



OPEN

Global identification of a marine diatom long noncoding natural antisense transcripts (NATs) and their response to phosphate fluctuations

Maria Helena Cruz de Carvalho^{1,2✉} & Chris Bowler¹

Often ignored and regarded as mere transcriptional noise, long noncoding RNAs (lncRNAs) are starting to be considered key regulators of gene expression across the Eukarya domain of life. In the model diatom *Phaeodactylum tricornutum*, we have previously reported the occurrence of 1,510 intergenic lncRNAs (lincRNAs), many of which displaying specific patterns of expression under phosphate fluctuation (Pi). Using strand-specific RNA-sequencing data we now expand the repertoire of *P. tricornutum* lncRNAs by identifying 2,628 novel natural antisense transcripts (NATs) that cover 21.5% of the annotated genomic loci. We found that NAT expression is tightly regulated by phosphate depletion and other naturally occurring environmental stresses. Furthermore, we identified 121 phosphate stress responsive NAT-mRNA pairs, the great majority of which showing a positive correlation (concordant pairs) and a small fraction with negative correlation (discordant pairs). Taken together our results show that NATs are highly abundant transcripts in *P. tricornutum* and that their expression is under tight regulation by nutrient and environmental stresses. Furthermore, our results suggest that in *P. tricornutum* Pi stress response NAT pairs predominantly regulate positively the expression of their cognate sense genes, the latter being involved in several biological processes underlying the control of cellular homeostasis under stress.

Diatoms are unicellular microalgae belonging to the stramenopile supergroup that emerged from an ancient secondary endosymbiosis over 200 My ago^{1,2}. They have since risen to prominence in the global ocean and are now amongst the major ocean primary producers³⁻⁵. Very good competitors when nutrients are abundant⁶, diatoms are also extremely resilient to nutrient fluctuations in their environment, being able to quickly resume growth as soon as the environment is again favorable⁷⁻⁹. Understanding of the molecular underpinnings regulating this resilience is nonetheless still very incomplete.

Long noncoding RNA (lncRNA) transcripts originate from unannotated regions of genomes and were long-time regarded as transcriptional noise. However, with the huge amount of data generated by next generation high throughput RNA sequencing, and subsequent functional characterization of several lncRNAs, they are now gaining momentum as key regulators of gene expression across the Eukarya domain of life. lncRNAs form a heterogeneous class of transcripts that are generally capped and polyadenylated, like mRNA transcripts, but have as common features a sequence longer than 200 nucleotides with no viable open reading frame (ORF)^{10,11}. Based on their relationship with protein coding genes, lncRNAs can be called intergenic (lincRNAs) when they are located between protein coding genes, intronic (incRNAs) when they are expressed from introns, or natural antisense transcripts (NATs, also referred to as *cis*-NATs) when they are expressed from the opposite strand of sense genes of the same genomic locus¹⁰. The latter group seems to be by far the most pervasive, having been described in both prokaryotes¹² and eukaryotes¹³. Historically, NATs have been neglected as byproducts of spurious transcription but enough evidence has now been gathered to show that NATs can de facto be important

¹Institut de Biologie de L'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France. ²Faculté des sciences et technologie, Université Paris Est-Créteil (UPEC), 94000 Créteil, France. ✉email: cruz@biologie.ens.fr

modulators of gene expression. NATs function by several different mechanisms, either locally, around the locus where they are expressed, or far from the site of transcription¹⁴. NATs may affect directly their cognate sense gene expression by transcription interference¹⁵, or indirectly by assisting chromatin remodeling^{16–19}. NATs may also form double stranded regulatory RNAs with their overlapping cognate transcripts and be processed by Dicer or Dicer-like enzymes, leading to RNA interference and gene silencing^{20–22}, or be involved in RNA editing and/or RNA masking¹⁴. Downstream effects of RNA masking include alternative splicing²³ and regulation of translation^{24,25}. Tissue- or developmental-specific expression of NATs seems to be an important determinant of the regulatory role of some NATs¹³. These noncoding transcripts have also been reported to be involved in the regulation of responses to various environmental signals, from yeast, to plants, to humans^{26–30}.

In the model pennate diatom, *Phaeodactylum tricornerutum*, we have previously reported the occurrence of ~1,500 lincRNAs, many of which being under phosphate (Pi), nitrogen (N)⁹ or pCO₂³¹ regulation. Using strand-specific RNA-sequencing data⁹ we now expand the repertoire of *P. tricornerutum* lincRNAs by identifying 2,628 novel NATs and characterizing their expression patterns under phosphate fluctuations as well as under other environmental constraints. Amongst the -Pi stress regulated NATs, approximately one third were correlated with their cognate sense protein coding genes, forming tight sense-antisense NAT-mRNA pairs (NAT pairs). Interestingly, the majority of the identified NAT pairs were positively correlated, forming concordant pairs, with only a small fraction being negatively correlated, forming discordant pairs. These results suggest a new level of gene regulation in *P. tricornerutum* whereby NAT pairs predominantly regulate positively the expression of their cognate sense protein coding genes.

Materials and methods

Diatom culture conditions. Axenic cultures of *P. tricornerutum* strain CCMP632 were obtained from the Center for the Culture of Marine Phytoplankton (Maine, USA). They were cultured in 250 mL filtered (Grade 1 Whatman) steam-sterilized artificial sea water (40 g/L, Sigma) supplemented with f/2 nutrients, elements and vitamins³² with the exception of silica (f/2-Si), in 1 L glass flasks under continuous shaking (100 rpm) at 20 °C under cool white fluorescent lights at 100 μmol.m⁻² s⁻¹ with a 12 h photoperiod. For phosphate fluctuation studies we used the 5 physiological time points previously described⁹. Equal aliquots of 4 day-old cultures from the same batch culture were inoculated in parallel in 250 mL fresh f/2-Si media (control conditions) and in 250 mL fresh f/2-Si media without phosphate supplement (Pi depleted) and cultured in the same conditions as described above. For Pi re-supplementation, 4-d-old Pi-depleted cultures were pelleted by centrifugation for 10 min at 1,500 g, resuspended in 250 ml of fresh f/2-Si medium and cultured in the same conditions described above. For nitrogen depletion studies, equal aliquots of 4 day-old cultures from the same batch culture were inoculated in parallel in 250 mL fresh f/2-Si media (control conditions) and in 250 mL fresh f/2-Si media without nitrogen supplement (N depleted) and cultured in the same conditions as described above. For heat and cold stress treatments, cultures were transferred for 20 min to a water bath at 30 °C and 10 °C, respectively. Salt shock treatment consisted in cultures being transferred to f/2 media with twice the amount of sea salt concentration (80 g/L) for 20 min. Culture growth was followed using a hemacytometer (Fisher Scientific, USA). Cultures (controlled and treated) were all started with the same initial cell densities of ~4.5 × 10⁵ cells/mL. Cells were harvested at the same time of day (6 h after beginning of the light period) by vacuum filtration on 0.2 μm polycarbonate filters (Millipore, United States), flash-frozen in liquid nitrogen and maintained at -80 °C until used. All the experiments were repeated at least 3 times.

Data sources. The strand-specific RNA Sequencing (ssRNA-seq) datasets used in this study were obtained from a previous publication⁹. Briefly, strand-specific (ss)RNA-seq reads were mapped to the *P. tricornerutum* genome using TopHat (version 2.0.8)³³ and then assembled using Cufflinks (v2.1.1)³⁴. Each ssRNA-seq sample was then merged and annotated using Cuffcompare (v2.1.1)³⁴. Expression levels of each gene were then calculated by the fragments per kilobase of exons per million fragments mapped (FPKM) using Cuffdiff (v2.1.1)³⁴. The Pearson correlation coefficients (PCC) were calculated between the FPKM of 2–3 replicates of each 5 physiological states corresponding to 2–3 independent experiments. In order to increase the depth of the reads for increased detection of noncoding RNA transcripts, and since the PCCs were high (>0.9), we pooled reads derived from the two most robust replicates as previously described⁹. A twofold variance in FPKM and a P value < 0.05 (Fisher's exact test) were used as cutoff to define differentially expressed (DE) genes.

Quantitative real-time reverse transcription PCR. Total RNA was extracted from *P. tricornerutum* flash frozen cell pellets using the Trizol method according to the manufacturer's instructions (Invitrogen). Extracted RNA was treated with TurboDNase I (Life Technologies AM2238) and then purified using and RNeasy spin column (Qiagen) with 0.5 volumes of ethanol, washed and then eluted with 50 μL room temperature molecular grade water (Qiagen). The quality of the purified RNA was assessed using the Agilent 2100 Bionalyzer. For cDNA synthesis, 500 ng RNA was incubated with SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. For quantitative RT-PCR analysis, cDNA was amplified using SYBR Premix ExTaq (Takara) with specific primers (Table S2) picked from a random sample of identified NATs. Primers were designed with the Primer-Blast program (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) defining PCR amplicon size between 70–150 bp and *Phaeodactylum tricornerutum* CCAP 1055/1 (taxid: 556484) as the reference organism to check for primer pair specificity. Quantitative 2-step PCR conditions were set as following: 95 °C for 10 s followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s and a final cycle of 95 °C for 10 s and 60 °C for 5 s in a Bio-Rad CFX96 real-time system or a Roche LightCycler 480 in a 384 well. *CDKA* and *HISTONE 4* mRNA levels were used for normalization³⁵.

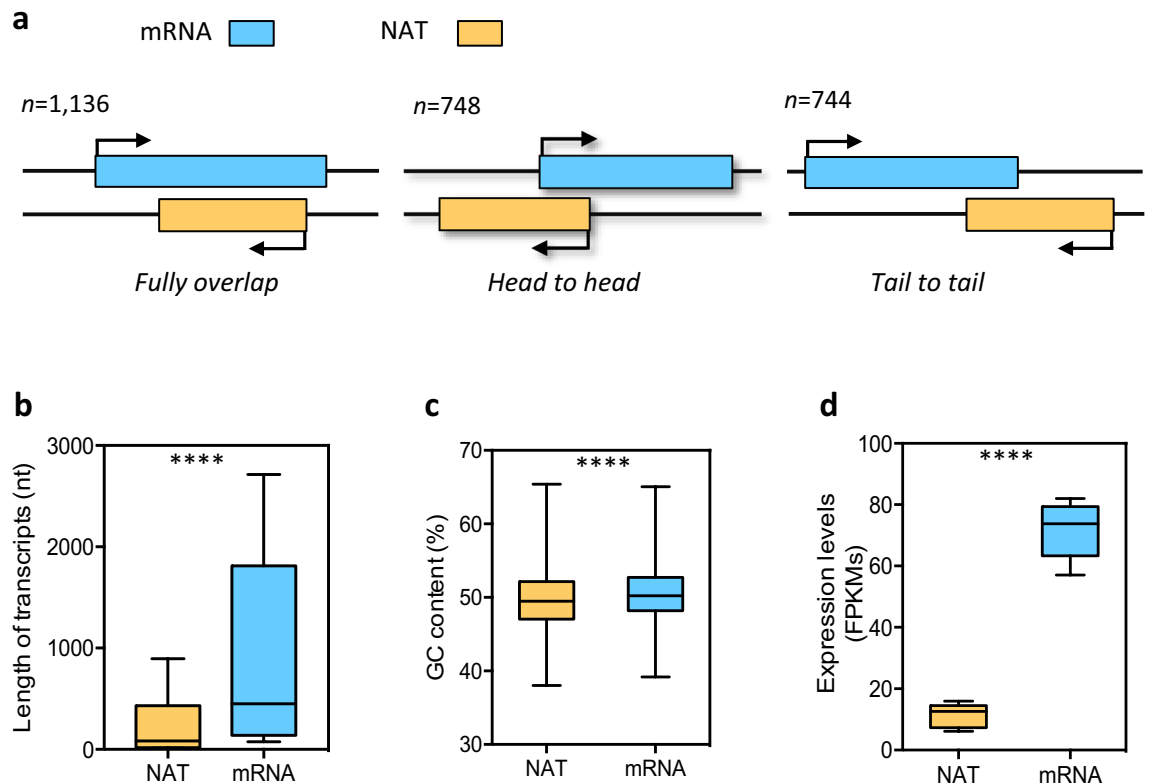


Figure 1. Characteristics of NAT and mRNA transcripts in *Phaeodactylum tricoratum*. **(a)** Schematic representation of the long noncoding natural antisense transcripts (NAT) identified in *P. tricoratum*. NATs include transcripts that are fully overlapped by their cognate protein coding genes (mRNA), or that have partial overlap at the 5' (head to head) or the 3' (tail to tail) ends. Arrows indicate the direction of transcription; n, number of NATs identified in each category. **(b)** length distribution in nucleotides (nt) **(c)** GC content (%) and **(d)** expression levels detected by RNA-Seq (FPKM) of NATs and mRNAs was assessed by a paired *t* test, (****) *P* value < 0.0001.

Identification of NATs and phosphate-regulated NAT-mRNA pairs. All assembled intergenic transcription units were collected as lincRNA candidates. Candidates with length ≥ 200 nt, predicted ORF ≤ 100 amino acids and with at least 50 nt overlap with annotated gene models were defined as NATs^{30,36}. Low abundance NATs (FPKMmax < 1) were excluded from the analysis. The length of transcripts was used to compare transcript length distribution. For ORF predictions, the spliced transcripts were used and sent to GenScan³⁷.

To detect putative NAT-mRNA pairs (expression correlation between sense-antisense NAT and overlapping gene models), the PCCs were calculated using FPKMs of all the phosphate-responsive NATs versus mRNAs encoded by their overlapping genes. We classified phosphate-regulated NAT pairs into two groups: concordant regulation and discordant regulation. For the identification of phosphate-regulated concordant NAT pairs, one transcript should change in expression level in the same direction as its partner during phosphate fluctuations and for discordant NAT pairs, expression levels of the sense and antisense transcripts should change in the opposite direction. Only those Pi responsive NAT-mRNA pairs with $r^2 \geq 0.6$ (positive or negative correlation) were considered as putative NAT pairs.

Results

Genome wide identification of NATs in *Phaeodactylum tricoratum*. Using high-depth strand-specific RNA sequencing (ssRNA-seq) data from at least 10 independent cDNA libraries from polyadenylated [poly(A)+] RNAs under normal or phosphate stress conditions (Accession no. GSE669970)⁹, we assembled all the reads to reconstruct the *P. tricoratum* transcriptome. After removing transcripts that were previously annotated as protein coding, transcripts with short length (< 200 nt), low abundance (FPKMmax < 1) and putative open reading frames (ORFs) of more than 100 amino acids, we identified a total of 4,138 lincRNAs. Amongst these were the 1,510 previously annotated lincRNAs⁹ plus 2,628 newly identified transcripts belonging to the natural antisense transcript family (NATs) (Supplementary Table S1), having at least 50 nt overlapped with annotated gene models (Phatr3). Around 43% (1,136) of the NATs identified were transcripts that fully overlapped the coding genes on the opposite strand and the remaining 57% (1,492), were partial overlaps, either head to head (748) or tail to tail (744) (Fig. 1a). Overall, the newly identified NATs cover 21.5% of the annotated genomic loci (Phatr3).

NATs in *P. tricornutum* represent cross-kingdom genomic features. We analyzed the average length and expression levels of the identified NATs and compared these to the annotated mRNAs detected in our transcriptomic data. NATs were significantly shorter in length than mRNAs (Fig. 1b) (P value < 0.0001 , paired t test) with a median length of 740 nt for NATs and 1,266 nt for mRNAs. Similarly, lincRNAs were also found to be shorter than mRNAs⁹. Interestingly, previously identified *P. tricornutum* lincRNAs which have a median size of 482 nt were significantly shorter than NATs (Supplementary Fig. S1). Another feature of functional lincRNAs is lower GC content than mRNAs³⁸, which was noted previously in *P. tricornutum* lincRNAs⁹ and was also found in *P. tricornutum* NATs (Fig. 1c). The expression levels of NATs and mRNAs, estimated by FPKM using Cuffdiff³⁴, revealed that the average expression levels of NATs was much lower than those of protein coding mRNAs (P -value < 0.0001 , paired t test) (Fig. 1d). Interestingly, significant differences in expression levels were found between NATs and lincRNAs in *P. tricornutum*, with NATs having the lowest expression levels of the three transcript categories (NATs, lincRNAs, mRNAs) (Supplementary Fig. S1).

NATs are highly abundant in the *P. tricornutum* phosphate responsive transcriptome. Diatoms are frequently subjected to nutrient fluctuations in their environment, and phosphate has been shown to be amongst the major nutrients affecting primary production in the ocean, namely in coastal areas where *P. tricornutum* is predominantly present^{39,40}. In order to assess the functional relevance of the identified NATs in *P. tricornutum* we sought to investigate their pattern of expression under Pi fluctuations. In order to do this we calculated the FPKM from the ssRNA-seq datasets obtained from diatom cultures under different Pi regimes (Pi starved for 4 and 8 days, Pi resupplemented for 4 days after 4 days depletion and control cultures grown for 4 and 8 days in full media)⁹. Of the 2,628 NATs identified, 1,481 NATs were differentially expressed (DE) (fold change ≥ 2 or ≤ 0.5 , P value < 0.05) under Pi depletion, which corresponds to $\sim 56\%$ of all the NATs identified in *P. tricornutum*. Amongst these DE NATs, 830 were up-regulated (fold change ≥ 2 , P value < 0.05) in response to Pi depletion (after 4 and/or 8 days Pi depleted) and 402 were specifically up-regulated by Pi since once Pi was re-supplied to the media, their expression levels returned to control values or lower (Fig. 2a,b; Supplementary Table S1). These results were validated by repeating Pi fluctuation experiments (Pi depletion and re-supply), extracting RNA from the cultures and performing RT-qPCR using specific primers of a randomly selected set of Pi depletion up-regulated NATs (Supplementary Table S2) and obtaining equivalent expression trend variation (Fig. 2c). Regarding Pi-specific regulation in the other gene categories, we detected 720 mRNAs and 209 lincRNAs (Fig. 2a). This reveals that the Pi-regulated noncoding transcriptome (NATs and lincRNAs) in *P. tricornutum* is almost equivalent in number to the Pi-regulated coding transcriptome (mRNAs) with 611 and 720 transcripts, respectively, with NATs being the dominant lincRNAs transcripts.

NAT expression is regulated by environmental stress. To further validate our results and investigate whether NATs in *P. tricornutum* were responsive to other stresses besides phosphate stress, we checked by RT-qPCR the relative expression levels of a randomly selected set of NATs using cDNA libraries originating from *P. tricornutum* cultures submitted to different treatments. The treatments were made with the intent to mimic recurrent nutrient or mineral stresses, such as nitrogen and iron depletion over the course of 4 days, thermal and salt stress designed to mimic changes in environmental temperature or salinity due to vertical mixing of currents. All the NATs tested were found to be stress-regulated either with an increased expression or a reduced expression, depending on the stress (Fig. 3). NAT_XLOC_014884, NAT_XLOC_014882 and NAT_XLOC_007148 were all up-regulated by Pi depletion but down-regulated under N or Fe depletion (Fig. 3). On the other hand, NAT_XLOC_009104 was up-regulated by Pi and also N depletion, suggesting it could be involved in common pathways shared by the two nutrient stresses, whereas NAT_XLOC_009726 and NAT_XLOC_004000 were up-regulated by N depletion only. This confirms the specificity of the NAT response and shows that despite being expressed at lower levels than mRNAs (Fig. 1d), diatom NATs are under tight environmental regulation and are not randomly expressed. Given the specificity and robustness of the response, it is unlikely that the noncoding transcriptome and, in particular, the novel NATs detected in the present study, are transcriptional noise or byproducts of spurious transcription.

NAT-mRNA pairs in *Phaeodactylum* are predominantly concordant. To detect putative expression correlations between sense-antisense NATs and overlapping gene models forming NAT (-mRNA) pairs, we calculated the Pearson correlation coefficients (PCCs) using the FPKMs of all the responsive NATs versus the mRNAs encoded by their overlapping genes. We defined Pi stress regulated NATs if their expression responded to the following pattern: up-regulation (fold change ≥ 2 , P value < 0.05) in Pi depleted conditions with subsequent down-regulation (fold change ≤ 0.5 , P value < 0.05) or return to control levels when Pi was resupplied to the media or the opposite; down-regulation (fold change ≥ 0.5 , P value < 0.05) in Pi depleted conditions with subsequent up-regulation (fold change ≤ 2 , P value < 0.05) or return to control levels when Pi was resupplied to the media. We defined concordant NAT pairs as those in which the expression levels of both transcripts changed in the same direction during phosphate fluctuations and for discordant NAT pairs, the expression levels of the two transcripts should change in opposite directions. We found that amongst the 810 Pi stress regulated NATs, 121 formed sense-antisense NAT pairs which reveals interdependency between both transcripts (Fig. 4). Correlated NAT-mRNA pairs included 34 tail-to-tail, 39 head-to-head and 48 fully overlapped (Supplementary Table S3). Amongst these interdependent NAT pairs, the majority were positively correlated forming concordant NAT pairs (99) and only a small fraction were discordant (22) (Fig. 4a, Supplementary Table S3). Within the concordant pairs, 25 were tail-to-tail, 36 head-to-head, and 38 were fully overlapping NAT-mRNA pairs. Within the discordant pairs, 9 were tail-to-tail, 3 head-to-head and 10 fully overlapped, showing that 5' end pairing is less frequent for discordant NAT-pairs in *P. tricornutum* than tail-to-tail and fully overlapping pairs.

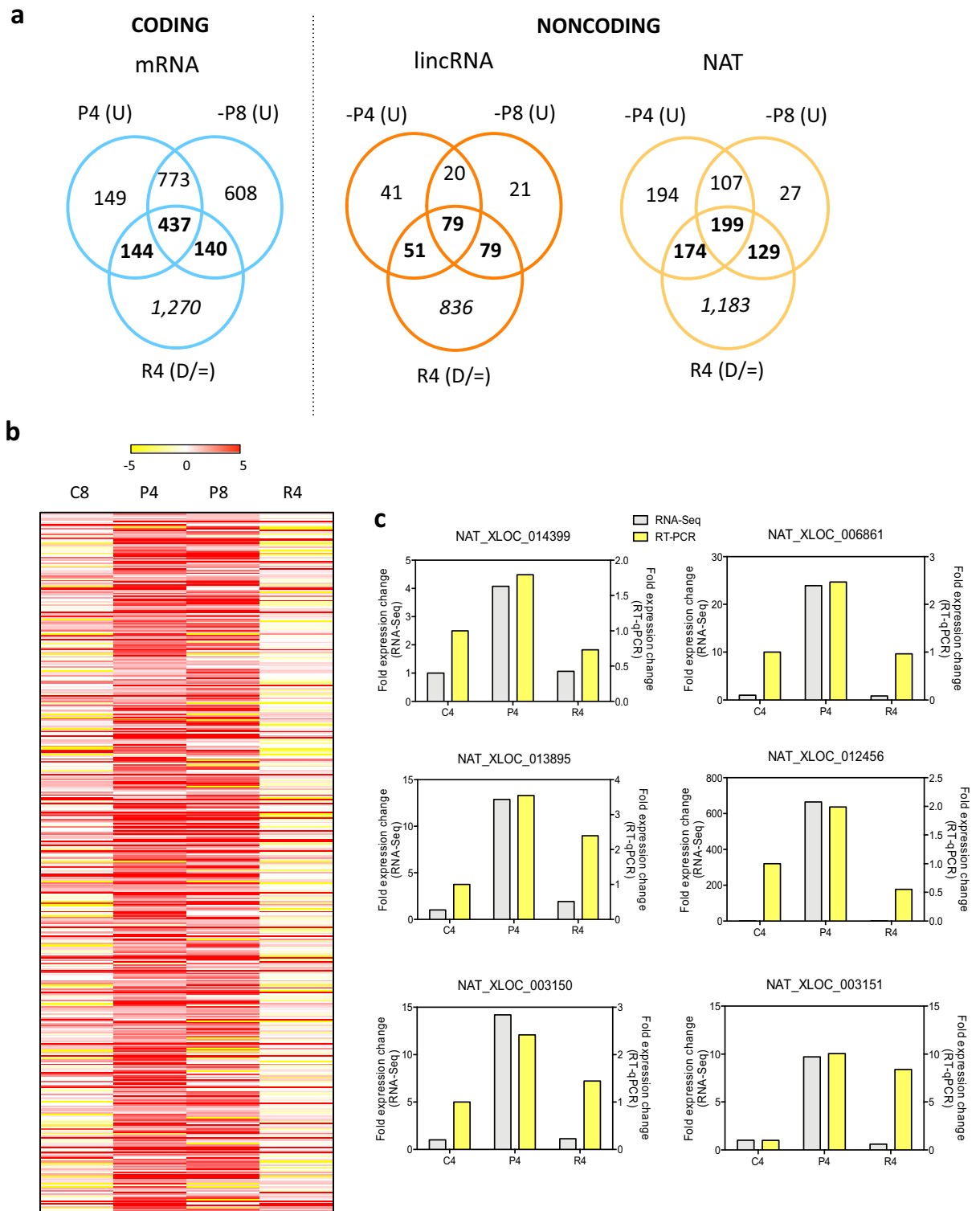


Figure 2. *P. tricornutum* NAT detection and expression under Pi fluctuations. **(a)** Venns representing the three categories of phosphate responsive *P. tricornutum* transcripts: mRNA (coding), NAT and lincRNA (noncoding) under phosphate fluctuations. **(b)** Heatmap of the NATs up-regulated by Pi depletion (> twofold, *P* value < 0.05). Rows represent log₂ expression change in relation to control 4 days, columns represent four different physiological states marked by Pi fluctuations **(c)** RT-qPCR validation of NAT RNA-Seq expression data under Pi fluctuation. P4, 4 days Pi stress; P8, 8 days phosphate stress; U, up-regulated; R4, 4 days recovery after 4 days Pi stress; D, down-regulated or =, unchanged transcripts.

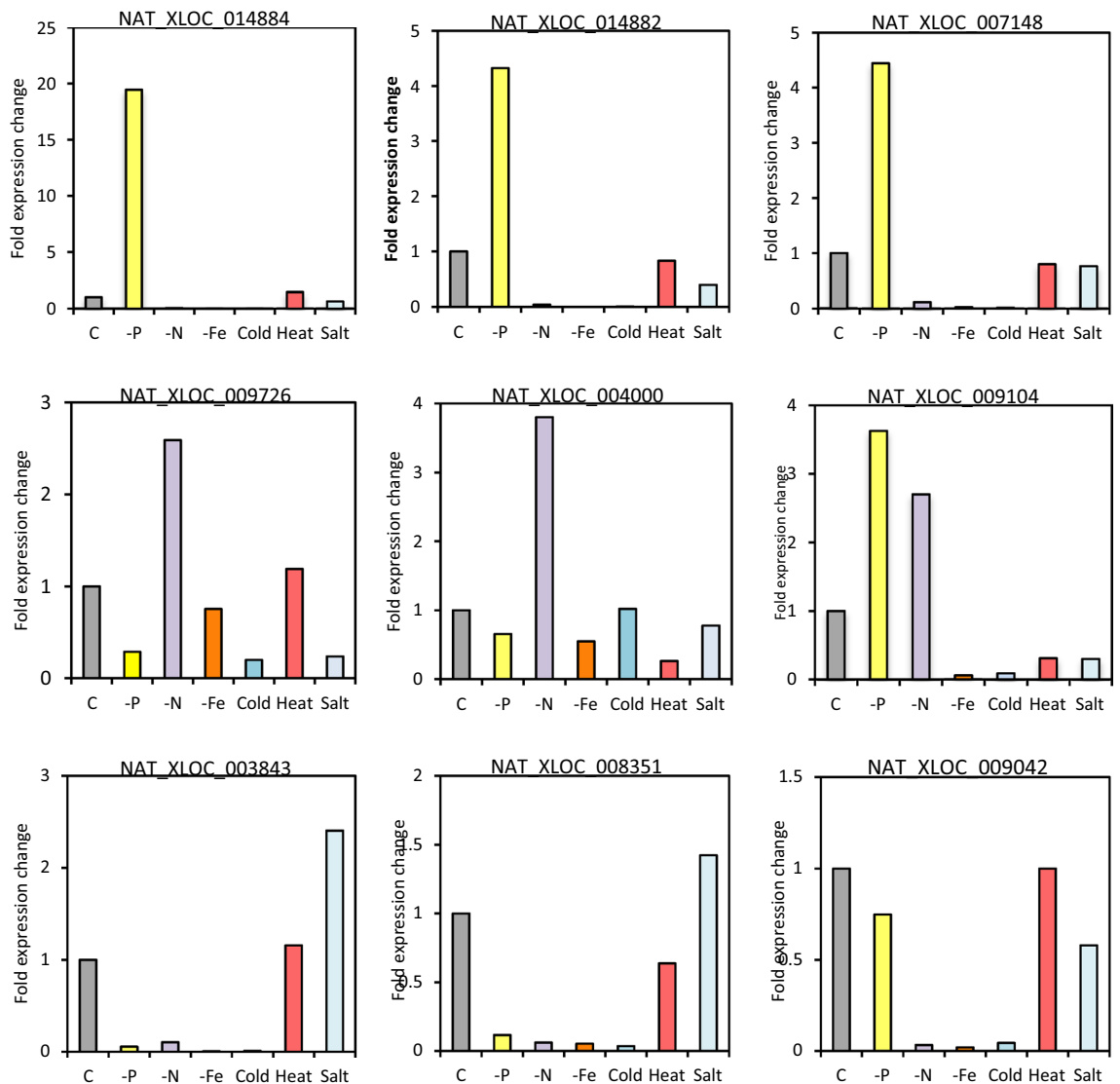


Figure 3. RT-qPCR stress expression analysis of NATs in *P. tricornutum* cultures submitted to different treatments mimicking recurrent environmental variations. Grown in C, full media; -P, media without phosphate; -N, media without nitrate; -Fe, media without iron; Cold, submitted to 10 °C for 20 min; Heat, submitted to 30 °C for 20 min; Salt, submitted to 2× salt in media for 1 h. All cultures were 4 days old at the time of sampling.

Insights into the putative regulatory functions of the Pi responsive NAT-mRNA pairs. We performed a search on the available gene ontology (GO) terms of the annotated transcripts with a correlated NAT (NAT-mRNA pairs with positive and negative correlation). Amongst the most abundant GO annotation categories were genes involved in the regulation of transcription or DNA-binding (Supplementary Table S4). These included a gene coding for a heat shock factor (HSF, 34415) which was upregulated under Pi stress and was negatively correlated to its associated NAT, a gene coding for a basic-leucine zipper (bZIP) transcription factor (47278) and a PHD-finger protein (44901), both of which being upregulated under Pi stress, accompanied by their respective cognate NATs (Supplementary Table S4). A mRNA coding for Histone 3 protein (50695) also formed a concordant NAT pair, which was upregulated under Pi stress. Transporter-related mRNAs were also present forming five concordant NAT pairs, four of which were upregulated under Pi stress (Supplementary Table S4). Amongst these was a mRNA coding for a phosphate transporter (40433) which has been previously shown to be Pi deficiency-specific⁹. The present data adds another layer of information, suggesting that this transporter gene could be under the positive control of its concordant NAT and supports an intricate role for the NAT-mRNA pair in phosphate homeostasis regulation under Pi stress. Another GO annotation category well represented amongst the NAT pairs were genes coding for metabolism-related proteins, mainly protein metabolism (Supplementary Table S4). Interestingly, these formed mainly discordant NAT pairs with the mRNAs being downregulated while their associated NATs were upregulated in response to Pi depletion (Supplementary Table S4). Genes involved in signaling pathways were also detected, including two mRNAs coding for a puta-

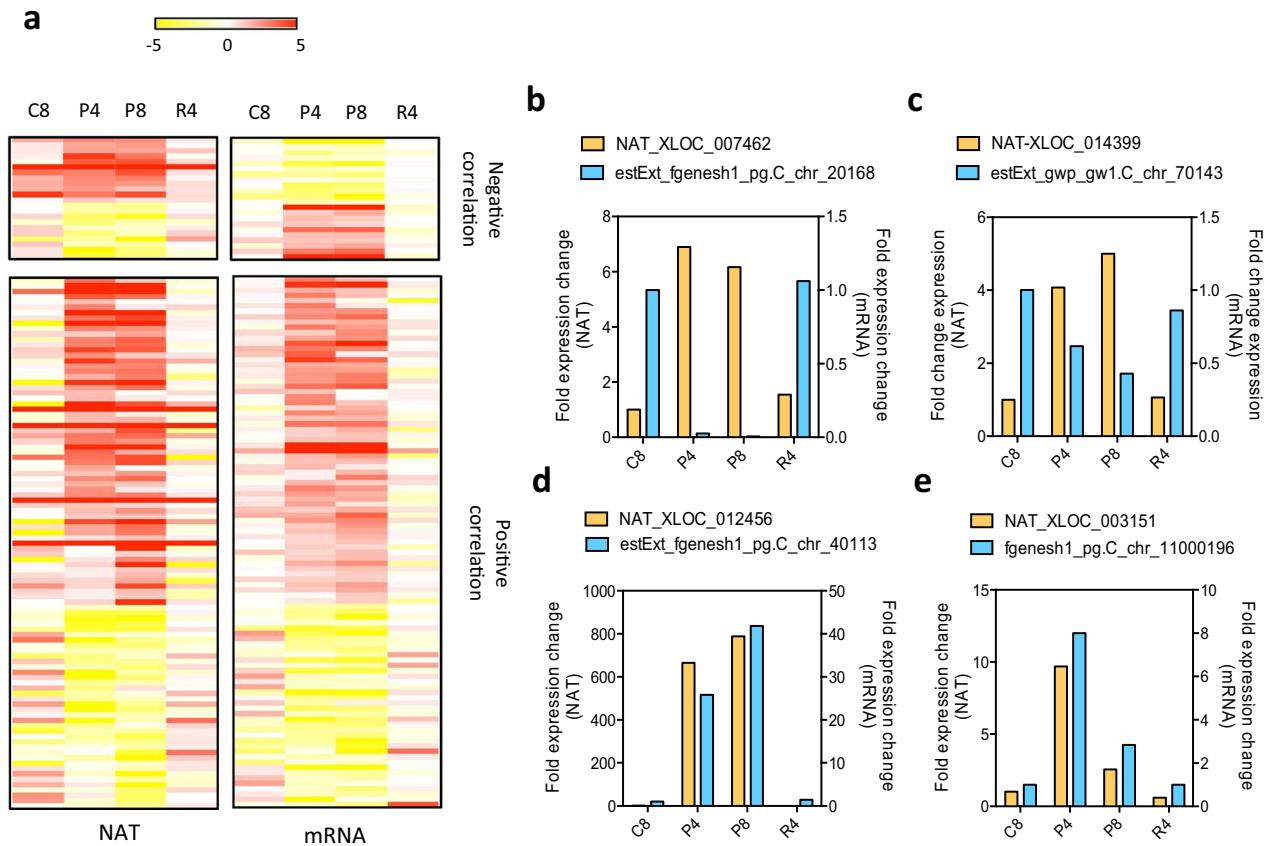


Figure 4. NAT-mRNAs pairs identified in *P. tricornutum* under Pi fluctuation. **(a)** Heat map of log₂ fold expression changes under Pi fluctuation of the NAT-mRNA pairs (with negative or positive correlation) identified in *P. tricornutum* ($r^2 > 0.6$, fold change ≥ 2 or ≤ 0.5 , P value < 0.05). **(b,c)** fold expression change of two discordant NAT-mRNA pairs; **(d,e)** fold expression change of two concordant lincNAT-mRNA pairs. P4, 4 days Pi depleted; P8, 8 days Pi depleted; R4, 4 days Pi resupply after 4 days Pi depletion; C8, 8 days control conditions.

tive protein containing a SET domain (36270) and one for a putative protein kinase (43421) (Supplementary Table S4).

Discussion

Diatoms are the most abundant eukaryotic phytoplankton in the contemporary ocean. One of the main reasons behind their ecological success is their extreme resilience to a changing environment, enabling them to quickly shift between states of exponential division under favorable conditions, to quiescent states when the environment is less favorable, and then back to exponential cell division when the environment is again favorable. Underpinning this ability is a highly responsive and dynamic molecular system regulating cellular function and metabolism in an extremely efficient way.

Long noncoding RNAs (lncRNAs) are transcripts that have been assigned many regulatory functions across the Eukarya domain of life and diatoms are excellent model systems to explore the occurrence and functional relevance of the long noncoding transcriptome in a context of stress resilience. In the present work we show that long noncoding natural antisense transcripts (NATs) are more pervasive in the *P. tricornutum* transcriptome than what was found in previous gene annotation efforts, with 21.5% of the annotated protein coding genes being overlapped by a NAT as opposed to the previously 1% reported⁴¹. This finding is in agreement with previous reports in several other eukaryotic genomes revealing the prevalence of NATs, with 27% of genomic loci in yeast⁴², close to 70% in Arabidopsis³⁰ and ~72% in mouse and human⁴³. It is likely that with the increased routine use of strand-specific deep RNA-sequencing techniques, more and more NATs will be uncovered and characterized across all eukaryotes.

NAT expression in *P. tricornutum* is specifically modulated by stress and several NATs were shown to be nutrient or environmental stress-specific (Fig. 4). This shows that the expression of NATs does not occur randomly and is instead highly regulated by external signals, strongly suggesting function. This is further supported by what has been found in other model systems such as yeasts, plants or animals^{16,17,24,44–47}. Like the previously characterized lincRNAs⁹, NATs in *P. tricornutum* have significantly lower expression levels than mRNAs (Fig. 1). This is a common cross-kingdom genomic feature of lincRNAs in general, and of NATs in particular, since these transcripts have been consistently shown to be expressed at much lower levels than mRNAs in plants^{30,46,48}, humans^{49,50} and yeast⁵¹. Unlike mRNAs that need to be converted to proteins to be active, NATs can act locally, and eventually

at the site of transcription, so it is reasonable to assume that fewer copies of NATs will be necessary to execute a function than those of mRNA. NATs can interact with DNA either directly through RNA–DNA interaction or indirectly through RNA–chromatin interactions⁴². For instance the NAT of CDKN1A, a tumor suppressor gene in humans, has been shown to recruit an enzymatic regulatory complex that induces H3K27me3 to suppress the sense promoter region⁴⁷. Another example is MAS, an antisense NAT of MADS AFFECTING FLOWERING 4 (MAF4), that mediates the recruitment of WDR5a leading to enhanced H3K4me3 in the MAF4 locus in *Arabidopsis*⁴⁶. Such local actions would not require a high number of transcripts to achieve function. In *P. tricornutum* we detected several NAT pairs that could be involved in the regulation of transcription, metabolism, signaling pathways and phosphate homeostasis in response to Pi depletion (Supplementary Table S4).

Despite being expressed at lower levels than mRNAs, NATs have been shown to be expressed in a very specific manner¹³. This is a common feature of lncRNAs and, interestingly, lncRNAs have also been shown to be comparatively more tissue-specific than mRNAs⁵². In unicellular organisms such as yeast, *Chlamydomonas* and in the present work, diatoms, NAT expression is under tight regulation by environmental signals^{44,53}. In *Chlamydomonas* NAT expression has been shown to be regulated in response to sulfur deprivation⁵³ and in yeast, NATs have been shown to be regulated by osmotic stress⁴⁴. NAT regulation by environmental stimuli has also been detected in land plants, where NATs have been reported to be induced by different stresses such as salt stress⁵⁴, light³⁰, drought and ABA⁴⁶ and cold⁵⁵.

Given the potential to form double-stranded RNA molecules, some diatom NATs may be the precursors of short interfering (si)RNAs, leading to gene silencing, as has been demonstrated in plants^{21,56}. Although no canonical miRNAs have been detected in diatoms^{57,58}, gene silencing through inverted or antisense transcripts is known to occur and is commonly used in functional studies to knock down transcription of target genes⁵⁹. This possibly operates through siRNA production which is supported by the fact that the *P. tricornutum* genome encodes one Dicer-like and an Argonaute protein⁵⁹. It has been largely assumed that sense-antisense overlapping transcripts are processed by Dicer or Dicer-like proteins to produce small interfering RNAs (siRNAs)²⁰. However, NATs have been shown to operate in many different ways besides the RNAi pathway. Overall, only a few reports have shown the involvement of NATs in the siRNA gene silencing pathway, suggesting that this is not the major mode of action of NATs³⁰.

Transcription interference which, like RNA interference also results in the negative correlation between NAT-pair expression, whereby the expression of one gene down-regulates the expression of the cognate gene on the opposite strand¹⁵, does not seem to be prevalent in *P. tricornutum* because only a small number of robust discordant NAT pairs were identified (Fig. 4a). This, nonetheless, cannot be excluded to occur in the case of the discordant NAT pairs detected in this study (Fig. 4; Supplementary Table S3). There are several modes of transcriptional interference, all of which suggest a negative regulation of NAT pairs. These include overlapped competing promoters for RNA Pol II occupation, dislodgement of one RNA Pol II by a second RNA Pol II on the opposite strand, occlusion of the initiation start site of one of the strands by RNA Pol II occupation on the other strand and, finally, collision of two RNA Pol II-transcribing opposite strands^{14,15}. The extent of this phenomenon of gene regulation by NATs is still debatable. Interestingly, in *Phaeodactylum* > 80% of the NAT pairs responsive to Pi stress are concordant, which is strikingly different from what has been demonstrated in *Arabidopsis* where concordant and discordant NAT pairs seem to be equally represented in response to light³⁰ or in yeast, where the majority of the NAT pairs identified are discordant^{60,61}. This reveals that *Phaeodactylum* Pi responsive NATs predominantly positively regulate the expression of their cognate sense protein coding genes.

In *Arabidopsis* both concordant and discordant NATs pairs have been functionally characterized. For example the NAT FLORE was found to form a discordant NAT pair with CDF5, a member of a plant specific transcription factor with a key regulatory role in connecting the circadian clock to periodic flowering-time control⁴⁵. Another example of a well characterized discordant NAT pair function is the epigenetic silencing of FLOWERING LOCUS C (FLC) by its NAT, COOLAIR¹⁸. On the other hand, MAS forms a concordant NAT pair with MAF4 and during vernalization leads to enhanced H3K4me3 in the MAF4 locus resulting in increased expression⁴⁶. A recent work using single-molecule RNA FISH revealed that the discordant sense-antisense transcripts, COOLAIR and FLC could co-occur in the same cell, but were mutually exclusive at individual loci⁶¹. It will thus be interesting to investigate the transcription patterns of the Pi responsive NAT pairs in *P. tricornutum* at the single-cell level.

The cells of multicellular eukaryotes benefit from the protective advantages of complex homeostasis. By contrast, unicellular diatoms are subject to constant environmental change which affects their entire organism. The speed at which a unicellular species can respond to its environment directly determines its chances for survival⁶². It is therefore likely that unicellular species employ an intricate regulatory network allowing for a rapid shift between processes leading to adaptation in response to environmental signals. Amongst the NAT pairs responsive to Pi fluctuations in *P. tricornutum* were mRNAs coding for proteins involved in signaling pathways, transcription regulation and phosphate transport. This suggests an important role for NATs in the orchestration of the Pi stress response.

In summary, with this work we show that NATs, with their highly specific expression profiles, could be major actors of cellular processes, particularly serving to regulate gene networks in stress adaptation pathways in *P. tricornutum*. Functional studies currently underway are likely to provide further information and novel insights on the importance of NATs in this marine diatom environmental stress response and resilience to a changing environment.

Data availability

The sequencing data used in this study is available at the Gene Expression Omnibus (GEO) database <https://www.ncbi.nlm.nih.gov/geo> (Accession No. GSE66997).

Received: 23 January 2020; Accepted: 6 August 2020

Published online: 24 August 2020

References

- Sims, P. A., Mann, D. G. & Medlin, L. K. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* **45**, 361–402 (2006).
- Reyes-Prieto, A., Weber, A. P. M. & Bhattacharya, D. The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* **41**, 147–168 (2007).
- Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Glob. Biogeochem. Cycles* **9**, 359–372. <https://doi.org/10.1029/95GB01070> (1995).
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
- Benoiston, A. S. *et al.* The evolution of diatoms and their biogeochemical functions. *Philos. Trans. R. Soc. B: Biol. Sci.* **372**, 20160397. <https://doi.org/10.1126/science.aac8721> (2017).
- Morel, F. M. M. & Price, N. M. The biogeochemical cycles of trace metals in the oceans. *Science* **300**, 944–947. <https://doi.org/10.1126/science.1083545> (2003).
- Armburst, E. V. The life of diatoms in the world's oceans. *Nature* **459**, 185–192. <https://doi.org/10.1038/nature08057> (2009).
- Bowler, C., Vardi, A. & Allen, A. E. Oceanographic and biogeochemical insights from diatom genomes. *Annu. Rev. Mar. Sci.* **2**, 333–365. <https://doi.org/10.1146/annurev-marine-120308-081051> (2010).
- Cruz de Carvalho, M. H., Sun, H.-X., Bowler, C. & Chua, N.-H. Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytol.* **210**, 497–510. <https://doi.org/10.1111/nph.13787> (2015).
- Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902> (2012).
- Bonasio, R. & Shiekhattar, R. Regulation of transcription by long noncoding RNAs. *Annu. Rev. Genet.* **48**(1), 433–455. <https://doi.org/10.1146/annurev-genet-120213-092323> (2014).
- Georg, J. & Hess, W. R. Widespread antisense transcription in prokaryotes. *Microbiol. Spectrum* <https://doi.org/10.1128/microbiolspec.RWR-0029-2018> (2018).
- Wight, M. & Werner, A. The functions of natural antisense transcripts. *Essays Biochem.* **54**, 91–101. <https://doi.org/10.1042/BST0381144> (2013).
- Rosikiewicz, W. & Makalowska, I. Biological functions of natural antisense transcripts. *Acta Biochim. Pol.* https://doi.org/10.18388/abp.2016_1350 (2017).
- Shearwin, K., Callen, B. & Egan, J. Transcriptional interference: a crash course. *Trends Genet.* **21**, 339–345 (2005).
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. & Stutz, F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131**, 706–717. <https://doi.org/10.1016/j.cell.2007.09.014> (2007).
- Camblong, J. *et al.* Trans-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*. *Genes Dev.* **23**, 1534–1545. <https://doi.org/10.1101/gad.522509> (2009).
- Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* **462**, 1–5. <https://doi.org/10.1038/nature08618> (2009).
- Kotake, Y. N. T. *et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* **30**, 1956–1962. <https://doi.org/10.1038/onc.2010.568> (2011).
- Zhang, X. *et al.* Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biol.* **13**, R20. <https://doi.org/10.1186/gb-2012-13-3-r20> (2012).
- Zhang, X. *et al.* Mechanisms of small RNA generation from cis-NATs in response to environmental and developmental cues. *Mol. Plant* **6**, 704–715. <https://doi.org/10.1093/mp/sst051> (2013).
- Napoli, S., Piccinelli, V., Mapelli, S. N., Pisignano, G. & Catapano, C. V. Natural antisense transcripts drive a regulatory cascade controlling c-MYC transcription. *RNA Biol.* **14**, 1742–1755. <https://doi.org/10.1080/15476286.2017.1356564> (2017).
- Morrissy, A. S., Griffith, M. & Marra, M. A. Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res.* **21**, 1203–1212. <https://doi.org/10.1101/gr.113431.110> (2011).
- Jabnousse, M. *et al.* A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *Plant Cell* **25**, 4166–4182. <https://doi.org/10.1105/tpc.113.116251> (2013).
- Deforges, J. *et al.* Control of cognate sense mRNA translation by cis-natural antisense RNAs. *Plant Physiol.* <https://doi.org/10.1104/pp.19.00043> (2019).
- Lavorgna, G. *et al.* In search of antisense. *Trends Biochem. Sci.* **29**, 88–94. <https://doi.org/10.1016/j.tibs.2003.12.002> (2004).
- Werner, A. Natural antisense transcripts. *RNA Biol.* **2**, 53–62. <https://doi.org/10.4161/rna.2.2.1852> (2005).
- Charon, C., Moreno, A. B., Bardou, F. & Crespi, M. Non-protein-coding RNAs and their interacting RNA-binding proteins in the plant cell nucleus. *Mol. Plant* **3**, 729–739. <https://doi.org/10.1093/mp/ssq037> (2010).
- van Dijk, E. L. *et al.* XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **114**, 1–6. <https://doi.org/10.1038/nature10118> (2011).
- Wang, H. *et al.* Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Genome Res.* **24**, 444–453. <https://doi.org/10.1101/gr.165555.113> (2014).
- Huang, R. *et al.* A potential role for epigenetic processes in the acclimation response to elevated pCO₂ in the model diatom *Phaeodactylum tricornutum*. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2018.03342> (2019).
- Guillard, R. R. & Ryther, J. H. Studies of marine planktonic diatoms: I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Can. J. Microbiol.* **8**, 229–239. <https://doi.org/10.1139/m62-029> (1962).
- Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53. <https://doi.org/10.1016/j.cell.2012.06.041> (2013).
- Siaut, M. *et al.* Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**, 23–35. <https://doi.org/10.1016/j.gene.2007.05.022> (2007).
- Liu, J. *et al.* Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell* **24**, 4333–4345. <https://doi.org/10.1105/tpc.112.102855> (2012).
- Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94. <https://doi.org/10.1006/jmbi.1997.0951> (1997).
- Niazi, F. & Valadkhan, S. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* **18**, 825–843. <https://doi.org/10.1261/rna.029520.111> (2012).
- De Martino, A., Meichenin, A., Shi, J., Pan, K. & Bowler, C. Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions 1. *J. Phycol.* **43**, 992–1009. <https://doi.org/10.1111/j.1529-8817.2007.00384.x> (2007).

40. De Martino, A., Amato, A. & Bowler, C. Mitosis in diatoms: rediscovering an old model for cell division. *BioEssays* **31**, 874–884. <https://doi.org/10.1002/bies.200900007> (2009).
41. Rastogi, A. et al. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci. Rep.* 1–14 (2018).
42. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell. Biol.* **10**, 637–643. <https://doi.org/10.1038/nrm2738> (2009).
43. Katayama, S. et al. Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566. <https://doi.org/10.1126/science.1112009> (2005).
44. Nadal-Ribelles, M. et al. Control of Cdc28 CDK1 by a stress-induced lncRNA. *Mol. Cell* **53**, 549–561. <https://doi.org/10.1016/j.molcel.2014.01.006> (2014).
45. Henriques, R. et al. The antiphasic regulatory module comprising CDF5 and its antisense RNA FLORE links the circadian clock to photoperiodic flowering. *New Phytol.* **216**, 854–867. <https://doi.org/10.1111/j.1365-3113.2007.03309.x> (2017).
46. Zhao, X. et al. Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat. Commun.* **9**, 362. <https://doi.org/10.1105/tpc.107.055046> (2018).
47. Morris, K. V., Santoso, S., Turner, A.-M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.* **4**, e1000258. <https://doi.org/10.1371/journal.pgen.1000258.s006> (2008).
48. Bai, Y. et al. Identification and characterization of pineapple leaf lncRNAs in crassulacean acid metabolism (CAM) photosynthesis pathway. *Sci. Rep.* **9**, 4333. <https://doi.org/10.1073/pnas.1015452108> (2019).
49. Khalil, A. M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667–11672 (2009).
50. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789. <https://doi.org/10.1101/gr.132159.111> (2012).
51. Uhler, J. P., Hertel, C. & Svejstrup, J. Q. A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc. Nat. Acad. Sci. USA* **104**, 8011–8016. <https://doi.org/10.1073/pnas.0702431104> (2007).
52. Cabili, M. N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927. <https://doi.org/10.1101/gad.17446611> (2011).
53. Li, H. et al. Genome-wide long non-coding RNA screening, identification and characterization in a model microorganism. *Sci. Rep.* <https://doi.org/10.1038/srep34109> (2016).
54. Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R. & Zhu, J.-K. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**, 1279–1291. <https://doi.org/10.1016/j.cell.2005.11.035> (2005).
55. Legeay, M., Aubourg, S., Renou, J.-P. & Duval, B. Large scale study of anti-sense regulation by differential network analysis. *BMC Syst. Biol.* <https://doi.org/10.1186/s12918-018-0613-7> (2018).
56. Celton, J.-M. et al. Widespread anti-sense transcription in apple is correlated with siRNA production and indicates a large potential for transcriptional and/or post-transcriptional control. *New Phytol.* **203**, 287–299. <https://doi.org/10.1111/nph.12787> (2014).
57. Lopez-Gomollon, S. et al. Global discovery and characterization of small non-coding RNAs in marine microalgae. *BMC Genom.* **15**, 697. <https://doi.org/10.1186/1471-2164-15-697> (2014).
58. Rogato, A. et al. The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricornutum*. *BMC Genom.* **15**, 698. <https://doi.org/10.1186/1471-2164-15-698> (2014).
59. De Riso, V. et al. Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucl. Acids Res.* **37**, e96–e96. <https://doi.org/10.1093/nar/gkp448> (2009).
60. Xu, Z. et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037. <https://doi.org/10.1093/nar/gkm683> (2009).
61. Rosa, S., Duncan, S. & Dean, C. Mutually exclusive sense–antisense transcription at FLC facilitates environmentally induced gene repression. *Nat. Commun.* **7**, 1–7. <https://doi.org/10.1038/ncomms13031> (2016).
62. de Nadal, E., Ammerer, G. & Posas, F. Controlling gene expression in response to stress. *Nat. Rev. Gen.* **594**, 100. <https://doi.org/10.1038/nrg3055> (2011).

Acknowledgements

The authors wish to thank Hai-Xi Sun, Federico Ibarbalz and Ouardia Ait-Mohamed for the bioinformatic support and suggestions, and Damien Le Moigne for his help with some of the RT-qPCR reactions. This work was supported by the Seventh Research Program of the European Union FP7/2007–2013 under Marie Curie Grant PEOF-GA-301466 (to MHCC) and also benefited from funding from the DiaLincs award from the French Government (ANR 19-CE43-0011-01) (to MHCC). CB acknowledges funding from the European Research Council Advanced Award Diatomite and Diatomic under the European Union’s Horizon 2020 research and innovation programme (grant agreement Nos. 294823 and 835067), as well as the French Government “Investissements d’Avenir” programs MEMO LIFE (ANR-10-LABX-54), Université de Recherche Paris Sciences et Lettres (Université PSL) (ANR-1253 11-IDEX-0001-02), and OCEANOMICS (ANR-11-BTBR-0008).

Author contributions

MHCC designed the research, performed the research, analysed the data and wrote the manuscript. CB supplied resources, read and edited the manuscript. Both authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-71002-0>.

Correspondence and requests for materials should be addressed to M.H.C.d.

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

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020