# scientific reports

OPEN

# Linking extreme seasonality and gene expression in Arctic marine protists

Magdalena Wutkowska [1,2,3✉], Anna Vader [1], Ramiro Logares [4], Eric Pelletier [5,6] & Tove M. Gabrielsen [1,7]

At high latitudes, strong seasonal differences in light availability affect marine organisms and regulate the timing of ecosystem processes. Marine protists are key players in Arctic aquatic ecosystems, yet little is known about their ecological roles over yearly cycles. This is especially true for the dark polar night period, which up until recently was assumed to be devoid of biological activity. A 12 million transcripts catalogue was built from 0.45 to 10 µm protist assemblages sampled over 13 months in a time series station in an Arctic fjord in Svalbard. Community gene expression was correlated with seasonality, with light as the main driving factor. Transcript diversity and evenness were higher during polar night compared to polar day. Light-dependent functions had higher relative expression during polar day, except phototransduction. 64% of the most expressed genes could not be functionally annotated, yet up to 78% were identified in Arctic samples from *Tara* Oceans, suggesting that Arctic marine assemblages are distinct from those from other oceans. Our study increases understanding of the links between extreme seasonality and biological processes in pico- and nanoplanktonic protists. Our results set the ground for future monitoring studies investigating the seasonal impact of climate change on the communities of microbial eukaryotes in the High Arctic.

Solar radiation is a dominant energy source for life on Earth, and an important driver of evolution[1]. In the ocean, phytoplankton, mostly cyanobacteria and photosynthetic microbial eukaryotes, contribute half of the net primary production on Earth[2]. Light availability in the ocean declines with depth and forces a vertical distribution of species, with phototrophic organisms dwelling in the epipelagic zone (< 200 m depth). The further from the equator, the more pronounced the annual changes in light regime, which at high latitudes is the strongest environmental driver of marine plankton phenology[3]. During the Arctic polar night, the sun does not rise above the horizon for 4–6 months. The opposite happens during polar day, when the sun stays above the horizon for an equally long period. Extreme seasonality introduces profound limitations to biological processes in polar regions, and for centuries researchers perceived polar night as a period devoid of biological activity. Recent studies have reported substantial biological activity during the polar night; however, most of these studies focused on macroorganisms, predominantly zooplankton[4–6].

Our understanding of communities of marine microbial eukaryotes in the Arctic is primarily based on studies limited to a single sampling time point or cruises sampling along transects once or infrequently. However, disentangling the dynamics of changing community composition of organisms requires time series stations sampled at regular intervals[7,8]. The world's northernmost marine time series station (IsA) in Adventfjorden, Isfjorden, Svalbard (Fig. 1), has been continuously sampled since December 2011[9]. This endeavour generated metabarcoding-based knowledge regarding which marine microbial eukaryotes are present and active throughout the year[10,11]. Seasonal dynamics of microbial eukaryotes can be analysed through many ecologically important measures and indices, such as diversity, biomass, cell counts, functions, etc. In general, cell counts and biomass of microbial eukaryotes during polar night are lower compared to polar day, also at IsA[12], whereas diversity of operational taxonomic units (OTUs) is inversely proportional to this trend being higher during polar night[10]. However, the proportion of plastid-bearing to heterotrophic cells is lower during polar night (reviewed in[4]).

[1]Department of Arctic Biology, The University Centre in Svalbard, Longyearbyen, Norway. [2]Department of Arctic and Marine Biology, UiT – The Arctic University of Norway, Tromsø, Norway. [3]Institute of Soil Biology and Biogeochemistry, Biology Centre CAS, České Budějovice, Czechia. [4]Institute of Marine Sciences (ICM), CSIC, Barcelona, Catalonia, Spain. [5]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France. [6]CNRS Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, Paris, France. [7]Department of Natural Sciences, University of Agder, Kristiansand, Norway. ✉email: magda.wutkowska@gmail.com
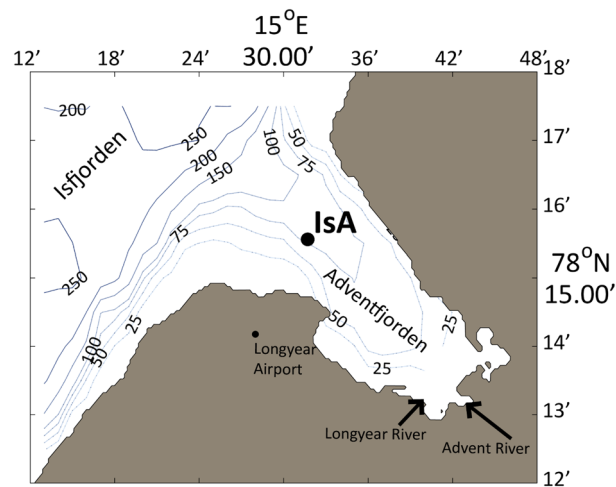
**Figure 1.** Location of the Isfjorden Adventfjorden (IsA) time series station in Svalbard. The map was generated using MATLAB R2013b.

Studies on the response of marine polar microbial communities to light/dark cycles are rare and cover a shorter timespan than the duration of the polar night[13]. Typical studies on dark survival of photosynthetic unicellular organisms are performed in laboratory conditions on single-species cultures. Some of the key Arctic microeukaryotic phototrophs were found to produce rRNA during polar night, which is a hint of them being metabolically active in these conditions[10,14]. Most of the primary production in the Arctic Ocean is performed by marine microbial eukaryotes when enough solar radiation is available[15,16]. Outside this period these cells are assumed to use accumulated resources[17], decrease their metabolism[18,19] or remain dormant[20]. However, many species of microbial eukaryotes instead of passively surviving prolonged darkness might switch their feeding strategy[21,22], as is the case with mixotrophs[23].

Pico- and nanoeukaryotes play important roles in the marine environment, including photo-, heterotrophy or parasitism, with some species that can switch between these trophic modes[24–26]. These fractions of phytoplankton are challenging for classic microscopic taxonomy assignments or elucidating their roles. The analyses of their gene expression are especially helpful in understanding what molecular processes they use to respond to environmental heterogeneity[24,27]. Fluctuating environments might promote more stochasticity in gene expression of individual cells within populations and contribute to higher fitness and higher survival in stressful conditions[28]. Nevertheless, community-level gene expression obtained by '-omics' methods was demonstrated as an effective predictor of current marine biogeochemical state[29]. In other words, the snapshot of metabolic functions performed by the community is tightly linked with environmental gradients present in the ecosystem at a given time.

We targeted the 0.45–10 μm size fraction of the microbial eukaryotic community from the IsA time series station to determine the dynamics of gene expression throughout a polar year, from which more than 12,000,000 eukaryote transcripts catalogue was built. Previous studies described higher diversity of microbial eukaryotes during polar night; thus, we hypothesize that the transcript diversity follows this trend. Given that light is the most important structuring force of community composition[3], we hypothesise that the light regime plays an essential role in controlling cellular processes in microbial eukaryotes. The presence of active phototrophic microbial eukaryotes during the polar night and their quick ecophysiological response to the return of light was confirmed by several studies[10,13,14,30]. Hence, we hypothesise that genes involved in light-dependent processes, such as light-harvesting, are expressed also during polar night.

## Materials and methods

### Study site and sampling.
The biological and environmental samples were collected at local noon at 11-time points between 14 December 2011 and 10 January 2013 from the Isfjorden Adventfjorden time series station (IsA); located on the west coast of Spitsbergen, Svalbard (N 78°15.6, E 15°31.8, Fig. 1). At each of the 11 sampling dates, 30 l of seawater was sequentially sampled from 25 m depth using a 10 l Niskin bottle (KC Denmark), and immediately processed on board. Samples were kept in dark and cold conditions while prefiltered by gravity through 10 μm nylon mesh (KC Denmark) and then onto 8–12 47 mm 0.45 μm Durapore filters (Millipore) using vacuum pumps. Each filter was fixed in 600 μl LB buffer (RNAqueous Total RNA Isolation Kit, Invitrogen, Thermo Fisher Scientific) 5–20 min after sampling, flash-frozen in liquid nitrogen and stored at −80 °C.

At each sampling date, a vertical profile of environmental variables was obtained using a handheld SAIV 204 STD/CTD probe. However, in this paper we present only the data for 25 m depth where the samples were taken from (see[10] for complete profile). Photosynthetically active radiation (PAR), size-fractionated chlorophyll *a* and nutrient concentrations (nitrate/nitrite, phosphate, silicate), were obtained as described in[10].

### mRNA extraction and amplification.
Total RNA was extracted with the RNAqueous Total RNA Isolation Kit (Invitrogen, Thermo Fisher Scientific) according to manufacturer's recommendation. Samples were

thawed on ice, vortexed and kept on ice during RNA extraction. The thawed lysate was added to a tube with 200 µm molecular biology-grade zirconium beads from pre-filled tubes. Extracts from filters collected on the same day were pooled together. After removal of DNA using the TURBO DNA-free Kit (Invitrogen, Thermo Fisher Scientific), RNA was tested for PCR inhibitors and DNA contamination by running PCRs with general eukaryotic 18S rRNA primers with and without a reverse transcription step. The DNA-free RNA was subsequently cleaned and concentrated by precipitation in 5 M ammonium acetate and absolute ethanol, using glycogen as an RNA carrier. RNA was amplified using the MessageAMP II aRNA Amplification Kit (Invitrogen, Thermo Fisher Scientific) according to manufacturer's recommendations, extending the in vitro transcription step to 14 h. Amplified samples were dissolved in 100 µl of nuclease-free water and frozen at -80 °C. Amplified mRNA was sent to GATC (Constance, Germany) where the libraries were prepared and sequenced on Illumina HiSeq 2500/4000, producing 150 bp paired-end reads.

**Data processing.** Generated sequences were processed in four main steps: pre-processing, metatranscriptome co-assembly, mapping of reads from individual metatranscriptomes onto the assembly and finally annotation of assembled transcript isoforms (Supplementary Figure S1).The quality of the data was assessed with FastQC v.0.11.5[31]. The pre-processing step aimed to remove unwanted sequences from the metatranscriptomes. First, Illumina adapters were removed using BBDuk v. 37.36[32]. Overrepresented sequences in each metatranscriptome reported by FastQC, consisting predominantly of poly(A) and poly(T) fragments, were removed with BBDuk. The same software was used to remove PhiX control reads. Although we used poly(A) selection to capture only eukaryotic mRNA during sample preparation, rRNA may remain in the samples[33]. Thus, we used SortMeRNA 2.0[34] to remove sequences that mapped to rRNA. Lastly, BBDuk was used to remove sequences of quality score < 20 and read length < 25 bp (because the next step by default uses k-mers of that length). The resulting 328 M read pairs without Illumina adapters, overrepresented sequences, PhiX control reads, rRNA and sequences of low quality (< 20) and length > 25 bp will be referred to as 'clean reads'.

Detailed statistics on the initial library size of each metatranscriptome, and its change after each of the pre-processing steps can be found in Supplementary Table S1. The clean reads from 11 metatranscriptomes were de novo co-assembled into a gene catalogue using Trinity [35,36]. Digital normalization removed 10 million read pairs with a median k-mer abundance of < 2 (–min_cov 2) and > 50 (–max_cov 50) prior to the co-assembly. The initial assembly step of Trinity – *Inchworm*, ran on 199 million of all read pairs with no further normalisation. The assembled output ran through the remaining part of the co-assembly, first constructing de Bruijn graphs (*Chrysalis* step) and then resolving them (*Butterfly* step). Expression levels were estimated by mapping clean reads against the gene catalogue in RSEM 1.3.0[37]. Due to varying numbers of reads in each of the metatranscriptomes (Supplementary Table S1) and to ensure between-sample comparison[38] we used a relative measure of transcripts per million reads (TPM).

**Annotations.** De novo assembly produced unique 12,245,433 transcript isoforms, with clean reads mapping at least once to 11,010,859 isoforms. Most transcripts were characterized by low sum of relative abundance across samples (8 transcripts with > 10,000 TPM, 154 with > 1000 TPM, 3483 with > 100 TPM, 68,166 with > 10 TPM and 2,390,862 with > 1 TPM; Supplementary Figure S3). To increase the robustness of analyses and avoid stochasticity due to low abundance transcripts, further analyses were carried out on a core dataset that contained 68,166 of the most abundant transcript isoforms for which the sum of TPM in all the samples was greater than 10 (Supplementary Figure S1; from now on we will refer to the transcript isoforms as "transcripts").

Coding regions were predicted using TransDecoder 5.1.0 (https://github.com/TransDecoder/TransDecoder/). To avoid false positive predictions, which are reported to drastically increase for shorter ORFs, only sequences longer than 100 amino acids were used for functional annotation. The core dataset was functionally annotated using Trinotate 3.3.1 with default parameters[39]. Similarities between the de novo assembled transcripts/predicted coding regions and proteins in the UniProt database[40] were assessed using blastx/blastp, with max_target_seq = 1 and e-value = 1e-3 (BLAST +)[41]. Protein domains were identified with HMMER3[42] against the Pfam database (31.0 release)[43]. Functional annotations were retrieved with Trinotate based on blast results against Pfam and protein domains identified using eggNOG 3.0[44], The Gene Ontology (GO)[45] and Kyoto Encyclopedia of Genes and Genomes (KEGG)[46,47]. We focused on the most abundant GO terms dataset corresponding to biological processes, molecular functions, and cellular compartments with an arbitrary value of > 5000 TPM for each GO term.

Taxonomy was assigned to clean reads using the TaxMapper search tool and corresponding database with default settings[48]. Reads were mapped to two taxonomic levels: seven main eukaryotic lineages (supergroups, e.g., Alveolata) and 28 groups within these lineages (e.g., Apicomplexa, Chromerida, Ciliophora, Dinophyta and Perkinsea within the Alveolata supergroup).

To validate the process of the de novo assembly, we mapped transcripts in our core dataset against metatranscriptomic data from the *Tara* Oceans expeditions, including *Tara* Oceans Polar Circle sampled in 2013. The reads mapping pipeline used is the same as described previously[49]. Briefly, reads from each *Tara* Oceans' metatranscriptomic read set were mapped onto transcript isoforms in our core dataset using bwa[50] and 95% identity over at least 80% of the length of the read picking the best match (or in case of several putative best matches—a random one). Annotation of the top ten abundant transcripts was done by a blastn search of the NCBI nucleotide database followed by a manual curation.

**Statistical methods.** The map of Adventfjorden was generated using MATLAB R2013b. All statistical analyses were performed in R v3.5.2[51], and data were visualised using *tidyverse* v1.2.1[52]. Diversity indices (richness, Shannon-Wiener index, and Pielou's evenness) were calculated using vegan package v2.5-4[53]. Principal component analysis (PCA) of environmental factors was calculated on centred and scaled data with *prcomp*

function (*stats* package v3.5.3) and visualised using *factoextra* v1.0.5. To explore differences between transcript abundances a Bray–Curtis dissimilarity matrix (*vegdist* function in *vegan* package v2.5–4[53] was constructed and clustered using a 'complete' method within *hclust* function (*stats* package v3.5.3). *Pvclust* package was used to assign support to the clustering topology[54]. To identify the strongest contribution of individual transcript isoforms to clustering patterns, we applied the *simper* function on the transcript matrix.

To explore GO annotations, for each metatranscriptome, we summarised relative counts for each transcript that was assigned to a specific GO term. We explored each of the three categories of GO terms: molecular functions, biological processes, and cellular components. For each category, a Bray–Curtis dissimilarity matrix of GO abundance tables was used to calculate global non-metric multidimensional scaling (GNMDS)[55]. The *envfit* function (*vegan* package) was used to fit environmental parameters onto the GNMDS ordination. Analysis of similarities (*ANOSIM*; *vegan* package) was used to test if there were differences between polar day and polar night associated with light. The *simper* function (*vegan* package) was used with 999 permutations to elucidate GO terms that contributed the most to the difference between polar night and polar day within the three GO categories. In this analysis, the September sample was excluded due to being from a time of a transition between polar day and polar night. *Simper* analysis identified GO terms that differed between polar day and polar night. These features then considered "overrepresented" if the differences in means were statistically significant.

## Results

### Seasonality.
Our study spanned over 13 months and included two polar nights (three and two samples, respectively), one polar day (five samples) and one sample from September coinciding with the transition period between polar day and polar night. Environmental parameters showed a seasonal pattern (Table 1, Supplementary Figure S2). This is a representative trend for the IsA time series station that displays a yearly recurrent pattern. Photosynthetically active radiation (PAR) at 25 m depth was detectable only between April and September 2012. Within this period the highest values were measured in April and the beginning of May 2012, followed by the lowest detected values at the end of May and June. Hydrography of Arctic fjords can be influenced by water masses originating from distinct sources and thus displaying different physiochemical properties categorised based on temperature and salinity[56,57]. Locally formed cold water (< 1° C; LW) was present in the first half of the year (December 2011 to May 2012) with warmer intermediate water (> 1° C; IW) influenced by land runoff and oxygen-rich Atlantic water dominated in the second half (June 2012 to January 2013). The coldest temperature was in January 2012 and the warmest in September 2012. Overall, nutrient concentrations (nitrate/nitrite, phosphate, and silicates) were heavily depleted from the onset of spring bloom until the end of polar day (from May to August; Table 1). However, silicates started to be depleted already in April (Table 1). Chlorophyll *a* was detectable throughout the year with a peak value in May and a second smaller peak in August. In all samples except those collected in May, most of the chlorophyll *a* was present in the small phytoplankton fraction (< 10 μm). Detailed descriptions of the IsA system, based on enhanced frequency of sampling can be found in[10,12,58,59].

### Seasonal transcript diversity.
The diversity and evenness of transcripts was higher during polar night (n = 5) than during polar day (n = 5; Fig. 2). The mean number of transcripts collected during polar day was similar to the value obtained in September, during the transition period between polar day and polar night (the average number of transcripts during polar day $\mu_{PD}$ = 1,178,988, the standard deviation of transcripts during polar day $\sigma_{PD}$ = 273,108 and 1,272,116, respectively), whereas average transcript diversity during polar night was

| | Light parameters | | | Chlorophyll *a* biomass | | Nutrient concentrations | | | Physical parameters | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sampling date | PAR [μmol m$^{-2}$ s$^{-1}$] | Declination [°] | Day length [h] | Total chl *a* [μg/l] | chl *a* > 10 μm [μg/l] | NO$_3$ and NO$_2$ [μM] | PO$_4$ [μM] | Si(OH)$_4$ [μM] | Water mass | Temp [°C] | Salinity |
| 2011-Dec-14 | BD | −23.20 | 0 | 0.024 | 0.003 | 2.81 | 0.15 | 1.51 | LW | 0.9 | 34.32 |
| 2012-Jan-17 | BD | −20.81 | 0 | 0.04 | 0.01 | NA | NA | NA | LW | −1.3 | 34.18 |
| 2012-Jan-28 | BD | −18.29 | 0 | 0.036 | 0.016 | 7.22 | 0.3 | 4.48 | LW | −0.8 | 34.24 |
| 2012-Apr-26 | 6.1 | 13.72 | 24 | 1.42 | 1.261 | 4.49 | 0.38 | 1.65 | LW | −0.2 | 34.56 |
| 2012-May-10 | 6.9 | 17.79 | 24 | 3.794 | 3.163 | 1.54 | 0.24 | 0.28 | LW | 0.3 | 34.57 |
| 2012-May-30 | 0.5 | 21.86 | 24 | NA | NA | BD | 0.23 | 0.85 | LW | 0.3 | 34.43 |
| 2012-Jul-06 | 0.3 | 22.62 | 24 | 0.449 | 0.039 | 0.21 | 0.08 | 0.43 | IW | 2.1 | 34.24 |
| 2012-Aug-06 | 3.4 | 16.51 | 24 | 1.04 | 0.024 | 0.25 | 0.12 | 1.42 | IW | 2.9 | 34.12 |
| 2012-Sep-18 | 1.7 | 1.62 | 13 | 0.206 | 0.042 | 2.6 | 0.3 | 2.45 | IW | 3.8 | 34.36 |
| 2012-Nov-29 | BD | −21.58 | 0 | 0.047 | 0.009 | 5.42 | 0.45 | 3.03 | IW | 1.8 | 34.39 |
| 2013-Jan-10 | BD | −21.89 | 0 | 0.02 | 0.007 | 5.41 | 0.57 | 2.3 | IW | 1.5 | 34.67 |

**Table 1.** Environmental parameters corresponding to each metatranscriptome sampled at 25 m depth. PAR (photosynthetically active radiation) was measured as close to local noon as possible; declination was calculated for local noon, and day length refers to the number of hours when the sun is above the horizon. Chlorophyll *a* biomass is reported for 2 size fractions: total (filtered on GF/F glass microfiber filters (Whatman, England) and > 10 μm (filtered on Isopore membrane polycarbonate filters (Millipore, USA)). Water masses: LW—local water and IW—intermediate water. Other abbreviations: BD—below detection, NA—not available. The data were originally published in[56,57].
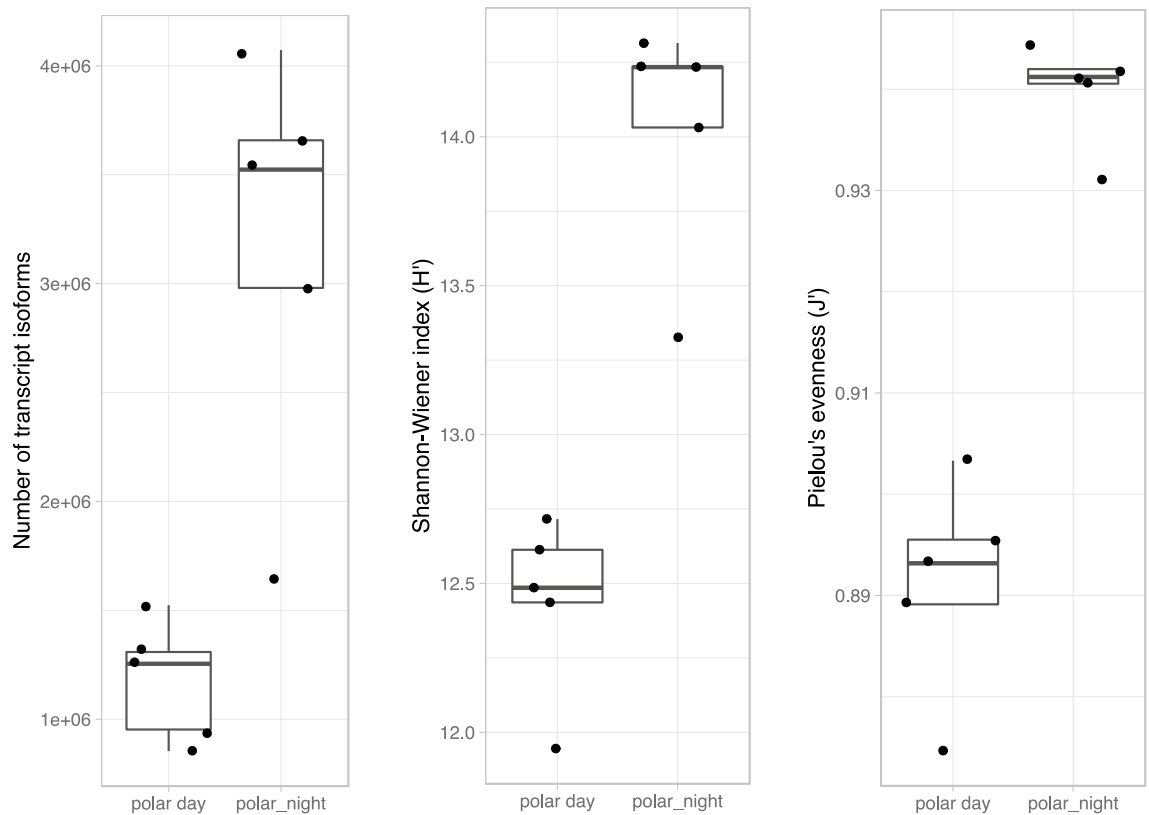
**Figure 2.** Diversity indices in the total dataset from polar day (n = 5) and polar night (n = 5). Data obtained in September, only sample taken during transition period between polar day and polar night, are not plotted.

2.7 times higher. However, the sample collected in January 2012 outlied significantly from the other polar night metatranscriptomes, containing ca. 1,600,000 transcript isoforms, a similar value to samples from polar day and September. Both the September 2012 and January 2012 samples that had low numbers of transcripts also had significantly lower depth of sequencing than the other samples (Supplementary Table S1).

We found a clear difference between metatranscriptomes from polar day and polar night with the September sample clustering with the polar night samples with high support (> 99% of both unbiased and bootstrap probability; Fig. 3). The polar day samples formed distinct subclusters. The core dataset containing almost 70,000 of the most abundant transcripts showed similar or identical clustering, indicating that the pattern was not altered by the high contribution of rare transcripts. Further functional descriptions were therefore conducted using the core dataset.

We identified the transcripts with the strongest contribution to the differences between the main clusters (Supplementary Table S2). Ten of the transcripts contributing to the difference between polar night and polar day were also the most abundant transcripts in our dataset. Only the most abundant transcript out of these was functional annotated by our pipeline and was classified as cytochrome b (Supplementary Table S3). Manual annotation of the remaining nine transcript isoforms revealed that six mapped to subunit III of cytochrome oxidases from different clones of *Karlodinium micrum* (Supplementary Table S3). The remaining three transcripts were either less than 25% percent identical to any NCBI sequence or no significant similarity was found.

**Activity of expressed genes in a seasonal perspective.** *Through the lens of biological processes.* Annotation of the core dataset gene catalogue against the GO database resulted in 24,643 transcripts with at least one annotation (36% of the core dataset). Environmental variables fitted into biological processes (GO category) dissimilarity matrix confirmed the importance of light as a structuring parameter (i.e., day length ($R^2_{GNMDS}$ = 0.88, $p$ = 0.019), declination ($R^2_{GNMDS}$ = 0.85, p = 0.025) and PAR ($R^2_{GNMDS}$ = 0.54, $p$ = 0.082). On the other hand, the analysis did not support water mass ($R^2_{GNMDS}$ = 0.04, $p$ = 0.974) and temperature ($R^2_{GNMDS}$ = 0.20, $p$ = 0.475) as important explanatory variables in structuring biological processes.

The most abundant GO terms within biological processes belonged to housekeeping genes encoding proteins involved in translation, microtubule-based processes, respiratory electron transport chain or protein folding etc. (Fig. 4A). Majority of the most abundant biological processes were overrepresented in polar day samples, such as respiratory electron transport chain or cytoplasmic translation (Fig. 4A). Some GO terms were more uniformly distributed throughout the year, such as cell or mitotic nuclear division (Fig. 4A). Finally, a few of the most abundant GO terms were overrepresented during polar night. This was the case for one-carbon metabolic processes (mean number of TPM in polar night samples, $\mu_{PN}$ = 1974, $\mu_{PD}$ = 1134 in polar day samples, $p$ = 0.03), response to stress ($\mu_{PN}$ = 1482 in polar night, $\mu_{PD}$ = 498 in polar day, $p$ = 0.01) and phototransduction ($\mu_{PN}$ = 936
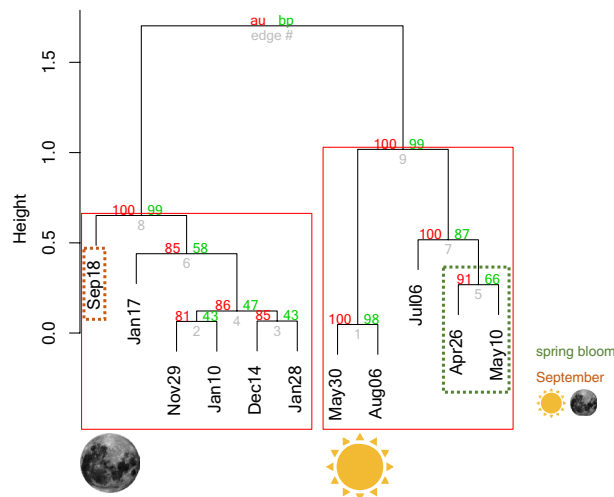
**Figure 3.** Grouping of the samples according to the similarity in their transcript composition based on the core dataset. Approximately unbiased (au) and bootstrap probability (bp) values strongly support the clustering (au and bp > 80). Note that two main highly supported groups are delineated according to the light regime: polar day and times of the year with night present, i.e., polar night and September. The polar day cluster was divided into two groups with strong support (au and bp > 99).

in polar night, $\mu_{PD} = 323$ in polar day, $p = 0.03$). Most transcripts within one-carbon metabolic processes mapped to adenosylhomocysteinase and S-adenosylmethionine synthase. The latter catalyses hydrolysis of L-methionine into S-adenosyl-L-methionine which is an essential source of different chemical groups, e.g. methyl groups used for epigenetic modifications including DNA methylation[60,61]. Whereas adenosylhomocysteinase catalyses one of the next reactions in methionine metabolism: hydrolysis of S-adenosyl-L-homocysteinase to adenosine and L-homocysteine[62] and has been connected to silicon[63] and vitamin starvation in diatoms[64]. All transcripts in response to stress mapped to chaperone proteins, and most (451 out of 456) mapped to different types of heat shock proteins, especially HSP90 (406 transcript isoforms).

Almost all light-dependent biological processes were relatively more abundant in polar day samples. This was especially true for GO terms connected to photosynthesis. However, most categories were also present during polar night albeit in low numbers. Three terms were more abundant in polar night, such as eye photoreceptor cell development, phototaxis, and especially phototransduction. Phototransduction category contained 208 transcripts mapping to green- and blue-light absorbing proteorhodopsins.

Most transcripts contributing to less abundant GO terms, but overrepresented during polar night (Fig. 5), mapped to multipurpose proteins, mainly chaperones (HSP72 and HSP71 in protein refolding, HSP72 in negative regulation of cellular response to heat or response to virus). Phagocytosis and response to other organism categories consisted mostly of transcripts assigned to calreticulin, a multipurpose protein acting as calcium-level regulator and chaperone in endoplasmic reticulum[65]. Pathogenesis contained mostly tripeptidyl-peptidase transcripts and acidic proteases probably involved in virulence response[66]. Response to cycloheximide, a naturally occurring fungicide, contained transcripts mapping to 60S ribosomal protein L44.

*Through the lens of molecular functions.* Most GO terms within molecular functions were overrepresented in polar day (Fig. 4B). Analyses indicated light, but not water masses, to be an important structuring factor of the most abundant molecular functions of the community (R2$_{GNMDS}$ = 0.795, $p$ = 0.005 versus R$^2$$_{GNMDS}$ = 0.017, $p$ = 0.897, respectively). Only DNA binding ($\mu_{PN}$ = 6766 and $\mu_{PD}$ = 4714, $p$ = 0.024), adenosylhomocysteinase activity ($\mu_{PN}$ = 1585 and $\mu_{PD}$ = 889, $p$ = 0.017), photoreceptor ($\mu_{PN}$ = 936 and $\mu_{PD}$ = 315, $p$ = 0.055) and light-activated ion channel activity ($\mu_{PN}$ = 883 and $\mu_{PD}$ = 247, $p$ = 0.025) were overrepresented in polar night.

DNA binding is a broad category of gene products identified as reacting selectively in a non-covalent manner with DNA. We identified 1651 transcripts containing mostly major basic nuclear proteins, histones, cold shock proteins *etc*. Light-activated ion channels and photoreceptors consisted mostly of identified proteorhodopsins; additionally, photoreceptors contained also transcripts mapping to centrins. Centrins are calcium-binding proteins involved in centrosome and microtubule functioning[67], as well as regulation of signalling and molecular translocation[68]. Among less abundant molecular functions overrepresented during polar night, we found that the transcripts mapped mostly to multipurpose proteins, similarly to biological processes. Chromatin binding consisted of diverse proteins, with most transcripts mapping to 60S ribosomal protein L29. Fumarate reductase (NADH) activity consisted of transcripts mapping to an enzyme that catalyses the reversible anaerobic reduction of succinate to fumarate, generating NADH and protons[69]. Sialic acid, phospholipase and oligosaccharide binding contained transcripts mapping mainly to e-selectin, a protein involved in an inflammatory response that changes properties of the cell surface[70]. Mapping to heat shock-related 70 kDa proteins was found in glycolipid binding, whereas nucleolin in nucleosome binding. Nucleolins are also plurifunctional proteins that play important roles in viral infections[71]. 17-beta-hydroxysteroid dehydrogenase (NAD +) activity and
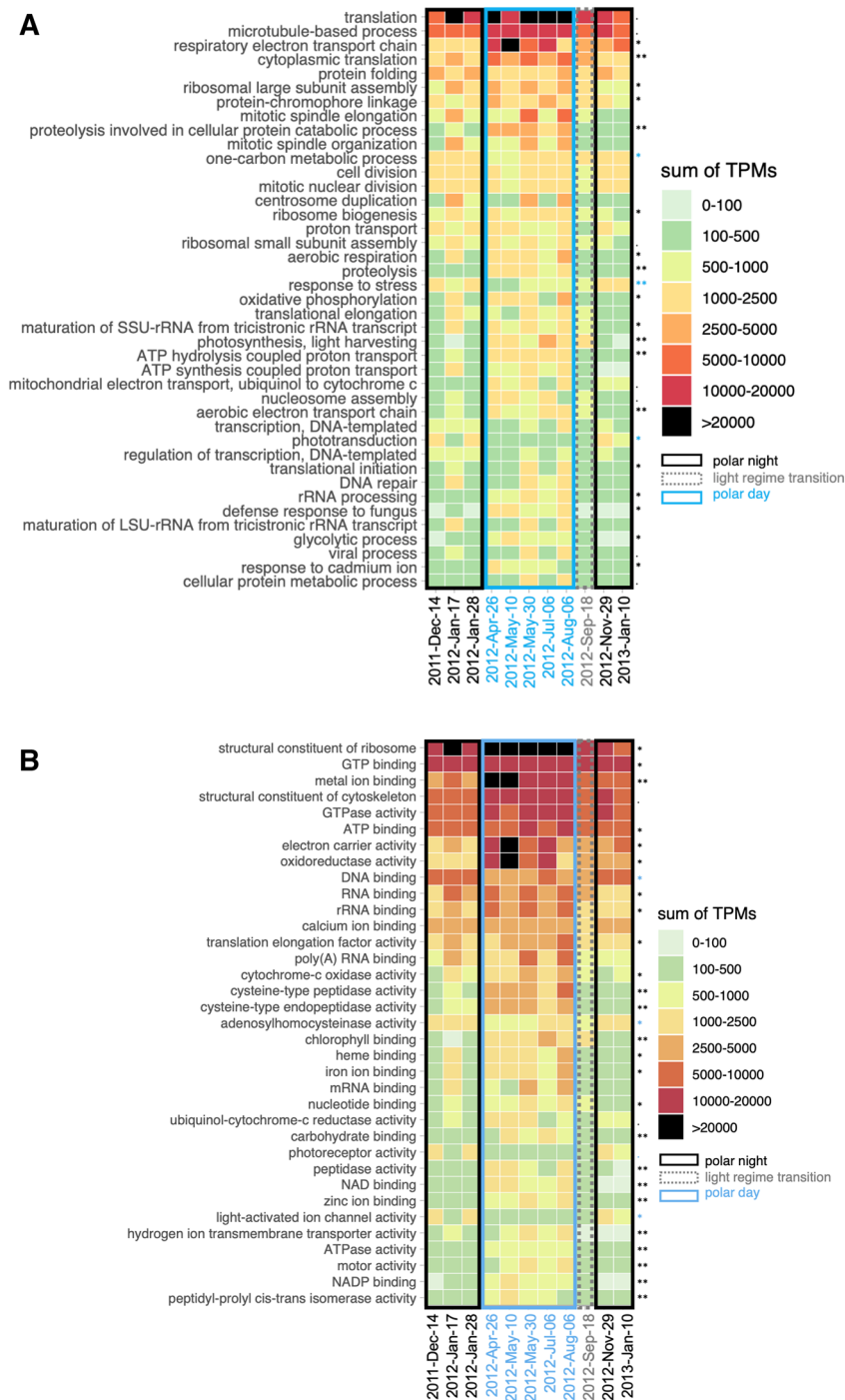
**Figure 4.** The most abundant GO terms within the core dataset corresponding to biological processes (**A**) and molecular functions (**B**) featuring GO terms with >5000 TPM (sum in all 11 samples). Asterisks indicate functions that differed between polar day (blue asterisk) and polar night (black asterisk) using the *simper* function (September was excluded from this analysis). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

3alpha,7alpha,12alpha- trihydroxy-5beta-cholest-24-enoyl-CoA hydratase activity contained the same transcript isoforms that mapped to peroxisomal multifunctional enzymes taking part in β-oxidation of lipids[72] but could also be necessary in fungal pathogenesis[73].
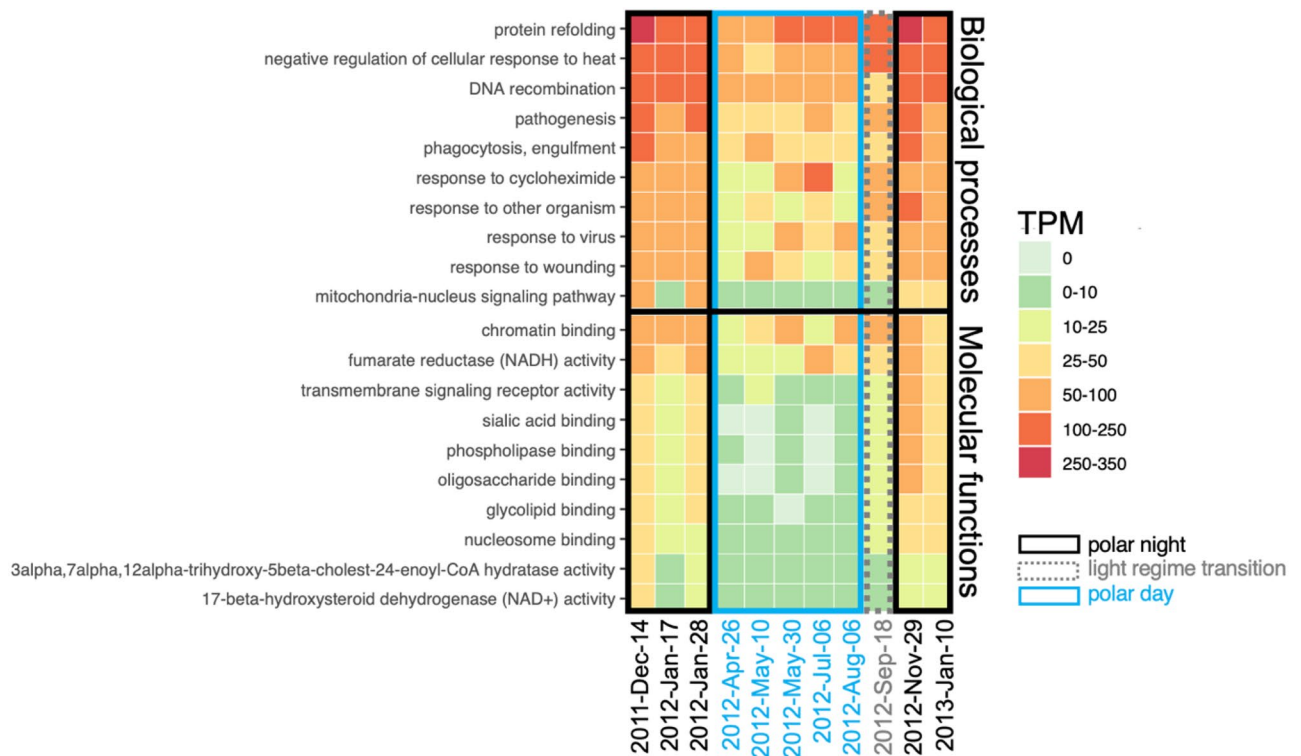
**Figure 5.** Top 10 less abundant GO terms (sum of TPM in all metatranscriptomes < 5000) with mean number of transcripts higher during polar night than during polar day ($p < 0.05$).

*Low level of functional annotations.* Levels of functional annotation were overall low, regardless of the database used. Mapping to UniProt (with blastp), Pfam, TmHMM, GO (based on Pfam) resulted in < 10% of transcript annotation, while eggNOG and KEGG gave 10–20% successful annotation. Only UniProt (with blastx) and GO (with blastp) annotated 38% and 36% of transcripts, respectively. However, mapping our assembled transcripts to the *Tara* Oceans datasets showed that most of our transcripts had hits, matching especially samples from the Arctic (Fig. 7). Up to 75% of our transcript isoforms mapped to the surface samples at station 196, north of Alaska, up to 78% mapped to the deep chlorophyll maximum layer at station 173, northeast of Novaya Zemlya, and up to 74% to the mesopelagic zone at station 201 in west part of Baffin Bay. The mean proportion of transcripts mapping to surface samples from *Tara* Oceans stations located north of 60°N was much higher than for stations in the temperate and tropical regions ($\mu_{\uparrow 60N} = 64\%$, $\sigma_{\uparrow 60N} = 9\%$ and $\mu_{\downarrow 60N} = 21\%$, $\sigma_{\downarrow 60N} = 8\%$, respectively). This was also true for samples from the deep chlorophyll maximum depth ($\mu_{\uparrow 60N} = 69\%$, $\sigma_{\uparrow 60N} = 12\%$ and $\mu_{\downarrow 60N} = 23\%$, $\sigma_{\downarrow 60N} = 9\%$, respectively) and mesopelagic depths 67% ($\sigma_{\uparrow 60N} = 6\%$).

*Taxonomic composition.* The ratio of reads that could be assigned to taxonomic groups was low, however similar throughout the year, ranging from 33 to 42% of all reads in each metatranscriptome. This left the majority of reads without a taxonomic annotation (58–67%). The proportion of taxonomically unannotated reads was independent of light regime and number of sequences per sample. The most represented supergroup in each sample was Alveolata, predominantly Dinophyceae and Ciliophora (Fig. 7). Dinophyceae dominated in metatranscriptomes from polar night (32% on 17th January 2012 up to 49% on 14th December 2011) and September (33%), while ciliates were more abundant during polar day (18–34% vs. 8–10% in polar night). Many taxonomic groups had low relative transcript abundance throughout the year, never exceeding 2% of the taxonomically assigned reads (Apusozoa, Bigyra, Cercozoa, Chromerida, Euglenozoa, Fornicata, Fungi, Glaucocystophyceae, Heterolobosea, Parabasalia, Perkinsea, Pseudofungi and Rhodophyta).

## DISCUSSION

Climate change is already influencing Arctic marine ecosystems[74] and different future scenarios of their development have been suggested. However, predicting the influence of climate change on polar ecosystems is challenging without a deep understanding of the structure and function of its components[75]. Thus, the responses of microbial communities to these changes cannot be predicted without understanding which biological and molecular activities are taking place and how they impact biogeochemical cycles. Differences in gene expression could change the outcome of trophic interactions in an ecosystem, potentially altering the energy and nutrient flow to higher trophic levels[26,75]. In this study we went beyond reporting detected species or their molecular proxies by examining community-level molecular engagement in biological activities and molecular functions.

Our study offers a first description of the key processes performed by the microbial eukaryotic community over seasons in the Arctic fjord.

The strong seasonality at high latitudes affects microbial eukaryotes by influencing cell counts, biomass distribution, community composition, dominating carbon acquisition mode and various biodiversity measures[10,12,76]. Therefore, seasonal gradients profoundly affect the overall pool of present genes and their products, i.e., gene transcripts or proteins. Higher OTU richness of marine protists during polar night compared to polar day was described independently in distant parts of the Arctic marine waters (e.g.,[10,77]). The same patterns were shown for other Arctic marine microorganisms, such as bacteria and archaea[78]. This likely panarctic phenomenon could originate from physically driven mixing throughout the water column which could enable the detection of diverse microorganisms at various atypical water depths during polar night [e.g.,[14]]. In this sense, mixing increases species diversity at different water depths. Temperature and salinity profiles during polar night are uniform throughout the water column at IsA[10] and thus could support this explanation. We did not find unequivocal evidence for increased functional diversity in microbial eukaryotes' transcript, i.e., expression of a wider array of genes needed for survival. In line with previously published evidence for higher richness of microbial eukaryotes during polar night, we showed that the diversity and evenness of their transcripts were also higher during polar night (Figs. 2, 6). The proportions of transcripts belonging to predominantly photosynthetic protists such as diatoms, haptophytes and chlorophytes, were consistently lower during polar night, confirming lower representation in the community and perhaps also lower overall activity of organisms in these groups[10,14]. However, despite the high diversity of OTUs and transcripts, cell counts, and therefore biomass of protists remained low throughout polar night[12,79–81].

The active status of Dinophyta, the most abundant taxonomic group during polar night[10], was reflected in high levels of transcripts involved in the mitochondrial electron transport chain. Cytochrome b and different variants of subunit III of cytochrome c oxidases (COX3), that were abundant in our metatranscriptomes, enable electron transport, by supporting an electrochemical gradient across the membrane and hence contribute to energy generation in the cells[82–84]. Retrieval of mitochondrial genes in our study was unexpected due to poly(A) selection step during sample preparation, however oligo– and polyadenylation of these genes has been described in Dinophyta[85], and other protists, such as Apicomplexa[86].

During polar night the contribution of single species to the overall low pool of biomass is more even than at any other time of the year, especially spring bloom[10,12]. This includes crucial primary producers, such as *Micromonas polaris*, which were encountered as active at different depths of Arctic marine habitats during polar night[14]. The persistence of low levels of light-dependent biological processes in primary producers during polar night is likely due to the persistence and perhaps even maintenance of a functional photosynthetic apparatus kept ready to be activated once the light comes back[13,30,87]. Therefore, an overrepresentation of eukaryotic proteorhodopsins mapping exclusively from Dinophyta during polar night was rather unexpected. Bacterial proteorhodopsins are known as an alternative pathway to photosynthesis for harnessing solar energy in the
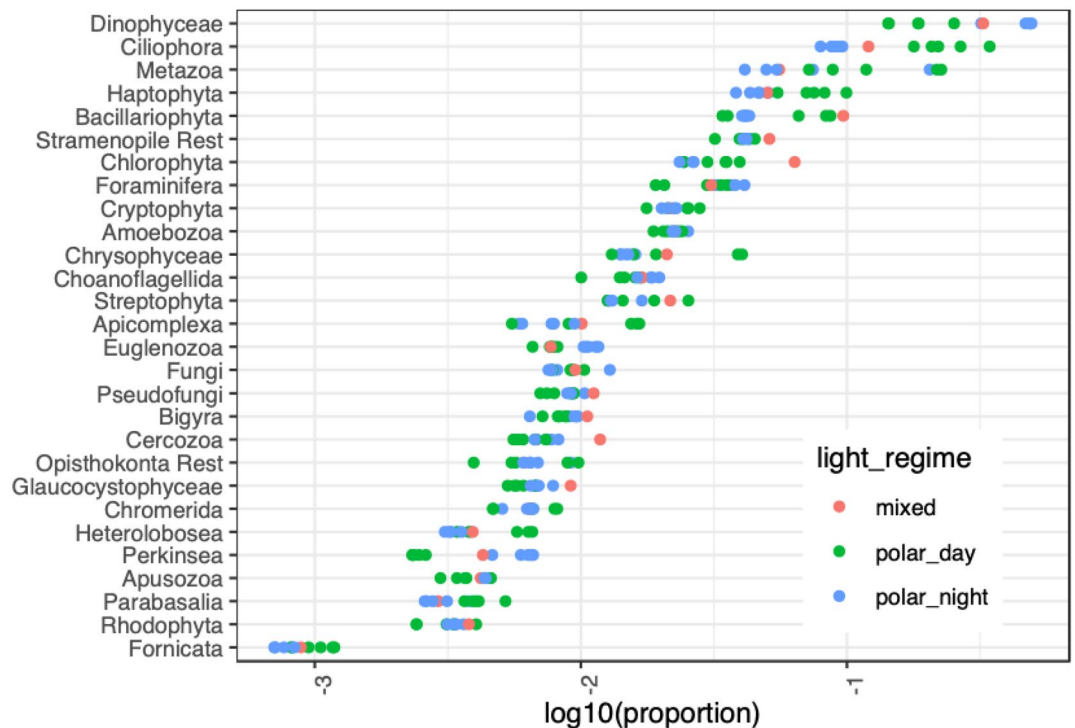


**Figure 6.** Taxonomic assignment shown as the proportion of clean reads assigned to a taxonomic group with Taxmapper. Each dot represents the proportion of reads in one sample.
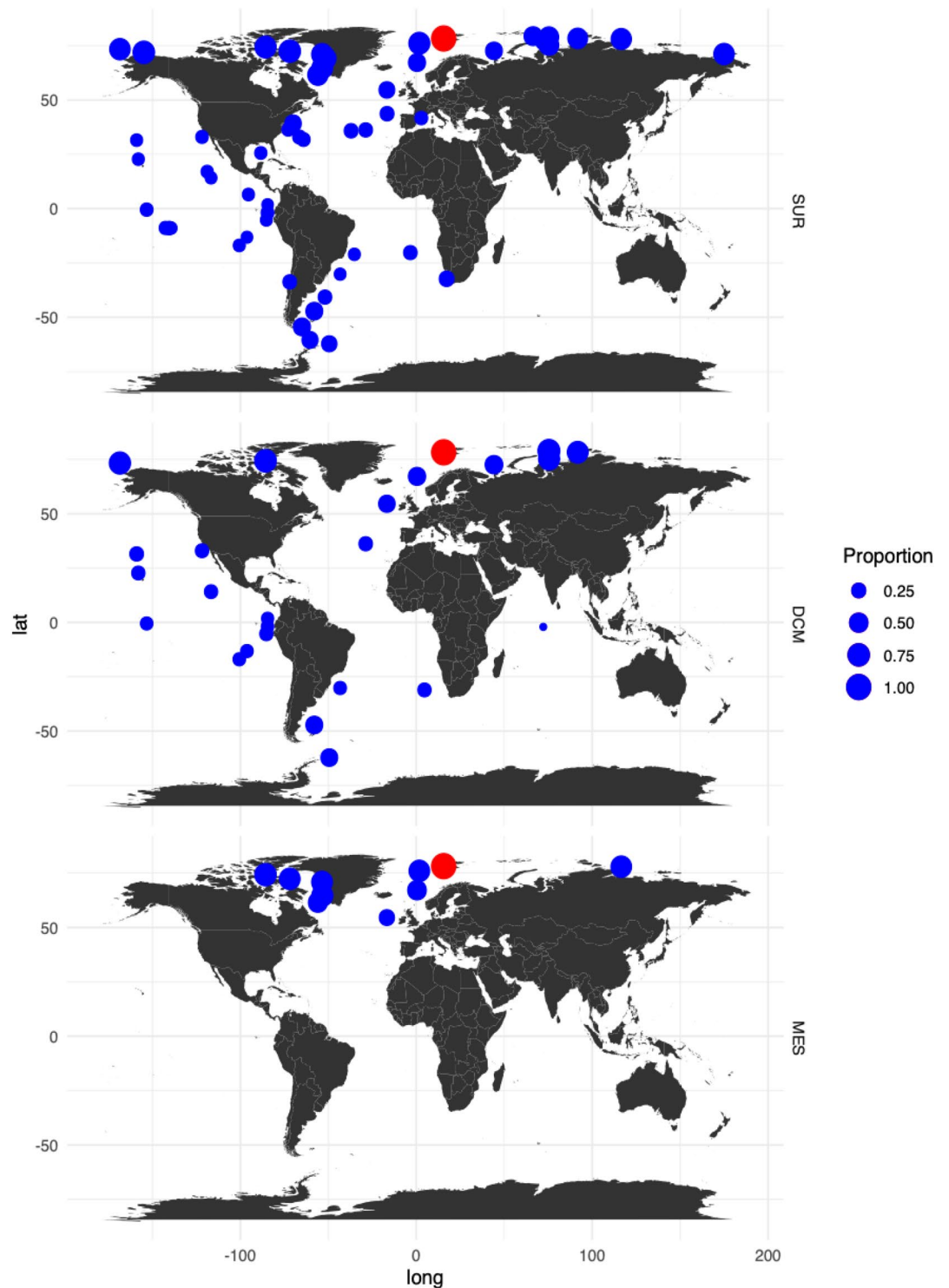
**Figure 7.** Proportion of transcripts isoforms from the IsA core dataset (red circle) with matches in metatranscriptomes from Tara Oceans stations (blue circles). The Tara Oceans dataset to which our core transcriptome was mapped originated from the 0.8–2000 μm plankton size fraction collected at different depths (SUR—surface waters, DCM—deep chlorophyll maximum and MES—mesopelagic waters). The highest proportion of matches between our dataset and the Tara Oceans samples was from stations above the polar circle. The map was created using *tidyverse* package v1.2.1[52] in R v3.5.2[51].

ocean[88]. It is not clear what is their function in microbial eukaryotes[89]. However, recently, it was suggested that they are involved in G protein-coupled receptor-based signalling in Dinophyta[90]. In marine diatoms microbial rhodopsins have been shown to acidify plastids through proton pumping, and a role as a carbon dioxide concentration mechanism inside of the cells has been proposed[91].

Gene expression is likely to be more strictly controlled in many organisms during polar night due to the overall lower availability of energy in the ecosystem[4]. In our dataset we found several GO functions that might hint at expression of genes that are involved in energy conservation. An increased expression of histones or major binding nuclear proteins or similar genes could serve as a way to control gene expression by binding and thus preventing DNA from being transcribed[92]. On the other hand, it may also point towards cellular division and the need to produce new histones for new cells[93]. GO term classification of transcripts overrepresented during polar night covers mostly categories such as response to stress, cellular signalling, modifications in cytoskeleton, pathogenesis, etc., through proteins that are known to be multifunctional. Multifunctionality might be an important strategy for efficient use of resources that could limit some groups of organisms during polar night. Other functions overrepresented during polar night involve adenosylhomocysteinase which could play an important role in increasing the lifespan of microbial eukaryotes by controlling the concentration of methionine[94]. In general, biochemical reactions involved in methionine degradation are the main source of methyl groups used in gene silencing by DNA methylation. This could be another argument for strict control of gene expression during polar night[94]. Overrepresentation of different types of chaperon and heat shock protein transcripts during polar night may be connected to the high demand for energy conservation by assuring correct assembly, maintenance and stability of proteins' structures within the cell[95,96]. Moreover, heat shock proteins could influence increased cell survival by several mechanisms able to attenuate apoptosis[97].

Among the most expressed transcripts in our study were a few functionally annotated sequences reaching up to 38% of the total number of transcripts which coincides with similar numbers of taxonomically annotated reads in our study. Metatranscriptomic studies often report low levels of functional annotations (down to 19%) that might be a result of various factors, such as the complexity of the studied environment[98], available reference databases[49], choice of algorithms, bioinformatic tools and parameters used for data analysis[99], *etc*. To date, the most extensive marine global survey examining expressed eukaryotic genes based on *Tara* Oceans 2009–2012 reported 51.2% unannotated clusters of expressed genes[49]. Although the overall rates of annotation in our study were low, the data mapped successfully against the *Tara* Oceans dataset (including *Tara* Oceans Polar Circle campaign in 2013) by matching to 78% of transcript isoforms, specifically in polar regions (Fig. 7). Therefore, we conclude that de novo assembled transcripts in our bioinformatic pipeline were robust and contained novel genes that are less likely to be found at lower latitudes (Fig. 7). We hypothesize that the proportion of successfully mapping transcripts in our study would have been higher if the *Tara* Oceans campaign in the Arctic was extended beyond June–October to collect samples during polar night.

The high proportions of transcripts mapping to *Tara* Oceans' metatranscriptomes from the Arctic suggest a distinct genetic makeup of microbial eukaryotes in this part of the world, in agreement with the proposed existence of a distinct Arctic community structure[100]. Perhaps the different genetic makeup of eukaryotic communities in high latitudes could reflect necessary adaptations to Arctic seasonality that are not present in potential invasive microbial eukaryotes moving northwards with progression of climate warming. This coincides with recently revealed biodiversity hotspot of DNA viruses in the Arctic Ocean[101]. Since viruses and mobile elements carried by them are known as powerful agents of evolution in all living cells[102], we speculate that viruses could potentially contribute to increased diversification of genes in microbial communities in the Arctic[103], resulting in the observed low transcript similarity to other parts of the ocean. The fact of distinct genetic make-up of microbial eukaryotes and viruses coexisting within the same ecosystem requires a separate dedicated line of research, and regardless of possible links between the two groups, the distinct genetic makeup of Arctic microbial eukaryotes warrants further exploration.

Polar night seems to work as a reset stage for Arctic marine environments, possibly enforcing shifts to heterotrophy in the absence of light and allowing protist survival as low biomass populations. Moreover, changing proportions of transcripts annotated to taxonomic groups as well as fluctuating abundances of functional categories point out that community-level metabolic state changes together with shifting community composition. The two polar nights in our study showed a striking similarity in the taxonomic and functional composition of transcripts that might reflect a specific, recurrent impact of environmental filtering imposed by seasonal light regime and temperature. A long-term monitoring of taxonomic and transcriptional dynamics could evaluate to which extent other factors such as the inflow of warmer water masses or the arrival of species moving northwards, influence the strength of light regime filtering and development of future eukaryotic communities.

## Data availability

## References

1. Judson, O. P. The energy expansions of evolution. *Nat. Ecol. Evol.* **1**(6), 0138. https://doi.org/10.1038/s41559-017-0138 (2017).
2. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**(5374), 237–240. https://doi.org/10.1126/science.281.5374.237 (1998).
3. Boyce, D. G., Petrie, B., Frank, K. T., Worm, B. & Leggett, W. C. Environmental structuring of marine plankton phenology. *Nat. Ecol. Evol.* **1**(10), 1484–1494. https://doi.org/10.1038/s41559-017-0287-3 (Oct.2017).

4.  Berge, J. *et al.* In the dark: A review of ecosystem processes during the Arctic polar night. *Prog. Oceanogr.* **139**, 258–271. https://doi.org/10.1016/j.pocean.2015.08.005 (Dec.2015).
5.  Berge, J. *et al.* Unexpected levels of biological activity during the polar night offer new perspectives on a warming arctic. *Curr. Biol.* **25**(19), 2555–2561. https://doi.org/10.1016/j.cub.2015.08.024 (Oct.2015).
6.  Błachowiak-Samołyk, K. *et al.* Winter Tales: The dark side of planktonic life. *Polar Biol.* **38**(1), 23–36. https://doi.org/10.1007/s00300-014-1597-4 (2015).
7.  Bunse, C. & Pinhassi, J. Marine bacterioplankton seasonal succession dynamics. *Trends Microbiol.* **25**(6), 494–505. https://doi.org/10.1016/j.tim.2016.12.013 (2017).
8.  Moreira, D. & López-García, P. Time series are critical to understand microbial plankton diversity and ecology. *Mol. Ecol.* **28**(5), 920–922. https://doi.org/10.1111/mec.15015 (2019).
9.  Buttigieg, P. L. *et al.* Marine microbes in 4D—using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Curr. Opin. Microbiol.* **43**, 169–185. https://doi.org/10.1016/j.mib.2018.01.015 (2018).
10. Marquardt, M., Vader, A., Stübner, E. I., Reigstad, M. & Gabrielsen, T. M. Strong seasonality of marine microbial eukaryotes in a high-Arctic fjord (Isfjorden, in West Spitsbergen, Norway). *Appl. Environ. Microbiol.* **82**(6), 1868–1880. https://doi.org/10.1128/AEM.03208-15 (2016).
11. Marquardt, M. *et al.* Vertical export of marine pelagic protists in an ice-free high-Arctic fjord (Adventfjorden, West Spitsbergen) throughout 2011–2012. *Aquat. Microb. Ecol.* **83**(1), 65–82. https://doi.org/10.3354/ame01904 (2019).
12. Kubiszyn, A. M. *et al.* The annual planktonic protist community structure in an ice-free high Arctic fjord (Adventfjorden, West Spitsbergen). *J. Mar. Syst.* **169**, 61–72. https://doi.org/10.1016/j.jmarsys.2017.01.013 (2017).
13. Kvernvik, A. C. *et al.* Fast reactivation of photosynthesis in arctic phytoplankton during the polar night. *J. Phycol.* **54**(4), 461–470. https://doi.org/10.1111/jpy.12750 (2018).
14. Vader, A., Marquardt, M., Meshram, A. R. & Gabrielsen, T. M. Key Arctic phototrophs are widespread in the polar night. *Polar Biol.* **38**(1), 13–21. https://doi.org/10.1007/s00300-014-1570-2 (2014).
15. Metfies, K., von Appen, W.-J., Kilias, E., Nicolaus, A. & Nöthig, E.-M. Biogeography and photosynthetic biomass of Arctic marine pico-eukaroytes during summer of the record sea ice minimum 2012. *PLOS ONE* **11**(2), e0148512. https://doi.org/10.1371/journal.pone.0148512 (2016).
16. Sherr, E. B., Sherr, B. F., Wheeler, P. A. & Thompson, K. 'Temporal and spatial variation in stocks of autotrophic and heterotrophic microbes in the upper water column of the central Arctic Ocean', *Deep Sea Res. Part Oceanogr. Res. Pap.* **50**(5), 557–571. https://doi.org/10.1016/S0967-0637(03)00031-1 (2003).
17. Schaub, I., Wagner, H., Graeve, M. & Karsten, U. Effects of prolonged darkness and temperature on the lipid metabolism in the benthic diatom Navicula perminuta from the Arctic Adventfjorden, Svalbard. *Polar Biol.* **40**(7), 1425–1439. https://doi.org/10.1007/s00300-016-2067-y (2017).
18. Smayda, T. J. & Mitchell-Innes, B. Dark survival of autotrophic, planktonic marine diatoms. *Mar. Biol.* **25**(3), 195–202. https://doi.org/10.1007/BF00394965 (1974).
19. Toseland, A. *et al.* The impact of temperature on marine phytoplankton resource allocation and metabolism. *Nat. Clim. Change* **3**(11), 979–984. https://doi.org/10.1038/nclimate1989 (Nov.2013).
20. McMinn, A. & Martin, A. Dark survival in a warming world. *Proc. R. Soc. B Biol. Sci.* **280**(1755), 20122909. https://doi.org/10.1098/rspb.2012.2909 (2013).
21. McKie-Krisberg, Z. M. & Sanders, R. W. Phagotrophy by the picoeukaryotic green alga Micromonas: implications for Arctic Oceans. *ISME J.* **8**(10), 1953–1961. https://doi.org/10.1038/ismej.2014.16 (2014).
22. Sanders, R. W. & Gast, R. J. Bacterivory by phototrophic picoplankton and nanoplankton in Arctic waters. *FEMS Microbiol. Ecol.* **82**(2), 242–253. https://doi.org/10.1111/j.1574-6941.2011.01253.x (Nov.2012).
23. Stoecker, D. K. & Lavrentyev, P. J. Mixotrophic plankton in the polar seas: A pan-arctic review. *Front. Mar. Sci.* **5**, 292. https://doi.org/10.3389/fmars.2018.00292 (2018).
24. Caron, D. A. *et al.* Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* **15**(1), 6–20. https://doi.org/10.1038/nrmicro.2016.160 (2017).
25. de Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**(6237), 1261605–1261605. https://doi.org/10.1126/science.1261605 (2015).
26. Worden, A. Z. *et al.* Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**(6223), 1257594–1257594. https://doi.org/10.1126/science.1257594 (2015).
27. Keeling, P. J. *et al.* The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**(6), e1001889. https://doi.org/10.1371/journal.pbio.1001889 (2014).
28. Acar, M., Mettetal, J. T. & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40**(4), 471–475. https://doi.org/10.1038/ng.110 (2008).
29. Coles, V. J. *et al.* Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* **358**(6367), 1149–1154. https://doi.org/10.1126/science.aan5712 (2017).
30. van der Poll, W. H., Abdullah, E., Visser, R. J. W., Fischer, P. & Buma, A. G. J. Taxon-specific dark survival of diatoms and flagellates affects Arctic phytoplankton composition during the polar night and early spring. *Limnol. Oceanogr.* **65**(5), 903–914. https://doi.org/10.1002/lno.11355 (2020).
31. Andrews, S. 'FastQC: A quality control tool for high throughput sequence data'. 2010. [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
32. Bushnell, B. 'BBMap'. 2017. [Online]. Available: sourceforge.net/projects/bbmap/
33. Zhao, W. *et al.* Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**(1), 419. https://doi.org/10.1186/1471-2164-15-419 (2014).
34. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611 (2012).
35. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084 (2013).
36. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**(7), 644–652. https://doi.org/10.1038/nbt.1883 (2011).
37. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**(1), 323. https://doi.org/10.1186/1471-2105-12-323 (2011).
38. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**(1), 13. https://doi.org/10.1186/s13059-016-0881-8 (2016).
39. Bryant, D. M. *et al.* A tissue-mapped axolotl De Novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **18**(3), 762–776. https://doi.org/10.1016/j.celrep.2016.12.063 (2017).
40. The UniProt Consortium, 'UniProt: the universal protein knowledgebase', *Nucleic Acids Res.*, vol. 45, no. D1, pp. D158–D169, 2017, doi:https://doi.org/10.1093/nar/gkw1099.
41. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform.* **10**(1), 421. https://doi.org/10.1186/1471-2105-10-421 (2009).

42. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**(12), e121–e121. https://doi.org/10.1093/nar/gkt263 (2013).
43. Finn, R. D. *et al.* The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344 (2016).
44. Powell, S. *et al.* eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* **40**(D1), D284–D289. https://doi.org/10.1093/nar/gkr1060 (2012).
45. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**(1), 25–29. https://doi.org/10.1038/75556 (2000).
46. Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
47. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**(D1), D457–D462. https://doi.org/10.1093/nar/gkv1070 (2016).
48. Beisser, D. *et al.* TaxMapper: An analysis tool, reference database and workflow for metatranscriptome analysis of eukaryotic microorganisms. *BMC Genomics* **18**(1), 787. https://doi.org/10.1186/s12864-017-4168-6 (2017).
49. Carradec, Q. *et al.* A global ocean atlas of eukaryotic genes. *Nat. Commun.* **9**(1), 373. https://doi.org/10.1038/s41467-017-02342-1 (2018).
50. Li, H. 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', ArXiv13033997 *Q-Bio*, May 2013, Accessed 14 May 2020. [Online]. Available: http://arxiv.org/abs/1303.3997
51. R Core Team, 'R: A Language and Environment for Statistical Computing'. R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/
52. Wickham, H. 'tidyverse: Easily Install and Load the "Tidyverse"'. Nov. 21, 2019. Accessed: May 14, 2020. [Online]. Available: https://CRAN.R-project.org/package=tidyverse
53. Oksanen, J. et al. 'vegan: Community Ecology Package'. Sep. 01, 2019. Accessed: May 14, 2020. [Online]. Available: https://CRAN.R-project.org/package=vegan
54. Suzuki, R. & Shimodaira, H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**(12), 1540–1542. https://doi.org/10.1093/bioinformatics/btl117 (2006).
55. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**(1), 1–27. https://doi.org/10.1007/BF02289565 (1964).
56. Svendsen, H. *et al.* The physical environment of Kongsfjorden-Krossfjorden, an Arctic fjord system in Svalbard. *Polar Res.* **21**(1), 133–166. https://doi.org/10.1111/j.1751-8369.2002.tb00072.x (2002).
57. Nilsen, F., Cottier, F., Skogseth, R. & Mattsson, S. Fjord–shelf exchanges controlled by ice and brine production: The interannual variation of Atlantic Water in Isfjorden, Svalbard. *Cont. Shelf Res.* **28**(14), 1838–1853. https://doi.org/10.1016/j.csr.2008.04.015 (2008).
58. Brandner, M. M., Stübner, E., Reed, A. J., Gabrielsen, T. M. & Thatje, S. Seasonality of bivalve larvae within a high Arctic fjord. *Polar Biol.* **40**(2), 263–276. https://doi.org/10.1007/s00300-016-1950-x (2017).
59. Stübner, E. I., Søreide, J. E., Reigstad, M., Marquardt, M. & Blachowiak-Samolyk, K. Year-round meroplankton dynamics in high-Arctic Svalbard. *J. Plankton Res.* **38**(3), 522–536. https://doi.org/10.1093/plankt/fbv124 (2016).
60. Cantoni, G. L. Biological methylation: Selected aspects. *Annu. Rev. Biochem.* **44**(1), 435–451. https://doi.org/10.1146/annurev.bi.44.070175.002251 (1975).
61. Fontecave, M., Atta, M. & Mulliez, E. S-adenosylmethionine: Nothing goes to waste. *Trends Biochem. Sci.* **29**(5), 243–249. https://doi.org/10.1016/j.tibs.2004.03.007 (2004).
62. De La Haba, G. & Cantoni, G. L. The enzymatic synthesis of S-adenosyl-L-homocysteine from adenosine and homocysteine. *J. Biol. Chem.* **234**(3), 603–608 (1959).
63. Hildebrand, M., Higgins, D. R., Busser, K. & Volcani, B. E. Silicon-responsive cDNA clones isolated from the marine diatom Cylindrotheca fusiformis. *Gene* **132**(2), 213–218. https://doi.org/10.1016/0378-1119(93)90198-C (1993).
64. Alexander, H. *et al.* Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl. Acad. Sci.* **112**(44), E5972–E5979. https://doi.org/10.1073/pnas.1518165112 (2015).
65. Michalak, M., Corbett, E. F., Mesaeli, N., Nakamura, K. & Opas, M. Calreticulin: one protein, one gene, many functions. *Biochem. J.* **344**(Pt 2), 281–292 (1999).
66. Reichard, U. *et al.* Sedolisins, a new class of secreted proteases from Aspergillus fumigatus with Endoprotease or tripeptidyl-peptidase activity at acidic pHs. *Appl. Environ. Microbiol.* **72**(3), 1739–1748. https://doi.org/10.1128/AEM.72.3.1739-1748.2006 (2006).
67. Satisbury, J. L. Centrin, centrosomes, and mitotic spindle poles. *Curr. Opin. Cell Biol.* **7**(1), 39–45. https://doi.org/10.1016/0955-0674(95)80043-3 (1995).
68. Gießl, A., Trojan, P., Rausch, S., Pulvermüller, A. & Wolfrum, U. Centrins, gatekeepers for the light-dependent translocation of transducin through the photoreceptor cell connecting cilium. *Vision Res.* **46**(27), 4502–4509. https://doi.org/10.1016/j.visres.2006.07.029 (2006).
69. Tielens, A. G. M. & Van Hellemond, J. J. The electron transport chain in anaerobically functioning eukaryotes. *Biochim. Biophys. Acta BBA - Bioenerg.* **1365**(1–2), 71–78. https://doi.org/10.1016/S0005-2728(98)00045-0 (1998).
70. Bhatia, S. K., Swers, J. S., Camphausen, R. T., Wittrup, K. D. & Hammer, D. A. Rolling adhesion kinematics of yeast engineered to express selectins. *Biotechnol. Prog.* **19**(3), 1033–1037. https://doi.org/10.1021/bp025756b (2003).
71. Hiscox, J. A. The nucleolus: A gateway to viral infection?. *Arch. Virol.* **147**(6), 1077–1089. https://doi.org/10.1007/s00705-001-0792-0 (2002).
72. Winkler, U., Säftel, W. & Stabenau, H. A new type of a multifunctional β-oxidation enzyme in Euglena. *Plant Physiol.* **131**(2), 753–762. https://doi.org/10.1104/pp.013151 (2003).
73. Klose, J. & Kronstad, J. W. The multifunctional β-oxidation enzyme is required for full symptom development by the biotrophic maize pathogen ustilago maydis. *Eukaryot. Cell* **5**(12), 2047–2061. https://doi.org/10.1128/EC.00231-06 (2006).
74. Wassmann, P., Duarte, C. M., Agustí, S. & Sejr, M. K. Footprints of climate change in the Arctic marine ecosystem: FOOTPRINTS of climate change. *Glob. Change Biol.* **17**(2), 1235–1249. https://doi.org/10.1111/j.1365-2486.2010.02311.x (2011).
75. Murphy, E. J. *et al.* Understanding the structure and functioning of polar pelagic ecosystems to predict the impacts of change. *Proc. R. Soc. B Biol. Sci.* **283**(1844), 20161646. https://doi.org/10.1098/rspb.2016.1646 (2016).
76. Randelhoff, A. *et al.* Arctic mid-winter phytoplankton growth revealed by autonomous profilers. *Sci. Adv.* **6**(39), eabc2678. https://doi.org/10.1126/sciadv.abc2678 (2020).
77. Onda, D. F. L. *et al.* Seasonal and interannual changes in ciliate and Dinoflagellate species assemblages in the Arctic Ocean (Amundsen Gulf, Beaufort Sea, Canada). *Front. Mar. Sci.* https://doi.org/10.3389/fmars.2017.00016 (2017).
78. Ladau, J. *et al.* Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**(9), 1669–1677. https://doi.org/10.1038/ismej.2013.37 (2013).
79. Rokkan Iversen, K. & Seuthe, L. Seasonal microbial processes in a high-latitude fjord (Kongsfjorden, Svalbard): I. Heterotrophic bacteria, picoplankton and nanoflagellates. *Polar Biol.* **34**(5), 731–749. https://doi.org/10.1007/s00300-010-0929-2 (2011).
80. Seuthe, L., Rokkan Iversen, K. & Narcy, F. Microbial processes in a high-latitude fjord (Kongsfjorden, Svalbard): II. Ciliates and dinoflagellates. *Polar Biol.* **34**(5), 751–766. https://doi.org/10.1007/s00300-010-0930-9 (2011).

81. Weslawski, J. M., Kwasniewski, S. & Wiktor, J. Winter in a Svalbard fiord ecosystem. *Arctic* **44**(2), 115–123. https://doi.org/10.14430/arctic1527 (1991).
82. Esposti, M. D. *et al.* Mitochondrial cytochrome b: Evolution and structure of the protein. *Biochim. Biophys. Acta BBA - Bioenerg.* **1143**(3), 243–271. https://doi.org/10.1016/0005-2728(93)90197-N (1993).
83. Belevich, I., Verkhovsky, M. I. & Wikström, M. Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase. *Nature* **440**(7085), 829–832. https://doi.org/10.1038/nature04619 (2006).
84. Michel, H. Cytochrome *c* oxidase: Catalytic cycle and mechanisms of proton pumping-A discussion. *Biochemistry* **38**(46), 15129–15140. https://doi.org/10.1021/bi9910934 (1999).
85. Chaput, H., Wang, Y. & Morse, D. Polyadenylated transcripts containing random gene fragments are expressed in dinoflagellate mitochondria. *Protist* **153**(2), 111–122. https://doi.org/10.1078/1434-4610-00090 (2002).
86. Gillespie, D. E., Salazar, N. A., Rehkopf, D. H. & Feagin, J. E. The fragmented mitochondrial ribosomal RNAs of Plasmodium falciparum have short A tails. *Nucleic Acids Res.* **27**(11), 2416–2422. https://doi.org/10.1093/nar/27.11.2416 (1999).
87. Hoppe, C. J. M. 'Always ready? Primary production of Arctic phytoplankton at the end of the polar night', *Limnol. Oceanogr. Lett.*, p. lol2.10222, 2021. https://doi.org/10.1002/lol2.10222.
88. Gómez-Consarnau, L. *et al.* Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci. Adv.* **5**(8), eaaw8855. https://doi.org/10.1126/sciadv.aaw8855 (2019).
89. Slamovits, C. H., Okamoto, N., Burri, L., James, E. R. & Keeling, P. J. A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat. Commun.* **2**(1), 183. https://doi.org/10.1038/ncomms1188 (2011).
90. Mojib, N. & Kubanek, J. Comparative transcriptomics supports the presence of G protein-coupled receptor-based signaling in unicellular marine eukaryotes. *Limnol. Oceanogr.* **65**(4), 762–774. https://doi.org/10.1002/lno.11345 (2020).
91. Yoshizawa, S. *et al.* Proton-pumping rhodopsins in marine diatoms. *Microbiology* https://doi.org/10.1101/2022.01.18.476826 (2022).
92. Prado, F., Jimeno-González, S. & Reyes, J. C. Histone availability as a strategy to control gene expression. *RNA Biol.* **14**(3), 281–286. https://doi.org/10.1080/15476286.2016.1189071 (2017).
93. Salomé, P. A. & Merchant, S. S. A series of fortunate events: Introducing chlamydomonas as a reference organism. *Plant Cell* **31**(8), 1682–1707. https://doi.org/10.1105/tpc.18.00952 (2019).
94. Parkhitko, A. A., Jouandin, P., Mohr, S. E. & Perrimon, N. Methionine metabolism and methyltransferases in the regulation of aging and lifespan extension across species. *Aging Cell* **18**(6), 55. https://doi.org/10.1111/acel.13034 (2019).
95. Frydman, J. Folding of newly translated proteins in vivo: The role of molecular chaperones. *Annu. Rev. Biochem.* **70**(1), 603–647. https://doi.org/10.1146/annurev.biochem.70.1.603 (2001).
96. Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**(6294), aac4354. https://doi.org/10.1126/science.aac4354 (2016).
97. Verbeke, P. Heat shock response and ageing: Mechanisms and applications. *Cell Biol. Int.* **25**(9), 845–857. https://doi.org/10.1006/cbir.2001.0789 (2001).
98. Jiang, Y., Xiong, X., Danska, J. & Parkinson, J. Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome* **4**(1), 2. https://doi.org/10.1186/s40168-015-0146-x (2016).
99. Celaj, A., Markle, J., Danska, J. & Parkinson, J. Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. *Microbiome* **2**(1), 39. https://doi.org/10.1186/2049-2618-2-39 (2014).
100. Chaffron, S. *et al.* Environmental vulnerability of the global ocean epipelagic plankton community interactome. *Sci. Adv.* **7**(35), eabg1921. https://doi.org/10.1126/sciadv.abg1921 (2021).
101. Gregory, A. C. *et al.* Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**(5), 1109-1123.e14. https://doi.org/10.1016/j.cell.2019.03.040 (2019).
102. Koonin, E. V. Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R. Soc. B Biol. Sci.* **371**(1701), 20150442. https://doi.org/10.1098/rstb.2015.0442 (2016).
103. Anesio, A. M. & Bellas, C. M. Are low temperature habitats hot spots of microbial evolution driven by viruses?. *Trends Microbiol.* **19**(2), 52–57. https://doi.org/10.1016/j.tim.2010.11.002 (2011).

## Author contributions

M.W.: Formal analysis, Writing—Original Draft, Writing—Review & Editing, Visualisation. A.V.: Formal analysis, Resources, Writing—Original Draft, Writing—Review & Editing. R.L.: Formal analysis, Resources, Writing—Review & Editing. E.P.: Formal analysis, Resources, Writing—Review & Editing. T.M.G.: Resources, Writing—Review & Editing, Funding acquisition.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-41204-3.

**Correspondence** and requests for materials should be addressed to M.W.

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

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