Microbial plankton occurrence database in the North American Arctic region: synthesis of recent diversity of potentially toxic and harmful algae

Nicolas Schiffrine[1], Fatma Dhifallah[1], Kaven Dionne[1, a], Michel Poulin[2], Sylvie Lessard[1], André Rochon[1], Michel Gosselin[1]

[1]Institut des sciences de la mer de Rimouski (ISMER), Université du Québec à Rimouski, Rimouski, QC G5L 3A1, Canada
[2]Recherche et collections, Musée canadien de la nature, C.P. 3443, Succ. D, Ottawa, ON K1P 6P4, Canada
[a]present address: Réseau Québec Maritime, Université du Québec à Rimouski, Rimouski, QC G5L 3A1, Canada

Correspondence to: Nicolas Schiffrine (nicolas_schiffrine@uqar.ca)

Abstract. The Arctic Ocean is currently undergoing significant transformations due to climate change, leading to profound changes in its microbial planktonic communities. These communities consist of a wide range of organisms, including photoautotrophic prokaryotes and eukaryotes, as well as heterotrophic, phagotrophic, and mixotrophic protistan species. Here, for simplicity, we refer to these single-celled species as phytoplankton. Within this diversity, potentially toxic and/or harmful algal species (hereafter abbreviated as "HA") are of particular concern. These organisms have the potential to spread into Arctic waters, posing threats to both human and ecosystem health. Despite their importance, the spatial and temporal distribution of phytoplankton communities, including HA species, in the North American Arctic, remains poorly understood. To address this gap, we compiled and synthesized the largest possible body of data from different databases, individual published and unpublished datasets, and partitioned it into nine regions based on the Large Marine Ecosystem classification. Our dataset contains 385 800 individual georeferenced data points and 18 268 unique sampling events, revealing greater diversity than previously thought, with 1445 unique taxa. Heterokontophyta (which notably included diatoms) and Dinoflagellata were the most dominant phyla. Our results indicate distinct spatial patterns of diversity, with the highest diversity observed in Atlantic-influenced regions of the North American Arctic. For most of the HA species recorded in our database, no evidence was found for an increase in the northernmost latitude where HA species are observed over the years, meaning that there is no substantial spread of HA species into the North American part of the Arctic. Our study challenges the traditional view of the Arctic as being unsuitable for toxin-producing and harmful algae and highlights the importance of extensive and long-term sampling efforts to understand the region's biodiversity. Overall, our findings provide new insights into the spatial patterns and biodiversity of phytoplankton and other protists in the North American Arctic and have implications for understanding the ecological functioning and response of this region to ongoing climate change.

1 Introduction

The Arctic Ocean has become a focal point for climate change research due to its vulnerability to rapid and significant alterations in the environment (Meredith et al., 2019). As a result, the Arctic has been subjected to a growing number of scientific investigations aimed at understanding how these transformations affect the region's ecosystems, people, and global climate. In 2007, the Intergovernmental Panel on Climate Change recognized the Arctic as a region among the most vulnerable to climate change, highlighting the urgent need for further research in this area. Since then, a large body of scientific literature has emerged that explores the effects and implications of climate change on the Arctic marine ecosystem.

One of the most striking aspects of Arctic change is the dramatic acceleration of warming. Within the last 43 years, the Arctic has encountered warming at a pace almost four times swifter than the rest of the globe (Rantanen et al., 2022). Another notable aspect of these changes is the reduction in ice cover, both in extent and thickness (Kacimi and Kwok, 2022). The pace of this trend has accelerated markedly in recent decades, with Arctic sea-ice extent in September declining by 12.8% per decade from 1979 to 2018 (Hanna et al., 2021). This sharp decline in sea-ice extent, combined with an earlier onset of spring melt and delayed late-season freeze-up, also suggests that the Arctic may become ice-free during the summer months within the next few decades (Stroeve et al., 2012; Serreze and Stroeve, 2015). Moreover, these changes are influencing the formation of Arctic melt ponds, which have increased by 15% in June over the past decade due to climate change (Rösel et al., 2012). This increase alters solar radiation dynamics and sea-ice conditions, as melt ponds have a lower albedo, leading to further ice melt and reinforcing a positive ice-albedo feedback loop. As the Arctic transitions from multi-year ice to first-year ice (Kacimi and Kwok, 2022), with wider melt pool coverage and greater energy absorption, the consequences for the energy balance are far-reaching. Additionally, shifts in ice melt dynamics significantly affect the hydrographic structure of the water column, enhancing vertical stratification and reducing nutrient supply, particularly nitrate, to the surface layer (Yamamoto-Kawai et al., 2009; Carmack et al., 2016). More recently, Ardyna and Arrigo (2020) have pointed out that the Arctic Ocean is experiencing some of the fastest rates of ocean acidification, adding to the array of challenges faced by this rapidly changing region.

In the Arctic, changes in the physical and chemical properties of the water column have a profound impact on microbial plankton communities, which consist of a diverse array of microorganisms, each uniquely adapted to its ecological niche. Among these communities, phytoplankton is a prominent group that play a central role in the marine ecosystem. In a broader context defined in this paper, the term "phytoplankton" includes not only photoautotrophic but also heterotrophic, phagotrophic, and mixotrophic species. Phytoplankton is a cornerstone of Arctic ecosystems, where complex interactions between physico-chemical changes and biological factors drive significant shifts in their phenology, abundance, and species composition (e.g., Ardyna and Arrigo, 2020). The reduction in sea-ice cover has led in earlier phytoplankton blooms in some regions (Kahru et al., 2011), and increased open water periods, potentially leading to a second bloom in autumn (Ardyna et

64 al., 2014). In the Barents Sea, shifts in the current surface velocities have driven poleward intrusions of *Gephyrocapsa huxleyi*
65 (Lohmann) Reinhardt (previously called *Emiliania huxleyi* (Lohmann) Hay & Mohler; Bendif et al., 2019, 2023), a temperate
66 marine calcifying phytoplankton species (Neukermans et al., 2018; Oziel et al., 2020). Conversely, in the less productive
67 waters of the Canada Basin, increased freshwater inflow has led to a transition from nanophytoplankton/diatom communities
68 to picophytoplankton due to altered nutrient availability in the surface layer (Li et al., 2009). The observed changes in physico-
69 chemical conditions in the Arctic may also increase the potential risk of proliferation of potentially toxic and/or harmful algal
70 species (hereafter abbreviated as HA). Numerous HA species have already been detected in several Arctic regions (Bates et
71 al., 2020; McKenzie et al., 2020). Notably, various toxin-producing diatoms of the genus *Pseudo-nitzschia* Heragallo have
72 been documented in Iceland, Western Greenland, Baffin Bay, Barrow Strait, Beaufort Sea, Bering Strait, and subarctic regions
73 around Norway (Pućko et al., 2019; Bates et al., 2020). Similarly, toxic dinoflagellate species belonging to the genera
74 *Alexandrium* Halim and *Dinophysis* Ehrenberg have been detected (Okolodkov and Dodge, 1996; Pućko et al., 2019; Bates et
75 al., 2020; Bruhn et al., 2021; Dhifallah et al., 2021). Olsen et al. (2019) recently documented a red tide of the harmful
76 phototrophic ciliate *Mesodinium rubrum* (Lohmann) Leegard at the interface between ice and water in newly formed pack ice
77 north of Svalbard during early spring. Their findings suggest that ephemeral blooms of this species are increasingly probable
78 under the context of thinning Arctic sea ice. There is also evidence of a dominance shift towards the harmful prymnesiophyte
79 *Phaeocystis pouchetii* (Hariot) Lagerheim in the European Arctic sector due to the intensification of the "Atlantification"
80 phenomenon (Nöthig et al., 2015). Moreover, the increase in maritime traffic due to growing economic and tourism
81 development in the Arctic may elevate the risk of introducing non-native species, including HA species (Chan et al., 2019;
82 Dhifallah et al., 2021). These shifts could have significant implications for the future of Arctic marine ecosystems, impacting
83 the transfer of energy and organic matter through the pelagic food web.

84

85 The paucity of quantitative data on the taxonomic composition of Arctic phytoplankton species poses a challenge to
86 understanding their spatial and temporal variability. Additionally, the complex biogeography of the polar region exacerbates
87 this issue. One way to address these challenges and track potential changes in phytoplankton community structure, dynamics
88 and phenology is through the use of long-term datasets. The emergence of digital archives of biological data, such as the Global
89 Biodiversity Information Facility (GBIF; https://www.gbif.org/) and the Ocean Biogeographic Information System (OBIS;
90 https://www.obis.org/), has enabled the identification of significant patterns in the global distribution of phytoplankton
91 diversity as well as the occurrence, toxicity, and associated risks posed by HA species (Righetti et al., 2019; Hallegraeff et al.,
92 2021). Despite numerous studies that have utilized long-term datasets to monitor changes in Arctic phytoplankton diversity
93 and dominance, most of these studies have been conducted in specific regions of the Arctic (Blais et al., 2017; Freyria et al.,
94 2021). Furthermore, previous reports on Arctic phytoplankton diversity have not included essential information such as
95 geographic coordinates and dates, which limits the ability to assess potential changes in diversity and dominance (e.g., Poulin
96 et al., 2011). To date, there has been no effort to combine data from various sources, such as OBIS, GBIF, published and
97 unpublished datasets, into a unified database for the North American Arctic sector. This study aims to fill this gap by creating

98 the largest database of its kind at a pan-American scale. This database will facilitate the investigation of global trends in the

99 biogeography, diversity, and composition of phytoplankton taxa across the North American region of the Arctic Ocean.

100 **2 Data and methods**

101 **2.1 Data acquisition**

102 Our database consists of phytoplankton occurrences (i.e., presences and abundances greater than zero) based on web-based

103 search engines and queries in online databases, such as OBIS (https://obis.org), GBIF (https://www.gbif.org) and PANGAEA

104 (https://www.pangaea.de/). Occurrence data from OBIS (last accessed: November 20, 2020) and GBIF (last accessed:

105 November 16, 2020) were first downloaded using the keywords "Chromista" and "Plantae"; from 45° N to 90° N and from

106 40° W to 180° W, without temporal restriction. Occurrence data from PANGAEA (last accessed: November 2020) were

107 collected using the keywords: "Chromista", "Phytoplankton", "Taxonomy", "Harmful algal bloom", "Arctic Ocean", "Polar"

108 and several combinations of these keywords. We supplemented the data with records from ArcticNet campaigns and individual

109 studies (see data in Schiffrine et al., 2024). To indicate the source of each record (GBIF, OBIS, ArcticNet, or individual

110 studies), we added the "sourceArchive" column (Table 1). We standardized the column names to ensure compatibility between

111 different datasets, resulting in a comprehensive dataset of 909 094 data points (Schiffrine et al., 2024).

112 **2.2 Biogeographic classification**

113 Our global database was divided into hexagonal bins using the R package *dggridR* (https://github.com/r-barnes/dggridR;

114 Barnes and Sahr, 2020), with a resolution of 2591.40183 km^2. The chosen grid resolution strikes a balance between providing

115 sufficient spatial resolution to capture ecological patterns and minimizing computational requirements. Each grid cell was then

116 assigned a corresponding Large Marine Ecosystem (LME) region using the spatial polygons obtained from the "*mr_shp*"

117 function of the R package *mregions* (Chamberlain and Schepers, 2021). Conserving only grid labelled as "arctic" according to

118 the LME classification, this new dataset contains 4458 grid cells partitioned into nine different regions and 550 033 data points.

119

120

121

122

123

124

125

126

127

128   **Table 1: Detailed description of the columns.**

Variable name	Definition
verbatimScientificName	Original scientific name recorded
ModifiedName	Taxonomic name after modification or correction to improve accuracy or consistency
ReduceName	Taxonomic name after reduction to a higher taxonomic rank when the original name contained multispecies or complex designations
parse.name	Taxonomic name used for verification with AlgaeBase and/or WoRMS, obtained through parsing and formatting processes to ensure compatibility and consistency with the databases
Open_Nomenclature	Uncertainty or provisional status of taxonomic identification
scientificName	Scientific name according to AlgaeBase and/or WoRMS
acceptedNameUsage	Currently accepted name according to AlgaeBase and/or WoRMS
eventDate	Date and time of the event
year	Integer representing the year in which the event occurred
month	Integer representing the month in which the event occurred
day	Integer representing the day of the month on which the event occurred
DayOfYear	Day of the year in which the event occurred
decimalLongitude	Geographic longitude in decimal degrees
decimalLatitude	Geographic latitude in decimal degrees
depth	Depth in meters at which the event occurred
individualCount	Number or enumeration value representing the quantity of organisms
Comments	Additional comments or notes about the record
dataset_id	Identifer for the dataset
datasetName	Name identifying the dataset from which the record is derived
basisOfRecord	Nature of the record, based on the Darwin Core terms
sourceArchive	Source from which the records were obtained
institutionCode	Name or acronym of the institution having custody of the object or information referred to in the record
seqnum	Identifier for the grid cell
lme_name	Name of the Large Marine Ecosystem (LME) region
abbrev_lme_name	Abbreviation of the Large Marine Ecosystem (LME) region name
scientificNameID	Identifier for scientific name details

acceptedNameUsageID	Identifier for the current accepted scientific name details
URI	Set of identifiers constructed according to the generic syntax for Uniform Resource Identifiers
acceptedNameUsageAuthorship	The authorship information for the acceptedNameUsage formatted according to the conventions of the applicable nomenclatural Code
taxonomicStatus	Status of the use of the scientificName as a label for a taxon
nomenclaturalStatus	Status related to the original publication of the name and its conformance to the relevant rules of nomenclature
taxonRank	Taxonomic rank of the most specific name in the acceptedNameUsage
taxonRankID	Identifier for the taxonomic rank of the most specific name in the acceptedNameUsage
kingdom	Full scientific name of the kingdom in which the taxon is classified
phylum	Full scientific name of the phylum or division in which the taxon is classified
class	Full scientific name of the class in which the taxon is classified
order	Full scientific name of the order in which the taxon is classified
family	Full scientific name of the family in which the taxon is classified
genus	Full scientific name of the genus in which the taxon is classified
bibliographicCitation	Bibliographic reference for the resource
isMarine	Boolean flag indicating whether the taxon is a marine organism, i.e. can be found in/above sea water
isBrackish	Boolean flag indicating whether the taxon is an organism that can be found in brackish water
isFreshwater	Boolean flag indicating whether the taxon occurs in freshwater habitats, i.e. can be found in/above rivers or lakes
isTerrestrial	Boolean flag indicating the taxon is a terrestial organism, I.e. occurs on land as opposed to the sea
isFossil	Boolean flag indicating whether the taxon is an extinct organism
isToxic	Boolean flag indicating whether the taxon is a toxic organism
isHarmful	Boolean flag indicating whether the taxon is a harmful organism
bibliographicCitation_ToxicHarmfulStatut	Bibliographic reference for the resource's toxic and harmful status
modified	Date on which the resource was changed
database	Database source used for the scientific name verification

129

2.3 Data quality control

Each record underwent a verification process to ensure the accuracy of taxonomic identification. First, we used the AlgaeBase database, and the API key provided by the AlgaeBase team to validate each record as an accepted name (http://www.algaebase.org/; AlgaeBase. World-wide electronic publication, 2023). If a record was not validated through this process, we performed a secondary verification using the "*wm_records*" function from the R package *worrms* (Chamberlain and Vanhoorne, 2023), using the World Register of Marine Species database (WoRMS; http://www.marinespecies.org; Ahyong et al., 2023). If the taxonomic identification could not be found in either of these databases, we assigned the record to the next higher taxonomic classification level (n = 39). These modifications were specified in the "ReduceName" column (Table 1). In order to maintain data quality and avoid loss of information, we manually adjusted a total of 249 taxonomic names, with the modified names indicated in the "ModifiedName" column (Table 1). Taxonomic records that included qualifiers such as "aff." (n = 40) and "cf." (n = 95) were categorized at the species level in our dataset to simplify taxonomic classification. While this simplification enhances dataset accessibility, it is crucial to acknowledge the potential introduction of errors due to a certain degree of uncertainty associated with species identification. To maintain transparency, qualifiers originally denoted by "cf." and "aff." were thoughtfully preserved in the "Open_Nomenclature" column (Table 1). This approach strikes a balance between simplification and taxonomic rigor, enabling users to recognize the initial uncertainty in identification and facilitating further investigation or refinement of taxonomic assignments as necessary. Taxonomic records with qualifiers such as "sp." (n = 193) or "spp." (n = 324), as well as those indicating a "group" (e.g., *Pseudo-nitzschia seriata* group; n = 27), "complex" (e.g., *Gymnodinium/Gyrodinium* complex; n = 3), or containing multiple species names (e.g., *Pseudo-nitzschia delicatissima/Pseudo-nitzschia pseudodelicatissima*; n = 12), were categorized at the genus level in the dataset. These qualifiers are denoted in the "Open_Nomenclature" column (Table 1). Less than 1% of the records in our dataset could not be identified in either the AlgaeBase or WoRMS databases. The original taxonomic names were retained in the "verbatimScientificName" (Table 1), allowing for traceability to the harmonized names.

2.4 Potentially toxic and harmful phytoplankton and other protists list

In the context of this study, the term "HA" is used as an abbreviation to refer collectively to potentially toxic and/or harmful algal species. Our comprehensive global list of HA species consists of the IOC-UNESCO taxonomic reference list (http://www.marinespecies.org/hab/; last accessed: June 20, 2023; Lundholm et al., 2009). We subsequently supplemented this list by incorporating the taxa list of Bates et al. (2019, 2020), which notably included the Ciliophora, *Mesodinium rubrum*. We chose to retain *M. rubrum* due to its significant ecological implications (Olsen et al., 2019; McKenzie et al., 2020). We excluded the dinoflagellate *Protoperidinium crassipes* (Kofoid) Balech from Bates et al. (2020) since this species appears to act more as a toxin vector than a toxin producer (e.g., Tillmann et al., 2009). Each taxon was verified as described in the previous section and merged based on the "acceptedNameUsage" column before integration into our global database. We assigned the toxic or harmful status to each record, following the criteria of Lundholm et al. (2009) and Bates et al. (2019,

162   2020). This compiled list includes 113 Dinoflagellata, 49 Heterokontophyta, 43 Cyanobacteriota, 11 Haptophyta, and one

163   Ciliophora species. Of these, 205 species have been identified as toxic (indicated by a flag in the "isToxic" column; Table 1),

164   seven are considered harmful (indicated by a flag in the "isHarmful" column; Table 1), and five species remain under debate

165   regarding their toxic or harmful status (flagged in both "isToxic" and "isHarmful" columns; Table 1).

166   **2.5 Data merger and synthesis**

167   The filters implemented during the data merging and synthesis process aimed to ensure the quality and relevance of the dataset.

168   The filters applied were as follows:

169   ▪ Records without year information were removed to ensure data quality and enable meaningful temporal analysis, as

170   the absence of this crucial temporal component would limit the dataset's usability for studying time-dependent

171   patterns or trends.

172   ▪ Records with depths greater than 2500 meters were excluded, considering the specific characteristics and depth ranges

173   of the Arctic region based on bathymetry data.

174   ▪ Records classified as "fossil only" or "fossil" in either the AlgaeBase or WoRMS databases (e.g., "isFossil" column;

175   Table 1) were excluded to focus only on currently occurring phytoplankton species. However, records classified as

176   freshwater or brackish according to the AlgaeBase or WoRMS databases (i.e., "isFreshwater" and "isBrackish"

177   columns; Table 1) were retained to account for their ecological relevance and potential responses to changing Arctic

178   conditions, given the Arctic's connection to freshwater and brackish coastal regions.

179   ▪ Records not found in either AlgaeBase or WoRMS were excluded to ensure the inclusion of taxonomically validated

180   and accepted names.

181   ▪ Taxa belonging to specific kingdoms (i.e., Animalia, Fungi, Acritarcha), phyla (i.e., Foraminifera, Oomycote,

182   Rhodophyta, Retaria), and classes (i.e., Phaeophyceae, Ulvophyceae) were excluded to maintain the focus on

183   phytoplankton. Note that in the context of this study, the term "phytoplankton" is used in a broad sense to include

184   photoautotrophic prokaryotes and eukaryotes as well as heterotrophic, phagotrophic, and mixotrophic protistan

185   species.

186   ▪ Records identified at a taxonomic level higher than genus were removed from the dataset to ensure consistent and

187   accurate taxonomic classification at the genus level. Retaining records at the genus level allows for a more detailed

188   understanding of the composition of phytoplankton communities in the study area.

189   ▪ Duplicate records were removed, using the following columns: "day", "month", "year", "depth", "decimalLatitude",

190   "decimalLongitude", "verbatimScientificName", "scientificName", "acceptedNameUsage", "basisOfRecord" and

191   "individualCount" (Table 1). This step ensured that each unique sampling event was represented by a single record

192   in the dataset.

193

194   After applying these filters, the dataset contains 385 800 individual georeferenced data points and 18 268 unique sampling

195   events (Fig. 1A; Schiffrine et al., 2024). To access the comprehensive diversity of HA species, we further subset the database

196   based on the "isToxic" and "isHarmful" columns (Table 1), resulting in a dataset with a total of 48 555 georeferenced data

197   points of HA species and 6744 unique sampling events (Fig. 1B; Schiffrine et al., 2024).

198

199
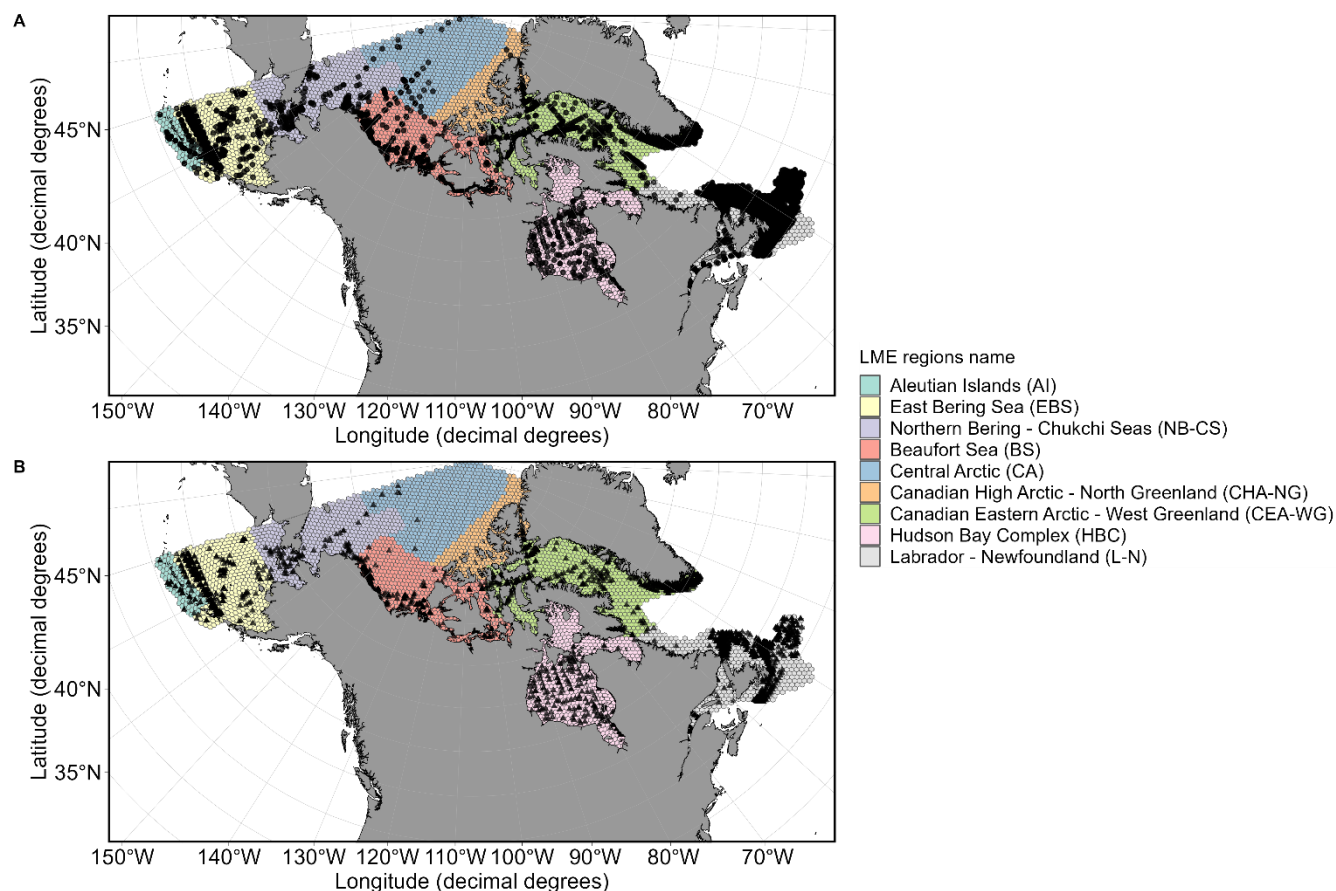
Figure 1: (A) Global distribution of phytoplankton occurrence and (B) potentially toxic or harmful phytoplankton (HA) species records. Abbreviations of the Large Marine Ecosystems (LME) regions are in parentheses.

2.7 Data analysis

203   The size of each LME region was determined by calculating the total number of grid cells (i.e., n_{total}). For each LME region

204   and each month, the number of grid cells containing phytoplankton records was counted and summed per year (i.e., $n_{sampled}$).

205   This value was then divided by n_{total} to estimate the percentage of the region that was sampled, or sample coverage, that specific

206   year within each LME region. The same method was applied to the HA sub-dataset, where $n_{HA\ sampled}$ represents the number of

207   grid cells containing HA records summed per year. Mapping and statistical analysis were performed in R (Team and R

208 Development Core Team, 2019), using *ggOceanMaps* (https://mikkovihtakari.github.io/ggOceanMaps/; Vihtakari, 2021),
209 *vegan* (Oksanen et al., 2020), *fossil* (Vavrek, 2011) and *betapart* (Baselga and Orme, 2012) packages.

3 Results and discussion

3.1 Spatiotemporal coverage

212 The use of long-term datasets has significantly improved our understanding of phytoplankton species distribution and diversity,
213 as well as the underlying drivers of these patterns at both local (McKenzie et al., 2020; Nohe et al., 2020) and global scales
214 (Righetti et al., 2019; Benedetti et al., 2021; Hallegraeff et al., 2021). However, currently available databases on phytoplankton
215 occurrences only provide limited information on the Arctic Ocean. Although PhytoBase (i.e., Righetti et al., 2020) is one of
216 the most comprehensive and up-to-date sources of information on phytoplankton occurrence, data above 60° N are generally
217 underrepresented. Furthermore, the recent study by Hallegraeff et al. (2021) did not specifically address the evolution of HA
218 blooms in the Arctic Ocean, but instead grouped the North American region of the Arctic Ocean within the broader region of
219 "East Coast America".

220

221 Despite the existence of several published phytoplankton lists specifically focused on the North American sector of the Arctic
222 Ocean, there is currently a lack of a comprehensive and freely available standardized database accessible to the scientific
223 community. To fill this gap, our project aimed to compile and integrate a large and diverse collection of data from multiple
224 sources. The objective was to create a comprehensive database covering the distribution of phytoplankton, including HA
225 species, across the North American sector of the Arctic Ocean. Our efforts greatly expanded the spatial and temporal coverage
226 of phytoplankton data across all LME regions in this sector of the Arctic Ocean compared to PhytoBase. Our database covers
227 an impressive time span of 132 years, from 1888 to 2020, with 95% of the data collected after 1963. Sampling was mainly
228 concentrated between the months of June to September, consistent with the seasonal dynamics of phytoplankton in the Arctic
229 Ocean. The spatial distribution of the records was highly unbalanced with 82% of data records falling in the Labrador—
230 Newfoundland region alone, followed by the Canadian Eastern Arctic—West Greenland with 8%. The remaining regions
231 contribute smaller proportions, ranging from 0.1% to 5% of the data records. The dataset covers a depth range from 0 to 1010
232 meters, providing insight into the distribution of phytoplankton at different oceanic depths. It is noteworthy that 95% of the
233 data are derived from surface sampling. It is also important to note that 17% of the dataset has no depth information available.
234 Consequently, caution should be exercised when interpreting the full vertical distribution of phytoplankton. This is particularly
235 important as the Arctic marine environment is characterized by the presence of subsurface chlorophyll *a* maximum (SCM)
236 (e.g., Martin et al., 2012). Due to the scarcity of data and the emphasis on surface data, there is the potential for underestimating
237 a significant portion of biodiversity. Regarding the types of records within the dataset (i.e., basisOfRecord column, Table 1),
238 the majority, constituting 71%, were derived from "HumanObservation", which includes instances where evidence of an
239 occurrence was obtained through field notes, literature, or records of occurrences without physical or machine-recorded

240 evidence. "PreservedSpecimen" accounted for 19% of the dataset, representing samples that have been treated with fixatives
241 for preservation.

3.2 Taxonomic coverage

243 A total of 1445 unique taxa were recorded in our study. This number falls within the range reported by Archambault et al.
244 (2010) and Poulin et al. (2011) for the same region (i.e., 1657 and 1229 taxa, respectively). It's essential to acknowledge that
245 both Archambault et al. (2010) and Poulin et al. (2011) conducted their analyses based on literature reviews predominantly
246 reliant on microscopic observations. In a more recent comprehensive pan-Arctic taxonomic inventory, Lovejoy et al. (2017)
247 reported 2241 taxa primarily using genomic data. The disparity in reported taxa between our study and Lovejoy et al. (2017)
248 can be attributed to the fundamental differences in our respective approaches—our reliance on microscopic observations versus
249 Lovejoy et al.'s use of genomic data. Genomic techniques possess the capacity to identify a broader spectrum of species,
250 including those of smaller size or less conspicuous under microscopic examination, such as the Mamiellophyceae *Micromonas*
251 *polaris* Simon, Foulon & Marin. Microscopic observations, which constitute a substantial portion of our dataset, inherently
252 introduce certain biases. They may overlook rare or small species (<3 µm) and encounter challenges related to precise species
253 identification, compounded by considerations such as the choice of fixative (e.g., acidic Lugol's solution or formalin; Sournia,
254 1978). As a result, our study may not offer a fully comprehensive representation of total species richness, particularly
255 concerning rare or molecularly detectable taxa.

257 In this study, Heterokontophyta and Dinoflagellata were the most commonly occurring phyla, accounting for approximately
258 40% and 36% of total occurrences, respectively. Within the Heterokontophyta phylum, which notably included diatoms (Guiry
259 et al., 2023), the genus *Chaetoceros* Ehrenberg was the most frequently observed, followed by *Thalassiosira* Cleve, which
260 accounted for 24% and 14% of total Heterokontophyta occurrences, respectively. *Tripos* Bory and *Gyrodinium* Kofoid &
261 Swezy were the two most abundant genera in the Dinoflagellata phylum, accounting for 20% and 16% of total Dinoflagellata
262 occurrences, respectively. The observed predominance of Heterokontophyta in this study, particularly the genera *Chaetoceros*
263 and *Thalassiosira,* is in line with the general understanding of Arctic phytoplankton diversity (Poulin et al., 2011; Lovejoy et
264 al., 2017). On the other hand, the findings for Dinoflagellata phylum contrast with prior research that has highlighted the
265 predominance of the genus *Protoperidinium* Bergh (Okolodkov and Dodge, 1996). The exceptionally high occurrence of
266 *Tripos* and *Gyrodinium* should be interpreted with caution. These two genera are mainly observed in the Labrador—
267 Newfoundland region, where the majority of the data collected originates from the Continuous Plankton Recorder (CPR). It is
268 important to note that CPR uses a large mesh size (270 µm) (e.g., Richardson et al., 2006), resulting in an over-representation
269 of larger taxa, such as *Tripos* and *Gyrodinium*.

271 Cyanobacteria seem to be important only in the southern fringes of the subarctic, for instance in the oceanic Bering Sea (e.g.,
272 Sakshaug, 2004), leading to the notion that they are nearly absent from the northernmost latitudes (Vincent, 2002). However,

273 our study yielded remarkable results, revealing the presence of 27 distinct Cyanobacteria taxa. This finding challenges the
274 prevailing understanding of Cyanobacteria rarity in polar marine waters, which has been attributed to the low temperatures of
275 the polar zones (e.g., Flombaum et al., 2013). The 27 taxa encompass a diverse range of Cyanobacteria, with *Synechococcus*
276 Nägeli being the most frequently detected genera (91% of the Cyanobacteria occurrence). This observation provides new
277 insights into the presence and diversity of Cyanobacteria in the Arctic and highlights the need for further research to
278 comprehend their ecological significance. Furthermore, our findings support previous studies suggesting that the presence of
279 Cyanobacteria in the Arctic may be underestimated and indicate that certain Cyanobacteria populations are indigenous to these
280 waters (Paulsen et al., 2016).

281

282 There is a commonly accepted understanding that phytoplankton diversity follows a latitudinal gradient, characterized by the
283 highest diversity near the equatorial region and declining as we move towards the poles (Ibarbalz et al., 2019; Righetti et al.,
284 2019), primarily influenced by the reduction in ocean temperatures (Ibarbalz et al., 2019). While environmental conditions
285 undoubtedly contribute to these patterns of diversity, the scarcity of data may also account for the observed low diversity.
286 Righetti et al. (2020) reported a total of 1704 phytoplankton species, including 239 species within the same grid used in our
287 study. However, our study detected an additional 1361 taxa, with 532 belonging to Heterokontophyta and 362 to
288 Dinoflagellata. The findings of our study challenge the conventional notion of a low-diversity Arctic ecosystem. Additionally,
289 our results indicate that previous research may have significantly underestimated the biodiversity of Arctic phytoplankton
290 (Righetti et al., 2019). Such underestimations can skew our understanding of the latitudinal gradient of phytoplankton diversity.

291 **3.3 Difference in species richness according to the Arctic LME regions**

292 Given the substantial variation in sampling coverage across LME regions over time (Figure S1), it is important to consider the
293 potential impact on species representation in our dataset. The variation in sample coverage across regions may introduce bias
294 and affect species coverage. Regions with greater sampling coverage, such as the Labrador—Newfoundland region (Figure
295 S1), are likely to have a higher probability of capturing a wider range of species and provide a more comprehensive
296 representation of local biodiversity. In contrast, regions with limited sampling coverage, such as the Aleutian Islands, Central
297 Arctic, or Canadian High Arctic—North Greenland regions (Figure S1), may underestimate species richness, potentially
298 leading to an incomplete understanding of the true species composition in these areas. To address this concern, we used the
299 Chao2 index, a widely used nonparametric method for estimating species richness in a community (Chao and Shen, 2003).
300 Application of the Chao2 index allows us to assess alpha diversity, i.e., diversity on a local scale, within each LME region,
301 especially when working with frequency counts or presence/absence data. The Chao2 index shows significant differences in
302 diversity among the LME regions (Fig. 2). The Canadian Eastern Arctic—West Greenland and Beaufort Sea regions exhibited
303 the highest values of the Chao2 index, indicating a higher species richness and diversity within their phytoplankton
304 communities (Fig. 2). In contrast, the Aleutian Islands and Central Arctic regions had the lowest values, suggesting lower
305 species richness and diversity in these areas (Fig. 2). The East Bering Sea, Hudson Bay Complex, Labrador—Newfoundland,

306   and Northern Bering—Chukchi Seas regions also showed moderate to high Chao2 index values, indicating varying levels of

307   species richness and diversity across these regions (Fig. 2).

308
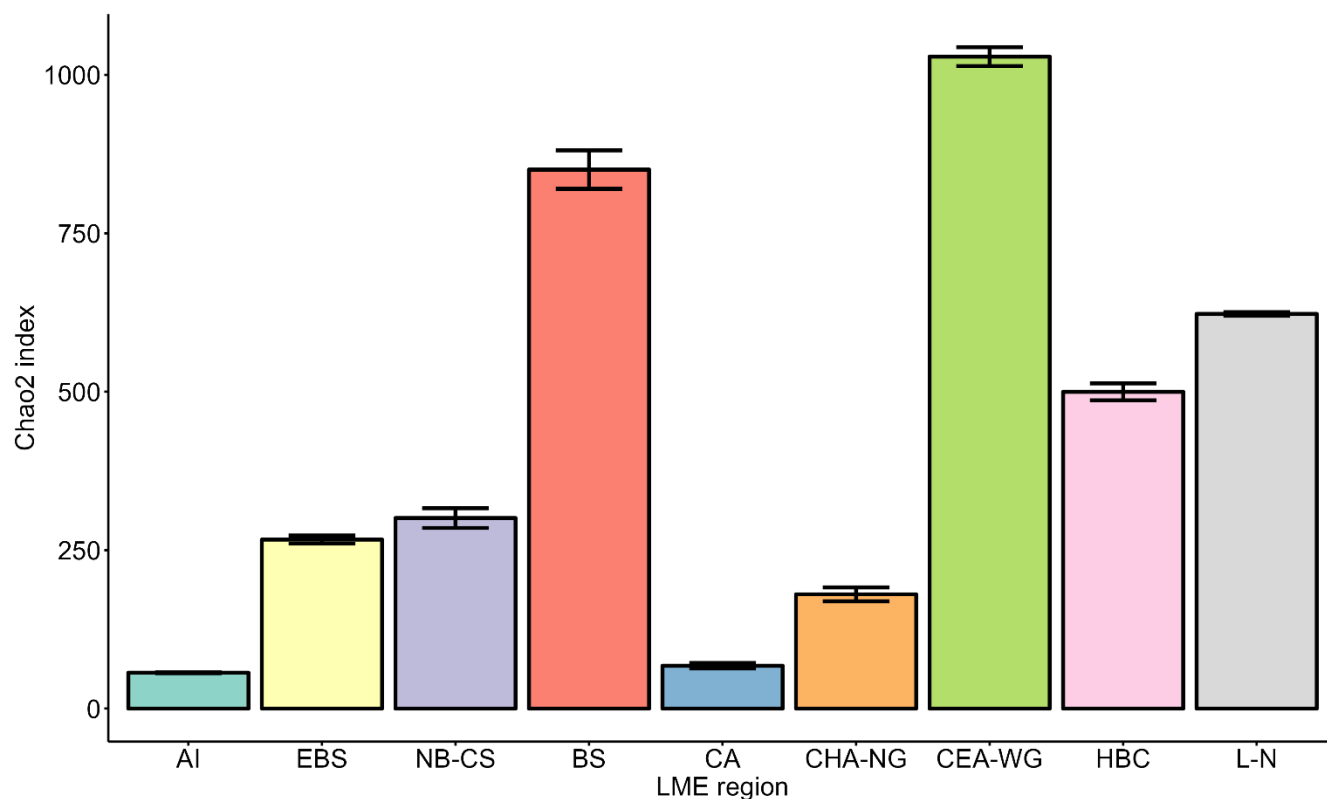
Figure 2: Chao2 index for each Large Marine Ecosystem (LME) region. Error bars represent the standard deviation (i.e., SD). LME regions are labeled as follows: AI (Aleutian Islands), EBS (East Bering Sea), NB-CS (Northern Bering—Chukchi Seas), BS (Beaufort Sea), CA (Central Arctic), CHA-NG (Canadian High Arctic—North Greenland), CEA-WG (Canadian Eastern Arctic—West Greenland), HBC (Hudson Bay Complex) and L-N (Labrador—Newfoundland).

314

315   To further analyze local diversity, we used species accumulation curves (SACs) to illustrate the number of species sampled

316   relative to the level of sampling effort (Thompson and Withers, 2003). SACs typically reach an asymptote when sufficient

317   sampling effort is achieved, enabling us to estimate the comprehensiveness of species richness detection. In our analysis, we

318   computed the SACs based on the number of species observed in each grid cell for each month of every year, yielding valuable

319   insights into species richness and the extent of saturation (i.e., completeness of species richness detection) between regions

320   (Fig. 3). The Hudson Bay Complex and Labrador—Newfoundland regions exhibited saturation at cumulative richness levels

321   of around 400 and 600 taxa (Fig. 3), respectively, indicating that a significant portion of the taxa present in these regions had

322   been sampled. Conversely, the SACs for other regions did not reach a plateau, suggesting that the sampling effort was

323 insufficient to capture the complete diversity (Fig. 3). This pattern was particularly pronounced in the northernmost regions,

324 such as Central Arctic and Canadian High Arctic—North Greenland (Fig. 3).

325

326
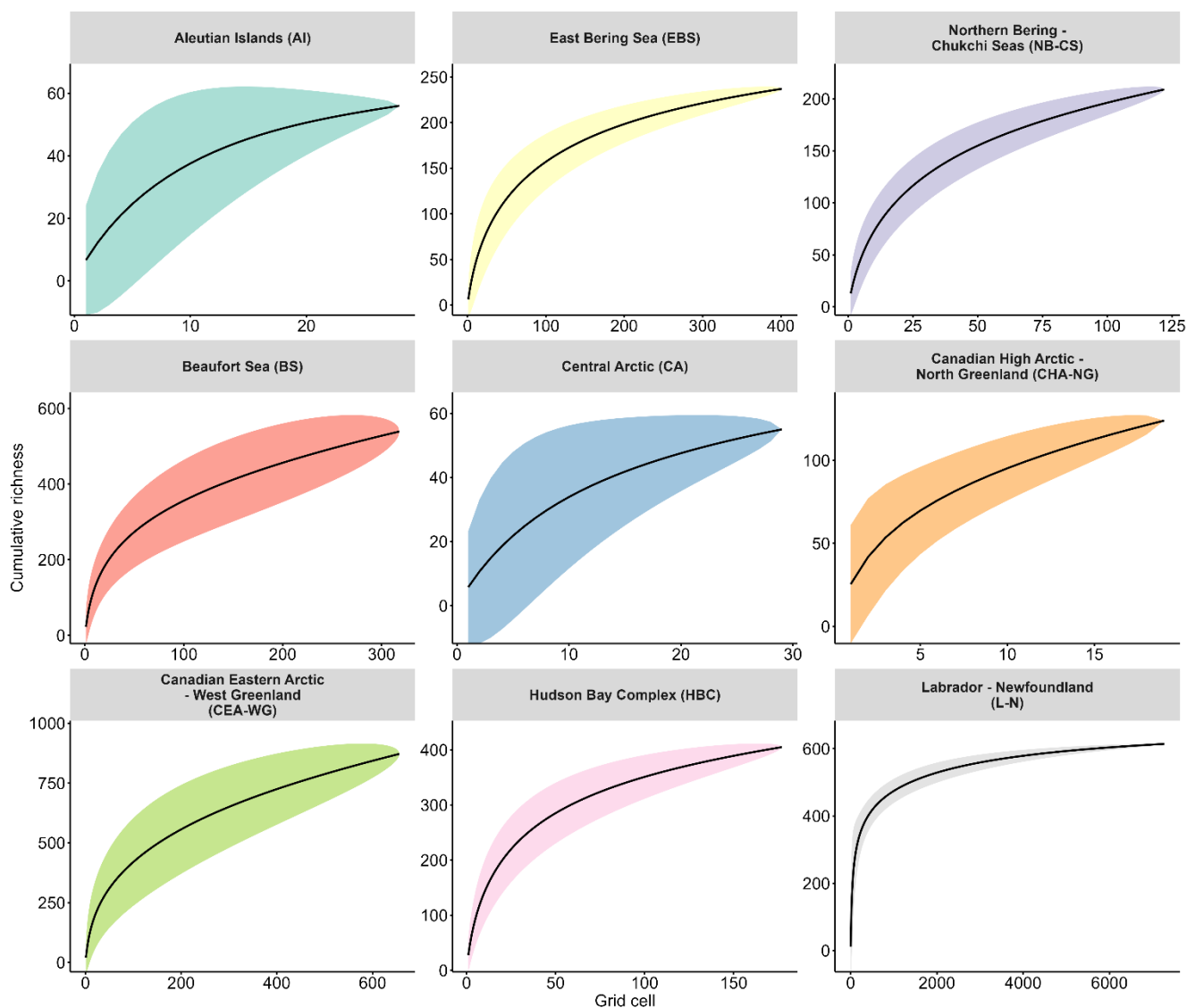
Figure 3: Cumulative richness as a function of grid cell for each month of every year with 95% confidence intervals (color area) for each Large Marine Ecosystem (LME) region. Abbreviations of the LME regions are in parentheses.

329

330 The Central Arctic region, known for its extreme environmental conditions, such as low-nutrient concentrations and prolonged

331 annual sea-ice cover (Codispoti et al., 2013), exhibits lower Chao2 index compared to other regions (Fig. 2). However, the

332 SAC did not saturate (Fig. 3), indicating that the actual diversity in this region may be higher than observed in this study.

14

333 Despite the inflow of nutrient-rich water through the Bering Strait (Torres-Valdés et al., 2013), which contributes to the high

334 productivity of phytoplankton in the Pacific regions such as East Bering Sea and Northern Bering—Chukchi Seas (Tremblay

335 et al., 2015), these regions display a relatively low Chao2 index (Fig. 2). One possible explanation for this observation is that

336 the sampling effort conducted in these regions may not have been sufficient to capture the complete range of species diversity,

337 leading to an underestimation of richness. This is supported by the SACs (Fig. 3), which show that the curves for both regions

338 do not reach a plateau, indicating that the sampling effort was insufficient to fully capture the diversity present in these areas.

339 The Labrador—Newfoundland region displays an intermediate Chao2 index (Fig. 2) despite the SACs indicating that the

340 majority of the species have been recorded (Fig. 3). This observation may be attributed to the sampling methods. The majority

341 of the data from this region is derived from the CPR program, which has a large mesh size (270 µm), fixed sampling depth (5–

342 10 m), and high sampling speed (15–20 knots). This collecting technique may result in under-sampling smaller and fragile

343 species, as well as those not constantly present in the surface mixed layer (Richardson et al., 2006). The Chao2 index values

344 for both the Beaufort Sea and Canadian Eastern Arctic—West Greenland regions were remarkably high (Fig. 2). This finding

345 aligns with the traditional view of the Canadian Eastern Arctic—West Greenland region as one of the most productive areas

346 in the North American Arctic (Hill et al., 2013). However, it comes as a surprising result for the Beaufort Sea region, which is

347 generally considered a low-productivity area, particularly in its northern part (Hill et al., 2013). The unexpected high Chao2

348 index value for the Beaufort Sea region may be attributed to the majority of samples being collected from nearshore areas

349 (Figure S2). These nearshore regions are known for their high productivity (Ardyna et al., 2017), likely due to their exposure

350 to nutrient-rich waters that support the growth and diversity of phytoplankton communities. Nevertheless, SACs for both

351 regions indicate that sampling efforts in these areas are incomplete (Fig. 3). This implies that the diversity may be

352 underestimated and underscores the significance of further sampling to achieve a more accurate understanding of the local

353 biodiversity in both the Beaufort Sea and Canadian Eastern Arctic—West Greenland regions.

354

355 The beta diversity (β) assessment provides valuable insights into the dissimilarity of species composition between multiple

356 samples, enabling researchers to understand the variation in biodiversity across different spatial scales (Whittaker, 1972). In

357 this study, we used the Sørensen dissimilarity index ($\beta_{SØR}$) as the β diversity index to determine the proportion of species not

358 shared between LME regions. The $\beta_{SØR}$ values range from 0 to 1, where 0 indicates identical taxonomic composition at all

359 sites, and 1 represents completely different sets of taxa (Baselga, 2010). Our analysis revealed the subdivision of the LME

360 regions into three distinct clusters based on their species composition (Fig. 4). The first cluster, known as the "Pacific Cluster",

361 includes the Aleutian Islands, East Bering Sea, and Northern Bering—Chukchi Seas regions. The second cluster, referred to

362 as the "Northern Arctic Cluster", encompasses the Central Arctic and Canadian High Arctic—North Greenland regions. Lastly,

363 the third cluster, named the "Mixed Arctic Cluster", consists of the Beaufort Sea, Canadian Eastern Arctic—West Greenland,

364 Hudson Bay Complex, and Labrador—Newfoundland regions. The grouping of the Aleutian Islands, East Bering Sea, and

365 Northern Bering—Chukchi Seas regions is anticipated owing to their common water supply and circulation patterns, which

366 involve receiving water inflows from the Pacific Ocean through the Bering Strait (Rudels and Carmack, 2022). Consequently,

367 this leads to comparable environmental conditions and nutrient inputs, which, in turn, explain the observed similarities in
368 phytoplankton composition. Similarly, the Central Arctic and Canadian High Arctic—North Greenland regions share common
369 water circulation patterns in the Arctic Ocean (Rudels and Carmack, 2022). This common circulation pattern, along with their
370 environmental characteristics, contributes to similarities in their phytoplankton composition. The inclusion of the Beaufort Sea
371 region with Atlantic-dominant regions (i.e., Canadian Eastern Arctic—West Greenland, Hudson Bay Complex, And
372 Labrador—Newfoundland) into one unique cluster may initially seem contradictory due to its geographical location outside
373 the Atlantic side of the Arctic and its lack of direct influence from Atlantic waters (Rudels and Carmack, 2022). However, this
374 clustering is based on similarities in species composition rather than geographical or environmental variables proximity.
375 Despite its location, the Beaufort Sea region exhibits a higher resemblance in phytoplankton species composition to the
376 Canadian Eastern Arctic—West Greenland, Hudson Bay Complex, and Labrador—Newfoundland regions compared to other
377 regions in the dataset. This unexpected similarity may be attributed to oceanic circulation patterns and water mass transport
378 mechanisms that connect these regions (Rudels and Carmack, 2022). These circulation patterns and transport mechanisms may
379 facilitate the dispersal of phytoplankton species from the Beaufort Sea to Atlantic-dominated regions, thereby influencing the
380 observed similarities in species composition. Such dispersal due to circulation patterns has been documented by Reid et al.
381 (2007), who observed the spread of the diatom *Neodenticula seminae* (Simonsen & T. Kanaya) Akiba & Yanagisawa from the
382 Northwest Arctic to the Atlantic side, possibly through the Canadian Arctic Archipelago and/or Fram Strait. This observation
383 provides additional support for the concept of shared phytoplankton species composition influenced by oceanic circulation.
384 These findings provide important insights into the biogeographical patterns of phytoplankton communities in the Arctic LME
385 regions and highlight the significance of considering both geographic and ecological factors when interpreting these patterns.

386 **3.4 Potentially toxic and harmful phytoplankton diversity**

387 The presence of HA species has been a well-known concern in temperate marine and freshwater ecosystems, but their
388 occurrence in the marine Arctic ecosystem is relatively new. With the ongoing climate change in the Arctic Ocean, there is a
389 high probability that the frequency of HA occurrences will increase. Additionally, the expansion of HA distribution from other
390 regions due to increased ship traffic in the Arctic may further exacerbate this problem (e.g., Chan et al., 2019). HA species
391 present substantial risks to both human and ecosystem health. The phycotoxins produced by some of these organisms can
392 bioaccumulate in higher trophic level organisms, including mollusks, seabirds, and marine mammals. When transferred to
393 higher trophic levels, these phycotoxins can result in massive mortalities, neurological or gastrointestinal adverse effects if
394 consumed in concentrations that surpass safe thresholds. In the Alaskan sector of the Bering Sea, concentrations of these toxins
395 that could pose a health risk to local populations have been detected in the tissues of shellfish (Gao et al., 2019). This issue is
396 particularly important as northern populations depend on traditional harvest of fish, shellfish, and marine mammals as
397 subsistence food.
398

399
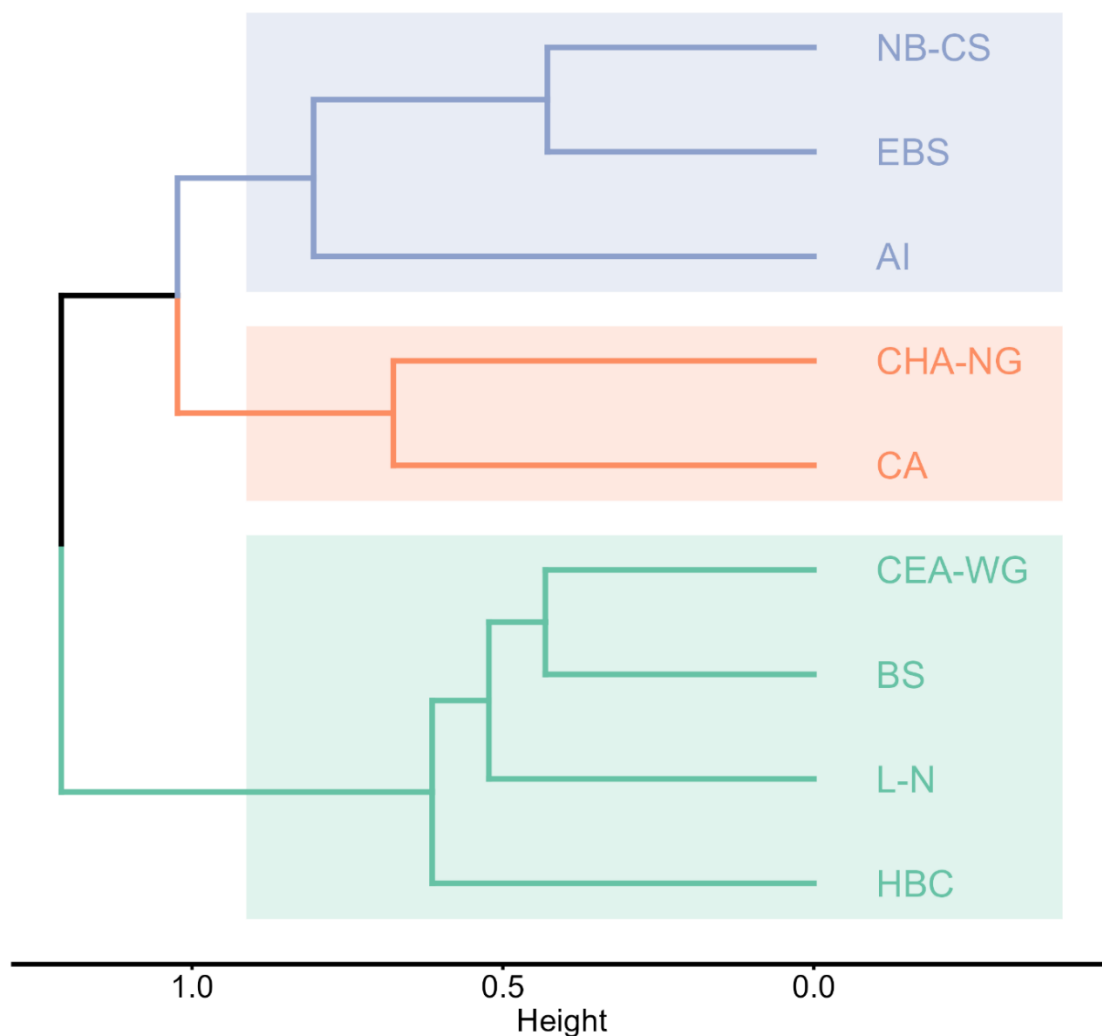
Figure 4: Cluster analysis of β diversity Sørensen dissimilarity index ($\beta_{SØR}$) between the different Large Marine Ecosystem (LME) region obtained with Ward's cluster method. LME regions are labeled as follows: NB-CS (Northern Bering—Chukchi Seas), EBS (East Bering Sea), AI (Aleutian Islands), CHA-NG (Canadian High Arctic—North Greenland), CA (Central Arctic), CEA-WG (Canadian Eastern Arctic—West Greenland), BS (Beaufort Sea), L-N (Labrador—Newfoundland) and HBC (Hudson Bay Complex).

405

Of the 217 HA species compiled from Lundholm et al. (2009) and Bates et al. (2019, 2020) (see section 2.4 for details), our

database identified 59 species. Notably, our study detected a higher number of HA species compared to previous studies

conducted by Poulin et al. (2011) and Pućko et al. (2019), who reported 36 and 27 species, respectively, after updating their

species lists with revised taxonomy. It is noteworthy that both studies primarily aggregated data from literature reviews based

410 on microscopic observations. Our study contributed an additional 25 species, including 16 species from the Dinoflagellata

411 phylum and seven species from the Heterokontophyta phylum. It is important to note that at least 11 species reported by Poulin

412 et al. (2011) were not detected in the present work, as they occur in other Arctic regions not covered in our study, such as

413 *Alexandrium minutum* Halim observed in the Russian and Scandinavian regions. Surprisingly, we observed the presence of

414 *Pyrodinium bahamense* Plate in our dataset, an occurrence hosted on GBIF (https://www.gbif.org/dataset/b42d7c7f-43e5-

415 4e24-abd7-fab3b4fceb09) and published by MGnify (https://www.ebi.ac.uk/metagenomics; e.g., Mitchell et al., 2020). This

416 observation is indeed intriguing, as *P. bahamense* is a species typically associated with warm tropical waters in regions such

417 as the Caribbean, Central America, the Persian Gulf, and the Red Sea (Morquecho, 2019). Its cyst stage distribution is more

418 extensive, encompassing tropical to subtropical coastal areas in both the Atlantic and Pacific regions (Morquecho, 2019). The

419 occurrence of this species in Arctic waters is unexpected and warrants caution in interpretation. MGnify, primarily designed

420 for microbiome data analysis, may introduce potential uncertainties when applied to species identification. Therefore, it is

421 important to exercise caution and thoroughly assess the accuracy of species identification and their geographic distribution in

422 such cases. In addition, it is worth noting that the publication referenced by MGnify (e.g., Joli et al., 2018) does not mention

423 the presence of *P. bahamense*, a fact also confirmed by the authors (pers. comm.), further adding to the uncertainty of this

424 occurrence. Despite the uncertainties and surprising nature of this observation, several reasons may have led to the decision to

425 retain this occurrence in the dataset. Retaining all data, even unexpected or unusual observations, is standard practice in

426 scientific research. It ensures transparency and data integrity, allowing other researchers to access and evaluate the data in its

427 entirety. In addition, the presence of *P. bahamense* in the Arctic, if confirmed, could have significant ecological implications.

428 Thus, the preservation of these data will allow further study of the species' behavior, its adaptation to Arctic conditions, and

429 its potential impact on local ecosystems. This case highlights the complexity of ecological research and the need for careful

430 consideration when working with large datasets and automated analysis tools such as MGnify in unique environmental contexts

431 such as the Arctic.

432

433 Many of the species highlighted in our study are of particular concern for the Arctic Ocean due to their production of

434 phycotoxins. Based on Lundholm et al. (2009) and Bates et al. (2019, 2020), we identified 48 potentially toxin-producing

435 species, as indicated by the "isToxic" flag (Table 1). Of the 73 accepted species included in the genus *Pseudo-nitzschia*

436 (AlgaeBase. World-wide electronic publication, 2023; last access October 2023), 28 are known to produce domoic acid

437 (Lundholm et al., 2009; Bates et al., 2019), with nine of these toxin-producing species being present in our database.

438 Meanwhile, at least 16 of the 45 accepted species in the genus *Alexandrium* are known to be toxic (Lundholm et al., 2009;

439 AlgaeBase. World-wide electronic publication, 2023; last access October 2023), but only five species have been recorded in

440 our database. The dinoflagellate genus *Dinophysis* has 276 phototrophic and heterotrophic accepted species worldwide

441 (AlgaeBase. World-wide electronic publication, 2023; last access October 2023), and 10 of these species have been found to

442 produce various toxins (Lundholm et al., 2009). Additionally, 14 out of the 133 accepted species of the genus *Prorocentrum*

443 Ehrenberg have been confirmed to produce a range of toxins (Lundholm et al., 2009; AlgaeBase. World-wide electronic

Earth System
Open Access
Science
Discussions
Data

444 publication, 2023; last access October 2023). Our database contains at least six and five species from the genera *Dinophysis*

445 and *Prorocentrum*, respectively.

446

447 While studies have detected phycotoxins in the North American Arctic (Baggesen et al., 2012; Li et al., 2016; Elferink et al.,

448 2017; Gao et al., 2019; Hubbard et al., 2023; Pućko et al., 2023), there are no reports of HA events at high latitudes (>60° N)

449 in the Harmful Algal Event Database (HAEDAT; http://haedat.iode.org/index.php; last access October 2023). HAEDAT's

450 criteria for a HA event are strict, including toxin accumulation in seafood above safe levels, discoloration or scum in the water

451 causing ecosystem or socioeconomic damage, negative effects on humans, animals, or other organisms, or precautionary

452 closures of harvesting areas based on predefined thresholds of toxic phytoplankton cells in the water. This suggests that these

453 events may not meet the HAEDAT criteria and raises questions about the relevance of these criteria for polar areas.

454

455 Approximately 50% of all HA occurrences are primarily represented by only five species: *Pseudo-nitzschia delicatissima*

456 (Cleve) Heiden, *P. seriata* (Cleve) Peragallo, *Dinophysis acuminata* Claparède & Lachmann, *Prorocentrum cordatum*

457 (Ostenfeld) Dodge, and *Mesodinium rubrum* (Lohmann) Leegard (Figure S3). Among these species, *P. delicatissima*, *P.*

458 *seriata*, *P. cordatum*, and *M. rubrum* showcase an extensive geographical distribution, aligning with their broad prevalence

459 across different regions, including the Arctic (Figure S4) (Lassus et al., 2016; Bates et al., 2018, 2020). In contrast, *D.*

460 *acuminata* demonstrates a more confined geographical range (Figure S4). While the presence of *M. rubrum*, which serves as

461 prey for *D. acuminata* (Reguera et al., 2012), could potentially influence the distribution of this species, it cannot provide a

462 comprehensive explanation for the constrained range of *D. acuminata*. This notion is underscored by the relatively broader

463 geographic distribution observed for *M. rubrum* (Figure S4). Temperature is also unlikely to be another restricting factor, as

464 *D. acuminata* demonstrates tolerance to a temperature range from 4 to 10 °C, which corresponds to the temperatures observed

465 in the region where this species occurred. The observed limited distribution of *D. acuminata* may therefore arise from a

466 complex interplay of ecological and environmental factors that collectively shape its spatial pattern, a certainty that remains

467 elusive in the scope of the present study.

468

469 Climate change is expected to cause HA species to move towards northern latitudes, resulting in their increased prevalence in

470 the North American Arctic region. Although evidence suggests that the harmful prymnesiophyte species *Phaeocystis pouchetii*

471 may outcompete diatoms in the European sector of the Arctic due to the intensification of the "Atlantification" phenomenon

472 (Nöthig et al., 2015), the extent of this northward progression of HA species in other Arctic regions, notably the North

473 American sector remains relatively unexplored. In the subsequent sections, we will address this topic through examining the

474 increase in the northernmost latitude at which HA species are observed over the years. The analysis of the temporal variation

475 in the maximum latitude of HA species for each month and year (i.e., max. Lat_{HA}) reveals a relative gradual increase over time

476 (Fig. 5A). However, this trend is likely influenced by heightened oceanographic research and expeditions in higher latitudes,

477 as evidenced by the strong correlation between max. Lat_{HA} and the maximum recorded latitude (max. $Lat_{recorded}$; ρ = 0.9; p-

478   value <0.01; Fig. 5B). Nonetheless, this association appears to exhibit variability depending on the species (Table 2). It is

479   noteworthy that among the species analyzed, there are 24 with insufficient available data to calculate the correlation (Table 2).

480   For 12 species, there is a very weak Spearman rank correlation (-0.2 <ρ <0.2; Table 2), indicating no meaningful linkage

481   between max. Lat_{HA} and max. $Lat_{recorded}$. One such example is the dinoflagellate species *Karenia mikimotoi* (Miyake &

482   Kominami ex Oda) Hansen & Moestrup, which consistently maintains a near-constant max. Lat_{HA} despite increasing max.

483   $Lat_{recorded}$ (Fig. 5C). This pattern suggests that while the sampling efforts expand northward, *K. mikimotoi* seems to be restricted

484   to a specific latitude range. This limited latitudinal distribution is possibly attributed to its temperature tolerance range (4–

485   30°C) (Li et al., 2019). The colder temperatures in the North American Arctic align with the lower thermal limit of this species,

486   likely acting as a thermal barrier to the dispersal of *K. mikimotoi*. Conversely, 12 species demonstrate a strong positive

487   correlation (ρ> 0.6; Table 2) that emphasizes a significant relationship between max. Lat_{HA} and max. $Lat_{recorded}$. For instance,

488   the max. Lat_{HA} of the raphidophyte species *Heterosigma akashiwo* (Hada) Hada ex Hara & Chihara appears to be closely

489   linked to max. $Lat_{recorded}$ (Fig. 5C; Table 2). This suggests the possibility of *H. akashiwo* being a permanent resident of the

490   North American Arctic algal community. However, there remains uncertainty about whether the species observed in the

491   database correspond to those found in temperate regions, as records of *H. akashiwo* in our database are identified with qualifiers

492   such as "cf." or "aff." (Bérard-Therriault et al., 1999; Schiffrine et al., 2024), indicating some uncertainty in their identification.

493   In addition, Arctic conditions may not be conducive to its growth (Edvardsen and Imai, 2006; Mehdizadeh Allaf, 2023). In

494   particular, toxin production is lowest at 30 °C, and blooms of *H. akashiwo* have been observed at temperatures ≥ 15 °C

495   (Edvardsen and Imai, 2006; Mehdizadeh Allaf, 2023), suggesting that toxin production in the Arctic might be significantly

496   reduced due to lower temperatures. The findings concerning the constrained latitudinal distribution of *K. mikimotoi* and the

497   potential permanent residency of *H. akashiwo* in North American Arctic waters highlight the significance of investigating

498   environmental factors and biological traits that shape the distribution and abundance of HA species Arctic Ocean and adjacent

499   seas. In particular, gaining insights into thermal limits, growth requirements, and toxin production of these species can provide

500   valuable information on their responses to the evolving Arctic climate and potential risks to human health and ecosystems.

501   Further research is needed to investigate the population dynamics and ecological roles of these HA species within the Arctic

502   context, as well as their interactions with other marine organisms and the physical environment.

503   **4 Data availability**

504   The dataset described in this work is published in the Zenodo repository: https://zenodo.org/records/10557176 (Schiffrine et

505   al., 2024).

506

507
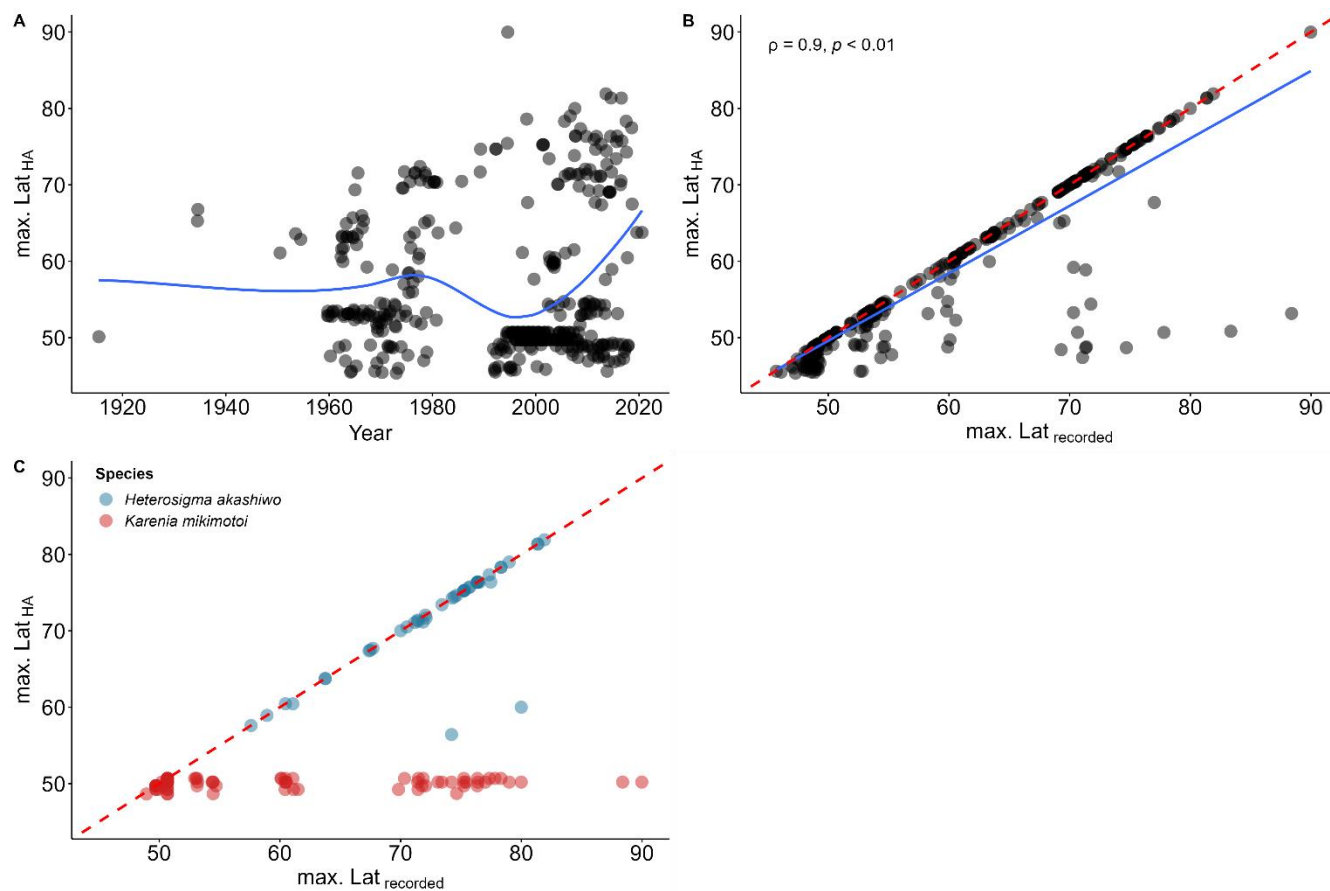
Figure 5: (A) Temporal variation of the maximum latitude of HA species for each month and year (max. Lat$_{HA}$); the blue line represents the local polynomial regression fitting. (B) Relationship between maximum latitude of HA species for each month and year (max. Lat$_{HA}$) and the maximum latitude recorded for each month and year (max. Lat$_{recorded}$); the blue line represents the linear model; the red dashed line represents the 1:1 slope. (C) Relationship between maximum latitude of HA species for each month and year (max. Lat$_{HA}$) and the maximum latitude recorded for each month and year (max. Lat$_{recorded}$) for *Heterosigma akashiwo* (yellow circle) and *Karenia mikimotoi* (blue circle); the red dashed line represents the 1:1 slope.

5 Code availability

The code used in this study is publicly accessible on Zenodo https://zenodo.org/records/10557176 (Schiffrine et al., 2024).
This repository contains the scripts and tools used for various aspects of our study, including data conversion, data quality
control, analysis, and visualization.

520 **Table 2: Summary of the Spearman rank correlation (ρ) analysis between maximum latitude of HA species for each month and year**
521 **(max. Lat$_{HA}$) and the maximum latitude recorded for each month and year (max. Lat$_{recorded}$) for each HA taxon.**

Phylum	Taxon	ρ	p-value
Ciliophora	*Mesodinium rubrum* Lohmann	0.381	***
Cyanobacteriota	*Dolichospermum spiroides* (Klebhan) Wacklin, L.Hoffmann & Komárek	—	—
	Planktothrix agardhii (Gomont) Anagnostidis & Komárek	—	—
	Alexandrium catenella (Whedon & Kofoid) Balech	—	—
	Alexandrium monilatum (J.F.Howell) Balech	—	—
	Alexandrium ostenfeldii (Paulsen) Balech & Tangen	-0.042	NS
	Alexandrium pseudogonyaulax (Biecheler) Horiguchi ex K.Yuki & Y.Fukuyo	-0.177	NS
	Alexandrium tamarense (Lebour) Balech	0.026	NS
	Amphidinium carterae Hulburt	-0.097	NS
	Amphidinium klebsii Kofoid & Swezy	—	—
	Amphidinium operculatum Claparède & Lachmann	—	—
	Dinophysis acuminata Claparède & Lachmann	0.322	***
	Dinophysis acuta Ehrenberg	0.46	***
	Dinophysis norvegica Claparède & Lachmann	0.144	*
	Dinophysis ovum F.Schütt	—	—
	Dinophysis tripos Gourret	-0.7	*
	Gonyaulax spinifera (Claparède & Lachmann) Diesing	0.078	NS
	Gymnodinium catenatum H.W.Graham	—	—
	Hematodinium Chatton & Poisson	—	—
	Karenia mikimotoi (Miyake & Kominami ex Oda) Gert Hansen & Moestrup	0.164	**
	Lingulodinium polyedra (F.Stein) J.D.Dodge	—	—
	Margalefidinium fulvescens (M.Iwataki, H.Kawami & Matsuoka) F.Gómez, Richlen & D.M.Anderson	—	—
	Margalefidinium polykrikoides (Margalef) F.Gómez, Richlen & D.M.Anderson	—	—
	Noctiluca scintillans (Macartney) Kofoid & Swezy	—	—
	Phalacroma rotundatum (Claparéde & Lachmann) Kofoid & J.R.Michener	0.014	NS
	Prorocentrum concavum Y.Fukuyo	—	—

	Prorocentrum cordatum (Ostenfeld) J.D.Dodge	0.561	***
	Prorocentrum emarginatum Y.Fukuyo	—	—
	Prorocentrum lima (Ehrenberg) F.Stein	0.038	NS
	Prorocentrum mexicanum Osorio-Tafall	—	—
	Prorocentrum micans Ehrenberg	-0.536	*
	Prorocentrum rhathymum A.R.Loeblich III, Sherley & R.J.Schmidt	—	—
	Protoceratium reticulatum (Claparède & Lachmann) Bütschli	0.515	***
	Pyrodinium bahamense L.Plate	—	—
Haptophyta	*Chrysochromulina leadbeateri* Estep, Davis, Hargreaves & Sieburth	—	—
	Haptolina ericina (Parke & Manton) Edvardsen & Eikrem	0.005	NS
	Haptolina hirta (Manton) Edvardsen & Eikrem	0.867	***
	Phaeocystis pouchetii (Hariot) Lagerheim	0.638	***
	Prymnesium parvum N.Carter	-0.489	*
	Prymnesium polylepis (Manton & Parke) Edvardsen, Eikrem & Probert	-0.014	NS
	Pseudohaptolina birgeri (Hällfors & Niemi) Ribeiro & Edvardsen	—	—
Heterokontophyta	*Aureococcus anophagefferens* Hargraves & Sieburth	—	—
	Chaetoceros concavicornis Mangin	0.4	***
	Chaetoceros convolutus Castracane	0.175	**
	Chaetoceros debilis Cleve	0.387	***
	Corethron pennatum (Grunow) Ostenfeld	0.69	***
	Dictyocha fibula Ehrenberg	0.688	***
	Heterosigma akashiwo (Hada) Hada ex Y.Hara & M.Chihara	0.868	***
	Leptocylindrus minimus Gran	0.367	***
	Octactis speculum (Ehrenberg) F.H.Chang, J.M.Grieve & J.E.Sutherland	0.673	***
	Pseudo-nitzschia australis Frenguelli	—	—
	Pseudo-nitzschia delicatissima (Cleve) Heiden	0.681	***
	Pseudo-nitzschia fraudulenta (Cleve) Hasle	—	—
	Pseudo-nitzschia granii (Hasle) Hasle	—	—
	Pseudo-nitzschia obtusa (Hasle) Hasle & Lundholm	0.881	***
	Pseudo-nitzschia pseudodelicatissima (Hasle) Hasle	0.703	***
	Pseudo-nitzschia pungens (Grunow ex Cleve) Hasle	0.68	***
	Pseudo-nitzschia seriata (Cleve) H.Peragallo	0.695	***
	Pseudo-nitzschia turgidula (Hustedt) Hasle	0.803	*

522 *NS, p-value> 0.05; *, p-value <0.05; **, p-value <0.01; ***, p-value <0.001; — indicates not tested.*

523

6 Conclusion

525 Several databases exist that document the occurrence of phytoplankton in temperate marine ecosystems. However, these
526 resources often have limited representation of polar ecosystems or may lack such data entirely. Given the substantial
527 environmental changes in the Arctic Ocean and their impact on phytoplankton, it is crucial to expand our understanding of
528 Arctic phytoplankton biodiversity and biogeography. This study compiled various sources of digital biological records, both
529 published and unpublished, to create a comprehensive dataset for North American Arctic marine phytoplankton occurrences.
530 This dataset encompasses 384 645 individual georeferenced data points and 18 266 unique sampling events, covering 1422
531 species, including key phyla like Heterokontophyta, Dinoflagellata, Haptophyta, Ciliophora, and others. This effort addresses
532 the historical limitations of Arctic phytoplankton data, which were often confined to specific regions or lacked comprehensive
533 geographical and date-referenced records (Poulin et al., 2011).

534

535 Our study provides the largest database to date on the occurrence of phytoplankton species in the North American part of the
536 Arctic. This dataset can serve as a valuable resource for investigating the biogeography and phenology of phytoplankton in the
537 region, particularly when integrated with other published datasets. Through the application of geostatistical methods, our
538 database contributes to a refined understanding of potential changes in Arctic phytoplankton communities in the future.
539 Additionally, by supplementing our dataset with information concerning the toxicity or harmful nature of species, it facilitates
540 assessments of the potential proliferation of toxic and harmful species within the Arctic Ocean.

541

542 Moreover, in light of the comprehensive analysis conducted in this study, it becomes evident that routine phycotoxin
543 monitoring should encompass the North American Arctic. Historically, this region has been overlooked, driven by the
544 assumption that high-latitude Arctic areas are not prone to significant toxic algal blooms or phycotoxin contamination.
545 Nevertheless, our research, which sheds light on the diversity, distribution, and prevalence of HA species within this distinct
546 area, emphatically underscores the urgency to reassess this perspective. The conspicuous presence of HA species underscores
547 the imperative for a comprehensive and proactive monitoring strategy.

Author contributions

549 NS led the study, the data processing and archiving, and the writing. FD, MP, SL, AR, and MG provided the quality-controlled
550 data from different Arctic regions. KD initiated the collection of data in Iqaluit and surrounding areas of Frobisher Bay as part
551 of the Baseline program in 2019. All authors reviewed and commented the manuscript.

Open Access

Earth System
Science
Data

Discussions

Competing interests

The authors declare that no competing interests are present.

Acknowledgements

References

Ahyong, S., Boyko, C. B., Bailly, N., Bernot, J., Bieler, R., Brandão, S. N., Daly, M., De Grave, S., Gofas, S., Hernandez, F., Hughes, L., Neubauer, T. A., Paulay, G., Boydens, B., Decock, W., Dekeyzer, S., Vandepitte, L., Vanhoorne, B., Adlard, R., Agatha, S., Ahn, K. J., Akkari, N., Alvarez, B., Amler, M. R. W., Amorim, V., Anderberg, A., Andrés-Sánchez, S., Ang, Y., Antić, D., Antonietto, L. S. ., Arango, C., Artois, T., Atkinson, S., Auffenberg, K., Baldwin, B. G., Bank, R., Barber, A., Barbosa, J. P., Baron-Szabo, R. C., Bartsch, I., Bellan-Santini, D., Bergh, N., Berta, A., Bezerra, T. N., Blanco, S., Blasco-Costa, I., Blazewicz, M., Błędzki, L. A., Bock, P., Bonifacino, M., Böttger-Schnack, R., Bouchet, P., Boury-Esnault, N., Bouzan, R., Boxshall, G., Bray, R., Bruce, N. L., Bruneau, A., Budaeva, N., Bueno-Villegas, J., Cairns, S., Calvo Casas, J., Camilo de Oliveira, J. P., Carballo, J. L., Cárdenas, P., Carstens, E., Chan, B. K., Chan, T. Y., Cheng, L., Choong, H., Christenhusz, M., Churchill, M., Collins, A. G., Collins, G. E., Collins, K., Consorti, L., Copilaș-Ciocianu, D., Corbari, L., Cordeiro, R., Cornils, A., Costa Corgosinho, P. H., Coste, M., Costello, M. J., Crandall, K. A., Cremonte, F., Cribb, T., Cutmore, S., Dahdouh-Guebas, F., Daneliya, M., Dauvin, J. C., Davie, P., De Broyer, C., de Lima Ferreira, P., de Mazancourt, V., de Voogd, N. J., Decker, P., Defaye, D., Dekker, H., d'Hondt, J. L., et al.: World Register of Marine Species (WoRMS), 2023.

Archambault, P., Snelgrove, P. V. R., Fisher, J. A. D., Gagnon, J.-M., Garbary, D. J., Harvey, M., Kenchington, E. L., Lesage, V., Levesque, M., Lovejoy, C., Mackas, D. L., McKindsey, C. W., Nelson, J. R., Pepin, P., Piché, L., and Poulin, M.: From Sea to Sea: Canada's Three Oceans of Biodiversity, PLoS ONE, 5, e12182, https://doi.org/10.1371/journal.pone.0012182, 2010.

Ardyna, M. and Arrigo, K. R.: Phytoplankton dynamics in a changing Arctic Ocean, Nature Climate Change, 10, 892–903, https://doi.org/10.1038/s41558-020-0905-y, 2020.

581   Ardyna, M., Babin, M., Gosselin, M., Devred, E., Rainville, L., and Tremblay, J.-É.: Recent Arctic Ocean sea ice loss triggers
582   novel fall phytoplankton blooms, Geophysical Research Letters, 41, 6207–6212, https://doi.org/10.1002/2014GL061047,
583   2014.

584   Ardyna, M., Babin, M., Devred, E., Forest, A., Gosselin, M., Raimbault, P., and Tremblay, J.-É.: Shelf-basin gradients shape
585   ecological phytoplankton niches and community composition in the coastal Arctic Ocean (Beaufort Sea), Limnology and
586   Oceanography, 62, 2113–2132, https://doi.org/10.1002/lno.10554, 2017.

587   Baggesen, C., Moestrup, Ø., Daugbjerg, N., Krock, B., Cembella, A. D., and Madsen, S.: Molecular phylogeny and toxin
588   profiles of *Alexandrium tamarense* (Lebour) Balech (*Dinophyceae*) from the west coast of Greenland, Harmful Algae, 19,
589   108–116, https://doi.org/10.1016/j.hal.2012.06.005, 2012.

590   Barnes, R. and Sahr, K.: dggridR: Discrete Global Grids, 2020.

591   Baselga, A.: Partitioning the turnover and nestedness components of beta diversity, Global Ecology and Biogeography, 19,
592   134–143, https://doi.org/10.1111/j.1466-8238.2009.00490.x, 2010.

593   Baselga, A. and Orme, C. D. L.: betapart: an R package for the study of beta diversity, Methods in Ecology and Evolution, 3,
594   808–812, https://doi.org/10.1111/j.2041-210X.2012.00224.x, 2012.

595   Bates, S., Beach, D., Comeau, L., Haigh, N., Lewis, N., Locke, A., Martin, J., McCarron, P., Mckenzie, C., Michel, C., Miles,
596   C., Poulin, M., Quilliam, M., Rourke, W., Scarratt, M., Starr, M., and Wells, T.: Marine harmful algal blooms and phycotoxins
597   of concern to Canada., Canadian Technical Report of Fisheries and Aquatic Sciences, 3384, x + 322, 2020.

598   Bates, S. S., Hubbard, K. A., Lundholm, N., Montresor, M., and Leaw, C. P.: *Pseudo-nitzschia*, *Nitzschia*, and domoic acid:
599   New research since 2011, Harmful Algae, 79, 3–43, https://doi.org/10.1016/j.hal.2018.06.001, 2018.

600   Bates, S. S., Lundholm, N., Hubbard, K. A., Montresor, M., and Leaw, C. P.: Toxic and Harmful Marine Diatoms, in: Diatoms:
601   Fundamentals and Applications, John Wiley & Sons, Ltd, 389–434, https://doi.org/10.1002/9781119370741.ch17, 2019.

602   Bendif, E. M., Nevado, B., Wong, E. L. Y., Hagino, K., Probert, I., Young, J. R., Rickaby, R. E. M., and Filatov, D. A.:
603   Repeated species radiations in the recent evolution of the key marine phytoplankton lineage *Gephyrocapsa*, Nature
604   Communications, 10, 4234, https://doi.org/10.1038/s41467-019-12169-7, 2019.

605   Bendif, E. M., Probert, I., Archontikis, O. A., Young, J. R., Beaufort, L., Rickaby, R. E., and Filatov, D.: Rapid diversification
606   underlying the global dominance of a cosmopolitan phytoplankton, The ISME Journal, 17, 630–640,
607   https://doi.org/10.1038/s41396-023-01365-5, 2023.

608   Benedetti, F., Vogt, M., Elizondo, U. H., Righetti, D., Zimmermann, N. E., and Gruber, N.: Major restructuring of marine
609   plankton assemblages under global warming, Nature Communications, 12, 5226, https://doi.org/10.1038/s41467-021-25385-
610   x, 2021.

611 Bérard-Therriault, L., Poulin, M., and Bossé, L.: Guide d'identification du phytoplancton marin de l'estuaire et du golfe du
612 Saint-Laurent: incluant également certains protozoaires, Presses scientifiques du CNRC, Ottawa, Ont., 387 pp., 1999.

613 Blais, M., Ardyna, M., Gosselin, M., Dumont, D., Bélanger, S., Tremblay, J.-É., Gratton, Y., Marchese, C., and Poulin, M.:
614 Contrasting interannual changes in phytoplankton productivity and community structure in the coastal Canadian Arctic Ocean,
615 Limnology and Oceanography, 62, 2480–2497, https://doi.org/10.1002/lno.10581, 2017.

616 Bruhn, C. S., Wohlrab, S., Krock, B., Lundholm, N., and John, U.: Seasonal plankton succession is in accordance with
617 phycotoxin   occurrence   in   Disko   Bay,   West   Greenland,   Harmful   Algae,   103,   101978,
618 https://doi.org/10.1016/j.hal.2021.101978, 2021.

619 Carmack, E. C., Yamamoto-Kawai, M., Haine, T. W. N., Bacon, S., Bluhm, B. A., Lique, C., Melling, H., Polyakov, I. V.,
620 Straneo, F., Timmermans, M. L., and Williams, W. J.: Freshwater and its role in the Arctic Marine System: Sources,
621 disposition, storage, export, and physical and biogeochemical consequences in the Arctic and global oceans, Journal of
622 Geophysical Research: Biogeosciences, 121, 675–717, https://doi.org/10.1002/2015JG003140, 2016.

623 Chamberlain, S. and Schepers, L.: mregions: Marine Regions Data from "Marineregions.org," 2021.

624 Chamberlain, S. and Vanhoorne, B.: worrms: World Register of Marine Species (WoRMS) Client, 2023.

625 Chan, F. T., Stanislawczyk, K., Sneekes, A. C., Dvoretsky, A., Gollasch, S., Minchin, D., David, M., Jelmert, A., Albretsen,
626 J., and Bailey, S. A.: Climate change opens new frontiers for marine species in the Arctic: Current trends and future invasion
627 risks, Global Change Biology, 25, 25–38, https://doi.org/10.1111/gcb.14469, 2019.

628 Chao, A. and Shen, T.-J.: Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample,
629 Environmental and Ecological Statistics, 10, 429–443, https://doi.org/10.1023/A:1026096204727, 2003.

630 Codispoti, L. A., Kelly, V., Thessen, A., Matrai, P., Suttles, S., Hill, V., Steele, M., and Light, B.: Synthesis of primary
631 production in the Arctic Ocean: III. Nitrate and phosphate based estimates of net community production, Progress in
632 Oceanography, 110, 126–150, https://doi.org/10.1016/j.pocean.2012.11.006, 2013.

633 Dhifallah, F., Rochon, A., Simard, N., McKindsey, C. W., Gosselin, M., and Howland, K. L.: Dinoflagellate communities in
634 high-risk Canadian Arctic ports, Estuarine, Coastal and Shelf Science, 107731, https://doi.org/10.1016/j.ecss.2021.107731,
635 2021.

636 Edvardsen, B. and Imai, I.: The ecology of harmful flagellates within Prymnesiophyceae and Raphidophyceae, in: Ecology of
637 Harmful Algae, edited by: Granéli, E. and Turner, J. T., Springer, Berlin, Heidelberg, 67–79, https://doi.org/10.1007/978-3-
638 540-32210-8_6, 2006.

639 Elferink, S., Neuhaus, S., Wohlrab, S., Toebe, K., Voß, D., Gottschling, M., Lundholm, N., Krock, B., Koch, B. P., Zielinski,
640 O., Cembella, A., and John, U.: Molecular diversity patterns among various phytoplankton size-fractions in West Greenland

641 in late summer, Deep Sea Research Part I: Oceanographic Research Papers, 121, 54–69,
642 https://doi.org/10.1016/j.dsr.2016.11.002, 2017.

643 Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W.,
644 Veneziano, D., Vera, C. S., Vrugt, J. A., and Martiny, A. C.: Present and future global distributions of the marine Cyanobacteria
645 *Prochlorococcus* and *Synechococcus*, PNAS, 110, 9824–9829, https://doi.org/10.1073/pnas.1307701110, 2013.

646 Freyria, N. J., Joli, N., and Lovejoy, C.: A decadal perspective on north water microbial eukaryotes as Arctic Ocean sentinels,
647 Scientific Reports, 11, 8413, https://doi.org/10.1038/s41598-021-87906-4, 2021.

648 Gao, C., Lin, S., Chen, M., Hong, J., and Liu, C.: Prevalence of phycotoxin contamination in shellfish from the Northern
649 Bering Sea and the Chukchi Sea, Toxicon, 167, 76–81, https://doi.org/10.1016/j.toxicon.2019.06.001, 2019.

650 AlgaeBase. World-wide electronic publication: https://www.algaebase.org/about/, last access: 22 June 2023.

651 Guiry, M. D., Moestrup, Ø., and Andersen, R. A.: Validation of the phylum name *Heterokontophyta*, Notulae Algarum, 1–5,
652 2023.

653 Hallegraeff, G. M., Anderson, D. M., Belin, C., Bottein, M.-Y. D., Bresnan, E., Chinain, M., Enevoldsen, H., Iwataki, M.,
654 Karlson, B., McKenzie, C. H., Sunesen, I., Pitcher, G. C., Provoost, P., Richardson, A., Schweibold, L., Tester, P. A., Trainer,
655 V. L., Yñiguez, A. T., and Zingone, A.: Perceived global increase in algal blooms is attributable to intensified monitoring and
656 emerging bloom impacts, Communications Earth & Environment, 2, 1–10, https://doi.org/10.1038/s43247-021-00178-8, 2021.

657 Hanna, E., Nolan, J. E., Overland, J. E., and Hall, R. J.: Climate Change in the Arctic, in: Arctic Ecology, John Wiley & Sons,
658 Ltd, 57–79, https://doi.org/10.1002/9781118846582.ch3, 2021.

659 Hill, V. J., Matrai, P. A., Olson, E., Suttles, S., Steele, M., Codispoti, L. A., and Zimmerman, R. C.: Synthesis of integrated
660 primary production in the Arctic Ocean: II. In situ and remotely sensed estimates, Progress in Oceanography, 110, 107–125,
661 https://doi.org/10.1016/j.pocean.2012.11.005, 2013.

662 Hubbard, K. A., Villac, M. C., Chadwick, C., DeSmidt, A. A., Flewelling, L., Granholm, A., Joseph, M., Wood, T., Fachon,
663 E., Brosnahan, M. L., Richlen, M., Pathare, M., Stockwell, D., Lin, P., Bouchard, J. N., Pickart, R., and Anderson, D. M.:
664 Spatiotemporal transitions in *Pseudo-nitzschia* species assemblages and domoic acid along the Alaska coast, PLOS ONE, 18,
665 e0282794, https://doi.org/10.1371/journal.pone.0282794, 2023.

666 Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., Coelho, L. P., Endo, H., Gasol, J. M., Gregory,
667 A. C., Mahé, F., Rigonato, J., Royo-Llonch, M., Salazar, G., Sanz-Sáez, I., Scalco, E., Soviadan, D., Zayed, A. A., Zingone,
668 A., Labadie, K., Ferland, J., Marec, C., Kandels, S., Picheral, M., Dimier, C., Poulain, J., Pisarev, S., Carmichael, M., Pesant,
669 S., Acinas, S. G., Babin, M., Bork, P., Boss, E., Bowler, C., Cochrane, G., de Vargas, C., Follows, M., Gorsky, G., Grimsley,
670 N., Guidi, L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H., Pesant, S.,

671 Poulton, N., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M. B., Sunagawa, S., Wincker, P., Babin, M., Boss, E.,
672 Iudicone, D., Jaillon, O., Acinas, S. G., Ogata, H., Pelletier, E., Stemmann, L., Sullivan, M. B., Sunagawa, S., Bopp, L., de
673 Vargas, C., Karp-Boss, L., Wincker, P., Lombard, F., Bowler, C., and Zinger, L.: Global Trends in Marine Plankton Diversity
674 across Kingdoms of Life, Cell, 179, 1084-1097.e21, https://doi.org/10.1016/j.cell.2019.10.008, 2019.

675 Joli, N., Gosselin, M., Ardyna, M., Babin, M., Onda, D. F., Tremblay, J.-É., and Lovejoy, C.: Need for focus on microbial
676 species following ice melt and changing freshwater regimes in a Janus Arctic Gateway, Sci Rep, 8, 1–11,
677 https://doi.org/10.1038/s41598-018-27705-6, 2018.

678 Kacimi, S. and Kwok, R.: Arctic snow depth, ice thickness, and volume from ICESat-2 and CryoSat: 2018–2021, Geophysical
679 Research Letters, 49, e2021GL097448, https://doi.org/10.1029/2021GL097448, 2022.

680 Kahru, M., Brotas, V., Manzano-Sarabia, M., and Mitchell, B. G.: Are phytoplankton blooms occurring earlier in the Arctic?,
681 Global Change Biology, 17, 1733–1739, https://doi.org/10.1111/j.1365-2486.2010.02312.x, 2011.

682 Lassus, P., Chomérat, N., Hess, P., and Nézan, E.: Toxic and harmful microalgae of the World Ocean / Micro-algues toxiques
683 et nuisibles de l'océan mondial, International Society for the Study of Harmful Algae / Intergovernmental Oceanographic
684 Commission of UNESCO., Denmark, 523 pp., 2016.

685 Li, A., Chen, H., Qiu, J., Lin, H., and Gu, H.: Determination of multiple toxins in whelk and clam samples collected from the
686 Chukchi and Bering seas, Toxicon, 109, 84–93, https://doi.org/10.1016/j.toxicon.2015.11.016, 2016.

687 Li, W. K. W., McLaughlin, F. A., Lovejoy, C., and Carmack, E. C.: Smallest Algae Thrive As the Arctic Ocean Freshens,
688 Science, 326, 539–539, https://doi.org/10.1126/science.1179798, 2009.

689 Li, X., Yan, T., Yu, R., and Zhou, M.: A review of *Karenia mikimotoi*: Bloom events, physiology, toxicity and toxic
690 mechanism, Harmful Algae, 90, 101702, https://doi.org/10.1016/j.hal.2019.101702, 2019.

691 Lovejoy, C., von Quillfeldt, C., Hopcroft, R. R., Poulin, M., Thaler, M., Arendt, K., Debes, H., Gíslason, Á., and Kosobokova,
692 K.: State of the Arctic Marine Biodiversity Report: chapter 3.2: Plankton, in: State of the Arctic Marine Biodiversity Report,
693 62–83, 2017.

694 Lundholm, N., Churro, C., Escalera, L., Fraga, S., Hoppenrath, M., Iwataki, M., Larsen, J., Mertens, K., Moestrup, Ø., and
695 Zingone, A.: IOC-UNESCO Taxonomic Reference List of Harmful Micro Algae. Accessed at
696 https://www.marinespecies.org/hab, https://doi.org/10.14284/362, 2009.

697 Martin, J., Tremblay, J.-É., and Price, N. M.: Nutritive and photosynthetic ecology of subsurface chlorophyll maxima in
698 Canadian Arctic waters, Biogeosciences, 9, 5353–5371, https://doi.org/10.5194/bg-9-5353-2012, 2012.

699 McKenzie, C. H., Bates, S. S., Martin, J. L., Haigh, N., Howland, K. L., Lewis, N. I., Locke, A., Peña, A., Poulin, M., Rochon,
700 A., Rourke, W. A., Scarratt, M. G., Starr, M., and Wells, T.: Three decades of Canadian marine harmful algal events:

701 Phytoplankton and phycotoxins of concern to human and ecosystem health, Harmful Algae, 101852,
702 https://doi.org/10.1016/j.hal.2020.101852, 2020.

703 Mehdizadeh Allaf, M.: *Heterosigma akashiwo*, a Fish-Killing Flagellate, Microbiology Research, 14, 132–147,
704 https://doi.org/10.3390/microbiolres14010012, 2023.

705 Meredith, M., Sommerkorn, M., Cassotta, C., Derksen, A., Ekaykin, A., Hollowed, A., Kofinas, A., Mackintosh, A.,
706 Melbourne-Thomas, J., Muelbert, M. M. C., Ottersen, G., Pritchard, H., and Schuur, E. A. G.: Polar Regions (chapitre 28), in:
707 IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by: Pörtner, Hans-Otto, D.C. Roberts, V.
708 Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama,
709 N.M. Weyer, Cambridge University Press, 2019.

710 Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M. R., Kale, V., Potter, S. C.,
711 Richardson, L. J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., and Finn,
712 R. D.: MGnify: the microbiome analysis resource in 2020, Nucleic Acids Research, 48, D570–D578,
713 https://doi.org/10.1093/nar/gkz1035, 2020.

714 Morquecho, L.: Pyrodinium bahamense One the Most Significant Harmful Dinoflagellate in Mexico, Front. Mar. Sci., 6,
715 https://doi.org/10.3389/fmars.2019.00001, 2019.

716 Neukermans, G., Oziel, L., and Babin, M.: Increased intrusion of warming Atlantic water leads to rapid expansion of temperate
717 phytoplankton in the Arctic, Global Change Biology, 24, 2545–2553, https://doi.org/10.1111/gcb.14075, 2018.

718 Nohe, A., Goffin, A., Tyberghein, L., Lagring, R., De Cauwer, K., Vyverman, W., and Sabbe, K.: Marked changes in diatom
719 and dinoflagellate biomass, composition and seasonality in the Belgian Part of the North Sea between the 1970s and 2000s,
720 Science of The Total Environment, 716, 136316, https://doi.org/10.1016/j.scitotenv.2019.136316, 2020.

721 Nöthig, E.-M., Bracher, A., Engel, A., Metfies, K., Niehoff, B., Peeken, I., Bauerfeind, E., Cherkasheva, A., Gäbler-Schwarz,
722 S., Hardge, K., Kilias, E., Kraft, A., Mebrahtom Kidane, Y., Lalande, C., Piontek, J., Thomisch, K., and Wurst, M.:
723 Summertime plankton ecology in Fram Strait—a compilation of long- and short-term observations, Polar Research, 34, 23349,
724 https://doi.org/10.3402/polar.v34.23349, 2015.

725 Okolodkov, Y. B. and Dodge, J. D.: Biodiversity and biogeography of planktonic dinoflagellates in the Arctic Ocean, Journal
726 of Experimental Marine Biology and Ecology, 202, 19–27, https://doi.org/10.1016/0022-0981(96)00028-7, 1996.

727 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G.
728 L., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H.: vegan: Community Ecology Package, 2020.

729 Olsen, L. M., Duarte, P., Peralta-Ferriz, C., Kauko, H. M., Johansson, M., Peeken, I., Różańska-Pluta, M., Tatarek, A., Wiktor,
730 J., Fernández-Méndez, M., Wagner, P. M., Pavlov, A. K., Hop, H., and Assmy, P.: A red tide in the pack ice of the Arctic
731 Ocean, Scientific Reports, 9, 9536, https://doi.org/10.1038/s41598-019-45935-0, 2019.

732 Oziel, L., Baudena, A., Ardyna, M., Massicotte, P., Randelhoff, A., Sallée, J.-B., Ingvaldsen, R. B., Devred, E., and Babin,
733 M.: Faster Atlantic currents drive poleward expansion of temperate phytoplankton in the Arctic Ocean, Nature
734 Communications, 11, 1705, https://doi.org/10.1038/s41467-020-15485-5, 2020.

735 Paulsen, M. L., Doré, H., Garczarek, L., Seuthe, L., Müller, O., Sandaa, R.-A., Bratbak, G., and Larsen, A.: *Synechococcus* in
736 the Atlantic Gateway to the Arctic Ocean, Frontiers in Marine Science, 3, 191, https://doi.org/10.3389/fmars.2016.00191,
737 2016.

738 Poulin, M., Daugbjerg, N., Gradinger, R., Ilyash, L., Ratkova, T., and von Quillfeldt, C.: The pan-Arctic biodiversity of marine
739 pelagic and sea-ice unicellular eukaryotes: a first-attempt assessment, Marine Biodiversity, 41, 13–28,
740 https://doi.org/10.1007/s12526-010-0058-8, 2011.

741 Pućko, M., Dionne, K., and Michel, C.: Occurrence of toxin-producing marine algae in the Canadian Arctic and adjacent
742 waters, Canadian Manuscript Report Fisheries And Aquatic Sciences 3180., 2019.

743 Pućko, M., Rourke, W., Hussherr, R., Archambault, P., Eert, J., Majewski, A. R., Niemi, A., Reist, J., and Michel, C.:
744 Phycotoxins in bivalves from the western Canadian Arctic: The first evidence of toxigenicity., Harmful Algae, 102474,
745 https://doi.org/10.1016/j.hal.2023.102474, 2023.

746 Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.:
747 The Arctic has warmed nearly four times faster than the globe since 1979, Communications Earth & Environment, 3, 1–10,
748 https://doi.org/10.1038/s43247-022-00498-3, 2022.

749 Reguera, B., Velo-Suárez, L., Raine, R., and Park, M. G.: Harmful *Dinophysis* species: A review, Harmful Algae, 14, 87–106,
750 https://doi.org/10.1016/j.hal.2011.10.016, 2012.

751 Reid, P. C., Johns, D. G., Edwards, M., Starr, M., Poulin, M., and Snoeijs, P.: A biological consequence of reducing Arctic ice
752 cover: arrival of the Pacific diatom Neodenticula seminae in the North Atlantic for the first time in 800 000 years, Global
753 Change Biology, 13, 1910–1921, https://doi.org/10.1111/j.1365-2486.2007.01413.x, 2007.

754 Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., Lindley, J. A., Sims, D. W., Stevens, D., and Witt, M.: Using
755 continuous plankton recorder data, Progress in Oceanography, 68, 27–74, https://doi.org/10.1016/j.pocean.2005.09.011, 2006.

756 Righetti, D., Vogt, M., Gruber, N., Psomas, A., and Zimmermann, N. E.: Global pattern of phytoplankton diversity driven by
757 temperature and environmental variability, Science Advances, 5, eaau6253, https://doi.org/10.1126/sciadv.aau6253, 2019.

758 Righetti, D., Vogt, M., Zimmermann, N. E., Guiry, M. D., and Gruber, N.: PhytoBase: A global synthesis of open-ocean
759 phytoplankton occurrences, Earth System Science Data, 12, 907–933, https://doi.org/10.5194/essd-12-907-2020, 2020.

760 Rösel, A., Kaleschke, L., and Birnbaum, G.: Melt ponds on Arctic sea ice determined from MODIS satellite data using an
761 artificial neural network, The Cryosphere, 6, 431–446, https://doi.org/10.5194/tc-6-431-2012, 2012.

762 Rudels, B. and Carmack, E.: Arctic Ocean Water Mass Structure and Circulation, Oceanography,
763 https://doi.org/10.5670/oceanog.2022.116, 2022.

764 Sakshaug, E.: Primary and Secondary Production in the Arctic Seas, in: The Organic Carbon Cycle in the Arctic Ocean, edited
765 by: Stein, R. and MacDonald, R. W., Springer Berlin Heidelberg, Berlin, Heidelberg, 57–81, https://doi.org/10.1007/978-3-
766 642-18912-8_3, 2004.

767 Schiffrine, N., Dhifallah, F., Dionne, K., Poulin, M., Lessard, S., Rochon, A., and Gosselin, M.: Microbial plankton occurrence
768 database in the North American Arctic region: synthesis of recent diversity of potentially toxic and harmful algae – Code and
769 Dataset, https://doi.org/10.5281/zenodo.10498859, 2024.

770 Serreze, M. C. and Stroeve, J.: Arctic sea ice trends, variability and implications for seasonal ice forecasting, Philosophical
771 Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, 20140159,
772 https://doi.org/10.1098/rsta.2014.0159, 2015.

773 Sournia, A.: Phytoplankton Manual, Unesco, 1978.

774 Stroeve, J. C., Serreze, M. C., Holland, M. M., Kay, J. E., Malanik, J., and Barrett, A. P.: The Arctic's rapidly shrinking sea
775 ice cover: A research synthesis, Climatic Change, 110, 1005–1027, https://doi.org/10.1007/s10584-011-0101-1, 2012.

776 Team, R. D. C. and R Development Core Team, R.: R: A Language and Environment for Statistical Computing, R Foundation
777 for Statistical Computing, Vienna, Austria, 409 pp., https://doi.org/10.1007/978-3-540-74686-7, 2019.

778 Thompson, G. G. and Withers, P. C.: Effect of species richness and relative abundance on the shape of the species accumulation
779 curve, Austral Ecology, 28, 355–360, https://doi.org/10.1046/j.1442-9993.2003.01294.x, 2003.

780 Tillmann, U., Elbrächter, M., Krock, B., John, U., and Cembella, A.: Azadinium spinosum gen. et sp. nov. (Dinophyceae)
781 identified as a primary producer of azaspiracid toxins, European Journal of Phycology, 44, 63–79,
782 https://doi.org/10.1080/09670260802578534, 2009.

783 Torres-Valdés, S., Tsubouchi, T., Bacon, S., Naveira-Garabato, A. C., Sanders, R., McLaughlin, F. A., Petrie, B., Kattner, G.,
784 Azetsu-Scott, K., and Whitledge, T. E.: Export of nutrients from the Arctic Ocean, Journal of Geophysical Research: Oceans,
785 118, 1625–1644, https://doi.org/10.1002/jgrc.20063, 2013.

786 Tremblay, J.-É., Anderson, L. G., Matrai, P., Coupel, P., Bélanger, S., Michel, C., and Reigstad, M.: Global and regional
787 drivers of nutrient supply, primary production and CO2 drawdown in the changing Arctic Ocean, Progress in Oceanography,
788 139, 171–196, https://doi.org/10.1016/j.pocean.2015.08.009, 2015.

789 Vavrek, M. J.: fossil: palaeoecological and palaeogeographical analysis tools, Palaeontologia Electronica, 14, 1T, 2011.

790 Vihtakari, M.: ggOceanMaps: Plot Data on Oceanographic Maps using "ggplot2," 2021.

791 Vincent, W. F.: Cyanobacterial Dominance in the Polar Regions, in: The Ecology of Cyanobacteria: Their Diversity in Time
792 and Space, edited by: Whitton, B. A. and Potts, M., Springer Netherlands, Dordrecht, 321–340, https://doi.org/10.1007/0-306-
793 46855-7_12, 2002.

794 Whittaker, R. H.: Evolution and Measurement of Species Diversity, TAXON, 21, 213–251, https://doi.org/10.2307/1218190,
795 1972.

796 Yamamoto-Kawai, M., McLaughlin, F. A., Carmack, E. C., Nishino, S., Shimada, K., and Kurita, N.: Surface freshening of
797 the Canada Basin, 2003–2007: River runoff versus sea ice meltwater, J. Geophys. Res., 114, C00A05,
798 https://doi.org/10.1029/2008JC005000, 2009.

799