Changes in functional composition and gene expression in eukaryotic plankton at the Atlantic-Arctic Polar front

Paul Frémont[*1,2], Corinne Da Silva[1], Lucia Campese[3], Émilie Villar[1], Achal Rastogi[1,4], Jean-Marc Aury[1], Chris Bowler[5,2], Lee Karp Boss[6], Patrick Wincker[1,2], Eric Pelletier[1,2], Marion Gehlen[7], Daniele Iudicone[3], Olivier Jaillon[*1,2]

[*]Corresponding authors: Paul Frémont (pfremont@genoscope.cns.fr) and Olivier Jaillon (ojaillon@genoscope.cns.fr)

[1]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université d'Evry, Université Paris-Saclay, Evry, France. [2]Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans, Paris, France. [3]Stazione Zoologica Anton Dhorn, Naples, Italy. [4]Current address: Corteva Agriscience, The V Ascendas, Atria Block, 12th Floor, Madhapur, Hyderabad 500081, India. [5]Institut de Biologie de l'Ecole Normale Supe´rieure (IBENS), Ecole Normale Supérieure, Centre National De La Recherche Scientifique (CNRS), Institut National De La Sante´ Et De La Recherche Médicale (INSERM), Universite´ Paris Sciences et Lettres (PSL), Paris, France. [6]School of Marine Sciences, University of Maine, Orono, Massachusetts, USA. [7]LSCE-IPSL, CEA, CNRS, Université Paris-Saclay,Gif-sur-Yvette, France.

Abstract

Planktonic communities are shaped by ocean currents and participate in multiple global biogeochemical cycles. Unraveling how the functions of communities respond to strong environmental gradients will improve our understanding of the interactions between climate change and marine ecosystems. Here, we investigate changes in functions and gene expression of eukaryotic plankton transiting between the North Atlantic Ocean (NAO) in winter and the Arctic Ocean (AO) in spring/summer using metatranscriptomes and metagenomes from *Tara* Oceans. In Arctic communities, functions involved in maintaining the protein pool and translation machinery appear to be more active. Four major phylogenetically distant algal groups are abundant in both basins and show similar strategies at transcriptional level, including increased expression of some functions related to cold acclimation. These results shade lights on gene expression strategies shared by cosmopolitan phototrophs of widely separated lineages.

Introduction

Planktonic microorganisms thrive in all oceans of the Earth under diverse conditions ranging from highly oligotrophic, warm saline tropical gyres to highly seasonal, low-salinity and cold polar waters. They are key players in many biogeochemical processes such as primary production via photosynthesis (*1*), carbon export to the deep ocean (*2–4*) and essentially sustain the entire

oceanic trophic system (*5*). Many studies based either on satellite (*6, 7*) or genomic data (*4, 8, 9*) revealed a mainly latitudinal and temperature driven biogeographical partitioning of the oceans. While some species have strictly defined environmental niches (*4, 10*), some display cosmopolitan biogeographies, such as *Bathycoccus prasinos* which is present in both polar and temperate oceans (*11*) or *Pelagomonas calceolata* that is cosmopolitan in all except polar oceans (*12*).

Understanding how microorganisms acclimate and adapt (or not) to highly differentiated environments as they are passively transported by currents to new physico-chemical conditions is a critical question in the context of climate change. As an example, the transition between the North Atlantic and Arctic waters is made through the North Atlantic Current (NAC) that is the most important boundary current on Earth, with more than 40 Sv (40 million m^3/s) flow rate of water in the south and 20 Sv as it crosses the more eastward Mid-Atlantic Ridge (*13*). Major inflows of the Arctic Ocean, mainly a landlocked sea, are therefore the warm and saline Atlantic Waters (AW) that enter via the Barents Sea Opening and eastern Fram Strait. The warming of AW, the arctic sea-ice loss (*14, 15*), along with the shoaling of the intermediate-depth Atlantic Water layer and the increased winter ventilation in the eastern Eurasian Basin of the Arctic are making it structurally similar to the Western Eurasian Basin (*16, 17*). This suite of processes has been called "atlantification" of the Eurasian Basin and it is bringing the Arctic into a new climate state, characterized by a much greater role of Atlantic inflows (*18*). Documented range expansions attributed to warming include new appearances of temperate dinoflagellates and coccolithophores in high-latitude subpolar regions (*18, 19*), while other planktonic taxa are predicted to extend their ranges poleward under future climate conditions (*20, 21*).

This context raises the question of physiological acclimation to this new environment of incoming species and consequences on this ecosystem. While the physiological response of plankton to various environmental changes has been relatively well explored *in vitro* (*10, 22, 23*), studies are only just beginning to explore key metabolisms to explore photosynthesis, nitrogen metabolism (*24*), nutrient limitation (*25*) as well as community-level responses to iron availability (*26*) using *in situ* omics data. Genomic adaptation and acclimation to cold environments has also been investigated in the diatom *Fragilariopsis cylindrus* whose genome is notably enriched in metal-binding proteins, as well as ice-binding proteins and other stress related proteins associated with photosynthesis (*27*). In this diatom, many genes were also found to have bi-allelic differential expression in the Southern Ocean (*27*).

As the Arctic warms, atlantification might turn the Arctic into a CO_2 source instead of a sink, which would have major consequences on the global carbon budget (*28, 29*). For example in some

autotrophic plankton, when temperatures increase, the metabolism switches to heterotrophy and its rate increases (*30*) in agreement with the metabolic theory of ecology (MTE) (*31*). In the context of the atlantification, it appears critical to explore functional characterization of North Atlantic plankton communities entering the Arctic Ocean and to compare the functioning of both basins. Here we study metagenomic and metatranscriptomic data from eukaryotic plankton communities to characterize their differences in composition and function along their transport through the North Atlantic and into the Arctic Ocean basin. Significant changes in gene expression levels of certain functions, consistent across various lineages, emerge in light of the strong environmental clines along this route.

Results

Oceanographic context of the NAO/AO transition

The *Tara* schooner sampled the North Atlantic Ocean between January and March 2012 and the Arctic Ocean between May and September 2013 (*Tara* Oceans sampling sites 143 to 196). While the sampling route did not sample both basins the same year, there is a continuum in the seasons sampled. Both basins have a relatively high seasonality, yet recent studies (*4, 8, 9*) suggest that relatively stable plankton biogeographies exist at the global scale, which is confirmed here (see below). In terms of connectivity, values of minimum transport time by currents (Tmin) computed from previous studies (*8, 9*) between each proximal station (defined by this minimum transport time) confirm that the flow of currents is directed from the Atlantic into the Arctic (Figure 1). To enter the Arctic Ocean the boat first followed the western route via the Greenland Sea south of the Fram strait (*Tara* Oceans stations 158 to 163, Tmin=3.2 months). The boat then left the main current and bifurcated to the Barents Sea route (*Tara* Oceans stations 163 to 168, Tmin=7.5 months, stations 158 to 168 orange dashed arrow, Tmin=6.1 months). From station_168 onward, the boat sampled different Arctic regions under the influence of Atlantic inflows with the exceptions of stations 191 to 196 which are impacted by Pacific inflows (Figure 1a and Supplementary Figure 1b). The polar front is identified by a steep temperature gradient located at the time of sampling (June 2013) between station 158 and 163 (orange arrow Figure 1b). We further discuss environmental differences between the two oceanic basins and in light of current connectivity in Supplementary Discussion 1. Considering this strong environmental change along the NAC we first studied the genomic differentiation of the eukaryote communities of the two basins.

Genomic biogeography of the NAO/AO transition

We updated a collection of ~158 million eukaryotic unigenes (longest assembled gene sequences) constructed from metatranscriptomic data (*32*) that included the *Tara* Oceans Polar Circle stations. For this study, we selected 69.2 million unigenes from the 0.8-2000 µm size

fraction that is enriched in unicellular eukaryotes and small zooplankton. To explore biogeographical structure, we studied changes in plankton community composition by estimating 2 types of β diversity (βD) among stations in metagenomes and metatranscriptomes (see Methods, equations 4 and 5). These β diversity estimates were highly correlated (see Methods). A strong differentiation was found between the Arctic and the Atlantic although each basin appears to be relatively heterogeneous (Figure 2a). In the Arctic, this heterogeneity doesn't correspond to the different oceanographic regions (*33*). We also note that, despite being influenced by Pacific waters (*34*) stations 193 to 196 do not cluster together (Figure 2a). From the β-diversity matrix clustering (Figure 2a), we classify stations 143-155 as NAO and 158-196 as AO. We can then define and analyze unigenes that are either specific to a single basin (NAO or AO) or common to both basins (Methods). The taxonomic composition of these specific and common unigenes, according to their relative abundances (Methods) is presented (Figure 2c and 2d) and discussed in the Supplementary Discussion 2 (Figure 2 and Supplementary Figure 5). In total, 10% (n=6.9 million) of expressed unigenes were common to both basins, whereas 55% (n=38.2 million) are unique to the North Atlantic and 35% (n=24.1 million) are unique to the Arctic. The relative abundances of common unigenes reveal a clear transition between the two basins corresponding to the polar front (Figure 2d). Stations 155, 158 (NAO side of the front) and 163 (AO side of the front) are found enriched with common unigenes (51.2% of the read abundance on average and 42.3 % of the unigenes in number, Figure 2d, Supplementary Figures 3-4). In the Arctic, the common unigenes represent a larger proportion of both reads and number of unigenes (Figure 2d and Supplementary Fig 3-4). This higher relative proportion of genes found only in the Atlantic compared to the Arctic is likely due to a selective environmental bottleneck at the polar front and lower species diversity in the Arctic (Supplementary Figure 4). Overall, these results confirm the mixing of different communities from different water masses, as is well known in oceanic fronts (35, 36). They also illustrate the existence of a cline of genomic composition of communities at the polar front. This implies the existence of species that successfully cross the environmental front (studied below).

Linking genomic biogeography with physical connectivity

To assess the role of ocean circulation on this biogeography, we then compared the minimum transport times (Tmin) separating sample sites with the corresponding β-diversity estimates (Figure 2b, Supplementary Figure 1a). In contrast to the results of Richter et al. (8), a maximum correlation is found between 4 and 8 years of transport time instead of 1.5 years. These are typical times that separate Atlantic and Arctic stations (light blue dots Figure 2b). The strong correlation with beta diversity indicates that the influence of currents on community composition previously

137 identified in the North Atlantic (8) extends far north beyond the Arctic Circle. However, this
138 maximum correlation is probably also due to the strong differentiation of the Atlantic and polar
139 communities.
140 We also found that the calculated environmental distance (ED) between sampling sites shows
141 the same maximum correlation of around 4-8 years with transport time (Supplementary
142 Discussion 1 and Supplementary Figure 1c) but also with β-diversity (βD) (Supplementary Figure
143 2b-c). These data indicate that β-diversity and currents are most strongly correlated (maximum
144 correlation at 0.75 vs. 0.67 for Tmin vs. ED and 0.5 for ED vs. βD), in contrast to Richter et al. (*8*)
145 in which environmental distance is better correlated with β-diversity. These results highlight the
146 strong and not readily decomposable interactions between the physicochemical environment,
147 currents, and planktonic communities that characterize the seascape (*37*).
148 **Functional divergences between the eukaryotic communities of NAO and AO**
149 The major differences in the composition of the Atlantic and polar eukaryotic communities suggest
150 that they function differently. To investigate this question we used Gene Ontology (GO) functional
151 annotations of unigenes. Our unigene set contains 4,695 different Gene Ontology (GO) terms
152 (*35*). Statistical analysis reveals that 66 terms are significantly more abundant at the community
153 level in the Arctic and 11 in the Atlantic (examples in Figure 3 and comprehensive list in
154 Supplementary Table 1, t-test, Hommel correction $p<0.05$). Out of these 66 GO terms in Arctic
155 communities, 20 are related to RNA processing and 11 are related to RNA methylation (6 to tRNA
156 methylation) (Figure 3a, Supplementary Table 1). RNA methylation has been reported to be
157 important in cold stressed plants (*36*). As an example among these 11 GO terms, 'tRNA (adenine-
158 N1-)-methyltransferase activity' is known to be important for tRNA stability (*37*) and has also been
159 reported to be involved in mRNA methylation and transcriptome protection in granules that are
160 formed within the cytoplasm of stressed cells (*38*). Finally, 'tRNA methylation' is also well
161 documented as being important in translational fidelity (*39*). In particular, stressful conditions have
162 been reported to change the balance between free tRNA and aa-tRNA (*40*) which is critical for the
163 accuracy of protein synthesis (*41*).
164 Many other significantly more abundant GO terms in the Arctic are related to regulation and
165 downstream processes of gene expression, such as regulation of gene expression ('snoRNA
166 binding' (Figure 3b), 'ncRNA processing', 'mitochondrial RNA metabolic process') and ribosome
167 biogenesis ('ribosome biogenesis' (Figure 3c), 'preribosome, small subunit precursor', and
168 'ribonucleoprotein complex biogenesis'). ncRNA, snoRNA (a type of ncRNA) and mitochondrial
169 RNA metabolism are known to play important roles in regulation of gene expression (*42, 43*).

170 snoRNAs are involved in chemical modification of RNAs, including by methylation, while ncRNA
171 (small RNAs) have been shown to play a critical role in cold acclimation in *Arabidopsis* (*44*).
172 Ribosomes are known to be cold and heat shock sensors (*45*) and their dynamics and stability
173 are altered in cold conditions (*46*). Functions associated with protein post-translational
174 modifications were also found to be significantly more abundant in the Arctic ('protein peptidyl-
175 prolyl isomerization', 'peptidyl-proline modification', 'serine-type exopeptidase activity' and
176 'serine-type carboxypeptidase activity' (Figure 3d), 'intracellular protein transmembrane
177 transport').
178 Taken together, these abundant Arctic functions raise the hypothesis that an important set of
179 reversible epigenetic processes at the transcriptional and translational level (RNAs), as well as
180 other post translational modifications are more active in the Arctic. At the community level, these
181 gene expression regulatory processes appear to be important to thrive in the Arctic during spring
182 and summer. In the Arctic, we find other abundant functions that are known to be important in
183 cold environments, such as maintenance of chromosome integrity as well as many functions
184 related to lipid catabolism (Figure 3e), membrane fluidity, and organization of organelle
185 membranes (Supplementary Discussion 3). In addition to these more abundant functions, our
186 results suggest a more rapid protein turnover in spring/summer Arctic plankton communities
187 (Supplementary Discussion 3).
188 While the functions we describe appear as being significantly more abundant in the Arctic when
189 we conduct the analysis at the community scale (all taxa together), this was not the case at the
190 individual taxon level (with the exception of some of the differentially more abundant taxa in the
191 Arctic) (Supplementary Figure 6). Furthermore, taxa contribute very unevenly to the abundance
192 of these functions (Figure 3). For example, the Ciliophora appear to be important contributors to
193 'lipid catabolic processes' (Figure 3) whereas the Dinophyceae are important contributors of the
194 GO term 'RNA methylation'. These differences are probably due to differences in the gene
195 repertoire of the different taxa. This community effect could be due to higher statistical power
196 when all taxa are grouped together. In any case, it may correspond to a similar transcriptional
197 response that occurs in many groups and becomes significant at the community level and to
198 expression regulation.
199 We then examined expression changes at a community level, which would be expected to be
200 shared by many taxa. In this case, a higher ratio of metatranscriptomic to metagenomic
201 abundance should be observed in the Arctic. We found most of these functions more expressed
202 at the community level, however this signal seems to come mainly from a few dominant groups
203 (Mamiellales, Dinophyceae, Haptophyta, Opisthokonta and unclassified Eukaryotes)

204 (Supplementary Fig. 6). On the other hand we note the absence of Bacillariophyta (diatoms) in
205 these groups. These functions, which are particularly important in the Arctic environment, are
206 overexpressed in a large number of taxa that therefore likely share the same strategy. In addition,
207 many of the associated GO terms are functionally consistent with each other (see Methods,
208 Supplementary Figures 6 and 7). Without excluding a seasonality effect and despite some
209 variations within the Arctic (Figure 3 d,b,e,f for instance), these functions appear more abundant
210 over the whole spatio-temporal sampling of the Arctic basin (from May to September, see
211 distribution plots of Figure 3).
212 Given the spatiotemporal continuity of the sampled sites, our results therefore indicate that we
213 can detect real functional differences between eukaryotic communities before and after the polar
214 front. We also detected 11 more abundant GO terms in Arctic metagenomes that would reflect an
215 enrichment of these functions in the genetic repertoire of Arctic species. Eight of these are also
216 functions related to gene expression (Supplementary Table 1). Only two are related to lipid
217 metabolism ("Sphingolipid metabolic process" and "Ceramide metabolic process") and are part of
218 the set of more abundant functions in the Arctic metatranscriptomes.
219 To complete the GO term analysis, we used the ALDEx2 package (*47, 48*) to compare PFAM
220 abundances between the AO and NAO (Supplementary Discussion 4 and 5). Interestingly, almost
221 all PFAMs associated with cold stress and cold acclimation were more abundant (but not
222 necessarily more highly expressed) in the Arctic (Supplementary Fig. 8 and Supplementary
223 Discussion 4).
224 Conversely, we also searched for GO terms that were more abundant in the Atlantic. Considering
225 all taxa, 11 GO terms were significantly more abundant in the metatranscriptomes and 85 in the
226 metagenomes (3 in common with all 11) (Supplementary Table 1 for the exhaustive list,
227 Supplementary Discussion 5, Supplementary Fig. 16). Overall, there relate to functions in central
228 carbon metabolism and lipid biosynthesis (contrasting with lipid catabolism in the Arctic). We also
229 found specific functions of photosynthesis (cyclic electron transfer chain and light harvesting
230 system) that contrast with the more abundant central functions of photosynthesis (PFAMs) in the
231 Arctic (Supplementary Discussion 5).
232 As discussed above, AO and NAO are strongly connected, and our data indeed confirm the
233 presence of many cosmopolitan taxa. The functional differences that emerged in our
234 metatranscriptomic data may correspond in these species to altered gene expression of some
235 key functions.
236 **Transcriptional changes in major eukaryotic plankton groups in the NAO/AO transition**

237 We then investigated gene expression changes in plankton groups that were found in abundance
238 in both basins. We hypothesize that in these groups the expression of certain genes may be
239 modified along the environmental gradient around the polar front. We specifically searched for
240 gene abundance patterns significantly correlated with different environmental parameters and
241 analyzed their functional and taxonomic annotation. We selected a set of ~485,000 unigenes that
242 were present in both metatranscriptomes and metagenomes from at least 5 stations covering both
243 basins. We determined an expression value for each unigene at each station from a
244 transformation of the metatranscriptomic abundance values and normalization by metagenomics
245 (see Methods and Supplementary Figs. 8-9 for reasoning).
246 We found ~88,500 instances of significant correlations, positive or negative, between gene
247 expression profiles and physical parameters (temperature, sunlight duration (SSD), and salinity).
248 These physical environmental parameters were highly correlated with each other, however
249 temperature is the parameter significantly correlated with the largest number of gene expression
250 profiles (Supplementary Figure 10, ~19,500 negative and ~12,700 positive, FDR=0.05,
251 Supplementary Figures 11-12). Among the unigenes negatively correlated with temperature, we
252 found a particularly large number annotated as Mamiellales, diatoms and Pelagophyceae (Figure
253 4a, Supplementary Figure 13). Therefore, we further analyzed the taxonomic and functional
254 distribution of these unigenes correlated negatively or positively with temperature by a clustering
255 and dimension reduction approach. For all algal taxa tested (Bacillariophyta, Mamiellales,
256 Phaeocystales and Pelagophyceae), all genes whose expression level was negatively correlated
257 with temperature overlapped strongly (Figure 4b) and clustered together (Figure 4c), apart from
258 copepods. These results suggest that the temperature cline encountered by plankton in this ocean
259 system leads to a transcriptomic response of a core set of functions in a similar way in eukaryotic
260 micro- and picophytoplankton, independently of their phylogeny. Nevertheless, beyond this set of
261 core functions, other functions appeared to follow the same response only in certain lineages
262 (Figure 4b). These were numerous for the Mamiellales, Bacillariophyta and Pelagophyceae, but
263 few in the Phaeocystales. We note that Mamiellales and diatoms have the most similar repertoire
264 of cold-responsive genes, in agreement with the functional convergence between these groups
265 recently proposed (*49*).
266 To better characterize this transcriptional response, we analyzed whether enrichments could be
267 detected in certain functions. To do so, we applied a rationale of metagenomic co-variation of
268 metatranscriptomic data to reconstitute *in silico* more or less partial gene repertoires of a species
269 according to a procedure we had already used (Methods). We thus obtained a new set of 1,000
270 Metagenomic-based transcriptomes, or MGTs. Of these, six algal MGTs (probably corresponding

271 to 5 species) with a sufficient number of unigenes significantly negatively correlated with
272 temperature contained a significant number of enriched functions (see Methods). Two of these
273 MGTs are annotated as *Bathycoccus* (Mamilellales), one as *Micromonas* (Mamiellales), one as
274 *Phaeocystis*, one as *Aureococcus* (Pelagophyceae), and one as *Minutocellus* (a pennate diatom).
275 In these MGTs, we found enriched functions known to be directly related to cold acclimation, light
276 or osmotic stress, as well as a large number of functions related to the regulation of gene
277 expression (Figure 5). In addition, the enriched functions contain processes related to energy
278 consumption, DNA division/replication, motility, nucleotide biosynthesis, protein folding and DNA
279 repair (Figure 5). We discuss these functions in more detail in Supplementary Discussion 6.
280 Overall, our results suggest that some eukaryotic lineages of unicellular planktonic algae have
281 common mechanisms of transcriptional changes when moving from the Atlantic to the Arctic basin
282 during the seasons we studied. However, lineage-specific properties emerge, particularly with
283 respect to cold and stress response functions that are more diverse. In addition, regulation and
284 other processes related to gene expression, from signal transduction to post-translational
285 modification, are found to be highly active at the community and species levels, highlighting their
286 likely critical roles in the AO and NAO/AO transition.

287 **Discussion**

288 Using meta-omic approaches, we have characterized a differentiation of eukaryotic plankton into
289 two types of communities that occurs during its transport by currents from the North Atlantic in
290 winter to the Arctic Ocean in spring and summer. These comparisons show that the two
291 communities are functionally and compositionally divergent but share certain lineages that persist
292 despite the strong environmental front they encounter. We compared the rates of change in
293 community composition with that of gene expression in polar and non-polar samples as was done
294 on the prokaryotic component of the same plankton samples (*24*). Similar to what has been
295 reported for prokaryotes (*24*), we observe a very clear modification around 15°C with a greater
296 contribution of community turnover in polar samples compared to temperate samples
297 (Supplementary Fig. 19). However, community turnover is still significantly lower than gene
298 expression changes, unlike in prokaryotes (*24*). This could suggest that gene expression changes
299 are higher in eukaryotes, but this may be a consequence of the large proportion of non-coding
300 regions in some eukaryotic genomes resulting in less sampling of eukaryotic communities at the
301 genomic than transcriptomic level (many zeros). In our analysis, the functional divergences are
302 explained both by changes in gene abundance, and thus taxa, but also by the level of gene
303 expression. The functions that differentiate the two types of communities are related to gene
304 expression (from signal transduction to post-translational modification), lipid metabolism

305 (catabolism and membrane organization), protein trafficking, and telomere maintenance. Life in
306 the polar environment appears to require tight regulation of post-transcriptional mechanisms,
307 including RNA methylation, which is essential for RNA stability (*37, 50*) and protein translation
308 fidelity (*40, 41, 44*). For those species that successfully cross the polar front, which we refer to as
309 persistent species, the genes for which transcriptional changes are observed carry functions also
310 related to gene expression (from signal transduction to post-translational modification) as well as
311 cold/stress responses, energy consumption processes, protein folding/ubiquitination, cell
312 division/DNA replication, nucleotide biosynthesis, and DNA repair. To explain gene expression
313 variations, we tested an important set of environmental parameters including physical factors and
314 nutrient concentrations. However, nutrient concentration values tested were not sufficiently
315 homogeneous and representative. Among the remaining physical factors we tested (temperature,
316 day length and salinity), temperature was the most important factor in explaining variations in
317 gene expression.
318 A similar transcriptional response is described across the major algal lineages; Mamiellales,
319 Bacillariophyta, Dinophyceae and Phaeocystales all show a transcriptional response to cold that
320 is similar among them but clearly different from that of copepods. This algal transcriptional
321 response is characterized by an increase in all of these groups in the expression of certain
322 functions and by an increase in the expression of other functions only in specific groups. Functions
323 known to be directly related to cold stress acclimation in many land plants are enriched in these
324 data. This suggests that these cold acclimation mechanisms are shared across photosynthetic
325 lineages. It is difficult at this stage to infer the nature of the evolutionary events from which these
326 properties are derived and which are now shared by phylogenetically distant lineages. It is likely
327 to be a combination of convergence and ancient adaptations.
328 Overall our study suggests that a strong transcriptional response involving key universal and life-
329 necessary but also more specific functions such as stress and cold responsive genes is critical
330 for plankton to enter and keep on living in the Arctic environment (in spring and summer here).
331 Importantly the temperature gradient we investigate is concentrated in the few stations of the
332 transition and no more than two years of transport time (Figure 1a) showing that persisting
333 plankton have the time to respond. Indeed acclimation responses are believed to be quite fast, in
334 order of days to week (*27, 51*). While the transcriptional response we report seem critical to adapt
335 it might not be sufficient to explain the survival of persisting algae. Indeed organisms might also
336 need to adapt by directly changing the sequence of their proteins as recently suggested in
337 *Bathycoccus* (*11*) and in the SAR supergroup (*52*) or by mechanisms such as allele specific
338 expression recently reported in the copepod *Oithona similis* (*53*) and in the diatom *Fragilariopsis*

339 *cylindrus* (*27*). Therefore, the signal we observe is probably due to a mix of acclimation and
340 adaptation responses that we can't disentangle here. Critically, adaptation might require more
341 time (*54*) than acclimation although adaptation can occur relatively fast in large populations (*55,*
342 *56*). Investigating both transcriptional and sequence changes at a broader scale and for more
343 organisms seem now critical to better understand how plankton can thrive in so many climates,
344 what differentiate specialists from more ubiquitous species, changes from one to another (*55*) and
345 their response to a changing climate as well as consequences on ecosystems (*28*). Particularly,
346 in the context of atlantification it might be critical to assess which species have the potential to
347 adapt and acclimate to the new environment that is being formed. The strong connectivity allowed
348 by the Gulf Stream followed by the North Atlantic Current (*13*) is believed to play a critical role in
349 these processes (*18*) in the context of global warming (*57, 58*). Nevertheless, this circulation might
350 collapse with global warming (*58–60*).
351

352 ## Materials and Methods

353 **Marine Atlas of *Tara* Oceans Unigenes (MATOU-v1.2)**

354 We updated the original version of the Marine Atlas of *Tara* Oceans Unigenes (*32*) by including
355 samples encompassing the Arctic seas. This extension includes a total of 581 samples (+140)
356 from 6 different size fractions divided between 0.8 and 2000 µm. In total all samples are collected
357 at 88 locations across the oceans (previously 68) at three depths (Surface (SRF), Deep
358 Chlorophyll Maximum (DCM) and Mesopelagic zones (MES)). We followed the same
359 methodology as previously described for the sequencing, assembly, mapping, taxonomic and
360 functional annotation of the updated atlas (*32*). The new catalog totals ~158.5 million unigenes
361 increasing by ~41.7 million unigenes the previous version (+35.7%).

362 **Metagenomics-based Transcriptomes (MGT-v1.2)**

363 We generated an updated version of the Metagenomics-based Transcriptomes (MGTs) that
364 cluster unigenes based on their relative abundance covariations across *Tara* Oceans samples
365 into relevant transcriptomic and more or less taxonomically coherent units. As for the MATOU-
366 v1.2 catalog, *Tara* Oceans Polar Circle samples were added to rebuild the collection from scratch.
367 Clustering and taxonomic assignment of MGTs were computed using the same methodology as
368 Vorobev et al. (*61*). Briefly, the metagenomic RPKM-based (reads per kilo base per million of
369 mapped reads) abundance matrix of sufficiently occurring unigenes was submitted to the canopy
370 clustering algorithm described in Nielsen et al. (*62*) . This is a density based clustering that does
371 not take into account the sequence composition, as opposed to most binning tools. We used a

max Pearson's correlation difference of 0.1 to define clusters and then clusters were merged if canopy centroids' distances were smaller than 0.05 (250k iterations, default parameters). This updated version extends the collection to a total of 1000 MGTs counting at least 1000 unigenes each (previously 924 MGT with min. 500 unigenes, +8.2%) encompassing 9,547,691 unigenes (previously 6,946,068 +37.5%).

Environmental parameters and distance

We use a set of 11 environmental parameters from three different sources (*In situ*, World Ocean Atlas 2018, PISCES-v2 model and the U.S. Naval Observatory), all believed to be relevant ecologically. We use three *physical* variables: Temperature and Salinity (*in situ Tara* measurements (*63*)) and Sea Sunshine Duration (SSD, Calculated from Astronomical Applications of the U.S. Naval Observatory). Two macronutrients are from *in situ Tara* Oceans measurements: dissolved Silica ($Si0_2$), and phosphate (PO_4^{3-}) (*64*). Nitrate (NO_3^-) is extracted from the World Ocean Atlas 2018 (*65*). From these we derive N* and Si* calculated as follows:

$$N^* = [NO_3^-] - 16 * [PO_4^{3-}] \qquad (1)$$

$$Si^* = [SiO_2] - [NO_3^-] \qquad (2)$$

N* reflect the excess of Nitrate compared to phosphate and inversely the excess of phosphate (expressed in the order or magnitudes of nitrate concentration according to a constant ratio of 1/16 (*66–68*)). This metric was shown to be importantly correlated with the habitat of diazotrophic organisms (*68*). The second Si* was shown to reflect competition between diatoms and other organisms. We use one micronutrient (Iron) simulated from the PISCES-v2 model (*69, 70*). Finally two seasonality indexes (temperature and nitrate) are calculated from the World Ocean Atlas 2018, reflecting seasonal maximal seasonal fluctuation over the years at the sampling locations:

$$SI(X(x,y)) = \frac{\max_t(X(x,y,t)) - \min_t(X(x,y,t))}{\max_S(\max_t(X(x,y,t)) - \min_t(X(x,y,t)))} \qquad (3)$$

X is either the nitrate concentration or the temperature. The indices t and S respectively refer to time (months here) and spatial (*Tara* samples here) coordinates. All variables are extracted at the month and sampling location of the *Tara* stations considered here.

An environmental distance is calculated as the Euclidean distance (*71*) of standardized parameters (z-values (*72*)):

$$ED(i,j,P) = \sqrt{\sum_{p \in P} \left(x_{ip} - x_{jp}\right)^2}$$

401 $ED(i,j,P)$ is the environmental distance between station i and j for a set of parameters P (here
402 we used the whole set).

403 **Calculating -omic community dissimilarity**

404 Following Richter et al. (*8*), metagenomic and metatranscriptomic community distance between
405 pairs of samples was estimated using shotgun metagenomes and metatranscriptomes for size
406 fraction 0.8-2000 µm. We used an improved -omic comparison method (SimkaMin (*73*)) that
407 computes standard ecological distances by replacing species counts by counts of DNA sequence
408 k-mers (segments of length k). SimkaMin is less computationally intensive than the previous
409 Simka thanks to an efficient data subsampling scheme used to estimate Bray-Curtis and Jaccard
410 dissimilarities (*73*). $\beta\ diversity$ for -omic reads is defined with the following equation within Simka:

411 $$\beta\ diversity = \frac{b+c}{2a+b+c} \qquad (4)$$

412 Where a is the number of distinct k-mers shared between two samples, and b and c are the
413 number of distinct k-mers specific to each sample. We represented the distance between each
414 pair of samples on a heatmap using the heatmap.2 function of the R-package gplots_2.17.0 (*74*).
415 We also calculate β-diversity between station i and j based on MATOU-v2 unigenes abundances
416 as follow:

417 $$\beta\ diversity_{ij} = \frac{\sum_k \left| RA_k^i - RA_k^j \right|}{\sum_k \left| RA_k^i + RA_k^j \right|} \qquad (5)$$

418 Where RA_k^i is the relative abundance of unigenes k in station i. We find high correlation between
419 unigenes based distance and simkaMin based distance (r^2=0.96 for metatranscriptomes-based
420 distances) as well as between metatranscriptomes and metagenomes distances (r^2=0.99 and
421 0.97 respectively for simkaMin and unigenes Bray-Curtis).

422 **Taxonomic description of the metatranscriptomes and metagenomes**

423 Each sample of our dataset was described taxonomically by taxonomical annotation of unigenes.
424 We harmonized unigene taxonomic annotation by selecting a set of ecologically relevant and
425 sufficiently abundant groups: Hexanauplia, Bacillariophyta, Pelagophyceae, Crysophyceae,
426 Coccosphaerales, Phaeocystales, Ciliophora, Tunicata, Mamiellales, Dinophyceae, Bacteria,
427 Streptophyta, Cnidaria, Rhizaria, Insecta, Cryptophyta, Euglenozoa, Amoebozoa, Craniata,
428 Fungi, Viruses, Archaea, Eukaryota (unclassified), root, and respectively other Haptophyta,
429 Opisthokonta, Stramenopiles, Alveolata and Viridiplantae. Each unigene was then assigned a
430 geographical annotation: either unique to the Atlantic basin, unique to the Arctic basin, or shared
431 between the two (when a unigene was found in at least one station of the two basins). To find
432 taxonomical groups/clades that were significantly relatively more abundant in one of the two

basins, we applied a Wilcoxon test on the mean of the logarithm of the relative abundance of each taxon across the station of the two basins and applied a Hommel correction (*75*) on the p-values for multiple hypothesis correction testing. Note that for this analysis we used the logarithms of abundances and not the centered log ratio (used for the unigene expression analysis; see below) as the former might bias some abundances as it is centered on zero in the case of skewed distributions (which is the case here).

Functional characterization: comparing the Arctic and Atlantic basin

For each unigene of the MATOU-v1.2 catalog we converted the PFAM (*76*) (protein family domain annotation) to the Gene Ontology annotation (*35, 77*). We used the full GO dataset including the three main categories of the catalog: molecular function, cellular component and biological processes. Unigenes can have several PFAMs or PFAMs can map on several GOs. To take this into account we duplicated the functional annotation and therefore each unigene is assigned to each level of its Gene Ontology hierarchy and can have several annotations. For each GO term found across our dataset (a total of 4,695 terms from a total of 44,000 in the database), we defined its relative abundance within each sample of our data by summing the relative abundance of all the unigenes annotated as such. We applied a Student test (*78*) on the means of the logarithms of the relative abundances of each GO term between the two basin means, applied Hommel multiple comparison (*75*) correction, and found significantly relatively more abundant functions (p-val<0.05 after correction). Here we used only the logarithm of the relative abundance and not the centered-log-ratio as the multiplicity and unevenness of levels in the GO hierarchy did not allow us to define an average logarithm consistent across the hierarchy.

We also used ALDEx2 R package (*47, 48*) to compute significantly more abundant PFAMs between North Atlantic and Arctic metagenomes. ALDEx2 uses standard Bayesian techniques (*79*) to infer the posterior distribution of $[p_1, p_2, \dots, p_i, \dots, p_n]$ as the product of the multinomial likelihood (where pi i the proportions of gene i (here we use PFAMs) in a sample) with a $Dirichlet(\frac{1}{2}, \frac{1}{2}, \dots)$ prior. ALDEx2 then uses the full Dirichlet posterior probabilities distribution to draw multiple estimates (Monte-Carlo realizations) of each p_i such that $[p_1, p_2, \dots, p_i, \dots, p_n] \sim Dirichlet([n_1, n_2, \dots, n_i, \dots, n_n] + \frac{1}{2})$. Drawing p_i from the full posterior allows accounting for the large variance and extreme non-normality of the marginal distributions of p_i when the associated n_i are small. p_i are then transformed using the centered-log-ratio (see next paragraph for this transformation). ALDEx2 therefore models the abundance of each gene (PFAMs for us) as a random-effect ANOVA model accounting for the fact that within-condition abundance/expression is a non-negligible random variable:

$$q_{ijk} = \mu_{ij} + \nu_{ijk} + \tau_{ijk} + \varepsilon_{ijk} \qquad (6)$$

Where q_{ijk} is the adjusted log-abundance/expression, μ_{ij} is the expected expression of gene i within each condition j, ν_{ijk} is the sample specific expression change for replicate k, τ_{ijk} is the sampling variation from inferring abundance from read counts and ε_{ijk} the remaining nonspecific error, independent and identically distributed within k. To find significantly more abundant PFAMs, ALDEx2 then performs a Welch t-test on each q_{ijk} estimates (drawn from the Monte Carlo realization) and corrected for multiple comparison by the Benjamini-Hochberg method (*80*)(*80*)7, we plotted the median of the following metrics to visualize between condition differences as well as within condition variability:

1. The absolute fold difference between-conditions

$$\Delta_A(i,j,j') = W(i,j) - W(i',j') \qquad (7)$$

Where $W(i,j) = \sum_{k=1}^{n} q_{ijk}$ is the within condition mixture. Index i indicates the PFAMs index, index j indicates the condition index (Atlantic or Arctic) and finally k indexes the replicate of a given condition.

2. The between sample, within condition difference

$$\Delta_W(i,j) = \max_{k!=k'} \left| q_{ijk} - q_{ijk'} \right| \qquad (8)$$

Unigene expression analysis and response to environmental parameters

We defined expression of each of unigene as the difference between the centered-log-ratio of the relative abundance of the unigene from metatranscriptomic reads and the centered-log-ratio of its relative abundance from metagenomic reads:

$$E = clr(MetaT) - clr(MetaG) \qquad (9)$$

The clr was used as it was shown to be stable to subsampling of the dataset (*81*). It is defined as follows:

$$q_i = \log(p_i) - \frac{1}{m}\sum_{k=1}^{m} \log(p_k) \qquad (10)$$

Where $p_i$is the relative abundance of unigene i and we have $\sum_{k=1}^{m} q_k = 0$. Only unigenes which were found in at least five stations and present in both basins (at least one expression value available in the two basins) were considered, totaling ~485,000 unigenes. This considerably decreased the number of unigenes on which we could do the analysis but we believe this is the most robust way to analyze unigene expression within metaomic samples by taking into account both the metatranscriptomic and metagenomic abundances and having sufficient statistical power (at least 5 stations). To validate this approach, we computed for each unigene the Erb rho

correlation coefficient (*82*) between $clr(MetaT)$ and $(MetaG)$. Importantly, they correlated well with each other although more or less strongly depending on the taxa, justifying the validity of this approach (Supplementary Fig. 8-9).

For each unigene expression profile we computed two Pearson correlation coefficients with each environmental parameter: one correlating the 'real' data and one correlating 'randomized' data with the given environmental parameter. The distribution of the values of the second test was used as a null model. To find unigenes whose expression was significantly correlated to a variable, we compared the distribution of correlation between 'real' data and 'shuffled' data. We set a false discovery rate (FDR (*83, 84*)) of 0.05, *i.e.,* a unigene was considered significantly differently expressed if its correlation was superior to the maximum correlation of the first 2.5% (at each tail of the distribution) of the ordered correlations of the 'randomized' dataset. Then, to have a more stringent view of the functional landscape of significantly correlated or significantly differentially expressed (basin comparison), we performed an enrichment statistical test (hypergeometric test (*85*), Hochberg correction, p-value=0.05). For this test, we used only the deepest GO annotation of each unigene found within significantly correlated unigenes in comparison to the functional GO annotations of all the unigenes from the given taxonomical group or MGT. All these steps were performed for each taxonomical group and for the set of MGTs.

Functional clustering of the GO catalog using the Levenshtein distance

Extensive efforts have been made to cluster the Gene Ontology catalog using multiple Semantic similarity Measures Approaches (*86–88*). We propose here a classification based on Levenshtein distance (*89*). To do so, each GO term found within our dataset (4695 GOs) was represented by a huge integer describing its position and ontology in the hierarchy of the Gene Ontology atlas. Let *go* be the GO with the following hierarchy annotation: GO:000XXX;GO00XXXX;GO0000XX;…;GO000XX (in the order of the hierarchy) the descriptive number of the GO will be 000XXX00XXX0000XXX…000XX. Then we computed the Levenshtein distance between two GOs *go1* and *go2* as follows:

$$lev(go1, go2) = \begin{cases} |go1| & if\ |go1| = 0 \\ |go2| & if\ |go2| = 0 \\ lev\big(tail(go1, go2)\big) & if\ go1[0] = go2[0] \\ 1 + min \begin{cases} lev(tail(go1), go2) \\ lev\big(go1, tail(go2)\big) \\ lev\big(tail(go1), tail(go2)\big) \end{cases} & otherwise \end{cases} \quad (11)$$

We then used UPGMA (*90*)hierarchical clustering to cluster the distance matrix using correlation as distance. To optimize the hierarchical clustering, we found a seed that minimized the Silhouette

criteria for the set of clusters retrieved. After clustering, groups were defined as the largest clusters with the bootstrap probability superior to 0.95. Clustering with only 1 bootstrap generally performs better than doing 1,000 bootstraps to maximize the Silhouette criteria. After finding the optimal seed (the one with the highest Silhouette), we added a cleaning step of the clusters by allocating wrongly clustered GOs (negative silhouette) to clusters containing unique GOs. This second step decreases slightly the average silhouette coefficient although it also decreases the number of misclassified GOs (relatively to the Levenshtein distance). Therefore we privileged accuracy over average efficiency of the clustering.

Community composition and expression based distances

We used estimated centered log ratios from the ALDEx2 package (*47, 48*) based on the PFAM abundances to compute pairwise community composition distances and expression distances between each pair of stations. Expression is defined as the difference between centered log ratios of metatranscriptomic abundances and metagenomic abundances (Equation 9). Then the distance between each pair of station is defined as the euclidean distance over all PFAMs :

$$D_{ij} = \sqrt{\sum_{PFAMs} (x_i^{PF} - x_j^{PF})^2}$$

Where x_i^{PF} is either the abundance (centered log ratio) or the expression value of PFAM PF in station_i. Following Salazar et al. (*24*), we then computed, for temperature bins each containing 5 stations, the median ratio between abundance based distances and expression based distances of all pairwise distances from stations found in each bin. A Wilcoxon test and holm correction for multiple comparison ($p<0.05$) to test whether this ratio is significantly different than 1 is then applied for each bin.

Author Contributions

P.F., O.J., M.G. and D.I. conceived the study. P.F. computed the results, compiled, and analyzed the data. P.F. wrote the initial draft of the paper. C.D.S. and E.P.: MATOU and unigene computing. J.M.A.: Bioinformatics Control Quality of Sequencing. E.P.: MGT computing. L.C. and A.R. helped with preliminary analysis and data compilation. E.V.: SimkaMin computing. P.F., O.J., D.I., M.G., E.P., L.K.B., C.B. and L.C. discussed the results and wrote the paper.

References

1. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science (1979)* (1998), doi:10.1126/science.281.5374.237.

2. L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. I. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. de Vargas, D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, Plankton networks driving carbon export in the oligotrophic ocean. *Nature* (2016), doi:10.1038/nature16942.

3. S. A. Henson, R. Sanders, E. Madsen, Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochem Cycles* (2012), doi:10.1029/2011GB004099.

4. P. Frémont, M. Gehlen, M. Vrac, J. Leconte, T. O. Delmont, P. Wincker, D. Iudicone, O. Jaillon, Restructuring of plankton genomic biogeography in the surface ocean under climate change. *Nat Clim Chang* (2022), doi:10.1038/s41558-022-01314-8.

5. F. Azam, T. Fenchel, J. Field, J. Gray, L. Meyer-Reil, F. Thingstad, The Ecological Role of Water-Column Microbes in the Sea. *Mar Ecol Prog Ser* (1983), doi:10.3354/meps010257.

6. G. Reygondeau, W. W. L. Cheung, C. C. C. Wabnitz, V. W. Y. Lam, T. Frölicher, O. Maury, Climate Change-Induced Emergence of Novel Biogeochemical Provinces. *Front Mar Sci*. **7** (2020), doi:10.3389/fmars.2020.00657.

7. A. R. Fay, G. A. McKinley, Global open-ocean biomes: Mean and temporal variability. *Earth Syst Sci Data* (2014), doi:10.5194/essd-6-273-2014.

8. D. J. Richter, R. Watteaux, T. Vannier, J. Leconte, P. Frémont, G. Reygondeau, N. Maillet, N. Henry, G. Benoit, O. da Silva, T. O. Delmont, A. Fernàndez-Guerra, S. Suweis, R. Narci, C. Berney, D. Eveillard, F. Gavory, L. Guidi, K. Labadie, E. Mahieu, J. Poulain, S. Romac, S. Roux, C. Dimier, S. Kandels, M. Picheral, S. Searson, T. O. Coordinators, S. Pesant, J.-M. Aury, J. R. Brum, C. Lemaitre, E. Pelletier, P. Bork, S. Sunagawa, F. Lombard, L. Karp-Boss, C. Bowler, M. B. Sullivan, E. Karsenti, M. Mariadassou, I. Probert, P. Peterlongo, P. Wincker, C. de Vargas, M. Ribera d'Alcalà, D. Iudicone, O. Jaillon, Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *Elife*. **11**, e78129 (2022).

9. G. Sommeria-Klein, R. Watteaux, F. M. Ibarbalz, J. Pierella Karlusich Juan, D. Iudicone, C. Bowler, H. Morlon, Global drivers of eukaryotic plankton biogeography in the sunlit ocean. *Science (1979)*. **374**, 594–599 (2021).

10. A. C. Palmisano, C. W. Sullivan, PHYSIOLOGY OF SEA ICE DIATOMS. I. RESPONSE OF THREE POLAR DIATOMS TO A SIMULATED SUMMER-WINTER TRANSITION. *J Phycol*. **18** (1982), doi:10.1111/j.1529-8817.1982.tb03215.x.

11. J. Leconte, Y. Timsit, T. O. Delmont, M. Lescot, G. Piganeau, P. Wincker, O. Jaillon, *bioRxiv*, in press, doi:10.1101/2021.07.13.452163.

12. N. Guérin, M. Ciccarella, E. Flamant, P. Frémont, S. Mangenot, B. Istace, B. Noel, S. Romac, C. Bachy, M. Gachenot, E. Pelletier, A. Alberti, O. Jaillon, C. Cruaud, P. Wincker, J.-M. Aury, Q. Carradec, *bioRxiv*, in press, doi:10.1101/2021.10.25.465678.

13. T. Rossby, The North Atlantic Current and surrounding waters: At the crossroads. *Reviews of Geophysics*. **34** (1996), doi:10.1029/96RG02214.

14. M. Steele, T. Boyd, Retreat of the cold halocline layer in the Arctic Ocean. *J Geophys Res Oceans*. **103** (1998), doi:10.1029/98jc00580.

15. I. v. Polyakov, L. A. Timokhov, V. A. Alexeev, S. Bacon, I. A. Dmitrenko, L. Fortier, I. E. Frolov, J. C. Gascard, E. Hansen, V. v. Ivanov, S. Laxon, C. Mauritzen, D. Perovich, K. Shimada, H. L. Simmons, V. T. Sokolov, M. Steele, J. Toole, Arctic ocean warming contributes to reduced polar ice cap. *J Phys Oceanogr*. **40** (2010), doi:10.1175/2010JPO4339.1.

16. I. v. Polyakov, A. v. Pnyushkov, M. B. Alkire, I. M. Ashik, T. M. Baumann, E. C. Carmack, I. Goszczko, J. Guthrie, V. v. Ivanov, T. Kanzow, R. Krishfield, R. Kwok, A. Sundfjord, J. Morison, R. Rember, A. Yulin, Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. *Science (1979)*. **356** (2017), doi:10.1126/science.aai8204.

17. I. v. Polyakov, T. P. Rippeth, I. Fer, M. B. Alkire, T. M. Baumann, E. C. Carmack, R. Ingvaldsen, V. v. Ivanov, M. Janout, S. Lind, L. Padman, A. v. Pnyushkov, R. Rember, Weakening of cold halocline layer exposes sea ice to oceanic heat in the eastern arctic ocean. *J Clim*. **33** (2020), doi:10.1175/JCLI-D-19-0976.1.

18. L. Oziel, A. Baudena, M. Ardyna, P. Massicotte, A. Randelhoff, J. B. Sallée, R. B. Ingvaldsen, E. Devred, M. Babin, Faster Atlantic currents drive poleward expansion of temperate phytoplankton in the Arctic Ocean. *Nat Commun*. **11** (2020), doi:10.1038/s41467-020-15485-5.

19. G. Neukermans, L. Oziel, M. Babin, Increased intrusion of warming Atlantic water leads to rapid expansion of temperate phytoplankton in the Arctic. *Glob Chang Biol*. **24** (2018), doi:10.1111/gcb.14075.

20. D. A. Hutchins, F. Fu, Microorganisms and ocean global change. *Nat Microbiol* (2017), doi:10.1038/nmicrobiol.2017.58.

21. J. Goldsmit, C. W. McKindsey, R. W. Schlegel, D. B. Stewart, P. Archambault, K. L. Howland, What and where? Predicting invasion hotspots in the Arctic marine realm. *Glob Chang Biol*. **26** (2020), doi:10.1111/gcb.15159.

22. M. Moulager, A. Monnier, B. Jesson, R. Bouvet, J. Mosser, C. Schwartz, L. Garnier, F. Corellou, F. Y. Bouget, Light-dependent regulation of cell division in Ostreococcus: Evidence for a major transcriptional input. *Plant Physiol*. **144** (2007), doi:10.1104/pp.107.096149.

23. J. S. O-Neill, G. van Ooijen, L. E. Dixon, C. Troein, F. Corellou, F. Y. Bouget, A. B. Reddy, A. J. Millar, Circadian rhythms persist without transcription in a eukaryote. *Nature*. **469** (2011), doi:10.1038/nature09654.

24. G. Salazar, L. Paoli, A. Alberti, J. Huerta-Cepas, H. J. Ruscheweyh, M. Cuenca, C. M. Field, L. P. Coelho, C. Cruaud, S. Engelen, A. C. Gregory, K. Labadie, C. Marec, E. Pelletier, M. Royo-Llonch, S. Roux, P. Sánchez, H. Uehara, A. A. Zayed, G. Zeller, M. Carmichael, C. Dimier, J. Ferland, S. Kandels, M. Picheral, S. Pisarev, J. Poulain, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O.

Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell* (2019), doi:10.1016/j.cell.2019.10.014.

25. L. J. Ustick, A. A. Larkin, C. A. Garcia, N. S. Garcia, M. L. Brock, J. A. Lee, N. A. Wiseman, J. Keith Moore, A. C. Martiny, Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. *Science (1979)*. **372** (2021), doi:10.1126/science.abe6301.

26. L. Caputi, Q. Carradec, D. Eveillard, A. Kirilovsky, E. Pelletier, J. J. Pierella Karlusich, F. Rocha Jimenez Vieira, E. Villar, S. Chaffron, S. Malviya, E. Scalco, S. G. Acinas, A. Alberti, J. M. Aury, A. S. Benoiston, A. Bertrand, T. Biard, L. Bittner, M. Boccara, J. R. Brum, C. Brunet, G. Busseni, A. Carratalà, H. Claustre, L. P. Coelho, S. Colin, S. D'Aniello, C. da Silva, M. del Core, H. Doré, S. Gasparini, F. Kokoszka, J. L. Jamet, C. Lejeusne, C. Lepoivre, M. Lescot, G. Lima-Mendez, F. Lombard, J. Lukeš, N. Maillet, M. A. Madoui, E. Martinez, M. G. Mazzocchi, M. B. Néou, J. Paz-Yepes, J. Poulain, S. Ramondenc, J. B. Romagnan, S. Roux, D. Salvagio Manta, R. Sanges, S. Speich, M. Sprovieri, S. Sunagawa, V. Taillandier, A. Tanaka, L. Tirichine, C. Trottier, J. Uitz, A. Veluchamy, J. Veselá, F. Vincent, S. Yau, S. Kandels-Lewis, S. Searson, C. Dimier, M. Picheral, P. Bork, E. Boss, C. de Vargas, M. J. Follows, N. Grimsley, L. Guidi, P. Hingamp, E. Karsenti, P. Sordino, L. Stemmann, M. B. Sullivan, A. Tagliabue, A. Zingone, L. Garczarek, F. d'Ortenzio, P. Testor, F. Not, M. R. d'Alcalà, P. Wincker, C. Bowler, D. Iudicone, Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. *Global Biogeochem Cycles*. **33** (2019), doi:10.1029/2018GB006022.

27. T. Mock, R. P. Otillar, J. Strauss, M. McMullan, P. Paajanen, J. Schmutz, A. Salamov, R. Sanges, A. Toseland, B. J. Ward, A. E. Allen, C. L. Dupont, S. Frickenhaus, F. Maumus, A. Veluchamy, T. Wu, K. W. Barry, A. Falciatore, M. I. Ferrante, A. E. Fortunato, G. Glöckner, A. Gruber, R. Hipkin, M. G. Janech, P. G. Kroth, F. Leese, E. A. Lindquist, B. R. Lyon, J. Martin, C. Mayer, M. Parker, H. Quesneville, J. A. Raymond, C. Uhlig, R. E. Valas, K. U. Valentin, A. Z. Worden, E. V. Armbrust, M. D. Clark, C. Bowler, B. R. Green, V. Moulton, C. van Oosterhout, I. v. Grigoriev, Evolutionary genomics of the cold-Adapted diatom Fragilariopsis cylindrus. *Nature*. **541** (2017), doi:10.1038/nature20803.

28. C. M. Duarte, S. Agustí, P. Wassmann, J. M. Arrieta, M. Alcaraz, A. Coello, N. Marbà, I. E. Hendriks, J. Holding, I. García-Zarandona, E. Kritzberg, D. Vaqué, Tipping elements in the Arctic marine ecosystem. *Ambio*. **41** (2012), doi:10.1007/s13280-011-0224-7.

29. P. Friedlingstein, M. O'Sullivan, M. W. Jones, R. M. Andrew, J. Hauck, A. Olsen, G. P. Peters, W. Peters, J. Pongratz, S. Sitch, C. le Quéré, J. G. Canadell, P. Ciais, R. B. Jackson, S. Alin, L. E. O. C. Aragão, A. Arneth, V. Arora, N. R. Bates, M. Becker, A. Benoit-Cattin, H. C. Bittig, L. Bopp, S. Bultan, N. Chandra, F. Chevallier, L. P. Chini, W. Evans, L. Florentie, P. M. Forster, T. Gasser, M. Gehlen, D. Gilfillan, T. Gkritzalis, L. Gregor, N. Gruber, I. Harris, K. Hartung, V. Haverd, R. A. Houghton, T. Ilyina, A. K. Jain, E. Joetzjer, K. Kadono, E. Kato, V. Kitidis, J. I. Korsbakken, P. Landschützer, N. Lefèvre, A. Lenton, S. Lienert, Z. Liu, D. Lombardozzi, G. Marland, N. Metzl, D. R. Munro, J. E. M. S. Nabel, S. I. Nakaoka, Y. Niwa, K. O'Brien, T. Ono, P. I. Palmer, D. Pierrot, B. Poulter, L. Resplandy, E. Robertson, C. Rödenbeck, J. Schwinger, R. Séférian, I. Skjelvan, A. J. P. Smith, A. J. Sutton, T. Tanhua, P. P. Tans, H. Tian, B. Tilbrook, G. van der Werf, N. Vuichard, A. P.

Walker, R. Wanninkhof, A. J. Watson, D. Willis, A. J. Wiltshire, W. Yuan, X. Yue, S. Zaehle, Global Carbon Budget 2020. *Earth Syst Sci Data*. **12** (2020), doi:10.5194/essd-12-3269-2020.

30. J. M. Holding, C. M. Duarte, J. M. Arrieta, R. Vaquer-Suyner, A. Coello-Camba, P. Wassmann, S. Agustí, Experimentally determined temperature thresholds for Arctic plankton community metabolism. *Biogeosciences*. **10** (2013), doi:10.5194/bg-10-357-2013.

31. J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, G. B. West, "Toward a metabolic theory of ecology" in *Ecology* (2004), vol. 85.

32. Q. Carradec, E. Pelletier, C. da Silva, A. Alberti, Y. Seeleuthner, R. Blanc-Mathieu, G. Lima-Mendez, F. Rocha, L. Tirichine, K. Labadie, A. Kirilovsky, A. Bertrand, S. Engelen, M. A. Madoui, R. Méheust, J. Poulain, S. Romac, D. J. Richter, G. Yoshikawa, C. Dimier, S. Kandels-Lewis, M. Picheral, S. Searson, S. G. Acinas, E. Boss, M. Follows, G. Gorsky, N. Grimsley, L. Karp-Boss, U. Krzic, S. Pesant, E. G. Reynaud, C. Sardet, M. Sieracki, S. Speich, L. Stemmann, D. Velayoudon, J. Weissenbach, O. Jaillon, J. M. Aury, E. Karsenti, M. B. Sullivan, S. Sunagawa, P. Bork, F. Not, P. Hingamp, J. Raes, L. Guidi, H. Ogata, C. de Vargas, D. Iudicone, C. Bowler, P. Wincker, A global ocean atlas of eukaryotic genes. *Nat Commun* (2018), doi:10.1038/s41467-017-02342-1.

33. M. Royo-Llonch, P. Sánchez, C. Ruiz-González, G. Salazar, C. Pedrós-Alió, M. Sebastián, K. Labadie, L. Paoli, F. M. Ibarbalz, L. Zinger, B. Churcheward, M. Babin, P. Bork, E. Boss, G. Cochrane, C. de Vargas, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Setmmann, M. B. Sullivan, S. Chaffron, D. Eveillard, E. Karsenti, S. Sunagawa, P. Wincker, L. Karp-Boss, C. Bowler, S. G. Acinas, T. O. Coordinators, Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol*. **6**, 1561–1574 (2021).

34. P. Wassmann, K. N. Kosobokova, D. Slagstad, K. F. Drinkwater, R. R. Hopcroft, S. E. Moore, I. Ellingsen, R. J. Nelson, E. Carmack, E. Popova, J. Berge, The contiguous domains of Arctic Ocean advection: Trails of life and death. *Prog Oceanogr*. **139** (2015), doi:10.1016/j.pocean.2015.06.011.

35. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: Tool for the unification of biology. *Nat Genet*. **25** (2000), , doi:10.1038/75556.

36. Y. Wang, C. Pang, X. Li, Z. Hu, Z. Lv, B. Zheng, P. Chen, Identification of tRNA nucleoside modification genes critical for stress response and development in rice and Arabidopsis. *BMC Plant Biol*. **17** (2017), doi:10.1186/s12870-017-1206-0.

37. C. Zhang, G. Jia, Reversible RNA Modification N1-methyladenosine (m1A) in mRNA and tRNA. *Genomics Proteomics Bioinformatics*. **16** (2018), , doi:10.1016/j.gpb.2018.03.003.

38. M. Alriquet, G. Calloni, A. MartÄ±Ânez-LimoÂn, R. Delli Ponti, G. Hanspach, M. Hengesbach, G. G. Tartaglia, R. M. Vabulas, The protective role of m1A during stress-induced granulation. *J Mol Cell Biol*. **12** (2020), doi:10.1093/jmcb/mjaa023.

713 39. C. Valadon, O. Namy, The importance of the Epi-transcriptome in translation fidelity. *Noncoding*
714 *RNA*. **7** (2021), , doi:10.3390/ncrna7030051.

715 40. Y. L. J. Pang, R. Abo, S. S. Levine, P. C. Dedon, Diverse cell stresses induce unique patterns of tRNA
716 up- and down-regulation: tRNA-seq for quantifying changes in tRNA copy number. *Nucleic Acids*
717 *Res*. **42** (2014), doi:10.1093/nar/gku945.

718 41. A. Fluitt, E. Pienaar, H. Viljoen, Ribosome kinetics and aa-tRNA competition determine rate and
719 fidelity of peptide synthesis. *Comput Biol Chem*. **31** (2007),
720 doi:10.1016/j.compbiolchem.2007.07.003.

721 42. T. J. Nicholls, J. Rorbach, M. Minczuk, Mitochondria: Mitochondrial RNA metabolism and human
722 disease. *Int J Biochem Cell Biol*. **45**, 845–849 (2013).

723 43. G. Dieci, M. Preti, B. Montanini, Eukaryotic snoRNAs: A paradigm for gene expression flexibility.
724 *Genomics*. **94**, 83–88 (2009).

725 44. B. Tiwari, K. Habermann, M. A. Arif, H. L. Weil, A. Garcia-Molina, T. Kleine, T. Mühlhaus, W. Frank,
726 Identification of small RNAs during cold acclimation in Arabidopsis thaliana. *BMC Plant Biol*. **20**
727 (2020), doi:10.1186/s12870-020-02511-3.

728 45. R. A. VanBogelen, F. C. Neidhardt, Ribosomes as sensors of heat and cold shock in Escherichia
729 coli. *Proc Natl Acad Sci U S A*. **87** (1990), doi:10.1073/pnas.87.15.5589.

730 46. D. O. Bayles, M. H. Tunick, T. A. Foglia, A. J. Miller, Cold shock and its effect on ribosomes and
731 thermal tolerance in Listeria monocytogenes. *Appl Environ Microbiol*. **66** (2000),
732 doi:10.1128/AEM.66.10.4351-4355.2000.

733 47. A. D. Fernandes, J. M. Macklaim, T. G. Linn, G. Reid, G. B. Gloor, ANOVA-Like Differential
734 Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One*. **8** (2013),
735 doi:10.1371/journal.pone.0067019.

736 48. A. D. Fernandes, J. N. S. Reid, J. M. Macklaim, T. A. McMurrough, D. R. Edgell, G. B. Gloor,
737 Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA
738 gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*.
739 **2** (2014), doi:10.1186/2049-2618-2-15.

740 49. T. O. Delmont, M. Gaia, D. D. Hinsinger, P. Frémont, C. Vanni, A. Fernandez-Guerra, A. M. Eren, A.
741 Kourlaiev, L. d'Agata, Q. Clayssen, E. Villar, K. Labadie, C. Cruaud, J. Poulain, C. da Silva, M.
742 Wessner, B. Noel, J.-M. Aury, S. Sunagawa, S. G. Acinas, P. Bork, E. Karsenti, C. Bowler, C. Sardet,
743 L. Stemmann, C. de Vargas, P. Wincker, M. Lescot, M. Babin, G. Gorsky, N. Grimsley, L. Guidi, P.
744 Hingamp, O. Jaillon, S. Kandels, D. Iudicone, H. Ogata, S. Pesant, M. B. Sullivan, F. Not, K.-B. Lee,
745 E. Boss, G. Cochrane, M. Follows, N. Poulton, J. Raes, M. Sieracki, S. Speich, C. de Vargas, C.
746 Bowler, E. Karsenti, E. Pelletier, P. Wincker, O. Jaillon, Functional repertoire convergence of
747 distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics*. **2**,
748 100123 (2022).

749 50. T. Ohira, T. Suzuki, Precursors of tRNAs are stabilized by methylguanosine cap structures. *Nat*
750 *Chem Biol*. **12** (2016), doi:10.1038/nchembio.2117.

51. S. Abu-Ghosh, Z. Dubinsky, D. Iluz, Acclimation of thermotolerant algae to light and temperature interaction. *J Phycol*. **56** (2020), doi:10.1111/jpy.12964.

52. T. O. Delmont, E. Kiefl, O. Kilinc, O. C. Esen, I. Uysal, M. S. Rappé, S. Giovannoni, A. M. Eren, Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *Elife* (2019), doi:10.7554/eLife.46497.

53. R. Laso-Jadart, K. Sugier, E. Petit, K. Labadie, P. Peterlongo, C. Ambroise, P. Wincker, J. L. Jamet, M. A. Madoui, Investigating population-scale allelic differential expression in wild populations of Oithona similis (Cyclopoida, Claus, 1866). *Ecol Evol* (2020), doi:10.1002/ece3.6588.

54. L. Schlüter, K. T. Lohbeck, J. P. Gröger, U. Riebesell, T. B. H. Reusch, Long-term dynamics of adaptive evolution in a globally important phytoplankton species to ocean acidification. *Sci Adv*. **2** (2016), doi:10.1126/sciadv.1501660.

55. P. Jin, S. Agustí, Fast adaptation of tropical diatoms to increased warming with trade-offs. *Sci Rep*. **8** (2018), doi:10.1038/s41598-018-36091-y.

56. A. J. Irwin, Z. v. Finkel, F. E. Müller-Karger, L. Troccoli Ghinaglia, Phytoplankton adapt to changing ocean environments. *Proceedings of the National Academy of Sciences* (2015), doi:10.1073/pnas.1414752112.

57. D. J. R. Thornalley, D. W. Oppo, P. Ortega, J. I. Robson, C. M. Brierley, R. Davis, I. R. Hall, P. Moffa-Sanchez, N. L. Rose, P. T. Spooner, I. Yashayaev, L. D. Keigwin, Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years. *Nature*. **556** (2018), doi:10.1038/s41586-018-0007-4.

58. L. Johannes, D. P. D., Risk of tipping the overturning circulation due to increasing rates of ice melt. *Proceedings of the National Academy of Sciences*. **118**, e2017989118 (2021).

59. K. H. Kilbourne, A. D. Wanamaker, P. Moffa-Sanchez, D. J. Reynolds, D. E. Amrhein, P. G. Butler, G. Gebbie, M. Goes, M. F. Jansen, C. M. Little, M. Mette, E. Moreno-Chamarro, P. Ortega, B. L. Otto-Bliesner, T. Rossby, J. Scourse, N. M. Whitney, Atlantic circulation change still uncertain. *Nat Geosci*. **15**, 165–167 (2022).

60. S. Rahmstorf, J. E. Box, G. Feulner, M. E. Mann, A. Robinson, S. Rutherford, E. J. Schaffernicht, Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat Clim Chang* (2015), doi:10.1038/nclimate2554.

61. A. Vorobev, M. Dupouy, Q. Carradec, T. O. Delmont, A. Annamalé, P. Wincker, E. Pelletier, Transcriptome reconstruction and functional analysis of eukaryotic marine plankton communities via high-throughput metagenomics and metatranscriptomics. *Genome Res*. **30** (2020), doi:10.1101/gr.253070.119.

62. H. B. Nielsen, M. Almeida, A. S. Juncker, S. Rasmussen, J. Li, S. Sunagawa, D. R. Plichta, L. Gautier, A. G. Pedersen, E. le Chatelier, E. Pelletier, I. Bonde, T. Nielsen, C. Manichanh, M. Arumugam, J. M. Batto, M. B. Quintanilha Dos Santos, N. Blom, N. Borruel, K. S. Burgdorf, F. Boumezbeur, F. Casellas, J. Doré, P. Dworzynski, F. Guarner, T. Hansen, F. Hildebrand, R. S. Kaas, S. Kennedy, K. Kristiansen, J. R. Kultima, P. Leonard, F. Levenez, O. Lund, B. Moumen, D. le Paslier, N. Pons, O.

Pedersen, E. Prifti, J. Qin, J. Raes, S. Sørensen, J. Tap, S. Tims, D. W. Ussery, T. Yamada, P. Renault, T. Sicheritz-Ponten, P. Bork, J. Wang, S. Brunak, S. D. Ehrlich, A. Jamet, A. Mérieux, A. Cultrone, A. Torrejon, B. Quinquis, C. Brechot, C. Delorme, C. M'rini, W. M. de Vos, E. Maguin, E. Varela, E. Guedon, F. Gwen, F. Haimet, F. Artiguenave, G. Vandemeulebrouck, G. Denariaz, G. Khaci, H. Blottière, J. Knol, J. Weissenbach, J. E. T. van Hylckama Vlieg, J. Torben, J. Parkhill, K. Turner, M. van de Guchte, M. Antolin, M. Rescigno, M. Kleerebezem, M. Derrien, N. Galleron, N. Sanchez, N. Grarup, P. Veiga, R. Oozeer, R. Dervyn, S. Layec, T. Bruls, Y. Winogradski, Z. Erwin, Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol*. **32** (2014), doi:10.1038/nbt.2939.

63. L. Guidi, M. Picheral, S. Pesant, C. Tara Oceans Consortium, P. Tara Oceans Expedition, "Environmental context of all samples from the Tara Oceans Expedition (2009-2013), about sensor data in the targeted environmental feature" in (PANGAEA, 2017; https://doi.org/10.1594/PANGAEA.875576).

64. L. Guidi, P. Morin, L. Coppola, J.-É. Tremblay, S. Pesant, C. Tara Oceans Consortium, P. Tara Oceans Expedition, "Environmental context of all samples from the Tara Oceans Expedition (2009-2013), about nutrients in the targeted environmental feature" in (PANGAEA, 2017; https://doi.org/10.1594/PANGAEA.875575).

65. H. Garcia, K. W. Weathers, C. R. Paver, I. Smolyar, T. P. Boyer, R. A. Locarnini, M. M. Zweng, A. V. Mishonov, O. K. Baranova, D. Seidov, J. R. Reagan, World Ocean Atlas 2018. Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate). *NOAA Atlas NESDIS 84*. **84** (2018).

66. P. Brun, M. Vogt, M. R. Payne, N. Gruber, C. J. O'Brien, E. T. Buitenhuis, C. le Quéré, K. Leblanc, Y. W. Luo, Ecological niches of open ocean phytoplankton taxa. *Limnol Oceanogr*. **60** (2015), doi:10.1002/lno.10074.

67. N. Gruber, J. L. Sarmiento, Global patterns of marine nitrogen fixation and denitrification. *Global Biogeochem Cycles*. **11** (1997), doi:10.1029/97GB00077.

68. C. Deutsch, J. L. Sarmiento, D. M. Sigman, N. Gruber, J. P. Dunne, Spatial coupling of nitrogen inputs and losses in the ocean. *Nature*. **445** (2007), doi:10.1038/nature05392.

69. O. Aumont, C. Ethé, A. Tagliabue, L. Bopp, M. Gehlen, PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geosci Model Dev* (2015), doi:10.5194/gmd-8-2465-2015.

70. A. Tagliabue, O. Aumont, R. Death, J. P. Dunne, S. Dutkiewicz, E. Galbraith, K. Misumi, J. K. Moore, A. Ridgwell, E. Sherman, C. Stock, M. Vichi, C. Völker, A. Yool, How well do global ocean biogeochemistry models simulate dissolved iron distributions? *Global Biogeochem Cycles*. **30** (2016), doi:10.1002/2015GB005289.

71. P. C. Legendre, L. Legendre, Numerical Ecology, Volume 24. *(Developments in Environmental Modelling)* (1988), doi:10.1017/CBO9781107415324.004.

72. S. Nielsen, Numerical Ecology. Legendre P. and Legendre L., second ed., Elsevier, Amsterdam, p. 853, 1998. *Ecological Modelling - ECOL MODEL*. **132**, 303–304 (2000).

73. G. Benoit, M. Mariadassou, S. Robin, S. Schbath, P. Peterlongo, C. Lemaitre, SimkaMin: Fast and resource frugal de novo comparative metagenomics. *Bioinformatics*. **36** (2020), doi:10.1093/bioinformatics/btz685.

74. G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, B. Venables, Package "gplots": Various R programming tools for plotting data. *R package version 2.17.0.* (2016).

75. R. J. Simes, An improved bonferroni procedure for multiple tests of significance. *Biometrika*. **73** (1986), doi:10.1093/biomet/73.3.751.

76. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, M. Punta, Pfam: The protein families database. *Nucleic Acids Res*. **42** (2014), , doi:10.1093/nar/gkt1223.

77. S. Carbon, E. Douglass, B. M. Good, D. R. Unni, N. L. Harris, C. J. Mungall, S. Basu, R. L. Chisholm, R. J. Dodson, E. Hartline, P. Fey, P. D. Thomas, L. P. Albou, D. Ebert, M. J. Kesling, H. Mi, A. Muruganujan, X. Huang, T. Mushayahama, S. A. LaBonte, D. A. Siegele, G. Antonazzo, H. Attrill, N. H. Brown, P. Garapati, S. J. Marygold, V. Trovisco, G. dos Santos, K. Falls, C. Tabone, P. Zhou, J. L. Goodman, V. B. Strelets, J. Thurmond, P. Garmiri, R. Ishtiaq, M. Rodríguez-López, M. L. Acencio, M. Kuiper, A. Lægreid, C. Logie, R. C. Lovering, B. Kramarz, S. C. C. Saverimuttu, S. M. Pinheiro, H. Gunn, R. Su, K. E. Thurlow, M. Chibucos, M. Giglio, S. Nadendla, J. Munro, R. Jackson, M. J. Duesbury, N. Del-Toro, B. H. M. Meldal, K. Paneerselvam, L. Perfetto, P. Porras, S. Orchard, A. Shrivastava, H. Y. Chang, R. D. Finn, A. L. Mitchell, N. D. Rawlings, L. Richardson, A. Sangrador-Vegas, J. A. Blake, K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. M. Sitnikov, M. A. Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Hayles, J. Bähler, E. R. Bolton, J. L. de Pons, M. R. Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F. Laulederkind, C. Plasterer, M. A. Tutaj, M. Vedi, S. J. Wang, P. D'Eustachio, L. Matthews, J. P. Balhoff, S. A. Aleksander, M. J. Alexander, J. M. Cherry, S. R. Engel, F. Gondwe, K. Karra, S. R. Miyasato, R. S. Nash, M. Simison, M. S. Skrzypek, S. Weng, E. D. Wong, M. Feuermann, P. Gaudet, A. Morgat, E. Bakker, T. Z. Berardini, L. Reiser, S. Subramaniam, E. Huala, C. N. Arighi, A. Auchincloss, K. Axelsen, G. Argoud-Puy, A. Bateman, M. C. Blatter, E. Boutet, E. Bowler, L. Breuza, A. Bridge, R. Britto, H. Bye-A-Jee, C. C. Casas, E. Coudert, P. Denny, A. Es-Treicher, M. L. Famiglietti, G. Georghiou, A. N. Gos, N. Gruaz-Gumowski, E. Hatton-Ellis, C. Hulo, A. Ignatchenko, F. Jungo, K. Laiho, P. le Mercier, D. Lieberherr, A. Lock, Y. Lussi, A. MacDougall, M. Ma-Grane, M. J. Martin, P. Masson, D. A. Natale, N. Hyka-Nouspikel, S. Orchard, I. Pedruzzi, L. Pourcel, S. Poux, S. Pundir, C. Rivoire, E. Speretta, S. Sundaram, N. Tyagi, K. Warner, R. Zaru, C. H. Wu, A. D. Diehl, J. N. Chan, C. Grove, R. Y. N. Lee, H. M. Muller, D. Raciti, K. van Auken, P. W. Sternberg, M. Berriman, M. Paulini, K. Howe, S. Gao, A. Wright, L. Stein, D. G. Howe, S. Toro, M. Westerfield, P. Jaiswal, L. Cooper, J. Elser, The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res*. **49** (2021), doi:10.1093/nar/gkaa1113.

78. Student, The Probable Error of a Mean. *Biometrika*. **6**, 1–25 (1908).

79. E. T. Jaynes, *Probability Theory* (Cambridge University Press, 2003; https://www.cambridge.org/core/product/identifier/9780511790423/type/book).

80. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. **57** (1995), doi:10.1111/j.2517-6161.1995.tb02031.x.

81. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome datasets are compositional: And this is not optional. *Front Microbiol*. **8** (2017), , doi:10.3389/fmicb.2017.02224.

82. I. Erb, C. Notredame, How should we measure proportionality on relative gene expression data? *Theory in Biosciences*. **135** (2016), doi:10.1007/s12064-015-0220-8.

83. B. Efron, R. Tibshirani, J. D. Storey, V. Tusher, Empirical bayes analysis of a microarray experiment. *J Am Stat Assoc*. **96** (2001), doi:10.1198/016214501753382129.

84. V. G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. **98** (2001), doi:10.1073/pnas.091062498.

85. I. Rivals, L. Personnaz, L. Taing, M. C. Potier, Enrichment or depletion of a GO category within a class of genes: Which test? *Bioinformatics*. **23** (2007), , doi:10.1093/bioinformatics/btl633.

86. A. Brionne, A. Juanchich, C. Hennequet-Antier, ViSEAGO: A Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity. *BioData Min*. **12** (2019), doi:10.1186/s13040-019-0204-1.

87. T. Slimani, Description and Evaluation of Semantic Similarity Measures Approaches. *Int J Comput Appl*. **80** (2013), doi:10.5120/13897-1851.

88. G. Wang, D. H. Oh, M. Dassanayake, GOMCL: A toolkit to cluster, evaluate, and extract non-redundant associations of Gene Ontology-based functions. *BMC Bioinformatics*. **21** (2020), doi:10.1186/s12859-020-3447-4.

89. V. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. **10** (1966).

90. R. R. Sokal, A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull*. **38** (1958).

91. https://www.whoi.edu/know-your-ocean/ocean-topics/how-the-ocean-works/ocean-circulation/arctic-ocean-circulation/.

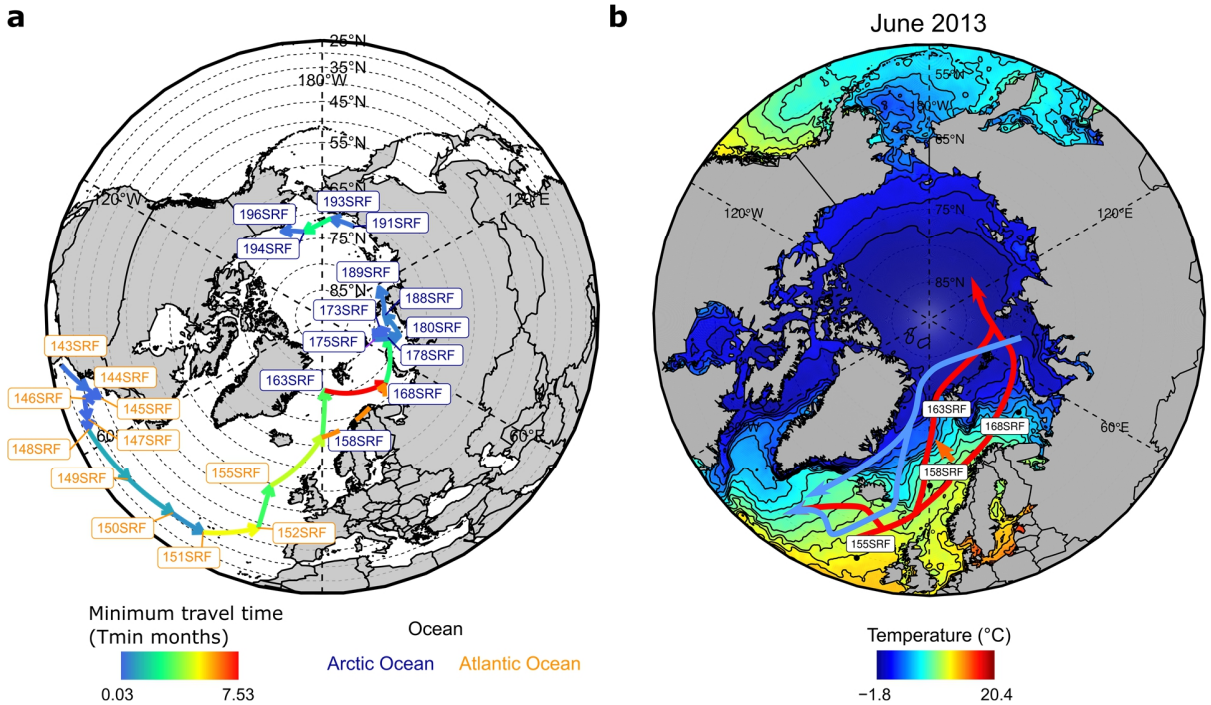


Figure 1 | Lagrangian connectivity of the North Atlantic Ocean and Arctic Ocean in the *Tara* Oceans and *Tara* Oceans Polar Circle expeditions and polar front at the time of sampling. (**a**) *Tara* Oceans and *Tara* Oceans Polar Circle stations considered in this study and shortest Lagrangian path (minimum travel time Tmin) between stations (starting at Station_143). (**b**) Mean Sea Surface Temperature in June 2013 around the time of sampling of stations 155 to 168 (between May and July). The orange arrow indicates approximately the polar front between Stations 158 and 163 where a steep temperature gradient was observed (black lines). Main inflow and outflow currents are represented, respectively, by red arrows (warm Atlantic inflows) and blue arrows (cold Arctic outflows). They are adapted from the Woods Hole Oceanographic Institution website (*91*). The orange arrow indicates the polar front.

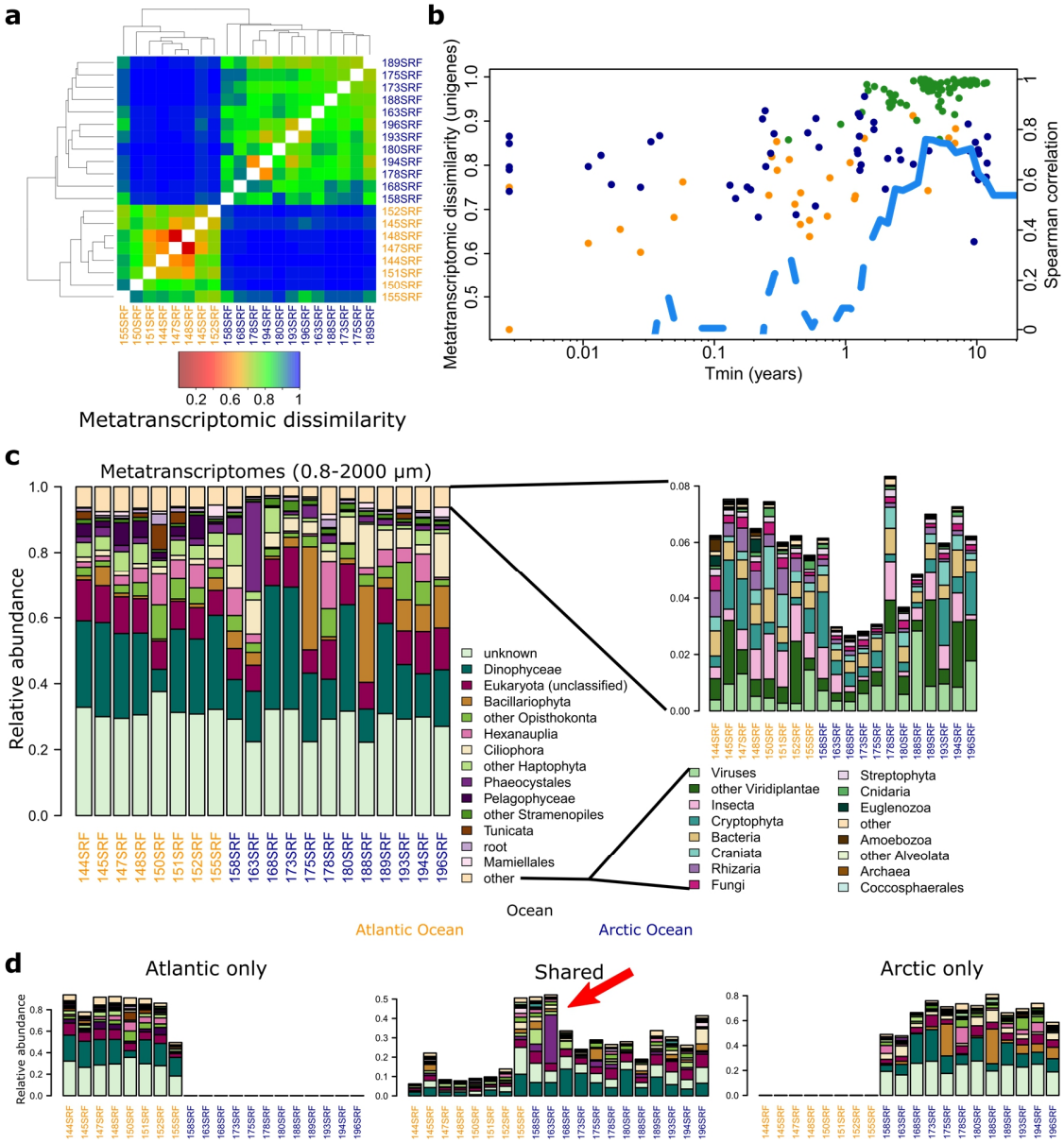


Figure 2 | Metatranscriptomic and taxonomical composition connectivity with respect to ocean currents of surface ocean plankton (size fraction 0.8-2000 µm). (**a**) Matrix of metatranscriptomic pairwise β diversity estimates based on unigene relative abundances (Bray-Curtis dissimilarity index) clustered using the UPGMA (*90*) algorithm. (**b**) Metatranscriptomic β diversity as a function of Lagrangian connectivity between pairs of *Tara* Oceans and *Tara* Oceans Polar Circle stations (minimum transport time) and cumulative Spearman correlation coefficient between the two metrics (blue line). Orange, dark blue and green points respectively indicate Atlantic, Arctic and Atlantic-Arctic pairs of stations. (**c**) Left: global metatranscriptomic taxonomical composition based on unigene relative abundances of major plankton taxa (taxonomic level can

vary) in the North Atlantic and Arctic Oceans. Right: Zoom on the 16 least abundant taxa (other). (**d**) Decomposition between basin-specific unigenes and shared unigenes between the two basins (classified based on presence or absence of expression). A clear transition enriched in shared genes appears in Stations 153, 158 and 163.

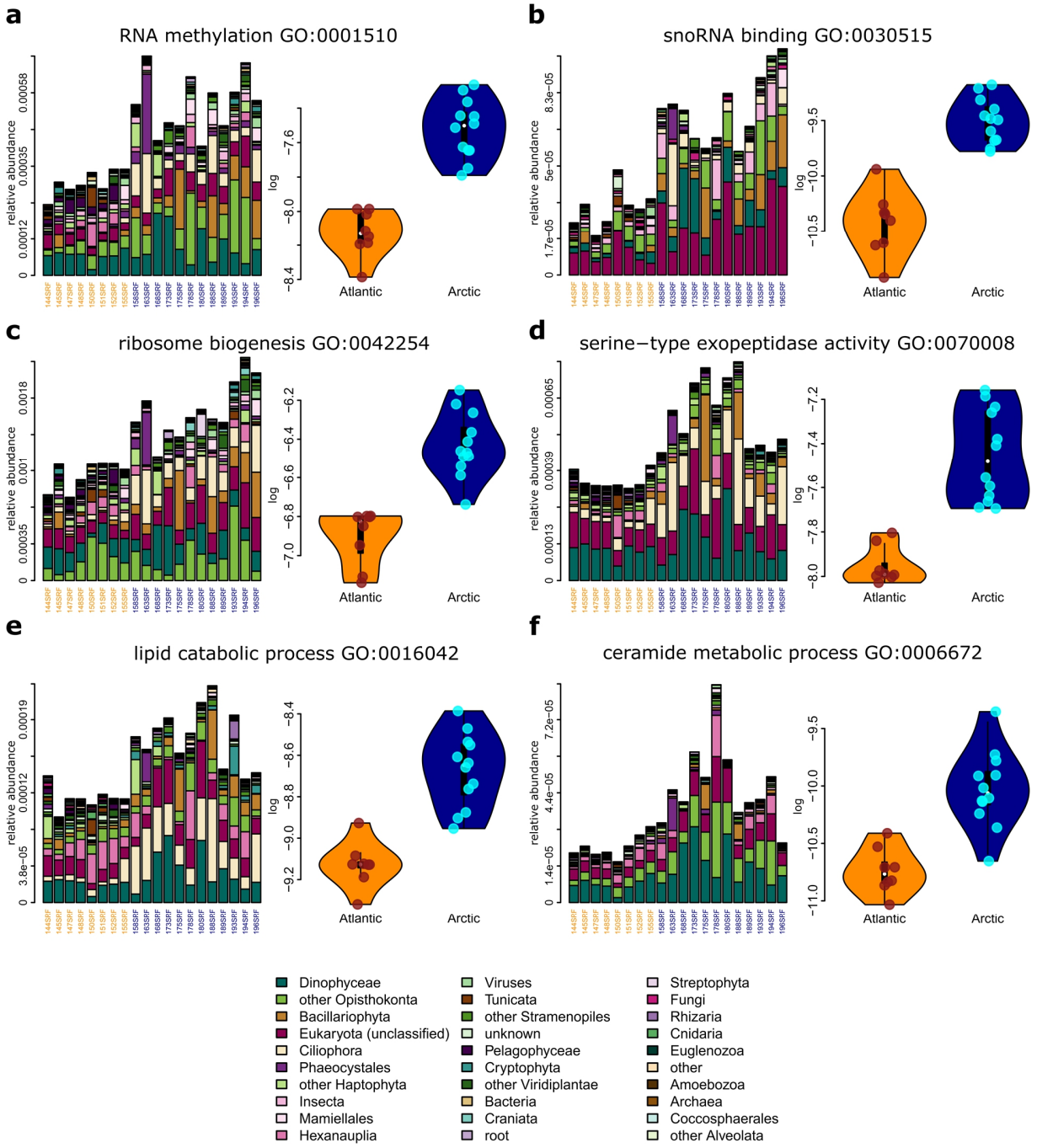


Figure 3 | Examples of Gene Ontology (GO) terms that are more abundant in Arctic metatranscriptomes. Each panel (**a-f**) represents (left) the sum of the metatranscriptomic relative abundances of unigenes associated with the abovementioned GO terms across considered taxa and (right) the distribution of this relative abundance in the Atlantic (dark blue) versus Arctic basin (purple). Each of these GO terms is significantly more abundant in the Arctic basin.

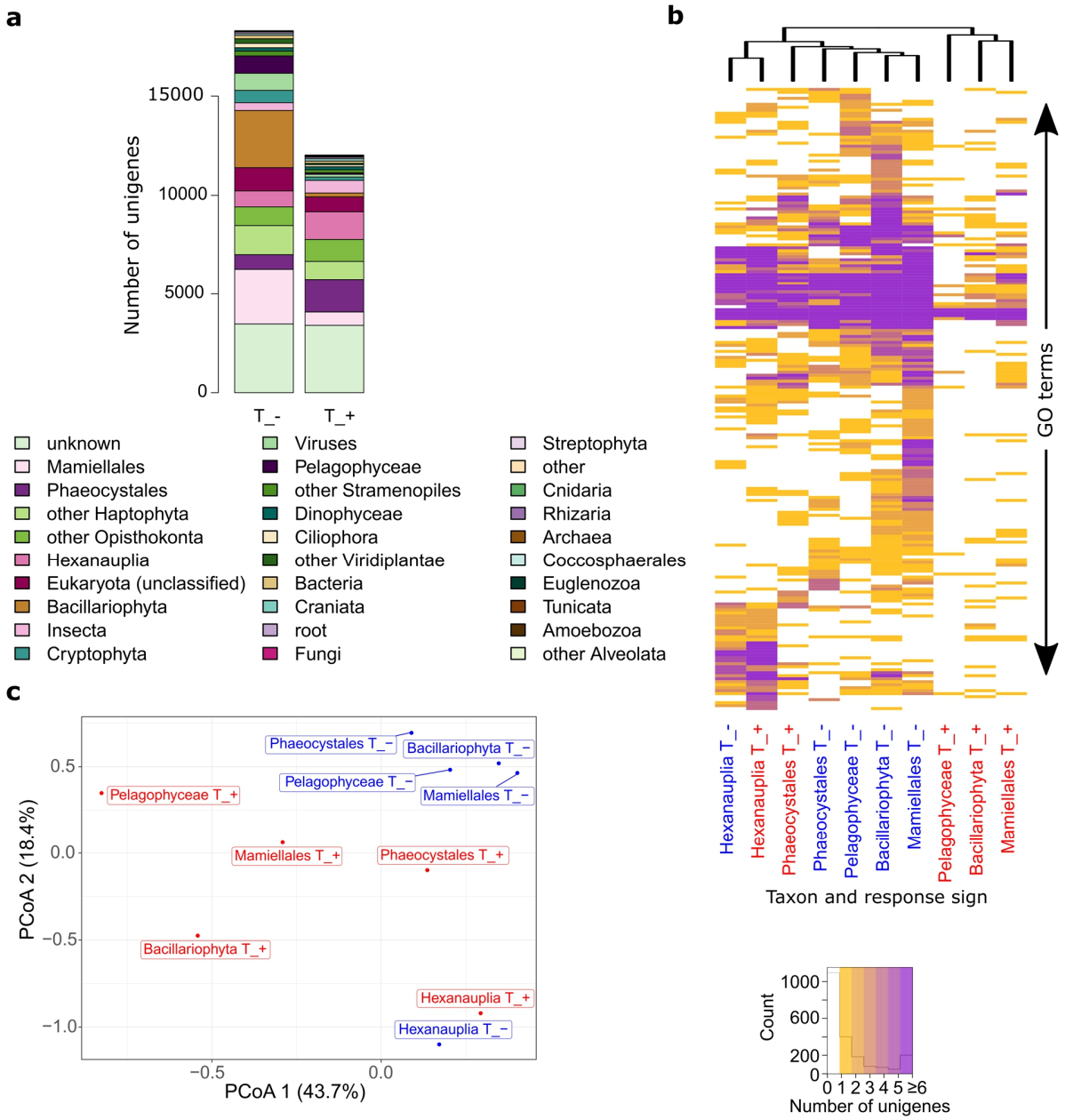
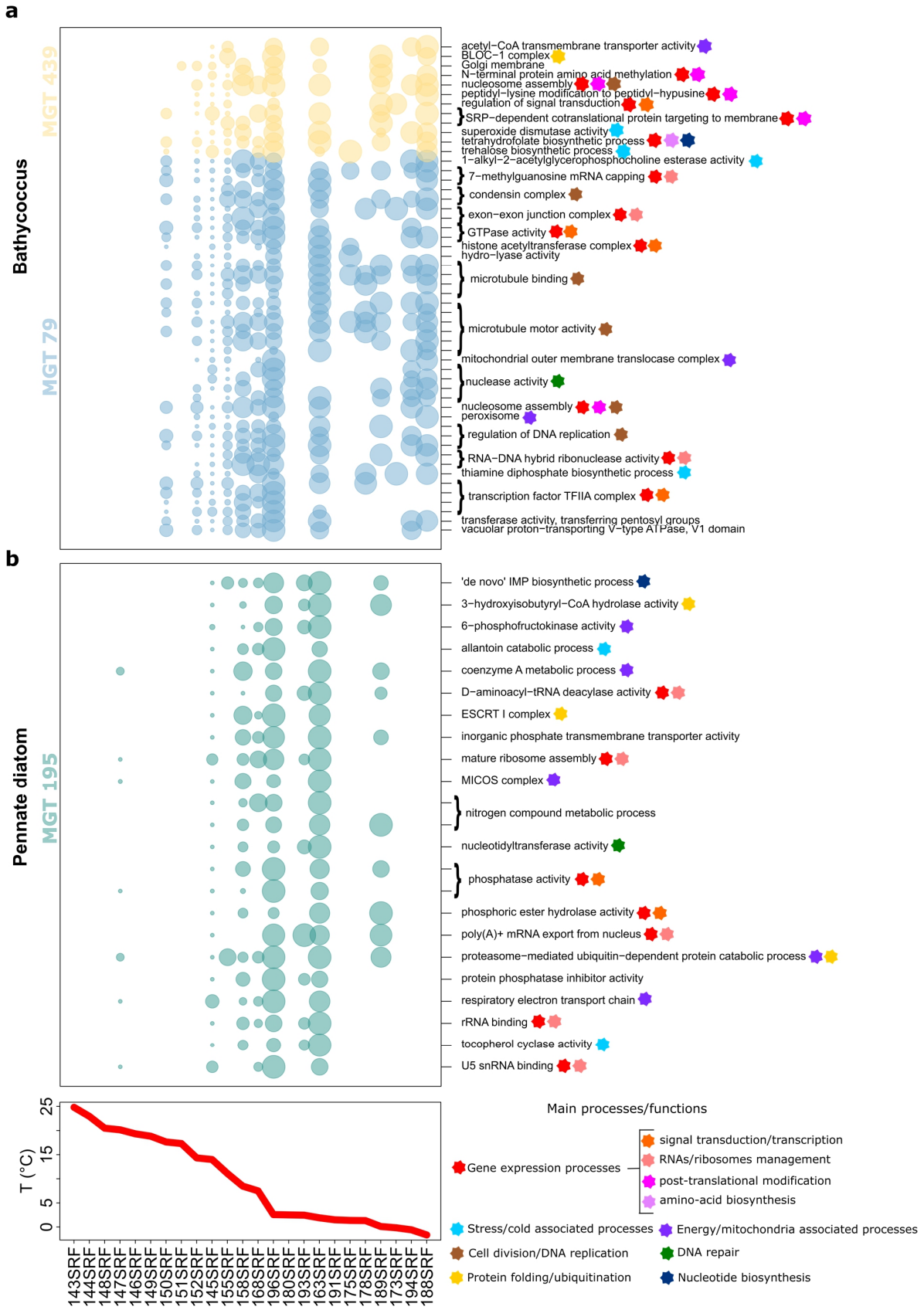


Figure 4 | Functional response to temperature at the unigene expression level for different taxa at the NAO/AO transition. (**a**) Taxonomic decomposition of unigenes significantly correlated with temperature, either negatively (left bar) or positively (right bar). (**b**) Heatmap of the number of unigenes and associated GO terms that correlate with temperature across main taxa. Each column represents the number of unigenes associated to a GO term (in line) for a given taxa and a correlation sign (blue: negative, red: positive). Columns are clustered based on Bray-Curtis dissimilarity index. (**c**) Bray-Curtis-based Principal Coordinate Analysis (PCoA) of the (**b**) panel. Noticeably the decrease in the functional response of the different phytoplankton groups

to temperature is similar while Hexanauplia (copepods) have a different functional landscape of response to temperature changes.


a
Bathycoccus
MGT 439
MGT 79
acetyl−CoA transmembrane transporter activity
BLOC−1 complex
Golgi membrane
N−terminal protein amino acid methylation
nucleosome assembly
peptidyl−lysine modification to peptidyl−hypusine
regulation of signal transduction
SRP−dependent cotranslational protein targeting to membrane
superoxide dismutase activity
tetrahydrofolate biosynthetic process
trehalose biosynthetic process
1−alkyl−2−acetylglycerophosphocholine esterase activity
7−methylguanosine mRNA capping
condensin complex
exon−exon junction complex
GTPase activity
histone acetyltransferase complex
hydro−lyase activity
microtubule binding
microtubule motor activity
mitochondrial outer membrane translocase complex
nuclease activity
nucleosome assembly
peroxisome
regulation of DNA replication
RNA−DNA hybrid ribonuclease activity
thiamine diphosphate biosynthetic process
transcription factor TFIIA complex
transferase activity, transferring pentosyl groups
vacuolar proton−transporting V−type ATPase, V1 domain
b
Pennate diatom
MGT 195
'de novo' IMP biosynthetic process
3−hydroxyisobutyryl−CoA hydrolase activity
6−phosphofructokinase activity
allantoin catabolic process
coenzyme A metabolic process
D−aminoacyl−tRNA deacylase activity
ESCRT I complex
inorganic phosphate transmembrane transporter activity
mature ribosome assembly
MICOS complex
nitrogen compound metabolic process
nucleotidyltransferase activity
phosphatase activity
phosphoric ester hydrolase activity
poly(A)+ mRNA export from nucleus
proteasome−mediated ubiquitin−dependent protein catabolic process
protein phosphatase inhibitor activity
respiratory electron transport chain
rRNA binding
tocopherol cyclase activity
U5 snRNA binding
T (°C)
0
5
15
25
143SRF
144SRF
148SRF
147SRF
146SRF
149SRF
150SRF
151SRF
152SRF
145SRF
155SRF
158SRF
168SRF
196SRF
180SRF
193SRF
163SRF
191SRF
175SRF
178SRF
189SRF
173SRF
194SRF
188SRF
Main processes/functions
Gene expression processes
signal transduction/transcription
RNAs/ribosomes management
post-translational modification
amino-acid biosynthesis
Stress/cold associated processes
Energy/mitochondria associated processes
Cell division/DNA replication
DNA repair
Protein folding/ubiquitination
Nucleotide biosynthesis

943 **Figure 5 | Expression profiles and Gene Ontology annotation of enriched unigenes**
944 **negatively correlated with temperature for three algal Metagenomics-based**
945 **Transcriptomes.** Expression profiles (left) of unigenes whose Gene Ontology terms (right) are
946 enriched within unigenes significantly negatively correlated to temperature are displayed for two
947 algal taxa from three Metagenomics-based Transcriptomes (MGTs): (**a**) *Bathycoccus* (MGT 79
948 and 439) (**b**) a pennate diatom (MGT 195, annotated as a Cymatosiraceae). For this MGT
949 unigenes, unassigned functions (unknown) are enriched; however they are too numerous to be
950 represented here.

951

952

953

954

Supplement of

Changes in functional composition and gene expression in eukaryotic plankton at the Atlantic-Arctic Polar front

Frémont et al.

Correspondence to: Paul Frémont (pfremont@genoscope.cns.fr), Olivier Jaillon (ojaillon@genoscope.cns.fr)

Supplementary Discussion 1 | Environmental differences between the North Atlantic Ocean (NAO) and the Arctic Ocean (AO) and link with current connectivity.

Using *in situ* environmental parameters from the *Tara* schooner, the North Atlantic Ocean and the Arctic Ocean are clearly distinguishable (Supplementary Figure 1a). The North Atlantic is warmer, saltier and has a stronger temperature seasonal variability while the Arctic Ocean is richer in nutrients, has longer daylight duration (in the considered seasons) and a higher nutrient seasonality (Supplementary Figure 1a). Stations 191 and 193 in the Arctic appear as outliers. These stations are more coastal and we note that stations 191 to 196 are influenced by Pacific waters (*1*). Correlating environmental distance and minimum travel time between stations, a maximum of correlation is found between 4 and 8 years of transport (Supplementary Figure 1b) and although the transition seem continuous environmentally (blue points to red points), high variability is also found within Arctic samples (green points and high values red points).

Supplementary Discussion 2 | Difference in taxonomic composition of the NAO/AO metatranscriptomes and metagenomes

We explore the differences in taxonomic composition of North Atlantic and Arctic Oceans metatranscriptomes and metagenomes. Composition of the metatranscriptomes (Figure 2 b-d) differs significantly from the metagenomes composition (Supplementary Figure 3-4). Strikingly, while Dinophyceae is the 1st most abundant taxa (after unknowns) in the metatranscriptomes, it represents a much lesser fraction of the metagenomes (10th most abundant taxa, 6th in terms of number of unigenes). In the metatranscriptomes, Ciliophora and Other Stramenopiles are significantly more relatively abundant in the Arctic than in the Atlantic while Pelagophyceae, Rhizaria, Amoebozoa, Other Alveolata, Archaea and Coccosphaerales are significantly more abundant in the Atlantic (Wilcoxon test, Hommel correction $p<0.05$, Supplementary Figure 5). In metagenomes, Ciliophora is significantly relatively more abundant in the Arctic while Pelagophyceae, Amoebozoa and 'other' are significantly relatively more abundant in the Atlantic (Wilcoxon test, Hommel correction $p<0.05$, Supplementary Figure 5). Importantly, a large part of metagenomes unigenes are not necessarily expressed in the metatranscriptomes representing a continuously decreasing part from western to eastern parts (from ~30% of the relative abundance in station 144 to ~10% in station 155) of the North Atlantic and then diminishing in the Arctic (Supplementary Figure 4). This might reflect the presence of dead cells in the Atlantic winter or at least less active cells. Indeed this North Atlantic temperate transect was sampled from January to March so one might expect an increasing activity of phytoplankton population as the spring bloom is approaching.

Supplementary Discussion 3 | Other functions more abundant in the Arctic metatranscriptomes.

Other functions more abundant in the Arctic metatranscriptomes are directly linked to lipid metabolism ('lipid catabolic process' (Figure 3e), 'fatty-acyl-CoA binding'), to organization of organelles membranes ('intrinsic component of organelle membrane' and 'integral component of organelle membrane') and membrane fluidity ('Ceramide metabolic process' and 'Sphingolipid metabolic process', also more abundant in metagenomes). The important abundance of functions linked to lipid catabolic process most likely reflect the need for membrane remodeling and free fatty acid removal in cold environments (*2*, *3*). Some enzymes involved in lipid catabolism have been reported to be involved in freezing tolerance in *A.thaliana* (*4*, *5*) and in response to cold (*6*) and N stress in algae (*3*). The regulation of membrane fluidity is directly linked to cold environment acclimation as for example shown in the chloroplasts of the green algae *Nannochloropsis* (*7*) and *Chlamydomonas reinhardtii* (*8*) and more generally in all plants especially with an important role of sphingolipids (*9*) and particularly ceramides (*10*). We also find functions linked to protein trafficking such as 'protein transmembrane transport' and 'mitochondrial inner membrane presequence translocase complex' more abundant in the Arctic. In contrast, functions linked to protein retention in their cellular compartments like 'maintenance of location', 'maintenance of protein localization in organelle' and 'protein retention in ER lumen' are found more abundant in North Atlantic metagenomes suggesting a less rapid turnover of proteins in winter Atlantic waters (Supplementary Table 1, Supplementary Discussion 2). Finally, functions linked to telomere maintenance such as 'telomere maintenance' and 'telomeric DNA binding' are also found more abundant in the Arctic metatranscriptomes which might reflect the need to maintain the integrity of chromosomes in extreme environments (*11*, *12*).

Supplementary Discussion 4 | PFAMs more abundant in the Arctic Ocean in link with cold stress acclimation

We investigate the different functioning of the spring/summer Arctic Ocean and the winter Atlantic Ocean by analyzing the differential abundances of a set of 22 PFAMs reported to be important in cold/hot stress response. We use the ALDEx2 R package (*13*) to characterize more abundant and overexpressed PFAMs between the two basins. The relevance of each PFAM to cold/hot stress acclimation is summarized in Supplementary Table 2. First, we visualize the inter-basin dispersion in function of the intra-basin dispersion of metagenomic and metatranscriptomic abundances and expression values fitted by the ALDEx2 package (Supplementary Figure 8). On a total of 11950 PFAMs, we find 1214 (respectively 758) and 4461 (respectively 3657) PFAMs significantly more abundant (Benjamini-Hochberg corrected Welch test, $p<0.05$) respectively in

the North Atlantic and the Arctic Ocean metagenomes (respectively metatranscriptomes). Most more abundant PFAMs in the North Atlantic have a high maximum within basin dispersion while PFAMs more abundant in the Arctic have lower maximum within basin dispersion. This observation might reflect the higher variability of the Arctic basin (Figure 2 and Supplementary Figure 3). Zooming on specific taxa (Supplementary Figure 8), we often find a bimodal pattern of within condition dispersion both in more abundant PFAMs in the Atlantic and Arctic. We formulate three hypotheses to explain these patterns. First, some specific lineages/species whose genomes are enriched in some functions probably differentiate the two basins. Second, some specific lineages enriched in specific functions might be sampled only in few stations (*e.g.* Diatoms which are particularly abundant in few arctic stations). Third, some functions might be highly overexpressed. On the other hand, expression dispersion is comparatively lower compared to abundances (Supplementary Figure 8) and only a few PFAMs are found to be highly overexpressed. We find relatively a lot of PFAMs highly overexpressed especially in the Arctic mainly in Dinoflagellates (Supplementary Figure 8f) suggesting an important role of expression regulation in these organisms. These results needs however to be taken with caution as here we normalize the metatranscriptomic abundances by metagenomic abundances that are not necessarily from the same unigenes between metagenomes and metatranscriptomes especially in Dinoflagellates (high genomic complexity of this group).

Regarding cold acclimation functions, we find that nearly all of them are significantly more abundant in the Arctic Ocean in both metagenomes and metatranscriptomes but not overexpressed (Supplementary Figure 9a). Strikingly this signal is the strongest when considering the overall abundance of these PFAMs (*i.e.* within the whole community) while specific taxons display various patterns (Supplementary Figure 9a-g). Particularly, these PFAMs tend to be more abundant in Mamiellales, Bacillariophyta, Pelagophyceae, Dinophyceae and Unknown while not necessarily in Phaeocystales and Ciliophora (which are more abundant in the Arctic). Noticeably, some of these PFAMs are overexpressed in Dinophyceae (Supplementary Figure 9f). Among the PFAMs we consider, the ice binding domain (PF11999) is largely the most excessively more abundant PFAM both at the community level and for specific taxons (~2^{15} higher abundance in the Arctic) while other PFAMs are more abundant at a lesser extent (~$2^{3 \text{ to } 5}$ higher abundance in the Arctic) (Supplementary Figure 9). Overall, these results might reflect the different strategies of the different plankton organisms to adapt and acclimate to an extreme environment such as the Arctic Ocean.

Supplementary Discussion 5 | Functions more abundant in Atlantic metagenomes and metatranscriptomes

We find 85 and 11 more abundant Gene Ontology (GO) terms (3 of which are common) respectively in Atlantic metagenomes and metatranscriptomes in comparison to the Arctic in size fraction 0.8-2000 µm (Supplementary Figure 10 and exhaustive lists in Supplementary Table 1). The first functions that jump to the eye when looking at more abundant functions in Atlantic metagenomes are central carbon metabolism functions. Indeed a few enzymes and reactions from the glycolysis are found more abundant in the Atlantic such as: "glycerol-3-phosphate (G3P) dehydrogenase [NAD+] activity" and "phosphoglycerate kinase activity". In addition "pentose-phosphate shunt, non-oxidative branch" (Supplementary Figure 10b) which is an alternative to glycolysis is also a more abundant GO terms in Atlantic metagenomes. The pentose phosphate shunt oxidative branch is known to reduce oxidative stress and support lipobiogenesis (*14*) while the non-oxidative branch ends to the beginning and middle of the glycolysis (generating a G3P and a fructose 6 phosphate molecule) and generates 5-carbon sugars used for the synthesis of nucleotides, nucleic acids and amino acids (*15*, *16*). The term "Fatty acid biosynthetic process" and other functions linked to fatty acid biosynthesis like "acetyl-CoA carboxylase complex" (which is biotin dependent for which 2 GO terms are found more abundant) are also found more abundant in Atlantic metagenomes although these terms are dominated by bacterial genes (Supplementary Figure 16c). Overall, the high abundance of central carbon metabolism function in the Atlantic might reflect the overall prevalence of heterotrophy in this ocean as expected (*17*) especially as it was mostly sampled in winter. In addition, "nitrogenous compound fermentation" (Supplementary Figure 16d) is also found more abundant in the Atlantic. Interestingly, their abundance is dominated by bacteria followed by unclassified eukaryotes (Supplementary Figure 10d) which might reflect the presence of unknown eukaryotes capable of fermentation. The high abundance of these terms as well as "anaerobic amino acid catabolic process" might reflect a higher abundance of organisms capable of living in anaerobic condition in the Atlantic (Figure 1b). Although Tara samples are from the surface, the presence of such organisms might be due to the winter sampling and might come from the deep mixing of oceanic waters in winter. More intriguingly, five GO terms linked to photosynthesis are found to be more abundant in Atlantic metagenomes and three of which are also more abundant in the metatranscriptomes: "photosynthetic electron transport chain", "photosynthetic electron transport in photosystem II", "electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity" (Supplementary Figure 10e), "photosynthesis light reaction" and "light-harvesting complex" (the two latter more abundant only in metagenomes). The high abundance of these functions in the Atlantic in winter is interesting as it suggests that algae are adapting to low light levels and saving energy. Indeed, the high abundance of photoreceptors (light reaction)

suggests the need to harvest light in a more efficient way. Second the cyclic electron transport path of photosynthesis is known to be used to balance the energy (*18*) and protect photosystems from photoinhibition (*19*). The cyclic electron flow (CEF) is also known to be involved in various acclimation responses in plants (*20*). We verify that these functions correspond to only a few PFAMs that are mostly more abundant (or not) in the Atlantic metagenomes and metatranscriptomes using ALDEx2 package (*13*, *21*) (Supplementary Figure 18): PF00124 (Photosynthetic reaction centre Photo_RC), PF00421 (PSII, photosystem II), PF00556 (Light Harvesting Complex LHC), PF03967 (Photosynthetic reaction centre, H-chain N-terminal region PRCH), and PF11623 (NADH dehydrogenase subunit S NdhS). In fact most other PFAMs associated to photosynthesis appear overall more abundant in Arctic metagenomes which is expected although it depends on taxons (Supplementary Figure 18).

Among other functions that we find more abundant in Atlantic metagenomes or metatranscriptomes, we can cite functions linked to mobility of genomic elements such as "transposase activity" and "transposition, DNA-mediated" (more abundant in metatranscriptomes, dominated by unknown followed by bacteria) or bacteria specific functions (Supplementary Table 1).

Supplementary Discussion 6 | Details of enriched function in persisting species.

In MGT 79 (Bathycoccus, Figure 5a) a unigene involved in thiamine (Vitamin B1) diphosphate biosynthetic process is negatively correlated to temperature, known to be important in early stage response to cold and osmotic stress in *Arabidopsis* (*22*) and in the poplar (*23*). In MGT 195, a pennate diatom (Figure 5b), the expression profiles of a gene involved in tocopherol cyclase activity (also in MGT 42 Aureococcus, Supplementary Figure 16b) and one in allantoin catabolic process are significantly negatively correlated to temperature. Tocopherol cyclase activity is known to be important in cold (*24*) but also salt (*25*) and high light acclimation (*26*) in plants. Allantoin on the other hand is a metabolic intermediate of purine catabolism that often accumulates in stressed plants (*27*) and was reported to enhance cold stress tolerance in rice grains (*28*). In MGT 4 (Supplementary Figure 16c), superoxide dismutase (SOD) (also in MGT 439, Figure 5a), heat shock protein binding and glutamyl-tRNA reductase activity are reported to be enriched GO terms in negatively correlated unigenes. SOD is known to increase stress tolerance in potato through mitochondrial activity (*29*) but also in wheat (*30*) and barley (*31*). Some heat shock proteins are known to respond to cold stress in *Synechococcus* (*32*), some plants (*33*) but also fishes (*34*). Finally, Glutamyl-tRNA reductase, the key enzyme for 5-aminolevulinic acid (ALA) synthesis, has been shown to be highly expressed in cold stress response and to remain at relatively high levels after acclimation although it varies across plants (*35*). In MGT 439 (Figure

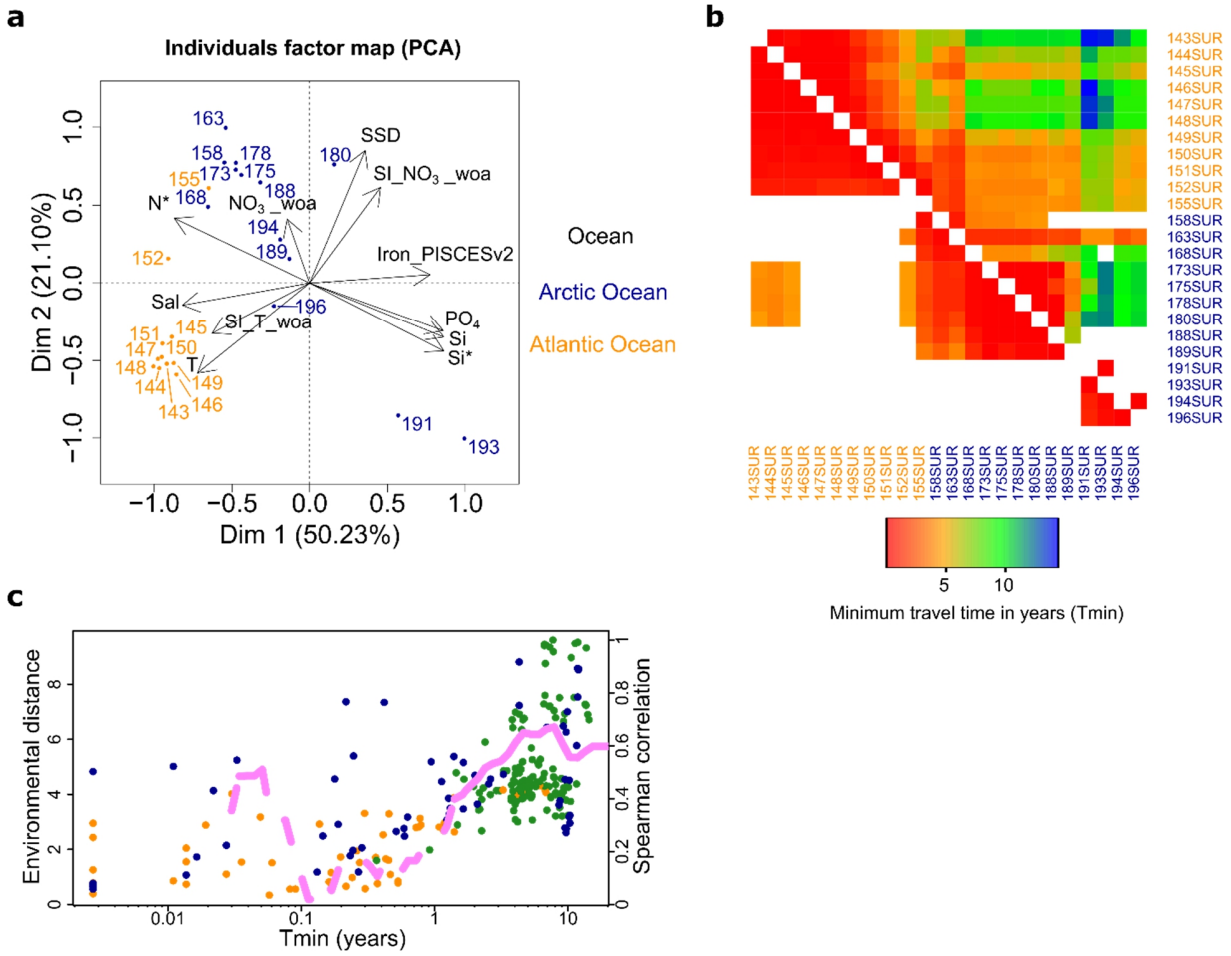
5a), trehalose biosynthetic process is also an enriched GO term in negatively correlated unigenes linked to cold acclimation (*36*). Finally, phospholipase A2 known to have a role in cold acclimation in wheat (*37*) is found to be enriched in MGT 76 (Supplementary Figure 16a). Zooming out at the higher taxonomy levels from Figure 2, other or same functions associated to cold stress are also found enriched such as tocopherol cyclase activity (Pelagophyceae) (*24–26*), polyketide metabolic process (Dinophyceae) (*38*, *39*) or pantothenate biosynthesis (Pelagophyceae) (*40*) (Supplementary Figure 17). Overall, we find a wide variety of functions whose expressions are negatively correlated to temperature and known to be important in cold acclimation both in terrestrial plants and aquatic organisms highlighting the quasi universality and diversity of these functions for cold acclimation in the *Plantae* kingdom.

In addition to these functions directly linked to cold acclimation, many enriched functions are linked to gene expression processes and regulation, energy consumption and mitochondria related processes, division/DNA replication, motility, nucleotide biosynthesis, protein folding and DNA repair (Figure 5). In particular, GO terms linked to the triggering and control of gene expression such as 'transcription factor TFIIA complex' (MGT 79), 'transcription by RNA polyremase II' (MGT 76), 'transcription corepressor activity' (MGT 76) but also to post transcriptional modification and protein biosynthesis such as 'histone acetyltransferase' (MGT 79 and 76), 'translation initiation factor activity' (MGT 4), 'poly(A)+ mRNA export from nucleus' (MGT 195), 'cellular amino acid biosynthetic process' (MGT 4), are found to be enriched. This suggests that the entrance in the Arctic Ocean triggers a tightly genetically controlled transcriptional burst that allows the persistence of some organisms (mainly algae here) to the new physico-chemical conditions they encounter. Indeed, enriched functions include all levels of genetic expression control and processes including signal transduction (phosphorylation, GTPase, phosphatidylinositol phosphorylation…), transcription regulation, RNAs/ribosomes management, amino-acid biosynthesis and post-translation modifications (Figure 5, Supplementary Figure 16-17).

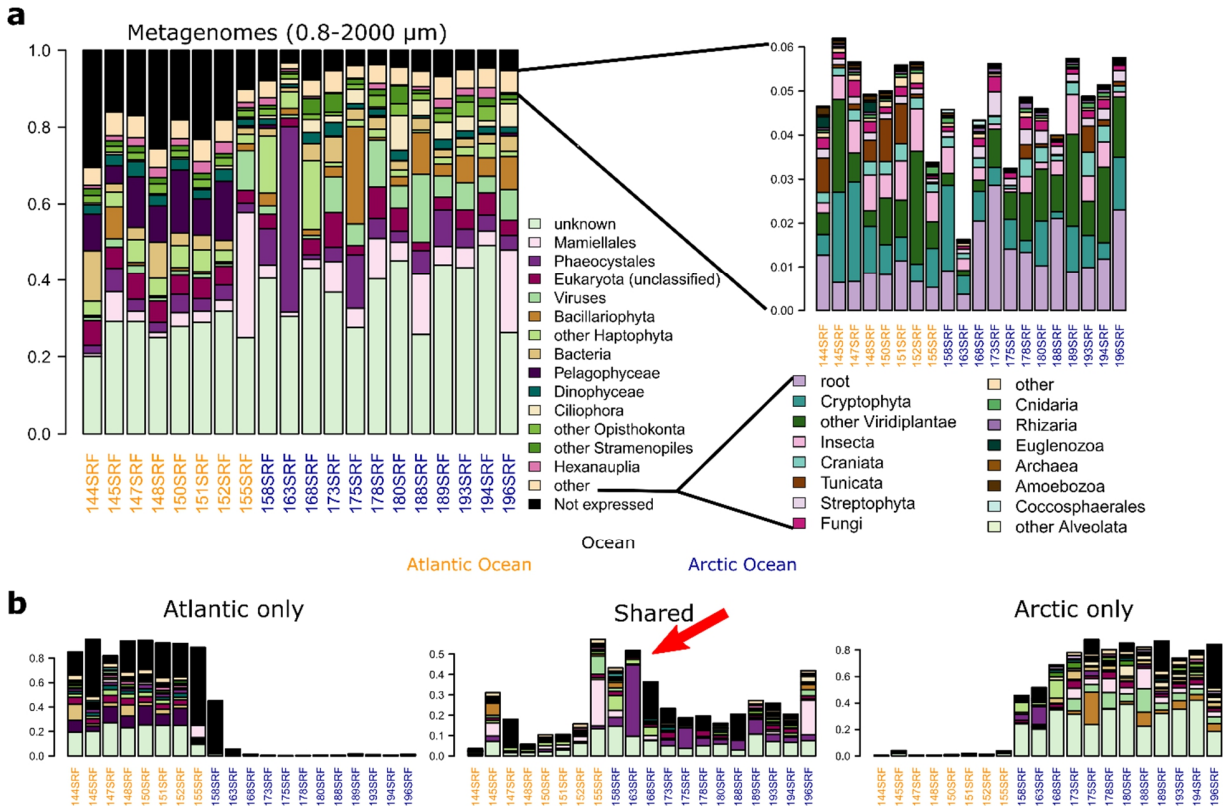
Overall many functions linked to energy and respiratory processes are found enriched in the 6 MGTs such as 'respiratory electron transport chain complex' and 'proton transporting ATP synthase activity, rotational mechanism'. This effect might be due to an increase in metabolic activities due to the spring and summer season (*2*). These seasons are known to lead to loss of synchrony of cell division of microalgae (*41*, *42*) in agreement with enriched functions in link with cell division and DNA replication.

Critically, many enriched GO terms are associated to DNA replication and consequently cell division such as 'regulation of DNA replication' (MGT 79 and Phaeocystales) and 'DNA replication

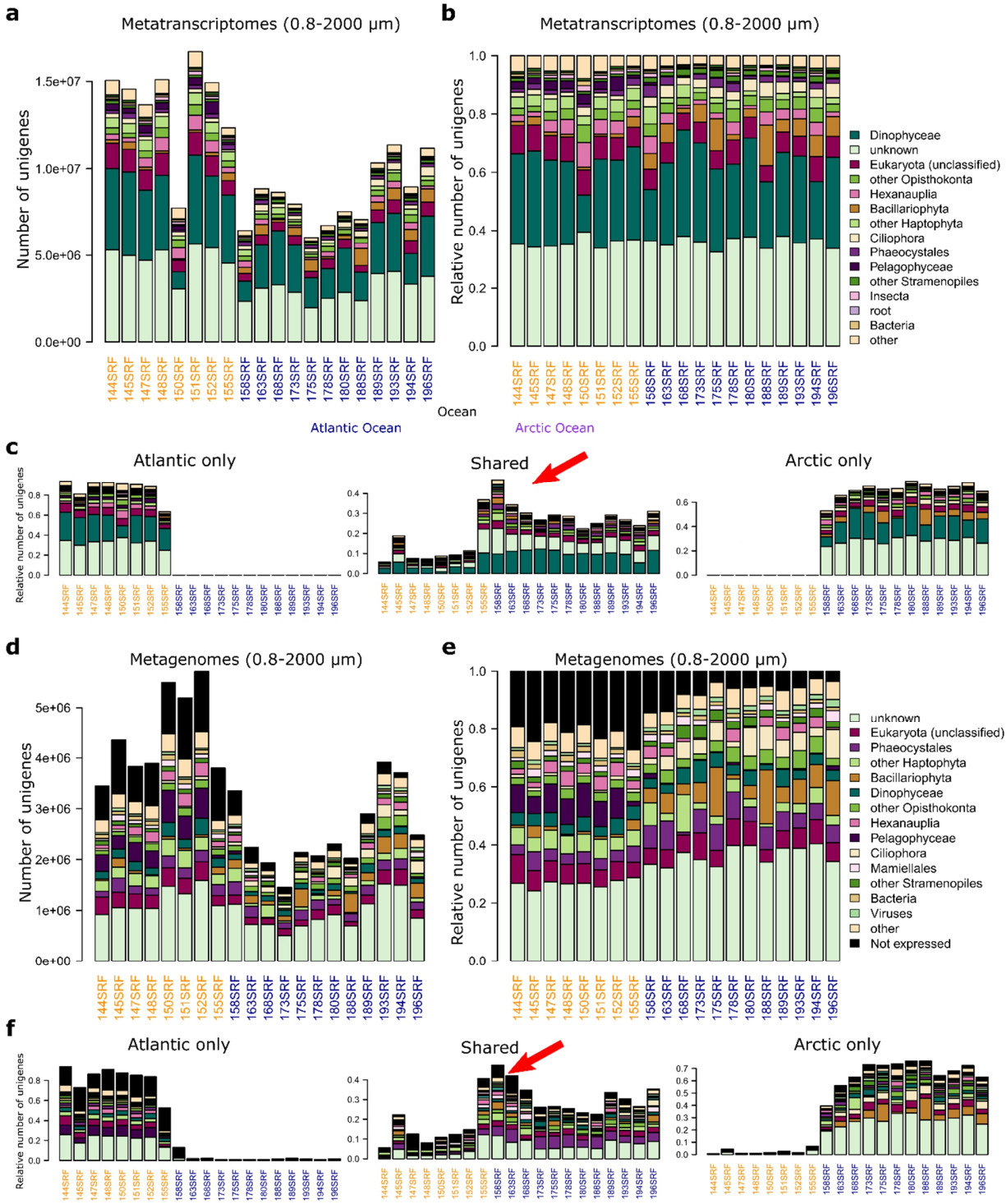
initiation' (MGT 76 and Mamiellales) or Ndc80 complex (Bacillariophyta). This might reflect the diminution in synchrony of cell division in link with the circadian clock (*41*, *42*) as days are getting longer as cells enter the Arctic. In addition, it might show the capacity of persisting organism to keep on dividing as they enter in the polar waters.



Supplementary Figure 1 | Environmental differentiation of the NAO and AO basins in link with their Lagrangian connectivity by currents. (**a**) Principal Component Analysis (PCA) of the environmental parameters characterizing the stations. (**b**) Lagrangian minimum travel time connectivity matrix. The upper matrix represents the minimum travel time by currents from row stations to column stations and vice versa. (**c**) Environmental distance in function of Lagrangian minimum travel time (Tmin) and cumulative Spearman correlation (brown line).

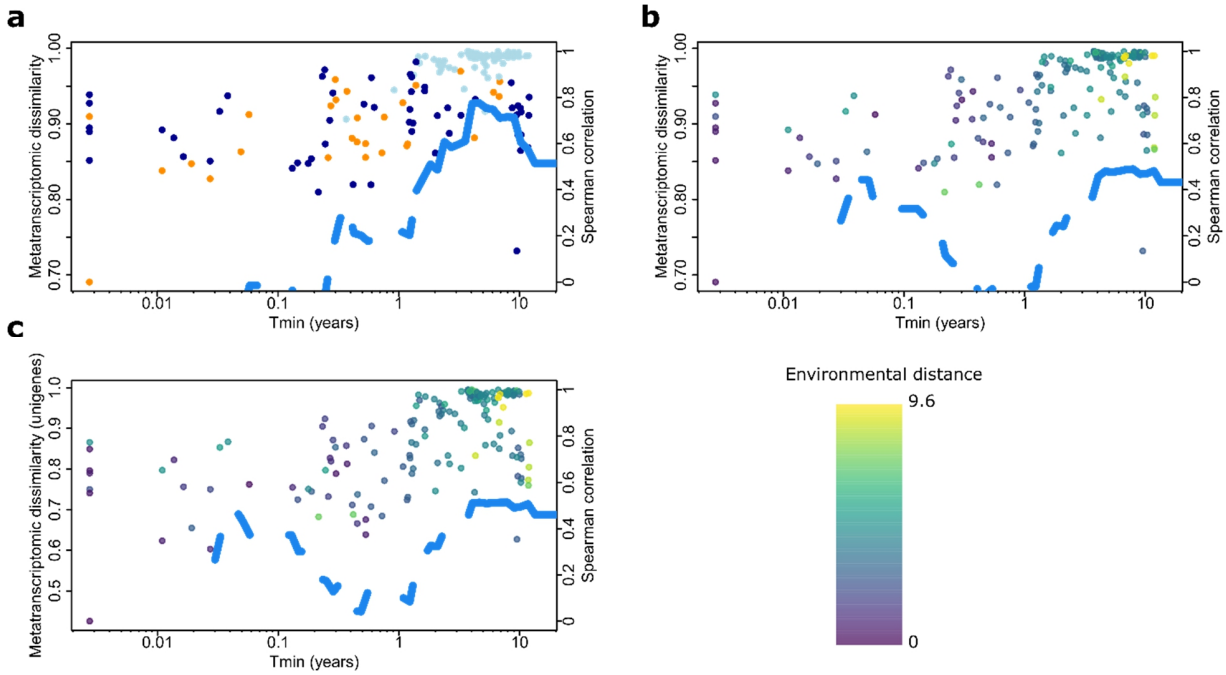


Supplementary Figure 2 | Taxonomic composition of surface ocean plankton metagenomes of size fraction 0.8-2000 µm. (**a**) Left: global metagenomic taxonomic composition based on expressed unigenes relative abundances of major plankton taxons (taxonomic level can vary) in the North Atlantic and Arctic Ocean. Right: Zoom on the 16 least abundant taxons (other). (**b**) Decomposition between basins specific unigenes and shared unigenes between the two basins (classified based on presence or absence of expression). A clear transition enriched in shared genes appears in station 153, 158 and 163.

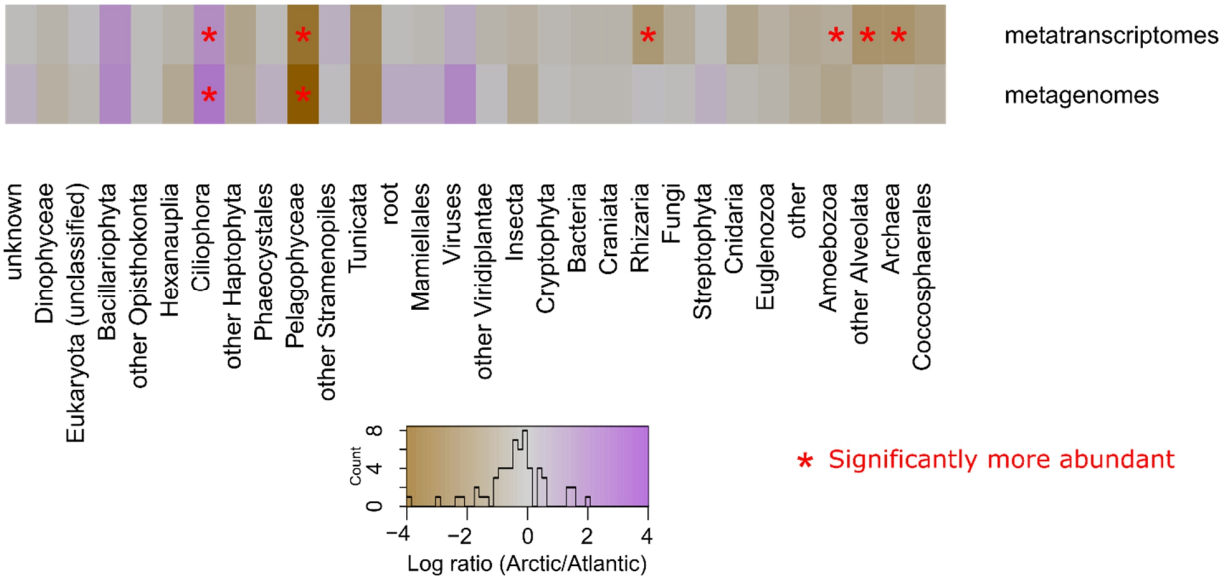


Supplementary Figure 3 | Taxonomic composition in unigenes of surface ocean plankton metatranscriptomes and metagenomes of size fraction 0.8-2000 µm. (**a,d**) Metatranscriptomic/metagenomic taxonomic composition in number of unigenes of North Atlantic and Arctic Ocean sampling stations. (**b, e**) Metatranscriptomic/metagenomic taxonomic

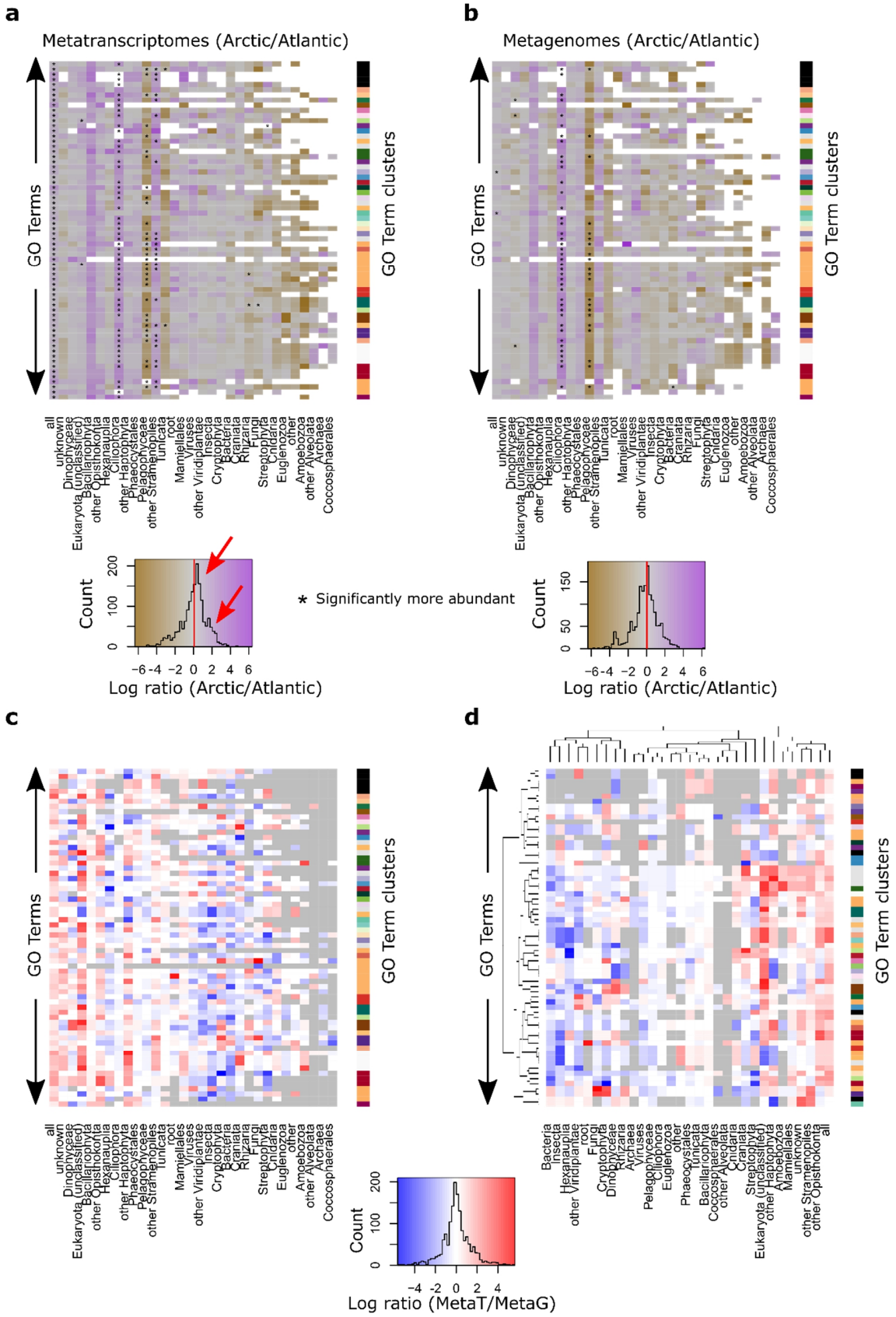
composition in relative number of unigenes of North Atlantic and Arctic Ocean sampling stations. (**c, f**) Decomposition of metatranscriptomes/metagenomes in relative number of unigenes between basins specific unigenes and shared unigenes between the two basins (classified based on presence or absence of expression).



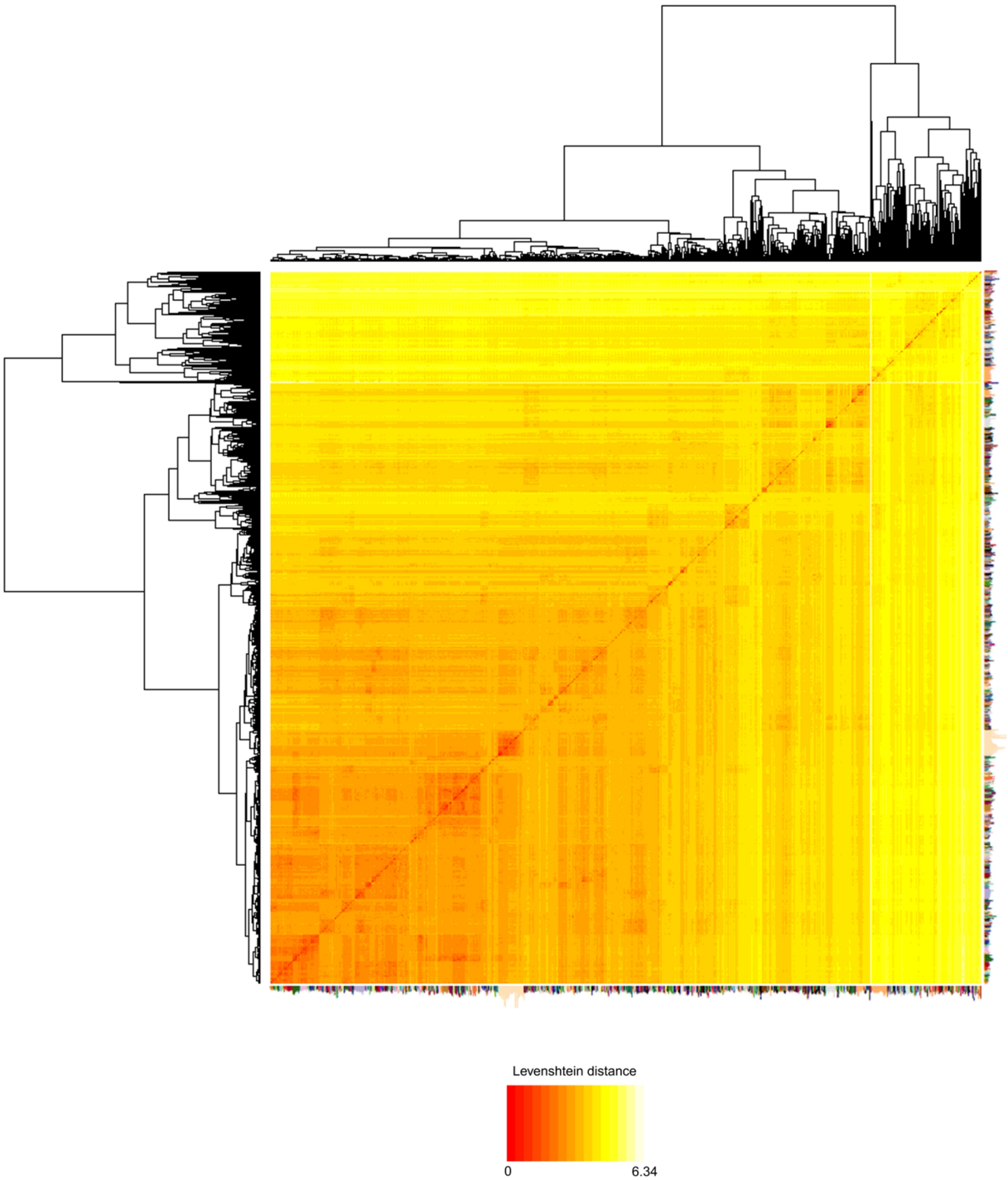
Supplementary Figure 4 | Metatranscriptomic β diversity in link with the Lagrangian and environmental connectivity of the NAO and AO basins. (**a**) Metatranscriptomic β diversity (SimkaMin (*43*)) in function of Lagrangian connectivity between pairs of Tara station (minimum transport time) and cumulative spearman correlation coefficient between the two metrics (blue line). Metatranscriptomic β diversity based on (**b**) SimkaMin (*43*) and (**c**) unigenes metatranscriptomic abundance-based Bray-Curtis (*44*) in function of Lagrangian connectivity between pairs of Tara station and cumulative spearman correlation coefficient between the environmental distance (points color) and β diversity (blue line).



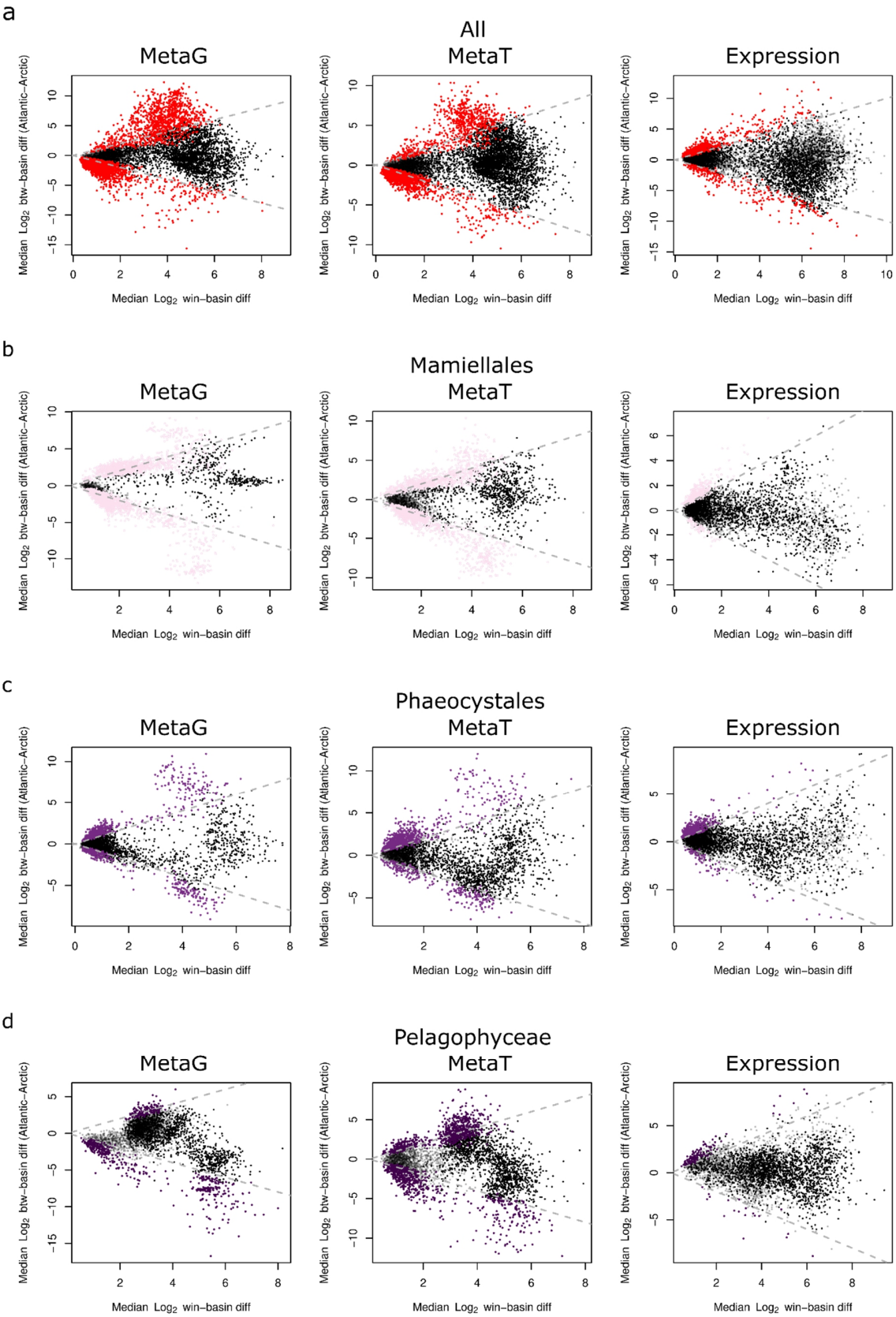
Supplementary Figure 5 | Median $\log_e$-fold change between Arctic Ocean and North Atlantic Ocean metatranscriptomes and metagenomes relative abundances of main plankton groups. The heatmap represents for each taxon the difference between their median logarithms of relative abundance between both basins (Arctic-Atlantic). Significantly more abundant taxons (Wilcoxon test, Hommel correction $p<0.05$) either in metatranscriptomes (top row) or metagenomes (bottom row) are indicated with a red star. Here we use the logarithms of abundances and not the centered log ratio as the former might bias some abundance values as it is centered on zero in the case of skewed distributions (which is the case here).

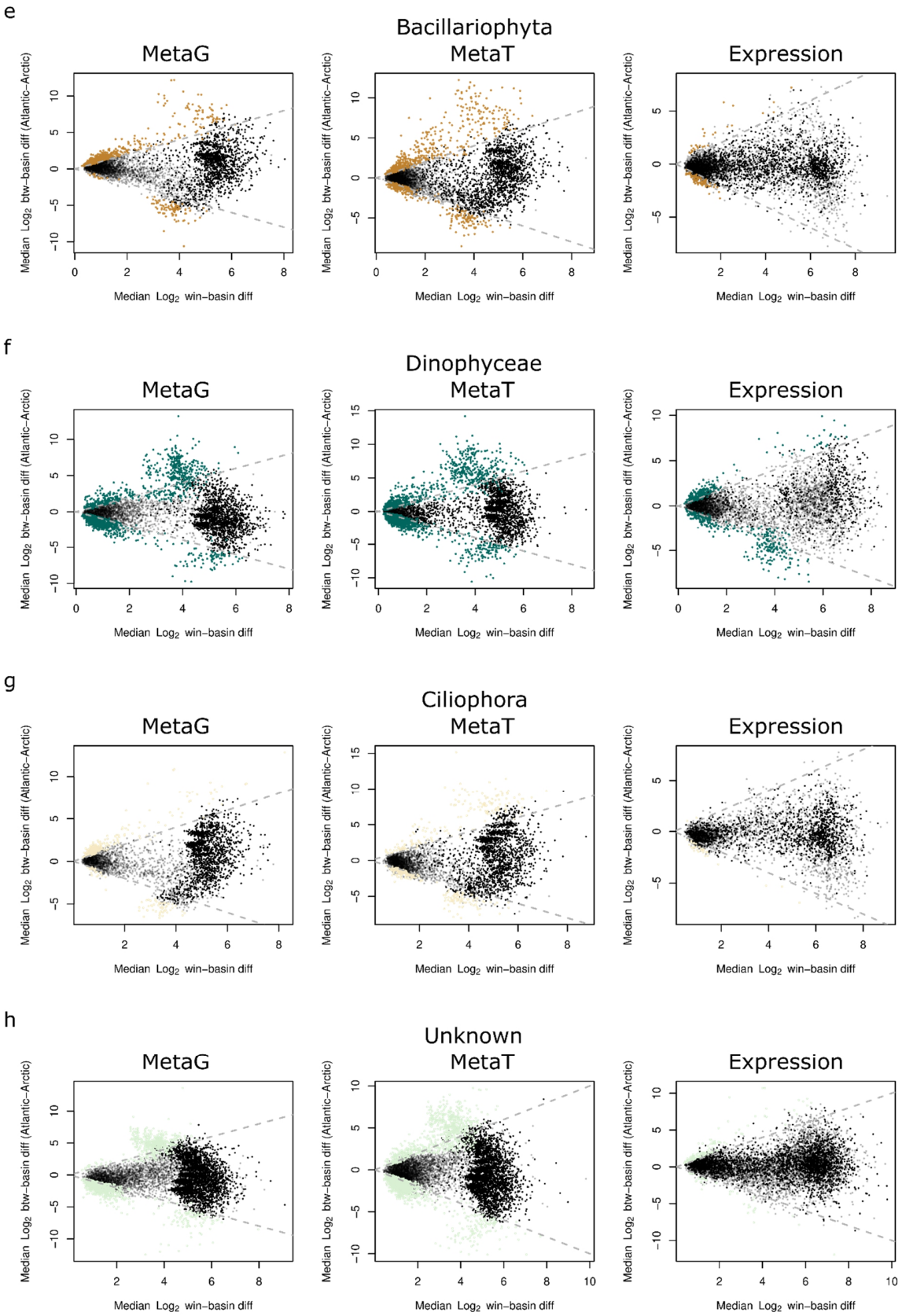

a
Metatranscriptomes (Arctic/Atlantic)
GO Terms
GO Term clusters
b
Metagenomes (Arctic/Atlantic)
GO Terms
GO Term clusters
Count
Log ratio (Arctic/Atlantic)
* Significantly more abundant
Count
Log ratio (Arctic/Atlantic)
c
GO Terms
GO Term clusters
Count
Log ratio (MetaT/MetaG)
d
GO Terms
GO Term clusters

Supplementary Figure 6 | Median log_e-fold change in metatranscriptomic and metagenomic relative abundances between Arctic Ocean and North Atlantic Ocean of significantly more abundant Gene Ontology (GO) terms in Arctic metatranscriptomes for all taxa and individual taxa and their ratios. (**a,b**) The heatmaps represent for all taxons (left columns of **a-c**) and each taxon in decreasing order of abundance (other columns) the difference between the median logarithms of relative abundance between both basins (Arctic-Atlantic) for Gene Ontology terms significantly more abundant in the Arctic metatranscriptomes. This ratio is plotted either for metatranscriptomes (**a**) or metagenomes (**b**). Significantly more abundant GO terms (Wilcoxon test, Hommel correction $p<0.05$) are indicated with a black star. GO terms (rows) are colored and ordered according to the GO cluster they belong to, reflecting close functional annotation (see Supplementary Figure 7). The first 6 GO terms (6 first rows) don't belong to any cluster. (**c**) Log_e ratio of metatranscriptomes (**a**) and metagenomes (**b**) relative abundance ratio matrices and (**d**) hierarchical clustering of this matrix.



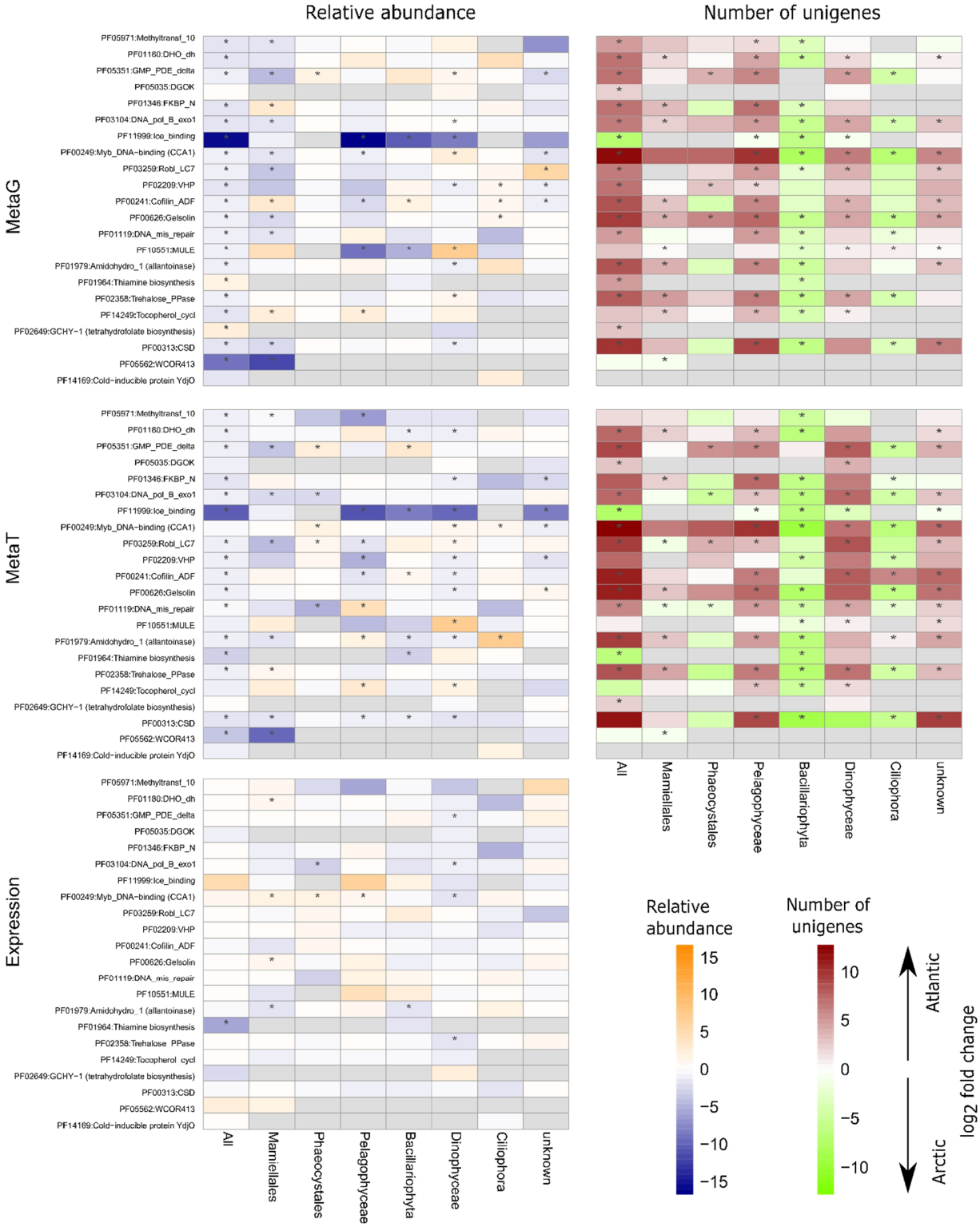
Supplementary Figure 7 | Clustering of the Gene Ontology (GO) terms based on the Levenshtein distance. Each GO term present in our unigene database was included in the clustering. The hierarchy of each GO term is converted in a string and pairwise Levenshtein distance(*45*) are computed. The resulting matrix is clustered using UPGMA(*46*) clustering (*Methods*). GO terms (lines and columns) are colored according to the cluster they belong to.


a
All
MetaG
MetaT
Expression
b
Mamiellales
MetaG
MetaT
Expression
c
Phaeocystales
MetaG
MetaT
Expression
d
Pelagophyceae
MetaG
MetaT
Expression
Median Log2 btw–basin diff (Atlantic–Arctic)
Median Log2 win–basin diff



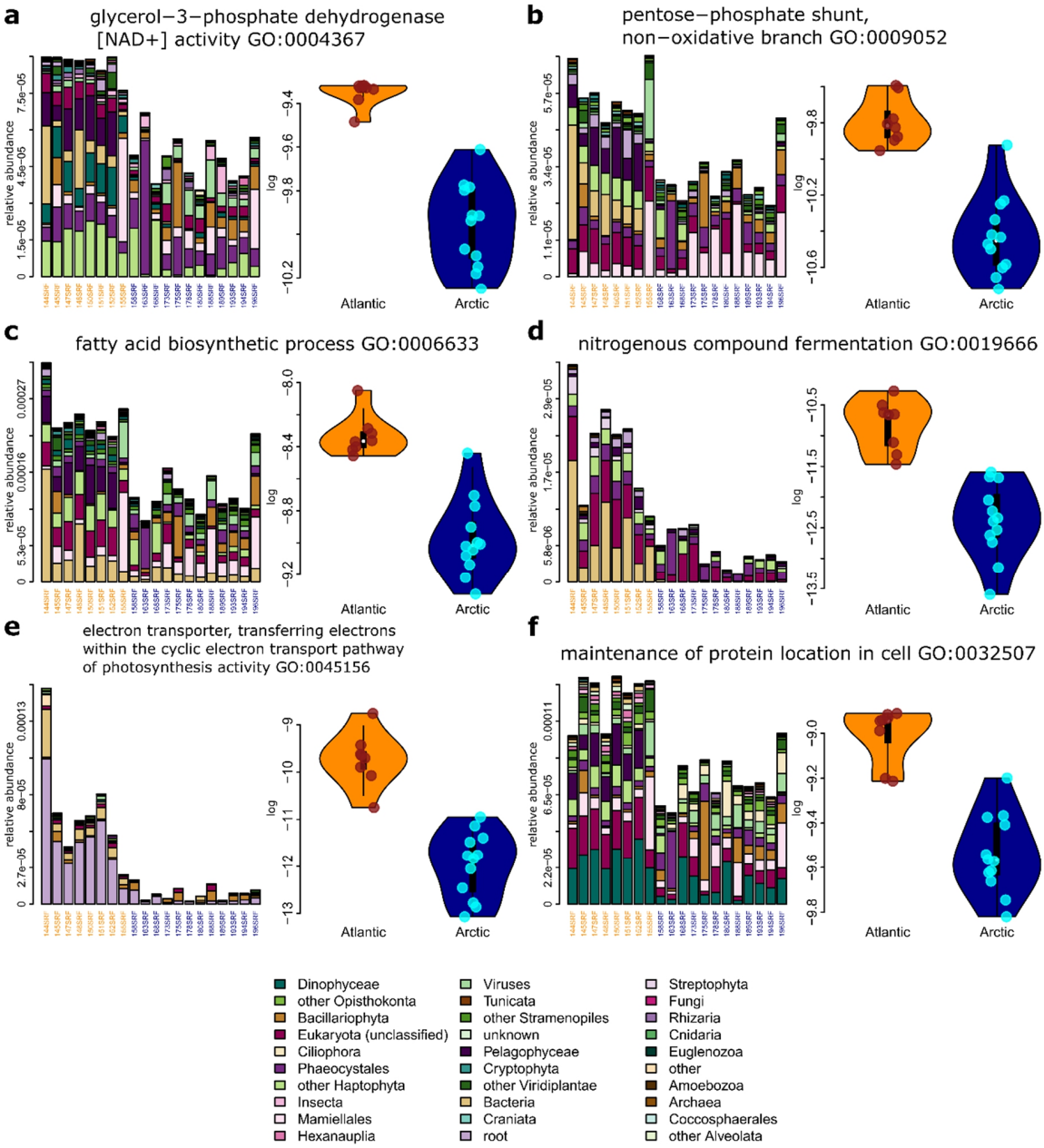
Suplementary Fig. 8 | Differential metagenomic, metatranscriptomic abundances and

expression of PFAM domains between the North Atlantic basin and Arctic basin considering all taxons and specific taxons. We use the ALDEx2 package (*13, 21*) to compute differentially abundant functions (PFAMs) in either metagenomes (left columns), metatranscriptomes (middle column) and expression (metatranscriptomic abundance normalized by metagenomic abundance, right column) between the North Atlantic Ocean and the Arctic Ocean. For each PFAM the total. For each panel, the median $\log_2$ difference between basins (equation 7) is plotted in function of the median $\log_2$ dispersion (for each PFAM) (equation 8) and significantly more abundant PFAMs are colored (bottom part of the plot in the Arctic and top part of the plot in the Atlantic). The dotted lines indicate the limits at which inter-basin dispersion is superior to the intra-basin dispersion.

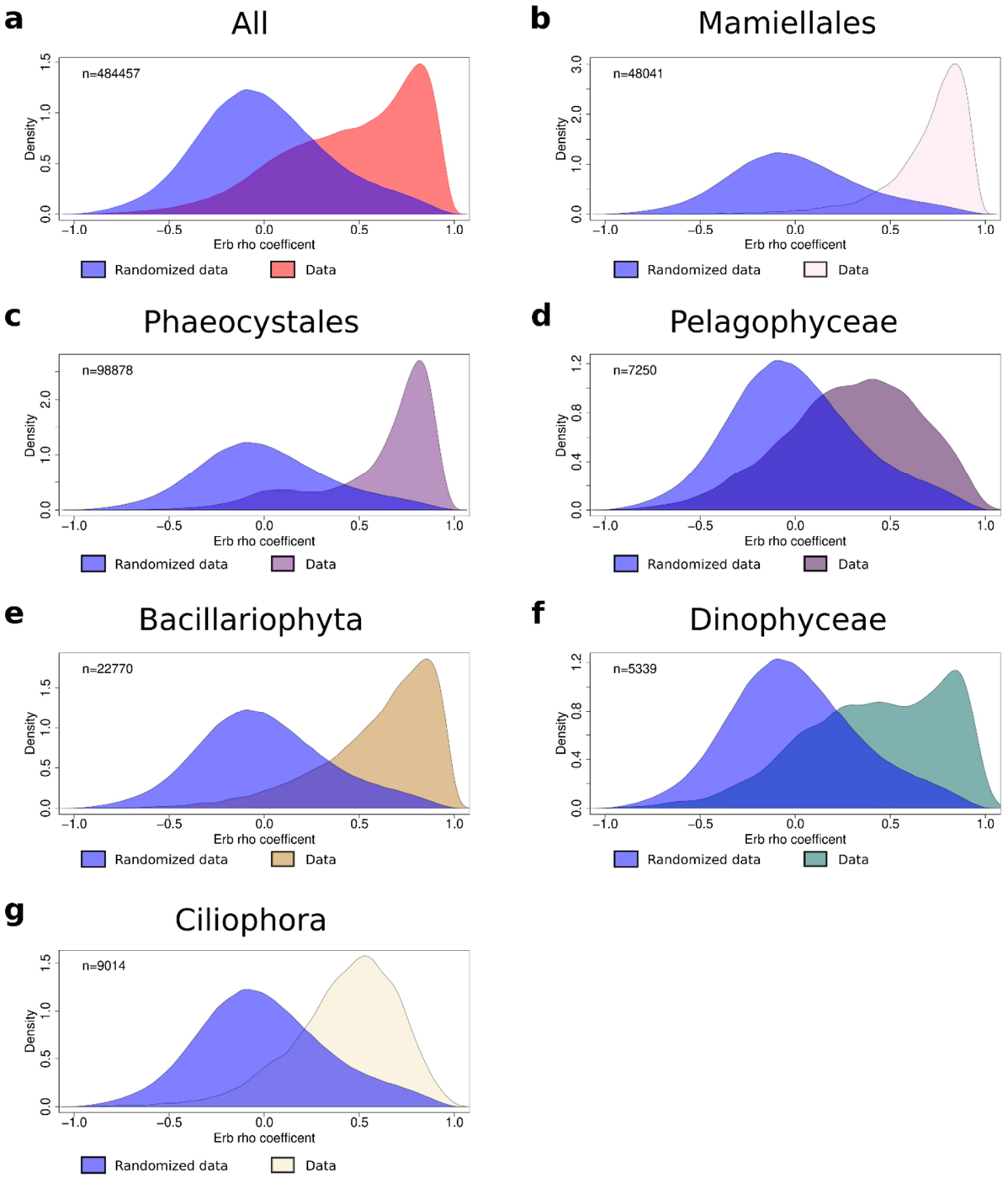


Supplementary Figure 9 | Differential abundance in metatranscriptomes and metagenomes and differential expression of PFAMs associated to temperature response. Right column: Median log2-fold change between Atlantic and Arctic metagenomes, metatranscriptomes

(equation 6) and expression (metatranscriptomes normalized by metagenomes) of PFAMs associated to temperature response based on estimated centered log ratios from ALDEx2 package for all taxons, Mamiellales, Phaeocystales, Pelagophyceae, Bacillariophyta, Dinophyceae, Ciliophora and Unknown. Significant differences are highlighted by stars. Left column: Median log2-fold change of the number of unigenes between Atlantic and Arctic metagenomes and metatranscriptomes for aforementioned taxa.

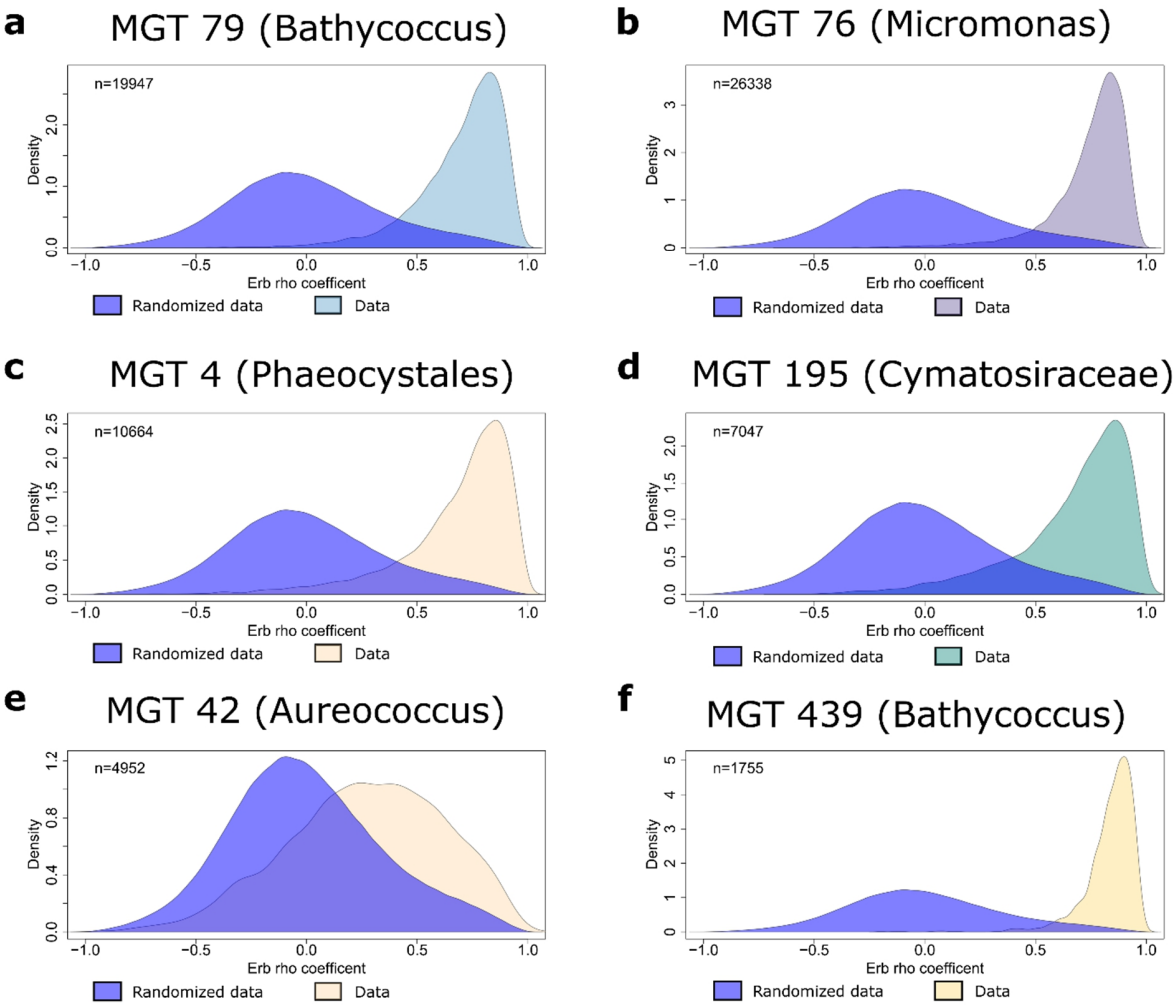


Supplementary Figure 10 | Examples of Gene Ontology (GO) terms more abundant in the Arctic metagenomes. Each panel (**a-f**) represents (left) the sum of the metagenomic relative abundance of unigenes associated to the abovementioned GO term across considered taxons and (right) the distribution of this relative abundance in the Atlantic (dark blue) versus Arctic basin (purple). Each of these GO terms is significantly more abundant in the Arctic basin.

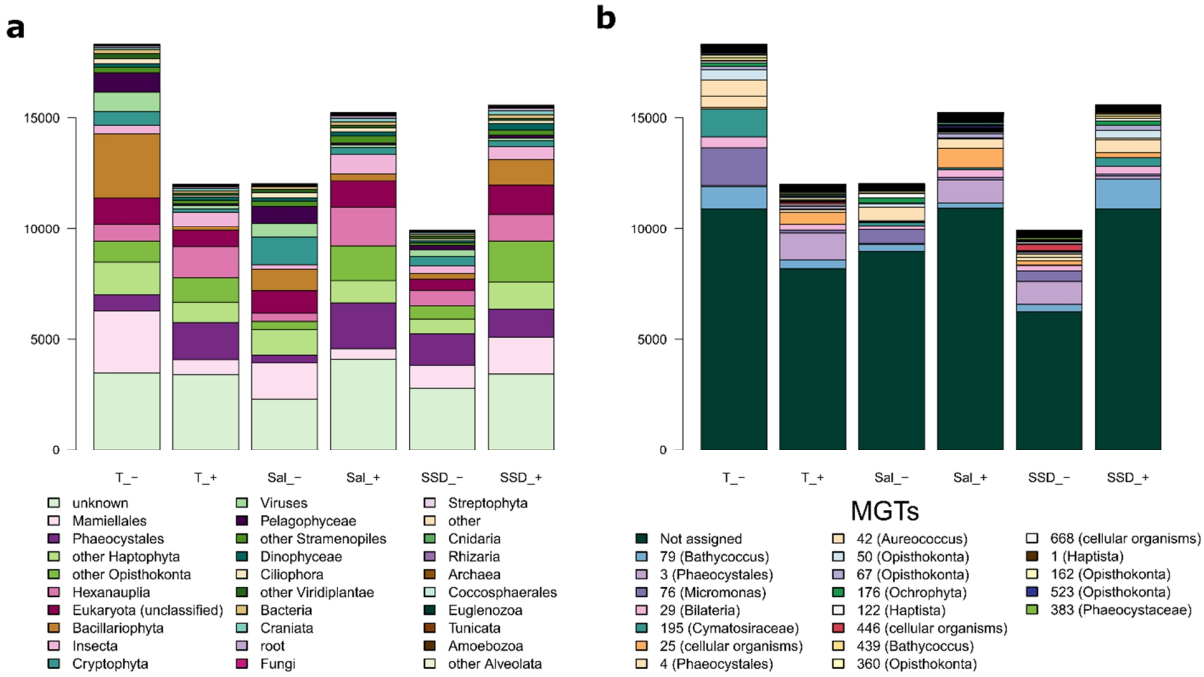


Supplementary Figure 11 | Distribution of Erb rho(*47*) **correlation index between centered log ratios of randomized or real metatranscriptomic and metagenomic relative abundance of unigenes present in at least 5 stations.** The correlation index distribution is represented for all taxons (**a**) and individual taxon (**b-g**). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. Importantly, metatranscriptomic and

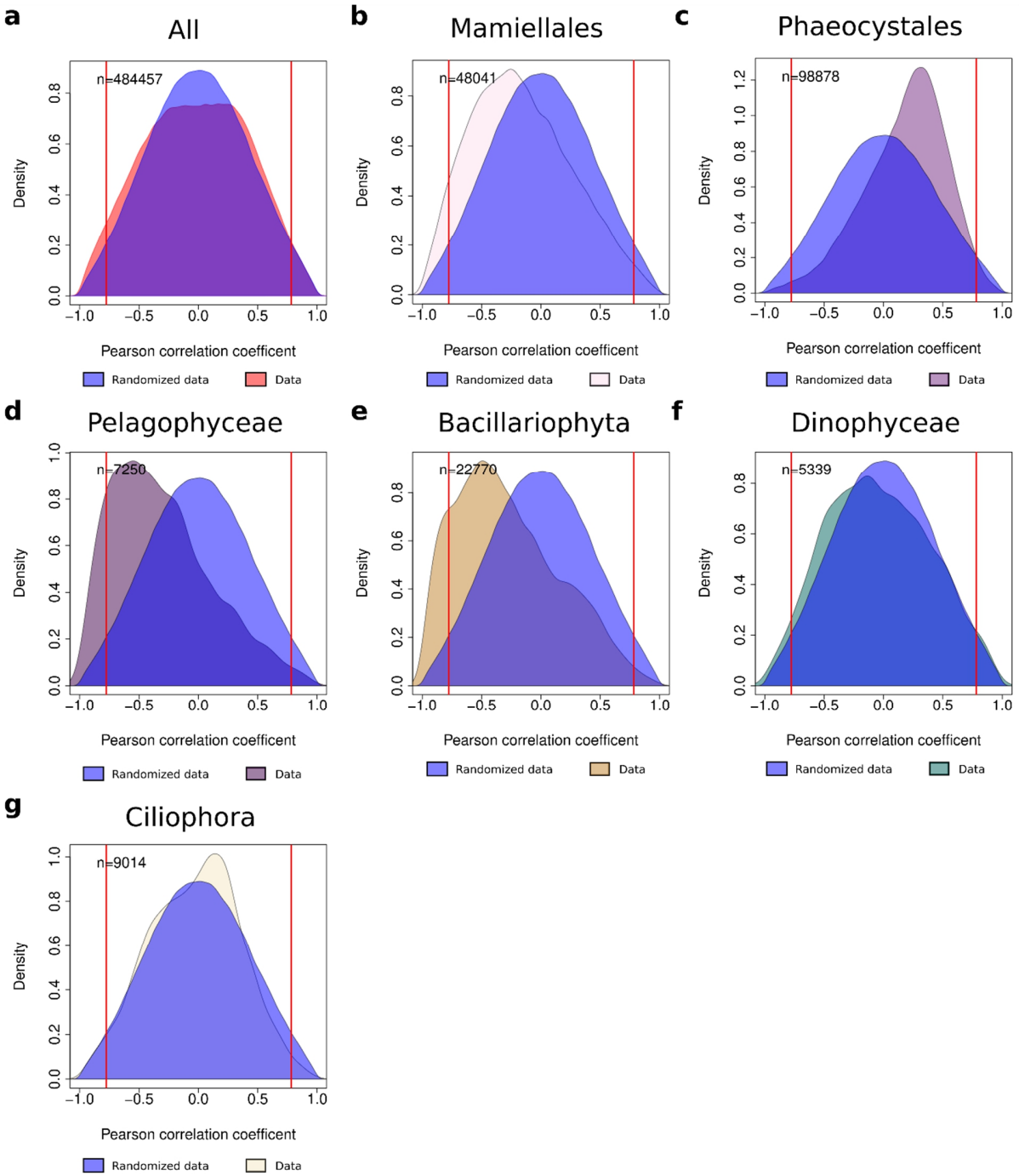
metagenomic centered log ratios (clr) are relatively well correlated although more or less strongly depending on the taxon. These good correlations justify the use of the metagenomic clr to normalize the metatranscriptomic clr to define unigene expression (Figure 4, 5, Supplementary Figure 12-13-14).



Supplementary Figure 12 | Distribution of Erb rho (*47*) **correlation index between centered log ratios of randomized or real metatranscriptomic and metagenomic relative abundance of unigenes present in at least 5 stations of 6 algal Metagenomics-based Transcriptomes (MGTs).** The correlation index distribution is plotted for 6 MGTs (**a-f**). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. Importantly, metatranscriptomic and metagenomic centered log ratios (clr) are relatively well correlated although more or less strongly depending on the taxon. These good correlations justify the use of the metagenomic clr to normalize the metatranscriptomic clr to define unigene expression (Figure 4, 5, Supplementary Figure 12-13-14).

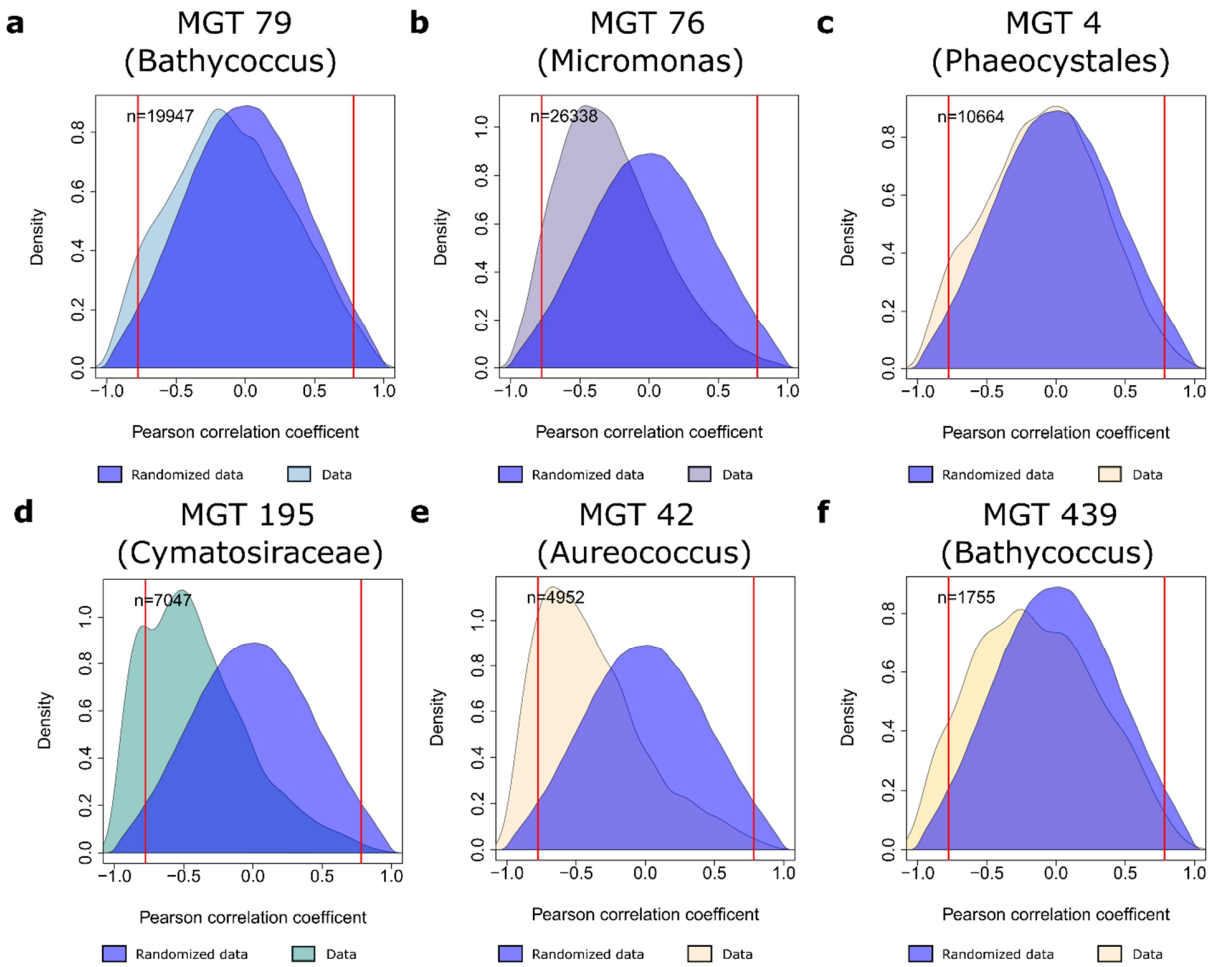


Supplementary Figure 13 | Taxonomic decomposition of unigenes significantly correlated with physical environmental parameters, either negatively or positively. (**a**) Taxonomic and (**b**) Metagenomics-based Transcriptomes (MGTs) annotation of unigenes. Temperature is the parameter for which the highest number of unigenes expression profile is correlated to temperature. Salinity (Sal) and day length (SSD) are highly correlated to temperature so we only analyze unigenes functions which are correlated to temperature (Figure 4-5).

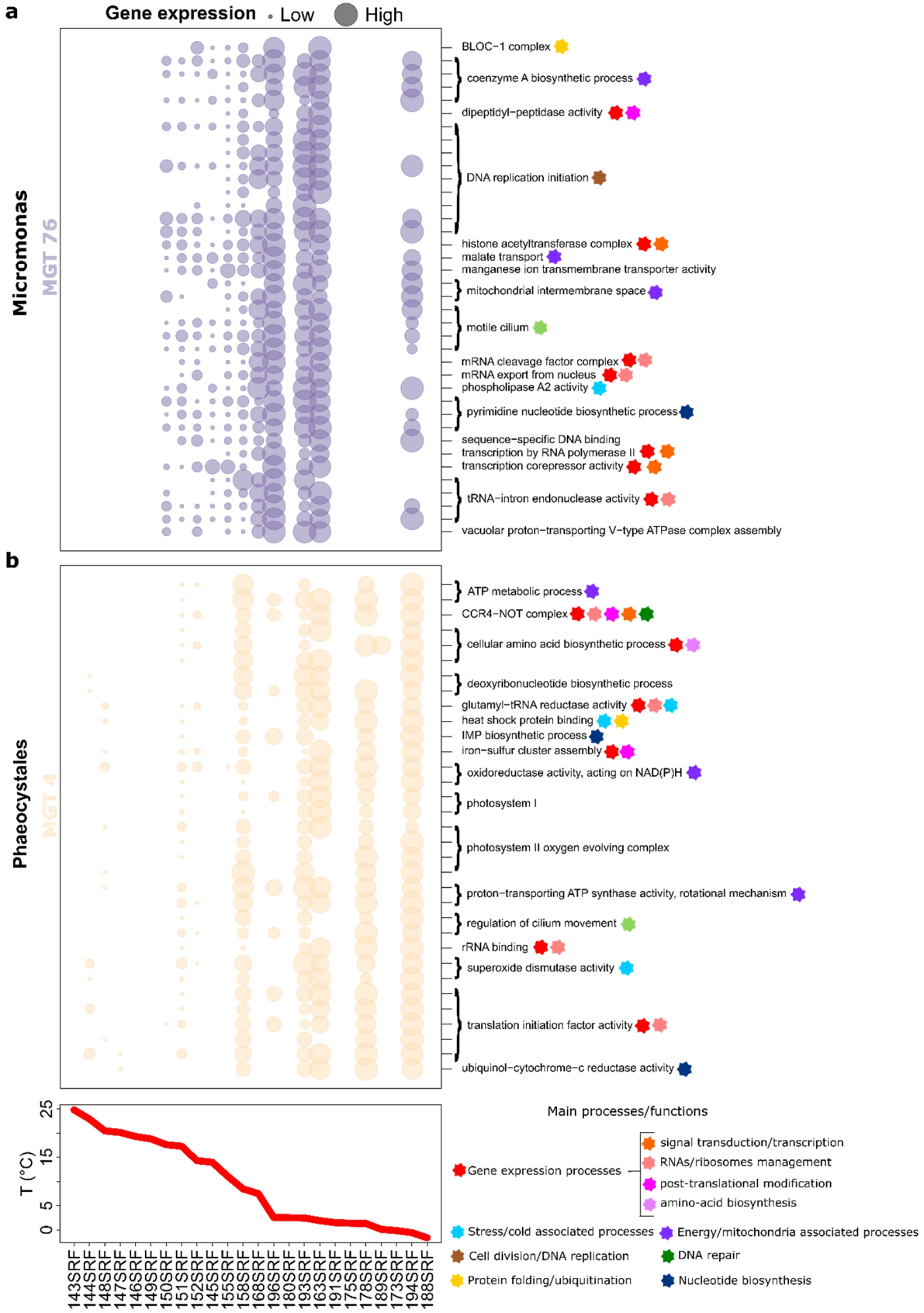


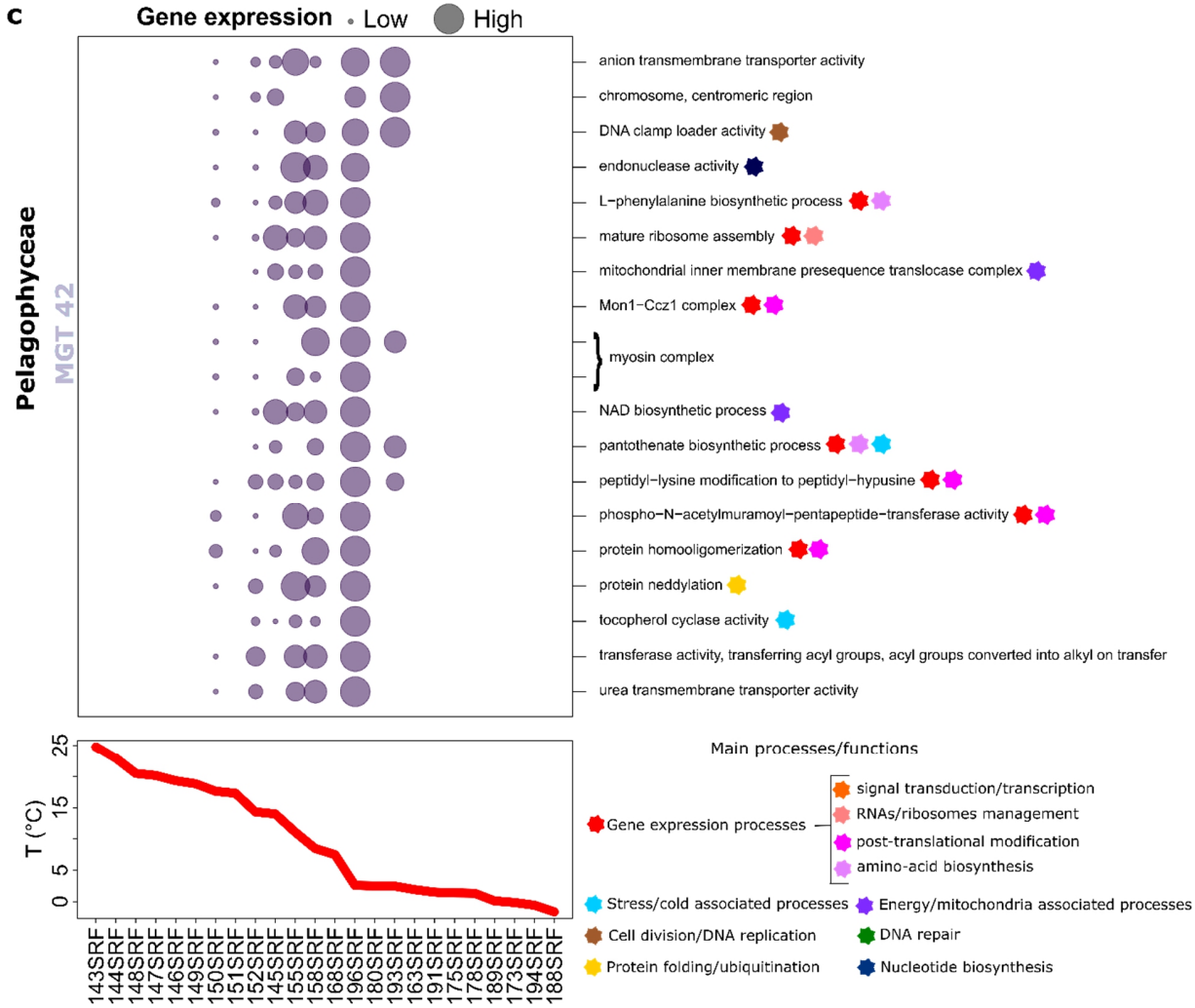
Supplementary Figure 14 | Distribution of Pearson correlation coefficient between temperature and randomized or real expression values of unigenes present in at least 5 stations. The Pearson correlation distribution is represented for all taxons (**a**) and individual taxon (**b-g**). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. The expression value of a unigene in a given station is calculated as the

difference between the centered log ratios of its metatranscriptomic relative abundance and metagenomic abundance. For some taxon correlation distribution are clearly distinguishable from the randomized data distribution indicating a clear response to temperature. It is particularly the case for Mamiellales, Pelagophyceae and Bacillariophyta that display skewed distribution enriched in unigenes whose expression profile is negatively correlated with temperature.

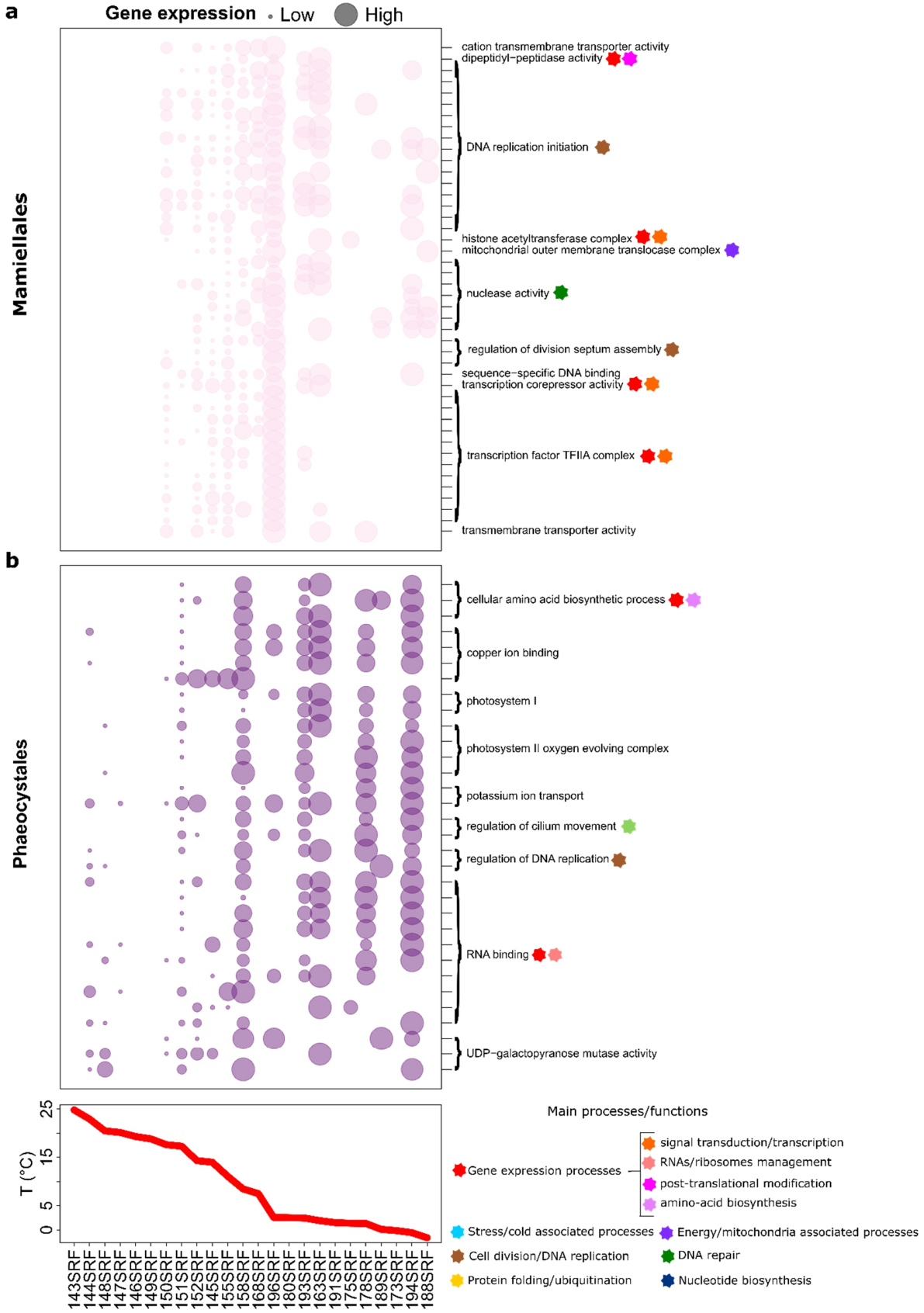


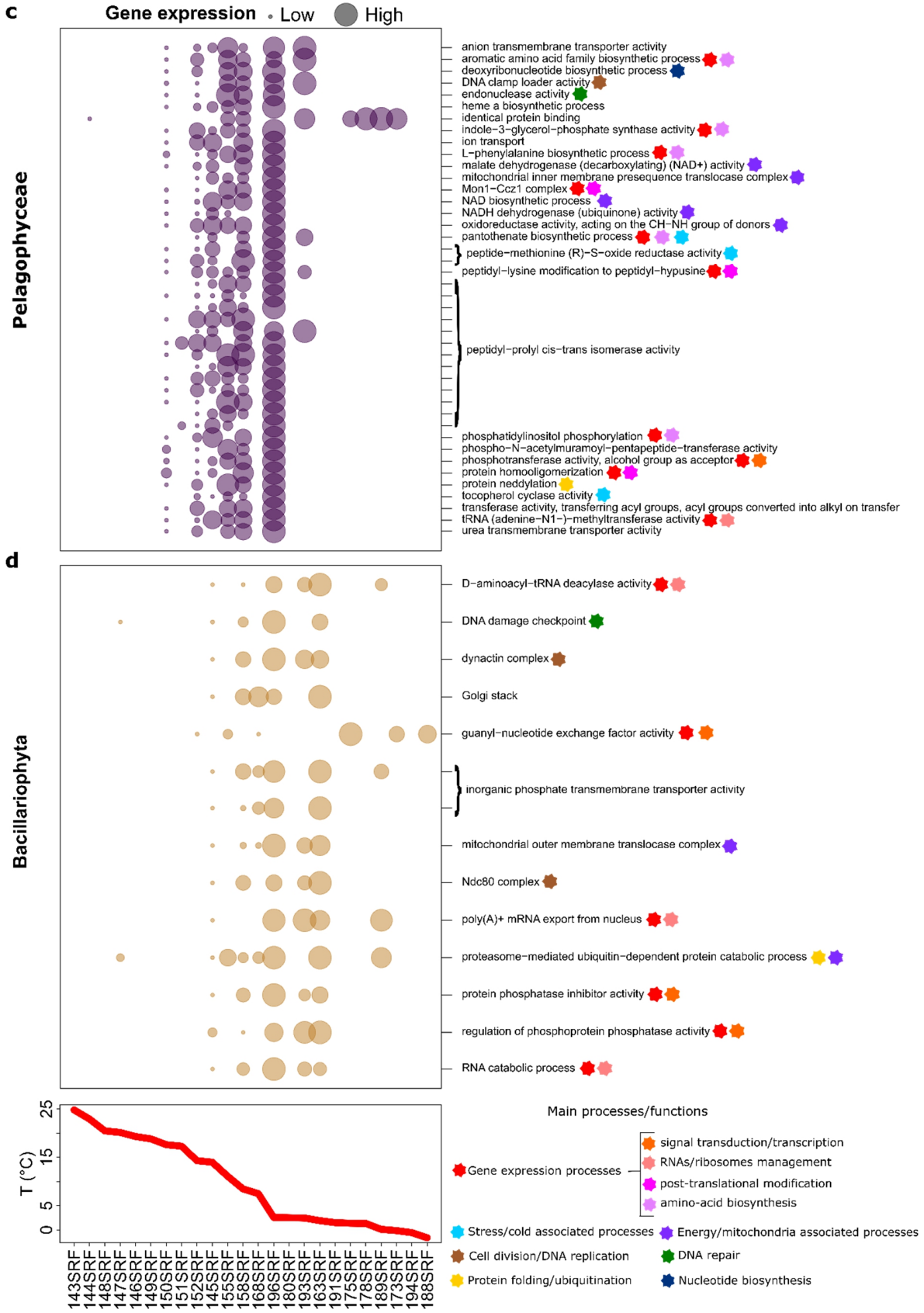
Supplementary Figure 15 | Distribution of Pearson correlation coefficient between temperature and randomized or real expression values of unigenes present in at least 5 stations for 6 algal Metagenomics-based Transcriptomes. The Pearson correlation distribution is represented for 6 algal MGTs (**a-d**). In blue, randomized data correlation distribution is presented and in other colors is the real data distribution. The expression value of a unigene in a given station is calculated as the difference between the centered log ratios of its metatranscriptomic relative abundance and metagenomic abundance. For some MGT correlation distribution are clearly distinguishable from the randomized data distribution indicating a clear response to temperature. It is particularly the case for MGT 76, MGT 195, MGT 42 and MGT 439 that display skewed distribution enriched in unigenes whose expression profile is negatively correlated with temperature.

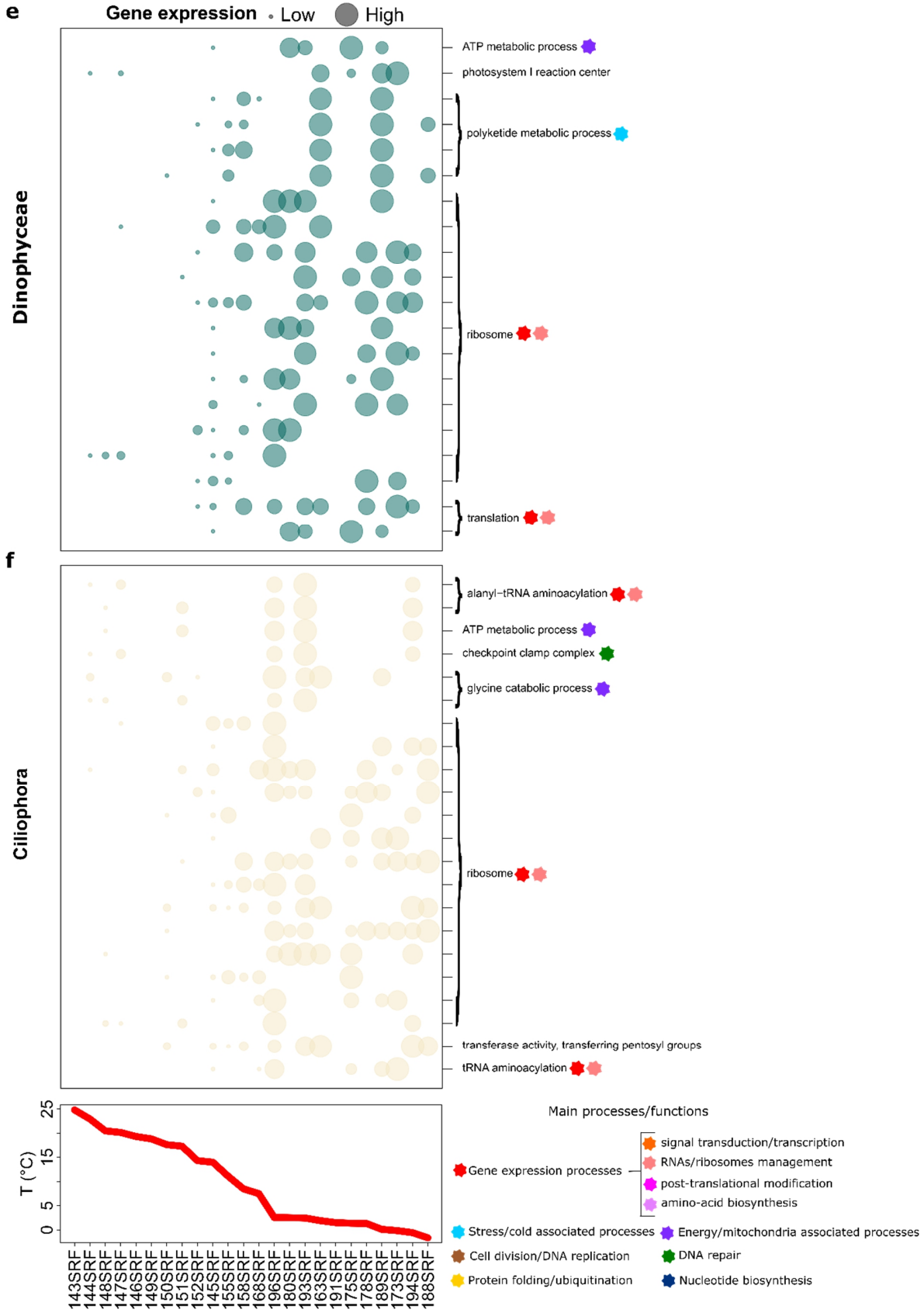

a
Gene expression
Low
High
Micromonas
MGT 76
BLOC−1 complex
coenzyme A biosynthetic process
dipeptidyl−peptidase activity
DNA replication initiation
histone acetyltransferase complex
malate transport
manganese ion transmembrane transporter activity
mitochondrial intermembrane space
motile cilium
mRNA cleavage factor complex
mRNA export from nucleus
phospholipase A2 activity
pyrimidine nucleotide biosynthetic process
sequence−specific DNA binding
transcription by RNA polymerase II
transcription corepressor activity
tRNA−intron endonuclease activity
vacuolar proton−transporting V−type ATPase complex assembly
b
Phaeocystales
MGT 4
ATP metabolic process
CCR4−NOT complex
cellular amino acid biosynthetic process
deoxyribonucleotide biosynthetic process
glutamyl−tRNA reductase activity
heat shock protein binding
IMP biosynthetic process
iron−sulfur cluster assembly
oxidoreductase activity, acting on NAD(P)H
photosystem I
photosystem II oxygen evolving complex
proton−transporting ATP synthase activity, rotational mechanism
regulation of cilium movement
rRNA binding
superoxide dismutase activity
translation initiation factor activity
ubiquinol−cytochrome−c reductase activity
T (°C)
25
15
5
0
143SRF
144SRF
148SRF
147SRF
146SRF
149SRF
150SRF
151SRF
152SRF
145SRF
155SRF
158SRF
168SRF
196SRF
180SRF
193SRF
163SRF
191SRF
175SRF
178SRF
189SRF
173SRF
194SRF
188SRF
Main processes/functions
Gene expression processes
signal transduction/transcription
RNAs/ribosomes management
post-translational modification
amino-acid biosynthesis
Stress/cold associated processes
Energy/mitochondria associated processes
Cell division/DNA replication
DNA repair
Protein folding/ubiquitination
Nucleotide biosynthesis



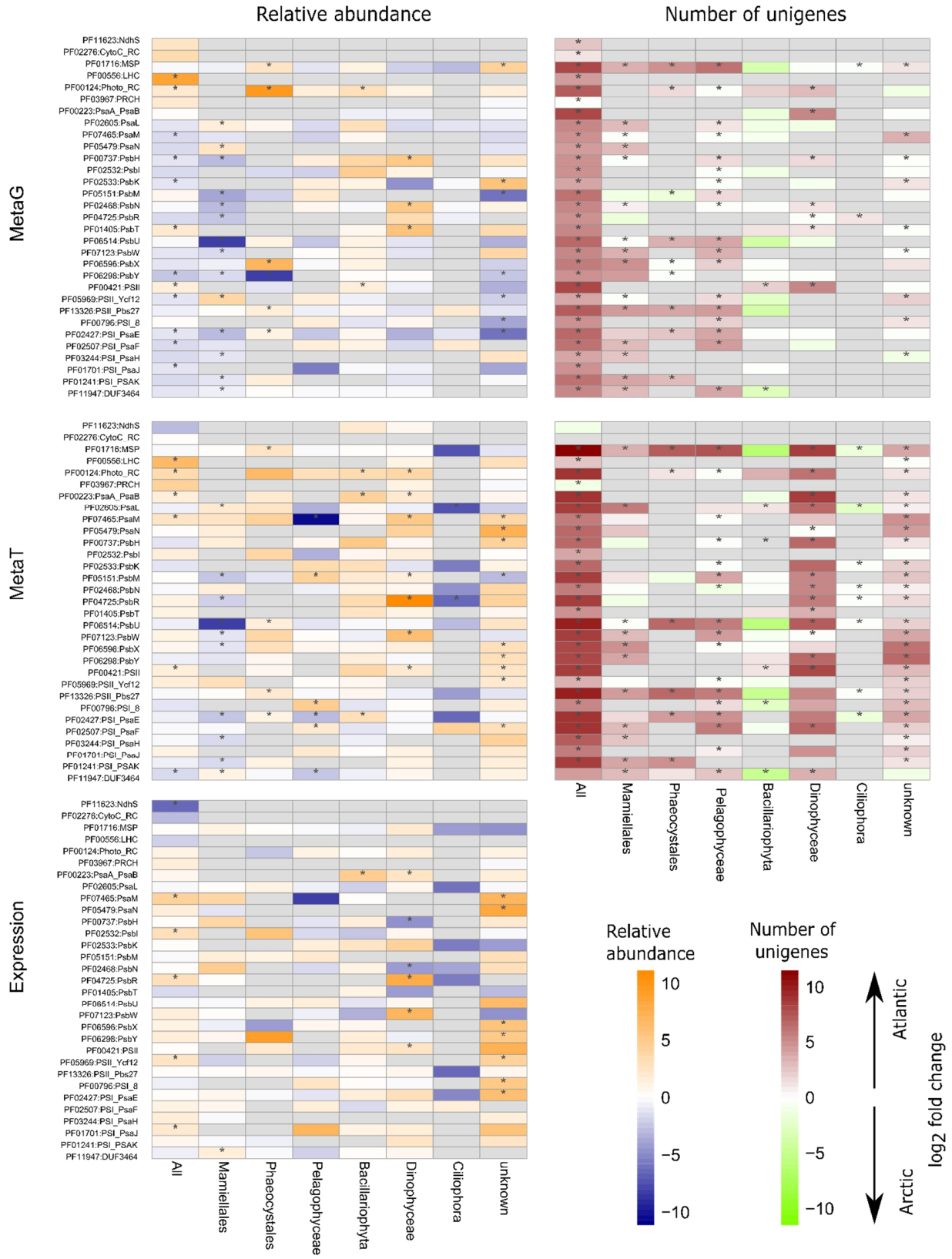
Supplementary Figure 16 | Expression profiles and Gene Ontology annotation of enriched unigenes negatively correlated with temperature for three algal Metagenomics-based Transcriptomes. Expression profiles of unigenes whose Gene Ontology terms (right) are enriched within unigenes significantly negatively correlated to temperature are displayed for three algal taxa from three Metagenomics-based Transcriptomes (MGTs): (**a**) Micromonas (MGT 76) (**b**) Phaeocystales (MGT 4). For MGT 4, unigenes associated to the 'ribosome' GO term were also enriched but removed as they were too numerous (24). (**c**) Pelagophyceae (MGT 42). Note that stations are ordered according to the temperature gradient.


a
Gene expression
Low
High
Mamiellales
cation transmembrane transporter activity
dipeptidyl−peptidase activity
DNA replication initiation
histone acetyltransferase complex
mitochondrial outer membrane translocase complex
nuclease activity
regulation of division septum assembly
sequence−specific DNA binding
transcription corepressor activity
transcription factor TFIIA complex
transmembrane transporter activity
b
Phaeocystales
cellular amino acid biosynthetic process
copper ion binding
photosystem I
photosystem II oxygen evolving complex
potassium ion transport
regulation of cilium movement
regulation of DNA replication
RNA binding
UDP−galactopyranose mutase activity
T (°C)
0
5
15
25
143SRF
144SRF
148SRF
147SRF
146SRF
149SRF
150SRF
151SRF
152SRF
145SRF
155SRF
158SRF
168SRF
196SRF
180SRF
193SRF
163SRF
191SRF
175SRF
178SRF
189SRF
173SRF
194SRF
188SRF
Main processes/functions
Gene expression processes
signal transduction/transcription
RNAs/ribosomes management
post-translational modification
amino-acid biosynthesis
Stress/cold associated processes
Energy/mitochondria associated processes
Cell division/DNA replication
DNA repair
Protein folding/ubiquitination
Nucleotide biosynthesis


c
Gene expression
Low
High
Pelagophyceae
anion transmembrane transporter activity
aromatic amino acid family biosynthetic process
deoxyribonucleotide biosynthetic process
DNA clamp loader activity
endonuclease activity
heme a biosynthetic process
identical protein binding
indole-3-glycerol-phosphate synthase activity
ion transport
L-phenylalanine biosynthetic process
malate dehydrogenase (decarboxylating) (NAD+) activity
mitochondrial inner membrane presequence translocase complex
Mon1-Ccz1 complex
NAD biosynthetic process
NADH dehydrogenase (ubiquinone) activity
oxidoreductase activity, acting on the CH-NH group of donors
pantothenate biosynthetic process
peptide-methionine (R)-S-oxide reductase activity
peptidyl-lysine modification to peptidyl-hypusine
peptidyl-prolyl cis-trans isomerase activity
phosphatidylinositol phosphorylation
phospho-N-acetylmuramoyl-pentapeptide-transferase activity
phosphotransferase activity, alcohol group as acceptor
protein homooligomerization
protein neddylation
tocopherol cyclase activity
transferase activity, transferring acyl groups, acyl groups converted into alkyl on transfer
tRNA (adenine-N1-)-methyltransferase activity
urea transmembrane transporter activity
d
Bacillariophyta
D-aminoacyl-tRNA deacylase activity
DNA damage checkpoint
dynactin complex
Golgi stack
guanyl-nucleotide exchange factor activity
inorganic phosphate transmembrane transporter activity
mitochondrial outer membrane translocase complex
Ndc80 complex
poly(A)+ mRNA export from nucleus
proteasome-mediated ubiquitin-dependent protein catabolic process
protein phosphatase inhibitor activity
regulation of phosphoprotein phosphatase activity
RNA catabolic process
T (°C)
25
15
5
0
143SRF
144SRF
148SRF
147SRF
146SRF
149SRF
150SRF
151SRF
152SRF
145SRF
155SRF
158SRF
168SRF
196SRF
180SRF
193SRF
163SRF
191SRF
175SRF
178SRF
189SRF
173SRF
194SRF
188SRF
Main processes/functions
Gene expression processes
signal transduction/transcription
RNAs/ribosomes management
post-translational modification
amino-acid biosynthesis
Stress/cold associated processes
Energy/mitochondria associated processes
Cell division/DNA replication
DNA repair
Protein folding/ubiquitination
Nucleotide biosynthesis

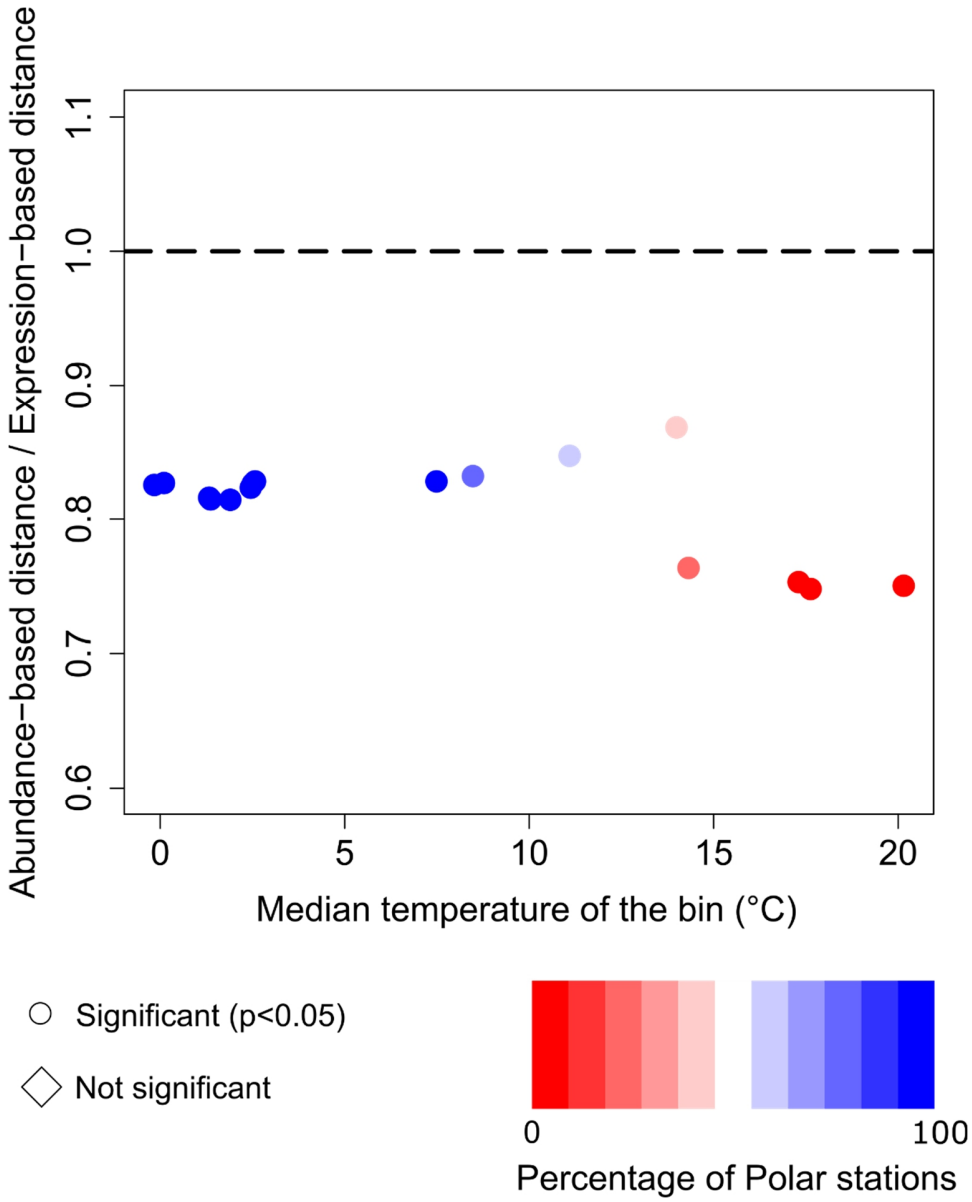

e
Gene expression
Low
High
Dinophyceae
ATP metabolic process
photosystem I reaction center
polyketide metabolic process
ribosome
translation
f
Ciliophora
alanyl−tRNA aminoacylation
ATP metabolic process
checkpoint clamp complex
glycine catabolic process
ribosome
transferase activity, transferring pentosyl groups
tRNA aminoacylation
T (°C)
25
15
5
0
143SRF
144SRF
148SRF
147SRF
146SRF
149SRF
150SRF
151SRF
152SRF
145SRF
155SRF
158SRF
168SRF
196SRF
180SRF
193SRF
163SRF
191SRF
175SRF
178SRF
189SRF
173SRF
194SRF
188SRF
Main processes/functions
Gene expression processes
signal transduction/transcription
RNAs/ribosomes management
post-translational modification
amino-acid biosynthesis
Stress/cold associated processes
Energy/mitochondria associated processes
Cell division/DNA replication
DNA repair
Protein folding/ubiquitination
Nucleotide biosynthesis

Supplementary Figure 17 | Expression profiles and Gene Ontology annotation of enriched unigenes negatively correlated with temperature for six algal taxa. Expression profiles of unigenes whose Gene Ontology terms (right) are enriched within unigenes significantly negatively correlated to temperature are displayed for six algal taxa: (**a**) Mamiellales (unigenes associated to the 'nucleus' GO term were also enriched but removed as they were too numerous (27) as well as the unknown), (**b**) Phaeocystales (unigenes associated to the 'ribosome' GO term were also enriched but removed as they were too numerous (40)) (**c**) Pelagophyceae (**d**) Bacillariophyta (**e**) Dinophyceae and (**f**) Ciliophora. Note that stations are ordered according to the temperature gradient.



Supplementary Figure 18 | Differential abundance in metatranscriptomes and

metagenomes and differential expression of PFAMs associated to photosynthesis across all and different taxons using ALDEx2 package (*13, 21*)**.** Right column: Median log2-fold change between Atlantic and Arctic metagenomes, metatranscriptomes (equation 6) and expression (metatranscriptomes normalized by metagenomes) of PFAMs associated to photosynthesis based on estimated centered log ratios from ALDEx2 package for all taxons, Mamiellales, Phaeocystales, Pelagophyceae, Bacillariophyta, Dinophyceae, Ciliophora and Unknown. Significant differences are highlighted by stars. Left column: Median log2-fold change of the number of unigenes between Atlantic and Arctic metagenomes and metatranscriptomes for aforementioned taxa.



Supplementary Figure 19 | Relative contribution of changes in community composition and gene expression changes to changes in metatranscriptomic composition along the temperature gradient of the NAO-AO transition. The median ratio between an abundance-based distance and an expression based distance (see Methods) is plotted for temperature bins each containing 5 stations. A Wilcoxon test and holm correction ($p<0.05$) is performed for each bin to test whether the median ratio is significantly different than 1.

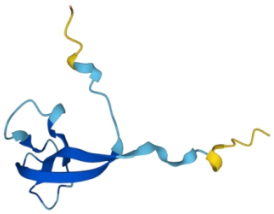
	Significantly more abundant functions	
Location	**Metagenomes**	**Metatranscriptomes**
Arctic	"cellulose binding" "sphingolipid metabolic process" "tetraacyldisaccharide 4'-kinase activity" "ceramide metabolic process" "polynucleotide dephosphorylation" "polynucleotide 5' dephosphorylation" "Mon1-Ccz1 complex" "polynucleotide phosphatase activity" "polynucleotide 5'-phosphatase activity" "transcription, RNA-templated" "RNA-directed 5'-3' RNA polymerase activity"	"hydrolase activity, acting on acid carbon-carbon bonds" "cis-trans isomerase activity" "ribonucleoprotein complex biogenesis" "pseudouridine synthase activity" "hydrolase activity, acting on acid carbon-carbon bonds, in ketonic substances" "preribosome, small subunit precursor" "ribosome biogenesis" "anatomical structure homeostasis" "fumarylacetoacetase activity" "snoRNA binding" "telomere organization" "ribonuclease H2 complex" "lipid catabolic process" "telomeric DNA binding" "S-adenosylmethionine-dependent methyltransferase activity" "intrinsic component of organelle membrane" "microtubule anchoring" "methyltransferase complex" "macromolecule methylation" "integral component of organelle membrane" "vacuole" "sphingolipid metabolic process" "urea metabolic process" "microbody" "tRNA methyltransferase complex" "peroxisome" "tRNA (m1A) methyltransferase complex" "protein transmembrane transport" "chromosome, telomeric region" "ceramide metabolic process" "vacuolar part" "nucleic acid phosphodiester bond hydrolysis" "mitochondrial RNA metabolic process" "peptidyl-amino acid modification" "RNA processing" "RNA modification" "rRNA metabolic process" "peptidyl-proline modification" "protein peptidyl-prolyl isomerization" "pseudouridine synthesis" "ncRNA processing" "serine-type exopeptidase activity" "nuclease activity" "nucleolus"

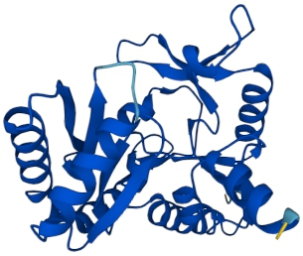
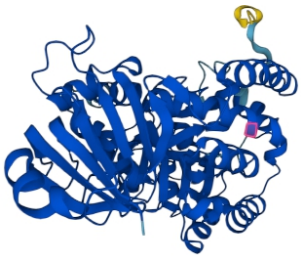
		"vacuolar membrane" "urea catabolic process" "RNA methylation" "serine-type carboxypeptidase activity" "endonuclease activity" "tRNA processing" "intracellular protein transmembrane transport" "peptidyl-prolyl cis-trans isomerase activity" "tRNA modification" "tRNA dihydrouridine synthesis" "rRNA processing" "telomere maintenance" "nuclear chromosome, telomeric region" "vacuolar proton-transporting V-type ATPase complex" "RNA methyltransferase activity" "tRNA methylation" "mitochondrial inner membrane presequence translocase complex" "transcription by RNA polymerase I" "tRNA dihydrouridine synthase activity" "tRNA methyltransferase activity" "tRNA (adenine) methyltransferase activity" "tRNA (adenine-N1-)-methyltransferase activity"
Atlantic	"glycosaminoglycan binding" "transferase activity, transferring aldehyde or ketonic groups" "ligase activity, forming carbon-carbon bonds" "argininosuccinate synthase activity" "hydroxymethylbilane synthase activity" "1-deoxy-D-xylulose-5-phosphate synthase activity" "CoA carboxylase activity" "phosphotransferase activity, carboxyl group as acceptor" "maintenance of location" "acetyl-CoA carboxylase activity" "chorismate synthase activity" "fatty acid synthase activity" "ribose-5-phosphate isomerase activity" "mannosamine metabolic process" "cellulase activity" "ER retention sequence binding" "N-acylglucosamine-6-phosphate 2-epimerase activity" "uridylyltransferase activity" "3-oxoacyl-[acyl-carrier-protein] synthase activity" "sulfate adenylyltransferase activity" "N-acetylmannosamine metabolic process" "photosynthesis, light reaction" "light-harvesting complex" "cytokinetic process" "regulation of cytoskeleton organization"	"transposition" "transposase activity" "chlorophyll binding" "photosynthetic electron transport chain" "ion channel complex" "photosynthetic electron transport in photosystem II" "electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity" "protein splicing" "transposition, DNA-mediated" "single strand break repair" "deoxyribonucleotide catabolic process"

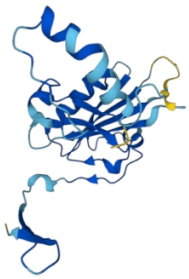
	"sulfate adenylyltransferase (ATP) activity" "fermentation" "DNA helicase complex" "glycopeptide alpha-N-acetylgalactosaminidase activity" "maintenance of protein location" "maintenance of location in cell" "oxidoreductase activity, acting on CH or CH2 groups, with an iron-sulfur protein as acceptor" "oxidoreductase activity, acting on diphenols and related substances as donors, with copper protein as acceptor" "regulation of microtubule-based movement" "regulation of cilium movement" "glycerol-3-phosphate dehydrogenase [NAD+] activity" "methylenetetrahydrofolate reductase (NAD(P)H) activity" "regulation of Toll signaling pathway" "plastoquinol--plastocyanin reductase activity" "5-methyltetrahydrofolate-dependent methyltransferase activity" "4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity" "terpenoid metabolic process" "acetyl-CoA carboxylase complex" "photosynthetic electron transport chain" "nitrogenous compound fermentation" "alditol phosphate metabolic process" "glycerol-3-phosphate metabolic process" "glutamyl-tRNA reductase activity" "peptidoglycan turnover" "photosynthetic electron transport in photosystem II" "proteasome accessory complex" "chorismate metabolic process" "electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity" "cell septum assembly" "phosphoglycerate kinase activity" "fatty acid metabolic process" "isoprenoid biosynthetic process" "maintenance of protein location in cell" "phycobilisome" "terpenoid biosynthetic process" "glycerol-3-phosphate catabolic process" "maintenance of protein localization in organelle" "methionine metabolic process" "dipeptidyl-peptidase activity" "maintenance of protein localization in endoplasmic reticulum" "L-lysine metabolic process"	

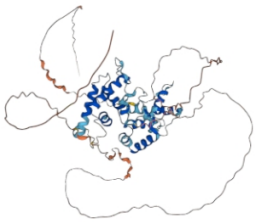
	"protein polyubiquitination" "protein retention in ER lumen" "biotin transport" "exonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters" "fatty acid biosynthetic process" "aspartate family amino acid catabolic process" "exodeoxyribonuclease activity" "monocarboxylic acid transmembrane transporter activity" "lysine catabolic process" "anaerobic amino acid catabolic process" "exodeoxyribonuclease activity, producing 5'-phosphomonoesters" "chlorophyll biosynthetic process" "methionine biosynthetic process" "L-lysine catabolic process" "pentose-phosphate shunt, non-oxidative branch" "queuosine salvage" "methionine synthase activity" "L-lysine catabolic process to acetate" "biotin transmembrane transporter activity"	

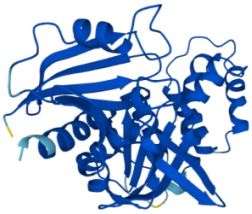
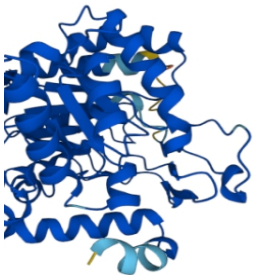
Supplementary Table 1 | List of more abundant GO terms in both North Atlantic and Arctic Ocean metagenomes and metatranscriptomes

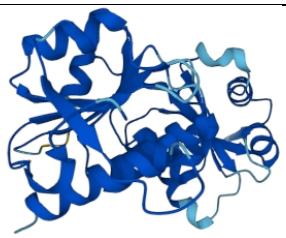
PFAM	Molecular function and structure	Relevance to cold/hot stress acclimation	References
PF14169: Cold−inducible protein YdjO	Molecular: Unknown Structure (example of YdjO) (48): 	Strongly cold-shock-induced in a transcriptome analysis of *B. subtilis*	(49)
PF05562: WCOR413	Molecular: putative transport of substances or ions, mediation of protein/protein interaction Structure: Highly hydrophobic protein with six transmembrane regions	A low-temperature-regulated protein from wheat resembling (25% to 56% identity) a stress-responsive gene from the resurrection plant Xerophyta viscosa Baker (XVSAP1). Hypothesis: "*It is possible that XVSAP1 may be involved in the transport of substances or ions across the plasma membrane as a region stretching from amino-terminal glycine residue of a growing polypeptide. Proteins so modified have diverse functions and the myristate appears to be critical for mediating protein/ protein and/or protein/membrane interactions*"(50)	(50, 51)
PF00313: CSD	Molecular: Part of this domain is highly similar to the RNP-1 RNA-binding motif(52). Structure: Protein domain of about 70 amino acids which has been found in prokaryotic and eukaryotic DNA-binding proteins(53, 54).	"*may have a function in activating transcription or unwinding or masking RNA molecules.*"(54)	(52–54)
PF02649: GCHY−1	Molecular: tetrahydrofolate biosynthesis. type I GTP cyclohydrolase Structure: highly conserved glutamate residue at position 216 in folE2: likely to be the substrate ligand. The metal ligand is likely to be the cysteine at position 147. Zn^{2+} dependent(55).	Cold-upregulated in a thermophilic alga: "*Glycine decarboxylase, which feeds methylene-tetrahydrofolate into the folate cycle was 1.7-fold upregulated at 28°C. In contrast, transcript frequency for formate-tetrahydrofolate ligase, which would drain the folate cycle, was reduced to 0.3-fold at 28°C*"(56) "*Both* in vitro *and* in vivo *experiments indicate that OtNOS, unlike mammalian NOS, efficiently uses tetrahydrofolate as a cofactor in Arabidopsis plants. The modulation of NO production to alleviate abiotic stress disturbances in higher plants highlights the potential of genetic manipulation to influence NO metabolism as a tool to improve plant fitness under adverse growth conditions.*"(57)	(55–57)
PF14249: Tocopherol_cycl	Molecular:tocopherol cyclase activity. Secondary metabolism Structure: 3 β-feuillet and 1 α-helix, strong confidence except the tip of the β and α(48) (N and C-terminal domains)	Abstract: "*As expected, the synthesis of tocopherols (vitamin E) from 2,3-dimethyl-5-phytyl-1,4-benzoquinol by the recombinant protein demonstrated that the EgSXD1 gene product is a TC. Northern hybridizations with RNA isolated from cell suspension cultures clearly showed a fast EgSXD1 downregulation by cold.*"(24)	(24)
PF01964: Thiamine biosynthesis	Molecular: Thiamine biosynthesis, "*Escherichia coli K-12 synthesizes thiamine pyrophosphate (vitamin B1)*	"*The expression of Arabidopsis genes involved in the thiamine diphosphate biosynthesis pathway, including that of* THI1, THIC, TH1 *and* TPK, *was analyzed for 48 h in seedlings*	(22, 58)

	de novo. Two precursors [4-methyl-5-(beta-hydroxyethyl)thiazole monophosphate and 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate] are coupled to form thiamine monophosphate, which is then phosphorylated to make thiamine pyrophosphate."(22) Structure: A β-feuillet surrounded by α-helixes and α-β-α on the side (C-terminal)(58)	*subjected to NaCl or sorbitol treatment. These genes were found to be predominantly up-regulated in the early phase (2-6 h) of the stress response."*(22)	
PF02358: Trehalose_PPase	Molecular: Trehalose phosphatase Structure (example of Trehalose phosphatase)(48): 	Abstract: *"A mutant Escherichia coli strain unable to produce trehalose died much faster than the wild type at 4°C. Transformation of the mutant with the otsA/otsB genes, responsible for trehalose synthesis, restored trehalose content and cell viability at 4°C. After temperature downshift from 37°C to 16°C ("cold shock"), trehalose levels in wild-type cells increased up to 8-fold."*(59)	(59)
PF01979: Amidohydro_1	Molecular: Allantoinase, Catalyzes the conversion of allantoin (5-ureidohydantoin) to allantoic acid by hydrolytic cleavage of the five-member hydantoin ring(60). Structure (example of allantoinase)(48): 	*"Allantoin is a metabolic intermediate of purine catabolism that often accumulates in stressed plants*(27)*."*	(27, 60, 61)
PF10551: MULE	Molecular: Mutator-like element (MULE) transposases. Zn dependent(62) Structure: This domain was identified by Babu and colleagues(62). "*Typical structural features of MULEs include long TIRs (>100 bp) and 8–10 bp flanking TSDs*(63)"	NONE (control)	(62, 63)

PF01119: DNA_mis_repair	Function : DNA repair(64). Structure: "*This family represents the C-terminal domain of the mutL/hexB/PMS1 family. This domain has a ribosomal S5 domain 2-like fold.*"(64)	Title: DNA Damage Inducible Protein 1 is Involved in Cold Adaption of Harvested Cucumber Fruit(65)	(64, 65)
PF00626: Gelsolin	Function : Gelsolin is a cytoplasmic, calcium-regulated, actin-modulating protein that binds to the barbed ends of actin filaments, preventing monomer exchange (end-blocking or capping). It can promote nucleation (the assembly of monomers into filaments), as well as sever existing filaments(66). Structure (example of Gelsolin)(48) : uncertain 	*"Cold-acclimation had a negligible effect on steady state protein expression in the heart, with just 7 differentially abundant (DA) proteins (0.5% of the quantified cardiac proteome). Furthermore, only gelsolin isoform X1 (log2FC = -0.289) was not influenced by a confounding effect of available development stage"*	(66)
PF00241: Cofilin_ADF	Function: "*They bind both actin-monomers and filaments and promote rapid filament turnover in cells by depolymerising/fragmenting actin filaments. ADF/cofilins bind ADP-actin with higher affinity than ATP-actin and inhibit the spontaneous nucleotide exchange on actin monomers*"(67) Structure: "*In molecular biology, ADF-H domain (actin-depolymerising factor homology domain) is an approximately 150 amino acid motif that is present in three phylogenetically distinct classes of eukaryotic actin-binding proteins*(68)."	*"Based on the above information and on the identification of a wheat cold-regulated ADF homolog, we designed the present experiments to determine the function and regulation of TaADF."*(69)	(67–69)
PF02209: VHP	Molecular: Bundles, nucleates, caps, and severs actin in a Ca2+-dependent	*"Compared with Table I, only eight spots, which represents seven proteins (38.9%) including a villin (Os03g24220), an actin 3 (Os03g61970), an actin 7 (Os05g01600), an aminotransferase (Os08g41990), an ATP-dependent Clp protease (Os04g32560),*	(70–72)

	manner(70). Structure: Villin headpiece domain. "*The core of the GH domain consists of a five-stranded b-sheet sandwiched between one long and one short helix*"(71)	*a putative subtilisin (Os02g53860), and a sucrose synthase (Os03g28330), showed similar up- or down-regulated patterns."*(72)	
PF03259: Robl_LC7	Molecular: associated to flagellar outer arm of dynein. Structure: See Koonin et al.(73)	This family includes proteins that may modulate specific dynein functions but it may also play a structural or regulatory role. Dynein were shown to be regulated in cold conditions and act as a *"thermally tunable motor"*(74)	(73, 74)
PF00249: Myb_DNA−binding	Molecular: Myb-like DNA-binding domain (transcription factor) Structure(48): 	Putative role in oxidative stress response in *Arabisdopsis Thaliana*: regulation of ROS in a circadian manner (probable regulation by CCA1)(75)	(75)
PF11999: Ice_binding	Molecular: Ice binding, antifreeze glycoproteins, they bind to crystals of ice Structure (example of Ice binding protein 1)(48): 	"*Sub-zero temperatures put plants at risk of damage associated with the formation of ice crystals in the apoplast. Some freeze-tolerant plants mitigate this risk by expressing ice-binding proteins (IBPs), that adsorb to ice crystals and modify their growth."*(76)	(48, 76)
PF03104: DNA_pol_B_exo1	Molecular: 3' to 5' exonuclease activity Structure: ribonuclease H type fold	*"Previous studies [15] showed changes in RNase activity in wheat leaves and roots following exposure to low temperatures (2 8C) for long periods (60 days)."*(77)	(77)
PF01346: FKBP_N	Molecular: peptidyl-prolyl isomerase (foldase/chaperone), post translation modification Structure(78): Domain amino terminal to FKBP-type	Some peptidyl prolyl isomerase have been shown to be important in protein folding under cold stress condition by increasing the rate of cis-trans isomerization(79)	(78, 79)
PF05035: DGOK	Molecular: 2-dehydro-3-deoxygalactonokinase activity Structure (example of	NONE (control)	(48)

	deoxygalactonokinase)(48): 		
PF05351: GMP_PDE_delta	Molecular: *"PDE delta subunit is thought to be a specific soluble transport factor for certain prenylated proteins and Arl2-GTP a regulator of PDE-mediated transport."*(80) Structure (exemple of GMP_PDE_delta containing protein) (48): 	*"The enzyme is stable only under specific conditions, and the activation property of the enzyme is lost relatively easy. Irreversible modifications occur at temperatures below 0° and above 30°C, and at pH below 6.0. Several other conditions such as high ion concentrations, temperatures just above 0°C and pH above 8.0 lead to reversibel modifications of enzyme activity."*(81)	(80, 81)
PF01180: DHO_dh	Molecular: cytosolic dihydroorotate dehydrogenase Structure (example of dihydroorotate dehydrogenase) (48): 	*"In the present paper, we identified and cloned OsDHODH1 encoding a putative cytosolic dihydroorotate dehydrogenase (DHODH) in rice. Expression analysis indicated that OsDHODH1 is upregulated by salt, drought and exogenous abscisic acid (ABA), but not by cold"*(82) *"Even though there is no data on the effect of temperature on DhoDH, we can expect that its thermal sensitivity will affect more mitochondrial functions than only the reduction of the Q pool."*(83)	(82, 83)
PF05971: Methyltransf_10	Molecular: RNA methyltransferase Structure (example of Methyltransferase)(48):	RNA methylation is known to be important in stress response especially cold acclimation (stability of RNA)(84)	(84, 85)

Supplementary table 2 | Summary table of function and structure of a set of PFAMs known to be involved in cold/hot stress acclimation.

References

1. P. Wassmann, K. N. Kosobokova, D. Slagstad, K. F. Drinkwater, R. R. Hopcroft, S. E. Moore, I. Ellingsen, R. J. Nelson, E. Carmack, E. Popova, J. Berge, The contiguous domains of Arctic Ocean advection: Trails of life and death. *Prog Oceanogr*. **139** (2015), doi:10.1016/j.pocean.2015.06.011.

2. A. C. Palmisano, C. W. Sullivan, PHYSIOLOGY OF SEA ICE DIATOMS. I. RESPONSE OF THREE POLAR DIATOMS TO A SIMULATED SUMMER-WINTER TRANSITION. *J Phycol*. **18** (1982), doi:10.1111/j.1529-8817.1982.tb03215.x.

3. F. Kong, I. T. Romero, J. Warakanont, Y. Li-Beisson, Lipid catabolism in microalgae. *New Phytologist*. **218** (2018), doi:10.1111/nph.15047.

4. W. Li, R. Wang, M. Li, L. Li, C. Wang, R. Welti, X. Wang, Differential degradation of extraplastidic and plastidic lipids during freezing and post-freezing recovery in Arabidopsis thaliana. *J Biol Chem*. **283**, 461–468 (2008).

5. B. Usadel, O. E. Bläsing, Y. Gibon, F. Poree, M. Höhne, M. Günter, R. Trethewey, B. Kamlage, H. Poorter, M. Stitt, Multilevel genomic analysis of the response of transcripts, enzyme activities and metabolites in Arabidopsis rosettes to a progressive decrease of temperature in the non-freezing range. *Plant Cell Environ*. **31**, 518–547 (2008).

6. I. A. Guschina, J. L. Harwood, Lipids and lipid metabolism in eukaryotic algae. *Prog Lipid Res*. **45**, 160–186 (2006).

7. S. S. Gill, S. Willette, B. Dungan, J. M. Jarvis, T. Schaub, D. M. van Leeuwen, R. St Hilaire, F. Omar Holguin, Suboptimal temperature acclimation affects kennedy pathway gene expression, lipidome and metabolite profile of nannochloropsis salina during PUFA enriched TAG synthesis. *Mar Drugs*. **16** (2018), doi:10.3390/md16110425.

8. L. Valledor, T. Furuhashi, A. M. Hanak, W. Weckwerth, Systemic cold stress adaptation of chlamydomonas reinhardtIII. *Molecular and Cellular Proteomics*. **12** (2013), doi:10.1074/mcp.M112.026765.

9. E. Huby, J. A. Napier, F. Baillieul, L. v. Michaelson, S. Dhondt-Cordelier, Sphingolipids: towards an integrated view of metabolism during the plant stress response. *New Phytologist*. **225** (2020), , doi:10.1111/nph.15997.

10. J. H. T. M. Fabri, N. P. de Sá, I. Malavazi, M. del Poeta, The dynamics and role of sphingolipids in eukaryotic organisms upon thermal adaptation. *Prog Lipid Res*. **80** (2020), , doi:10.1016/j.plipres.2020.101063.

11. D. Carrero, J. G. Pérez-Silva, V. Quesada, C. López-Otín, Differential mechanisms of tolerance to extreme environmental conditions in tardigrades. *Sci Rep*. **9** (2019), doi:10.1038/s41598-019-51471-8.

12. J. J. Luxton, M. J. McKenna, A. Lewis, L. E. Taylor, K. A. George, S. M. Dixit, M. Moniz, W. Benegas, M. J. Mackay, C. Mozsary, D. Butler, D. Bezdan, C. Meydan, B. E. Crucian, S. R. Zwart, S. M. Smith,

C. E. Mason, S. M. Bailey, Telomere Length Dynamics and DNA Damage Responses Associated with Long-Duration Spaceflight. *Cell Rep*. **33** (2020), doi:10.1016/j.celrep.2020.108457.

13. A. D. Fernandes, J. N. S. Reid, J. M. Macklaim, T. A. McMurrough, D. R. Edgell, G. B. Gloor, Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. **2** (2014), doi:10.1186/2049-2618-2-15.

14. Y. P. Wang, L. S. Zhou, Y. Z. Zhao, S. W. Wang, L. L. Chen, L. X. Liu, Z. Q. Ling, F. J. Hu, Y. P. Sun, J. Y. Zhang, C. Yang, Y. Yang, Y. Xiong, K. L. Guan, D. Ye, Regulation of G6PD acetylation by SIRT2 and KAT9 modulates NADPH homeostasis and cell survival during oxidative stress. *EMBO Journal*. **33** (2014), doi:10.1002/embj.201387224.

15. L. Jin, Y. Zhou, Crucial role of the pentose phosphate pathway in malignant tumors (review). *Oncol Lett*. **17** (2019), doi:10.3892/ol.2019.10112.

16. H. Liu, D. Huang, D. L. McArthur, L. G. Boros, N. Nissen, A. P. Heaney, Fructose induces transketolase flux to promote pancreatic cancer growth. *Cancer Res*. **70** (2010), doi:10.1158/0008-5472.CAN-09-4615.

17. J. M. Holding, C. M. Duarte, J. M. Arrieta, R. Vaquer-Suyner, A. Coello-Camba, P. Wassmann, S. Agustí, Experimentally determined temperature thresholds for Arctic plankton community metabolism. *Biogeosciences*. **10** (2013), doi:10.5194/bg-10-357-2013.

18. N. T. Miller, M. D. Vaughn, R. L. Burnap, Electron flow through NDH-1 complexes is the major driver of cyclic electron flow-dependent proton pumping in cyanobacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. **1862**, 148354 (2021).

19. W. Huang, Y.-J. Yang, H. Hu, S.-B. Zhang, Different roles of cyclic electron flow around photosystem I under sub-saturating and saturating light intensities in tobacco leaves. *Front Plant Sci*. **6** (2015), doi:10.3389/fpls.2015.00923.

20. M. Suorsa, Cyclic electron flow provides acclimatory plasticity for the photosynthetic machinery under various environmental conditions and developmental stages. *Front Plant Sci*. **6** (2015), doi:10.3389/fpls.2015.00800.

21. A. D. Fernandes, J. M. Macklaim, T. G. Linn, G. Reid, G. B. Gloor, ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One*. **8** (2013), doi:10.1371/journal.pone.0067019.

22. M. Rapala-Kozik, N. Wolak, M. Kujda, A. K. Banas, The upregulation of thiamine (vitamin B 1) biosynthesis in Arabidopsis thaliana seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response. *BMC Plant Biol*. **12** (2012), doi:10.1186/1471-2229-12-2.

23. S. Ferreira, K. Hjernø, M. Larsen, G. Wingsle, P. Larsen, S. Fey, P. Roepstorff, M. Salomé Pais, Proteome profiling of Populus euphratica Oliv. upon heat stress. *Ann Bot*. **98** (2006), doi:10.1093/aob/mcl106.

24. W. el Kayal, G. Keller, C. Debayles, R. Kumar, D. Weier, C. Teulieres, C. Marque, Regulation of tocopherol biosynthesis through transcriptional control of tocopherol cyclase during cold hardening in Eucalyptus gunnii. *Physiol Plant*. **126** (2006), doi:10.1111/j.1399-3054.2006.00614.x.

25. S. Q. Ouyang, S. J. He, P. Liu, W. K. Zhang, J. S. Zhang, S. Y. Chen, The role of tocopherol cyclase in salt stress tolerance of rice (Oryza sativa). *Sci China Life Sci*. **54** (2011), doi:10.1007/s11427-011-4138-1.

26. R. Szymańska, J. Kruk, Plastoquinol is the main prenyllipid synthesized during acclimation to high light conditions in arabidopsis and is converted to plastochromanol by tocopherol cyclase. *Plant Cell Physiol*. **51** (2010), doi:10.1093/pcp/pcq017.

27. H. Takagi, Y. Ishiga, S. Watanabe, T. Konishi, M. Egusa, N. Akiyoshi, T. Matsuura, I. C. Mori, T. Hirayama, H. Kaminaka, H. Shimada, A. Sakamoto, Allantoin, a stress-related purine metabolite, can activate jasmonate signaling in a MYC2-regulated and abscisic acid-dependent manner. *J Exp Bot*. **67** (2016), doi:10.1093/jxb/erw071.

28. P. Wang, C. H. Kong, B. Sun, X. H. Xu, Distribution and function of allantoin (5-ureidohydantoin) in rice grains. *J Agric Food Chem*. **60** (2012), doi:10.1021/jf2051043.

29. M. M. Seppänen, K. Fagerstedt, The role of superoxide dismutase activity in response to cold acclimation in potato. *Physiol Plant*. **108** (2000), doi:10.1034/j.1399-3054.2000.108003279.x.

30. K. H. Baek, D. Z. Skinner, Differential expression of manganese superoxide dismutase sequence variants in near isogenic lines of wheat during cold acclimation. *Plant Cell Rep*. **25** (2006), doi:10.1007/s00299-005-0073-6.

31. C. Kayihan, F. Eyidogan, N. Afsar, H. A. Oktem, M. Yucel, Cu/Zn superoxide dismutase activity and respective gene expression during cold acclimation and freezing stress in barley cultivars. *Biol Plant*. **56** (2012), doi:10.1007/s10535-012-0143-x.

32. J. Porankiewicz, A. K. Clarke, Induction of the heat shock protein ClpB affects cold acclimation in the cyanobacterium Synechococcus sp. strain PCC 7942. *J Bacteriol*. **179** (1997), doi:10.1128/jb.179.16.5111-5117.1997.

33. N. Ukaji, C. Kuwabara, D. Takezawa, K. Arakawa, S. Yoshida, S. Fujikawa, Accumulation of small heat-shock protein homologs in the endoplasmic reticulum of cortical parenchyma cells in mulberry in association with seasonal cold acclimation. *Plant Physiol*. **120** (1999), doi:10.1104/pp.120.2.481.

34. L. E. Teigen, J. I. Orczewska, J. McLaughlin, K. M. O'Brien, Cold acclimation increases levels of some heat shock protein and sirtuin isoforms in threespine stickleback. *Comp Biochem Physiol A Mol Integr Physiol*. **188** (2015), doi:10.1016/j.cbpa.2015.06.028.

35. Y. Huang, D. Jin, C. Lu, X. Lan, P. Qiao, H. Li, Y. Chen, Proteomic responses associated with freezing tolerance in the callus of the Tibetan alpine plant Saussurea laniceps during cold acclimation. *Plant Cell Tissue Organ Cult*. **124** (2016), doi:10.1007/s11240-015-0876-2.

36. A. C. Seybold, D. A. Wharton, M. A. S. Thorne, C. J. Marshall, Investigating trehalose synthesis genes after cold acclimation in the Antarctic nematode Panagrolaimus sp. DAW1. *Biol Open*. **6**, 1953–1959 (2017).

37. D. Z. Skinner, Phospholipase A2 activity during cold acclimation of wheat. *Crop Sci*. **49** (2009), doi:10.2135/cropsci2008.06.0359.

38. X.-C. Wang, Q.-Y. Zhao, C.-L. Ma, Z.-H. Zhang, H.-L. Cao, Y.-M. Kong, C. Yue, X.-Y. Hao, L. Chen, J.-Q. Ma, J.-Q. Jin, X. Li, Y.-J. Yang, Global transcriptome profiles of Camellia sinensis during cold acclimation. *BMC Genomics*. **14**, 415 (2013).

39. T. Pang, C.-Y. Ye, X. Xia, W. Yin, De novo sequencing and transcriptome analysis of the desert shrub, Ammopiptanthus mongolicus, during cold acclimation using Illumina/Solexa. *BMC Genomics*. **14**, 488 (2013).

40. C. M. Paget, J.-M. Schwartz, D. Delneri, Environmental systems biology of cold-tolerant phenotype in Saccharomyces species adapted to grow at different temperatures. *Mol Ecol*. **23**, 5241–5257 (2014).

41. M. Moulager, A. Monnier, B. Jesson, R. Bouvet, J. Mosser, C. Schwartz, L. Garnier, F. Corellou, F. Y. Bouget, Light-dependent regulation of cell division in Ostreococcus: Evidence for a major transcriptional input. *Plant Physiol*. **144** (2007), doi:10.1104/pp.107.096149.

42. J. S. O-Neill, G. van Ooijen, L. E. Dixon, C. Troein, F. Corellou, F. Y. Bouget, A. B. Reddy, A. J. Millar, Circadian rhythms persist without transcription in a eukaryote. *Nature*. **469** (2011), doi:10.1038/nature09654.

43. G. Benoit, M. Mariadassou, S. Robin, S. Schbath, P. Peterlongo, C. Lemaitre, SimkaMin: Fast and resource frugal de novo comparative metagenomics. *Bioinformatics*. **36** (2020), doi:10.1093/bioinformatics/btz685.

44. P. J. Somerfield, Identification of the Bray-Curtis similarity index: Comment on Yoshioka (2008). *Mar Ecol Prog Ser* (2008), , doi:10.3354/meps07841.

45. V. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. **10** (1966).

46. R. R. Sokal, A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull*. **38** (1958).

47. I. Erb, C. Notredame, How should we measure proportionality on relative gene expression data? *Theory in Biosciences*. **135** (2016), doi:10.1007/s12064-015-0220-8.

48. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).

49. T. Kaan, G. Homuth, U. Mäder, J. Bandow, T. Schweder, Genome-wide transcriptional profiling of the Bacillus subtilis cold-shock response. *Microbiology (N Y)*. **148**, 3441–3455 (2002).

50. D. Garwe, J. A. Thomson, S. G. Mundree, Molecular characterization of XVSAP1, a stress-responsive gene from the resurrection plant Xerophyta viscosa Baker. *J Exp Bot*. **54**, 191–201 (2003).

51. X. Huang, A. Gaballa, M. Cao, J. D. Helmann, Identification of target promoters for the Bacillus subtilis extracytoplasmic function σ factor, σW. *Mol Microbiol*. **31**, 361–371 (1999).

52. D. Landsman, RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res*. **20**, 2861–2864 (1992).

53. J. Doniger, D. Landsman, M. A. Gonda, G. Wistow, The product of unr, the highly conserved gene upstream of N-ras, contains multiple repeats similar to the cold-shock domain (CSD), a putative DNA-binding motif. *New Biol*. **4**, 389–395 (1992).

54. P. G. Jones, M. Inouye, The cold-shock response--a hot topic. *Mol Microbiol*. **11**, 811–818 (1994).

55. B. Sankaran, S. A. Bonnett, K. Shah, S. Gabriel, R. Reddy, P. Schimmel, D. A. Rodionov, V. de Crécy-Lagard, J. D. Helmann, D. Iwata-Reuyl, M. A. Swairjo, Zinc-independent folate biosynthesis: genetic, biochemical, and structural investigations reveal new metal dependence for GTP cyclohydrolase IB. *J Bacteriol*. **191**, 6936–6949 (2009).

56. A. W. Rossoni, G. Schönknecht, H. J. Lee, R. L. Rupp, S. Flachbart, T. Mettler-Altmann, A. P. M. Weber, M. Eisenhut, Cold Acclimation of the Thermoacidophilic Red Alga Galdieria sulphuraria: Changes in Gene Expression and Involvement of Horizontally Acquired Genes. *Plant Cell Physiol*. **60**, 702–712 (2019).

57. N. Foresi, M. L. Mayta, A. F. Lodeyro, D. Scuffi, N. Correa-Aragunde, C. García-Mata, C. Casalongué, N. Carrillo, L. Lamattina, Expression of the tetrahydrofolate-dependent nitric oxide synthase from the green alga Ostreococcus tauri increases tolerance to abiotic stresses and influences stomatal development in Arabidopsis. *The Plant Journal*. **82**, 806–821 (2015).

58. S. Coquille, C. Roux, A. Mehta, T. P. Begley, T. B. Fitzpatrick, S. Thore, High-resolution crystal structure of the eukaryotic HMP-P synthase (THIC) from Arabidopsis thaliana. *J Struct Biol*. **184**, 438–444 (2013).

59. O. Kandror, A. DeLeon, A. L. Goldberg, Trehalose synthesis is induced upon exposure of Escherichia coli to cold and is essential for viability at low temperatures. *Proc Natl Acad Sci U S A*. **99**, 9727–9732 (2002).

60. https://www.uniprot.org/uniprot/Q8XGS6.

61. I. Ramazzina, L. Cendron, C. Folli, R. Berni, D. Monteverdi, G. Zanotti, R. Percudani, Logical identification of an allantoinase analog (puuE) recruited from polysaccharide deacetylases. *J Biol Chem*. **283**, 23295–23304 (2008).

62. M. M. Babu, L. M. Iyer, S. Balaji, L. Aravind, The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res*. **34**, 6505–6520 (2006).

63. K. Liu, S. R. Wessler, Transposition of Mutator-like transposable elements (MULEs) resembles hAT and Transib elements and V(D)J recombination. *Nucleic Acids Res*. **45**, 6644–6655 (2017).

64. S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. **48**, D265–D268 (2020).

65. B. Wang, G. Wang, S. Zhu, DNA Damage Inducible Protein 1 is Involved in Cold Adaption of Harvested Cucumber Fruit. *Front Plant Sci*. **10**, 1723 (2019).

66. A. G. WEEDS, J. GOOCH, B. POPE, H. E. HARRIS, Preparation and characterization of pig plasma and platelet gelsolins. *Eur J Biochem*. **161**, 69–76 (1986).

67. P. Lappalainen, M. M. Kessels, M. J. Cope, D. G. Drubin, The ADF homology (ADF-H) domain: a highly exploited actin-binding module. *Mol Biol Cell*. **9**, 1951–1959 (1998).

68. https://pfam.xfam.org/family/PF00241.

69. F. Ouellet, E. Carpentier, M. J. Cope, A. F. Monroy, F. Sarhan, Regulation of a wheat actin-depolymerizing factor during cold acclimation. *Plant Physiol*. **125**, 360–368 (2001).

70. E. Friederich, K. Vancompernolle, D. Louvard, J. Vandekerckhove, Villin function in the organization of the actin cytoskeleton. Correlation of in vivo effects to its biochemical activities in vitro. *J Biol Chem*. **274**, 26751–26760 (1999).

71. U. Ghoshdastider, D. Popp, L. D. Burtnick, R. C. Robinson, The expanding superfamily of gelsolin homology domain proteins. *Cytoskeleton (Hoboken)*. **70**, 775–795 (2013).

72. C. Huo, B. Zhang, H. Wang, F. Wang, M. Liu, Y. Gao, W. Zhang, Z. Deng, D. Sun, W. Tang, Comparative Study of Early Cold-Regulated Proteins by Two-Dimensional Difference Gel Electrophoresis Reveals a Key Role for Phospholipase Dα1 in Mediating Cold Acclimation Signaling Pathway in Rice. *Mol Cell Proteomics*. **15**, 1397–1411 (2016).

73. E. V Koonin, L. Aravind, *Curr Biol*, in press, doi:10.1016/s0960-9822(00)00774-0.

74. W. Hong, A. Takshak, O. Osunbayo, A. Kunwar, M. Vershinin, The Effect of Temperature on Microtubule-Based Transport by Cytoplasmic Dynein and Kinesin-1 Motors. *Biophys J*. **111**, 1287–1294 (2016).

75. L. A. Grace, D. C. J., M.-R. Bernd, K. S. A., S. J. H. M., D. P. P., CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proceedings of the National Academy of Sciences*. **109**, 17129–17134 (2012).

76. M. Bredow, V. K. Walker, Ice-Binding Proteins in Plants. *Front Plant Sci*. **8** (2017), doi:10.3389/fpls.2017.02153.

77. C. Spanò, C. Crosatti, R. Pacchini, P. Meletti, I. Grilli, Ribonucleases during cold acclimation in winter and spring wheats. *Plant Science*. **162**, 809–815 (2002).

78. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, M. Punta, Pfam: The protein families database. *Nucleic Acids Res*. **42** (2014), , doi:10.1093/nar/gkt1223.

79. C. Budiman, Y. Koga, K. Takano, S. Kanaya, FK506-Binding Protein 22 from a Psychrophilic Bacterium, a Cold Shock-Inducible Peptidyl Prolyl Isomerase with the Ability to Assist in Protein Folding. *Int J Mol Sci*. **12**, 5261–5284 (2011).

80. https://www.ncbi.nlm.nih.gov/Structure/cdd/PF05351.

81. R. Bulgakov, P. J. M. Van Haastert, Isolation and partial characterization of a cyclic GMP-dependent cyclic GMP-specific phosphodiesterase from Dictyostelium discoideum. *Biochimica et Biophysica Acta (BBA) - General Subjects*. **756**, 56–66 (1983).

82. W. Y. Liu, M. M. Wang, J. Huang, H. J. Tang, H. X. Lan, H. S. Zhang, The OsDHODH1 gene is involved in salt and drought tolerance in rice. *J Integr Plant Biol*. **51** (2009), doi:10.1111/j.1744-7909.2009.00853.x.

83. H. Lemieux, P. U. Blier, Exploring Thermal Sensitivities and Adaptations of Oxidative Phosphorylation Pathways. *Metabolites*. **12**, 360 (2022).

84. C. Zhang, G. Jia, Reversible RNA Modification N1-methyladenosine (m1A) in mRNA and tRNA. *Genomics Proteomics Bioinformatics*. **16** (2018), , doi:10.1016/j.gpb.2018.03.003.

85. https://www.rcsb.org/structure/6b91.