Linking extreme seasonality and gene expression in Arctic marine protists
Magdalena Wutkowska, Anna Vader, Ramiro Logares, Eric Pelletier, Tove Gabrielsen

To cite this version:
Magdalena Wutkowska, Anna Vader, Ramiro Logares, Eric Pelletier, Tove Gabrielsen. Linking extreme seasonality and gene expression in Arctic marine protists. Scientific Reports, 2023, 13, pp.14627. 10.1038/s41598-023-41204-3. cea-04321382

HAL Id: cea-04321382
https://cea.hal.science/cea-04321382
Submitted on 18 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Linking extreme seasonality and gene expression in Arctic marine protists

Magdalena Wutkowska1,2,3*, Anna Vader1, Ramiro Logares4, Eric Pelletier5,6 & Tove M. Gabrielsen1,7

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 evolution1. In the ocean, phytoplankton, mostly cyanobacteria and photosynthetic microbial eukaryotes, contribute half of the net primary production on Earth1. 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 phenology3. 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 zooplankton4–6.

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 intervals5–8. The world’s northernmost marine time series station (IsA) in Adventfjorden, Isfjorden, Svalbard (Fig. 1), has been continuously sampled since December 20119. This endeavour generated metabarcoding-based knowledge regarding which marine microbial eukaryotes are present and active throughout the year10,11. 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 IsA12, whereas diversity of operational taxonomic units (OTUs) is inversely proportional to this trend being higher during polar night10,12. However, the proportion of plastid-bearing to heterotrophic cells is lower during polar night (reviewed in13).

1Department of Arctic Biology, The University Centre in Svalbard, Longyearbyen, Norway. 2Department of Arctic and Marine Biology, UiT – The Arctic University of Norway, Tromsø, Norway. 3Institute of Soil Biology and Biogeochemistry, Biology Centre CAS, České Budějovice, Czechia. 4Institute of Marine Sciences (ICM), CSIC, Barcelona, Catalonia, Spain. 5Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France. 6CNRS Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, Paris, France. 7Department of Natural Sciences, University of Agder, Kristiansand, Norway. *email: magda.wutkowska@gmail.com
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\textsuperscript{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\textsuperscript{10,14}. Most of the primary production in the Arctic Ocean is performed by marine microbial eukaryotes when enough solar radiation is available\textsuperscript{15,16}. Outside this period these cells are assumed to use accumulated resources\textsuperscript{17}, decrease their metabolism\textsuperscript{18,19} or remain dormant\textsuperscript{20}. However, many species of microbial eukaryotes instead of passively surviving prolonged darkness might switch their feeding strategy\textsuperscript{21,22}, as is the case with mixotrophs\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{28}. Nevertheless, community-level gene expression obtained by ‘-omics’ methods was demonstrated as an effective predictor of current marine biogeochemical state\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{10} for complete profile). Photosynthetically active radiation (PAR), size-fractionated chlorophyll a and nutrient concentrations (nitrate/nitrite, phosphate, silicate), were obtained as described in\textsuperscript{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...
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.0.11.541. The pre-processing step aimed to remove unwanted sequences from the metatranscriptomes. First, Illumina adapters were removed using BBduk v. 37.36. 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 samples45. Thus, we used SortMeRNA 2.034 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 – Chrysalis, 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 Brujin graphs (Chrysalis step) and then resolving them (Butterfly step). Expression levels were estimated by mapping clean reads against the gene catalogue in RSEM 1.3.077. Due to varying numbers of reads in each of the metatranscriptomes (Supplementary Table S1) and to ensure between-sample comparison 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 3.1 with default parameters38. Similarities between the de novo assembled transcripts/predicted coding regions and proteins in the UniProt database46 were assessed using blastx/blastp with max_target_seq = 1 and e-value = 1e-3 (BLAST +)41. Protein domains were identified with HHMERS42 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 3.044, 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 settings48. 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 Perkinsina 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 previously77. Briefly, reads from each Tara Oceans’ metatranscriptomic read set were mapped onto transcript isoforms in our core dataset using bwa86 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 v3.5.231, and data were visualised using tidyverse v1.2.132. Diversity indices (richness, Shannon-Wiener index, and Pielou's evenness) were calculated using vegan package v2.5-483. Principal component analysis (PCA) of environmental factors was calculated on centred and scaled data with prcomp.
function \((\text{stats package v3.5.3})\) and visualised using \(\text{factoextra v1.0.5}\). To explore differences between transcript abundances a Bray–Curtis dissimilarity matrix \((\text{vegdist function in vegan package v2.5–4})\) was constructed and clustered using a 'complete' method within \(\text{kclus function (stats package v3.5.3)}\). \(\text{Pclust package was used to assign}}\) to the clustering topology\(^{54}\). To identify the strongest contribution of individual transcript isoforms to clustering patterns, we applied the \(\text{simper function on the transcript matrix}\).

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 \(\text{envfit function (vegan package)}\) was used to fit environmental parameters onto the GNMDS ordination. Analysis of similarities \((\text{ANOSIM; vegan package)}\) was used to test if there were differences between polar day and polar night associated with light. The \(\text{simper function (vegan package)}\) 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. \(\text{Simper analysis identified GO terms that differed between polar day and polar night. These features then considered “overrepresented” if the differences in means were statistically significant.}\)

**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 \((\text{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 \((< 1° \text{C}; \text{LW})\) was present in the first half of the year (December 2011 to May 2012) with warmer intermediate water \((> 1° \text{C}; \text{IW})\) 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 \((\text{nitrate/nitrite, phosphate, and silicates})\) were heavily depleted from the onset of spring bloom until the end of polar day \((\text{from May to August; Table 1)}\). However, silicates started to be depleted already in April (Table 1). Chlorophyll \(\text{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 \(\text{a was present in the small phytoplankton fraction (< 10 μm). Detailed descriptions of the IsA system, based on enhanced frequency of sampling can be found in 10,12,58,59.}\)}

**Seasonal transcript diversity.** The diversity and evenness of transcripts was higher during polar night \((n = 5)\) than during polar day \((n = 5; \text{Fig. 2})\). 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 \((\text{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

| Sampling date | Light parameters | Chlorophyll \(\text{a biomass} | Nutrient concentrations | Physical parameters |
|--------------|------------------|-----------------|----------------|-----------------|
|              | PAR \([\text{µmol m}^{-2} \text{s}^{-1}]\) | Declination \([°]\) | Day length \([\text{h}]\) | Total \(\text{chl a}\) \([\text{µg/l}]\) | \(\text{chl a} > 10 \mu\text{m}\) \([\text{µg/l}]\) | \(\text{NO}_2\) \([\mu\text{M}]\) | \(\text{NO}_3\) \([\mu\text{M}]\) | \(\text{PO}_4\) \([\mu\text{M}]\) | \(\text{Si(OH)}_4\) \([\mu\text{M}]\) | Water mass | Temp \([°\text{C}]\) | 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           | IW              | 0.3           | 34.43         |
| 2012-Jul-06  | 0.3               | 22.62           | 24             | NA              | NA              | 0.49          | 0.039         | 0.21           | 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            | LW              | 1.5           | 34.67         |

Table 1. Environmental parameters corresponding to each metatranscriptome sampled at 25 m depth. PAR \((\text{photosynthetically active radiation})\) 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 \(\text{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 10,12,58,59.}
2.7 times higher. However, the sample collected in January 2012 outliers 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, μPN = 1974, μPD = 1134 in polar day samples, p = 0.03), response to stress (μPN = 1482 in polar night, μPD = 498 in polar day, p = 0.01) and phototransduction (μPN = 936...
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-homocysteine 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 ($R^2_{GNMDS} = 0.795, p = 0.005$ 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 centrosomes. Centriols 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 anoxygenic 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
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 but could also be necessary in fungal pathogenesis.73
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 (μ↑60N = 64%, σ↑60N = 9% and μ↓60N = 21%, σ↓60N = 8%, respectively). This was also true for samples from the deep chlorophyll maximum depth (μ↑60N = 69%, σ↑60N = 12% and μ↓60N = 23%, σ↓60N = 9%, respectively) and mesopelagic depths 67% (σ↑60N = 6%).

Textomic 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 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. 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. 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\(^\text{10,12,76}\). 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., \(^\text{10,77}\)). The same patterns were shown for other Arctic marine microorganisms, such as bacteria and archaea\(^\text{78}\). 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., \(^\text{14}\)). 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 \(\text{IsA}\)\(^\text{10}\) 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\(^\text{10,14}\). However, despite the high diversity of OTUs and transcripts, cell counts, and therefore biomass of protists remained low throughout polar night\(^\text{12,79–81}\).

The active status of Dinophyta, the most abundant taxonomic group during polar night\(^\text{10}\), 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\(^\text{82–84}\). 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\(^\text{85}\), and other protists, such as Apicomplexa\(^\text{86}\).

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\(^\text{10,12}\). This includes crucial primary producers, such as Micromonas polaris, which were encountered as active at different depths of Arctic marine habitats during polar night\(^\text{14}\). 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\(^\text{13,30,87}\). 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 v1.2.132 in R v3.5.231.
It is not clear what is their function in microbial eukaryotes. However, recently, it was suggested that they are involved in G protein-coupled receptor-based signalling in Dinophyta. 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.

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. 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. On the other hand, it may also point towards cellular division and the need to produce new histones for new cells. 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. In general, biochemical reactions involved in methionine degradation are the main source of methyl groups used in gene expression mechanisms inside of the cells has been proposed. 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. On the other hand, it may also point towards cellular division and the need to produce new histones for new cells. 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.

The raw data generated for this study are deposited in ENA, project ID: PRJEB58729. Raw data from Tara Oceans are available at EBI and GenBank under project IDs PRJEB9738 and PRJEB9739.

Received: 27 February 2023; Accepted: 23 August 2023
Published online: 05 September 2023

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).
81. Weslawski, J. M., Kwasniewski, S. & Wiktor, J. Winter in a Svalbard fjord ecosystem. *Arctic* **44**(2), 115–123. https://doi.org/10.14330/arctic15277 (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/nature04618 (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/bi9109354 (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/10.1076/1434-4610-00090 (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/n27.11.2416 (1999).

87. Hoppe, C. J. M. ‘Always ready?’ Primary production of Arctic phytoplankton at the end of the polar night. *Limnol. Oceanogr. Lett.* p. lol10222, 2021. https://doi.org/10.1002/lol10222.

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/lio.11345 (2020).

91. Yotulisa, V. et al. Proton-pumping rhodopsins in marine diatoms. *Microbiology* https://doi.org/10.1016/S0962-8882(11)70153-3 (2011).

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/10.8080/RNA.2016.15476286.2016.1189071 (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.18.009952 (2019).

94. Parkhizko, 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* **16**(6), 1211. https://doi.org/10.1111/ace.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 aging: 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/s40168-015-0148-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/s40168-2014-1130 (2014).

100. Chaffron, S. et al. Environmental vulnerability of the global ocean epipelagic plankton community interactome. *Sci. Adv.* **7**(35), eabc1921. https://doi.org/10.1126/sciadv.eabc1921 (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).

Acknowledgements
The authors are thankful to all the people that contributed to the marine part of MicroFun project, collected and processed the samples. This work was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian Metacenter for High Performance Computing (NOTUR), and operated by the Department of Research Computing at USIT, the University of Oslo IT-department. http://www.hpc.uio.no/. We are thankful for help from Abel Cluster support team. We are also deeply thankful to the Tara Oceans Consortium who provided open and freely available data. We are thankful to Ragnheid Skogseth for providing a map of Adventfjorden. Finally, we would like to thank the two reviewers for their insightful comments that helped us to improve the manuscript.

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.

Funding
This research was funded by University Centre in Svalbard, as well as ConocoPhillips and Lundin Petroleum through The Northern Area Program and by the Norwegian Research Council (project number: 230970). The cost of sequencing was partly covered through Jan Christensens Legat. Open access funding was provided by UiT—The Arctic University of Norway, Tromso, Norway.

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.
