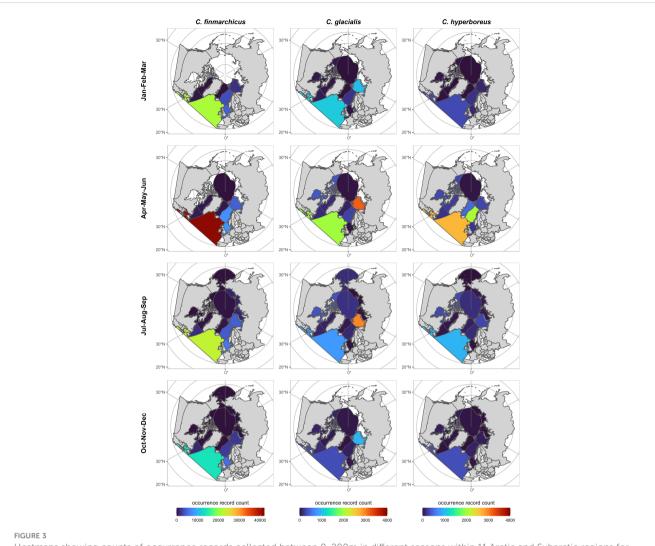
OBIS and GBIF datasets go through rigorous quality control procedures to meet standards for geographic location and taxonomy fields, however other fields are only recommended or optional to complete. Making fields on sampling protocol and life-stage mandatory, and/or creating a controlled vocabulary for these fields at the data submission stage would aid in the end user's data cleaning process and help standardize community-wide quality of data reporting (Vandepitte et al., 2011). Moreover, it would further align these data with findable, accessible, interoperable, and reusable (FAIR) principles (Ramirez et al., 2022).

4.2 Sampling biases

Throughout the 150 years of *Calanus* sampling captured by OBIS and GBIF records, we see the possible influence of regional and global sampling initiatives in the number of records being collected. Records appear in earnest alongside the launch of the CPR in 1931, and a step change in the numbers of all three species

correspond to the International Polar Year in 1957. C. *finmarchicus* records a peak at a similar time to the commencement of the GLOBEC programme in 1996, which focused on the North West Atlantic, while C. *hyperboreus* records peak in 2007, the same year as the most recent International Polar Year.

Our compilation reveals that sampling for *Calanus* has been most frequent at 40-60° north, in the upper 10m of the water column, after which records decline exponentially both spatially and with depth. This pattern is driven by the dominance of the Continuous Plankton Recorder (CPR) dataset, which has been surveying at a depth of ~7m across the North Sea and North Atlantic since 1931, providing an invaluable time series of basin-scale plankton dynamics (Batten et al., 2003; Richardson et al., 2006). As a consequence, data for the boreal species C. *finmarchicus* are much more numerous (approximately 8:1) than the other Arctic species. Given the value of CPR data to all aspects of plankton research and marine management (Brander et al., 2003), efforts to implement similar surveys at a wider range of latitudes, particularly in the high Arctic, would help resolve the core distribution patterns of *Calanus*, and how these are shaped by rapidly changing sea-ice conditions (Ershova et al., 2021; Tarling et al., 2021). For example, new northern CPR transects extending to a latitude of ~75°N begun over ten years ago and are providing temporally high-resolution
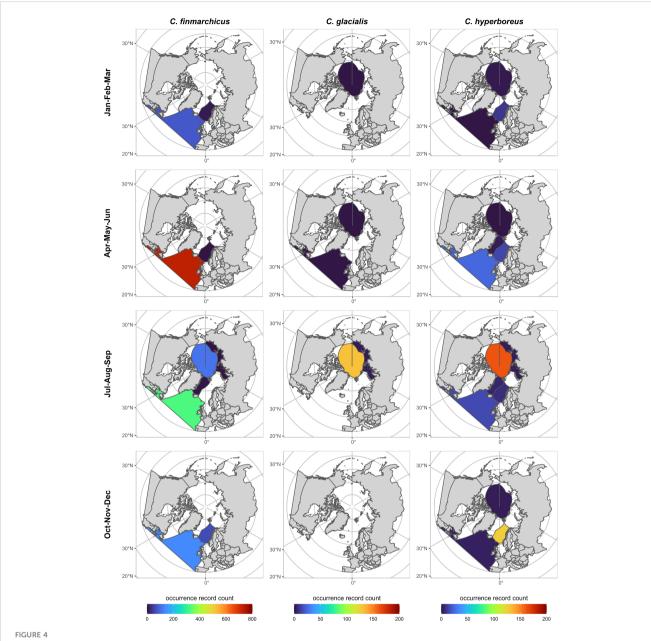
FIGURE 3
Heatmaps showing counts of occurrence records collected between 0-200m in different seasons within 11 Arctic and Subarctic regions for (left to right) *Calanus finmarchicus*, *C. glacialis* and *C. hyperboreus*. Note difference in scales between species. Records are from cleaned OBIS database only.

data on the *Calanus* community within the Nordic Seas (Strand et al., 2020).

Records with an average sampling greater than 400m accounted for only 3.5% of all observations in the merged and cleaned GBIF +OBIS dataset. Of those, even fewer (2%) were collected in months important for diapause. Analysis of deep records (which are not influenced by the CPR survey) reveals reduced spatial coverage of sampling than in the upper 200m. As well as being more restricted to areas with suitably deep bathymetry, another explanation for the reduced availability of records is that the data are an incomplete representation of sampling effort. The compilation of diapausing *Calanus* records by Kvile et al. (2019) included records of *C. glacialis* and *C. hyperboreus* located in Western Arctic regions that are missing from the OBIS data we present. Nevertheless, the deep pelagic is the most understudied and under-sampled sectors of the water column (Webb et al., 2010). Increasing the frequency of discrete, deep observations, particularly in winter months, should be a priority for future *Calanus* sampling and data digitization efforts.

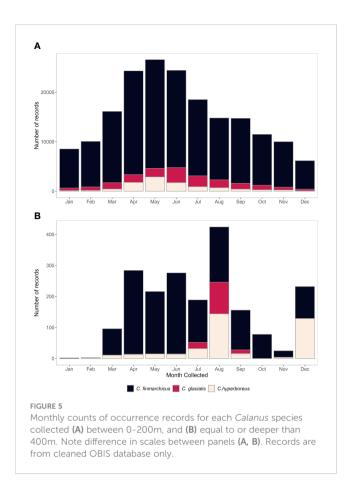## 4.3 Temperature and sea-ice associations

Matching each occurrence to seasonal and depth-specific temperature information, we show that shallow summer records fall within expected species-specific temperature ranges from previous observational and modeling studies (Albouy-Boyer et al., 2016; Strand et al., 2020). However, high densities of *C. glacialis* at ~12°C is higher than the previously suggested critical threshold of ~6°C for the presence of this species (Carstensen et al., 2012). Associating *Calanus* records to sea-ice conditions highlights the extent of our sampling bias towards ice-free conditions, with only ~5% of all OBIS records coming from regions with seasonal ice concentration ≥15%. While this pattern may well reflect lower likelihood of presence of the boreal species *C. finmarchicus*, for Arctic species known to be present in ice conditions, this pattern likely comes from the greater logistical constraints of sampling in ice-covered areas and greater sampling effort in ice-free months. Such geographical and environmental biases are vital to take in to account when using these data within distribution models to avoid

FIGURE 4
Heatmaps showing counts of occurrence records collected at depths equal to or deeper than 400m in different seasons within 11 Arctic and Subarctic regions for (left to right) *Calanus finmarchicus, C. glacialis* and *C. hyperboreus*. Note difference in scales between species. Records are from cleaned OBIS database only.

generating outputs that reflect habitat preference as well as sampling effort (Botella et al., 2020).

It is recognized that observations of overwintering *Calanus* remain incomplete and hinder our ability to understand this ecologically important life history strategy, or predict its vulnerability to changing conditions (Pierson et al., 2013; Baumgartner et al., 2017; Kvile et al., 2019; Bandara et al., 2021b). Our analysis finds that deep winter records fall within three temperature categories; a cold population (-1 to 2°C) containing all three species but dominated by *C. glacialis* and *C. hyperboreus*, an intermediate temperature population (~4 to 6°C) containing low densities of all three species, and a warm temperature population (~8°C) dominated by *C. finmarchicus*. These patterns align with previous diapause

temperature estimates of between -1 and +11°C (Kaartvedt, 1996). The separation of cold and warm diapausing populations may link to the presence of diapausers in cold Norwegian Sea Deep Water (<0°C) north of the Iceland-Faroe Ridge, and in warmer Atlantic Current Water (2 to 6°C) and Labrador Sea Water (2 to 4°C) to the south and west of the Iceland-Faroe Ridge (Heath et al., 2000). Optimal diapause temperature in the Northwest Atlantic was estimated to be <5°C (Saumweber and Durbin, 2006) and locations with colder diapausing temperatures have been found to have greater abundance and higher lipid:total weight ratios of *C. finmarchicus* (Heath et al., 2004), thus being important epicenters of overwintering stock and carbon sequestration (Jonasdottir et al., 2019). However, despite our finding that *C. glacialis* and *C. hyperboreus* dominate these cold overwintering

FIGURE 5
Monthly counts of occurrence records for each *Calanus* species collected **(A)** between 0-200m, and **(B)** equal to or deeper than 400m. Note difference in scales between panels **(A, B)**. Records are from cleaned OBIS database only.
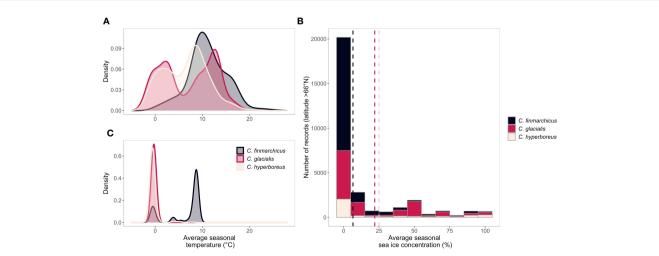
populations, their contribution to the lipid pump is largely unknown, and there remains high uncertainty in abundance estimates (Visser et al., 2017). The pan-Arctic assessment of winter depths and temperatures that we have compiled for these species will contribute to future efforts to fill this gap.

4.4 Limitations

When compiling data from multiple sources, regions and years, it is important to reflect on methodological limitations that may influence our findings. While OBIS and GBIF are the largest online portals for biodiversity information, this will not be a complete representation of all observations made. In fact, it is estimated that only 10% of biocollections worldwide have been digitized (Page et al., 2015), highlighting that biodiversity databases such as OBIS and GBIF are still in the initial stages of data compilation (Ball-Damerow et al., 2019). It is reassuring to find a high degree of overlap in the identity of *Calanus* records held within OBIS and GBIF databases but this overlap did not always extend to metadata fields; GBIF records contained more information on sampling protocol than OBIS, whilst the opposite was true for life-stage. The choice of database is therefore an important determinant of how readily accessible such information will be to researchers, and indicates that both databases can further improve how metadata fields are populated.

It is also possible that there are reporting errors within the data. While we are able to check some errors, such as reported collection depth being deeper than bathymetry, others may include misidentification of species as morphological criteria are known to be unreliable at distinguishing between *C. finmarchicus* and *C. glacialis* (Choquet et al., 2018). We also acknowledge limitations in our assignment of presence records to "shallow summer" or "deep winter" categories based upon their depth and month of collection. This simplistic approach allowed for broad-scale patterns to be assessed, however it will overlook some regional and intra- and inter-specific variability in timings of diapause entry and exit (Ji et al., 2012; Melle et al., 2014; Baumgartner et al., 2017).

Lastly, we note that our analyses based on presence-only data show where the species have been found, not necessarily where

FIGURE 6
The density of records associated to average seasonal temperatures for records collected **(A)** during the "shallow summer" period, defined by records collected between 0-200m in April-July only and **(C)** the "deep winter" period, defined by records collected at depths ≥ 400m in August-March only. **(B)** The number of records (all three *Calanus* species combined) collected above 66° north. Associated with different average seasonal sea ice concentrations. Dashed vertical lines represent the mean sea ice concentration association for each species. Records are from cleaned OBIS database and between years 1955-2017 only. Records for which average collection depth was missing or greater than bathymetric depth were removed.
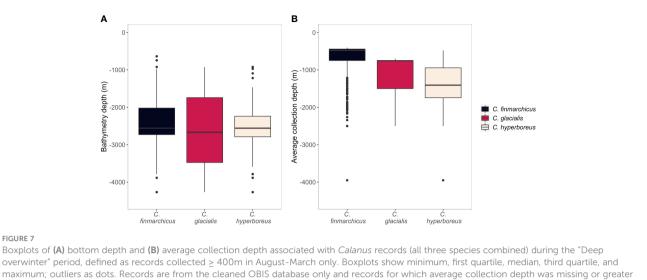
FIGURE 7
Boxplots of **(A)** bottom depth and **(B)** average collection depth associated with *Calanus* records (all three species combined) during the "Deep overwinter" period, defined as records collected ≥ 400m in August-March only. Boxplots show minimum, first quartile, median, third quartile, and maximum; outliers as dots. Records are from the cleaned OBIS database only and records for which average collection depth was missing or greater than bathymetric depth were removed.

sampling happened but the species was absent. Acknowledging the lack of absence data is important when attempting to disentangle patterns of sampling bias and the underlying habitat suitability of a species, both of which will have the effect of increasing or decreasing the probability of presence of a species in a given locality (Phillips et al., 2009). While the addition of absence data would allow biogeographic studies to more accurately define the realized distribution of a species, the reporting of absences in the marine and pelagic environment is rare. This is due to the difficulty in separating a true absence (an absence due to real habitat unsuitability or geographical hindrance) from an "undetected presence" which may be due to species mobility, net avoidance, net size, or other random sampling variation (Coro et al., 2016).

4.5 Future data mobilization efforts

Here we have highlighted the extent of sampling and subsequent digitization efforts for *Calanus* across the Arctic and sub-Arctic. Identifying these circumpolar-scale data needs is important to support cost-effective progress toward greater coverage of observations, which are vital for biodiversity assessments, habitat models and other ecological and conservation research. Based on our findings, we advocate efforts to i) widen the spatial extent of sampling during spring/summer months, ii) increase the frequency of sampling during autumn/winter, particularly at depths below 400m and within sea-ice regions, and iii) improve the quality and quantity of metadata reporting.

We are undergoing a revolution in our understanding of the Arctic ecosystem in the winter (Berge et al., 2015), shifting from the traditional view of bottom-up regulation to one of plasticity, opportunism and top-down control, increasing the motivation and importance of wintertime observations. At the same time, the

arrival of autonomous instruments in polar research, such as gliders, moorings, robotics and drones, have permitted a step-change in the capacity for year-round measurements of, for instance, ocean variables and sea-ice dynamics. Similar developments are a critical requirement to deliver 21st century polar ecology (Kennicutt et al., 2016) and could involve, for example, real-time monitoring of species using environmental DNA and *in-situ* plankton imaging (Vilgrain et al., 2021), or using argo floats to replicate the trajectory, depth, and conditions of overwintering habitat (Campbell, 2008). Harnessing new and powerful sampling protocols will become increasingly valuable to fill in the gaps in our knowledge of *Calanus* biogeography, across all seasons and environments.

In the meantime, short-term gains for these identified gaps could lie within existing data that are yet to be made accessible within online portals. Meyer et al. (2015) analyzed the completeness of digitally accessible information for over 21,000 terrestrial vertebrate species. The authors found that factors affecting inventory completeness of a region included distance to data contributors, locally available research funding and political commitment to data sharing. Ramirez et al. (2022) came to similar conclusions after assessing the completeness of marine records within European seas. Thus, solutions to fill data gaps in *Calanus* can come from efforts to incentivize researchers to share their existing data online, to streamline existing data entry mechanisms, and to prioritize the integration of non-western data sources (Meyer et al., 2015). Arctic and neighboring nations have a long history of international cooperation. Leveraging these partnerships to enhance data acquisition within underrepresented geographic areas, or to develop data sharing networks across stakeholders, institutions and individual scientists will lead to a collective benefit in improving the fitness of not only *Calanus* data, but all Arctic biodiversity.

Data availability statement

The annotated source code necessary to reproduce analyses and figures is available in the Supplementary Material. Full details of the datasets used and created within this study are available at https://doi.org/10.5285/A5CBFBE1-6FA6-44CF-96FE-DB92C32B69D0.

Author contributions

JJF: Conceptualization, data compilation, analyses, and writing. GAT: Conceptualization, supervision, writing, project administration and funding. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MD is currently organizing a Research Topic with the author GAT.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.908112/full#supplementary-material

References

Albouy-Boyer, S., Plourde, S., Pepin, P., Johnson, C. L., Lehoux, C., Galbraith, P. S., et al. (2016). Habitat modelling of key copepod species in the Northwest Atlantic ocean based on the Atlantic zone monitoring program. *J. Plankton Res.* 38, 589–603. doi: 10.1093/plankt/fbw020

Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., et al. (2019). Research applications of primary biodiversity databases in the digital age. *PloS One* 14(9):e0215794. doi: 10.1371/journal.pone.0215794

Banas, N. S., Moller, E. F., Nielsen, T. G., and Eisner, L. B. (2016). Copepod life strategy and population viability in response to prey timing and temperature: Testing a new model across latitude, time, and the size spectrum. *Front. Mar. Sci.* 3:225. doi: 10.3389/fmars.2016.00225

Bandara, K., Varpe, O., Maps, F., Ji, R. B., Eiane, K., and Tverberg, V. (2021a). Timing of *Calanus finmarchicus* diapause in stochastic environments. *Ecol. Model.* 460. doi: 10.1016/j.ecolmodel.2021.109739

Bandara, K., Varpe, O., Wijewardene, L., Tverberg, V., and Eiane, K. (2021b). Two hundred years of zooplankton vertical migration research. *Biol. Rev.* 96, 1547–1589. doi: 10.1111/brv.12715

Barnard, R., Batten, S., Beaugrand, G., Buckland, C., Conway, D. V. P., Edwards, M., et al. (2004). Continuous plankton records: Plankton atlas of the north Atlantic ocean, (1958-1999). II. biogeographical charts. *Mar. Ecol. Prog. Ser.* 11-75.

Batten, S. D., Clark, R., Flinkman, J., Hays, G. C., John, E., John, A. W. G., et al. (2003). CPR Sampling: the technical background, materials and methods, consistency and comparability. *Prog. Oceanogr.* 58, 193–215. doi: 10.1016/j.pocean.2003.08.004

Baumgartner, M. F., and Tarrant, A. M. (2017). The Physiology and Ecology of Diapause in Marine Copepods. *Annual review of marine science.* 9, 387–411. doi: 10.1146/annurev-marine-010816-060505

Beaugrand, G. (2009). Decadal changes in climate and ecosystems in the north Atlantic ocean and adjacent seas. *Deep-Sea Res. Part II: Topical Stud. Oceanogr.* 56, 656–673. doi: 10.1016/j.dsr2.2008.12.022

Beck, J., Boller, M., Erhardt, A., and Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inf.* 19, 10–15. doi: 10.1016/j.ecoinf.2013.11.002

Berge, J., Daase, M., Renaud, P. E., Ambrose, W. G., Darnis, G., Last, K. S., et al. (2015). Unexpected levels of biological activity during the polar night offer new perspectives on a warming Arctic. *Curr. Biol.* 25, 2555–2561. doi: 10.1016/j.cub.2015.08.024

Boakes, E. H., Mcgowan, P. J. K., Fuller, R. A., Ding, C. Q., Clark, N. E., O'connor, K., et al. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PloS Biol.* 8(6):e1000385. doi: 10.1371/journal.pbio.1000385

Botella, C., Joly, A., Monestiez, P., Bonnet, P., and Munoz, F. (2020). Bias in presence-only niche models related to sampling effort and species niches: Lessons for background point selection. *PloS One* 15(5):e0232078. doi: 10.1371/journal.pone.0232078

Brander, K. M., Dickson, R. R., and Edwards, M. (2003). Use of continuous plankton recorder information in support of marine management: Applications in fisheries, environmental protection, and in the study of ecosystem response to environmental change. *Prog. Oceanogr.* 58, 175–191. doi: 10.1016/j.pocean.2003.08.003

Campbell, R. W. (2008). Overwintering habitat of *Calanus finmarchicus* in the north Atlantic inferred from autonomous profiling floats. *Deep-Sea Res. Part I-Oceanogr. Res. Pap.* 55, 630–645. doi: 10.1016/j.dsr.2008.02.009

Carstensen, J., Weydmann, A., Olszewska, A., and Kwasniewski, S. (2012). Effects of environmental conditions on the biomass of *Calanus* spp. in the Nordic seas. *J. Plankton Res.* 34, 951–966. doi: 10.1093/plankt/fbs059

Choquet, M., Hatlebakk, M., Dhanasiri, A. K. S., Kosobokova, K., Smolina, I., Soreide, J. E., et al. (2017). Genetics redraws pelagic biogeography of *Calanus*. *Biol. Lett.* 13:20170588. doi: 10.1098/rsbl.2017.0588

Choquet, M., Kosobokova, K., Kwasniewski, S., Hatlebakk, M., Dhanasiri, A. K. S., Melle, W., et al. (2018). Can morphology reliably distinguish between the copepods *Calanus finmarchicus* and c-glacialis, or is DNA the only way? *Limnol. Oceanography-Methods* 16, 237–252. doi: 10.1002/lom3.10240

Conover, R. J., and Huntley, M. (1991). Copepods in ice-covered seas–distribution, adaptations to seasonally limited food, metabolism, growth patterns and life cycle strategies in polar seas. *J. Mar. Syst.* 2, 1–41. doi: 10.1016/0924-7963(91)90011-I

Coro, G., Magliozzi, C., Vanden Berghe, E., Bailly, N., Ellenbroek, A., and Pagano, P. (2016). Estimating absence locations of marine species from data of scientific surveys in OBIS. *Ecol. Model.* 323, 61–76. doi: 10.1016/j.ecolmodel.2015.12.008

Daase, M., Kosobokova, K., Last, K. S., Cohen, J. H., Choquet, M., Hatlebakk, M., et al. (2018). New insights into the biology of *Calanus* spp. (Copepoda) males in the Arctic. *Mar. Ecol. Prog. Ser.* 607, 53–69. doi: 10.3354/meps12788

Darnis, G., Robert, D., Pomerleau, C., Link, H., Archambault, P., Nelson, R. J., et al. (2012). Current state and trends in Canadian Arctic marine ecosystems: II. heterotrophic food web, pelagic-benthic coupling, and biodiversity. *Climatic Change* 115, 179–205. doi: 10.1007/s10584-012-0483-8

Ershova, E. A., Kosobokova, K. N., Banas, N. S., Ellingsen, I., Niehoff, B., Hildebrandt, N., et al. (2021). Sea Ice decline drives biogeographical shifts of key *Calanus* species in the central Arctic ocean. *Global Change Biol.* 27(10), 2128–2143 doi: 10.1111/gcb.15562

Falk-Petersen, S., Mayzaud, P., Kattner, G., and Sargent, J. R. (2009). Lipids and life strategy of Arctic *Calanus*. *Mar. Biol. Res.* 5, 18–39. doi: 10.1080/17451000802512267

Freer, J. J., Daase, M., and Tarling, G. A. (2022). Modelling the biogeographic boundary shift of *Calanus finmarchicus* reveals drivers of Arctic atlantification by subarctic zooplankton. *Global Change Biol.* 28(2), 429–440. doi: 10.1111/gcb.15937

Heath, M. R. (2003). The ascent migration of *Calanus finmarchicus* from overwintering depths in the faroe–Shetland channel. *Fisheries Oceanogr.* 8, 84–99. doi: 10.1046/j.1365-2419.1999.00013.x

Heath, M. R., Boyle, P. R., Gislason, A., Gurney, W. S. C., Hay, S. J., Head, E. J. H., et al. (2004). Comparative ecology of over-wintering *Calanus finmarchicus* in the northern north Atlantic, and implications for life-cycle patterns. *Ices J. Mar. Sci.* 61, 698–708. doi: 10.1016/j.icesjms.2004.03.013

Heath, M. R., Fraser, J. G., Gislason, A., Hay, S. J., Jonasdottir, S. H., and Richardson, K. (2000). Winter distribution of *Calanus finmarchicus* in the northeast Atlantic. *Ices J. Mar. Sci.* 57, 1628–1635. doi: 10.1006/jmsc.2000.0978

Heath, M. R., and Lough, R. G. (2007). A synthesis of large-scale patterns in the planktonic prey of larval and juvenile cod (*Gadus morhua*). *Fisheries Oceanogr.* 16, 169–185. doi: 10.1111/j.1365-2419.2006.00423.x

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., and Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci. United States America* 118 (6), e2018093118. doi: 10.1073/pnas.2018093118

Hirche, H. J. (1996). Diapause in the marine copepod, *Calanus finmarchicus* - a review. *Ophelia* 44, 129–143. doi: 10.1080/00785326.1995.10429843

Hirche, H. J. (1997). Life cycle of the copepod *Calanus hyperboreus* in the Greenland sea. *Mar. Biol.* 128, 607–618. doi: 10.1007/s002270050127

Hop, H., Wold, A., Meyer, A., Bailey, A., Hatlebakk, M., Kwasniewski, S., et al. (2021). Winter-spring development of the zooplankton community below Sea ice in the Arctic ocean. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.609480

IHO (1953). "Limits of oceans and seas," in *IHO special publication*, *23. 3 ed* (Monaco: International Hydrographic Organization).

Ji, R. B., Ashjian, C. J., Campbell, R. G., Chen, C. S., Gao, G. P., Davis, C. S., et al. (2012). Life history and biogeography of *Calanus* copepods in the Arctic ocean: An individual-based modeling study. *Prog. Oceanogr.* 96, 40–56. doi: 10.1016/j.pocean.2011.10.001

Jónasdóttir, S. H. (1999). Lipid content of *Calanus finmarchicus* during overwintering in the faroe–Shetland channel. *Fisheries Oceanogr.* 8, 61–72. doi: 10.1046/j.1365-2419.1999.00003.x

Jonasdottir, S. H., Wilson, R. J., Gislason, A., and Heath, M. R. (2019). Lipid content in overwintering *Calanus finmarchicus* across the subpolar Eastern north Atlantic ocean. *Limnol. Oceanogr.* 64, 2029–2043. doi: 10.1002/lno.11167

Kaartvedt, S. (1996). Habitat preference during overwintering and timing of seasonal vertical migration of *Calanus finmarchicus*. *Ophelia* 44, 145–156. doi: 10.1080/00785326.1995.10429844

Kennicutt, M. C., Kim, Y. D., Rogan-Finnemore, M., Anandakrishnan, S., Chown, S. L., Colwell, S., et al. (2016). Delivering 21st century Antarctic and Southern Ocean science. *Antarctic Sci.* 28, 407–423. doi: 10.1017/S0954102016000481

Krumhansl, K. A., Head, E. J. H., Pepin, P., Plourde, S., Record, N. R., Runge, J. A., et al. (2018). Environmental drivers of vertical distribution in diapausing *Calanus* copepods in the Northwest Atlantic. *Prog. Oceanogr.* 162, 202–222. doi: 10.1016/j.pocean.2018.02.018

Kvile, K. O., Ashjian, C., and Ji, R. B. (2019). Pan-Arctic depth distribution of diapausing *Calanus* copepods. *Biol. Bull.* 237, 76–89. doi: 10.1086/704694

Langoy, H., Nottestad, L., Skaret, G., Broms, C., and Ferno, A. (2012). Overlap in distribution and diets of Atlantic mackerel (*Scomber scombrus*), Norwegian spring-spawning herring (*Clupea harengus*) and blue whiting (*Micromesistius poutassou*) in the Norwegian Sea during late summer. *Mar. Biol. Res.* 8, 442–460. doi: 10.1080/17451000.2011.642803

Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H., et al. (2018). "World ocean atlas 2018, volume 1: Temperature," in *NOAA Atlas NESDIS 81*. Ed. A. Mishonov Silver Spring, MD: NOAA National Centers for Environmental Information, 52.

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Alban, J., et al. (2015). Estimating species diversity and distribution in the era of big data: To what extent can we trust public databases? *Global Ecol. Biogeogr.* 24, 973–984. doi: 10.1111/geb.12326

Maps, F., Record, N. R., and Pershing, A. J. (2014). A metabolic approach to dormancy in pelagic copepods helps explaining inter- and intra-specific variability in life-history strategies. *J. Plankton Res.* 36, 18–30. doi: 10.1093/plankt/fbt100

Marcer, A., Chapman, A. D., Wieczorek, J. R., Pico, F. X., Uribe, F., Waller, J., et al. (2022). Uncertainty matters: Ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. *Ecography*. 2022(9), e06025. doi: 10.1111/ecog.06025

Mauchline, J. (1998). *The biology of calanoid copepods* (San Diego, CA, USA: Academic Press).

Melle, W., Runge, J., Head, E., Plourde, S., Castellani, C., Licandro, P., et al. (2014). The north Atlantic ocean as habitat for *Calanus finmarchicus*: Environmental factors and life history traits. *Prog. Oceanogr.* 129, 244–284. doi: 10.1016/j.pocean.2014.04.026

Meyer, C., Kreft, H., Guralnick, R., and Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* 6, 8221. doi: 10.1038/ncomms9221

Moudry, V., and Devillers, R. (2020). Quality and usability challenges of global marine biodiversity databases: An example for marine mammal data. *Ecol. Inf.* 56, 101051. doi: 10.1016/j.ecoinf.2020.101051

Murphy, E. J., Cavanagh, R. D., Drinkwater, K. F., Grant, S. M., Heymans, J. J., Hofmann, E. E., et al. (2016). Understanding the structure and functioning of polar pelagic ecosystems to predict the impacts of change. *Proc. R. Soc. B-Biological Sci.* 283, 20161646. doi: 10.1098/rspb.2016.1646

Page, L. M., Macfadden, B. J., Fortes, J. A., Soltis, P. S., and Riccardi, G. (2015). Digitization of biodiversity collections reveals biggest data on biodiversity. *Bioscience* 65, 841–842. doi: 10.1093/biosci/biv104

Phillips, S., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., et al. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197. doi: 10.1890/07-2153.1

Pierson, J. J., Batchelder, H., Saumweber, W., Leising, A., and Runge, J. (2013). The impact of increasing temperatures on dormancy duration in *Calanus finmarchicus*. *J. Plankton Res.* 35, 504–512. doi: 10.1093/plankt/fbt022

Ramirez, F., Sbragaglia, V., Soacha, K., Coll, M., and Piera, J. (2022). Challenges for marine ecological assessments: Completeness of findable, accessible, interoperable, and reusable biodiversity data in European seas. *Front. Mar. Sci.* 8, 802235. doi: 10.3389/fmars.2021.802235

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

Rey-Rassat, C., Irigoien, X., Harris, R., and Carlotti, F. (2002). Energetic cost of gonad development in *Calanus finmarchicus* and c-helgolandicus. *Mar. Ecol. Prog. Ser.* 238, 301–306. doi: 10.3354/meps238301

Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., Lindley, J. A., Sims, D. W., et al. (2006). Using continuous plankton recorder data. *Prog. Oceanogr.* 68, 27–74. doi: 10.1016/j.pocean.2005.09.011

Sainmont, J., Andersen, K. H., Varpe, O., and Visser, A. W. (2014). Capital versus income breeding in a seasonal environment. *Am. Nat.* 184, 466–476. doi: 10.1086/677926

Saumweber, W. J., and Durbin, E. G. (2006). Estimating potential diapause duration in *Calanus finmarchicus*. *Deep-Sea Res. Part II: Topical Stud. Oceanogr.* 53, 2597–2617. doi: 10.1016/j.dsr2.2006.08.003

Soberon, J., and Peterson, A. T. (2004). Biodiversity informatics: managing and applying primary biodiversity data. *Philos. Trans. R. Soc. B-Biological Sci.* 359, 689–698. doi: 10.1098/rstb.2003.1439

Strand, E., Bagoien, E., Edwards, M., Broms, C., and Klevjer, T. (2020). Spatial distributions and seasonality of four *Calanus* species in the northeast Atlantic. *Prog. Oceanogr.* 185, 102344. doi: 10.1016/j.pocean.2020.102344

Tarling, G. A., Freer, J. J., Banas, N. S., Belcher, A., Blackwell, M., Castellani, C., et al. (2021). Has a key boreal *Calanus* copepod species become an Arctic resident? evidence and implications for Arctic food-webs. *Ambio* 51, 333–344. doi: 10.1007/s13280-021-01667-y

Vandepitte, L., Hernandez, F., Claus, S., Vanhoorne, B., De Hauwere, N., Deneudt, K., et al. (2011). Analysing the content of the European ocean biogeographic information system (EurOBIS): Available data, limitations, prospects and a look at the future. *Hydrobiologia* 667, 1–14. doi: 10.1007/s10750-011-0656-x

Veiga, A. K., Saraiva, A. M., Chapman, A. D., Morris, P. J., Gendreau, C., Schigel, D., et al. (2017). A conceptual framework for quality assessment and management of biodiversity data. *PloS One* 12 (6), e0178731. doi: 10.1371/journal.pone.0178731

Vilgrain, L., Maps, F., Picheral, M., Babin, M., Aubry, C., Irisson, J. O., et al (2021). Trait-based approach using *in situ* copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. *Limnol. Oceanogr.* 66(4), 1155–1167. doi: 10.1002/lno.11672

Visser, A. W., Gronning, J., and Jonasdottir, S. H. (2017). *Calanus hyperboreus* and the lipid pump. *Limnol. Oceanogr.* 62, 1155–1165. doi: 10.1002/lno.10492

Walsh, J. E., Chapman, W. L., Fetterer, F., and Stewart, J. S. (2019). *Gridded monthly Sea ice extent and concentration 1850 Onward, version 2* (Boulder, Colorado USA: NSIDC: National Snow and Ice Data Center).

Wassmann, P., Reigstad, M., Haug, T., Rudels, B., Carroll, M. L., Hop, H., et al. (2006). Food webs and carbon flux in the Barents Sea. *Prog. Oceanogr.* 71, 232–287. doi: 10.1016/j.pocean.2006.10.003

Webb, T. J., Vanden Berghe, E., and O'dor, R. (2010). Biodiversity's big wet secret: the global distribution of marine biological records reveals chronic under-exploration of the deep pelagic ocean. *PloS One* 5, e10223. doi: 10.1371/journal.pone.0010223

Yang, W. J., Ma, K. P., and Kreft, H. (2013). Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *J. Biogeogr.* 40, 1415–1426. doi: 10.1111/jbi.12108

Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., et al. (2007). How global is the global biodiversity information facility? *PloS One* 2(11), e1124. doi: 10.1371/journal.pone.0001124