Diurnal transcript profiling of the diatom *Seminavis robusta* reveals adaptations to a benthic lifestyle

Gust Bilcke[1,2,3,4*], Cristina Maria Osuna-Cruz[1,2,5*], Marta Santana Silva[1,2], Nicole Poulsen[6], Sofie D'hondt[3], Petra Bulankova[1,2], Wim Vyverman[3**†], Lieven De Veylder[1,2**†], Klaas Vandepoele[1,2,5**†].

1. Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 71, 9052, Ghent, Belgium
2. VIB Center for Plant Systems Biology, Technologiepark 71, 9052, Ghent, Belgium
3. Protistology and Aquatic Ecology, Department of Biology, Ghent University, 9000, Ghent, Belgium
4. Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000, Ghent, Belgium
5. Bioinformatics Institute Ghent, Ghent University, Technologiepark 71, 9052, Ghent, Belgium
6. B CUBE Center for Molecular Bioengineering, Technical University of Dresden, Tatzberg 41, 01307, Dresden, Germany

* Shared first authorship

** Shared senior authorship

Abstract

Coastal regions contribute an estimated 20% of annual gross primary production in the oceans, despite occupying only 0.03% of their surface area. Diatoms frequently dominate coastal sediments, where they experience large variations in light regime resulting from the interplay of diurnal and tidal cycles. Here, we report on an extensive diurnal transcript profiling experiment of the motile benthic diatom *Seminavis robusta*. Nearly 90% (23,328) of expressed protein-coding genes and 66.9% (1124) of expressed long intergenic non-coding RNAs (lincRNAs) showed significant expression oscillations and are predominantly phasing at night with a periodicity of 24h. Phylostratigraphic analysis found that rhythmic genes are enriched in deeply conserved genes, while diatom-specific genes are predominantly associated with midnight expression. Integration of genetic and physiological cell cycle markers with silica depletion data revealed potential new silica cell wall associated gene families specific to diatoms. Additionally, we observed 1752 genes with a remarkable semidiurnal (12-h) periodicity, while the expansion of putative circadian transcription factors may reflect adaptations to cope with highly unpredictable external conditions. Taken together, our results provide new insights into the adaptations of diatoms to the benthic environment and serve as a valuable resource for diurnal regulation in photosynthetic eukaryotes.

Introduction

Most organisms are exposed to a daily alternation of light and darkness as the result of the earth's rotation. This has led to the evolution of rhythmic cellular programs that are controlled by environmental cues as well as endogenous circadian clocks (Mittag, 2001). In particular, coordination with diurnal cycles is crucial for photosynthetic organisms, whose metabolism relies heavily on the availability of sunlight (de Los Reyes et al., 2017). Compared to multicellular plants, rhythmicity is especially pronounced in microalgae, where up to 80% of the transcriptome is subject to rhythmic oscillations (Zones et al., 2015; Ferrari et al., 2019).

Diatoms are prominent stramenopile microalgae in marine and freshwater environments. They account for up to 20% of annual global primary production and play an important role in biogeochemical cycling of nutrients in aquatic ecosystems, including silica (Malviya et al., 2016). Diatom cells are characterized by a unique bipartite silica cell wall, called the frustule. Although cell wall structure and morphology is widely used in diatom taxonomy and biodiversity studies, silica biomineralization is still poorly understood (Hildebrand and Lerch, 2015).

To date, only a handful of studies have explored rhythmic gene expression in diatoms, generally with a particular focus on the response of the diurnal transcriptome to nutrient limitation (Ashworth et al., 2013; Goldman et al., 2019; Chauton et al., 2013; Smith et al., 2016). The most extensive accounts come from the pennate diatom *Phaeodactylum tricornutum,* whose diurnal coordination of carbohydrate and cell cycle pathways was shown to be strongly affected by iron availability and for which the first diatom circadian regulator was recently characterized (Chauton et al., 2013; Smith et al., 2016; Annunziata et al., 2019). Furthermore, studies thus far have focused solely on planktonic diatom species, which typically inhabit well-mixed water columns. In contrast, no data are available on diurnal variation in gene expression of the extraordinarily diverse group of benthic pennate diatoms, dominating some of the most productive ecosystems on the globe (Nakov et al., 2019; Pinckney, 2018; Cahoon, 1999). Yet, benthic ecosystems are exposed to dynamic and often extreme light regimes, resulting from the combined effects of tidal phasing, sediment composition and water column turbidity (Barnett et al., 2015; Underwood and Kromkamp, 1999; Hope et al., 2020). Under laboratory conditions, benthic diatom species generally show maximal growth rates at irradiances of 10-50 µmol photons $m^{-2}s^{-1}$, similar to conditions of shading that are periodically encountered in benthic communities (Admiraal, 1977; Peña and

de la Peña, 2007; Palmisano et al., 2004; Cahoon et al., 1993; Gattuso et al., 2006). Not only light availability varies over the course of the diel cycle, also nutrient and oxygen availability in the sediment show rhythmic patterns (Eyre and Ferguson, 2002; Haro et al., 2019), adding to the complexity and dynamic nature of benthic habitats. Motile species from the raphid pennate diatom clade are particularly successful in these environments and use the secretion of mucilaginous exopolysaccharides (EPS) to actively migrate through the sediment at dusk and dawn to maximize photosynthesis while avoiding excessive grazing (Longphuirt et al., 2006; Haro et al., 2019; Underwood and Paterson, 2003; Hanlon et al., 2006). This suggests that benthic diatoms employ elaborate mechanisms to track the diurnal cycle in order to optimize photosynthesis, nutrient uptake and survival.

Long non-coding RNAs (lncRNAs) are an emerging class of transcripts longer than 200 nucleotides that are not translated into protein (Yamada, 2017). Advances in whole-transcriptome sequencing have revealed that a considerable portion of the non-coding DNA in plant and animal genomes make up lncRNAs, which are expressed at low and mostly tissue-specific levels (Yamada, 2017; Deniz and Erman, 2017). Despite lncRNAs still being largely unexplored, they appear to play a pivotal role in gene expression regulation and are increasingly implicated in various biological processes such as environmental stress and defense (Karlik et al., 2019). In addition, numerous lncRNAs are involved in diurnal regulation or the circadian clock in both plants and animals (Hazen et al., 2009; Hardeland, 2018). Although intergenic lncRNAs (lincRNAs) have been described in diatoms in the context of environmental stimuli (Carvalho et al., 2016; Huang et al., 2019), rhythmic oscillations in these genes have not been assessed yet.

In this study, we investigated the diurnal regulation of the benthic model diatom Seminavis robusta using RNA-sequencing (RNA-seq). High-resolution sampling of a two-day (48h) time series on three separate occasions allowed us to identify genes with a stable recurring expression oscillation. Next to the analysis of protein-coding genes, lincRNAs were annotated in the *S. robusta* genome to identify the first rhythmic lincRNAs in microalgae. We further performed phylogenomic and comparative transcriptomic analyses to answer outstanding questions regarding the conservation of rhythmic gene expression in benthic versus planktonic diatoms, the evolutionary history of rhythmic gene phasing and periodicity, and synchronization of organelle division to the diurnal cycle. Finally, we took advantage of available silicon depletion transcriptome datasets to uncover several new gene families with a potential role in diatom cell wall synthesis.

Results and discussion

The majority of *Seminavis robusta* genes exhibit a rhythmic expression pattern under diurnal growth conditions

We maintained *S. robusta* cultures in a 12h-light/12h-dark photoperiod to identify genes exhibiting periodic expression. Three replicate cultures were sampled over a 48-h time course at 4-h intervals beginning at ZT2 (Zeitgeber Time = hours since the onset of illumination), resulting in 12 time points over two complete diel cycles (***Figure 1A***). After RNA sequencing, more than one billion reads were obtained with an average of 28 million reads per sample and an average mapping rate of 87.15% (***Supplemental Data Set 1***). Multidimensional scaling revealed a concerted diurnal progression in gene expression, with samples harvested at the same time of each day clustering together, regardless of the experimental repeat (***Figure 1B***). Genes having significant oscillatory gene expression (q-value < 0.05) were classified into clusters by period (24h or 12h) and phase (ZT2, ZT6, ZT10, ZT14, ZT18 and ZT22) as well as into more general categories (day/night and dawn/dusk), as explained in the Methods section (***Figure 1C, Supplemental Figure 1-2***).

A total of 23,328 protein-coding genes showed a significant rhythmic expression pattern, which corresponds to 89% of protein-coding genes with a detectable diurnal expression and 64% of protein-coding genes encoded in the *S. robusta* genome (***Supplemental Table 1***). While similar high proportions of rhythmic genes have been described for various unicellular algae (Zones et al., 2015; Poliner et al., 2015; Monnier et al., 2010; Ferrari et al., 2019), the planktonic diatoms *P. tricornutum* and *T. pseudonana* displayed lower proportions in nutrient replete conditions, ranging between 28% and 55% (Chauton et al., 2013; Goldman et al., 2019; Ashworth et al., 2013), and this was supported by the re-analysis of a comparable, 24-h *P. tricornutum* RNA-seq dataset (***Supplemental Figure 3***) (Smith et al., 2016). The high level of rhythmicity in *S. robusta* compared to other diatom data sets might be explained by differences in experimental procedures and analyses, or could reflect an adaptation of benthic diatoms to cyclic changes in the abiotic conditions of their environment (Barnett et al., 2015; Saburova and Polikarpov, 2003; Eyre and Ferguson, 2002; Haro et al., 2019). Noteworthy, these pronounced transcriptomic oscillations occurred under a relatively low experimental irradiance of 30 µmol photons $m^{-2}s^{-1}$, allowing diurnal coordination even under conditions of shading by the overlying water column characteristic to the benthos (Gattuso et al., 2006; Underwood and Kromkamp, 1999).

Most rhythmic protein-coding genes displayed a 24-h (92.5%) or 12-h (7.5%) periodicity, and similar to reports in other diatoms, were predominantly phasing at dusk and at night (***Figure 1C, Supplemental Data Set 2***) (Smith et al., 2016; Ashworth et al., 2013; Goldman et al., 2019). We cross-referenced the 10,145 protein-coding genes that were not expressed during the diurnal cycle in a publicly available *S. robusta* expression atlas that reports on transcriptional activity during sexual reproduction, abiotic stress and diatom-bacterial interactions (Osuna-Cruz et al., 2020). More than half (5,315) of non-diurnal genes were transcriptionally active in at least one condition, although at lower expression levels than diurnally expressed genes (***Supplemental Figure 4A***). Calculation of Kendall's Tau as a measure of condition specificity showed that non-diurnal genes were expressed in a more condition-specific manner than diurnally expressed genes (Average Tau 0.82 and 0.49 respectively, Wilcoxon rank-sum test $p < 2.2e-16$) (***Supplemental Figure 4B***). Hence, while most diurnally expressed genes display a broad expression pattern not associated with a specific condition, genes that were not expressed during our diurnal experiment seem to be relevant under specific conditions such as sexual reproduction (***Supplemental Figure 4C***).

We next investigated the diurnal expression of 7,117 high-confidence lincRNAs that we annotated in the *S. robusta* genome leveraging different RNA-seq datasets (see Methods). The number of *S. robusta* lincRNAs is in line with the diversity of lincRNAs in the brown alga *Saccharina japonica (Yang et al., 2020)*, but exceeds that in the diatom *P. tricornutum* by about 5-fold (Carvalho et al., 2016). From a total of 1,680 diurnally expressed lincRNAs, 1,124 showed significant oscillations (66.9%, ***Figure 1C***). Similar to the protein-coding genes, rhythmic lincRNAs displayed repetitive peaks in expression associated with a specific time of the day and were predominantly phasing at night and at dusk (***Figure 1C***). Applying the guilt-by-association principle using co-expression networks and functional information of protein-coding genes, we could attribute a biological process to 250 rhythmic lincRNA genes, implying a cooperation between protein-coding and lincRNA genes (Liesecke et al., 2018) (***Supplemental Data Set 3***). Rhythmic lincRNAs were predicted to function in photosynthesis, carbohydrate metabolism, translation, proteolysis, cell cycle, regulation of transcription and intracellular signal transduction. In contrast to protein-coding genes, the majority of all lincRNAs (5,437, 76.4%) was not diurnally expressed. In general, lincRNAs were transcribed at lower levels than protein-coding genes (average maximal log2(TPM) of 3.0 and 5.1 respectively, Wilcoxon rank-sum test $p < 2.2e-16$, ***Supplemental Figure 4A)*** and showed higher condition specificity as expected (Average Tau 0.85 and 0.54 respectively, Wilcoxon rank-sum test $p < 2.2e-16$, ***Supplemental Figure 4B***). Thus, although a

considerable set of lincRNAs plays a role in diurnal conditions, the majority of lincRNAs is rather associated with specific stress environments, as previously reported in diatoms (***Supplemental Figure 4C***) *(Carvalho et al., 2016; Huang et al., 2019)*. Finally, only 103 lincRNAs displayed conservation in at least one other diatom species, of which 19 were significantly rhythmic. These results suggest a poor conservation of rhythmic lincRNAs between distantly related diatom species or a strong divergence in primary nucleotide sequence, as is widely observed for lncRNAs, interfering with reliable homology detection (Zampetaki et al., 2019).

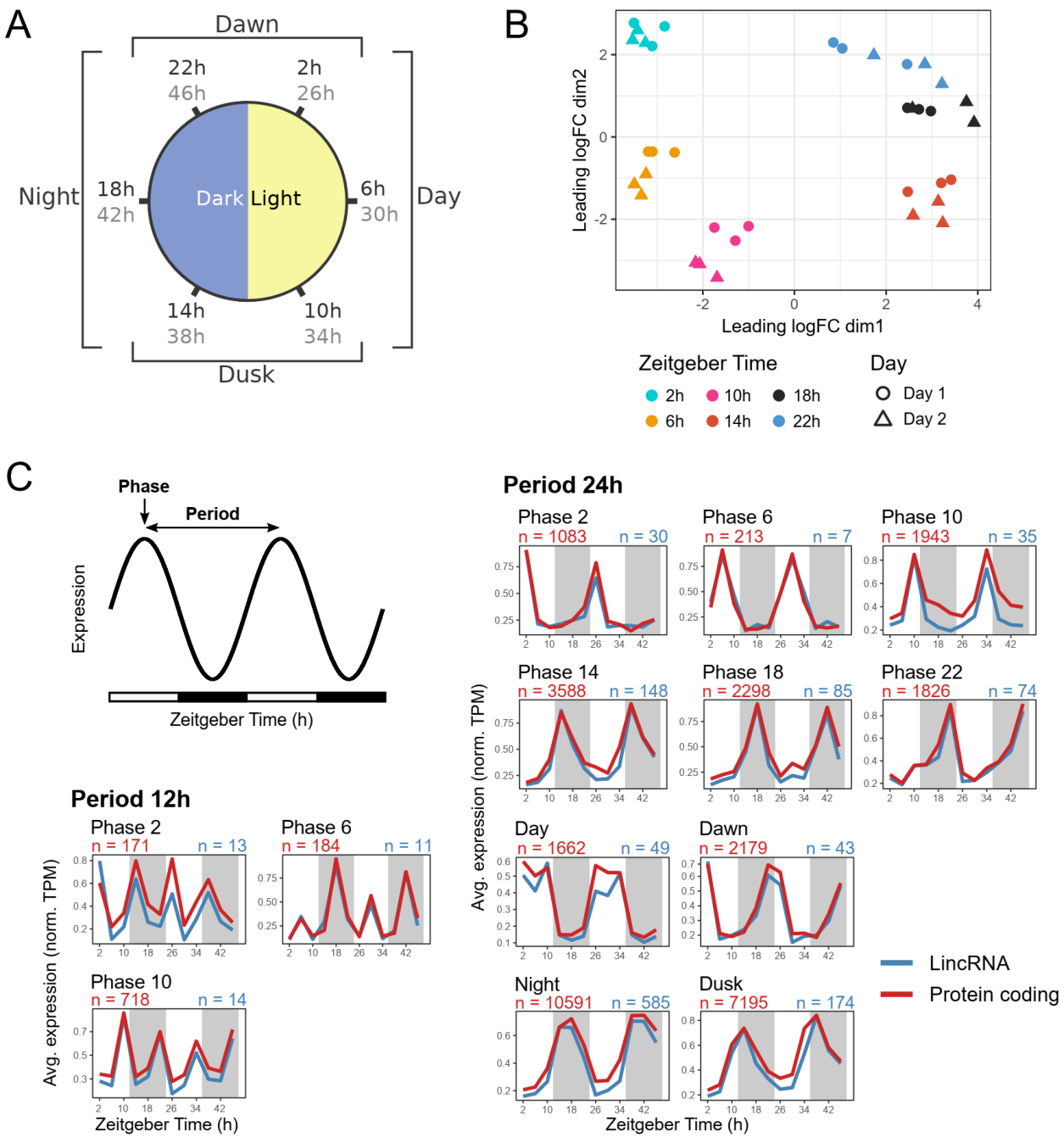

Figure 1: Overview of the diurnal gene expression experiment and classification of rhythmic genes in *S. robusta*. (A) Schematic representation of the experimental setup and classification. The blue and yellow shaded semicircles represent the dark and light part of the 12h/12h cycle, respectively. Samples acquired in the first 24h are printed in black, while the corresponding samples from 24h-48h are labelled in grey. Genes were classified in clusters for each phase (peak expression) as well as broader categories "Day", "Night", "Dawn" and "Dusk". **(B)** Multidimensional scaling (MDS) plot of 36 sequenced samples for the top 500 most differential genes of the diurnal time series. Colours indicate time of day (Zeitgeber Time, time since first illumination) and shapes represent if a sample belongs to the first or second day of the experiment. **Inset in (C)**: nomenclature of rhythmic expression waveforms. A waveform is shown that represents the fitted expression of a hypothetical gene during the time series, with the day and night treatment indicated in white and black bars below the waveform. Terms used throughout the paper to discuss properties of the waveform are indicated with arrows. **(C)** Average expression for 24-h and 12-h period genes, organized in temporal classes by their predicted phase, i.e. the Zeitgeber Time at which expression peaks. Expression in transcripts per million (TPM) was normalized to a maximum of 1. "n" represents the number of genes in each cluster, printed in red and blue for protein-coding genes and lincRNAs, respectively.

Evolutionary analysis of rhythmic genes points towards ancient daytime functions and diatom-specific expression at night

Phylostratigraphic analysis, which explores protein homology in a wide set of species to date the origin of specific genes, revealed that rhythmic protein-coding genes were enriched in eukaryote and stramenopile age, while non-significantly rhythmic genes displayed enrichment in species-specific genes (***Figure 2A***). The latter group of genes was mainly enriched in genes upregulated under different stress conditions in the *S. robusta* expression atlas (Osuna-Cruz et al., 2020), while the set of rhythmic genes was predominantly enriched in downregulated genes under these conditions (***Figure 2B***). These results suggest that while phylogenetically old rhythmic genes play a role in basal diurnal functions, young and non-rhythmic genes are more important in specific environmental contexts that are not part of the diurnal regime, a trend also observed in land plants and several microalgae (Ferrari et al., 2019). Considering the period of rhythmic genes, genes following a 24-h period were associated with older phylostrata than genes following a 12-h period (***Figure 2A***). Moreover, 24-h periodic genes peaking at ZT2, ZT6, ZT10 and ZT14 were significantly enriched in eukaryotic age, whereas genes peaking at ZT18 and ZT22 were associated with diatom and/or species-specific ages as well as genes with unknown functions (***Figure 2A, Supplemental Data Set 4***). Notably, no such association between phasing and gene age was found in glaucophyta, red or green algae (Ferrari et al., 2019). Taken together, genes phasing during the day and at the start of the night tend to perform deeply conserved eukaryotic functions, while previously undescribed processes specific to diatom biology appear to be occurring at midnight and during the second half of the night.

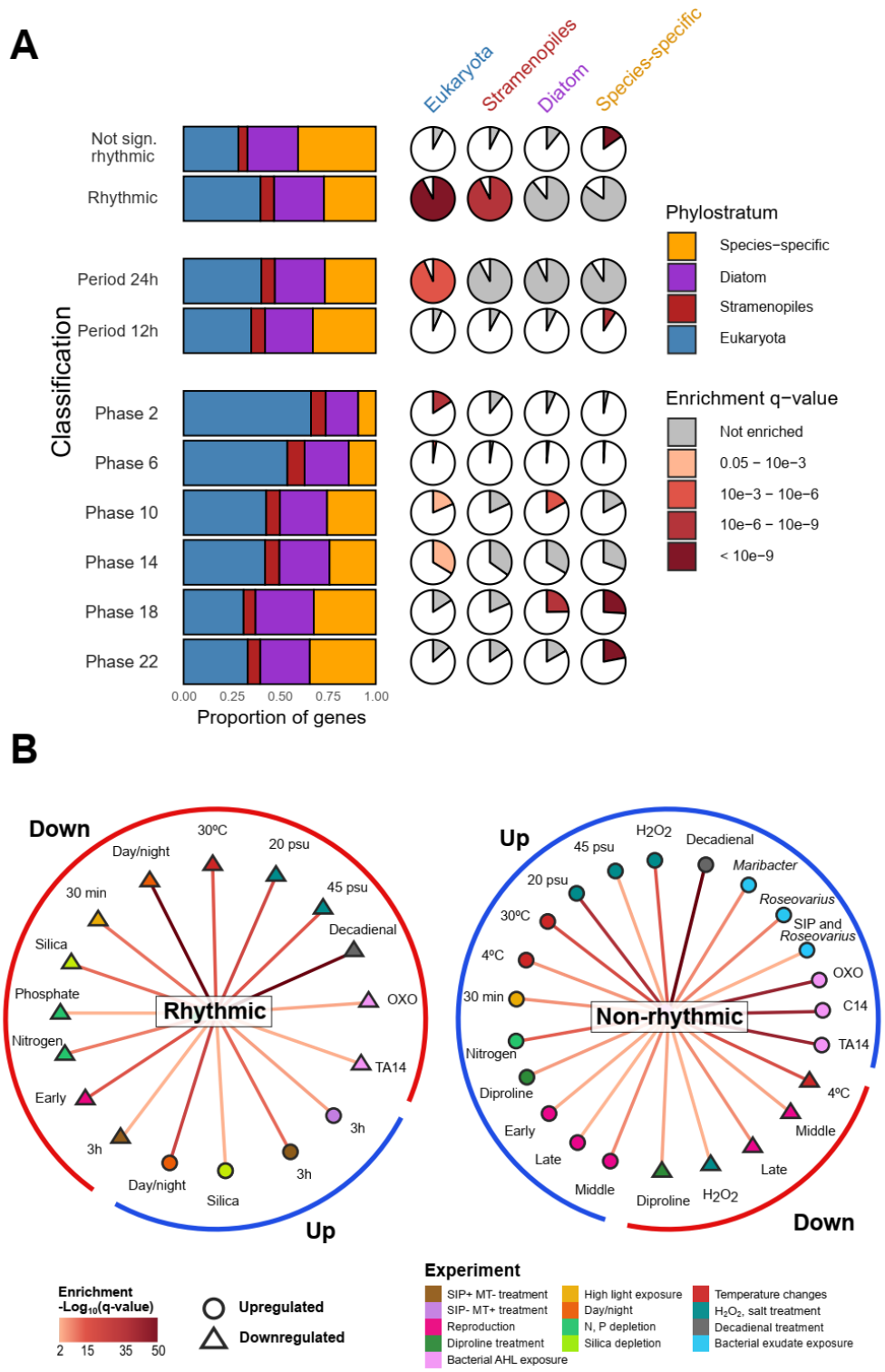

Figure 2: Phylostratigraphy and differential expression of rhythmic and non-significantly rhythmic protein-coding genes. (A) Phylostratigraphic trends of different rhythmic gene sets. Horizontal labels refer to the gene set, including rhythmicity (top), period (middle) and phase (bottom). The fraction of genes of each phylostratum in each gene set is displayed by different colours in the barplot. Pie charts show enrichment of phylostrata (columns) for each gene set (rows). The shaded area of the pie chart represents the proportion of genes that is in this gene set, while its colour displays the q-value of enrichment. Enrichment results for phylostrata can be found in ***Supplemental Data Set 4.*** **(B)** Network showing enrichment of differentially expressed genes over a wide range of conditions in the cluster of rhythmic (left) versus non-significantly rhythmic genes (right). Enrichment in upregulated genes is represented by circles and downregulated genes by triangles. The color of circles/triangles denotes the experiment, while the significance of the enrichment for a specified condition is highlighted by a color gradient from light red to dark red in –log10(q-value) scale in the edge of the network. Red and blue lines in the outer circle summarize the proportion of down- and upregulated conditions respectively.

Rhythmic gene expression is relatively conserved between two pennate species with different lifestyles

A comparative analysis based on orthology was carried out to explore the conservation of gene expression phasing in diatoms. We restricted our comparison to the nutrient replete dataset from the *P. tricornutum* iron availability study of Smith et al. (Smith et al., 2016), as the *T. pseudonana* diurnal studies only cover two time points per day (Ashworth et al., 2013; Goldman et al., 2019). We inspected the presence of phase shifts between *S. robusta* and *P. tricornutum* in 575 rhythmic one-to-one ortholog pairs with a 24-h period and found a significant association between the phasing of orthologs (***Figure 3A***). The majority (60%) of *P. tricornutum genes* phased at a comparable time of the day in both species, i.e. 1h earlier or 3h later than their *S. robusta* orthologs (Chi-square Goodness of Fit test, $p < 2.2e-16$, std. res. of 14.7 and 2.8 respectively, ***Figure 3A,B***). Inspecting the ortholog pairs classified into each combination of phases revealed that genes with small phase shifts are involved in metabolic processes such as ribosome biogenesis, photosynthesis, proteasomal degradation and respiration, while cell cycle regulators were generally expressed 5h later in *S. robusta* (***Figure 3C***). The similar phasing between the lowly silicified planktonic diatom *P. tricornutum* and the motile benthic *S. robusta,* despite considerable differences in lifestyle, indicates that the temporal regulation of basal metabolic pathways is relatively conserved.

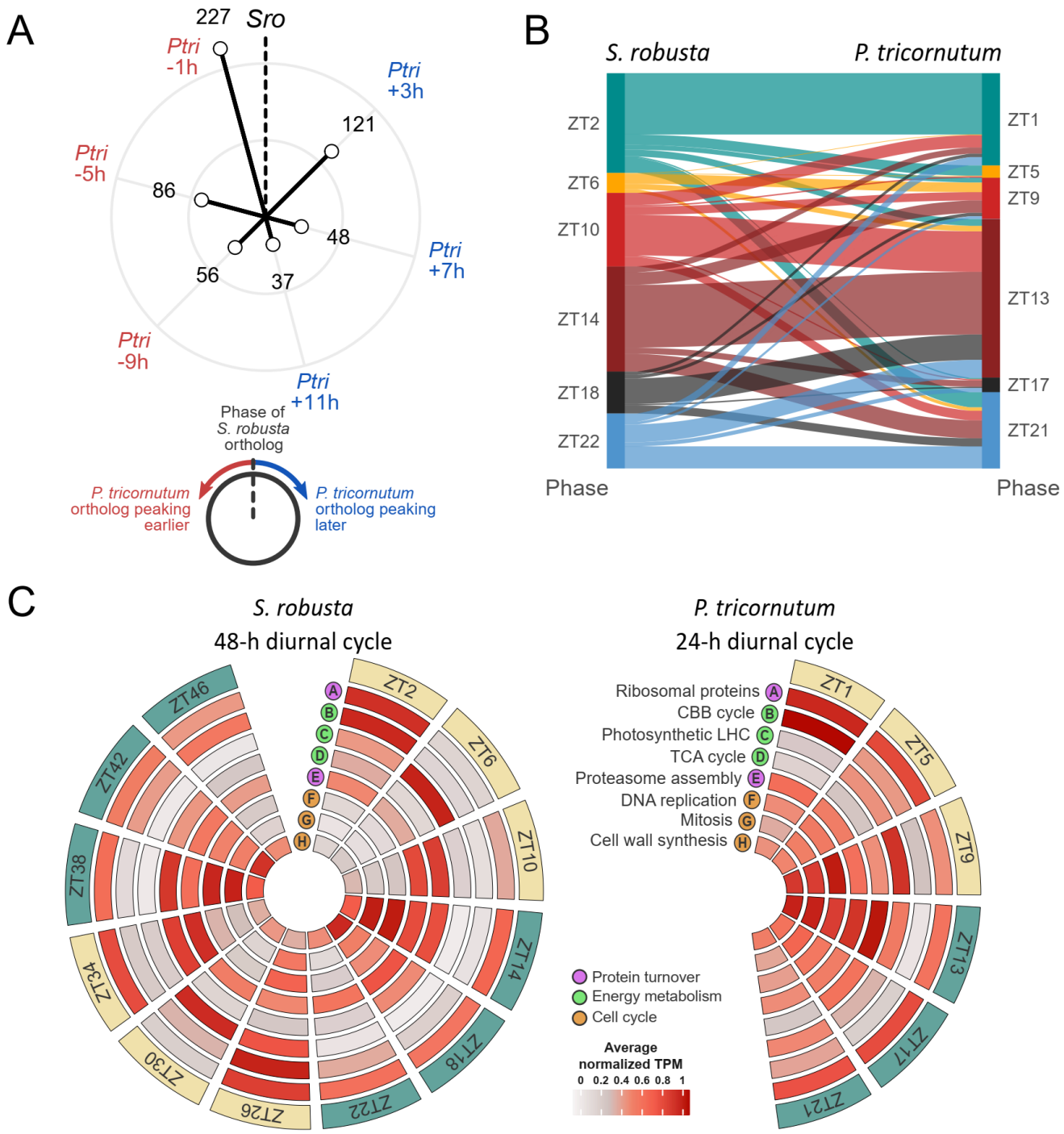

Figure 3: Comparative analysis diurnal expression of *S. robusta* and *P. tricornutum*. (A) Circular bar chart showing the timing of the phase of 575 *P. tricornutum ("Ptri")* rhythmic genes compared to their one-to-one *S. robusta ("Sro")* orthologs. The phase of the reference *S. robusta* ortholog is indicated with a dotted line, while the shift in phase of the corresponding *P. tricornutum* ortholog is shown on the outside of the circle. Each dataset consists of six time points per cycle in 4-h intervals which are offset by 1h, limiting the number of possible phase shifts to six. The number of genes for each phase shift is shown by the length of each bar and is printed in black numerals. Shifts in which the *P. tricornutum* gene is peaking relatively earlier or later than its *S. robusta* counterpart are indicated in red or blue respectively. **(B)** Sankey diagram of the phase of 575 one-to-one orthologs in *S. robusta* versus *P. tricornutum*. The number of ortholog pairs within a given combination of phases is shown by the width of the line connecting the phases. The phase of the *S. robusta* ortholog is coded by their colour. **(C)** Circular heatmap plots displaying in every row the average normalized expression (transcripts per million, TPM) of diurnally expressed genes belonging to selected processes in *S. robusta* and *P. tricornutum*, in a color gradient ranging from grey (0) to dark red (1). Gene lists for each of these processes can be retrieved from ***Supplemental Data Set 5***. In every column the time point in the diurnal cycle (Zeitgeber Time = time since illumination) is shown (green = dark, yellow = light). The *P. tricornutum* time point ZT25 represents the equivalent ZT1 in this figure. CBB refers to Calvin-Benson-Bassham, LHC to light harvesting complex and TCA to tricarboxylic acid cycle.

Daytime anabolic and nocturnal catabolic programs

Comparative orthology analysis with *P. tricornutum* indicated a generally similar timing of metabolic processes between both species (***Figure 3C***). To gather a more detailed overview of the respective timing of specific processes, we next analyzed the diurnal expression of well-known metabolic pathways involved in protein turnover, photosynthesis and respiration in the diurnal *S. robusta* protein-coding gene dataset.

Rhythmic expression of genes related to protein turnover was strongly coordinated in *S. robusta* (***Figure 4A, Supplemental Figure*** **5**). While tRNA-loading aminoacyl tRNA-ligases peaked early at ZT22, translation elongation and initiation factors were phasing at ZT2 (***Figure 4A***). Following protein synthesis, we found that both vesicle-mediated transport and proteasome-mediated proteolysis were enriched at dusk (***Figure 4A***). The phasing of proteasomal proteolytic machinery at dusk might point towards cell cycle or circadian clock control by ubiquitination (Yu et al., 2008; Stojkovic et al., 2014), or may simply reflect cyclic protein turnover in which proteins are translated during the day and are recycled at night, as reported in other diatoms (Ashworth et al., 2013; Chauton et al., 2013; Smith et al., 2016). Interestingly, we observed an enrichment of 12-h periodic genes for the GO terms “translation” and “proteolysis” (31 and 64 genes respectively), the former phasing at ZT2 as well as ZT14, the latter at the second half of each day and night (***Figure 4A***). While the set of 24-h period genes mainly consisted of proteasome-mediated proteolytic factors, 12-h periodic proteolytic genes contained putative lysosomal peptidases, pointing towards a temporal separation between the two main pathways for protein degradation in *S. robusta (Ciechanover, 2005)*.

Genes involved in autotrophic metabolism were consistently phasing at dawn and midday (***Figure 4B-D***). GO terms describing biosynthesis of chlorophyll and carotenoid pigments were enriched in genes peaking at ZT2, indicating that pigment biosynthesis is triggered by light onset for photosynthesis. In addition, all 73 expressed *S. robusta* light harvesting complex antenna proteins (LHC) showed significant diel rhythms. Phylogenetic analysis classified LHC genes into four distinct LHC clades: the light-harvesting fucoxanthin chl a/c LHCF, LHCR and LHCY and the photoprotective LHCX (***Figure 4B,C*** and ***Supplemental Figure 6***) (Blommaert et al., 2020; Nymark et al., 2013). In contrast to pigment synthesis phasing at dawn, we observed a sharp peak in expression of LHCF, LHCR and LHCY genes at ZT6, likely in preparation for chloroplast division (Smith et al., 2016; Gillard et al., 2008) (Figure 4C, Supplemental Figure 7). An expansion of photoprotective LHCX is thought to allow benthic diatoms to cope with exposure to highly variable irradiances (Blommaert et al.,

2020). Here, members of the LHCX family showed an intense joint expression peak at ZT2 the first day but lacked a second-day peak in expression, even though experimental conditions were equal to the first day. Photosynthetic electron transport chain components from both photosystem I and II were enriched in genes peaking during the day, the former more specifically at ZT2. In addition, Calvin-Benson-Bassham (CBB) cycle genes (Poliner et al., 2015) such as the CO_2 capturing enzyme rubisco (rbcS and rbcL), sedoheptulose-bisphosphatase (SBPase) and phosphoribulokinase (PRK) phased at dawn (***Figure 4D***). Genes encoding key enzymes in the biosynthesis of the storage polysaccharide chrysolaminarin were phasing at dawn as well, in line with a daytime accumulation and nocturnal depletion of chrysolaminarin (***Supplemental Figure 8***) (Caballero et al., 2016).

In antiphase with photosynthetic and anabolic genes, tricarboxylic acid (TCA) cycle enzymes showed a coherent expression pattern peaking at dusk (ZT10 and ZT14). Thus, in addition to the absence of photosynthetic oxygen production at night, the dominance of cellular respiration at this time further contributes to the nocturnal depletion of oxygen in benthic ecosystems (Clavier et al., 2008; Haro et al., 2019). Peak expression of glycolytic isozymes was incoherent and dependent on which organelle the gene product is targeted to (***Figure 4E, Supplemental Figure 9***). Apart from isozymes localized in the cytosol and chloroplast, several enzymes from the pay-off phase of the glycolytic pathway were targeted to the mitochondrion, a unique feature of stramenopile metabolism that is thought to allow simultaneous flux of incompatible reactions (Smith et al., 2012; Kroth et al., 2008; Bártulos et al., 2018). Cytosolic isozymes were either phasing at dawn or dusk, presumably supporting gluconeogenesis during the day and supplying pyruvate for respiration at night (Smith et al., 2016). On the other hand, chloroplast and mitochondrial-targeted glycolytic isozymes were co-expressed at ZT2 the first day (***Supplemental Figure 9***), substantiating the existence of a temporal separation of glycolytic pathways and a close coupling between mitochondrial glycolysis and the CBB cycle in diatoms (Smith et al., 2016; Goldman et al., 2019).

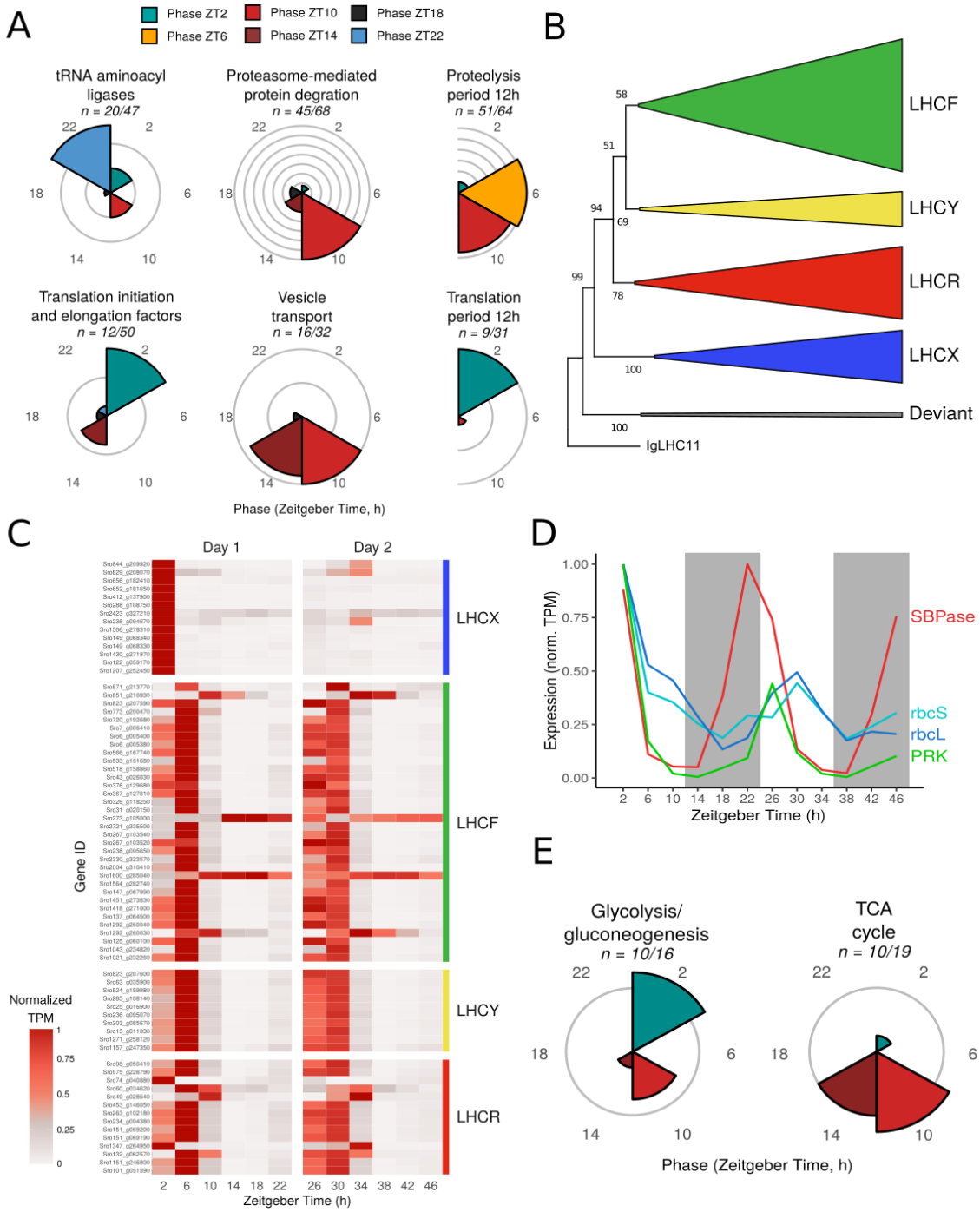

Figure 4: **Diurnal expression of genes from the basal cellular metabolism.** Gene sets plotted in this figure can be retrieved from ***Supplemental Data Set 5.*** **(A)** Circular bar plots summarizing the timing of gene sets related to protein turnover. The number of genes classified into each phase cluster is indicated by the radius of each bar, with each concentric grey background circle representing an increase of four genes. "n" reports the number of genes with a consistent peak out of all rhythmic genes in a given category (**Supplemental Figure 5**). The colour shade and the axis numbers surrounding the plot indicate the phase in Zeitgeber Time (time since first illumination). The phase of gene sets with a 12-h (semidiurnal) period for proteolysis and translation is indicated on semicircles. For glycolysis/gluconeogenesis, only cytosolic and mitochondrial targeted isozymes were included. **(B)** Schematic maximum likelihood phylogenetic tree of genes encoding light harvesting complex proteins (LHC) in *S. robusta* and *P. tricornutum*. Clades are collapsed and shown as triangles. The "atypical LHC" IgLHC11 gene from the haptophyte *Isochrysis galbana*, was included as an outgroup. Numbers indicate bootstrap support values. **(C)** Heatmap of the expression of LHC in *S. robusta* over a two-day day/night time series. The x-axis represents Zeitgeber Time and the y-axis contains significantly rhythmic genes, following the same color code as panel B for LHC classification. Vertical panels denote four categories of LHC proteins recognized in diatoms. TPM denotes transcripts per million. **(D)** Line plot showing normalized expression (TPM) over time (Zeitgeber Time) for four genes committed to the Calvin-Benson-Bassham (CBB) cycle: Rubisco (RbcS, RbcL), sedoheptulose-bisphosphatase (SBPase) and phosphoribulokinase (PRK). **(E)** Circular bar plots summarizing the timing of gene sets related to cellular respiration. TCA refers to tricarboxylic acid cycle and this panel follows the same style and color code as panel A.

Tight coordination of the cell division over the diurnal cycle reveals potential uncharacterized cell wall biosynthetic genes

The *S. robusta* cell cycle was highly synchronized to the diel cycle, supported by 87% (140/161) of cell cycle marker genes being significantly rhythmic and the oscillation of physiological measurements of cell division (***Figure 5A, Supplemental Data Set 5***). Coordinated cell cycle progression during the diel cycle has been observed in all major groups of microalgae, suggesting an evolutionary advantage such as the temporal separation of growth and cell division (Ferrari et al., 2019; Moulager et al., 2007; Poliner et al., 2015; Chisholm et al., 1978). The expression of the *S. robusta* homolog of the *P. tricornutum* G1 re-entry gene dsCyc2 at ZT2 implies that a majority of cells start the day in the G1 phase while the G1-associated cyclin-dependent kinase CDKA1 phased at ZT10 (Huysman et al., 2013, 2010). G1/S transition and S phase marker genes were phasing at ZT14 with an additional peak the second day at ZT2, the latter indicating a secondary round of cell cycling or a subpopulation dividing during the day (***Figure 5A***). Flow cytometry revealed that each peak in S phase expression is followed by an increase in cells with a duplicated genome 4h later (***Figure 5A "G2/M phase cells"***). Following genome replication, chromosome assembly markers and the mitotic cyclin-dependent kinase CDKA2 (Huysman et al., 2015) were maximally expressed at ZT14, while markers for metaphase-to-anaphase transition such as the anaphase promoting complex APC/C activator cdc20 (Huysman et al., 2014) phased at ZT18 (***Figure 5A***). Cell division was mainly restricted to the night, with the proportion of doublets (dividing cells) peaking at ZT18 and ZT22 and cell density almost doubling overnight (80% increase, ***Supplemental Figure 10***). Remarkably, while S phase markers and flow cytometry indicated a secondary round of cycling restricted to the second day, such an extra peak was not supported by mitotic and cytokinetic marker genes or an increase in the cell number (***Figure 5A,*** **Supplemental Figure 10**).

A preference for genome replication taking place around dusk and at night is observed in most lineages of microalgae and is thought to avoid DNA damage by UV or photosynthetic reactive oxygen species (ROS) during this sensitive phase of the cell cycle (Miyagishima et al., 2019; Nikaido and Johnson, 2000). However, the nocturnal peak of *S. robusta* S phase markers was delayed compared to other microalgal transcriptomes (Zones et al., 2015; Poliner et al., 2015; Ashworth et al., 2013; Monnier et al., 2010; Moulager et al., 2007), including *P. tricornutum*, where orthologs encoding S phase machinery that phase at night in *S. robusta* were already active in the late afternoon (***Figure 3C, Supplemental Data Set 6***).

Given the relatively low irradiance used in our experiment, the delay of S phase might be explained by the existence of a G1 phase commitment point which allows progression through the cell cycle only when sufficient photosynthetic energy has been captured. A dependence of S phase timing on light intensity in the pre-commitment period has been found in green algae but such a commitment point has not yet been demonstrated in diatoms (Vítová et al., 2011; Moulager et al., 2010). In addition, cell cycle progression in natural conditions might be restricted until cells have migrated at dusk, as was observed for several other species of motile benthic diatoms (Longphuirt et al., 2006; Saburova and Polikarpov, 2003).

A combination of phylogenetics and transit peptide prediction revealed that genes encoding bacterial division-like proteins were targeted to the chloroplast and the mitochondrion ***(Figure 5B, Supplemental Figure 11-12)***, pointing to the retention of an ancestral system for mitochondrial division in *S. robusta,* as was observed in various protists, including diatoms but with the notable exception of *P. tricornutum* (Leger et al., 2015; Nishikawa et al., 2012; Kiefel et al., 2004). While the eukaryotic chloroplast division gene DRP5B (***Supplemental Figure 13***) was phasing at ZT14 in accordance with chloroplast division being coordinated to the S-G2-M phases of the cell cycle (Gillard et al., 2008), all three bacterial-derived chloroplastic FstZ genes showed a maximum expression at ZT2 with additional expression at night (***Figure 5B***). The discrepancy between the timing of FtsZ expression and chloroplast division in *S. robusta* is highly unusual, given that the FtsZ peak of expression is coordinated with chloroplast division in other microalgae such as *Nannochloropsis oceanica, Ostreococcus tauri* and *Cyanidioschyzon merolae (Monnier et al., 2010; Poliner et al., 2015; Fujiwara et al., 2009).* All mitochondrial FtsZ and Min-system division genes were significantly rhythmic and showed a consistent daily burst at ZT14 and ZT18 (***Figure 5B***), suggesting that mitochondria either divide in coordination with maximal respiration at night, or with the cell cycle as was observed in the red alga *C. merolae* (Fujiwara et al., 2009), scenarios that, to the best of our knowledge, have not been described in diatoms.

The peak of cytokinesis at ZT18/ZT22 coincided with the expression of genes involved in cell wall formation, which takes placed in a specialized intracellular compartment termed the silica deposition vesicle (SDV): 23 rhythmic SDV-acidifying V-type ATPases, 11 SDV transmembrane silicanin-like genes, 24 diatom-specific SET domain protein methyltransferases (BacSETs), eight S-adenosylmethionine decarboxylase-aminoprolyltransferase (AMDC-APT) genes, four silicic acid transporters and 28 genes from

the frustulin family were phasing predominantly at ZT18 (***Figure 5A***)**.** Notably, an 11-transmembrane *S. robusta* homolog of the plant Silicon efflux transporter Lsi2 was also co-expressed at ZT18. This gene was downregulated during nutrient depletion and pheromone induced cell cycle arrest, supporting a role as an additional type of silicic acid transporter in diatoms as was hypothesized in *T. pseudonana* (***Supplemental Figure 14***) (Shrestha et al., 2012). The over-representation of uncharacterized, diatom-specific genes phasing simultaneously with known cell wall formation factors suggests that hitherto unknown cell wall biosynthesis gene families are active at ZT18 (***Figure 2A***). *T. pseudonana*, the best established model for cell wall biogenesis, displayed a lower expression of cell wall biosynthesis genes under silicon depletion conditions (Shrestha et al., 2012; Brembu et al., 2017). Inspection of genes downregulated after silicon depletion in *S. robusta* yielded a significant association with genes phasing at ZT18 ($q < 0.05$). Among the top six diatom-specific families in this dataset, we observed the known cell wall associated frustulins and silicanins as well as four new candidate cell wall formation families that we termed Siren1-4 ("Silica Responsive Night", ***Figure 6, Supplemental Data Set 5)***. The presence of a predicted signal peptide in Siren1 and Siren3 proteins supports a possible role in cell wall formation, as a signal peptide is required for SDV protein targeting (***Supplemental Figure 15***) (Nemoto et al., 2020; Shrestha et al., 2012). Although Siren2 and Siren4 do not possess a characteristic signal peptide and therefore may rather be involved in other processes such as late-stage mitosis, they may also participate in cytosolic processes that are suggested to influence silica morphogenesis (Tesson and Hildebrand, 2010; Schmid, 1994). Siren1 and Siren4 genes do not contain known protein domains, but Siren2 and Siren3 were annotated with ankyrin repeats and an alpha/beta hydrolase fold domain, respectively. During cell wall synthesis, the protein-interacting ankyrin repeats in the Siren2 family may regulate cytoskeletal organization or cell cycle progression, processes these motifs are often involved in (Li et al., 2006). Over the years, numerous proteins associated with diatom cell wall biosynthesis (e.g. silaffins, cingulins and silacidins) have been proposed, but only a handful have a broad distribution across diatoms, most notably the silicanins, frustulins, p150 proteins and BacSETs (Nemoto et al., 2020). We expand this list with the putative Siren3 and Siren4 cell wall families which are shared among pennate and centric diatoms (***Figure 6***). Homologs of Siren1 and Siren2, on the other hand, appeared to be uniquely present in pennate diatoms, which is exciting as no pennate-specific cell wall associated proteins have been described to date ***(Figure 6***) (Nemoto et al., 2020).

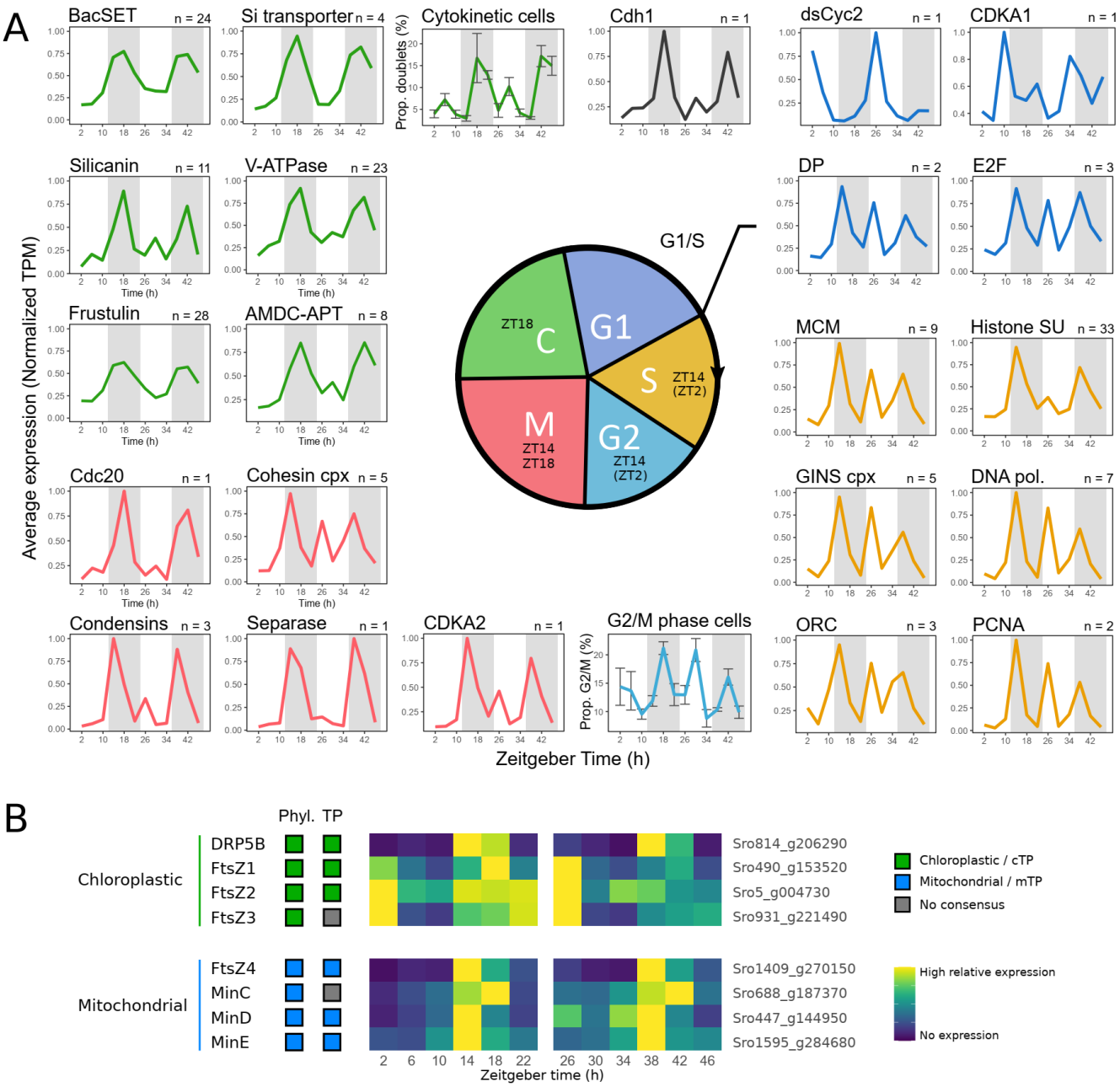

Figure 5: Diel transcriptional dynamics of marker genes for cell cycle and organelle division. (A) Line plots showing the average normalized expression (transcripts per million, TPM) over time (Zeitgeber Time = time since illumination) of significantly rhythmic marker genes for cell cycle progression, where the number of marker genes used in each plot is shown in the top right corner (n). The colour of the line refers to the cell cycle stage as indicated in the central cartoon, where also the consensus peak in expression for each stage is indicated, with the timing of an additional peak uniquely occurring during the second day indicated between brackets. Phases in the cartoon are not necessarily proportional to the length of these phases in the experiment. Line plots of S phase progression depicting the average fraction of G2/M phase cells ("G2/M phase cells") and the average proportion of microscopically visible cytokinetic cells ("Cytokinetic cells") are included, with error bars representing the standard error. SU = Subunit, cpx = complex, DP = Dimerization Partner, PCNA = Proliferating cell nuclear antigen, MCM = Minichromosome maintenance complex, ORC = Origin recognition complex, AMDC-APT = S-adenosylmethionine decarboxylase - aminoprolyltransferase family, BacSETs = SET domain protein methyltransferases. Gene sets plotted in this panel can be retrieved from ***Supplemental Data Set 5.*** **(B)** Classification and diurnal expression of organelle division genes. Coloured boxes on the left show the predicted target organelle (chloroplast = green, mitochondrion = blue) based on gene phylogeny and the presence of an N-terminal transit peptide. TPM normalized to a maximum of 1 over each day is shown as a heatmap for each of these genes following a color gradient from blue (no expression) to yellow (maximal expression). cTP = chloroplastic transit peptide, mTP = mitochondrial transit peptide.

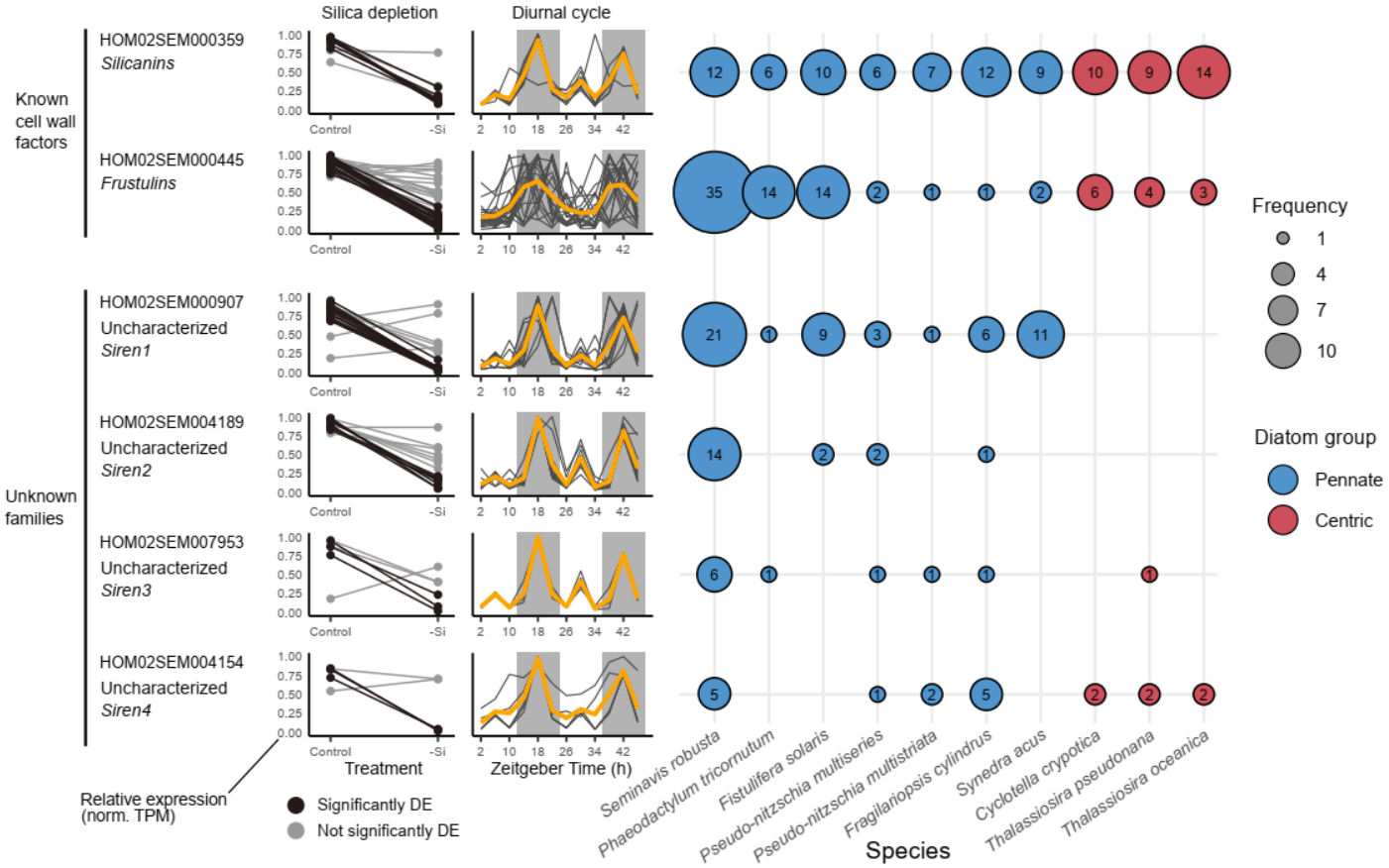

Figure 6: Identification of putative new diatom cell wall biosynthesis families using silica depletion and diurnal RNA-Seq datasets. The PLAZA Diatoms 1.0 family ID and description of six gene families are given on the left hand side, where families are classified as previously known cell wall families or unknown families which were termed "Siren" (**Si**lica **re**sponsive **n**ight). The relative expression of members of each gene family during silica depletion and in the two-day diurnal expression dataset is shown as TPM normalized for each gene to a maximum of 1 over the experiment. Solid lines connect observations from the same gene over conditions or time. Genes significantly differentially expressed in response to silica depletion are plotted with black lines, while an orange line shows the average relative expression of the gene family over the diurnal cycle experiment. Diurnal expression was only plotted for significantly rhythmic genes with a period of 24h. On the right, the presence of members of the gene family in the genome of 10 diatom species is shown, with circle radii and numbers inside each circle showing the number of genes belonging to each species in each family. Gene families in pennate diatoms are shown with a blue filled circle, while centric diatoms are indicated by a red shade. Diatom gene IDs belonging to the Siren families can be retrieved from ***Supplemental Data Set 5***

Linking the expansion of putative circadian genes and semidiurnal periodicity to benthic diatom biology

The Per/Arnt/Sim (PAS) signalling domain is a hallmark of many proteins that constitute the central circadian oscillator in animals, often alongside a basic helix-loop-helix (bHLH) DNA binding domain (Crane and Young, 2014). The bHLH-PAS TFs have recently been identified as the first known components of the stramenopile circadian clock, with the discovery of the time-keeping regulator bHLH1a-PAS (RITMO1) in the diatom *P. tricornutum (Annunziata et al., 2019)*. In *S. robusta*, we identified 13 *S. robusta* bHLH TFs, of which 10 contained a PAS domain, which is in stark contrast to *P. tricornutum,* encoding only three bHLH-PAS (Annunziata et al., 2019). Phylogenetic analysis revealed that each *P. tricornutum* bHLH TF corresponded to one *S. robusta* ortholog*,* except for bHLH2-PAS which expanded into seven gene copies (***Figure 7***). RITMO1 and bHLH3 orthologs showed conserved rhythmic expression, while bHLH1b-PAS displayed antiphase expression between both species (***Figure 7***). The multi-copy *S. robusta* bHLH2-PAS clade of which 6/7 were rhythmic, exhibited expression divergence, with rhythmic genes peaking either at dawn or broadly during the day (***Figure 7***). The expansion and expression divergence of bHLH-PAS TFs may suggest that the inherently variable and highly rhythmic benthic environment drove the evolution of extensive time-keeping mechanisms. Additionally, two circadian-regulated *P. tricornutum* bZIP-PAS TFs (bZIP5-PAS and bZIP7-PAS) showed a conserved diurnal expression in *S. robusta*, phasing at night and at dusk respectively (***Supplemental Figure 16***) (Annunziata et al., 2019). Together, the phasing of *S. robusta* bHLH-PAS and bZIP-PAS TFs covered the complete diurnal cycle, suggesting that a sequential expression of these TFs drives the circadian clock of this species (***Supplemental Figure 16***).

Two other types of presumed diatom circadian genes showed significant oscillations (Farré, 2020; Coesel et al., 2009) (***Supplemental Figure 17***). We observed a peak at night for three casein kinase 1 (CK1) homologs while the blue-light sensing cryptochrome/photolyase family 1 (CPF1) gene was peaking at dawn, analogous to *P. tricornutum* and *N. oceanica (Poliner et al., 2015)*. Genes encoding G-protein-coupled receptor (GPCR) rhodopsin-like proteins, which are prominently expanded in the *S. robusta* genome and conserved in several other benthic diatoms, were predominantly expressed at dusk and at night (**Supplemental Figure 18**) (Osuna-Cruz et al., 2020). Given the sensitivity of animal rhodopsins to extremely low levels of light, they may be involved in photoreception inside sediments or phototrophic biofilms which exhibit strong light attenuation on a millimeter scale (Reingruber et al., 2015; Congestri and Albertano, 2011; MacIntyre et al., 1996). Alternatively and in accordance with

their nocturnal expression, GPCR rhodopsin-like proteins may be used to perceive twilight or moonlight at night, a feature observed for rhodopsins in the compound eyes of insects (Schlichting et al., 2014, 2015).

Next to diurnally regulated genes, we observed that 7.5% of rhythmic genes displayed a semidiurnal rhythm, i.e. had a period of 12h (***Figure 1C***). Apart from translation factors and proteases/peptidases, we observed that a range of glycosyltransferases, sulfotransferases, sugar phosphate transporters and fasciclin-like adhesive proteins are enriched in 12-h periods and phase at ZT10/ZT22. In diatoms, such genes have been implicated in the production of EPS, which are typically produced during migration and adhesion (Aslam et al., 2018; McConville et al., 1999; Smith and Underwood, 1998). The semidiurnal expression of these genes might be linked to twice-daily vertical migration in subtidal and intertidal diatom communities, where diatoms burrow into the sediment at dusk and emerge at dawn (Haro et al., 2019; Longphuirt et al., 2006; Consalvey et al., 2004). Alternatively, semidiurnal rhythmicity may be an adaptation to the occurrence of tides in coastal benthic ecosystems, as there is typically a period of 12.4h from one high tide to the next one, and tides were observed to induce cyclic migration patterns in intertidal diatoms (Palmer and Round, 1967; Haro et al., 2019; Naylor, 2010). In several coastal animals, gene expression with a 12-h period has already been attributed to tidality (Tarrant et al., 2019; Schnytzer et al., 2018). Nonetheless, our laboratory experiment did allow for vertical migration and did not simulate tidality, making it hard to attribute this semidiurnal phenomenon to any of these scenarios. Finally, a hybrid-sensor PAS (*Sro442_g143910*) and a bZIP transcription factor (*Sro199_g084500,* bZIP1) showed significant 12-h oscillations phasing at ZT10/ZT22, both being credible candidate components of a potential circatidal clock as has been described in marine animals (Zhang et al., 2013; Tarrant et al., 2019; Takekata et al., 2012; Farré, 2020).

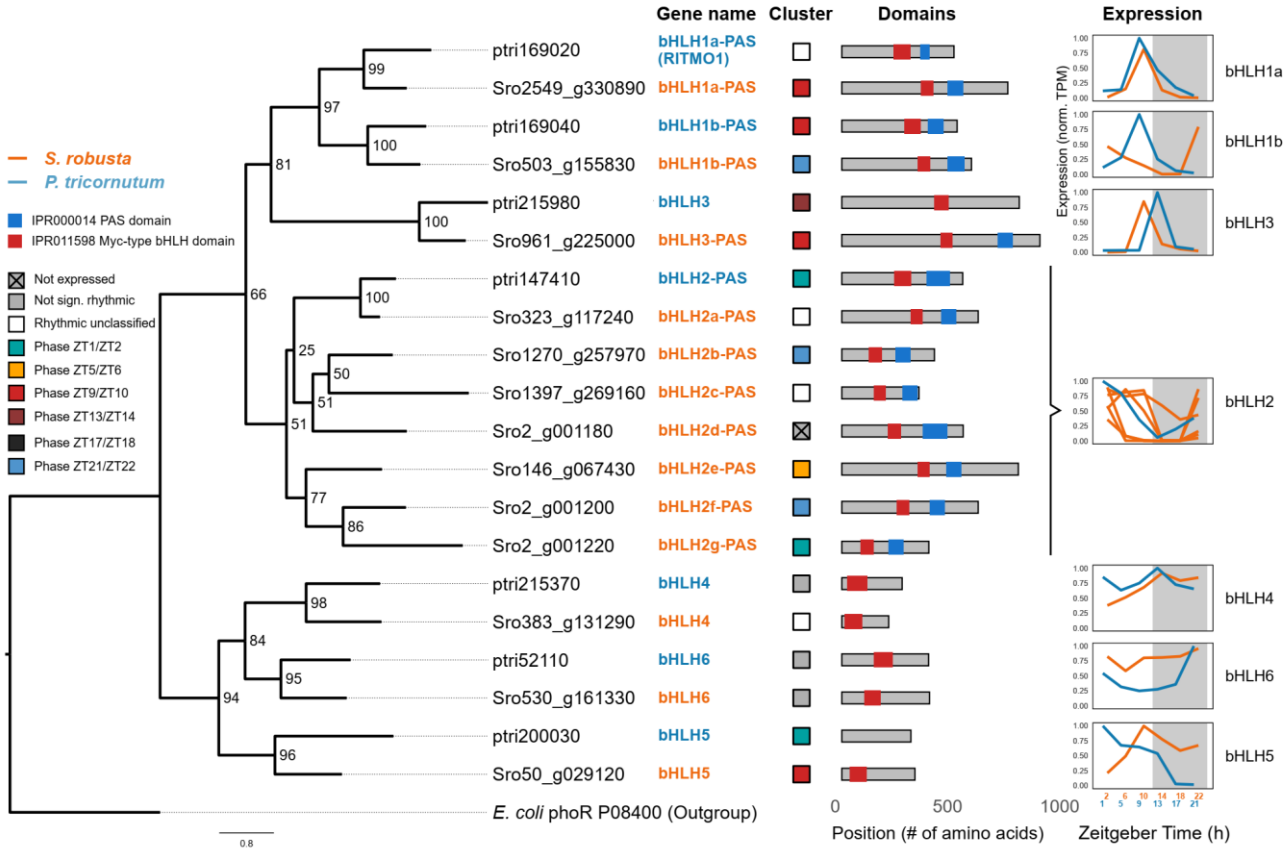

Figure 7: Phylogeny and comparative analysis of the diurnal expression of basic helix-loop-helix (bHLH) transcription factors in the pennate diatoms *S. robusta* and *P. tricornutum*. The left side shows the maximum likelihood phylogenetic tree (left) of bHLH transcription factors, with bootstrap support values shown at the branch points. The *E. coli* PAS-domain containing protein PhoR was used as an outgroup for rooting. For each gene in the tree, coloured boxes indicate the classification of diurnal oscillations ("cluster"): non-expressed, not-significantly rhythmic, rhythmic with unclassified phase and a specific phase cluster. The position of bHLH (red) and PAS (blue) InterPro domains on each gene (grey) is visualized under "domains". When multiple identical overlapping domains were predicted, the longest domain was selected for visualization. Normalized expression (transcripts per million, TPM) over time (Zeitgeber Time = time since illumination) is shown for each subclade at the right side of the plot (*S. robusta* = orange, *P. tricornutum* = blue). To allow comparison with the 1-day *P. tricornutum* time, *S. robusta* relative expression was averaged over both days and shown over the span of one day. The *P. tricornutum* time point ZT25 represents the equivalent ZT1.

Concluding remarks

Comprehensive profiling of the *S. robusta* transcriptome revealed that oscillating protein-coding genes are organized into highly coordinated transcriptional programs controlling cell cycle progression, protein turnover and carbohydrate metabolism. Moreover, the identification of rhythmic lincRNAs indicates that the non-coding transcriptome is not solely important in the response to environmental stressors in diatoms (Carvalho et al., 2016; Huang et al., 2019), but also plays a role in the diurnal cycle. The pronounced differences in expression levels of these processes over the course of the day should be considered when interpreting data from metatranscriptome initiatives such as TARA Oceans where samples are collected at a particular time of the day and at varying latitude with inherently different day lengths, as diel rhythms will likely have a profound impact on gene expression in each sample (Pesant et al., 2015; Carradec et al., 2018). The diurnal regulation of carbohydrate metabolism appears to be conserved among microalgae. Not only was autotrophic and heterotrophic transcription temporally segregated as the result of photosynthetic energy input being restricted to the day, we also observed anticipatory expression of anabolic and photosynthetic genes before sunrise, which suggests regulation by the circadian clock (Monnier et al., 2010; Zinser et al., 2009; Ashworth et al., 2013).

The strict diurnal orchestration of gene expression patterns and the behaviour of cells in *S. robusta* dataset revealed new potential cell wall biosynthetic families, a spatial and temporal segregation of the glycolytic pathway, and the first indications of synchronous mitochondrial division in diatoms. In addition, the expansion of circadian transcription factors and photoprotective LHCX, and the 12-h periodicity of putative EPS-producing enzymes all suggest that raphid pennate diatoms evolved multiple adaptations to the complex and variable benthic environment. Additional transcriptomic profiling of benthic diatoms would benefit from a set-up allowing for vertical migration while simulating tidal regimes. Similarly, the global transcriptomic response to different photoperiods has not been assessed in diatoms despite it being a highly relevant topic, as daylength at high latitudes vary greatly throughout the year due to seasonality, and circadian expression in diatoms is sensitive to these changes in periodicity (Annunziata et al., 2019).

Methods

Experimental conditions of the diel experiment

Seminavis robusta strain 85A (DCG 0105), obtained from the Belgian Coordinated Collection of Microorganisms (BCCM/DCG, http://bccm.belspo.be/about-us/bccm-dcg), was grown in artificial sea water (ASW) prepared using 34.5 g/L Tropic Marin BIO-ACTIF sea salt (Tropic Marin, Wartenberg, Germany) and 0.08 g/L sodium bicarbonate. After autoclaving, ASW was supplemented with Guillard's F/2 Marine Water enrichment solution (Algoid technologies) and an antibiotic mix (400 µg/ml penicillin, 400 µg/ml ampicillin, 100 µg/ml streptomycin and 50 µg/ml gentamicin)(Koedooder et al., 2019). Cultures were grown in 150 mL of medium in Cellstar Standard Cell Culture Flasks (Sigma-Aldrich) with a surface area of 175 cm², which were placed in a climate-controlled plant growth chamber (Weiss technik) at a temperature of 21°C and a light intensity of 30 µmol photons $m^{-2}s^{-1}$. Cultures were first acclimated to a 12h-light/12h-dark photoperiod for 5 days, after which they were sampled for transcriptomic analysis at 12 time points spanning two diurnal cycles (day, night, day, night). The first culture flask was harvested the first day after 2h of illumination (Zeitgeber Time 2h = ZT2). Subsequently, samples were harvested every 4h, resulting in 6 samples for the first cycle (2h, 6h, 10h, 14h, 18h, 22h) and 6 samples the next (26h, 30h, 34h, 38h, 42h, 46h). At each time point, a complete culture flask was harvested by scraping cells from the bottom surface using a cell scraper and homogenizing by gentle shaking. Cells were collected by filtration with a pore size of 3µm and flash frozen in liquid nitrogen to be stored at -80°C awaiting RNA extraction. For samples collected at night, this procedure was performed in the dark to avoid responses to light. The experiment was repeated three times on separate occasions so that each time point was sequenced in biological triplicate (36 samples).

Quantification of cell cycle progression

To assess S phase progression during the experiment, a subsample of each of the 36 sequenced samples was analysed using flow cytometry: 5 mL of culture was centrifuged for 5 min at 3,000 RPM. The pellet was fixated by adding 10 mL of ice cold 75% ethanol, after which samples were kept in the dark at 4 °C. Preparation of samples, staining with SYBR green and flow cytometry analysis on a Bio-Rad S3e cell sorter was performed as detailed by Bilcke et al. (2020). In parallel, cell density and the proportion of cytokinetic cells were determined at each time point from microscopic pictures taken with a Canon Powershot G11 camera mounted on a Primovert inverted microscope at 20x magnification. At night, pictures

were taken from a separate culture in order to not disturb the diurnal cycle. Manual cell counting was performed using ImageJ software. We defined cytokinetic cells as cells showing a visible cleavage furrow or developing frustule subdividing the daughter cells.

RNA preparation and sequencing

Total RNA was prepared using the RNeasy Plant Mini Kit (Qiagen). Culture was scraped from the filter into 1 mL of RLT+ME buffer mixed with ⌀1.0 mm SiC beads (BioSpec products) and were lysed on a beating mill (Retsch) for 1 min at 20 Hz. The lysate was transferred to a spin column and the rest of the protocol was performed according to manufacturer's instructions. RNA concentration was determined on the Nanodrop ND-1000 spectrophotometer (Thermo Scientific) and RNA integrity was assessed using a BioAnalyzer 2100 (Agilent). Library preparation was performed using the Illumina TruSeq Stranded mRNA Sample Prep Kit (Illumina), and included mRNA enrichment using poly-T oligo-attached magnetic beads. Paired-end 2x76 bp sequencing of 36 samples was performed on the Illumina HiSeq 4000 platform at the VIB Nucleomics core (www.nucleomics.be).

Long intergenic noncoding RNA (lincRNA) gene annotation and conservation analyses

LincRNAs were annotated in the *S. robusta* genome leveraging several RNA-seq datasets. The quality of adaptor-trimmed reads belonging to 161 samples from the previously published *S. robusta* gene expression atlas (Osuna-Cruz et al., 2020) and the 36 new diel cycle samples was assessed using FastQC v0.11.4 (Andrews, 2010). All reads were mapped to the *S. robusta* genome using STAR v2.5.2 (--sjdbGTFfeatureExon, --alignIntronMax 6000)(Dobin et al., 2013; Osuna-Cruz et al., 2020) followed by the removal of chimeric alignments and retention of uniquely mapping reads using samtools v1.3 (-F 4 -q 1)(Li et al., 2009). Next, these alignments were split per scaffold and genome-guided Trinity De novo Transcriptome v2.6.6 was executed (--genome_guided_bam, --genome_guided_max_intron 6000)(Grabherr et al., 2011), generating 221,735 transcripts. Selecting the major isoform of each transcript led to an initial set of 143,473 distinct transcript models which were mapped back to the *S. robusta* genome using gmap v2017-11-15 (-min-identity 0.9)(Wu and Watanabe, 2005). We focused our analysis on lincRNAs, whose expression quantification is the most reliable since they do not overlap with protein-coding gene regions. Therefore, transcripts overlapping with protein-coding genes were removed using bedtools v2.2.28 (-intersect)(Quinlan and Hall, 2010), retaining 33,849 transcripts. Afterwards, the FEELnc

v2018-07-12 pipeline was used to annotate lincRNA candidates (Wucher et al., 2017). In detail, we executed the FEELnc filter (--monoex=1) removing transcripts with a length of <200nt and bi-exonic transcripts having one exon with a length of <25 nt. Subsequently, we executed FEELnc codpot with --mode=intergenic and --mode=shuffle parameters. Transcripts with coding potential in any of these two runs were discarded (30,968 transcripts were kept). Transcripts suspected to be fragments of 5' or 3' UTRs, i.e. ending close to a protein-coding gene (distance <317 nt, average UTR length of protein-coding genes) were removed, retaining 18,262 transcripts. Additionally, potential remaining transcript redundancy was removed using the cuffmerge tool from cufflinks v2.2.1 (default parameters)(Trapnell et al., 2012). Finally, only transcripts without similarity to protein-coding genes were retained by comparing transcripts to known proteins in the Uniprot database using blastx v2.6.0 (Altschul et al., 1990; UniProt Consortium, 2019). LincRNA transcripts were discarded if their E-value was below a threshold value of 1.23e-05 that was experimentally determined as the minimal blastx E-value of ten randomly shuffled sequences per candidate transcript, obtaining a set of 14,782 candidate lincRNAs (13,470 mono-exonic and 1,312 multi-exonic). Finally, a subset of 7,117 high-confidence lincRNAs which show robust expression support (average TPM > 2 in at least one condition) in either the *S. robusta* gene expression atlas or the new diurnal expression dataset was selected for further consideration (***Supplemental Data Set 1***).

To assess conservation of the predicted lincRNAs in other diatoms, a blastn v2.6.0 (Altschul et al., 1990) search was executed against the diatom MMETSP transcriptomes (Keeling et al., 2014) and nine diatom genomes stored in PLAZA Diatoms 1.0 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_diatoms_01/). The same blastn search was executed on 10-fold shuffled sequences generated from candidate lincRNAs to determine a cut-off for significant hits (E-value < 3.8e-04).

Quantification of diel cycle gene expression and detection of rhythmic genes

Adaptor-trimmed reads from the 36 diel cycle samples were mapped to the longest isoform for each protein-coding gene (36,254) of the *S. robusta* genome merged with the set of candidate predicted lincRNAs (14,782) using Salmon v1.1, using following parameters for indexing: --keepDuplicates -k 31, and quantification: --seqBias --gcBias -l A (Patro et al., 2017)(***Supplemental Data Set 1***). To assess reproducibility between diel cycle replicates and corresponding time points over subsequent days, count estimates were imported using tximport (Soneson et al., 2015) and an unsupervised multidimensional scaling plot was

created as implemented in the EdgeR package for R (Robinson et al., 2010), in which distances between samples approximate log2 fold changes for the top 500 most distinctive genes.

Genes of which the diurnal expression in transcripts per million (TPM) averaged over the three replicates exceeded two in at least one time point were retained for further analysis, keeping 26,109 protein-coding and 1,680 lincRNA genes. Genes with a rhythmic diurnal expression were identified using the empirical Jonckheere-Terpstra-Kendall (empJTK) tool (Hutchison et al., 2015; Hughes et al., 2010). Gene-level TPMs of three replicates for each of the 12 time points were used as input, specifying the waveform family as “cosine”, the possible periods of the waveform as 24h and 12h, and allowing asymmetries ranging from 4h to 20h in steps of 4h. The false discovery rate (FDR) for the set of rhythmic genes was controlled at 0.05, after adjusting p-values for multiple testing using the Benjamini-Hochberg method implemented in empJTK as “GammaBH”.

Comparisons with rhythmic gene expression of the pennate diatom *Phaeodactylum tricornutum* were performed based on publicly available SOLiD RNA-seq data retrieved from the iron depletion study of Smith et al. (2016) (***Supplemental Table 2***). Triplicate RPKM values for 6 time points (ZT5, ZT9, ZT13, ZT17, ZT21, ZT25 = ZT1) concerning the “High” iron condition (400 pM Fe, nutrient replete conditions) were converted to TPM. After removal of duplicated, empty rows and lowly expressed genes with average TPM < 2 in all time points, 10,771 genes were used for rhythmicity detection with analogous empJTK settings as for *S. robusta*, but specifying possible phases as ZT1-ZT21 with a step size of 4.

Classification of rhythmic genes into clusters

Genes were classified with regard to rhythmicity in a hierarchical way (***Supplemental Data Set 2***). The first distinction concerns rhythmic versus non-significantly rhythmic genes, as determined by empJTK. At the next level, empJTK classified rhythmic genes as having a period of 24h (one cycle each day) or 12h (two cycles each day). Further, genes were classified according to their phase, i.e. the timing of maximum expression. Although the phase of the best fitting waveform is reported by empJTK, we restricted the phase classification to genes with a stably recurring maximum each day, taking advantage of our two-day dataset. In detail, genes were annotated with a phase if the timing of the maximal average TPM for each day (24h period) or half day (12h period) agreed with the predicted phase by empJTK. Finally, rhythmic genes were classified into broad temporal categories if their predicted and actual expression maxima and minima fall within and outside

predetermined time intervals respectively: day (ZT2, ZT6, ZT10), night (ZT14, ZT18, ZT22), dawn (ZT22, ZT2) and dusk (ZT10, ZT14). *P. tricornutum* rhythmic genes were classified into clusters following the same set of rules defined for *S. robusta*.

Functional annotation of protein-coding and lincRNA genes

Functional annotation of *S. robusta* protein-coding genes was obtained from PLAZA Diatoms 1.0 (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_diatoms_01/). In particular, Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and InterPro (IPR) domain annotation from InterProScan were retrieved (Osuna-Cruz et al., 2020). Phylostratification of protein-coding genes was calculated based on the taxonomic scope of their predicted TribeMCL gene family in PLAZA Diatoms 1.0 (Domazet-Lošo et al., 2017; Enright et al., 2002). Differential gene expression (DE) labels for *S. robusta* protein-coding genes were retrieved from Osuna-Cruz et al. (2020). Significant enrichment of phylostrata/gene family/DE/GO/IPR features for a gene set were inferred using a hypergeometric test, setting the minimum number of positive features in a set to two and controlling p-values for multiple testing on a 5% FDR level using the Benjamini-Hochberg procedure (***Supplemental Data Set 4***).

Combining the diurnal expression dataset with the biological process GO terms assigned to *S. robusta* protein-coding genes, a co-expression network was constructed to annotate lincRNAs based on the guilt-by-association principle (Li et al., 2017). Pearson correlations were computed in a pairwise manner between all genes and Highest Reciprocal Ranks (HRR) were calculated (Mutwil et al., 2010; Liesecke et al., 2018; Vaneechoutte and Vandepoele, 2019). The optimal HRR threshold that maximizes the weighted F1-score (harmonic mean of the precision and recall) was estimated at 12. Co-expression clusters were defined based on this cut-off and GO enrichment was executed per cluster by performing a hypergeometric test (q-value < 0.05).

Detection and rhythmicity of orthologous genes between *S. robusta* and *P. tricornutum*

Orthologous gene relationships between *S. robusta* and *P. tricornutum* were inferred using the PLAZA integrative orthology method (Van Bel et al., 2018) (***Supplemental Data Set 6***). Genes were considered orthologs when they were confirmed by the OrthoFinder gene family (ORTHO) and the Best-Hits-and-Inparalogs (BHIF) methods in both species, identifying 575 highly reliable ortholog pairs with a one-to-one relationship and a stable, 24-h periodic phase

in both species. For each pair, the phase shift in *P. tricornutum* as compared to *S. robusta* was calculated. Since harvesting points were offset 1h in *P. tricornutum* relative to *S. robusta*, the phase shift can take on following values: +3h (-21h), +7h (-17h), +11h (-13h), +15h (-9h), +19h (-5h) and +23h (-1h). The relationship of the phase between orthologs was visualized using a Sankey diagram as implemented in the ggalluvial package for R (http://corybrunson.github.io/ggalluvial).

Construction of phylogenetic gene trees

Maximum-likelihood phylogenetic trees were computed using the following protocol: i) multiple sequence alignment of genes/families of interest using MAFFT v7.187 (Katoh and Standley, 2013), ii) automatic editing of this multiple sequence alignment using trimal v1.4.1 (–gt 0.1)(Capella-Gutiérrez et al., 2009), iii) phylogenetic tree construction using IQ-TREE v1.7 (-bb 1000 -mset JTT,LG,WAG,Blosum62,VT,Dayhoff -mfreq F -mrate R) (Nguyen et al., 2015) and iv) visualization and midpoint rooting using FigTree v1.4.4 (Rambaut, 2010). The simplified light harvesting complex protein (LHC) tree was visualized using MEGA X (Kumar et al., 2018) (***Figure 4B***).

Prediction of subcellular localization

The localization of nuclear-encoded protein-coding genes in the *S. robusta* cell was computationally predicted for organelle division genes, glycolytic isozymes and genes from proposed novel cell wall associated gene families (***Supplemental Data Set 3***)(Moog, 2018). Mitochondrial transit peptide (mTP) prediction was performed with TargetP 2.0 in non-plant mode (http://www.cbs.dtu.dk/services/TargetP/) (Almagro Armenteros et al., 2019), whereas signal peptides (SP) were predicted using SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP-3.0/) (Bendtsen et al., 2004). Bipartite targeting signals (BTS) for import into the stramenopile secondary chloroplast were identified using ASAFind (https://rocaplab.ocean.washington.edu/tools/asafind/) (Gruber et al., 2015) and HECTAR (Gschloessl et al., 2008). Making use of the endosymbiotic origin of chloroplasts and mitochondria, localization of organelle fission genes was confirmed based on phylogenetic clustering with bacterial homologs in selected alphaproteobacteria and cyanobacteria species.

Identification of putative new cell wall associated gene families

The differential expression matrix depicting the response of *S. robusta* protein-coding genes to prolonged silica depletion was retrieved from the publicly available *S. robusta* gene expression atlas (Osuna-Cruz et al., 2020). An enrichment test was carried out as described before, testing for an association between the phasing of genes (ZT2-ZT22) and their response to silica depletion (upregulated, downregulated, not DE) in the differential expression analysis from Osuna-Cruz et al. (2020). Next, the enrichment of gene families in combined clusters of phase and silica response was determined, and gene families enriched in the set of genes with phase ZT18 and downregulated in response to silica depletion were withheld as potential silica biogenesis families. Six key families were selected after retaining only families with an enrichment q-value < 0.001 in the set of ZT18:downregulated genes, that have genes in more than three diatom species and no representatives outside diatoms. The novelty of uncharacterized families returned by this analysis was verified by cross-referencing with a list of known cell wall families (***Supplemental Data Set 3***).

Data availability statement

Raw Illumina RNA-seq reads are available at the European Nucleotide Archive (ENA) under accession number **PRJEB40442**. The improved *S. robusta* genome annotation including lincRNA genes annotated in this work is available through ORCAE (http://bioinformatics.psb.ugent.be/orcae/overview/Semro). Additional functional and homology information concerning genes mentioned in this work is available through the diatom-PLAZA platform (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_diatoms_01/).

Supplemental Data

Supplemental Figures

Supplemental Figure 1: Expression of 24-h period rhythmic genes after classification by phase.

Supplemental Figure 2: Expression of 12-h period rhythmic genes after classification by phase.

Supplemental Figure 3: Overview of the diurnal gene expression experiment and classification of rhythmic *Phaeodactylum tricornutum* genes.

Supplemental Figure 4: Overview of non-diurnal gene expression.

Supplemental Figure 5: Overview of rhythmicity in protein turnover expression in *S. robusta*.

Supplemental Figure 6: Maximum likelihood phylogenetic tree of *S. robusta* and *P. tricornutum* genes part of LHC superfamily.

Supplemental Figure 7: Heatmap of the expression of light harvesting complex proteins (LHC) in *P. tricornutum* over diel cycle.

Supplemental Figure 8: Diurnal expression of genes putatively involved in the synthesis of the storage carbohydrate chrysolaminarin.

Supplemental Figure 9: Heatmap of gene expression of rhythmic glycolytic isozymes.

Supplemental Figure 10: Culture density during the three repeats of the diurnal experiment.

Supplemental Figure 11: Maximum-likelihood phylogenetic tree of FtsZ genes.

Supplemental Figure 12: Maximum-likelihood phylogenetic tree of Min-system genes.

Supplemental Figure 13: Maximum-likelihood phylogenetic tree of DRP5A and DRP5B genes.

Supplemental Figure 14: Expression and structure of the plant-type silicon efflux transporter-like gene Sro979_g227220 in Seminavis robusta.

Supplemental Figure 15: Signal peptide and transmembrane domain prediction of known and putative new diatom cell wall families.

Supplemental Figure 16: Expression peaks of potential circadian transcription factors containing a PAS domain cover the diurnal cycle.

Supplemental Figure 17: Expression of putative circadian casein kinase 1 (CK1) and cryptochrome/photolyase family 1 (CPF1) genes.

Supplemental Figure 18: Expression and distribution of potential G-protein coupled receptor (GPCR) rhodopsin photoreceptors.

Supplemental Figure 19: Expression of chloroplast division genes in *P. tricornutum*.

Supplemental Tables

Supplemental Table 1: Classification of rhythmic genes in *S. robusta* and *P. tricornutum.*

Supplemental Table 2: Overview of available day/night-related transcriptomes in diatoms.

Supplemental Data Sets

Supplemental Data Set 1: Sequenced samples/mapping statistics (A), expression levels of diurnally expressed genes in transcripts per million (TPM) (B) and gene IDs of 7117 high-confidence lincRNAs that are assessed in this study (C).

Supplemental Data Set 2: Classification of rhythmicity, period and phase for all expressed genes.

Supplemental Data Set 3: Functional annotation: localization (A,B,C), lincRNA guilt-by-association annotation (D,E) and known cell wall families (F).

Supplemental Data Set 4: Results of enrichment analyses: phylostrata (A), silica/midnight (B), 12-h periodic (C) and full enrichment results for all feature types (Age, DE category, gene family, GO terms, InterPro domain) (D)

Supplemental Data Set 5: Curated lists of genes : metabolism (A-F), cell cycle (G), Sirens (H), Putative circadian (I) and rhodopsin-like genes (J)

Supplemental Data Set 6: Results orthology analysis.

Acknowledgements

We would like to thank Nils Kröger for his advice in validation of novel silica families. We are further grateful to Camilla Ferrari for helpful discussion and to Thomas Depuydt for his assistance with applying the guilt-by-association method.
G.B. and C.M.O-C. were supported by a UGent-BOF project (GOA 01G01715) and G.B. was additionally funded by an FWO aspirant grant (3F001916). N.P. has been supported by the Deutsche Forschungsgemeinschaft (PO 2256/1–1). P.B. received funds from an Erwin Schrödinger fellowship from the Austrian Science Fund (J3692-B22) and an FWO project (G0D6114N). This work was further supported with infrastructure funded by the research council of Ghent University (BOF/GOA 01G01715) as well as an EMBRC Belgium—FWO project (GOH3817N).

Author contributions

K.V., L.DV. and W.V. initiated and managed this project, while experiments were designed by K.V., L.DV., W.V., P.B., C.M.O-C. and G.B.. The RNA-seq experiments were executed by G.B. and C.M.O-C. Flow cytometry and RNA quality control was performed by G.B., the latter with the help of S.D.. LinRNA detection and functional annotation was carried out by C.M.O-C. with the help of M.S-S.. RNA-seq quantification, rhythmic gene detection, clustering, comparative transcriptomic analysis and subsequent downstream biological interpretation were performed jointly by C.M.O-C. and G.B.. N.P. assisted in the computational validation of proposed new cell wall families. The manuscript was written by G.B. and C.M.O-C. and reviewed by all co-authors.

References

Admiraal, W. (1977). Influence of light and temperature on the growth rate of estuarine benthic diatoms in culture. Marine Biology 39: 1–9.

Almagro Armenteros, J.J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance 2.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Andrews (2010). FastQC, a quality control tool for high throughput sequence data.

Annunziata, R. et al. (2019). bHLH-PAS protein RITMO1 regulates diel biological rhythms in the marine diatom *Phaeodactylum tricornutum*. Proc. Natl. Acad. Sci. U. S. A. 116: 13137–13142.

Ashworth, J., Coesel, S., Lee, A., Armbrust, E.V., Orellana, M.V., and Baliga, N.S. (2013). Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. Proc. Natl. Acad. Sci. U. S. A. 110: 7518–7523.

Aslam, S.N., Strauss, J., Thomas, D.N., Mock, T., and Underwood, G.J.C. (2018). Identifying metabolic pathways for production of extracellular polymeric substances by the diatom *Fragilariopsis cylindrus* inhabiting sea ice. ISME J. 12: 1237–1251.

Barnett, A., Méléder, V., Blommaert, L., Lepetit, B., Gaudin, P., Vyverman, W., Sabbe, K., Dupuy, C., and Lavaud, J. (2015). Growth form defines physiological photoprotective capacity in intertidal benthic diatoms. ISME J. 9: 32–45.

Bártulos, C.R., Rogers, M.B., Williams, T.A., Gentekaki, E., Brinkmann, H., Cerff, R., Liaud, M.-F., Hehl, A.B., Yarlett, N.R., Gruber, A., Kroth, P.G., and van der Giezen, M. (2018). Mitochondrial Glycolysis in a Major Lineage of Eukaryotes. Genome Biology and Evolution 10: 2310–2325.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340: 783–795.

Bilcke, G., Van den Berge, K., De Decker, S., Bonneure, E., Poulsen, N., Bulankova, P., Osuna-Cruz, C. M., Dickenson, J., Sabbe, K., Pohnert, G., Vandepoele, K., Mangelinckx, S., Clement, L., De Veylder, L., & Vyverman, W. (2020) Mating type specific transcriptomic response to sex inducing pheromone in the pennate diatom *Seminavis robusta.* ISME J.

Blommaert, L., Vancaester, E., Huysman, M.J.J., Osuna-Cruz, C.M., D'hondt, S., Lavaud, J., Lepetit, B., Winge, P., Bones, A.M., Vandepoele, K., Vyverman, W., and Sabbe, K. (2020). Light Regulation of LHCX Genes in the Benthic Diatom *Seminavis robusta*. Front. Mar. Sci. 7.

Brembu, T., Chauton, M.S., Winge, P., Bones, A.M., and Vadstein, O. (2017). Dynamic responses to silicon in *Thalasiossira pseudonana* - Identification, characterisation and classification of signature genes and their corresponding protein motifs. Scientific Reports 7.

Caballero, M.A., Jallet, D., Shi, L., Rithner, C., Zhang, Y., and Peers, G. (2016). Quantification of chrysolaminarin from the model diatom *Phaeodactylum tricornutum*. Algal Research 20: 180–188.

Cahoon, L. (1999). The role of benthic microalgae in neritic ecosystems. Oceanogr. Mar. Biol. Annu. Rev. 37: 47–86.

Cahoon, L.B., Beretich, G.R., Thomas, C.J., and McDonald, A.M. (1993). Benthic microalgal production at Stellwagen Bank, Massachusetts Bay, USA. Marine Ecology Progress Series 102: 179–185.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973.

Carradec, Q. et al. (2018). A global ocean atlas of eukaryotic genes. Nat. Commun. 9: 373.

Carvalho, M.H.C. de, de Carvalho, M.H.C., Sun, H.-X., Bowler, C., and Chua, N.-H. (2016). Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. New Phytologist 210: 497–510.

Chauton, M.S., Winge, P., Brembu, T., Vadstein, O., and Bones, A.M. (2013). Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated to light/dark cycles. Plant Physiol. 161: 1034–1048.

Chisholm, S.W., Azam, F., and Eppley, R.W. (1978). Silicic acid incorporation in marine diatoms on light:dark cycles: Use as an assay for phased cell division 1. Limnology and Oceanography 23: 518–529.

Ciechanover, A. (2005). Proteolysis: from the lysosome to ubiquitin and the proteasome. Nat. Rev. Mol. Cell Biol. 6: 79–87.

Clavier, J., Chauvaud, L., Cuet, P., Esbelin, C., Frouin, P., Taddei, D., and Thouzeau, G. (2008). Diel variation of benthic respiration in a coral reef sediment (Reunion Island, Indian Ocean). Estuarine, Coastal and Shelf Science 76: 369–377.

Coesel, S., Mangogna, M., Ishikawa, T., Heijde, M., Rogato, A., Finazzi, G., Todo, T., Bowler, C., and Falciatore, A. (2009). Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. EMBO Rep. 10: 655–661.

Congestri, R. and Albertano, P. (2011). Benthic Diatoms in Biofilm Culture. The Diatom World: 227–243.

Consalvey, M., Paterson, D.M., and Underwood, G.J.C. (2004). The ups and downs of life in a benthic biofilm: migration of benthic diatoms. Diatom Research 19: 181–202.

Crane, B.R. and Young, M.W. (2014). Interactive features of proteins composing eukaryotic circadian clocks. Annu. Rev. Biochem. 83: 191–219.

Deniz, E. and Erman, B. (2017). Long noncoding RNA (lincRNA), a new paradigm in gene expression control. Funct. Integr. Genomics 17: 135–143.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Domazet-Lošo, T., Carvunis, A.-R., Albà, M.M., Šestak, M.S., Bakaric, R., Neme, R., and Tautz, D. (2017). No evidence for phylostratigraphic bias impacting inferences on patterns of gene emergence and evolution. Mol. Biol. Evol. 34: 843–856.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30: 1575–1584.

Eyre, B.D. and Ferguson, A.J.P. (2002). Comparison of carbon production and decomposition, benthic nutrient fluxes and denitrification in seagrass, phytoplankton, benthic microalgae- and macroalgae-dominated warm-temperate Australian lagoons. Marine Ecology Progress Series 229: 43–59.

Farré, E.M. (2020). The brown clock: circadian rhythms in stramenopiles. Physiologia Plantarum.

Ferrari, C., Proost, S., Janowski, M., Becker, J., Nikoloski, Z., Bhattacharya, D., Price, D., Tohge, T., Bar-Even, A., Fernie, A., Stitt, M., and Mutwil, M. (2019). Kingdom-wide comparison reveals the evolution of diurnal gene expression in Archaeplastida. Nat. Commun. 10: 737.

Fujiwara, T., Misumi, O., Tashiro, K., Yoshida, Y., Nishida, K., Yagisawa, F., Imamura, S., Yoshida, M., Mori, T., Tanaka, K., Kuroiwa, H., and Kuroiwa, T. (2009). Periodic gene expression patterns during the highly synchronized cell nucleus and organelle division cycles in the unicellular red alga *Cyanidioschyzon merolae*. DNA Res. 16: 59–72.

Gattuso, J.-P., Gentili, B., Duarte, C.M., Kleypas, J.A., Middelburg, J.J., and Antoine, D. (2006). Light availability in the coastal ocean: impact on the distribution of benthic photosynthetic organisms and their contribution to primary production. Biogeosciences 3: 489–513.

Gillard, J. et al. (2008). Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom *Seminavis robusta*. Plant Physiol. 148: 1394–1411.

Goldman, J.A.L., Schatz, M.J., Berthiaume, C.T., Coesel, S.N., Orellana, M.V., and Armbrust, E.V. (2019). Fe limitation decreases transcriptional regulation over the diel cycle in the model diatom *Thalassiosira pseudonana*. PLoS One 14: e0222325.

Grabherr, M.G. et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652.

Gruber, A., Rocap, G., Kroth, P.G., Armbrust, E.V., and Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J. 81: 519–528.

Gschloessl, B., Guermeur, Y., and Cock, J.M. (2008). HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinformatics 9: 393.

Hanlon, A.R.M., Bellinger, B., Haynes, K., Xiao, G., Hofmann, T.A., Gretz, M.R., Ball, A.S., Osborn, A.M., and Underwood, G.J.C. (2006). Dynamics of extracellular polymeric substance (EPS) production and loss in an estuarine, diatom-dominated, microalgal biofilm over a tidal emersion-immersion period. Limnology and Oceanography 51: 79–93.

Hardeland, R. (2018). On the Relationships between LncRNAs and Other Orchestrating Regulators: Role of the Circadian System. Epigenomes 2: 9.

Haro, S., Bohórquez, J., Lara, M., Garcia-Robledo, E., González, C.J., Crespo, J.M., Papaspyrou, S., and Corzo, A. (2019). Diel patterns of microphytobenthic primary production in intertidal sediments: the role of photoperiod on the vertical migration circadian rhythm. Sci. Rep. 9: 13376.

Hazen, S.P., Naef, F., Quisel, T., Gendron, J.M., Chen, H., Ecker, J.R., Borevitz, J.O., and Kay, S.A. (2009). Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. Genome Biol. 10: R17.

Hildebrand, M. and Lerch, S.J.L. (2015). Diatom silica biomineralization: Parallel development of approaches and understanding. Semin. Cell Dev. Biol. 46: 27–35.

Hope, J.A., Paterson, D.M., and Thrush, S.F. (2020). The role of microphytobenthos in soft-sediment ecological networks and their contribution to the delivery of multiple ecosystem services. Journal of Ecology 108: 815–830.

Huang, R., Ding, J., Gao, K., de Carvalho, M.H.C., Tirichine, L., Bowler, C., and Lin, X. (2019). A Potential Role for Epigenetic Processes in the Acclimation Response to Elevated pCO2 in the Model Diatom *Phaeodactylum tricornutum. Frontiers in Microbiology* 9.

Hughes, M.E., Hogenesch, J.B., and Kornacker, K. (2010). JTK_CYCLE: An efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. Journal of Biological Rhythms 25: 372–380.

Hutchison, A.L., Maienschein-Cline, M., Chiang, A.H., Tabei, S.M.A., Gudjonson, H., Bahroos, N., Allada, R., and Dinner, A.R. (2015). Improved statistical methods enable greater sensitivity in rhythm detection for genome-wide data. PLoS Comput. Biol. 11: e1004094.

Huysman, M.J.J. et al. (2013). AUREOCHROME1a-Mediated Induction of the Diatom-Specific Cyclin dsCYC2 Controls the Onset of Cell Division in Diatoms (*Phaeodactylum tricornutum*). The Plant Cell 25: 215–228.

Huysman, M.J.J., Martens, C., Vandepoele, K., Gillard, J., Rayko, E., Heijde, M., Bowler, C., Inzé, D., Van de Peer, Y., De Veylder, L., and Vyverman, W. (2010). Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. Genome Biol. 11: R17.

Huysman, M.J.J., Martens, C., Vyverman, W., and De Veylder, L. (2014). Protein degradation during the diatom cell cycle: Annotation and transcriptional analysis of SCF and APC/C ubiquitin ligase genes in Phaeodactylum tricornutum. Marine Genomics 14: 39–46.

Huysman, M.J.J., Tanaka, A., Bowler, C., Vyverman, W., and De Veylder, L. (2015). Functional characterization of the diatom cyclin-dependent kinase A2 as a mitotic regulator reveals plant-like properties in a non-green lineage. BMC Plant Biology 15.

Karlik, E., Ari, S., and Gozukirmizi, N. (2019). LncRNAs: genetic and epigenetic effects in plants. Biotechnology & Biotechnological Equipment 33: 429–439.

Katoh, K. and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30: 772–780.

Keeling, P.J. et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12: e1001889.

Kiefel, B.R., Gilson, P.R., and Beech, P.L. (2004). Diverse eukaryotes have retained mitochondrial homologues of the bacterial division protein FtsZ. Protist 155: 105–115.

Koedooder, C., Stock, W., Willems, A., Mangelinckx, S., De Troch, M., Vyverman, W., and Sabbe, K. (2019). Diatom-bacteria interactions modulate the composition and productivity of benthic diatom biofilms. Front. Microbiol. 10: 1255.

Kroth, P.G. et al. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLoS ONE 3: e1426.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35: 1547–1549.

Leger, M.M., Petrů, M., Žárský, V., Eme, L., Vlček, Č., Harding, T., Lang, B.F., Eliáš, M., Doležal, P., and Roger, A.J. (2015). An ancestral bacterial division system is widespread in eukaryotic mitochondria. Proc. Natl. Acad. Sci. U. S. A. 112: 10239–10246.

Liesecke, F., Daudu, D., Dugé de Bernonville, R., Besseau, S., Clastre, M., Courdavault, V., de Craene, J.-O., Crèche, J., Giglioli-Guivarc'h, N., Glévarec, G., Pichon, O., and Dugé de Bernonville, T. (2018). Ranking genome-wide correlation measurements improves microarray and RNA-seq based global and targeted co-expression networks. Sci. Rep. 8: 10885.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000

Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Li, J., Mahajan, A., and Tsai, M.-D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry 45: 15168–15178.

Li, Y., Xu, J., Shao, T., Zhang, Y., Chen, H., and Li, X. (2017). RNA Function Prediction. Methods Mol. Biol. 1654: 17–28.

Longphuirt, S.N., Leynaert, A., Guarini, J.M., Chauvaud, L., Claquin, P., Herlory, O., Amice, E., Huonnic, P., and Ragueneau, O. (2006). Discovery of microphytobenthos migration in the subtidal zone. Marine Ecology Progress Series 328: 143–154.

de Los Reyes, P., Romero-Campero, F.J., Ruiz, M.T., Romero, J.M., and Valverde, F. (2017). Evolution of daily gene co-expression patterns from algae to plants. Front. Plant Sci. 8: 1217.

MacIntyre, H.L., Geider, R.J., and Miller, D.C. (1996). Microphytobenthos: The Ecological Role of the “Secret Garden” of Unvegetated, Shallow-Water Marine Habitats. I. Distribution, Abundance and Primary Production. Estuaries 19: 186.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A., and Bowler, C. (2016). Insights into global diatom distribution and diversity in the world’s ocean. Proceedings of the National Academy of Sciences 113: E1516–E1525.

McConville, M.J., Wetherbee, R., and Bacic, A. (1999). Subcellular location and composition of the wall and secreted extracellular sulphated polysaccharides/proteoglycans of the diatom *Stauroneis amphioxys* Gregory. Protoplasma 206: 188–200.

Mittag, M. (2001). Circadian rhythms in microalgae. Int. Rev. Cytol. 206: 213–247.

Miyagishima, S.-Y., Era, A., Hasunuma, T., Matsuda, M., Hirooka, S., Sumiya, N., Kondo, A., and Fujiwara, T. (2019). Day/Night Separation of Oxygenic Energy Metabolism and Nuclear DNA Replication in the Unicellular Red Alga *Cyanidioschyzon merolae*. mBio 10.

Monnier, A., Liverani, S., Bouvet, R., Jesson, B., Smith, J.Q., Mosser, J., Corellou, F., and Bouget, F.-Y. (2010). Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. BMC Genomics 11: 192.

Moog, D. (2018). In silico tools for the prediction of protein import into secondary plastids. Methods Mol. Biol. 1829: 381–394.

Moulager, M., Corellou, F., Vergé, V., Escande, M.-L., and Bouget, F.-Y. (2010). Integration of light signals by the retinoblastoma pathway in the control of S phase entry in the picophytoplanktonic cell *Ostreococcus*. PLoS Genet. 6: e1000957.

Moulager, M., Monnier, A., Jesson, B., Bouvet, R., Mosser, J., Schwartz, C., Garnier, L., Corellou, F., and Bouget, F.-Y. (2007). Light-Dependent Regulation of Cell Division in *Ostreococcus*: Evidence for a Major Transcriptional Input. Plant Physiology 144: 1360–1369.

Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöh, O., and Persson, S. (2010). Assembly of an interactive correlation network for the *Arabidopsis* genome using a novel heuristic clustering algorithm. Plant Physiol. 152: 29–43.

Nakov, T., Beaulieu, J.M., and Alverson, A.J. (2019). Diatoms diversify and turn over faster in freshwater than marine environments. Evolution 73: 2497–2511.

Naylor, E. (2010). Chronobiology of marine organisms.

Nemoto, M., Iwaki, S., Moriya, H., Monden, Y., Tamura, T., Inagaki, K., Mayama, S., and Obuse, K. (2020). Comparative Gene Analysis Focused on Silica Cell Wall Formation: Identification of Diatom-Specific SET Domain Protein Methyltransferases. Mar. Biotechnol. .

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32: 268–274.

Nikaido, S.S. and Johnson, C.H. (2000). Daily and circadian variation in survival from ultraviolet radiation in *Chlamydomonas reinhardtii*. Photochem. Photobiol. 71: 758–765.

Nishikawa, T., Moriyama, Y., Sato, M., Sano, T., Hasezawa, S., Ota, S., and Kawano, S. (2012). Isolation of mitochondrial and plastid ftsZ genes and analysis of the organelle targeting sequence in the diatom *Chaetoceros neogracile* (Diatoms, Bacillariophyceae). Phycological Research 60: 123–136.

Nymark, M., Valle, K.C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., Bones, A.M., and Brembu, T. (2013). Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum.* PLoS One 8: e58722.

Osuna-Cruz, C.M. et al. (2020). The *Seminavis robusta* genome provides insights into the evolutionary adaptations of benthic diatoms. Nat. Commun. 11: 3320.

Palmer, J.D. and Round, F.E. (1967). Persistent, vertical-migration rhythms in benthic microflora. VI. The tidal and diurnal nature of the rhythm in the diatom *Hantzschia virgata*. The Biological Bulletin 132: 44–55.

Palmisano, A.C., SooHoo, J.B., White, D.C., Smith, G.A., Stanton, G.R., and Burckle, L.H. (2004). Shade adapted benthic diatoms beneath Antarctic sea ice. Journal of Phycology 21: 664–667.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14: 417–419.

Peña, M.R. de la and de la Peña, M.R. (2007). Cell growth and nutritive value of the tropical benthic diatom, *Amphora* sp., at varying levels of nutrients and light intensity, and different culture locations. Journal of Applied Phycology 19: 647–655.

Pesant, S. et al. (2015). Open science resources for the discovery and analysis of Tara Oceans data. Scientific Data 2.

Pinckney, J.L. (2018). A Mini-Review of the Contribution of Benthic Microalgae to the Ecology of the Continental Shelf in the South Atlantic Bight. Estuaries and Coasts 41: 2070–2078.

Poliner, E., Panchy, N., Newton, L., Wu, G., Lapinsky, A., Bullard, B., Zienkiewicz, A., Benning, C., Shiu, S.-H., and Farré, E.M. (2015). Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark cycles. The Plant Journal 83: 1097–1113.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Rambaut (2010). FigTree.

Reingruber, J., Holcman, D., and Fain, G.L. (2015). How rods respond to single photons: Key adaptations of a G-protein cascade that enable vision at the physical limit of perception. Bioessays 37: 1243–1252.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140.

Saburova, M.A. and Polikarpov, I.G. (2003). Diatom activity within soft sediments: behavioural and physiological processes. Marine Ecology Progress Series 251: 115–126.

Schlichting, M., Grebler, R., Menegazzi, P., and Helfrich-Förster, C. (2015). Twilight dominates over moonlight in adjusting *Drosophila'*s activity pattern. J. Biol. Rhythms 30: 117–128.

Schlichting, M., Grebler, R., Peschel, N., Yoshii, T., and Helfrich-Förster, C. (2014). Moonlight detection by *Drosophila'*s endogenous clock depends on multiple photopigments in the compound eyes. J. Biol. Rhythms 29: 75–86.

Schmid, A.-M.M. (1994). Aspects of morphogenesis and function of diatom cell walls with implications for taxonomy. The Protistan Cell Surface: 43–60.

Schnytzer, Y., Simon-Blecher, N., Li, J., Waldman Ben-Asher, H., Salmon-Divon, M., Achituv, Y., Hughes, M.E., and Levy, O. (2018). Tidal and diel orchestration of behaviour and gene expression in an intertidal mollusc. Scientific Reports 8.

Shrestha, R.P., Tesson, B., Norden-Krichmar, T., Federowicz, S., Hildebrand, M., and Allen, A.E. (2012). Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. BMC Genomics 13: 1–16.

Smith, D.J. and Underwood, G.J.C. (1998). Exopolymer production by intertidal epipelic diatoms. Limnology and Oceanography 43: 1578–1591.

Smith, S.R., Abbriano, R.M., and Hildebrand, M. (2012). Comparative analysis of diatom genomes reveals substantial

differences in the organization of carbon partitioning pathways. Algal Research 1: 2–16.

Smith, S.R., Gillard, J.T.F., Kustka, A.B., McCrow, J.P., Badger, J.H., Zheng, H., Ashley M. New, Dupont, C.L., Obata, T., Fernie, A.R., and Allen, A.E. (2016). Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLOS Genetics 12: e1006490.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 4: 1521.

Stojkovic, K., Wing, S.S., and Cermakian, N. (2014). A central role for ubiquitination within a circadian clock protein modification code. Front. Mol. Neurosci. 7: 69.

Takekata, H., Matsuura, Y., Goto, S.G., Satoh, A., and Numata, H. (2012). RNAi of the circadian clock gene period disrupts the circadian rhythm but not the circatidal rhythm in the mangrove cricket. Biology Letters 8: 488–491.

Tarrant, A.M., Helm, R.R., Levy, O., and Rivera, H.E. (2019). Environmental entrainment demonstrates natural circadian rhythmicity in the cnidarian *Nematostella vectensis*. The Journal of Experimental Biology 222: jeb205393.

Tesson, B. and Hildebrand, M. (2010). Extensive and intimate association of the cytoskeleton with forming silica in diatoms: control over patterning on the meso- and micro-scale. PLoS One 5: e14300.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7: 562–578.

Underwood, G.J.C. and Kromkamp, J. (1999). Primary production by phytoplankton and microphytobenthos in estuaries. Advances in Ecological Research: 93–153.

Underwood, G.J.C. and Paterson, D.M. (2003). The importance of extracellular carbohydrate production by marine epipelic diatoms.

UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47: D506–D515.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F., and Vandepoele, K. (2018). PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. 46: D1190–D1196.

Vaneechoutte, D. and Vandepoele, K. (2019). Curse: building expression atlases and co-expression networks from public RNA-Seq data. Bioinformatics 35: 2880–2881.

Vítová, M., Bišová, K., Umysová, D., Hlavová, M., Kawano, S., Zachleder, V., and Cížková, M. (2011). *Chlamydomonas reinhardtii*: duration of its cell cycle and phases at growth rates affected by light intensity. Planta 233: 75–86.

Wucher, V. et al. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Res. 45: e57.

Wu, T.D. and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859–1875.

Yamada, M. (2017). Functions of long intergenic non-coding (Linc) RNAs in plants. J. Plant Res. 130: 67–73.

Yang, X., Li, L., Wang, X., Yao, J., and Duan, D. (2020). Non-coding RNAs participate in the regulation of CRY-DASH in the growth and early development of *Saccharina japonica* (Laminariales, Phaeophyceae). International Journal of Molecular Sciences 21: 309.

Yu, J.-W. et al. (2008). COP1 and ELF3 control circadian function and photoperiodic flowering by regulating GI stability. Mol. Cell 32: 617–630.

Zampetaki, A., Albrecht, A., and Steinhofel, K. (2019). Long Non-coding RNA Structure and Function: Is There a Link? Frontiers in Physiology 9:1201.

Zhang, L., Hastings, M.H., Green, E.W., Tauber, E., Sladek, M., Webster, S.G., Kyriacou, C.P., and Wilcockson, D.C. (2013). Dissociation of circadian and circatidal timekeeping in the marine crustacean *Eurydice pulchra.* Curr. Biol. 23: 1863–1873.

Zinser, E.R., Lindell, D., Johnson, Z.I., Futschik, M.E., Steglich, C., Coleman, M.L., Wright, M.A., Rector, T., Steen, R., McNulty, N., Thompson, L.R., and Chisholm, S.W. (2009). Choreography of the transcriptome, photophysiology, and

cell cycle of a minimal photoautotroph, *Prochlorococcus*. PLoS One 4: e5135.

Zones, J.M., Blaby, I.K., Merchant, S.S., and Umen, J.G. (2015). High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. Plant Cell 27: 2743–2769.