



OPEN ACCESS

EDITED BY

Xianbiao Lin,
Ocean University of China, China

REVIEWED BY

Genmei Lin,
Sun Yat-sen University, China
Prashasti Singh,
University of Tasmania, Australia
Clarence Wei Hung Sim,
Nanyang Technological University,
Singapore

*CORRESPONDENCE

Fraser Kennedy
✉ f.c.kennedy@utas.edu.au

RECEIVED 27 November 2025

REVISED 04 February 2026

ACCEPTED 27 February 2026

PUBLISHED 23 April 2026

CITATION

Zhai C, Kennedy F, Han M, Yu H, Sun J,
Liang Y, Wang M and McMinn A (2026)
Comparative genomics of an Antarctic
sea-ice diatom (*Nitzschia* sp.) provides
insights into potential polar adaptation.
Front. Microbiol. 17:1755917.
doi: 10.3389/fmicb.2026.1755917

COPYRIGHT

© 2026 Zhai, Kennedy, Han, Yu, Sun,
Liang, Wang and McMinn. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Comparative genomics of an Antarctic sea-ice diatom (*Nitzschia* sp.) provides insights into potential polar adaptation

Chuan Zhai^{1,2}, Fraser Kennedy^{1,2*}, Meiaoxue Han³, Hao Yu³,
Jianhua Sun³, Yantao Liang^{3,4}, Min Wang^{3,4,5} and
Andrew McMinn^{1,2,3}

¹Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS, Australia, ²The Australian Centre for Excellence in Antarctic Science, University of Tasmania, Hobart, TAS, Australia, ³Key Lab of Polar Oceanography and Global Ocean Change, College of Marine Life Sciences, Institute of Evolution and Marine Biodiversity, Frontiers Science Center for Deep Ocean Multispheres and Earth System, Ocean University of China, Qingdao, China, ⁴UMT-OUC Joint Centre for Marine Studies, Qingdao, China, ⁵Haide College, Ocean University of China, Qingdao, China

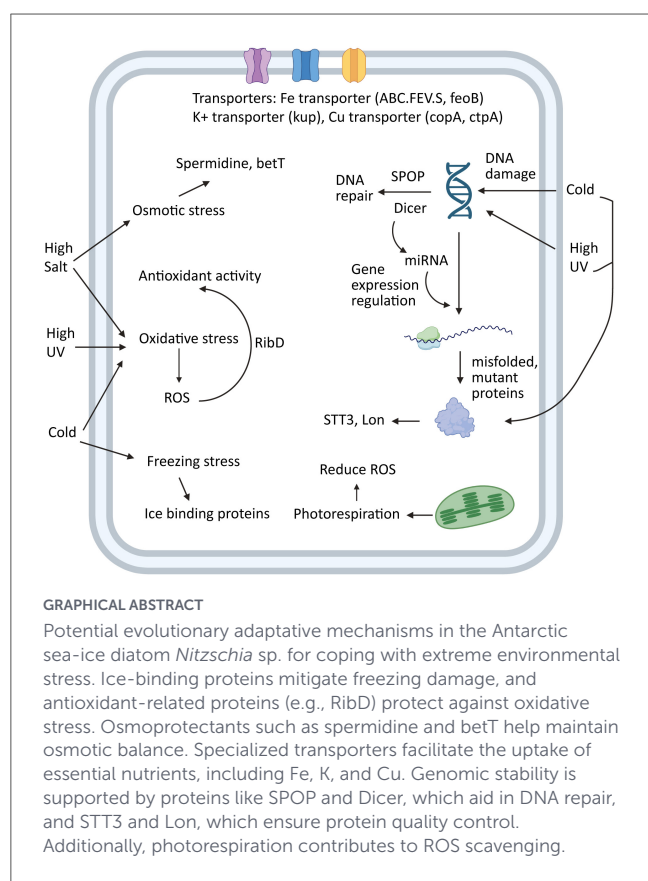
Sea-ice diatoms are key primary producers in Antarctic sea-ice ecosystems and the Southern Ocean. However, the genomic features that characterize these organisms remain incompletely understood. Here, we present a high-quality draft genome assembly of an Antarctic sea-ice diatom, *Nitzschia* sp. (most likely *Nitzschia stellata*), and examine its genomic characteristics using comparative analyses. We found that *Nitzschia* sp. and *Fragilariopsis cylindrus*, another Antarctic sea-ice diatom, share numerous orthologous genes that are absent from temperate diatoms. These shared genes are associated with biological processes such as oxidative stress response, DNA repair, protein quality control, osmotic regulation, and nutrient acquisition and transport—functions that may be relevant for coping with the combined stresses of the sea-ice environment. In addition, *Nitzschia* sp. shows contraction in several gene families, including those involved in proprotein convertase activity, clathrin-mediated endocytosis, BTB/POZ domain-containing proteins, solute carrier transporters, and voltage-gated potassium channels. We also identified a high-quality endogenous viral element (EVE) related to the viral family *Metaviridae*, as well as a high proportion of transposable elements (TEs), accounting for 27.25% of the genome assembly. Together, our results characterize the genomic features of *Nitzschia* sp. and highlight potential genomic strategies that may contribute to survival in sea-ice environments.

KEYWORDS

adaptation, diatom, EVE, genomic, polar, sea ice, TE

1 Introduction

Sea ice, covering approximately 7% of the global ocean, represents one of the Earth's largest biomes (Weeks, 2010). It serves as a critical habitat for many organisms, including primary producers that thrive either attached to the ice or within its brine inclusions. In the Antarctic, primary production within sea ice is estimated to range from 23 to 70 Tg C



per year, accounting for about 12% of the total production in the Antarctic sea-ice zone (Legendre et al., 1992; Arrigo et al., 1997; Saenz and Arrigo, 2014). Diatoms dominate these algal communities, forming the foundation of the Antarctic sea-ice ecosystem. These diatoms play an essential role in transferring energy and nutrients through Antarctic marine food webs, supporting diverse organisms, including krill—a keystone species in the Antarctic ecosystem (Schmidt et al., 2014). Moreover, diatoms significantly contribute to biogeochemical cycles via the export of organic carbon to deeper ocean layers during seasonal ice melt (Vancoppenolle et al., 2013; Duncan and Petrou, 2022).

The sea-ice environment is characterized by extreme and highly dynamic conditions, including sub-zero temperatures, elevated salinities, restricted gas exchange, and strongly fluctuating light regimes (Thomas and Dieckmann, 2002). To persist under these pressures, sea-ice algae have evolved diverse genomic and physiological strategies that enhance tolerance to cold, osmotic stress, and oxidative damage (Mock et al., 2017; Kennedy et al., 2019; Zhang et al., 2020; Zhang et al., 2023a). Genomic studies of Antarctic sea-ice green algae, such as *Microglona* sp. YARC and *Chlamydomonas* sp. ICE-L, have provided important insights into potential adaptive mechanisms, including gene family expansion, horizontal gene transfer, and enhanced capacities for photoprotection, lipid metabolism, and DNA repair (Zhang et al., 2020; Zhang et al., 2023a). However, these taxa are relatively minor components of Antarctic sea-ice algal assemblages. In contrast, diatoms dominate Antarctic sea-ice algal communities and represent some of the most ecologically abundant and productive microorganisms in polar oceans (Kang and Fryxell, 1992). Among them, *Fragilariopsis cylindrus* has emerged as a

model species, with genomic and transcriptomic analyses revealing extensive allelic divergence and environmentally responsive gene regulation under stressors such as iron limitation, prolonged darkness, freezing temperatures, and elevated CO₂ (Mock et al., 2017; Kennedy et al., 2019; Joli et al., 2024). These findings underscore the central role of diatoms in Antarctic sea ice and highlight the importance of expanding genomic investigations beyond a single model species.

Despite the ecological dominance of diatoms in Antarctic sea ice, our understanding of their genomic adaptation remains limited, with most available studies focused on *F. cylindrus*. Here, we present a high-quality metagenome-assembled genome (MAG) of *Nitzschia* sp. (most likely *Nitzschia stellata*), recovered from Antarctic fast ice in McMurdo Sound. *Nitzschia* species are well-documented as dominant members of bottom sea-ice diatom assemblages in this region, and *N. stellata* in particular has been repeatedly reported as one of the most abundant taxa in McMurdo Sound sea ice (Martin et al., 2012; McMinn et al., 2010; Kennedy et al., 2020). Consistent with these observations, *Nitzschia* sp. emerged as the most complete and abundant diatom MAG recovered from our sea-ice metagenome, motivating its selection for downstream genomic analyses. By integrating comparative genomic analyses of *Nitzschia* sp. and *F. cylindrus*, we aim to broaden the current genomic perspective on Antarctic sea-ice diatoms and to explore shared and lineage-specific genomic features that may underpin their persistence in polar sea-ice environments.

2 Materials and methods

2.1 Sampling, DNA extraction, sequencing

Sea-ice algae were sampled from the bottom 300 mm of 1.8 m thick land-fast (fast) ice at Cape Evans, McMurdo Sound, Antarctica (77°38'S, 166°24'E) in November 2022. Cape Evans lies on the western coast of Ross Island in southern McMurdo Sound, in a region characterized by seasonal fast ice and high ice-associated primary productivity. Fast ice at this site typically forms during austral autumn and persists through spring, providing a stable substratum for the development of dense under-ice algal communities. *N. stellata*, *Berkeleya adeliensis*, *Fragilariopsis* spp., *Navicula* spp. have been documented as dominant diatoms in bottom ice around McMurdo Sound (Martin et al., 2012; McMinn et al., 2010; Kennedy et al., 2020). At the time of our sampling, fast ice exhibited visible brownish ice algal biomass at the ice–water interface, indicative of substantial under-ice diatom communities. The site was selected as part of an ongoing McMurdo Sound sampling program and not exclusively based on *Nitzschia* dominance, although *Nitzschia* spp. are known to be common components of Antarctic bottom ice diatom assemblages in this region. Sea ice cores were extracted using a Mark V SIPRE corer (Kovaks, United States) and immediately transferred to a field laboratory, where they were melted naturally at 4°C overnight without the addition of seawater. The melted sea ice was sequentially filtered through a 20 µm mesh followed by 0.2 µm pore-size polycarbonate (PC) membrane filters (Millipore,

GTP14250) using a peristaltic pump. The initial 20 μm pre-filtration step was applied to remove large particles and debris prior to microbial collection. The 0.2 μm filters were flash-frozen in liquid nitrogen and stored at -80°C until DNA extraction. It should be noted that this filtration protocol was designed primarily for the collection of microbial communities and is therefore not optimal for comprehensive algal sampling, as the 20 μm pre-filtration step removes large diatom cells. In contrast, algal-focused sampling typically employs a size-fractionated filtration strategy, such as sequential filtration through a 200 μm mesh to remove zooplankton, followed by 20 μm filters to retain large algae and 3 μm filters to collect smaller algal cells. Total DNA on filters was extracted using the Water DNA Kit (Omega) according to the manufacturer's instructions. Library construction was performed using a NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA, United States), and high-throughput sequencing of the total DNA was conducted using the Illumina NovaSeq 6000 (paired-end sequencing, 2×150 bp) platform (Novogene Bioinformatics Technology Co., Ltd., Nanjing, China).

2.2 Diatom genome assembly and phylogenetics

The raw sequencing data were quality-controlled by the sequencing provider using Trimmomatic v0.36 (Bolger et al., 2014) and subsequently assembled using MEGAHIT v1.2.9 with the default k-mer series (Li et al., 2015). Contigs were classified using Tiara v1.0.3 with default parameters, a deep-learning-based tool designed to identify eukaryotic sequences in metagenomic datasets (Karlicki et al., 2022). Contigs classified as eukaryotic, organellar, or unknown were retained for subsequent binning. Binning was performed by metaBAT2, CONCOCT, and MaxBin2 in metaWRAP v1.3 (–concoct –maxbin2 –metabat2 –universal) (Uritskiy et al., 2018). Genome assembly completeness of bins was assessed using BUSCO v5.8.0 (Benchmarking Universal Single-Copy Orthologs) with parameters “–miniprot” and “–auto-lineage-euk” (Simão et al., 2015), and bins with genome completeness $> 90\%$ were initially classified as high quality. Among the three binning approaches, metaBAT2 produced the highest-quality bins for this eukaryotic metagenome, as indicated by higher BUSCO completeness scores. Bin dereplication was performed using dRep v3.4.2 (parameter: dRep compare) (Olm et al., 2017). High-quality diatom bins were subsequently curated and visualized using BlobToolKit on the Galaxy platform (Challis et al., 2020), during which contigs taxonomically inconsistent with the dominant lineage of the bin (i.e., diatoms) were removed to minimize contamination. The BlobToolKit workflow on Galaxy followed the tutorial provided at Galaxy Training Network. The taxonomy of the diatom genome was initially assigned using the Bin Annotation Tool v5.2.3 (BAT) on the Galaxy platform with default settings (Von Meijenfeldt et al., 2019), and a finer taxonomic resolution was subsequently obtained through phylogenetic tree reconstruction. The phylogenetic tree was constructed following a slightly modified version of the Snakemake workflow described by Manni et al. (2021). Briefly, first, reference genome assembly and corresponding GFF files of lineage-related diatoms (representative *Bacillariophyceae* diatoms)

were downloaded from NCBI datasets. Then, orthologous genes of these diatoms were identified using BUSCO, and proteins from each orthologous group were aligned with MAFFT (Katoh and Standley, 2013) and trimmed using trimAl (Capella-Gutiérrez et al., 2009). The resulting alignments were concatenated with AMAS (Borowiec, 2016), and a maximum-likelihood phylogeny was inferred from the concatenated superalignment using IQ-TREE 2 (Minh et al., 2020). In our modified workflow, parameters in the *config.yaml* file were adjusted as follows: “path_to_genomes” was set to “genomes,” “keep_duplicates” was set to “Yes,” and “dataset” was specified as “stramenopiles_odb10.” Given the high prevalence of duplicated genes in diatom genomes, duplicated BUSCO genes (“keep_duplicates” = “Yes”) were retained to ensure sufficient phylogenetic signal for tree reconstruction. All tools were run with the authors' default parameters, and the workflow was executed using the command “snakemake –cores 16 –use-conda.” The visualization of the phylogenetic tree was performed by iTOL (Letunic and Bork, 2024).

2.3 Gene annotation and gene family analysis

Protein-coding genes in the *Nitzschia* sp. genome were predicted using MAKER2 v2.31.11 on the Galaxy platform (Holt and Yandell, 2011; The Galaxy Community, 2024), following the protocol described at the Galaxy Training Network. Within the MAKER2 workflow, expressed sequence tags (ESTs) from the lineage-related diatom species (*F. cylindrus*) were downloaded from GenBank, and the protein reference database was obtained from OrthoDB v12 (Tegenfeldt et al., 2025). The MAKER2 workflow produced a final annotation in GFF format, from which exon, CDS, and protein sequences were extracted using GFFread (Pertea and Pertea, 2020). Exon sequences were used as input for BUSCO to assess the quality of the MAKER2 gene annotation, with BUSCO run in *transcriptome* mode using the *stramenopiles_odb10* dataset. Gene annotation quality was considered good when transcriptome BUSCO completeness was comparable to, or only slightly lower than, genome BUSCO completeness, indicating that the predicted gene models adequately captured the conserved gene content of the genome. Comparative genomic analyses were performed using OrthoVenn3, an integrated platform for whole-genome comparison, annotation, and evolutionary analysis (Sun et al., 2023). Eight *Bacillariophyceae* genomes were selected to enable a focused comparison between Antarctic sea-ice-associated diatoms and non-sea-ice diatoms from temperate or tropical environments. The two Antarctic sea-ice diatoms included were *Nitzschia* sp. (this study) and *F. cylindrus*. The six non-sea-ice diatoms (*Seminavis robusta*, *Nitzschia inconspicua*, *Cylindrotheca closterium*, *Fistulifera solaris*, *Phaeodactylum tricornutum*, and *Pseudo-nitzschia multistriata*) represent well-characterized model or reference diatom species isolated from coastal, benthic, or open-ocean environments in temperate or tropical regions. These species were selected based on the availability of high-quality reference proteomes, which are required for reliable orthologous gene clustering in OrthoVenn3. For all comparative species except *Nitzschia* sp., reference proteomes were obtained from UniProt and used as inputs for the OrthoVenn3 workflow. Within OrthoVenn3, orthologous clustering was performed with OrthoMCL with

inflation value of 1.5, phylogenetic reconstruction with FastTree2 under the LG + CAT model, and gene family expansion and contraction analyses with CAFE5, all using default parameters. Gene family annotation and functional prediction for shared unique, expanded, and contracted gene sets in *Nitzschia* sp. and *F. cylindrus* were conducted using InterPro and GhostKOALA (Blum et al., 2024; Kanehisa et al., 2016).

2.4 Endogenous viral elements and transposable elements

Endogenous viral elements (EVEs) in the *Nitzschia* sp. genome were analyzed using the “Viruses and Plasmids” workflow on the NMDC platform. This workflow detects viral and plasmid sequences in assembled scaffolds using geNomad (Camargo et al., 2024) and evaluates viral genome quality with CheckV (Nayfach et al., 2021). The workflow was executed using default parameters, and high-quality EVEs were defined as those with CheckV-estimated completeness > 90%. Predicted protein-coding genes within the high-quality EVE were initially generated by the NMDC workflow. To further ensure comprehensive gene prediction given the structural complexity and evolutionary divergence of EVEs, additional gene prediction was performed using Prodigal v2.6.3 (Hyatt et al., 2010) and NCBI ORF Finder. Both tools were run using genetic codes 1 (standard) and 11 (bacterial), allowing alternative start codons, with a minimum ORF length of 150 nt. This dual-genetic code approach was applied to account for potential heterogeneity in coding signals arising from viral origins and host integration. All predicted proteins were functionally annotated using InterPro (Blum et al., 2024), with all available member databases and sequence feature options enabled. The phylogenetic tree of the high-quality EVE was constructed using ViPTree, which infers proteomic trees based on normalized tBLASTx similarity scores (SG) and genomic distances (1-SG) using the BIONJ algorithm (Nishimura et al., 2017). For tree construction, ssRNA-RT reference viruses from the built-in ViPTree database, together with representative *Ortervirales* viruses downloaded from GenBank, were included to infer the phylogenetic placement of the high-quality EVE. The existence of long terminal repeats (LTR) in the high-quality EVE was checked by LTR_retriever v3.0.4 (Ou and Jiang, 2018). Transposable elements (TEs) within the eight diatom genomes were predicted by RepeatModeler v2.0.5 (Flynn et al., 2020). The classification and proportion of TEs in each diatom were identified and calculated using RepeatMasker v4.1.5. These tools were all run with default parameters.

3 Results

3.1 Diatom genome, phylogenetics, and genes

A draft MAG of a diatom was recovered from Antarctic sea-ice metagenomic data, yielding a total genome size of 50.9 Mb with an N50 of 11.5 kb and a GC content of 46.8% (Figure 1A).

BUSCO analysis indicated a 95% genome completeness, denoting a high-quality MAG assembly (Figure 1A). Taxonomic annotation using BAT classified this MAG within the family *Bacillariaceae*, and phylogenetic reconstruction placed it within the genus *Nitzschia* (Figure 2). Microscopic identification suggested the species was *N. stellata*, a dominant diatom found in our sea ice samples and in previous sea ice collections (Martin et al., 2012; Ryan et al., 2009; McMinn et al., 2010; Kennedy et al., 2020; Supplementary Table 1). *N. stellata* cells were observed as solitary or attached end-to-end, forming straight, zig-zag, or stellate colonies comprising 3–12 cells; valves were linear to lanceolate in valve view, tapering toward the poles, with an apical axis of 60–134 μm and a transapical axis of 6.5–10.5 μm (Figure 1B; Scott and Marchant, 2005). For clarity, this diatom MAG was hereafter referred to as *Nitzschia* sp. Gene prediction using MAKER2 identified 22,781 protein-coding genes in *Nitzschia* sp. A subsequent BUSCO assessment of the predicted transcripts confirmed a 95% completeness, supporting the accuracy of the gene annotation.

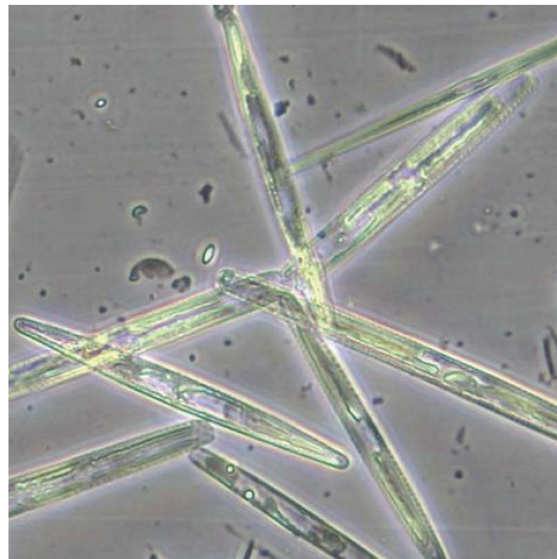
3.2 Comparative genomic analysis

Orthologous clustering identified 11,456 gene families in *Nitzschia* sp., including 621 gene families unique to this species when compared to seven other diatoms analyzed (Figure 3A). In comparison, *F. cylindrus* possesses 11,817 gene families, with 558 being species-specific (Figure 3A). Notably, *Nitzschia* sp. and *F. cylindrus* share 104 gene families absent in the other six non-polar species analyzed (Figure 3B). Based on KEGG annotations and manual literature searches, these unique genes are associated with diverse biological functions, including oxidative stress response, DNA repair, protein quality control, osmotic homeostasis, and nutrient uptake and transport (Table 1 and Supplementary Tables 2, 3).

The analysis of gene family expansions and contractions showed that *Nitzschia* sp. has 48 expanded and 126 contracted gene families (Figure 4). In contrast, *F. cylindrus* shows 53 expanded and 29 contracted gene families (Figure 4). The gene families expanded in both *Nitzschia* sp. and *F. cylindrus* are primarily associated with oxidative stress response, DNA repair, and protein quality control (Table 1). In *Nitzschia* sp., the expanded gene families additionally include those related to transcriptional and translational regulation, DNA repair, protein quality control, and osmotic homeostasis (Table 1 and Supplementary Tables 4, 5). Conversely, the contracted gene families in *Nitzschia* sp. encode proteins such as proprotein convertase subtilisin/kexin proteases, components of clathrin-mediated endocytosis, BTB/POZ domain-containing proteins, members of the solute carrier (SLC) family, and voltage-gated potassium channels (Table 2 and Supplementary Tables 6, 7).

3.3 Endogenous viral elements and transposable elements

Forty-six putative EVEs were identified within the *Nitzschia* sp. genome using geNomad, the majority of which were incomplete retroviral-like fragments. One high-quality EVE was identified within a 7,269 bp contig and was classified by CheckV as



Genome assembly and morphology of *Nitzschia* sp. (most likely *Nitzschia stellata*). **(A)** Snail plot of the *Nitzschia* sp. genome assembly (50.9 Mb). The longest scaffold is shown in red, while orange and light-orange arcs represent the N50 and N90 scaffold lengths, respectively. GC and AT contents distributions are depicted in dark blue and blue. BUSCO results in the upper right indicate genome completeness, including the proportions complete (Comp.), duplicated (Dupl.), fragmented (Frag.), and missing orthologs. **(B)** Light microscopy image of *N. stellata* cells. Cells occur either singly or attached end-to-end, forming straight, zig-zag, or stellate colonies of 3–12 cells. Valves are linear to lanceolate in valve view, tapering toward the poles, with an apical axis of 60–134 μm and a transapical axis of 6.5–10.5 μm .

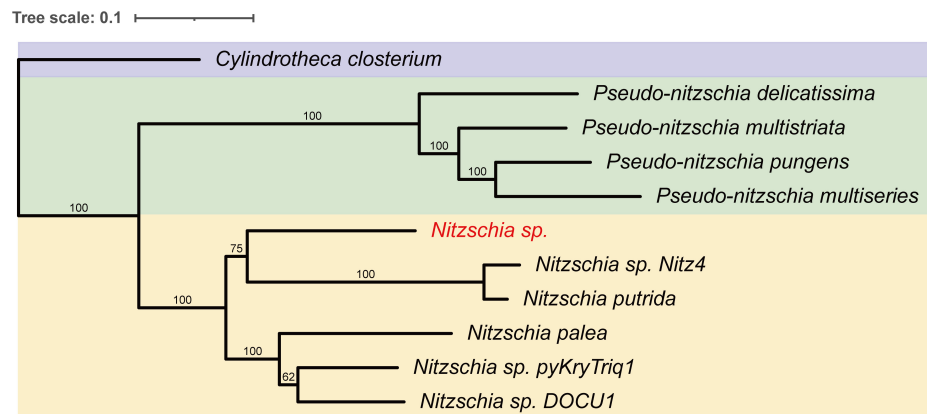


FIGURE 2

Maximum likelihood phylogenetic tree of the *Nitzschia* sp. metagenomic assembled genome (MAG) and ten additional *Bacillariaceae* diatom genomes, inferred from a concatenated alignment of the 24 orthologous genes. *Cylandrotheca closterium* was used as the outgroup species. Bootstrap support values are shown at each node. Clades are color-coded: purple for *Cylandrotheca*, green for *Pseudo-nitzschia*, and yellow for *Nitzschia*.

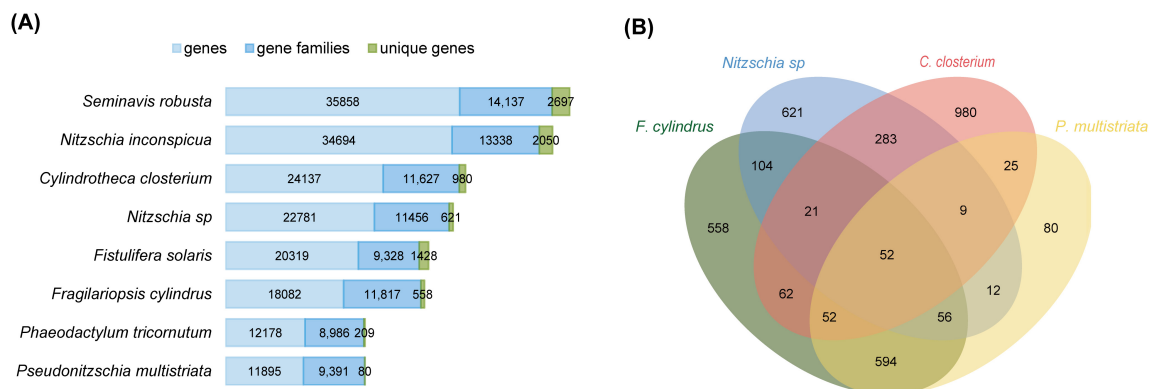


FIGURE 3

Comparative gene family analysis of *Nitzschia* sp. (A) Predicted gene counts, gene family numbers, and unique gene families for eight *Bacillariophyceae* diatom genomes. Sky blue bars represent the total number of predicted genes (annotated using MAKER2), blue bars denote the number of gene families, and green bars indicate unique gene families. Gene family inference was conducted using OrthoVenn3. (B) Numbers of shared and lineage specific gene families identified in four *Bacillariophyceae* diatoms. Gene family counts were inferred with OrthoVenn3. The Antarctic sea-ice isolates *Nitzschia* sp. and *F. cylindrus* are phylogenetically most closely related to *C. closterium* and *P. multistriata*, respectively.

high-quality with an estimated completeness of 100% (AAI-based, medium confidence). Gene predictions revealed that this element encodes a retrovirus-related Pol polyprotein containing conserved domains, including RVT_1 (reverse transcriptase), RT_RNaseH, RT_RNaseH_2, and Integrase_H2C2 (Figure 5A), which are essential and diagnostic components of reverse-transcribing viruses and retroelements. Phylogenetic analysis based on proteomic similarity placed this element within the *Metaviridae* family (i.e., Ty3/Gypsy-like LTR retroelements) (Figure 5B). Notably, although canonical *Metaviridae* elements typically encode gag/capsid proteins and long terminal repeats (LTRs), these features were not annotated in this EVE. Importantly, the CheckV completeness estimate here reflects the recovery of a near-complete Pol region, rather than the presence of all structural modules, which may be absent, highly diverged, or unrecognizable in endogenous viral elements. Furthermore, a high proportion of TEs (27.25% of the total genome size) was detected in *Nitzschia* sp., with LTR elements, particularly Ty1/Copia, being the most abundant

(Supplementary Table 8). Among the eight diatom genomes analyzed, the proportion of TEs varied considerably, with no clear correlation to geographical distribution (i.e., polar vs. non-polar) (Supplementary Table 8).

4 Discussion

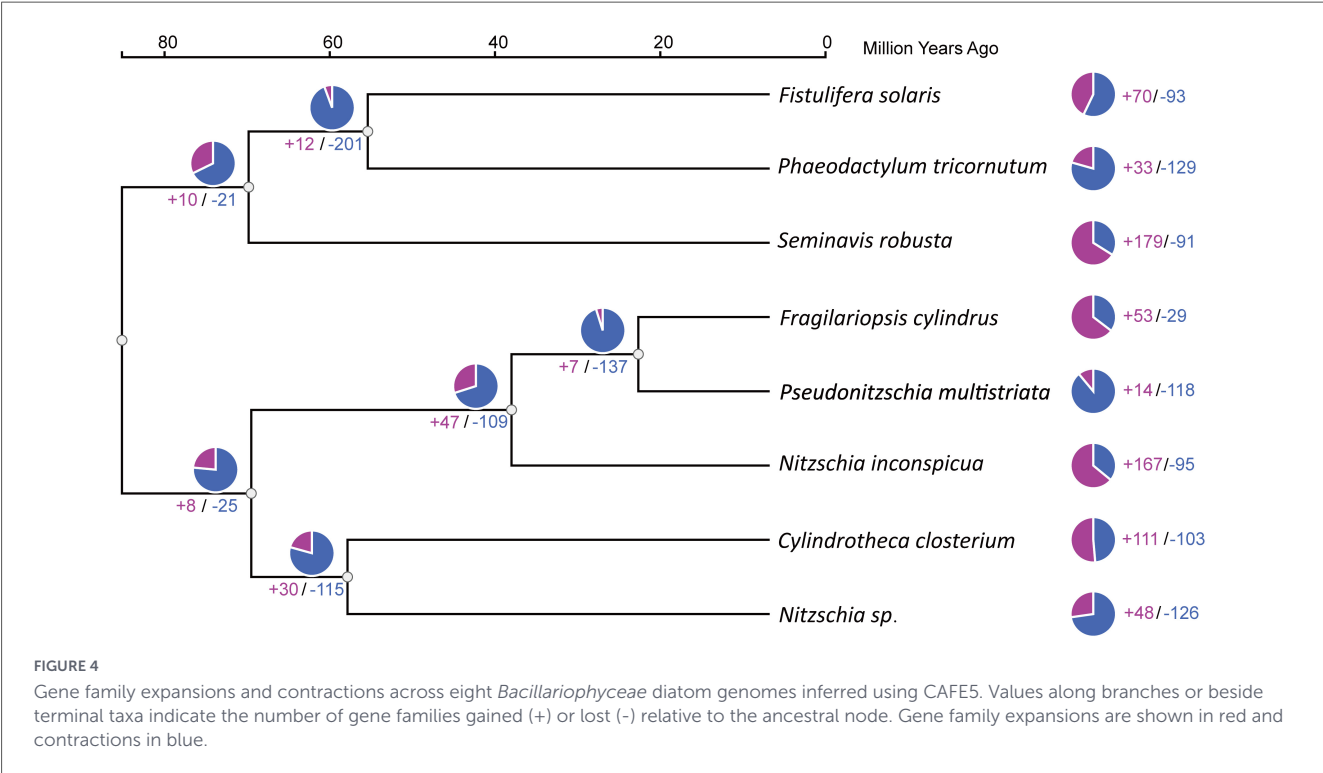
4.1 Genomic adaptations of two Antarctic sea-ice diatoms

The Antarctic sea ice environment is characterized by freezing temperatures, high salinity, and extreme light conditions, all of which impose oxidative, osmotic, and freezing stresses on resident microbial communities. Comparative genomic analyses here show that Antarctic sea-ice diatoms, including *Nitzschia* sp. and *F. cylindrus*, share a subset of expanded or unique gene

TABLE 1 Functional annotation of selected unique proteins shared between *Nitzschia* sp. and *F. cylindrus* (column 2), and selected expanded proteins in *Nitzschia* sp. (column 3).

Function	Unique proteins	Expanded proteins
Oxidative stress response	Spermidine synthase (K00797)	RibD* (K11752)
		SHMT* (K00600)
DNA repair, protein quality control	Chaperonin GroEL (K04077)	SMARCA3* (K15711)
	STT3 (K07151)	HMGB (K10802, K11295, K11296)
	SPOP (K10523)	IARS (K01870)
	Endoribonuclease Dicer (K11592)	ATP-dependent Lon protease* (K01338)
	SUPT16H (K25639)	ClpS (K06891)
Osmotic homeostasis	Spermidine synthase (K00797)	STP (K24193)
	Kup (K03549)	BetT (K02168)
Nutrient acquisition and transport	ABC.FEV.S (K02016)	STP (K24193)
	Kup (K03549)	Ferrous iron transport protein B (K04759)
Transcriptional and translational regulation	P-type Cu + transporter (K17686)	
	ABC-F (K06158)	HMGB (K10802, K11295, K11296)
	Endoribonuclease Dicer (K11592)	TUT (K13291)
	SUPT16H (K25639)	BED zinc finger (K24637)
Others	Ice-binding proteins (pfam: PF11999, PF20597)	SSRP1 (K09272)

Proteins marked with an asterisk (*) are also expanded in *F. cylindrus*. The corresponding KEGG or Pfam identifiers are provided for each protein.



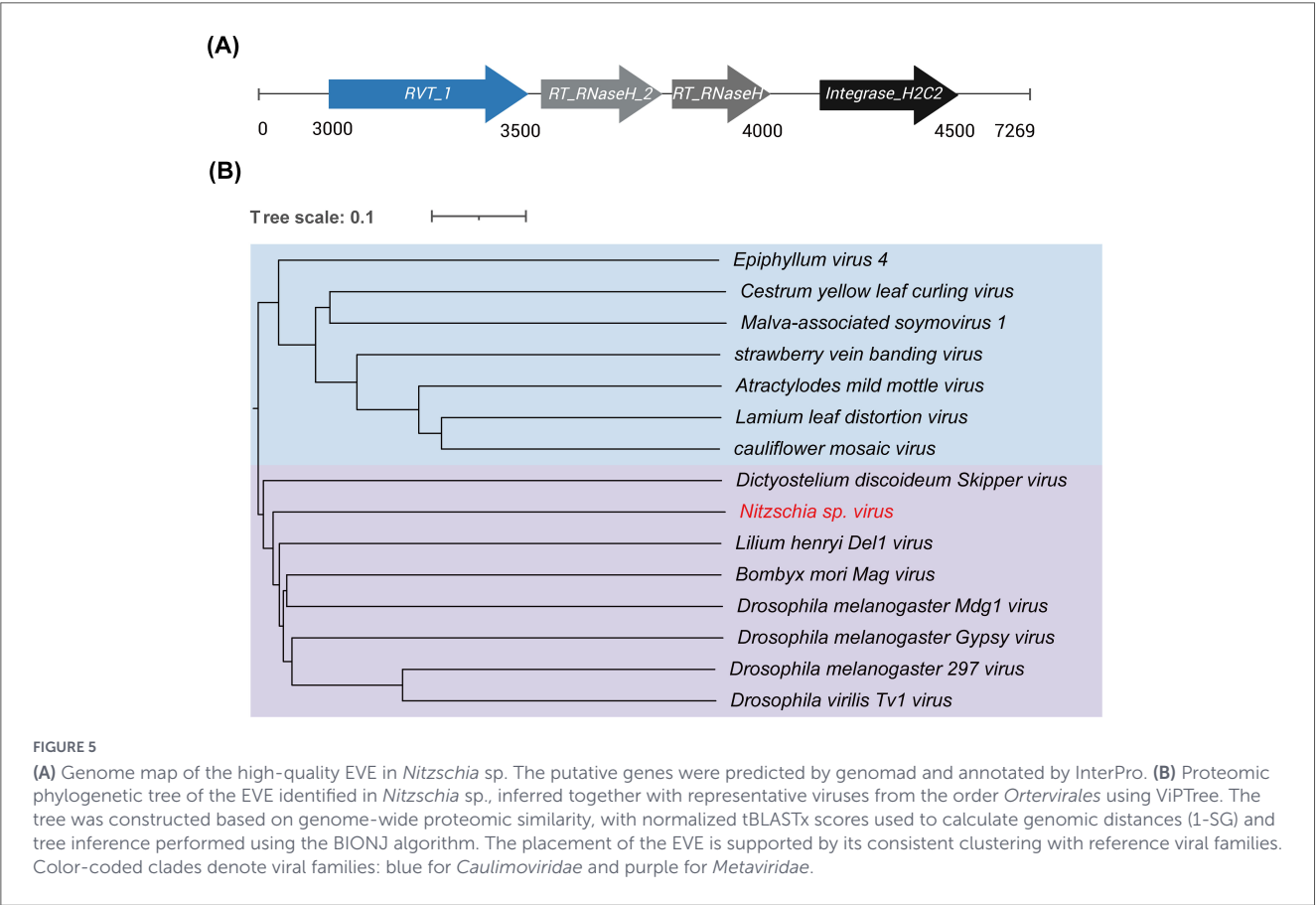
families relative to six non-polar diatom species (Tables 1, 2). These genes suggest several candidate functional pathways that may be relevant to survival and thrive in sea-ice environments. Oxidative stress, caused by excessive reactive oxygen species (ROS) production, can damage lipids, proteins, and nucleic acids, ultimately impairing cellular function and viability (Kennedy

et al., 2020). Results presented here show that both two sea-ice diatoms have expanded gene families encoding RibD (KEGG: K11752) and SHMT (KEGG: K00600), which may help mitigate the detrimental effects of oxidative stress. RibD participates in riboflavin biosynthesis, and riboflavin has been shown to contribute to cold tolerance in plants (Chen et al., 2006). For instance, riboflavin enhances non-enzymatic antioxidant

TABLE 2 Functional annotation of selected contracted proteins in *Nitzschia* sp.

Function	Proteins
Proprotein convertase subtilisin/kexin proteases (PCSK)	PCSK1 (K01359), PCSK5 (K08654), PCSK6 (K08672), PCSK7 (K08673)
Clathrin-mediated endocytosis	AP2S1 (K11827), LRP1 (K04550), LRP1B (K20049), PIP5K (K00889), SHKBP1 (K21953)
BTB/POZ domain-containing proteins	KCTD1_15 (K21754), KCTD3 (K21915), KCTD7_14 (K21917), KCTD8_12_16 (K21918), KCTD18 (K21921), KCTD21 (K21922)
Solute carrier family (SLC)	SLC4A1 (K06573), SLC4A2 (K13855), SLC4A3 (K13856), SLC4A10 (K13861), SLC34A (K14683), SLC47A (K03327)
Voltage-gated potassium channels	KCNC3 (K04889), KCNK16 (K04924), KCNKF (K05389), KCNJ2 (K04996), KCNJ4 (K04998), KCNJ6 (K05000), KCNJ9 (K05002)

The corresponding KEGG identifiers are provided for each protein.



activity, thereby reducing chilling injury and improving cold tolerance in zucchini fruit (Castro-Cegrí et al., 2024). It also helps regulate energy metabolism in strawberries during cold storage by suppressing respiration rate increases, limiting ROS accumulation, and maintaining soluble sugar content (Zhang et al., 2023b). SHMT (serine hydroxymethyltransferase) is a key enzyme in the photorespiratory pathway, a crucial mechanism for dissipating excess light energy and protecting cells from oxidative damage (Moreno et al., 2005; Yan et al., 2024). Expansion of photorespiration-related genes has also been observed in Antarctic green algae, hypothesizing that sea-ice algae may utilize photorespiration to reduce ROS production and mitigate oxidative damage (Zhang et al., 2019). In addition, both sea-ice diatoms possess a spermidine synthase (KEGG: K00797) and ice-binding protein genes (pfam: PF11999, PF20597), which are absent from six non-polar diatom species. Spermidine synthase produces

spermidine, a polyamine that plays an essential role in plant cold acclimation (Alcázar et al., 2011). Spermidine functions as both an osmoprotectant, stabilizing cellular structures under cold stress, and a ROS scavenger that mitigates oxidative damage at low temperatures (Alcázar et al., 2011). It also enhances gene expression and promotes protein repair in cyanobacteria exposed to winter chill-light stress (Zhu et al., 2015). Ice-binding proteins, commonly found in cold-adapted organisms, are thought to have been acquired in Antarctic algae through horizontal gene transfer from bacteria (Zhang et al., 2020). These proteins inhibit ice nucleation and limit crystal growth, thereby reducing freeze-induced damage and osmotic stress (Zhang et al., 2020). Importantly, it should be noted that, the functional interpretations presented in this section are based on comparative genomic patterns rather than direct functional or evolutionary evidence. Owing to the limited availability of functional studies in diatoms and algae, several

interpretations are necessarily inferred from higher-plant literature. Moreover, many of the genes discussed here are broadly conserved across eukaryotes and may represent general stress-response capacities rather than polar-specific adaptations. Accordingly, these interpretations are framed as putative hypotheses that require further validation through approaches such as selection analyses, transcriptomic profiling under sea-ice conditions, and experimental functional studies.

The extreme polar environment (e.g., freezing cold, UV exposure) can compromise microbial DNA integrity, disrupt gene expression, and impair protein synthesis. Our results show that *Nitzschia* sp. and *F. cylindrus* share several unique genes (e.g., SPOP, Dicer, STT3) not found in other six non-polar diatoms, which may contribute to enhanced DNA repair and protein folding capacities (Table 1). SPOP (Speckle-type POZ protein; KEGG: K10523) functions in the DNA damage response (DDR), a coordinated cellular system that detects and repairs genomic lesions (Zhang et al., 2014; Boysen et al., 2015). Dicer (KEGG: K11592), an endoribonuclease, also participates in repairing double-strand breaks (DSBs) (Chen et al., 2017; Yao et al., 2016) and generates small regulatory RNAs (miRNAs) that modulate gene expression and contribute to stress responses (Sunkar et al., 2007; Liu et al., 2008; Lv et al., 2010). STT3 (KEGG: K07151) plays a crucial role in N-glycosylation, a post-translational modification essential for proper protein folding, stability, and function (Koiwa et al., 2003). In *Arabidopsis*, STT3a is a key component of adaptive responses to salt and osmotic stress by promoting the refolding of unfolded proteins (Koiwa et al., 2003), and in *jojoba*, it may facilitate cold acclimation by enhancing protein folding and quality control in the endoplasmic reticulum (Gao et al., 2019). Additionally, the ATP-dependent Lon protease (KEGG: K01338), which mediates the selective degradation of misfolded or damaged proteins, is expanded in *Nitzschia* sp. and *F. cylindrus* (Table 1). This protease plays vital roles in protein quality control, DNA replication and repair, and diverse stress responses (Lee and Suzuki, 2008; He et al., 2016; Figaj et al., 2020; Kirthika et al., 2023).

Nitzschia sp. and *F. cylindrus* also share several unique genes related to nutrient transport and acquisition, including those encoding a P-type Cu^+ transporter, KUP, and ABC.FEV.S (Table 1). The P-type Cu^+ transporter (KEGG: K17686) is a membrane-bound ATPase that mediates the translocation of copper ions across cellular membranes. Copper is essential for a range of physiological processes, including photosynthetic and mitochondrial electron transport, oxidative stress management, cell wall metabolism, and hormone signaling (Yruela, 2009). KUP (KEGG: K03549) is a potassium uptake protein that enables efficient K^+ acquisition, particularly under hyperosmotic stress (Trchounian and Kobayashi, 1999). ABC.FEV.S (KEGG: K02016) is an iron complex transport system substrate-binding protein, which is a key component of the siderophore-dependent iron uptake system, responsible for recognizing and binding iron–ligand complexes (Delepeleire, 2019). Previous studies have also reported an expansion of iron transporters or transferrin genes in Antarctic sea-ice diatoms and green algae (Mock et al., 2017; Zhang et al., 2023a). Furthermore, increased iron availability has been shown to enhance the photosynthetic parameters (Fv/Fm, rETRmax, α) of the sea-ice diatoms *F. cylindrus* and *F. curta* (Pankowski and McMinn, 2009). Thus, it is hypothesized that the expansion of iron transporter genes can enhance iron uptake efficiency,

thereby improving photosynthetic performance and supporting the survival of sea-ice diatoms during the polar winter (Pankowski and McMinn, 2009; Zhang et al., 2023a).

4.2 Species-specific genomic innovations of *Nitzschia* sp.

In addition to the expanded gene families shared with *F. cylindrus*, *Nitzschia* sp. possesses several unique expansions. These expanded genes, including those encoding a BED zinc finger protein, HMGB proteins, ClpS, IARS, and BetT, are associated with transcriptional regulation, DNA repair, protein quality control, and osmotic homeostasis (Table 1). Collectively, these genes and their associated functional pathways are hypothesized to contribute to the adaptability of *Nitzschia* sp. to sea-ice environments; however, these putative roles require further validation using approaches such as selection analyses, transcriptomic profiling, and targeted functional studies. The BED zinc finger protein (KEGG: K24637) belongs to a novel superfamily of transcription factors that regulate a wide range of processes in plants, including growth, development, and cellular responses to abiotic stresses (Hussain et al., 2022). A related class of zinc finger proteins (C3HC4-type) has also been reported as expanded in Antarctic green algae (Zhang et al., 2023a). This expansion of zinc finger proteins may be attributed to the critical role of zinc in supporting the photosynthetic growth of phytoplankton in low-temperature polar oceans (Ye et al., 2022). HMGB proteins 1, 2, and 3 (KEGG: K10802, K11295, K11296) belong to the High Mobility Group superfamily and function as DNA chaperones, playing a role in various chromatin-associated processes, including transcription, replication, recombination, DNA repair, and stress response (Štros, 2010; Wang and Zhang, 2020). For instance, *HMGB2* and *HMGB3* expression in *A. thaliana* was significantly up-regulated in response to cold stress (Kwak et al., 2007). Isoleucyl-tRNA synthetase (IARS, KEGG: K01870) is an essential enzyme that possesses proofreading mechanisms to remove incorrectly activated amino acids, maintaining the high fidelity of the translation process (Yadavalli and Ibba, 2012). ClpS (KEGG: K06891) is an adaptor protein for the Clp protease (ClpP) that is involved in the degradation of misfolded and dysfunctional proteins, the regulation of short-lived proteins, and stress responses (Guo et al., 2002; Aljghami et al., 2022). The choline/glycine/proline betaine transport protein (betT, KEGG: K02168) functions in the uptake of compatible solutes such as choline, glycine betaine, and proline betaine (Chen and Beattie, 2008). The expansion of this gene is hypothesized to help *Nitzschia* sp. accumulate compatible solutes, thereby maintaining osmotic balance and protecting cellular structures in high-salinity and freezing environments.

Nitzschia sp. shows a larger number of inferred gene family contractions (126 vs. 29) than *F. cylindrus*. While this difference may partly reflect lineage-specific evolutionary history or adaptation to sea-ice environments, it is important to emphasize that the *Nitzschia* sp. genome analyzed here represents a draft metagenome-assembled genome, rather than a genome obtained from a clonal axenic culture. Consequently, some inferred gene family contractions may result from assembly fragmentation, uneven sequencing coverage, or binning artifacts, and the apparent absence of genes cannot be unequivocally interpreted as true

biological gene loss. To mitigate this limitation, our discussion focuses exclusively on contracted gene families that exhibit functional coherence, in which multiple genes associated with the same biological pathway or cellular process are inferred to be reduced (Figure 4 and Table 2). The likelihood that several independent genes from a single pathway are simultaneously missing due solely to random annotation incompleteness is relatively low, lending increased confidence to these contraction patterns as potential genomic trends rather than isolated artifacts. Nevertheless, these interpretations should be viewed as hypotheses that require validation using more complete genome assemblies derived from clonal cultures and/or functional evidence. Within this framework, several biologically coherent contraction patterns were observed in *Nitzschia* sp. (Table 2). Proprotein convertase subtilisin/kexin-like proteases (PCSKs), including PCSK1 (KEGG: K01359), PCSK5 (KEGG: K08654), PCSK6 (KEGG: K08672) and PCSK7 (KEGG: K08673), which are involved in the processing of various protein precursors (Bae et al., 2008), were inferred to be reduced in *Nitzschia* sp. This reduction may indicate a decreased requirement for processing specific protein precursors or a loss of redundant PCSK functions in *Nitzschia* sp. Additionally, genes encoding proteins involved in clathrin-mediated endocytosis, such as low-density lipoprotein receptor family members LRP1 (KEGG: K04550) and LRP1B (KEGG: K20049), were inferred to be contracted. Clathrin-mediated endocytosis is an essential process for internalizing extracellular molecules, facilitating functions such as neurotransmission, signal transduction, and plasma membrane regulation (McMahon and Boucrot, 2011). Interestingly, *Nitzschia* sp. possesses a distinct LDLR homologue, LRP5_6, which was not identified in the other seven analyzed diatoms. This suggests that *Nitzschia* sp. may employ specialized lineage-specific proteins to fulfill similar cellular functions. Voltage-gated potassium channels, such as KCNC3 (KEGG: K04889), KCNK16 (KEGG: K04924), KCNJ2 (KEGG: K04996), and KCNJ9 (KEGG: K05002), were also inferred to be contracted in *Nitzschia* sp. These channels are crucial for maintaining cell excitability and environmental sensing (Kim and Nimigean, 2016; Hedrich, 2019). The reduced repertoire of voltage-gated K⁺ channels suggest *Nitzschia* sp. may have altered ion transport or environmental signaling requirements.

4.3 Endogenous viral elements and transposable elements

Several EVEs and a large number of TEs were identified within the *Nitzschia* sp. genome. EVEs are remnants of viral DNA that have become integrated into the host genome over evolutionary time (Frank and Feschotte, 2017). EVEs, which are widespread in eukaryotes, can drive genomic innovation by donating novel genes and can sometimes act as defensive factors against related viruses (Feschotte and Gilbert, 2012; Veglia et al., 2023). In this study, a high-quality EVE, related to the *Metaviridae* viral family (i.e., Ty3/Gypsy LTR retroelements), was detected in the *Nitzschia* sp. genome. Its relatively high sequence completeness suggests a relatively recent integration or a well-preserved viral relic. At present, however, there is no direct evidence that this EVE is transcriptionally active or functionally relevant, and its presence should therefore be interpreted as a descriptive genomic feature rather than an adaptive trait. Determining whether this EVE has

any biological significance will require future transcriptomic and functional analyses. In addition, although gag/capsid proteins and LTRs are typically present in *Metaviridae* viruses, their absence in this EVE is not unexpected. Endogenous viral elements often undergo degeneration, recombination, or structural rearrangement following genomic integration, leading to the loss or divergence of certain viral modules over evolutionary time (Katzourakis and Gifford, 2010; Johnson, 2019). Also, limitations in current gene prediction tools may further hinder the detection of highly diverged gag or LTR sequences in EVEs.

Transposable elements (TEs) make a substantial contribution (27.25%) to the genome assembly of *Nitzschia* sp., which is within the upper range reported for diatom genomes (Zheng et al., 2025). TEs, or “jumping genes,” are mobile DNA segments capable of changing position within a genome, thereby influencing genome rearrangement and expansion (Hermann et al., 2014). In diatoms, LTR retrotransposons are the most abundant class of TEs (Hermann et al., 2014) and the above newly identified high-quality EVE also belongs to this clade. LTR retrotransposons have been implicated in stress responses in both plants and diatoms (Hermann et al., 2014; Pargana et al., 2019; Egue et al., 2015; Maumus et al., 2009). For instance, in the diatom *Leptocylindrus aporus*, LTR retrotransposon transcripts are significantly upregulated under cold stress, with TE-related transcripts linked to antioxidant activity (Pargana et al., 2019). In *P. tricornutum*, the LTR retrotransposon *Surcouf*, which contains regulatory motifs such as heat-shock elements, stress-response elements and CCAAT boxes, is strongly overexpressed under heat stress in parallel with small heat-shock proteins (Egue et al., 2015). Moreover, in *P. tricornutum*, LTR retrotransposons display increased transcriptional activity under nitrate limitation and in response to diatom-derived reactive aldehydes, which can trigger stress responses and cell death (Maurus et al., 2009). However, our study did not find a clear relationship between overall TE abundance and geographical distribution (i.e., polar versus non-polar diatoms). This lack of correlation suggests that TE accumulation alone is unlikely to be a driver of adaptation to polar or sea-ice environments. Instead, previous studies indicate that TE regulation, transcriptional activation, and specific TE lineages may be more relevant for stress responses (Maurus et al., 2009; Egue et al., 2015; Pargana et al., 2019). As transcriptional activity and epigenetic regulation of TEs were not assessed in the present study, no direct functional or adaptive role can be inferred here. Future transcriptomic and experimental studies will be necessary to determine whether TEs in *Nitzschia* sp. and other sea ice diatoms are stress-responsive or contribute to genome plasticity under sea-ice conditions.

In conclusion, this study presents a draft MAG of an Antarctic sea-ice diatom, *Nitzschia* sp., providing a genomic resource for exploring potential mechanisms underlying life in extreme polar environments. Comparative analyses reveal that *Nitzschia* sp. shares a number of unique or expanded genes with *F. cylindrus*, including genes associated with oxidative stress responses, DNA repair, protein quality control, osmotic regulation, and nutrient acquisition and transport. While these genes are broadly conserved across eukaryotes, their shared presence or expansions in sea-ice diatoms highlights functional capacities that may be relevant for coping with the combined stresses of the sea-ice habitat. In addition, *Nitzschia* sp. exhibits pronounced

contraction of several gene families, including those related to proprotein convertase subtilisin/kexin proteases, clathrin-mediated endocytosis, BTB/POZ domain-containing proteins, solute carrier transporters, and voltage-gated potassium channels. We also identified a high-quality EVE related to the viral family *Metaviridae*, as well as a substantial proportion of transposable elements (27.25% of the genome assembly). While no direct adaptive role can be inferred from these observations, they document notable features of genome composition and potential genome plasticity in this Antarctic diatom. Overall, this study characterizes the genomic features of *Nitzschia* sp. and provides a comparative framework for future studies. Further work integrating population genomics, transcriptomics, and experimental approaches will be necessary to determine how these genomic patterns contribute to ecological success and persistence of diatoms in Antarctic sea-ice ecosystems.

Data availability statement

The data presented in this study are publicly available in the NCBI (<https://www.ncbi.nlm.nih.gov/>), accession PRJNA1454863.

Author contributions

CZ: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. FK: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. MH: Data curation, Formal analysis, Methodology, Writing – review & editing. HY: Data curation, Formal analysis, Methodology, Writing – review & editing. JS: Methodology, Writing – review & editing. YL: Supervision, Writing – review & editing. MW: Supervision, Writing – review & editing. AM: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by the Australian Research Council Special Research Initiative, Australian Centre for Excellence in Antarctic Science (SR200100008), the National Natural Science Foundation of China (42120104006, 42176111, 42306156), the Ocean Negative Carbon Emissions, and

the Fundamental Research Funds for the Central Universities, China (202172002, 201812002).

Acknowledgments

We acknowledge the support of the High-Performance Biological Supercomputing Center at the Ocean University of China for this research.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer PS declared a shared affiliation with the authors CZ, FK and AM to the handling editor at the time of review.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2026.1755917/full#supplementary-material>

References

- Alcázar, R., Cuevas, J. C., Planas, J., Zarza, X., Bortolotti, C., Carrasco, P., et al. (2011). Integration of polyamines in the cold acclimation response. *Plant Sci.* 180, 31–38. doi: 10.1016/j.plantsci.2010.07.022
- Aljhami, M. E., Barghash, M. M., Majaesic, E., Bhandari, V., and Houry, W. A. (2022). Cellular functions of the ClpP protease impacting bacterial virulence. *Front. Mol. Biosci.* 9:1054408. doi: 10.3389/fmolb.2022.1054408

- Arrigo, K. R., Worthen, D. L., Lizotte, M. P., Dixon, P., and Dieckmann, G. (1997). Primary production in Antarctic sea ice. *Science* 276, 394–397. doi: 10.1126/science.276.5311.394
- Bae, J. A., Park, H. J., Seo, Y. M., Roh, J., Hsueh, A. J., and Chun, S. Y. (2008). Hormonal regulation of proprotein convertase subtilisin/kexin type 5 expression during ovarian follicle development in the rat. *Mol. Cell. Endocrinol.* 289, 29–37. doi: 10.1016/j.mce.2008.04.006
- Blum, M., Andreeva, A., Florentino, L. C., Chuguransky, S. R., Grego, T., Hobbs, E., et al. (2024). InterPro: The protein sequence classification resource in 2025. *Nucleic Acids Res.* 53, D444–D456. doi: 10.1093/nar/gkae1082
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660. doi: 10.7717/peerj.1660
- Boysen, G., Barbieri, C. E., Prandi, D., Blattner, M., Chae, S. S., Dahija, A., et al. (2015). SPOP mutation leads to genomic instability in prostate cancer. *Elife* 4:e09207. doi: 10.7554/eLife.09207
- Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., et al. (2024). Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* 42, 1303–1312. doi: 10.1038/s41587-023-01953-y
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Castro-Cegri, A., García, A., Garrido, D., and Palma, F. (2024). Riboflavin improves postharvest cold tolerance in zucchini fruit inducing non-enzymatic antioxidant response and phenolic metabolism. *Plant Physiol. Biochem.* 217:109270. doi: 10.1016/j.plaphy.2024.109270
- Challis, R., Richards, E., Rajan, J., Cochrane, G., and Blaxter, M. (2020). BlobToolKit – interactive quality assessment of genome assemblies. *G* 3, 1361–1374. doi: 10.1534/g3.119.400908
- Chen, C., and Beattie, G. A. (2008). *Pseudomonas syringae* BetT is a low-affinity choline transporter that is responsible for superior osmoprotection by choline over glycine betaine. *J. Bacteriol.* 190, 2717–2725. doi: 10.1128/jb.01585-07
- Chen, S. C., Chang, Y. C., Lin, C. H., Lin, C. H., and Liaw, S. H. (2006). Crystal structure of a bifunctional deaminase and reductase from *Bacillus subtilis* involved in riboflavin biosynthesis. *J. Biol. Chem.* 281, 7605–7613. doi: 10.1074/jbc.M510254200
- Chen, X., Li, W. F., Wu, X., Zhang, H. C., Chen, L., Zhang, P. Y., et al. (2017). Dicer regulates non-homologous end joining and is associated with chemosensitivity in colon cancer patients. *Carcinogenesis* 38, 873–882. doi: 10.1093/carcin/bgx059
- Delepelaire, P. (2019). Bacterial ABC transporters of iron containing compounds. *Res. Microbiol.* 170, 345–357. doi: 10.1016/j.resmic.2019.10.008
- Duncan, R. J., and Petrou, K. (2022). Biomolecular composition of sea ice microalgae and its influence on marine biogeochemical cycling and carbon transfer through polar marine food webs. *Geosciences* 12:38. doi: 10.3390/geosciences12010038
- Egue, F., Chenais, B., Tastard, E., Marchand, J., Hiard, S., Gateau, H., et al. (2015). Expression of the retrotransposons Surcouf and Blackbeard in the marine diatom *Phaeodactylum tricornutum* under thermal stress. *Phycologia* 54, 617–627. doi: 10.2216/15-52.1
- Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: Insights into viral evolution and impact on host biology. *Nat. Rev. Genet.* 13, 283–296. doi: 10.1038/nrg3199
- Figaj, D., Czaplewska, P., Przepióra, T., Ambroziak, P., Potrykus, M., and Skorko-Glonek, J. (2020). Lon protease is important for growth under stressful conditions and pathogenicity of the phytopathogen, bacterium *Dickeya solani*. *Int. J. Mol. Sci.* 21:3687. doi: 10.3390/ijms21103687
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Frank, J. A., and Feschotte, C. (2017). Co-option of endogenous viral sequences for host cell function. *Curr. Opin. Virol.* 25, 81–89. doi: 10.1016/j.coviro.2017.07.021
- Gao, F., Ma, P., Wu, Y., Zhou, Y., and Zhang, G. (2019). Quantitative proteomic analysis of the response to cold stress in jojoba, a tropical woody crop. *Int. J. Mol. Sci.* 20:243. doi: 10.3390/ijms20020243
- Guo, F., Esser, L., Singh, S. K., Maurizi, M. R., and Xia, D. (2002). Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. *J. Biol. Chem.* 277, 46753–46762. doi: 10.1074/jbc.M208104200
- He, L., Nair, M. K. M., Chen, Y., Liu, X., Zhang, M., Hazlett, K. R., et al. (2016). The protease locus of *Francisella tularensis* LVS is required for stress tolerance and infection in the mammalian host. *Infect. Immun.* 84, 1387–1402. doi: 10.1128/iai.00076-16
- Hedrich, R. (2019). Diatom signaling: A novel channel type identified. *Curr. Biol.* 29, R319–R321. doi: 10.1016/j.cub.2019.03.038
- Hermann, D., Egue, F., Tastard, E., Nguyen, D. H., Casse, N., Caruso, A., et al. (2014). An introduction to the vast world of transposable elements—what about the diatoms? *Diatom Res.* 29, 91–104. doi: 10.1080/0269249X.2013.877083
- Holt, C., and Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491
- Hussain, A., Liu, J., Mohan, B., Burhan, A., Nasim, Z., Bano, R., et al. (2022). A genome-wide comparative evolutionary analysis of zinc finger-BED transcription factor genes in land plants. *Sci. Rep.* 12:12328. doi: 10.1038/s41598-022-16602-8
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Johnson, W. E. (2019). Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat. Rev. Microbiol.* 17, 355–370. doi: 10.1038/s41579-019-0189-2
- Joli, N., Concia, L., Mocaer, K., Guterman, J., Laude, J., Guerin, S., et al. (2024). Hypometabolism to survive the long polar night and subsequent successful return to light in the diatom *Fragilariopsis cylindrus*. *New Phytol.* 241, 2193–2208. doi: 10.1111/nph.19387
- Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Kang, S. H., and Fryxell, G. A. (1992). *Fragilariopsis cylindrus* (Grunow) Krieger: The most abundant diatom in water column assemblages of Antarctic marginal ice-edge zones. *Polar Biol.* 12, 609–627. doi: 10.1007/BF00236984
- Karlicki, M., Antonowicz, S., and Karnkowska, A. (2022). Tiara: Deep learning-based classification system for eukaryotic sequences. *Bioinformatics* 38, 344–350. doi: 10.1093/bioinformatics/btab672
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Katzourakis, A., and Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet.* 6:e1001191. doi: 10.1371/journal.pgen.1001191
- Kennedy, F., Martin, A., Bowman, J. P., Wilson, R., and McMinn, A. (2019). Dark metabolism: A molecular insight into how the Antarctic sea-ice diatom *Fragilariopsis cylindrus* survives long-term darkness. *New Phytol.* 223, 675–691. doi: 10.1111/nph.15843
- Kennedy, F., Martin, A., Castriños, K., Cimoli, E., McMinn, A., and Ryan, K. G. (2020). Rapid manipulation in irradiance induces oxidative free-radical release in a fast-ice algal community (McMurdo Sound, Antarctica). *Front. Plant Sci.* 11:588005. doi: 10.3389/fpls.2020.588005
- Kim, D. M., and Nimigeon, C. M. (2016). Voltage-gated potassium channels: A structural examination of selectivity and gating. *Cold Spring Harb. Perspect. Biol.* 8:a029231. doi: 10.1101/cshperspect.a029231
- Kirthika, P., Lloren, K. K. S., Jawalagatti, V., and Lee, J. H. (2023). Structure, substrate specificity and role of Lon protease in bacterial pathogenesis and survival. *Int. J. Mol. Sci.* 24:3422. doi: 10.3390/ijms24043422
- Koiwa, H., Li, F., McCully, M. G., Mendoza, I., Koizumi, N., Manabe, Y., et al. (2003). The STT3a subunit isoform of the *Arabidopsis* oligosaccharyltransferase controls adaptive responses to salt/osmotic stress. *Plant Cell.* 15, 2273–2284. doi: 10.1105/tpc.013862
- Kwak, K. J., Kim, J. Y., Kim, Y. O., and Kang, H. (2007). Characterization of transgenic *Arabidopsis* plants overexpressing high mobility group B proteins under high salinity, drought or cold stress. *Plant Cell Physiol.* 48, 221–231. doi: 10.1093/pcp/pcp057
- Lee, I., and Suzuki, C. K. (2008). Functional mechanics of the ATP-dependent Lon protease—lessons from endogenous protein and synthetic peptide substrates. *Biochim. Biophys. Acta* 1784, 727–735. doi: 10.1016/j.bbapap.2008.02.010
- Legendre, L., Ackley, S. F., Dieckmann, G. S., Gulliksen, B., Horner, R., Hoshiai, T., et al. (1992). Ecology of sea ice biota. *Polar Biol.* 12, 429–444. doi: 10.1007/BF00243114
- Leticun, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52, W78–W82. doi: 10.1093/nar/gkae268
- Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Liu, H. H., Tian, X., Li, Y. J., Wu, C. A., and Zheng, C. C. (2008). Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA* 14, 836–843. doi: 10.1261/rna.895308
- Ly, D. K., Bai, X., Li, Y., Ding, X. D., Ge, Y., Cai, H., et al. (2010). Profiling of cold-stress-responsive miRNAs in rice by microarrays. *Gene* 459, 39–47. doi: 10.1016/j.gene.2010.03.011
- Manni, M., Berkeley, M. R., Seppey, M., and Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Curr. Protoc.* 1:e323. doi: 10.1002/cpz1.323
- Martin, A., McMinn, A., Heath, M., Hegseth, E. N., and Ryan, K. G. (2012). The physiological response to increased temperature in over-wintering sea ice algae and phytoplankton in McMurdo Sound, Antarctica and Tromsø Sound, Norway. *J. Exp. Mar. Biol. Ecol.* 428, 57–66. doi: 10.1016/j.jembe.2012.06.006
- Maumus, F., Allen, A. E., Mhiri, C., Hu, H., Jabbari, K., Vardi, A., et al. (2009). Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genom.* 10:624. doi: 10.1186/1471-2164-10-624

- McMahon, H. T., and Boucrot, E. (2011). Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* 12, 517–533. doi: 10.1038/nrm3151
- McMinn, A., Martin, A., and Ryan, K. (2010). Phytoplankton and sea ice algal biomass and physiology during the transition between winter and spring (McMurdo Sound, Antarctica). *Polar Biol.* 33, 1547–1556. doi: 10.1007/s00300-010-0844-6
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Mock, T., Otilar, R. P., Strauss, J., McMullan, M., Paajanen, P., Schmutz, J., et al. (2017). Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* 541, 536–540. doi: 10.1038/nature20803
- Moreno, J. I., Martín, R., and Castresana, C. (2005). *Arabidopsis* SHMT1, a serine hydroxymethyltransferase that functions in the photorespiratory pathway, influences resistance to biotic and abiotic stress. *Plant J.* 41, 451–463. doi: 10.1111/j.1365-313X.2004.02311.x
- Nayfach, S., Camargo, A. P., Schulz, F., Eloie-Fadrosch, E., Roux, S., and Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* 39, 578–585. doi: 10.1038/s41587-020-00774-7
- Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., and Goto, S. (2017). ViPTree: The viral proteomic tree server. *Bioinformatics* 33, 2379–2380. doi: 10.1093/bioinformatics/btx157
- Olm, M. R., Brown, C. T., Brooks, B., and Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868. doi: 10.1038/ismej.2017.126
- Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310
- Pankowski, A., and McMinn, A. (2009). Iron availability regulates growth, photosynthesis, and production of ferredoxin and flavodoxin in Antarctic sea ice diatoms. *Aquat. Biol.* 4, 273–288. doi: 10.3354/ab00116
- Pargana, A., Musacchia, F., Sanges, R., Russo, M. T., Ferrante, M. I., Bowler, C., et al. (2019). Intraspecific diversity in the cold stress response of transposable elements in the diatom *Leptocylindrus aporus*. *Genes* 11:9. doi: 10.3390/genes11010009
- Pertea, G., and Pertea, M. (2020). GFF utilities: GffRead and GffCompare. *F1000Res* 9:ISCB Comm J-304. doi: 10.12688/f1000research.23297.2
- Ryan, K. G., Cowie, R. O., Liggins, E., McNaughtan, D., Martin, A., and Davy, S. K. (2009). The short-term effect of irradiance on the photosynthetic properties of Antarctic fast-ice microalgal communities. *J. Phycol.* 45, 1290–1298. doi: 10.1111/j.1529-8817.2009.00764.x
- Saenz, B. T., and Arrigo, K. R. (2014). Annual primary production in Antarctic sea ice during 2005–2006 from a sea ice state estimate. *J. Geophys. Res. Oceans* 119, 3645–3678. doi: 10.1002/2013JC009677
- Schmidt, K., Atkinson, A., Pond, D. W., and Ireland, L. C. (2014). Feeding and overwintering of Antarctic krill across its major habitats: The role of sea ice cover, water depth, and phytoplankton abundance. *Limnol. Oceanogr.* 59, 17–36. doi: 10.4319/lo.2014.59.1.0017
- Scott, F. J., and Marchant, H. J. (2005). *Antarctic Marine Protists*. Canberra: Australian Biological Resources Study.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Štros, M. (2010). HMGB proteins: Interactions with DNA and chromatin. *Biochim. Biophys. Acta Gene Regul. Mech.* 1799, 101–113. doi: 10.1016/j.bbagr.2009.09.008
- Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., and Wang, Y. (2023). OrthoVenn3: An integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* 51, 397–403. doi: 10.1093/nar/gkad313
- Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J. K. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci.* 12, 301–309. doi: 10.1016/j.tplants.2007.05.001
- Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E. M., Kriventseva, E. V., et al. (2025). OrthoDB and BUSCO update: Annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res.* 53, D516–D522. doi: 10.1093/nar/gkae987
- The Galaxy, and Community. (2024). The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* 52, 83–94. doi: 10.1093/nar/gkae410
- Thomas, D. N., and Dieckmann, G. S. (2002). Antarctic sea ice—a habitat for extremophiles. *Science* 295, 641–644. doi: 10.1126/science.1063391
- Trchounian, A., and Kobayashi, H. (1999). Kup is the major K⁺ uptake system in *Escherichia coli* upon hyper-osmotic stress at a low pH. *FEBS Lett.* 447, 144–148. doi: 10.1016/S0014-5793(99)00288-4
- Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6:158. doi: 10.1186/s40168-018-0541-1
- Vancoppenolle, M., Meiners, K. M., Michel, C., Bopp, L., Brabant, F., Carnat, G., et al. (2013). Role of sea ice in global biogeochemical cycles: Emerging views and challenges. *Quat. Sci. Rev.* 79, 207–230. doi: 10.1016/j.quascirev.2013.04.011
- Veglia, A. J., Bistolas, K. S., Voolstra, C. R., Hume, B. C., Ruscheweyh, H. J., Planes, S., et al. (2023). Endogenous viral elements reveal associations between a non-retroviral RNA virus and symbiotic dinoflagellate genomes. *Commun. Biol.* 6:566. doi: 10.1038/s42003-023-04917-9
- Von Meijenfildt, F. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H., and Dutilh, B. E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* 20:217. doi: 10.1186/s13059-019-1817-x
- Wang, S., and Zhang, Y. (2020). HMGB1 in inflammation and cancer. *J. Hematol. Oncol.* 13:116. doi: 10.1186/s13045-020-00950-x
- Weeks, W. (2010). *On Sea Ice*. Alaska: University of Alaska Press.
- Yadavalli, S. S., and Ibba, M. (2012). Quality control in aminoacyl-tRNA synthesis: Its role in translational fidelity. *Adv. Protein Chem. Struct. Biol.* 86, 1–43. doi: 10.1016/B978-0-12-386497-0.00001-3
- Yan, M., Li, W., Zhou, Z., Pan, T., Li, L., Chai, M., et al. (2024). Genome-wide identification of the serine hydroxymethyltransferase gene revealed the function of GhSHMT1s involved in plant growth development and salt stress tolerance in cotton. *Ind. Crops Prod.* 222:119687. doi: 10.1016/j.indcrop.2024.119687
- Yao, Y., Bilichak, A., Golubov, A., and Kovalchuk, I. (2016). *Arabidopsis thaliana* siRNA biogenesis mutants have the lower frequency of homologous recombination. *Plant Signal. Behav.* 11:e1151599. doi: 10.1080/15592324.2016.1151599
- Ye, N., Han, W., Toseland, A., Wang, Y., Fan, X., Xu, D., et al. (2022). The role of zinc in the adaptive evolution of polar phytoplankton. *Nat. Ecol. Evol.* 6, 965–978. doi: 10.1038/s41559-022-01750-x
- Yruela, I. (2009). Copper in plants: Acquisition, transport and interactions. *Funct. Plant Biol.* 36, 409–430. doi: 10.1071/FP08288
- Zhang, D., Wang, H., Sun, M., Yang, J., Zhang, W., Han, S., et al. (2014). Speckle-type POZ protein, SPOP, is involved in the DNA damage response. *Carcinogenesis* 35, 1691–1697. doi: 10.1093/carcin/bgu022
- Zhang, X., Han, W., Fan, X., Wang, Y., Xu, D., Sun, K., et al. (2023a). Gene duplication and functional divergence of new genes contributed to the polar acclimation of Antarctic green algae. *Mar. Life Sci. Technol.* 5, 511–524. doi: 10.1007/s42995-023-00203-z
- Zhang, X., Xu, J., Wang, M., Gan, C., Ren, Y., Zhao, X., et al. (2023b). Riboflavin application regulates sugar and energy metabolism in strawberries during cold storage. *LWT* 187:115301. doi: 10.1016/j.lwt.2023.115301
- Zhang, Z., Qu, C., Yao, R., Nie, Y., Xu, C., Miao, J., et al. (2019). The parallel molecular adaptations to the Antarctic cold environment in two psychrophilic green algae. *Genome Biol. Evol.* 11, 1897–1908. doi: 10.1093/gbe/evz104
- Zhang, Z., Qu, C., Zhang, K., He, Y., Zhao, X., Yang, L., et al. (2020). Adaptation to extreme Antarctic environments revealed by the genome of a sea ice green alga. *Curr. Biol.* 30, 3330–3341. doi: 10.1016/j.cub.2020.06.029
- Zheng, T., Zhang, X., Liu, T., Liu, X., Bowler, C., and Lin, X. (2025). The roles of transposable elements and gene family dynamics in shaping diversity and evolution in diatoms. *bioRxiv[Preprint]* doi: 10.1101/2025.10.24.684338
- Zhu, X., Li, Q., Yin, C., Fang, X., and Xu, X. (2015). Role of spermidine in overwintering of cyanobacteria. *J. Bacteriol.* 197, 2325–2334. doi: 10.1128/jb.00153-15