A bHLH-PAS protein regulates light-dependent rhythmic processes in the marine diatom *Phaeodactylum tricornutum*

Rossella Annunziata[a,b,1], Andrés Ritter[a,1], Antonio Emidio Fortunato[a,c,1], Soizic Cheminant-Navarro[a], Nicolas Agier[a], Marie J. J. Huysman[a,d,e*], Per Winge[f], Atle Bones[f], François-Yves Bouget[h], Marco Cosentino Lagomarsino[a], Jean Pierre Bouly [a], and Angela Falciatore[a,2].

[a] Sorbonne Université, CNRS, Institut de Biologie Paris-Seine, Laboratory of Computational and Quantitative Biology, F-75005, Paris, France.

[b] Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy.

[c] ImmunRise Technologies, 75006 Paris, France.

[d]Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 927, B-9052 Ghent, Belgium.

[e]VIB Center for Plant Systems Biology, Technologiepark 927, B-9052 Ghent, Belgium.

[f]Department of Biology, NTNU Norwegian University of Science and Technology, 7491 Trondheim, Norway.

[h]Sorbonne Université, UMR 7621, Laboratoire d'Océanographie Microbienne, Observatoire Océanologique, Banyuls-sur-Mer, France.

[1] These authors contributed equally to this work.

*Disclaimer: The author (MJJH) is currently an employee of the European Research Council Executive Agency. The views expressed are purely those of the writer and may not in any circumstances be regarded as stating an official position of the European Commission.

[2]Corresponding Author:
Angela Falciatore
Diatom Functional Genomics Team
Laboratory of Computational and Quantitative Biology (LCQB)
Institut de Biologie Paris-Seine (IBPS)
Jussieu Campus – Sorbonne Université
4, Place Jussieu, 75005 Paris, France
Tel: +33 0144278142
Email: angela.falciatore@upmc.fr

Keywords: Diatoms, diurnal rhythms, transcription factors, cell division, gene expression.

ABSTRACT

Periodic light dark cycles govern the timing of basic biological processes in organisms inhabiting land as well as the sea, where life evolved. Although prominent marine phytoplanktonic organisms such as diatoms show robust diurnal rhythms in growth, cell cycle and gene expression, the molecular bases controlling these processes are still obscure. By exploring the regulatory landscape of diatom diurnal rhythms, we here unveil the key function of a *Phaeodactylum tricornutum* bHLH-PAS protein, named *Pt*bHLH1a, in the regulation of light-dependent rhythms. Peak expression of *PtbHLH1a* mRNA occurs at the end of the light period and it is adjusted to photoperiod changes. Ectopic over-expression of *Pt*bHLH1a results in lines with altered cell division and gene expression and showing a phase shift in diurnal responses, compared to the wild-type cells. Reduced oscillations in gene expression are also observed in continuous darkness, showing that the regulation of rhythmicity by *Pt*bHLH1a is not directly dependent on light inputs and cell division. *Pt*bHLH1a orthologs are widespread in both pennate and centric diatom genomes, hinting at a common function in many species. This study adds new elements to understand diatom biology and ecology and offers new perspectives to elucidate timekeeping mechanisms in marine organisms belonging to a major, but still underinvestigated branch of the tree of life.

INTRODUCTION

Owing to the Earth's rotation, life evolved in an environment governed by 24-hour periodic light-dark alternations. Most living organisms anticipate these changes by generating daily rhythms of many fundamental biological processes ranging from physiology to behaviour, which phase at optimal times of the day [1, 2]. Diurnal rhythms are accomplished by the coordinated action of signals derived from endogenous timekeepers, together with inputs derived by the environment and the metabolism [1-6]. The resulting synchronization of cellular and physiological events with the cyclic environment enhances the fitness of a variety of organisms [7-10]. Robust diel rhythms in growth, cell cycle, gene expression, pigment synthesis, phototactic movements or bioluminescence [11-19] have been also observed in a variety of phytoplanktonic organisms including diatoms. Diatoms represent the most species-rich group of algae in the ocean, which populate a wide range of aquatic environments [20-22]. These algae of Stramenopile lineage show peculiar genomic, metabolic and cellular features [23-27], and are evolutionary distant from the most studied model organisms in the field of biology rhythms [24, 28]. Diatoms have an impressive capacity to deal with

environmental changes thanks to sophisticated acclimation mechanisms [29-32]. Furthermore, important biological processes show robust diel rhythmicity in these algae [33, 34]. Recent genome wide analyses showed that 25% of the transcriptome is influenced by the light-dark cycles in the centric diatom *Thalasiossira pseudonana* [35]. Moreover, detailed diurnal studies in the pennate diatom *Phaeodactylum tricornutum* highlighted a strict temporal separation of transcriptional gene networks [36, 37]. A tight diurnal control of the *P. tricornutum* cell cycle has also been observed [33, 36, 38], with light onset triggering cell cycle progression through the transcriptional activation of the diatom-specific cyclin dsCYC2 by the blue light sensor Aureochrome-1a [38]. Furthermore, the functions of the transcription factor bZIP14 as diurnal activator of the TCA (tricarboxylic acid cycle) cycle, a process that is constrained to the late afternoon in diatoms, have been recently uncovered [39]. All together these studies highlight the complexity of diurnal activities regulation in *P. tricornutum*. However, the molecular mechanisms orchestrating such tight control are still unknown in diatoms, as well as in many phytoplanktonic organisms. Of note, except for cryptochromes [40-42], no orthologs of the circadian clock components discovered in bacteria, fungi, animals or plants have been found in the diatom genomes. Nonetheless, a number of proteins containing bHLH-PAS domains, which are involved in the regulation of rhythmic processes in animals [43, 44] have been identified [45].

By integrating transcriptomic, physiological and functional analyses, in this work we explore the regulatory landscape of *P. tricornutum* diurnal rhythms and uncover the functions of a bHLH-PAS encoding protein (bHLH1a) in the regulation of critical light-dependent rhythmic processes, such as cell division and diurnal transcription. Phylogenetic analyses reveal that bHLH1a orthologs are widely distributed in diatoms, allowing speculating a common function in many diatom species. These results open the way to new explorations aimed at dissecting the still cryptic molecular time keepers in diatoms.

RESULTS

Transcriptome profiling identifies potential regulators of diurnal rhythms in *P. tricornutum*

To pinpoint potential regulators of cellular rhythmicity in *P. tricornutum*, we analyzed a publicly available diurnal transcriptomic dataset [36] selecting 107 genes with robust diel oscillating expression. We focused our analysis on photoreceptors [38, 40-42, 46, 47] and transcription factors (TFs) [48], which might be involved in the generation of light-dependent rhythms and potential output genes implicated in rhythmic processes (pigment synthesis, cell

cycle regulation and photosynthesis) (Table S1). We further examined their transcriptional dynamics in 16 L (light):8 D (dark) photocycles for 32h. Hierarchical clustering (HCL) analysis of the resulting expression profiles revealed 4 clusters of co-expressed genes, termed A-D (Fig. 1A), with average expression profiles phasing at different moments of the day, from dawn to dusk (Fig. 1B). Cluster A peaked at the dark to light transition, suggesting a transcriptional anticipation of the light onset (Fig. 1A-B). This cluster gathers 18 genes including 14 TFs, mostly belonging to the Heat Shock Transcription Factor family (HSF), 2 DNA repair enzymes CPD photolyases (*CPD2* and *CPD3*) and one carotenoid synthesis enzyme (*PDS1*). Cluster B phased around 7h Zeitgeber Time (hours after illumination, ZT) and encompasses 36 genes (Fig. 1A-B). Out of the 18 TFs, 8 are sigma factors putatively involved in the regulation of chloroplast transcription. Moreover, the chlorophyll synthesis gene *POR1* and genes implicated in photoprotection (*LHCX1*, *Vdl2* and *Zep1*) were also included in cluster B, testifying for an active transcription of genes involved in chloroplast activity during the light period [30]. The blue light sensors Aurochrome1b and the cryptochromes *CPF1* and *CryP-like* also peaked at morning time. Cluster C holds genes phasing during early evening, around ZT9 (Fig. 1A-B), and comprises 9 TFs and 10 metabolism-related genes, including genes of the photosynthetic apparatus. Finally, cluster D includes 23 TF genes phasing before dusk, including the TCA cycle regulator *bZIP14* [36, 39], likely contributing to preparing cells for light to dark transition (Fig. 1A-B). The blue light *Aureochrome2* and *CPF4* and the Far-Red light sensing phytochrome (*DPH1*) also peaked at the end of the light period, consistently with previous observations [49, 50].

Put together, these results underline the existence of tight transcriptional programs phasing at discrete moments of the day in *P. tricornutum* and potentially timing specific cellular activities along the diurnal cycle.

***PtbHLH1a* expression is adjusted in a photoperiod-dependent manner**

Our analysis identified in clusters C and D two TFs holding a Per-ARNT-Sim (PAS) domain in conjunction with a bHLH DNA binding domain, respectively *Pt*bHLH1a (Phatr3_J44962) and *Pt*bHLH1b (Phatr3_J44963). Because of the involvement of bHLH-PAS proteins in the regulation of rhythmic processes in animals [43, 44], we decided to further examine the expression profiles of these diatom genes in cells growing under different photoperiods. *PtbHLH1a* expression peaked at ZT8 in 12L:12D photoperiod whereas under 16L:18D photoperiod this gene expression phased at ZT12, and then declined during the night to reach a trough at ZT0 (Fig. 1C). In both photoperiods, the maximum of expression of this gene was

observed around 4 hours before the end of the light period. Compared to *PtbHLH1a*, *PtbHLH1b* transcription started earlier. In cells entrained in 12L:12D cycles, *PtbHLH1b* maximum of expression phased at ZT8, whereas it peaked between ZT8/ZT12 in 16L:8D photoperiods (Fig. 1C). Thus, *PtbHLH1b* expression onset almost coincided in the two photoperiods although transcription dramatically dropped after ZT8 in 12L:12D, while remained at maximum levels up to ZT12 in long days (Fig. 1C).

We further tested robustness of *PtbHLH1a* and *PtbHLH1b* rhythmic expression patterns under stress conditions, by taking advantage of recent transcriptome datasets performed on *P. tricornutum* cells grown in 12L:12D cycles and in iron depletion conditions [37]. Iron homeostasis is diurnally regulated in phytoplankton [51] and it affects diurnal rhythmic processes such as cell cycle progression and diurnal gene expression in *P. tricornutum* [37]. Interestingly, *PtbHLH1a* and *PtbHLH1b* expression profiles showed similar patterns in both control and iron depleted conditions, with peaks of expression before dusk at ZT9 (Fig. S1), phenocopying what was observed in our experimental conditions (Fig. 1C).

Altogether these results demonstrate a robust control of *PtbHLH1a* and *PtbHLH1b* expression timing, which is not affected by iron deficiency and it is adjusted in a photoperiod-dependent manner. We decided to explore the involvement of *PtbHLH1a* in the regulation of diurnal light-dependent rhythmic processes considering its possible role in the dusk anticipation.

***PtbHLH1a* ectopic expression affects growth and cell cycle progression**

To determine *Pt*bHLH1a functions, we generated cell lines expressing HA-tagged bHLH1a under the regulation of the *Light harvesting complex protein family F2* promoter (*Lhcf2p*), activating transcription after 3h from light onset [52] (Fig. 2A). Gene expression analysis at ZT7 and ZT12 defined two independent lines, hereafter named as OE-1 and OE-2, showing over-expression of the *PtbHLH1a* gene already at ZT7 (Fig. 2B) compared to the wild type (Wt) strain. Remarkably, both *PtbHLH1a* over-expressing lines showed decreased growth rates compared to the Wt strain under 16L:8D cycles (Fig. 2C). The observed growth defects upon *PtbHLH1a* over-expression prompted us to thoroughly analyze cell cycle dynamics in lines OE-1 and OE-2. Accordingly, cell cultures were synchronized by 40h of dark treatment and harvested on an hourly basis for 12h after re-illumination. At T0, total DNA content measurements showed comparable proportions of cells in G1 phase reflecting the effective synchronization of both cell cultures (Fig. 2D). After 3h of illumination, a progressive reduction of G1 cell number was observed in the Wt, with a minimum reached after 10h of

illumination, following which the emergence of daughter cells increased again the percentage of cells in G1 (Fig. 2D). Interestingly, compared to Wt, OE-1 and OE-2 cells showed a slower exit from the G1 phase, with G1 cells reaching the lowest number only at the end of the time course, after 12h of illumination (Fig. 2D).

To further characterize the cell cycle deregulation caused by *PtbHLH1a* over-expression, we analyzed the expression profiles of specific cell cycle phase marker genes [33] in dark-synchronized Wt and OE-1 and OE-2 lines illuminated for 12h. The G1 phase gene markers *CDKA1* and *CDKD1* showed similar expression profiles (Fig. 2E) in the three lines. On the contrary, the G2/M marker *CYCB1* showed a drop in expression after 4h of irradiation in the OE-1 and OE-2 compared to the Wt (Fig. 2E). The expression of another G2/M phase marker, *CYCA/B1*, also resulted deregulated in OE-1 and OE-2, presenting reduced amplitude and peaking 2h later compared to Wt. Put together, these results suggest that the deregulation of *PtbHLH1a* affects growth, possibly by altering cell cycle transition from G1 to S or G2/M phases.

Cell division of many organisms is synchronized to the daily light-dark cycles through regulatory mechanisms integrating external light cues with internal oscillators [19, 53-56]. Synchronization of cell division is also observed in *P. tricornutum*, where divisions take place mostly toward the end of the light period and at night (Fig. S2, [19, 33, 55-58]). Based on these observations, we further inquired whether *bHLH1a* de-regulated expression could result in alterations of diel cell division dynamics. Cell growth of Wt and OE-1 cultures growing in 16L:8D photocycles was monitored every 2h for 4 days. In the Wt, diurnal growth displayed a characteristic “stair-step” shape dynamic, with a moderate increase in cell number at the beginning of the light period followed by an active increase from the late evening until the midnight (Fig. 3A). In contrast, this dynamic was altered in the OE-1 cells that showed a slower but more regular increase in cell number along day/night cycles. We next assessed daily cellular rhythmicity by following the course of the flow cytometer channel FL3 (excitation 488nm, emission 655-730 nm) that estimates chlorophyll a cellular content [57]. Cellular fluorescence displayed highly oscillating rhythms in 16L:8D grown cultures with a periodicity of approximately 24h (Fig. 3B, Table S2). Fluorescence in Wt cultures increased during daytime to phase around dusk at ZT14.68 ±0.3, then decreasing during the nighttime to reach a through around the early morning (Fig. 3B). Cellular fluorescence started to decline again before the dark period, concomitantly with an increase in cell concentrations (Fig. 3A) likely reflecting chloroplast partitioning to daughter cells during cell division [19]. Despite maintaining rhythmicity in the fluorescence dynamics, the OE-1 line displayed a remarkable

phase shift of around 1.5 h (ZT16.04 ±0.7) compared to the Wt. This phenotype may reflect a delayed or asynchronous cell division. We further investigated cellular phase responses in resetting experiment. With this aim, Wt and OE-1 cultures were grown in 16L:8D photoperiod, then transferred to 8L:16D and monitored for another 6 days. After transfer to 8L:16D photocycles, the phase of maximum fluorescence in Wt cells progressively re-synchronized with the newly imposed photoperiod, peaking at ZT7.67±0.72 after 3 days (Fig. 3C-D; Table S3). In contrast, after 3 days of re-entrainment under 8L:16D, the OE-1 line showed a 3 hour phase delay compared to Wt, peaking around dusk (Fig. 3C-D; Table S2). Put together, these results indicate that cellular rhythmicity is entrained in a photoperiod-dependent manner and that *Pt*bHLH1a deregulation alters the capacity of cells to set diurnal pattern phase.

_Pt_bHLH1a regulates pace of diel gene expression

We next questioned if *Pt*bHLH1a de-regulation could affect rhythmicity of other diurnal regulated processes such as gene expression. To this end, Wt and OE-1 lines were grown in 16L:8D photocycles and sampled every 3 hours during 25 hours. We selected for this analysis genes with strong diurnal transcription oscillation, including TFs (*bHLH1a*, *bHLH1b* and *bHLH3*) and rhythmic genes putatively involved in chlorophyll and carotenoid synthesis (*NADPH:protochlorophyllide oxidoreductase 2*, *Por2*, and *Violaxanthin de-epoxidase-related*, *Vdr*) (Fig. 1A, [36, 57]). As previously observed (Fig. 2B), total *PtbHLH1a* transcript levels, including endogenous and transgenic mRNAs, were increased compared to the Wt, and the expression phase anticipated at ZT7 in the OE-1 line (Fig. 3E). A decrease of endogenous *PtbHLH1a* transcripts was observed in the OE-1 line compared to the Wt, possibly reflecting a negative feed-back regulation of *Pt*bHLH1a on its own expression (Fig. 3E). A similar pattern was also observed for the *PtbHLH1b* gene, suggesting that *PtbHLH1a* and *PtbHLH1b* transcription may be controlled by the same regulatory circuit. Compared to the Wt, alterations in rhythmicity were observed in the OE-1 line for the *bHLH3* and *Por2* genes, that showed anticipated phases of expression, and for the *Vdr* gene presenting increased amplitude of expression, compared to the Wt (Fig. 3E).

We next questioned if the observed alterations in gene expression in the line with an ectopic *Pt*bHLH1a expression could be the consequence of the deregulation of cell division and cell cycle progression (Fig. 2 C-E and Fig. 3A). To test this hypothesis, we analyzed gene expression in dark conditions, when light inputs are lacking and cell cycle is arrested ([33] and (Fig. S3)). Because information about transcription dynamics in this condition is limited,

we first performed a survey of gene expression analysis of the previously selected 107 *P. tricornutum* diurnal rhythmic genes (Fig. 1A) in cells exposed to continuous dark for 30 hours. Comparative analysis of transcript profiles revealed that around 20% of the gene expression profiles show persistent oscillation in D:D, although displaying in some cases reduced amplitudes and-or shifted phases of expression compared to the 16L:8D condition (Fig. S4). Among the analyzed transcripts we also found genes that were severely affected by the absence of light, being strongly down-regulated or over-expressed when compared to the L:D condition (Fig. S5). The comparison of transcripts profiles in L:D and D:D allowed the identification of 20 genes showing the highest amplitude of expression in both conditions (for details see Materials and Methods), consisting of 16 putative TFs and 4 pigment related enzymes. We then analyzed the expression of some of these genes in constant darkness in Wt and *Pt*bHLH1a OE-1 cells for a period of 24h. In the Wt, the analyzed genes showed comparable transcript profiles in D:D and 16L:8D conditions (Fig. 3F, Fig. S6). Remarkably, 10 out of 13 tested genes displayed reduced amplitudes and shifts in the phase of expression in OE-1 compared to Wt in D:D (Fig. 3F, Fig. S4 and S6), revealing an even more pronounced deregulation in D:D compared to L:D (Fig. 3E and F). Taken together, these results suggest that *Pt*bHLH1a contributes to define diurnal gene expression patterns and that its activity is independent on direct light inputs and cell division.

_Pt_bHLH1a-like proteins are widely represented in the genome of marine algae

bHLH-PAS proteins were thought to be restricted to the animal (Opistokonta) lineage [59] until genome and transcriptome sequencing projects revealed bHLH-PAS family members in diatoms [45] and other microalgae [60]. Interestingly, when compared to animal bHLH-PAS, diatom proteins show peculiar features including a single predicted PAS domain (Fig. 4A) and a N-ter extension that is absent in the animal counterparts. These observations prompted us to further search bHLH-PAS proteins in the available transcriptomic and genomic databases of marine algae and animals, revealing ≈90 novel bHLH-PAS proteins from Rhodophyta, Cryptophyta, Stramenopila, Alveolata and basal Opistokonta organisms (Table S3). With one exception, all the newly identified proteins showed a single predicted PAS domain, short C-ter extensions and N-ter regions of variable length, resulting similar to the diatom bHLH-PAS proteins in terms of predicted structure (Fig. 4A). Notably, the only bHLH-PAS possessing two PAS domains like the animal proteins has been identified in *Galdieria sulphuraria* (Rhodophyta) and, to date, it represents the first TF of this family identified in Archaeplastida. All the identified sequences, including selected bHLH-PAS from

Opistokonta lineages, were used to perform a detailed phylogenetic analysis of the protein family using the bHLH and PAS domains. This analysis revealed at least three clades of algal bHLH-PAS proteins clearly separated from the Opistokonta TFs (Fig. 4B). Interestingly, domain organization and branching positions of proteins from basal Opistokonta (*Monosiga brevicollis*) and microalgae (*Guillardia theta* (Cryptophyta) and *Nannochloropsis* (Stramenopila)) (Fig. 4B) support a possible common origin for this TF family, from an ancestor featuring single bHLH and PAS domains. However, we cannot exclude the contribution of horizontal gene transfer and convergent evolution to the spreading and diversification of this family, as supported by the features of the *G. sulphuraria* bHLH-PAS protein, likely independently acquired by this alga. Based on our analysis, the majority of microalgal bHLH-PAS proteins falls into three separate clades: a first one hosting 9 TFs from diatoms and *Ectocarpus siliculosus*, a second one comprising *Pt*bHLH1a together with 35 proteins from diatoms and Alveolata (Dinoflagellata), and a third one grouping 41 proteins from Alveolata (Ciliophora and Dinoflagellata) and diatoms, including *Pt*bHLH1b (Fig. 4B).

Our results highlight a surprising diversification and widespread distribution of bHLH-PAS family members in different groups of algae. Moreover, the presence of *Pt*bHLH1a orthologs in the genome of dinoflagellate and diatom species suggests that these proteins may share similar functions in microalgae. The similarities in diel transcript regulation and timing of expression between *Pt*bHLH1a and the centric diatom *Thalassiosira pseudonana* ortholog, *TpHLH1* [35], further reinforce this hypothesis.

DISCUSSION

The daily cycle is marked by profound periodic light and temperature changes. These alternations have shaped most ecosystems on Earth including those evolving in the oceans. In most organisms, biological rhythms are generated by interconnected transcriptional-translational feed-back loops [61] involving TFs and integrating signals from the environment. Although this regulatory framework is conserved among eukaryotes, the actual molecular constituents of timing mechanisms seem to have emerged several times through evolution [1, 2, 62, 63]. Therefore, our current understanding of diurnal and circadian regulation, largely based on the study of terrestrial model organisms, is not always appropriate or relevant for marine organisms belonging to different branches of the tree of life. In this study, we shed light on the unknown components controlling diurnal patterns in diatoms, one of the most prominent phytoplanktonic groups in the Ocean. Diatoms show an impressive capacity to deal with light variations, thanks to efficient light perception and acclimation

mechanisms [26, 38, 42, 46, 64-67]. Here we show the existence of organized transcriptional programs controlling cellular activities along the daily cycle in *P. tricornutum*. In agreement with previous studies [36, 37], we identify a number of TFs exhibiting transcriptional oscillation and phasing at different moments during the 24h cycle, as novel candidates of diatom diurnal regulation. Furthermore, several photoreceptors have oscillating expression over the daytime (this study and [46, 50, 68]), suggesting that *P. tricornutum* coordinates diurnal transcription by integrating different environmental light inputs. We focus our investigation of diurnal rhythms regulators on a *P. tricornutum* bHLH-PAS protein, *Pt*bHLH1a, that appears among the most rhythmic transcripts and shows a photoperiod-dependent expression profile. *Pt*bHLH1a transcription anticipates the dark, by peaking 4 hours before the night onset in both short and long day photoperiods. The ability to entrain phase according to different photoperiods constitutes one defining criteria of a circadian clock regulated mechanism [69]. Interestingly, the pattern of *Pt*bHLH1a expression is maintained also in iron deprived cells, differently from what observed for many *P. tricornutum* transcripts [37]. These results suggest that this protein contributes to prepare the cells to the light-dark transitions and that its expression timing needs to be preserved, even under stressful conditions.

In relation, over-expression of *PtbHLH1a* results in altered diurnal progression of cell cycle. This phenotype is possibly explained by a delayed exit from the G1 phase in over-expressing lines compared to Wt, as also supported by the altered expression of cell cycle regulators [33]. Interestingly, a similar regulation of cell cycle occurs in mammalian cells, where the circadian clock controls the expression of G2 cycle-related genes to gate cell division at specific times of the day [70]. Previous data together with our results clearly show that cell division rate is accentuated in *P. tricornutum* during the first hours of the night. This observation is reminiscent of the circadian gating of cell cycle observed in many unicellular algae, where the clock constrains cell division at night time, therefore maximizing the energetic budget and avoiding deleterious effects from irradiation on chromatin [19, 34, 38, 70]. Remarkably, *PtbHLH1a* over-expression results in cells with deregulated diurnal division pattern, as further supported by our comparative cytometric analyses of cellular fluorescence. Daily cell division is indeed concomitant with the maximum cellular chlorophyll fluorescence that phases before dusk. Interestingly, *PtbHLH1a* over-expression induces a phase shift in this process towards the night period. Furthermore, we here demonstrate that diurnal cellular fluorescence is adjusted according to the photoperiod length and that alteration of *Pt*bHLH1a expression affects the resetting of the cell division phase. Collectively these results support

the existence of timing systems controlling cell division and suggest that *Pt*bHLH1a contributes to this regulation in *P. tricornutum*.

Besides cell division phenotypes, rhythmicity of several genes is affected upon overexpression of *Pt*bHLH1a. These genes show altered oscillations in phase and/or amplitude. Importantly, this phenotype is uncoupled from the cell cycle deregulation as demonstrated by transcript analysis in the absence of cell division at dark [33]. Remarkably, several analyzed genes maintain a rhythmic expression pattern in Wt cells for one day under continuous dark. In addition, the effect of *Pt*bHLH1a deregulation in gene transcription is much more pronounced when analyzed in D:D compared to L:D conditions. These results suggest on one hand that multiple regulatory inputs participate to the regulation of diurnal rhythmic gene expression [69], partially masking *Pt*bHLH1a contribution to this process in cyclic environments and, on the other hand, support *Pt*bHLH1a involvement in rhythms maintenance independently on light inputs. It is important to note that cells with deregulated *Pt*bHLH1a expression show slightly reduced growth rates. This result could be explained by the involvement of this protein in the regulation of cell division. However, it could also reflect the reduced capacity of cells to synchronize key metabolic processes according to the daily cycle as demonstrated for bacteria, plants and animals, which all display reduced fitness when losing their temporal endogenous programming [71-74].

By providing evidence of sophisticated timing mechanisms in diatoms, our study opens the way to future investigations of the networks generating diurnal and circadian rhythms in these organisms. As a bHLH-PAS containing protein, *Pt*bHLH1a likely acts as a TF controlling the expression of genes at the end of the light periods and participating to the regulation of the light to dark transitions. Furthermore, the TFs characterized in this study and showing altered expression patterns in *Pt*bHLH1a transgenic cells (*i.e.*, *bHLH1b*, *bHLH3*, *bZIP5*, *bZIP7*, *HSF1d*, *HSF1g*, *HSF3.3a*, *HSF4.7b*) represent direct or indirect targets of *Pt*bHLH1a activity and possible additional components of the network participating to diel rhythms regulation.

A series of proofs provided in this work also reinforce the hypothesis that *Pt*bHLH1a is one component of a still uncharacterized endogenous circadian clock in diatoms, either as part of a central oscillator or as a mediator of clock inputs or outputs. With the exception of CRY [42, 49], diatom genomes lack the canonical plant and animal circadian clock genes. However, *Pt*bHLH1a possesses bHLH and PAS protein domains that are also present in the CLOCK and BMAL proteins, the components of the central circadian oscillator of mammals [75, 76]. Interestingly, previous studies showed that the *P. tricornutum* animal-like blue light sensor Cpf1 can repress the transcriptional activity of these proteins in a heterologous

mammalian cell system [42], suggesting partial conservation in the regulatory program generating rhythmicity in animals and diatoms [77, 78]. Furthermore, as observed for the clock components, we show that *Pt*bHLH1a contributes to set the phase of output processes in cycling environment. Moreover, we show that diatoms maintain gene expression rhythmicity in constant darkness and that *Pt*bHLH1a contributes to this function. Even if further analyses under prolonged free running conditions will be necessary to conclusively assess the involvement of *Pt*bHLH1a in circadian regulation, this protein constitutes a promising entry point for the characterization of diatom molecular timekeepers. Finally, the discovery of the wide distribution of bHLH-PAS domain-containing proteins in diatoms, as well as in other algae, has the potential to shed new light on the evolution of biological rhythms. bHLH-PAS proteins might have independently acquired the function in rhythms regulation by convergent evolution. However, the existence of this function in an ancient heterotrophic marine ancestor that subsequently acquired plastids via endosymbiosis events [79] and before the conquer of land cannot be excluded. Regulators of cellular rhythmicity such as *Pt*bHLH1a may have played a critical role for diatom prominence in marine ecosystems, by synchronizing cellular events in optimal temporal programs and improving diatom ability to adapt to and anticipate environmental variations. Further characterization of *Pt*bHLH1a proteins in diatoms and other algae is expected to provide new insights into marine rhythms that are still poorly understood at molecular and mechanistic levels, despite their fundamental significance for growth, propagation and temporal distribution of phytoplankton, and in general of marine communities, in the oceans.

METHODS

Culture conditions

Wild-type *P. tricornutum* (*Pt*1 8.6; CCMP2561) cells and transgenic lines were grown in F/2 medium [80] without silica at 18°C, under shaking at 100 rpm, in 12L-12D or 16L-8D photocycles at 40 µmol photons m^{-2} s^{-1} of white light (Philips TL-D De Luxe Pro 950). For experiments in different photoperiods, cultures were pre-adapted to the different L:D cycles for 2 weeks before starting the experiment. For experiments in continuous darkness cells were pre-adapted in 16L:8D photocycles for 2 weeks, then transferred to D:D and collected every four hours for 24h. Growth measurements were performed using a MACSQuant Analyser flow cytometer (Miltenyi Biotec, Germany) by counting the cells based on the R1-A (630nm excitation, 670-700nm emission) versus the R1-H parameters. For detailed cell growth

analysis (multiple samples/day for several days), cells were harvested using a bench-top multi-channel liquid handling FC204 collector (Gilson, US), then stored at 4°C in continuous darkness until cytometric analysis. Phase rhythmicity assays were carried out by measuring the Chlorophyll fluorescence through the flow cytometer FL-3 parameter (488 nm excitation, 670-700 nm emission). For the re-entrainment experiment, cultures were initially entrained under 16L:8D photocycles at 40 µmol m^{-2} s^{-1} of white light, then transferred to 8L:16D photocycles at 80 µmol m^{-2} s^{-1} of white light for 6 days. All phase time, period and amplitude calculations were performed using the MFourfit curve-fitting method using the Biodare2 tool (biodare2.ed.ac.uk, [81]).

Cell cycle analysis

P. tricornutum and transgenic cells were synchronized in the G1 phase by 40h of prolonged darkness. After re-illumination, samples for cell cycle analysis were collected on an hourly basis for 12h. Cells were pelleted by centrifugation (4000 rpm, 15 minutes, 4°C), fixed in 70% EtOH and stored in the dark at 4°C until processing. Fixed cells were then washed three times with 1×PBS, stained with 4',6-diamidino-2-phenylindole (at a final concentration of 1 ng/ml) on ice for 30', then washed and resuspended in 1xPBS. After staining, samples were immediately analyzed with a MACsQuant Analyser flow cytometer (Miltenyi Biotec, Germany). For each sample 30,000 cells were analysed and G1 and G2 proportions were inferred by calculating the 2c and 4c peak areas at 450 nm (V1-A channel) using the R software. On the resulting histogram, a peak calling method was applied, based on a 1st derivative approach [82]. Peaks were first located in G1 and G2 reference samples and then used to locate the G1 and G2 cells in the other samples. The area under each peak was used as a proxy for the proportion of cells in each population.

RNA extraction and gene expression analyses

Total RNA was extracted and qRT-PCR performed as described in [46] using *PtRPS* and *PtTBP* as reference genes. Each independent replica of the qRT-PCR data was further normalized against the maximum expression value of each gene (*i.e.*, gene expression range lies between 0 and 1 in across the time series). Then, average expression and standard error was calculated and plotted. The full list of oligonucleotides used in this work can be found in Table S4. For the nCounter analysis, gene specific probes (Table S1) were designed and screened against the *P. tricornutum* annotated transcript database (JGI, genome version 2, Phatr2) for potential cross-hybridization. Total RNA extracts (100 ng) from three biological replicates were used for hybridization and transcript levels were measured using the nCounter

analysis system (Nanostring Technologies) at the UCL Nanostring Facility (London, UK) and at the Institut Curie technical platform (Paris, France) as previously described [83]. Expression values were first normalized against the internal spike-in controls, then against the geometric mean of the 3 reference genes *PtRPS*, *PtTBP* and *PtACTIN12*.

Selection of rhythmic transcripts and clustering analysis

To select genes with rhythmic expression in the light-dark cycle, we used microarray data from [36]. First, we identified all the genes belonging to the TFs, photoreceptors and metabolism-related categories (cell cycle progression, pigment synthesis and photosynthesis). Then, transcripts were ranked based on a defined t-value for each time point (mean gene expression of the replica/(1+s.d.)) and those showing t-value >+0.7 or <-0.7 across the time series, were retained. The nCounter expression data was normalized against the maximum expression value of each gene, in a similar way than qRT-PCR expression data. This normalization was carried independently on the 3 replicas and separately for each condition (L:D and D:D) with the average expression value used for the clustering analysis. (Table S2). Hierarchical clustering analysis was performed using MeV 4.9 [84] through Pearson correlation. Phase analysis was performed using the MFourfit curve-fitting method defining average expression phases for each cluster (biodare2.ed.ac.uk, [81]).

For the selection of rhythmic transcripts in D:D, expression values were normalized using as reference genes *PtRPS*, *PtTBP* and *PtACTIN12* and the genes with the highest values of standard deviation from the average expression (M value) over the two time courses (16L:8D and D:D) were selected. A threshold equal to 1 was set using the published P. tricornutum diurnal microarray dataset [36] as background; gene expression profiles were further empirically examined and false positive eliminated.

Generation of the *PtbHLH1a* overexpressing lines

The full length *PtbHLH1a* coding sequence was obtained by PCR amplification with the specific oligonucleotides *Pt*bHLH1a-DraI-Fw and *Pt*bHLH1a-XhoI-Rv on cDNA template using the Phusion high fidelity DNA polymerase (Thermo Fisher, USA). The PCR fragment was inserted into the pENTR1A vector (Invitrogen, USA) using the DraI/XhoI restriction sites, and recombined with the pDEST-C-HA vector [85]. Diatom transgenic lines were obtained by co-transformation of the final pDEST-C-HA-*PtbHLH1a* with Nourseothricin resistance plasmid as described [86] and tested for the presence of the transgene by PCR and qRT-PCR analysis (see Table S4 for oligonucleotide sequences).

Data mining, protein sequence and phylogenetic analysis

The *P. tricornutum* bHLH1a (Phatr3_J44962) protein sequence was used as query for blastP analyses on the JGI, NCBI and MMETSP public database [87]. Searches on the pfam database for proteins possessing both the HLH and PAS domains were also performed. The retrieved sequences were analyzed using the batch search on CDD (Conserved Domain Database) NCBI server to isolate only proteins presenting at least one HLH and one PAS domain. We identified 100 HLH-PAS proteins from 71 marine algal species. Multi sequence alignments were performed using MAFFT [88], including also 22 HLH-PAS proteins from relevant metazoan (*Homo sapiens*, *Mus musculus* and *Drosophila melanogaster*) and unicellular Opistokonta (*Monosiga brevicollis* and *Capsaspora owczarzaki*). Preliminary phylogenies were performed with MEGA 7 [20] to eliminate ambiguously aligned sequences, refining the alignment to 107 sequences and to a final length of 198 aa (<5% gap per position). Estimation of the best aminoacidic model given the data was performed with ProtTest 3.4.2 [89]. Phylogenetic analyses were performed with RAxML (1000 bootstraps) and MrBayes 3.2.6 (2.5 million generations, 2 runs, 25% burn-in) on the CIPRESS gateway [90]. The final tree was edited in FigureTree 1.4 (http://tree.bio.ed.ac.uk/software/Figure tree/). Genebank accession codes of the genes utilized in the bHLH-PAS phylogenetic analysis are reported in Table S3.

ACKNOWLEDGEMENTS

We thank M. Jaubert and L. De Veylder for critical suggestions, Kevin Saluan and Alessandro Manzotti for support with characterization of OE *Pt*bHLH1a lines, P. Oliveri for support with nCounter technology and D. Petroutsos and G. Finazzi for support in monitoring diatom physiology during experiments. This work was funded by the HFSP research grant (#RGY0082/2010), the Marie-Curie ITNs CALIPSO (FP7-PEPOPLE-2012 ITN 2013 GA 607607), and AccliPhot (FP7-PEPOPLE-2012- ITN; 316427), and a grant from the Gordon and Betty Moore Foundation GBMF 4966 to AF and AB.

COMPETING INTERESTS

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS: R.A., A.R., A.E.F., J.P.B. and A.F. designed research; R.A., A.R., A.E.F., S. C., and M.J.J.H. performed research; P.W. and A.B. provided the transcriptomic data used to select initial candidate TFs; FYB provided support for the cell

cycle analysis; R.A., A.R, A.E.F., N.A., F.Y.B., M.C.L., J.P.B. and A.F. analyzed the data; R.A., A.R., A.E.F. and A.F. wrote the paper. All the authors revised and approved the manuscript.

FIGURE LEGENDS

Fig. 1. Diurnal expression analysis of selected rhythmic *P. tricornutum* genes. A) nCounter expression analysis of 107 selected genes in cells grown under a 16L:8D photoperiod and sampled every 4 hours for 32 hours. Hierarchical clustering analysis defined four major groups (A-D) of co-regulated genes. In brown, blue and pink colors, transcription factors (TFs), metabolism and photoreceptor encoding genes were designated, respectively. Expression values were normalized using *PtRPS*, *PtTBP* and *PtACTIN12* as reference genes and represent the average of three biological replicates. Values are represented as relative to the maximum expression for each gene, where '1' corresponds to the highest expression value of the time series. The *PtbHLH1a* and *PtbHLH1b* gene expression profiles are indicated with arrowheads. B) Polar plot and table showing the average phases of expression (intended as expression peaks during the analyzed period) of the four gene clusters, calculated using the MFourfit averaged method. In the plot, petal length is proportional to the standard deviation of phases over biological replicates. Light and dark periods are schematized by white and grey regions respectively. N, number of genes within each cluster; Av. Period and Av. Phase, average phases and periods of expression in hours. C) Diurnal expression profiles of *PtbHLH1a* and *PtbHLH1b* in Wt cells grown under 12L:12D (blue line) and 16L:8D (red line) photocycles. qRT-PCR expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of three biological replicates ± s.e.m (standard error of the mean; black bars). For each gene, expression values are represented as relative to its maximum expression, corresponding in the graph to '1'. Light and dark periods are schematized by white and grey regions respectively.

Fig. 2. *PtbHLH1a* over-expression affects growth and cell cycle progression. A) Schematic representation of the *Lhcf2p:bHLH1a-3xHA:Lhcf1t* construct used to generate *Pt*bHLH1a over-expressing lines. *Lhcf2p*, Light Harvesting Complex F2 promoter; 3HA, triple hemagglutinin tag; *Lhcf1t*, *Lhcf1* terminator. B) Quantification of total *PtbHLH1a* transcripts in two independent *PtbHLH1a* over-expressing lines (OE-1, OE-2) compared to the wild type (Wt) strain. Cells were grown in 16L:8D photocycles and sampled at the ZT7 and ZT12 time points. qRT-PCR expression values were normalized using *PtRPS* and *PtTBP*

as reference genes and represent the average of two biological replicates (n=2) ± s.e.m (black bars). C) Growth curves of the Wt cells and the OE-1 and OE-2 over-expressing lines over 5 days in 12L:12D photocycle with calculated growth rates for each line. Values represent the average counts of three biological replicates (n=3) ± s.e.m (black bars). D) Cell cycle progression dynamics of Wt, OE-1 and OE-2 cell lines. Data represent the hourly proportion of cells in the G1 phase measured by flow cytometry over 12h of illumination following dark synchronization. Results are representative of three biological replicates ± s.e.m (black bars); *P<0.05, **P<0.01, t-test. E) Transcript expression profiles of G1 (*PtCDKA1*, *PtCDKD1*) and G2/M phase marker genes (*PtCYCB1*, *PtCYCA/B1*) in synchronized Wt, OE-1 and OE-2 cell lines over 12h of illumination. Expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of two independent biological replicates. For each gene, expression values are represented as relative to its maximum expression corresponding in the graph to '1'.

Fig. 3. *PtbHLH1a* over-expression alters rhythmic diel cell growth and gene expression. A) Diel cell growth measurements of Wt (yellow) and OE-1 cell line (red) during 4 days under 16L:8D photocycles. Values represent the mean cell count of three independent experiments ± s.e.m. White and grey regions represent illuminated and dark periods respectively. B) Diurnal oscillation of chlorophyll fluorescence in Wt compared to the OE-1 cell line. Cells were entrained under 16L:8D cycles and the FL-3 parameter was monitored for three days. FL-3 values represent the mean of three independent experiments ± s.e.m. and were normalized and aligned by the mean FL-3 value of each replicate. C) Phase re-entrainment analysis of fluorescence rhythms after photoperiod change in Wt compared to the OE-1 cell line. The FL-3 parameter was monitored in cultures grown in 16L:8D photocycles for one day and then transferred to 8L:16D for 6 days. D) Bar plot representation of the initial fluorescence phases in Wt and OE-1 cultures entrained under 16L:8D photocycles and after three days of 8L:16D regime change. Values represent the average of 2 biological replicates ± s.e.m. (black bars). E) qRT-PCR diurnal expression analysis of *PtbHLH1a*, *PtbHLH1b*, *PtbHLH3*, *PtPor2* and *PtVdr* transcripts in 16L:8D entrained Wt and OE-1 cultures. *bHLH1a-total* represents the sum expression of the *bHLH1a* endogenous and the transgene transcripts; *PtbHLH1a-endogenous* refers to the transcript levels of the endogenous gene only. Grey rectangles represent dark periods. Expression values represent the average of 2 biological replicates ± s.e.m. (black bars) and were normalized using *PtRPS* and *PtTBP* as reference genes. For each gene, expression values are represented as relative to its maximum expression

corresponding in the graph to '1'. F) nCounter expression analysis of *PtbHLH1a*, *PtbHLH1b*, *PtbHLH3*, *PtPor2* and *PtVdr* transcripts during 24 hours of dark free-running period in the Wt and OE-1 lines. Cells were previously entrained in 16L:8D cycles. Grey rectangles represent dark periods. Expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of 3 biological replicates ± s.e.m (black bars). For each gene, expression values are represented as relative to its maximum expression corresponding in the graph to '1'.

Fig. 4. bHLH-PAS protein family structure and phylogeny. A) Schematic representation of the domain architecture of bHLH-PAS proteins across Eukaryotes. Segmented line indicates possible absence of the second PAS domain in some Opistokonta species. The grey patterns before the bHLH domain and after the PAS domain schematize the variations in length of N-ter and C-ter regions in bHLH-PAS proteins of different organisms. B) Maximum Likelihood (ML) phylogenetic tree of the bHLH-PAS family. The Opistokonta clade is used as outgroup. Numbers refer to bootstrap values of the basal nodes using different methods: ML (RAxML, 1000 bootstraps) and Bayesian Inference (MrBayes, 2.5M generations, 25% burn-in). The asterisk highlights the position of *Monosiga brevicollis* bHLH-PAS; the arrows indicate the position of *Pt*HLH1a and *Pt*bHLH1b; the square indicates the position of *Tp*bHLH1. The colour code indicates the corresponding lineage for each bHLH-PAS protein, as emphasized in Fig 4-A.

REFERENCES

1. Pittendrigh, C.S., *Temporal Organization: Reflections of a Darwinian Clock-Watcher.* Annual Review of Physiology, 1993. **55**(1): p. 17-54.
2. Bell-Pedersen, D., et al., *Circadian rhythms from multiple oscillators: lessons from diverse organisms.* Nature reviews. Genetics, 2005. **6**(7): p. 544-56.
3. Fu, M. and X. Yang, *The sweet tooth of the circadian clock.* Biochemical Society Transactions, 2017. **45**(4): p. 871-884.
4. Haydon, M.J., et al., *Photosynthetic entrainment of the Arabidopsis thaliana circadian clock.* Nature, 2013. **502**: p. 689.
5. Brody, S., *The genetics of circadian rhythms. Introduction.* Advances in genetics, 2011. **74**: p. 1-12.
6. Dunlap, J.C., *Molecular Bases for Circadian Clocks.* Cell. **96**(2): p. 271-290.
7. Michael, T.P., et al., *Enhanced Fitness Conferred by Naturally Occurring Variation in the Circadian Clock.* Science, 2003. **302**(5647): p. 1049-1053.
8. Bass, J. and M.A. Lazar, *Circadian time signatures of fitness and disease.* Science, 2016. **354**(6315): p. 994-999.

9. Dodd, A.N., et al., *Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage.* Science, 2005. **309**(5734): p. 630-633.
10. Johnson, C.H. and S.S. Golden, *Circadian Programs in Cyanobacteria: Adaptiveness and Mechanism.* Annual Review of Microbiology, 1999. **53**(1): p. 389-409.
11. Brierley, A.S., *Diel vertical migration.* Current Biology, 2014. **24**(22): p. R1074-R1076.
12. Chisholm, S.W., *Temporal Patterns of Cell-Division in Unicellular Algae.* Canadian Bulletin of Fisheries and Aquatic Sciences, 1981(210): p. 150-181.
13. Naylor, E., *Chronobiology of marine organisms*. 2010, Cambridge ; New York: Cambridge University Press. x, 242 p.
14. Ottesen, E.A., et al., *Pattern and synchrony of gene expression among sympatric marine microbial populations.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(6): p. E488-E497.
15. Ottesen, E.A., et al., *Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages.* Science, 2014. **345**(6193): p. 207-12.
16. Raible, F., H. Takekata, and K. Tessmar-Raible, *An Overview of Monthly Rhythms and Clocks.* Frontiers in Neurology, 2017. **8**.
17. Tessmar-Raible, K., F. Raible, and E. Arboleda, *Another place, another timer: Marine species and the rhythms of life.* Bioessays, 2011. **33**(3): p. 165-172.
18. Häfker, N.S., et al., *Circadian Clock Involvement in Zooplankton Diel Vertical Migration.* Current Biology. **27**(14): p. 2194-2201.e3.
19. Moulager, M., et al., *Light-dependent regulation of cell division in Ostreococcus: evidence for a major transcriptional input.* Plant Physiol, 2007. **144**(3): p. 1360-9.
20. Malviya, S., et al., *Insights into global diatom distribution and diversity in the world's ocean.* Proc Natl Acad Sci U S A, 2016. **113**(11): p. E1516-25.
21. Field, C.B., et al., *Primary production of the biosphere: integrating terrestrial and oceanic components.* Science, 1998. **281**(5374): p. 237-40.
22. de Vargas, C., et al., *Eukaryotic plankton diversity in the sunlit ocean.* Science, 2015. **348**(6237).
23. Moustafa, A., et al., *Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms.* Science, 2009. **324**(5935): p. 1724-1726.
24. Baldauf, S.L., *The Deep Roots of Eukaryotes.* Science, 2003. **300**(5626): p. 1703-1706.
25. Bowler, C., et al., *The Phaeodactylum genome reveals the evolutionary history of diatom genomes.* Nature, 2008. **456**: p. 239.
26. Bailleul, B., et al., *Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms.* Nature, 2015. **524**: p. 366.
27. Maier, U.G., S. Zauner, and F. Hempel, *Protein import into complex plastids: Cellular organization of higher complexity.* European Journal of Cell Biology, 2015. **94**(7): p. 340-348.
28. Raible, F. and A. Falciatore, *It's about time: Rhythms as a new dimension of molecular marine research.* Marine genomics, 2014. **14**: p. 1-2.
29. Goss, R. and B. Lepetit, *Biodiversity of NPQ.* Journal of Plant Physiology, 2015. **172**: p. 13-32.
30. Bailleul, B., et al., *An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(42): p. 18214-9.
31. Allen, A.E., et al., *Evolution and metabolic significance of the urea cycle in photosynthetic diatoms.* Nature, 2011. **473**: p. 203.
32. Morrissey, J., et al., *A Novel Protein, Ubiquitous in Marine Phytoplankton, Concentrates Iron at the Cell Surface and Facilitates Uptake.* Current Biology. **25**(3): p. 364-371.
33. Huysman, M.J., et al., *Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling.* Genome biology, 2010. **11**(2): p. R17.
34. Ragni, M. and M.R. D'Alcalà, *Circadian variability in the photobiology of Phaeodactylum tricornutum: pigment content.* Journal of Plankton Research, 2007. **29**(2): p. 141-156.

35. Ashworth, J., et al., *Genome-wide diel growth state transitions in the diatom Thalassiosira pseudonana.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(18): p. 7518-23.
36. Chauton, M.S., et al., *Gene regulation of carbon fixation, storage, and utilization in the diatom Phaeodactylum tricornutum acclimated to light/dark cycles.* Plant Physiol, 2013. **161**(2): p. 1034-48.
37. Smith, S.R., et al., *Transcriptional Orchestration of the Global Cellular Response of a Model Pennate Diatom to Diel Light Cycling under Iron Limitation.* PLoS genetics, 2016. **12**(12): p. e1006490.
38. Huysman, M.J., et al., *AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (Phaeodactylum tricornutum).* The Plant cell, 2013. **25**(1): p. 215-28.
39. Matthijs, M., et al., *The transcription factor bZIP14 regulates the TCA cycle in the diatom Phaeodactylum tricornutum.* The EMBO journal, 2017. **36**(11): p. 1559-1576.
40. Juhas, M., et al., *A novel cryptochrome in the diatom Phaeodactylum tricornutum influences the regulation of light-harvesting protein levels.* FEBS Journal, 2014. **281**(9): p. 2299-2311.
41. König, S., et al., *The Influence of a Cryptochrome on the Gene Expression Profile in the Diatom Phaeodactylum tricornutum under Blue Light and in Darkness.* Plant and Cell Physiology, 2017. **58**(11): p. 1914-1923.
42. Coesel, S., et al., *Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity.* EMBO reports, 2009. **10**(6): p. 655-61.
43. Young, M.W. and S.A. Kay, *Time zones: a comparative genetics of circadian clocks.* Nature reviews. Genetics, 2001. **2**(9): p. 702-15.
44. Kewley, R.J., M.L. Whitelaw, and A. Chapman-Smith, *The mammalian basic helix-loop-helix/PAS family of transcriptional regulators.* Int J Biochem Cell Biol, 2004. **36**(2): p. 189-204.
45. Rayko, E., et al., *Transcription factor families inferred from genome sequences of photosynthetic stramenopiles.* The New phytologist, 2010. **188**(1): p. 52-66.
46. Fortunato, A.E., et al., *Diatom Phytochromes Reveal the Existence of Far-Red-Light-Based Sensing in the Ocean.* The Plant cell, 2016. **28**(3): p. 616-28.
47. Jaubert, M., et al., *Light sensing and responses in marine microalgae.* Current Opinion in Plant Biology, 2017. **37**(Supplement C): p. 70-77.
48. Rayko, E., et al., *Transcription factor families inferred from genome sequences of photosynthetic stramenopiles.* New Phytologist, 2010. **188**(1): p. 52-66.
49. Fortunato, A.E., et al., *Dealing with light: the widespread and multitasking cryptochrome/photolyase family in photosynthetic organisms.* Journal of plant physiology, 2015. **172**: p. 42-54.
50. Oliveri, P., et al., *The Cryptochrome/Photolyase Family in aquatic organisms.* Marine genomics, 2014. **14**: p. 23-37.
51. Botebol, H., et al., *Central role for ferritin in the day/night regulation of iron homeostasis in marine phytoplankton.* Proc Natl Acad Sci U S A, 2015. **112**(47): p. 14652-7.
52. Russo, M.T., et al., *The upstream regulatory sequence of the light harvesting complex Lhcf2 gene of the marine diatom Phaeodactylum tricornutum enhances transcription in an orientation- and distance-independent fashion.* Marine genomics, 2015. **24 Pt 1**: p. 69-79.
53. Johnson, C.H., *Circadian clocks and cell division: what's the pacemaker?* Cell Cycle, 2010. **9**(19): p. 3864-73.
54. Nozue, K., et al., *Rhythmic growth explained by coincidence between internal and external cues.* Nature, 2007. **448**(7151): p. 358-61.
55. Goto, K. and C.H. Johnson, *Is the cell division cycle gated by a circadian clock? The case of Chlamydomonas reinhardtii.* The Journal of Cell Biology, 1995. **129**(4): p. 1061-1069.
56. Sweeney, B.M. and J.W. Hastings, *Rhythmic Cell Division in Populations of Gonyaulax polyedra.* The Journal of Protozoology, 1958. **5**(3): p. 217-224.

57. Hunsperger, H.M., et al., *Differential Regulation of Duplicate Light-Dependent Protochlorophyllide Oxidoreductases in the Diatom Phaeodactylum tricornutum.* PLOS ONE, 2016. **11**(7): p. e0158614.
58. Miyagishima, S.Y., et al., *FtsZ-less prokaryotic cell division as well as FtsZ- and dynamin-less chloroplast and non-photosynthetic plastid division.* Front Plant Sci, 2014. **5**: p. 459.
59. Yan, J., et al., *Evolution, functional divergence and conserved exon-intron structure of bHLH/PAS gene family.* Mol Genet Genomics, 2014. **289**(1): p. 25-36.
60. Thiriet-Rupert, S., et al., *Transcription factors in microalgae: genome-wide prediction and comparative analysis.* BMC genomics, 2016. **17**: p. 282.
61. Hurley, J.M., J.J. Loros, and J.C. Dunlap, *Circadian Oscillators: Around the Transcription-Translation Feedback Loop and on to Output.* Trends Biochem Sci, 2016. **41**(10): p. 834-46.
62. Palmer, J.D., F.A. Brown, and L.N. Edmunds, *An introduction to biological rhythms*. 1976, New York: Academic Press. xvii, 375 p.
63. Bhadra, U., et al., *Evolution of circadian rhythms: from bacteria to human.* Sleep Medicine, 2017. **35**: p. 49-61.
64. Flori, S., et al., *Plastid thylakoid architecture optimizes photosynthesis in diatoms.* Nature communications, 2017. **8**: p. 15885.
65. Lepetit, B., et al., *The diatom Phaeodactylum tricornutum adjusts nonphotochemical fluorescence quenching capacity in response to dynamic light via fine-tuned Lhcx and xanthophyll cycle pigment synthesis.* The New phytologist, 2017. **214**(1): p. 205-218.
66. Depauw, F.A., et al., *Exploring the molecular basis of responses to light in marine diatoms.* Journal of experimental botany, 2012. **63**(4): p. 1575-91.
67. Juhas, M., et al., *A novel cryptochrome in the diatom Phaeodactylum tricornutum influences the regulation of light-harvesting protein levels.* The FEBS journal, 2014. **281**(9): p. 2299-311.
68. Banerjee, A., et al., *Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes.* Nucleic acids research, 2016. **44**(12): p. 5957-70.
69. Merrow, M., et al., *Entrainment of the Neurospora circadian clock.* Chronobiol Int, 2006. **23**(1-2): p. 71-80.
70. Matsuo, T., et al., *Control mechanism of the circadian clock for timing of cell division in vivo.* Science, 2003. **302**(5643): p. 255-9.
71. Woelfle, M.A., et al., *The adaptive value of circadian clocks: an experimental assessment in cyanobacteria.* Current biology : CB, 2004. **14**(16): p. 1481-6.
72. Beaver, L.M., et al., *Loss of circadian clock function decreases reproductive fitness in males of Drosophila melanogaster.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2134-9.
73. Dodd, A.N., et al., *Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage.* Science, 2005. **309**(5734): p. 630-3.
74. Takahashi, J.S., et al., *The genetics of mammalian circadian order and disorder: implications for physiology and disease.* Nature reviews. Genetics, 2008. **9**(10): p. 764-75.
75. Bunger, M.K., et al., *Mop3 is an essential component of the master circadian pacemaker in mammals.* Cell, 2000. **103**(7): p. 1009-1017.
76. Gekakis, N., et al., *Role of the CLOCK protein in the mammalian circadian mechanism.* Science, 1998. **280**(5369): p. 1564-9.
77. Nusinow, D.A., et al., *The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth.* Nature, 2011. **475**: p. 398.
78. Martín, G., et al., *Circadian Waves of Transcriptional Repression Shape PIF-Regulated Photoperiod-Responsive Growth in Arabidopsis.* Current Biology. **28**(2): p. 311-318.e5.
79. Archibald, J.M., *Endosymbiosis and Eukaryotic Cell Evolution.* Current biology : CB, 2015. **25**(19): p. R911-21.
80. Guillard, R.R.L., *Culture of Phytoplankton for Feeding Marine Invertebrates. In Culture of Marine Invertebrate Animals, D.R.*
Smith and M.H. Chanley, eds. Springer US. 1975.

81. Zielinski, T., et al., *Strengths and limitations of period estimation methods for circadian data.* PloS one, 2014. **9**(5): p. e96462.
82. Agier, N. and G. Fischer, *A Versatile Procedure to Generate Genome-Wide Spatiotemporal Program of Replication in Yeast Species.* Methods in molecular biology, 2016. **1361**: p. 247-64.
83. Geiss, G.K., et al., *Direct multiplexed measurement of gene expression with color-coded probe pairs.* Nat Biotechnol, 2008. **26**(3): p. 317-25.
84. Sharma, A.K., et al., *Distinct dissolved organic matter sources induce rapid transcriptional responses in coexisting populations of Prochlorococcus, Pelagibacter and the OM60 clade.* Environmental Microbiology, 2014. **16**(9): p. 2815-30.
85. Siaut, M., et al., *Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum.* Gene, 2007. **406**(1-2): p. 23-35.
86. De Riso, V., et al., *Gene silencing in the marine diatom Phaeodactylum tricornutum.* Nucleic acids research, 2009. **37**(14): p. e96.
87. Keeling, P.J., et al., *The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing.* Plos Biology, 2014. **12**(6).
88. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Molecular biology and evolution, 2013. **30**(4): p. 772-80.
89. Darriba, D., et al., *ProtTest 3: fast selection of best-fit models of protein evolution.* Bioinformatics, 2011. **27**(8): p. 1164-5.
90. Miller, M.A., Pfeiffer, W., and Schwartz, T. , *Creating the CIPRES Science Gateway for Inference of Large PhylogeneticTrees.* In Proceedings of the Gateway Computing Environments Workshop (GCE) (New Orleans, LA), 2010: p. 1-8.

Figure 1 - Annunziata et al.

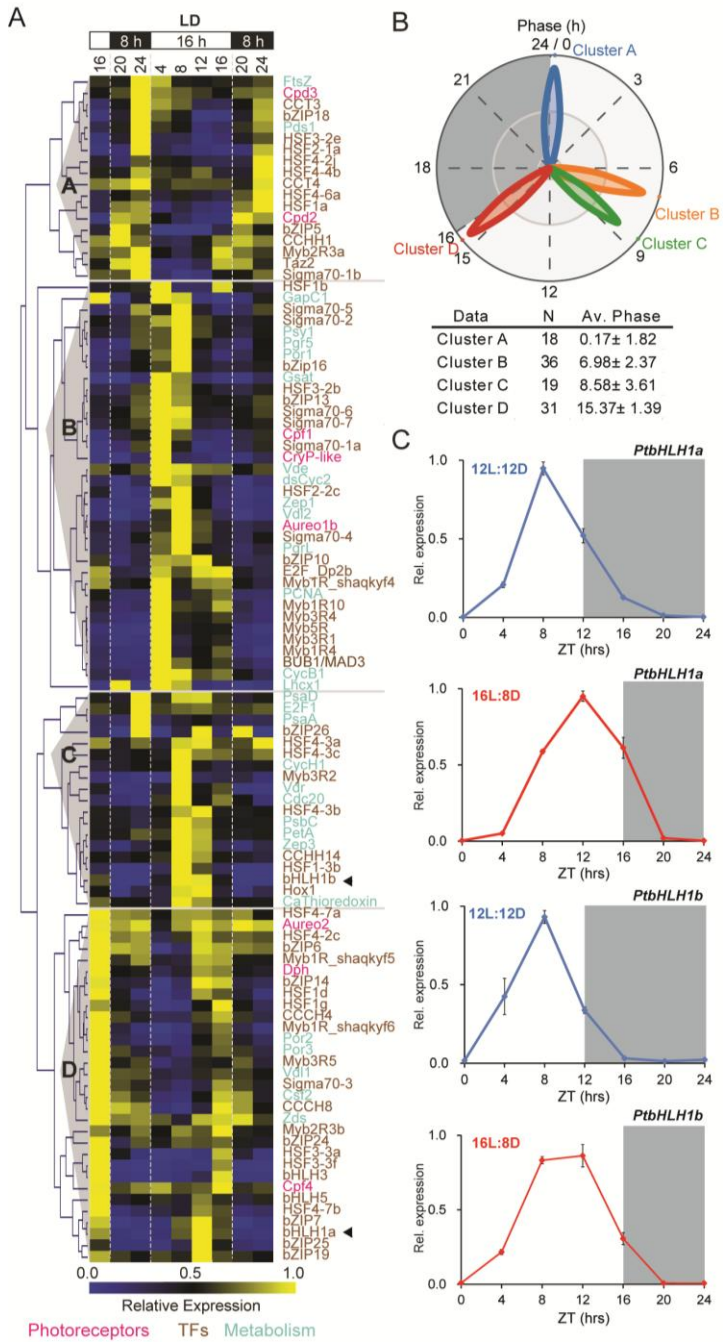

Fig. 1. Diurnal expression analysis of selected rhythmic *P. tricornutum* genes. A) nCounter expression analysis of 107 selected genes in cells grown under a 16L:8D photoperiod and sampled every 4 hours for 32 hours. Hierarchical clustering analysis defined four major groups (A-D) of co-regulated genes. In brown, blue and pink colors, transcription factors (TFs), metabolism and photoreceptor encoding genes were designated, respectively. Expression values were normalized using *PtRPS*, *PtTBP* and *PtACTIN12* as reference genes

and represent the average of three biological replicates. Values are represented as relative to the maximum expression for each gene, where ‘1’ corresponds to the highest expression value of the time series. The *PtbHLH1a* and *PtbHLH1b* gene expression profiles are indicated with arrowheads. B) Polar plot and table showing the average phases of expression (intended as expression peaks during the analyzed period) of the four gene clusters, calculated using the MFourfit averaged method. In the plot, petal length is proportional to the standard deviation of phases over biological replicates. Light and dark periods are schematized by white and grey regions respectively. N, number of genes within each cluster; Av. Period and Av. Phase, average phases and periods of expression in hours. C) Diurnal expression profiles of *PtbHLH1a* and *PtbHLH1b* in Wt cells grown under 12L:12D (blue line) and 16L:8D (red line) photocycles. qRT-PCR expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of three biological replicates ± s.e.m (standard error of the mean; black bars). For each gene, expression values are represented as relative to its maximum expression, corresponding in the graph to ‘1’. Light and dark periods are schematized by white and grey regions respectively.

Figure 2 - Annunziata et al.

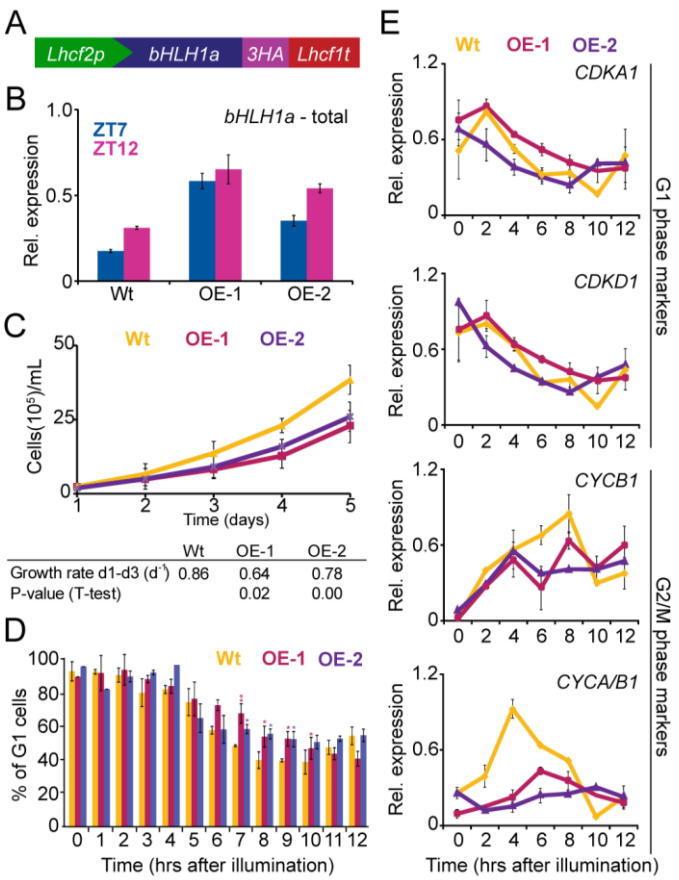

Fig. 2. *PtbHLH1a* over-expression affects growth and cell cycle progression. A) Schematic representation of the *Lhcf2p:bHLH1a-3xHA:Lhcf1t* construct used to generate *Pt*bHLH1a over-expressing lines. *Lhcf2p*, Light Harvesting Complex F2 promoter; 3HA, triple hemagglutinin tag; *Lhcf1t*, *Lhcf1* terminator. B) Quantification of total *PtbHLH1a* transcripts in two independent *PtbHLH1a* over-expressing lines (OE-1, OE-2) compared to the wild type (Wt) strain. Cells were grown in 16L:8D photocycles and sampled at the ZT7 and ZT12 time points. qRT-PCR expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of two biological replicates (n=2) ± s.e.m (black bars). C) Growth curves of the Wt cells and the OE-1 and OE-2 over-expressing lines over 5 days in 12L:12D photocycle with calculated growth rates for each line. Values represent the average counts of three biological replicates (n=3) ± s.e.m (black bars). D) Cell cycle progression dynamics of Wt, OE-1 and OE-2 cell lines. Data represent the hourly proportion of cells in the G1 phase measured by flow cytometry over 12h of illumination following dark synchronization. Results are representative of three biological replicates ± s.e.m (black bars); *P<0.05, **P<0.01, t-test. E) Transcript expression profiles of G1 (*PtCDKA1*, *PtCDKD1*) and

G2/M phase marker genes (*PtCYCB1*, *PtCYCA/B1*) in synchronized Wt, OE-1 and OE-2 cell lines over 12h of illumination. Expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of two independent biological replicates. For each gene, expression values are represented as relative to its maximum expression corresponding in the graph to ‘1’.

Figure 3 - Annunziata et al.

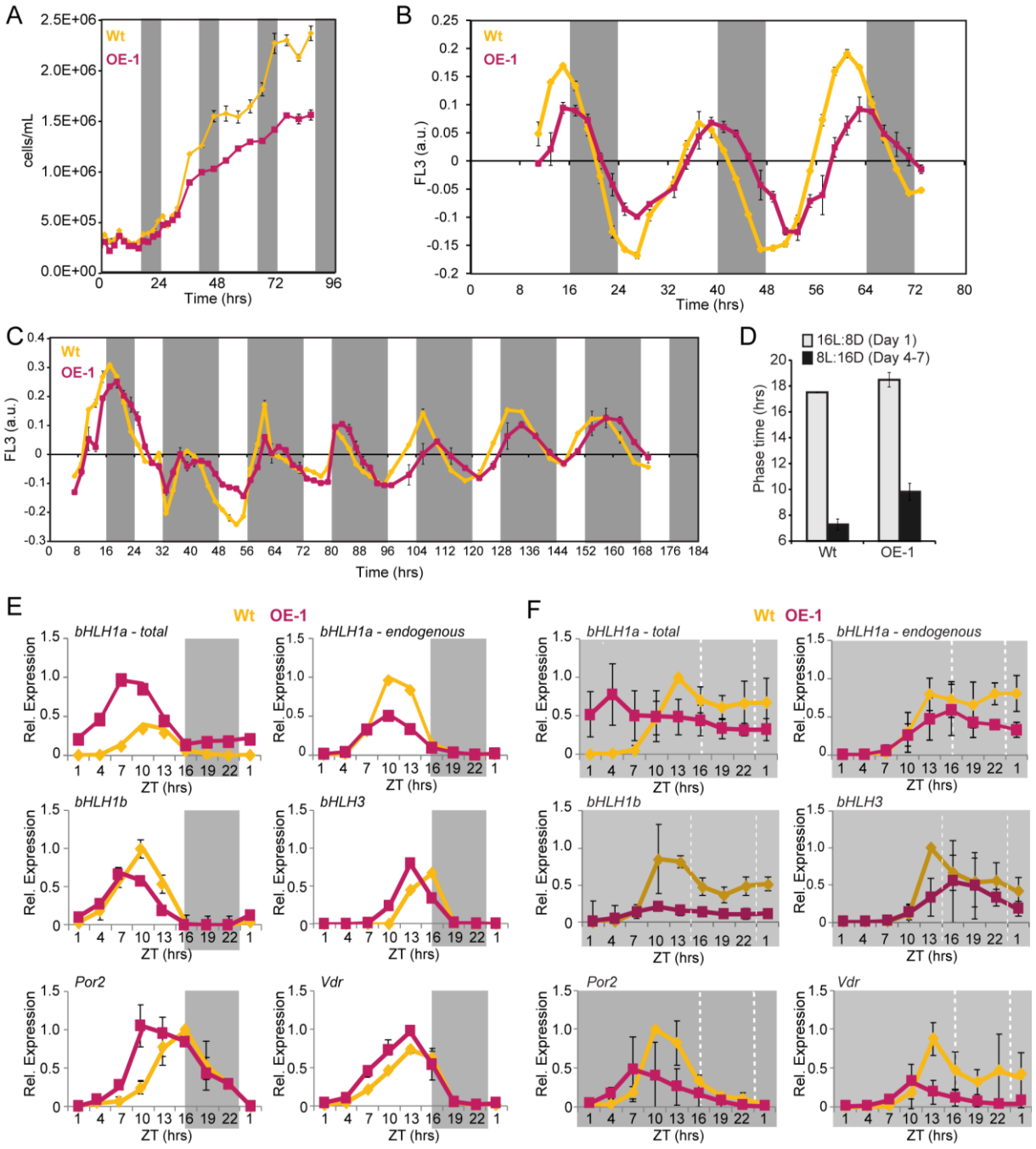

Fig. 3. *PtbHLH1a* over-expression alters rhythmic diel cell growth and gene expression. A) Diel cell growth measurements of Wt (yellow) and OE-1 cell line (red) during 4 days under 16L:8D photocycles. Values represent the mean cell count of three independent experiments ± s.e.m. White and grey regions represent illuminated and dark periods respectively. B) Diurnal oscillation of chlorophyll fluorescence in Wt compared to the OE-1 cell line. Cells were entrained under 16L:8D cycles and the FL-3 parameter was monitored for

three days. FL-3 values represent the mean of three independent experiments ± s.e.m. and were normalized and aligned by the mean FL-3 value of each replicate. C) Phase re-entrainment analysis of fluorescence rhythms after photoperiod change in Wt compared to the OE-1 cell line. The FL-3 parameter was monitored in cultures grown in 16L:8D photocycles for one day and then transferred to 8L:16D for 6 days. D) Bar plot representation of the initial fluorescence phases in Wt and OE-1 cultures entrained under 16L:8D photocycles and after three days of 8L:16D regime change. Values represent the average of 2 biological replicates ± s.e.m. (black bars). E) qRT-PCR diurnal expression analysis of *PtbHLH1a*, *PtbHLH1b*, *PtbHLH3*, *PtPor2* and *PtVdr* transcripts in 16L:8D entrained Wt and OE-1 cultures. *bHLH1a-total* represents the sum expression of the *bHLH1a* endogenous and the transgene transcripts; *PtbHLH1a-endogenous* refers to the transcripts levels of the endogenous gene only. Grey rectangles represent dark periods. Expression values represent the average of 2 biological replicates ± s.e.m. (black bars) and were normalized using *PtRPS* and *PtTBP* as reference genes. For each gene, expression values are represented as relative to its maximum expression corresponding in the graph to '1'. F) nCounter expression analysis of *PtbHLH1a*, *PtbHLH1b*, *PtbHLH3*, *PtPor2* and *PtVdr* transcripts during 24 hours of dark free-running period in the Wt and OE-1 lines. Cells were previously entrained in 16L:8D cycles. Grey rectangles represent dark periods. Expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of 3 biological replicates ± s.e.m (black bars). For each gene, expression values are represented as relative to its maximum expression corresponding in the graph to '1'.

Figure 4 - Annunziata et al.

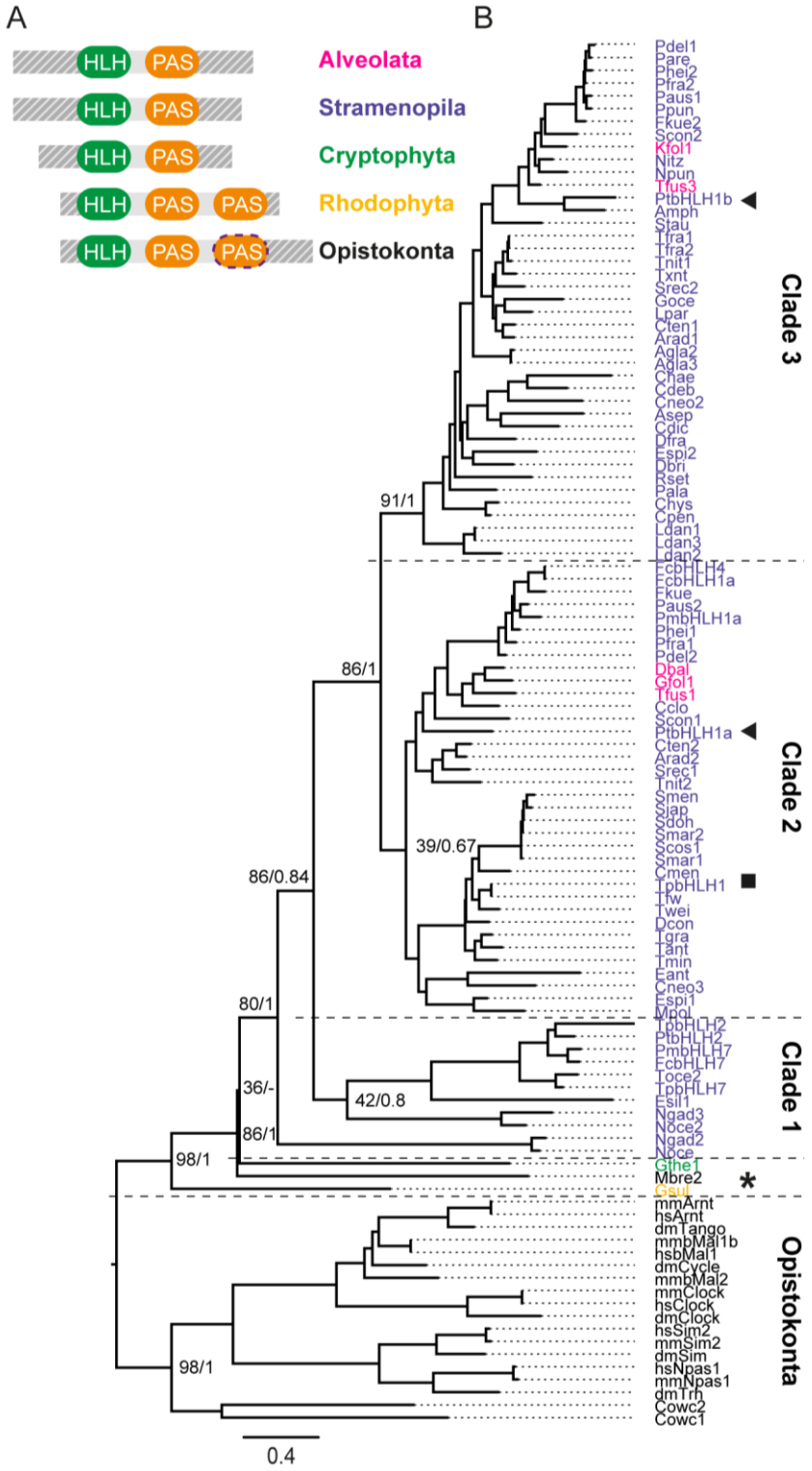

Fig. 4. bHLH-PAS protein family structure and phylogeny. A) Schematic representation of the domain architecture of bHLH-PAS proteins across Eukaryotes. Segmented line indicates possible absence of the second PAS domain in some Opistokonta species. The grey

patterns before the bHLH domain and after the PAS domain schematize the variations in length of N-ter and C-ter regions in bHLH-PAS proteins of different organisms. B) Maximum Likelihood (ML) phylogenetic tree of the bHLH-PAS family. The Opistokonta clade is used as outgroup. Numbers refer to bootstrap values of the basal nodes using different methods: ML (RAxML, 1000 bootstraps) and Bayesian Inference (MrBayes, 2.5M generations, 25% burn-in). The asterisk highlights the position of *Monosiga brevicollis* bHLH-PAS; the arrows indicate the position of *Pt*bHLH1a and *Pt*bHLH1b; the square indicates the position of *Tp*bHLH1. The colour code indicates the corresponding lineage for each bHLH-PAS protein, as emphasized in Fig 4-A.

SUPPLEMENTAL INFORMATION

Figure S1 – Annunziata et al.

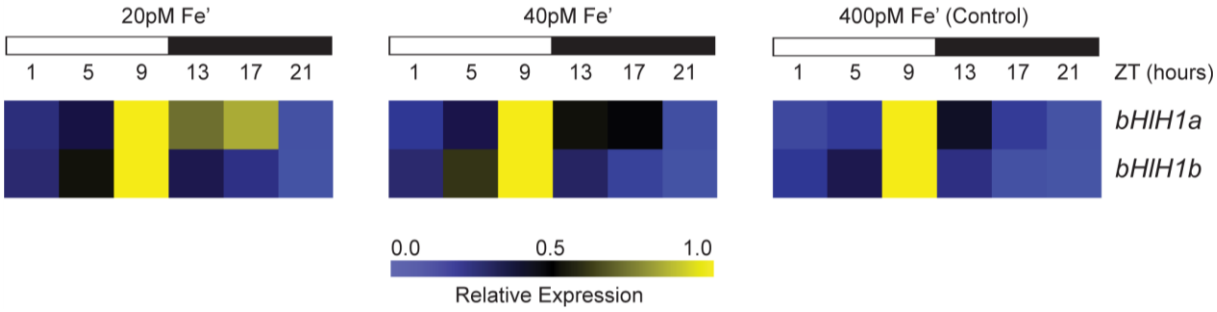

Fig. S1. Diel expression patterns of *PtbHLH-PAS* genes under Fe-depletion conditions in 12L:12D photoperiods. Diel expression patterns of *PtbHLH1a* and *bHLH1b* in normal (400 pM Fe') and iron depletion conditions (40 and 20 pM Fe') were obtained using transcriptome data extracted from (36). Light and dark periods are schematized by white and black rectangles. Expression values are represented as relative to the maximum expression for each gene, where '1' represents the highest expression value of the time series.

Figure S2 – Annunziata et al.

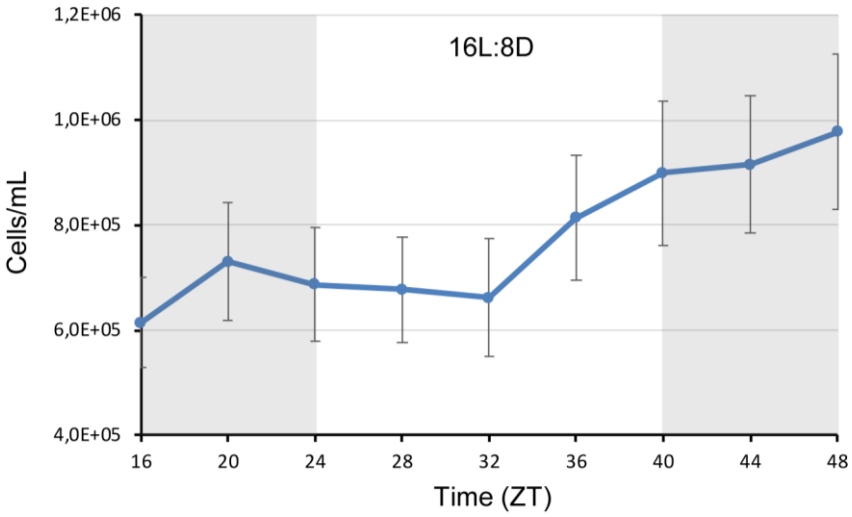

Fig. S2. Diurnal cell growth dynamics in *P. tricornutum*. Diel cell number measurements in wild type cultures grown under 16L:8D photoperiods. Values represent the mean counts of three independent biological replicates ± s.e.m. (black bars).

Figure S3 – Annunziata et al.

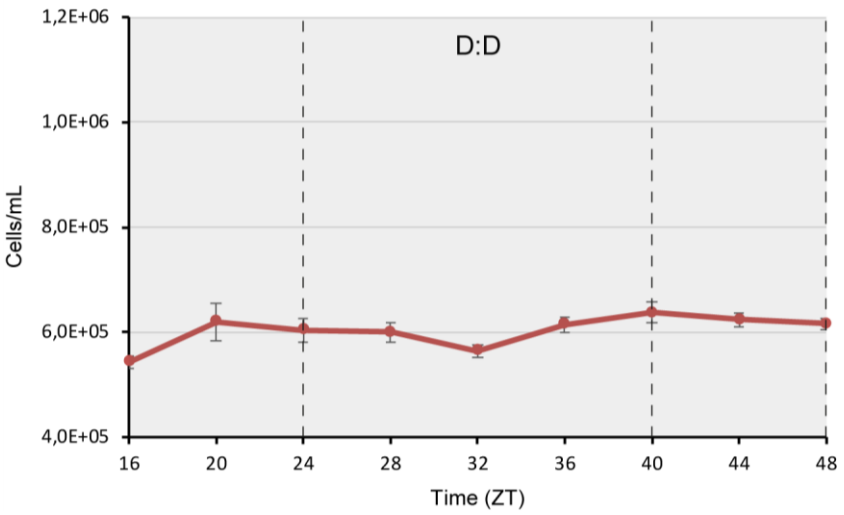

Fig. S3. Cell division is arrested in continuous dark conditions in *P. tricornutum*. Diurnal cell growth measurements of the wild type cultures entrained in 16L:8D and then transferred to continuous darkness condition (D:D) . Values represent the mean counts of three independent experiments ± s.e.m. (black bars).

Figure S4 – Annunziata et al.

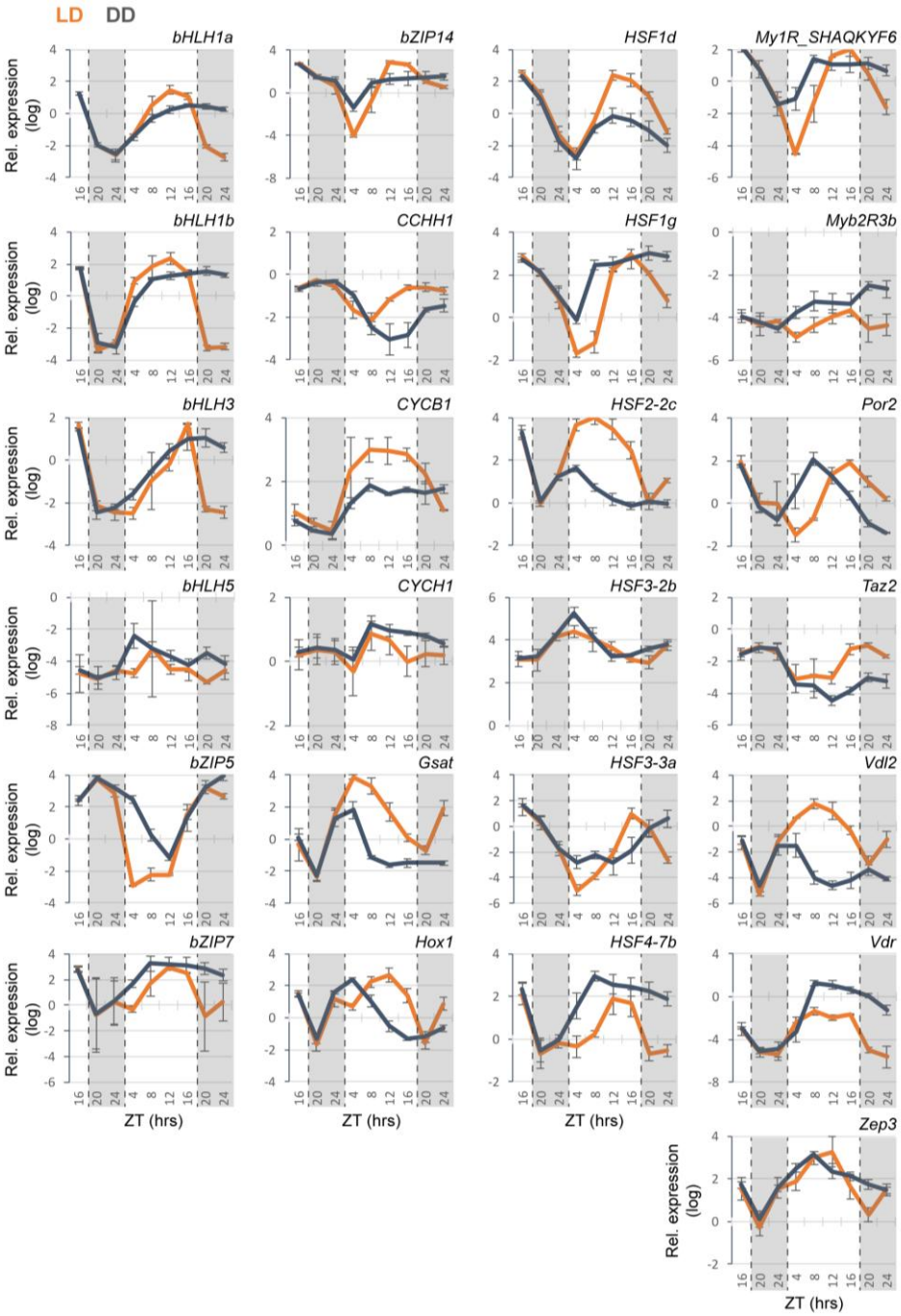

Fig. S4. nCounter expression analysis in Wt cells of genes maintaining rhythmic expression in D:D conditions compared to 16L:8D condition. Data represent the average expression of biological triplicates ±SD and are normalized using the *PtRPS*, *PtTBP* and *PtACTIN12* reference genes. For each gene, expression values are represented as relative to its maximum expression corresponding in the graph to '1'.. L:D, 16L:8D cycle; D:D, 16L:8D adapted cells shifted to D:D for 24h.

Figure S5 – Annunziata et al.

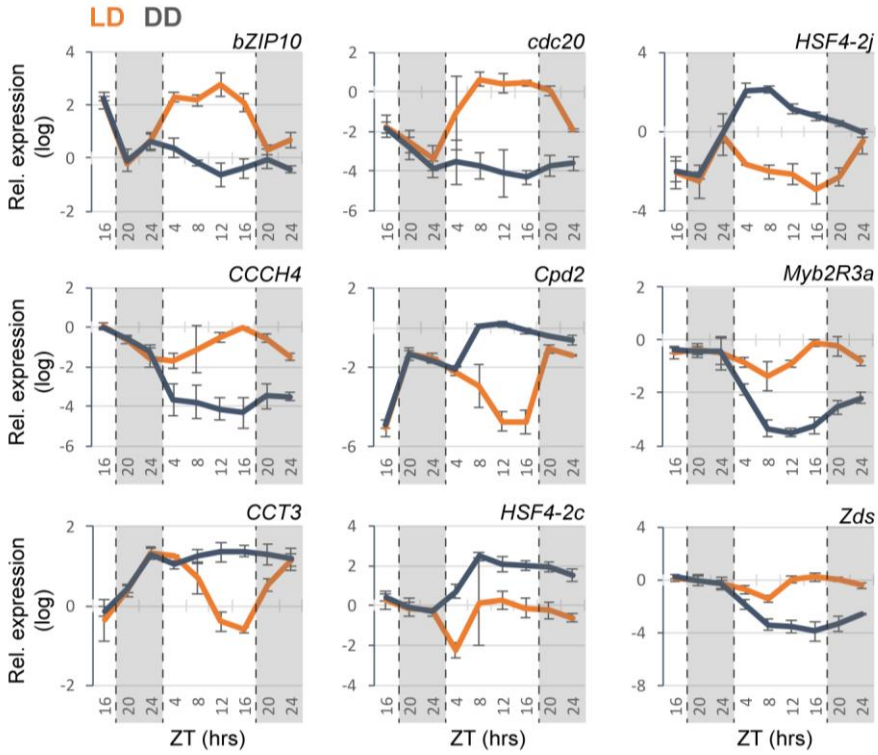

Fig. S5. nCounter expression analysis of selected genes with perturbed rhythmic expression in Wt cells in D:D conditions compared to 16L:8D condition. Expression values represent the average of three biological triplicates ±SD and are normalized using the *PtRPS*, *PtTBP* and *PtACTIN12* reference genes. For each gene, expression values are represented as relative to its maximum expression corresponding in the graph to '1'. L:D, 16L:8D cycle; D:D, 16L:8D adapted cells shifted to D:D for 24h.

Figure S6 – Annunziata et al.

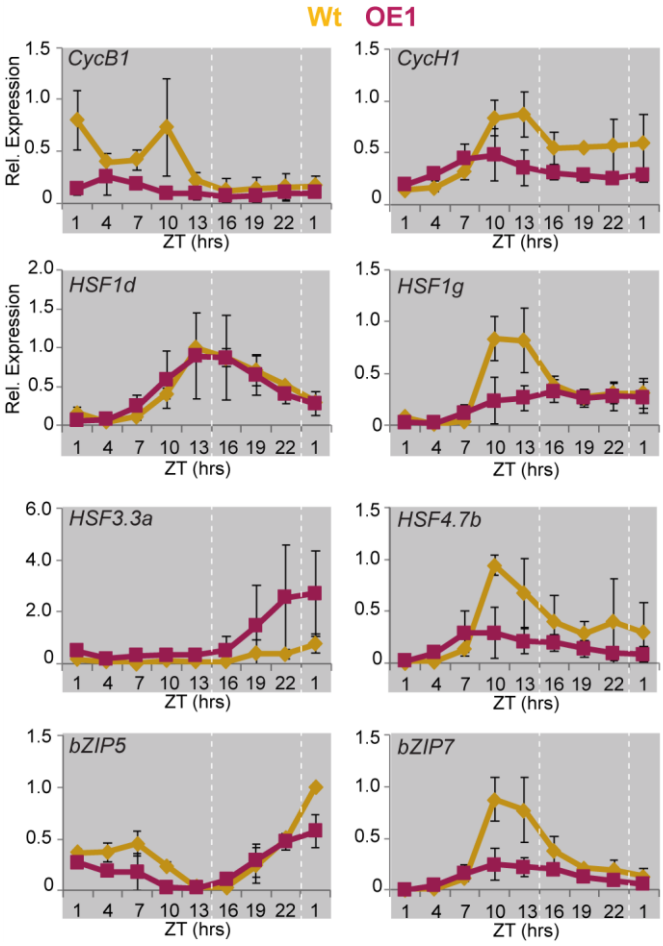

Fig. S6. nCounter analysis of selected rhythmic gene expression profiles in continuous darkness in Wt and *Pt*bHLH1a OE-1 cells. Cells were entrained in 16L:8D cycles, then transferred to D:D and collected every 3 hours for 24h. Expression values represent the average of three biological replicates (n=3) ± s.e.m (black bars) and have been normalized using *PtRPS* and *PtTBP*.

Table S1: Accession codes and sequence probes used in the nCounter analysis.

Category	Gene Name	NCBI Accession	Phatr3 Accession	nCounter Probe Target Sequence
Reference Gene	Actin12	XM_002182185.1	Phatr3_J29136	ATGTAAGCCTGCAGCCACTGAGGACTTGTGCTCGTAACCCTGATTTCGATATTCCAAAATGCGGCTCACTAAAAAGACATCGTAGTCCAGTGCCAGTCCC
Photoreceptor	Aureo1b	#N/A	Phatr3_J49458	TTATTGCCGAGTGCTGTACCTCGAGTCCGTTGTTTGAAGAAATGGACGGCATAGATCAAACGAAAGGGGCCAATCTGGAACGTGCCGATTTTTCATTGAT
Photoreceptor	Aureo2	XM_002183279.1	Phatr3_J15468	AGACGTTACTGTCACTCTCATGAACTACACTGCCGATGGTACACCATTTTGGAACAAGCTCTTTATTGCCGCATTGCGTGACGCGCAAAATAACATTGTC
TF	bHLH1a_PAS	XM_002179032.1	Phatr3_J44962	TACTTTGGTCAATATCAAAGTCAGTTTGGTACGAACAGCTCAGCATAGCCCTCGGTTTTTCAATGTGGCGTTGGTGCCATCAGACGATGCAGCGAAGCTG
TF	bHLH1b_PAS	XM_002178812.1	Phatr3_J44963	AGGAATTGACTACCGTGCTGTTTTCAATCATTGTCCATATGCCATGGGGGTGGCTTCATTGGACGGAAGAATACTGGTCTGCAACAGCTCTTTCGAATCT
TF	bHLH3	XM_002176560.1	Phatr3_J42586	ATCACTAGCGACACATTCAGATTCATCTGCGGGAGAAGGACGGTCTATTCCGTCAGTAGTTGGTACAAGTGTTAGCGGTTCAAGTCGCTCTGGGAACTGC
TF	bHLH5	XM_002177858.1	Phatr3_J43365	ACCCTCTTGCAGCGCAATCATTCGCTCCCGCATACGACTATGGGTCGATTTCTCAATGAACGCCATTCCTTTTCCTCAGACCTTTCGCTCTCGCTGGCGT
Metabolism	BUB1/MAD3	XM_002178639.1	Phatr3_J10954	CGTGTGCTGCATGTATGCGGACAAAACCGATCGTCCACTGGAAGTTTTCCAGCATCTACATCAGCAAAGGATCGGGAGCGATATTGCCGTGTTCTGGATG
TF	bZIP10	XM_002177776.1	Phatr3_J43744	AAGCTGCTCGGGAATCGCGTCGGCGGAAGAAGGTTATGATTGAAGAACTTCAACGCAGCGTGATTTTCTTCTCGCGTGCCAACGGAACCCTCAAACAACA
TF	bZIP13	XM_002181635.1	Phatr3_J47278	CGCGGCGGCTTCGCGTGCTAAGATTCGATGCCGCATCACCGAACTTGAAACCGAGGTTAGTGGTTGGAAGGATAAATACACACAAGCAATGGAACGCTTG
TF	bZIP14	XM_002179477.1	Phatr3_EG02108	ATGCAGAGCTTGGAGGGTCGAGCTACGGATCTCAAGCATGAGCAAATTCGTTTAAAACAGATCATTAACGAAAAGAACACGGCGAACATTCTGGTGGGGC
TF	bZIP16	XM_002181462.1	Phatr3_J47279	CTTTTGGAAGCAGTCACAAATCCGCAAACATGAACAGTTCCAACGCCCAGGAAGTCGCTCGTCAACTGAACGAAGATCCTGCGCTCGCTCGGAAGCGTCG
TF	bZIP18	XM_002176556.1	Phatr3_J42577	ACTCTTCTATAGCGGAATCCGAGGCCGATTTCAAAACTATAGCTCAGGCTGCTGTTTCGAATCTGATCATGTCGGCTGGAACTACCAAGGTTGAATCTGG
TF	bZIP19	XM_002183101.1	Phatr3_J48701	CCTGCACCTTCTCAAATCAGCTGTGCAACGAGTGCAACAAGAAGAAATATTGGCTGCCGAGGCATTACCTACACCTTTTATAGGCGTTTTGCGATTTGGC
TF	bZIP24	XM_002184633.1	Phatr3_J49887	TTGTCACGGGGCCAACAGACTATCTGGAAAGATCACTTCTTGCCCAAAGAGAAGCGGAAAGAACTCGTTCGGCTTTGTCAAGCATCGCGCTTCGGGGGAA
TF	bZIP25	XM_002185584.1	Phatr3_J46647	TCGTCGTTCACGATGGCATCTGTACCAACTCACCTAGTAGGAAATACTGTCATCACTGATACCAGTTATCAGCCAGTTTCGTCGCTATTACCTGCAGCGG
TF	bZIP26	XM_002182641.1	Phatr3_EG02494	GATTCTTTGACGGACTCAAGGAGCGACAACTTGTCATCGACATCTCAGGAGCTTTCGGTGCGACAGGAAAAGGTCGAAGCCGCTTTGAAATCGAAACCTC
TF	bZIP5_PAS	XM_002179145.1	Phatr3_J45142	CCCTTGTTTCCGAACAGCGCATTCGATCCGCAAAACAACGCTAGCAACAACGCACAATCACAGTCGCAAATGATGATGCAGAACGTGGCTATTATGAGTG
TF	bZIP6_PAS	XM_002184884.1	Phatr3_J50039	CGCGCAATCTAATACACAATATCAGCTAGACAGCCCACAACATCCCGAAAATACGATGTCGCAGTGTAGACCGCAGCAGGGTTCAAATAATTCGGCGAAC
TF	bZIP7_PAS	XM_002183297.1	Phatr3_J48800	CGTGAGAGACCTTGCCATGCGTAATCAAGGATTAGAGGCATTAGCGGCGATTGCGACAGCAAGCTCGTCTTCAGGGCACACGTTTAGTACCAACACGAAT
Metabolism	CaThioredoxin	XM_002184396.1	Phatr3_J49634	CTGGAGAGCCCTACCGTGCTAGTGTCCGTGTATTGGAAAAATGTCGTTGTTTTGCGTTCAGCGTAGAAGACATTCCTGACAGCTCCATTCTGAGTCTGAA
TF	CCCH4	XM_002181104.1	Phatr3_J13664	CATGTGCCGCCTTGGAGCGTCGCACATTAACATGTCTACTGGCGAAAACATTCGAAAGGACATTGATGGTCCGAAGCCCAAGCCAATCTTAAATACTCTA
TF	CCCH8	XM_002186091.1	Phatr3_J44042	TCCGAAAAATATGTGCGAAGAGTCCCACTTGTCGCGTATGTTCTTGGCTCATACGAGTATTGCGCCGACGGAAGTGGAAGAGATTGAGTTTGCGGGTGAA
TF	CCHH1_CCCH20	XM_002179821.1	Phatr3_EG02317	TAACTACAACAACGTTCCTCAGCTAAACGGTTTGAACCATCTGCCTTCTAAGAACCTATCCGACTTACATTCCCACTATCAAGATCATAGGAGTCCAAAC
TF	CCHH14	#N/A	Phatr3_J34600	GCCCAGTTTCTGCGGAGGCATCCCGTTCAATCCGCCGATCTGAGTTACGAAACGCACTACATGGCGTACCAAAAAGTTTACACTGATGCCCGAGTTGTTC
TF	CCT3	XM_002185983.1	Phatr3_J43850	AAACGCTGCCGCCGCGTTTGGAATAAGAAGATTCGCTACGGTTGCCGTAAAAATTTAGCAGACCGCCGGTTGCGCGTGAAGGGGCGATTCGTGAAACGTT
TF	CCT4	XM_002178133.1	Phatr3_J44285	TATGAATGCAGTCGACCGGGAATTGAGTTTCCATCAAACAACAGAGCTAAGCTGGTCGAGCACCAGCACGGCCTTTCGCCCATTCCTTTTCACGATAGTG
Metabolism	cdc20	XM_002180546.1	Phatr3_J12783	TTACGACGGAATTGCTTTGAAGAAAATCAGAACACTTCATGGGCACACAGGCCGAATATCGTCGTTGGGATGGAACCAACACTGGTTGAGTTCAGGCGCA
Photoreceptor	Cpd2	XM_002179343.1	Phatr3_J54342	TCTGGTGTGCGGCAATACGGCGGACAGTGTTTCGGAGCTCTGTAAAATTGCTTCTGAGATTGGCGCGAGTGCAGTCTATTGGAATCGCGAGATGACACCT
Photoreceptor	Cpd3	XM_002180035.1	Phatr3_J51952	CTCTGATTTTTCTCCGATACGCGAATACCGTCAATGGATGGAACTCCAGGCCGTACCTATCTTGGAGGAAGCGAAGGTCCCGTTTTATCAGGTTGATGCC
Photoreceptor	Cpf1	XM_002180059.1	Phatr3_J27429	GGATGCGTGCCAAAAATCCGAATCACTCCTTCCGATTTACGTAGTCGACCCTGAATTTCCCTTCGCGCAAACTGCTGGGTGCCGCGCCGGTACAATTCGT
Photoreceptor	Cpf4	XM_002184521.1	Phatr3_J55091	CTCGATCCCTTGTTGCGGCGCAAACGAGACTGCATTTCCGTAATGGGTCTACCTAATGATTTTGTAGACTCCATTGTCGAGGCAGCGTTTGAAGCAGCTG
Photoreceptor	CryP-like	XM_002178853.1	Phatr3_J34592	CACTTTATGGTGGCGAAACAGCCGGACTCGCACGATTGTTTGAAGGATTACTTCGAAACGCGCAACGGGATGCTGGGGCCGAACTATTCGACCAAATTCA
TF	CSF2	XM_002185522.1	Phatr3_J41601	GAACAGAACGAAAAGATTCGGGACGCTTATGCATTCCAACAGACGAAAATCGACGAGGCTCATCAACTTATTGTCAATTTGGGCATGGACGTTGCAGACT
Metabolism	CYCB1	XM_002180361.1	Phatr3_J46095	TTAGAGAAAAGTCAATCGCAAATCCTGAGCTCCGAGCGGTGAATAAAAAGTACAGCGGTCATCGATACGGCGGAGTTGCTTCGACCGTTCTGGTATTTGA
Metabolism	CYCH1	XM_002185709.1	Phatr3_J36892	TTTGCGATGGCTCGCGTTTGACGTTTCGGTTTCCCATCCGCACCGAGGAGTTGCCGCAATCGTCAACCATCTCGTGGATTGCTCACCTCAATGGTTTGGC
Photoreceptor	Dph	XM_002179026.1	Phatr3_J54330	CAACAATTCCATTACAACCAAAGAGCTGACGGAATGTGATCGTGAGCCTGTGCACTTGATCGCAAACGTACAAGGGGGTACCGGCCATTTGTTGTTCATT
Metabolism	dsCYC2	XM_002179247.1	Phatr3_J34956	GATATTGCCGTTACCGAACGCGAAATAATGACTGCTCTTTCCTGGCACTTGCATCCTCCAACCGCTATTGGCTTTTCACGAATGTACTGGTCTCTGCTGG
TF	E2F_Dp2b	XM_002181452.1	Phatr3_EG02016	TGATTTGAATAGAGCTGTGCAGGAAATGAGGGTACAAAAGCGCCGGATTTATGACATTACCAACGTTTTGGAAGGTATTGGGCTGATTACCAAGGATAGC
Metabolism	E2F1	XM_002176842.1	Phatr3_J43065	GTGTGGGTCTCATCGAAAAACGATCCAAAAACACAGTCGCTTGGAAAGGGAGTGAGCTTCTTCTTGGCTCGTCCTTTTCAAGTGCCGCCAAGCAGAGAAT
Metabolism	FtsZ	XM_002185088.1	Phatr3_J42361	GAAAAAGGAGAATGAAGCAAGCGATTGCCGCGATTGAGAGGTTGCGAGAGAATGTGGATACGGTCATCGTAGTGTCCAACGACAGACTGCTCGAGATTAT
Metabolism	GapC1	XM_002182255.1	Phatr3_J22122	TGAAGGGATTCCTCGGATACTCCGACGAACCGTTGGTCTCCACCGACTTTGAAGGTGACTTGCGCTCCTCCATCTTTGATGCCGATGCCGGTATCATGCT

Metabolism	Gsat	XM_002180931.1	Phatr3_J36347	ATTCGTTTGGTCCGCGCACACACGGGTCGGGAAAAGGTCATCAAGTTCGAAGGATGCTACCACGGACACGCCGATTCATTTTTGGTCCAAGCCGGATCCG
TF	Hox1	XM_002184935.1	Phatr3_EG02213	AACAAACATCGGTTCACATACTGCGTCATACCGGACAAGTCCCGACAGTCGAAGATGTTACGATCCTTACCAATGTTCCGGCGGAACGCATTCTCCGAAC
TF	HSF1.3b	XM_002179894.1	Phatr3_J35419	CTCCAGGGAACAACTCGATACCAGAGTTCCTCTATCAGCTAACAAAGATGCTGACGGATAACAACCGAGATATTATTGAATGGACAAATGGCAAGATTGA
TF	HSF1a	XM_002177362.1	Phatr3_J43051	CCACCAACGTCGGATAGCTCCGACTGGGGTTCTCCAATCCAATCATCCAAGCCAGCTGTTGATCTCTTAACGAGGAGTATGCCGCAATGGACAAATATAC
TF	HSF1b	XM_002181395.1	Phatr3_J47181	TTGTAGTGGACACTTCGATGCTCGAAGAGGTCATCAACGAAACGCAGCCACTCGTTGGAGACAAAGGTAGCTTCTCCCTCTTGACAAATAGAGCACCAGT
TF	HSF1d	XM_002178673.1	Phatr3_J44750	GTCACCATTCCAGCATCGGATGCGGCTCCTTACGCCATTTCTCCCGAAACCTCACCGAGAGACTTGACTCTGAGACACAACGAAATTCCCAACCTTGATG
TF	HSF1g	XM_002177016.1	Phatr3_J42514	ATAGAGAATCTGCAAGCACTAGGGACCATGACGACTACTTCATCGAAGCGTAGCATTACAGAGAGCGAAGATAGTTCCCTGTCAGACTCAGCATCAGGAG
TF	HSF2.1a	XM_002181681.1	Phatr3_J47360	AGCCCTTGCCGCTGGAATTTTGGGCTCGTTGTCGCCTCAAGTCAACCAACCACAGCCGTGGATGATCCAACACGCCCTGTCAATCGCTCAAGCGTTAGTT
TF	HSF2.2c	XM_002185405.1	Phatr3_J50481	GGACTGGAAACTGTAAGCGCGACATATTTCCGCAGCGCCTTCTAGCGATTTTGAGCGAATCATCCCTATCTGACATTATCACTTGGCTTCCACATGGACG
TF	HSF3.2b	XM_002186124.1	Phatr3_J44099	TTGACTCCTCGCGACCAGCAAGCACTTAGCGCGTTTCAACAGTCCCTCGGTGCATCCGAGAGTCAGTTCAAGTCTATGAGCTTTTCCACCACAACTCCTC
TF	HSF3.2e	XM_002179404.1	Phatr3_J45206	TTGGAACGGGAACGCTTTCCCTCAGATTAGCAATCCGCAACCTTCGCCTACTCCTCCAAGTTCGACGATGGTAAATACTTCGGACTTAGCAAGACTACTT
TF	HSF3.3a	XM_002179738.1	Phatr3_J45393	TCTGGGCAAATCTTTCGTATGGAAGGTTTGGCGACGCCGAGTAGTCAGGGAGTTGGCGTAACCACAGGCAGTTTGCCGCTTCGAGACCAAGACGCTGTTA
TF	HSF3.3f	XM_002179735.1	Phatr3_J45389	CATGCCAATTGAACACCCAATTCAGCCATACATGGAAGAGCTTAAGAGGGTCCATCTTGCAGCGGCGGCGGGAAGTCTCGAAAATGCGCTTCGAATTTTG
TF	HSF4.2c	XM_002182762.1	Phatr3_J48361	CGCACGTATATCACGATTTCGCACAAGTGCGGCCCGACGCGGATCCCATGATGATCGTCCGGAAAAAGACTGGAGGAGTCACTCAGCCATTTCCCGAAAA
TF	HSF4.2j	XM_002177548.1	Phatr3_J43363	GTCGGCGGACGGGCCTTTATCATTCACAAGCCTGACAAGTTCTTTAAGGATATAGTGCCGCTATATTTTCGCCAGTCGCGGCTGAGTTCATTCAAGCGGC
TF	HSF4.3a	XM_002184368.1	Phatr3_J49594	CGCGAACACTCGCGTTCCAATGTCGCGGAAGAACCTTCCAAGCGAACCTTTGTTCAGCATATTTACCATGACCATATGAACGACCCTGAGCAGGTGGATG
TF	HSF4.3b	XM_002183012.1	Phatr3_J48558	CTGTCTTGAGCCTATCGCAGCAGTGGCGCCGAATTTCTACGCTATGCCAACTACTGAAGGTGGAGACAAAGACCAATTATTCTCGATCCAAAATACGTGG
TF	HSF4.3c	XM_002183081.1	Phatr3_J48667	AAACTGTCAGTCATGCTTGACCATGTAAAAGCCGCAGGATTGGACGACGTTATTTCTTGGGCCAGTCATGGCCGCTGTTTTAGCATTCACAACCCAGACC
TF	HSF4.4b	XM_002184371.1	Phatr3_J55070	CCCATTCCGGTAAACCCGGTTTCCAACTTGGGCGTATCCAGAGCGCCTAGTTTTGGTAAGCTCGAACCAGGCCAATTCGCTGCTTATGGCGGTTGTATGA
TF	HSF4.6a	XM_002184347.1	Phatr3_J49557	CTCAAACACTGGCGATTGCCGTATCGCCTCGTTCCGCCATCCAGTGCACGCCAACATCGTGAATCCTTCGCACTTTCGCTCTGTCTCCAATAACGGGCAC
TF	HSF4.7a	XM_002185271.1	Phatr3_EG00092	CGGACATCCCGGGGCGGATTTGAGATCGACGCGATAGTGTCATTCTTGCGTTCCAAGAGTATTTCGCAACTCTACATTATCGGTGGTGACGGTACTCACC
TF	HSF4.7b	XM_002184277.1	Phatr3_J49596	CTCGAGGTGGTGTGGGAACTCCTTTTCCACTCAAGCTGCACGAAATGCTTCAAAATGTCGTACAAGACGGCTACGCTCACATTGTATCGTGGCAGCCTCA
Metabolism	Lhcx1	XM_002179724.1	Phatr3_J27278	AATCCTTGAGAATCTTCAGGGTTAAAGAGTGCATCCATCCAAAACTGGAGGTATGGCTGGTCACATGCACAGAGTTAACAACCAATCCCTTTTCAAGCGA
TF	Myb1R_SHAQKYF4	XM_002178425.1	Phatr3_J44331	CGTATCTGCCGAATCGGTCTCTGGTGCAAATCAAGTCCCACGCCCAAAAAGTGCTCAAGCGCATCGATCAAGGCGAACACGTCTTCCGACGTATCGAGGA
TF	Myb1R_SHAQKYF5	XM_002181623.1	Phatr3_J47256	ACGCTTCACCTGGCCCGAAGACTTACACCGGGATTTTGTATCGGCAATTTTTGATGTCGGCTTGAAACAGTCGTCACCGTCCACCATTTTGGAAACCATG
TF	Myb1R_SHAQKYF6	XM_002180835.1	Phatr3_J46535	TAACGACGAAGCGCAAATCCTTACCGACAAACTTGCCGTCCCGCAAAAAGGGTAAGACACGAAAGTCTGCCGCCTTGGTACAGCGTAATACTTCAGTTTC
TF	Myb1R10	XM_002181361.1	Phatr3_J37257	CATATCCCTCTTCGGGGCCCTCACTTCCCTGTACAAGTGAATCTTTGGGGGTGAAAGCGAATCGTAGAAAGAAACGTGCTAAGGACGAGACCTCGGTTCT
TF	Myb1R4	XM_002181410.1	Phatr3_J47205	AAAGTTCGTTGCAGGTTTAGACGGAACGTCTACCGCATTGTCAATGGGAACAATCCGGCGAGTAGGATTCACCCATATGCGATTGAACTTCTGGAAGCCG
TF	Myb2R3a	XM_002181911.1	Phatr3_J47726	TCCATCTTTCCTTTGAGCGACTGTGACGACCTGTGGAGGGAGAGTGATTTTTGTCACTGTGAAGTTGTCGGTACTCCGGCACTCGGCAGGTAAAATACTC
TF	Myb2R3b	XM_002182076.1	Phatr3_EG02275	GCGCACAACGGTTTAAAGTGATTCGGGATTCATTCACAACAGTCAATGGCTAATTGGCCTCCATCGTTCACTGGTTGACATTTGCGATCGGCCCAACACT
TF	Myb3R1	XM_002182738.1	Phatr3_J15016	GTATGATCCTGGAGTGTCATTTGACGTTGGGAAACCGCTGGGCCGAAATTGCCAAGCGTCTACCGGGAAGAACTGATAACGCTATTAAAAATCACTGGAA
TF	Myb3R2	XM_002180449.1	Phatr3_J6839	CGAATTCTCATAGGAGCACAGGACACAATGGGAAATCGATGGGCTGAGATTGCCAAGCGGCTGCCAGGACGTACGGACAATGCGATAAAGAATCGCTGGA
TF	Myb3R4	XM_002180741.1	Phatr3_J36337	GACAGATGCCGAAGACGCAATTGTAATGGATGCGGTCAATTCCAGCTCGGAGCAACCATTTACTCGTTGGTCAGACTTGGCGCAACGTCTACCCGGACGT
TF	Myb3R5	XM_002184658.1	Phatr3_J7959	GACGTCGTTATCTGAATTGCTTGGACCCAACAATCCGAAAAGACGAATGGACGGAGCAAGAGAAAGAAGCTTTGTTTGATGCACAGCGGCGCATGGGAAA
TF	Myb5R	XM_002185245.1	Phatr3_J50365	TCAACACCACATAAACGCTCGTTTCAGGCTTTGGATGTGTTAGTGAACGCGGTTGGTTTTCATCGAGAACCAAATGACCAAATGAAGCAGACGAAGATGT
Metabolism	PCNA	XM_002182225.1	Phatr3_J29196	AATGCAGGAACCGGTCGAACTCACCTTTGCCCTGCGCTACCTCAACTTTTTTACCAAGGCGACTCCGTTGAGTGGACACGTCATTATCTCCATGGCACCG
Metabolism	Pds1	XM_002180135.1	Phatr3_J35509	AGCATGGCCAAGGCACTCGATTTCATTGATCCCGACAAATTGAGTATGACGGTCGTTTTGACCGCCATGAATCGCTTTTTGAACGAAGACAACGGCCTCC
Metabolism	petA	GI:118411009	#N/A	TCTCTGGATCAGGAGCCATAACAGGGAAAATTAACTCTTGATGAGTCTTTCCTGCAATTGGTCCAACTACTAAGATATTATCAAATTCACTACTATACGG
Metabolism	Pgr5	XM_002178672.1	Phatr3_J44748	GTAACGTTCGCGCTTCTGATCGCTACCGCTGCCGCGTTTCGCGTGGCGGGCGGAGCCACCGGGGGTTCTTCCTTTGTGCTTTCCCGACGAACCAGCATCG
Metabolism	PgrL	XM_002177043.1	Phatr3_J42543	CCACAACAAGTAGCGTTTCCAATTTAGACAAGACTACGAACAAACGATGCGAGTCAACTCTGTGGTATTTGGCCTTTTGGTGTTCTCGGCCGTCAGCGAT
Metabolism	Por1	XM_002179653.1	Phatr3_J12155	TCGGGTGGTGAGATTTTTGAGAATCAGCAGTCCGACGCAGTACGGGATCTTCCTACCGCCAAGAAAATGTGGAAACTGAGTAGGGAAGCAGTTGGTCTTT
Metabolism	Por2	XM_002180956.1	Phatr3_J13001	AATCACCACTATTCCGTGAAAAGCGCCCGTGGTTCCGTAAATACTTTCCCATCTTTATGAAGTTTATTACGGGAGGATACGTTGGAGAGCACGAAGCCGG
Metabolism	Por3	XM_002177432.1	Phatr3_J43164	CGCGTTCTCAACACGAGGAAAAGCGGTGAGACTCATGATTCGTGATCCGACCAGTAGTCGTATCGAGGCACACCGTAGCGTGGGAGACGTAGACGTTCCC
Metabolism	psaA	GI:118410964	#N/A	GCTAAGAAGAACGCCCAAGTAGTTCCAATACCACCCAATAAATAATGAGCTAAACCAACTGCACGACCCTGAGTAATGCTTAATGCACGAGGTTGGATTG
Metabolism	psaD	GI:118410999	#N/A	TTAAATGTACGTAATTGGGTTCCCAATGCTAAACATTGCTCTTTGCGCGCTAAGTATAATAAATTTTCTCCACTACGCATGATAGCTGCTCCACCAATTG

Metabolism	psbC	HQ912250.1	#N/A	GGCCCAACAGGTCCTGAGGCTTCACAAGCACAAGCATTCACGTTTTTAGTTCGAGATCAACGTTTAGGTGCTAATGTTTCATCAGCACAGGGTCCAACTG
Metabolism	Psy1	XM_002178740.1	Phatr3_EG02349	ACCTGTCGTCGTGGGAATTGCGATTGGAGCGTTTGTGGCAATACGGACAAGTACAGGATGTCTTTGACCTGTGCTTGCTAGATCTTCGAATCCAGTATCC
Reference Gene	Rps	XM_002178225.1	Phatr3_J10847	CTCCGCGCATTTTTGCCCGGTTCCCATTTGACCGGACAGTTGCCCGACGAAGATCTCATTGGAAACAACTTGCAGCTTAAATTCCTCGAAGTCAACCAGG
TF	Sigma70.1a	XM_002182291.1	Phatr3_J14599	GACAATTAGGGGTATCTCGGGATCGGATTCGCCTTGTCGAAACCCGAGCACTGAACAAATTGCGACACCCGCAACAAAATTACAAGCTGCAGTCTTACGT
TF	Sigma70.1b	XM_002182124.1	Phatr3_J3388	TTCTTGGTCATTGGAAGAAACGGCCAAACAACTAGGGATGTCTCGGGATCGGATTCGCCTTGTCGAAGCCCGAGCACTGAATAAATTGCGACGTCCACAA
TF	Sigma70.2	XM_002178682.1	Phatr3_J5537	ACGTGGATTGCGATTCGCCACCTACGCCACTTACTGGGTCACCAATTCCGTACGCTTCTGCTTCCAAACCGCATCCACCGGATGTTTGCGGGTACCGGTA
TF	Sigma70.3	XM_002185461.1	Phatr3_J17029	TATGATGGTTCCTGATCTAAAGAAGCAAATTCGTCGCTCGCTTCGCGCCAAGGCTGCCCTGACGGAATCAAATATTCGACTCGTGGTATCCATTGCGAAA
TF	Sigma70.4	XM_002177232.1	Phatr3_J9312	AACCCGAAAGATCTCGGTCACGAAAGTCACGATTTGCTAGGGGACACAATCTCCGCCAGTAGTGCTCTTTTTGACGAAAGCACACCCGAAAAACGAGTCG
TF	Sigma70.5	XM_002182854.1	Phatr3_J14908	AAGAACTCATACAGGAAGGCAGTTTGGGATTGTTGCGCGCGGCCGAATTGTTTGATCCCGAGCGAGGACTGCGATTCAGCACTTACGCGGTAGTGTGGAT
TF	Sigma70.6	XM_002178042.1	Phatr3_J9855	CTGGAAAAATCAATATGGAAGCTATTCGGCAGACCATTGAAGAAGGACTGGAAGCCAAAAATCAGCTAGTCACTTCGAATCTGCGGATGGTACAGAGTGT
TF	Sigma70.7	XM_002185047.1	Phatr3_J50183	AATGATGACGCTCCTCTCGCATACGGTGGAACTCCGTCGTATCAAGCAAATCGAAGCTGATCTTGCTTTGACGTATCCATCTTCCCGGAAACCCACTACT
TF	Taz2	XM_002178935.1	Phatr3_J44807	CCTAATGTAGAGACTGTAGTGGTAAAAGCGAAGCGCGCAGCCTCGACTATTTGGACTTTACTTCATGCACAGAATTGCGTTTTGGGGGACAGATGCCCAC
Reference Gene	Tbp	XM_002186285.1	Phatr3_J10199	CTGATTTTTTCCAGCGGTAAAATGTGCGTGACCGGAGTCAAGAGCACACACAACGCCAATCTGGCGGCGAAAAAGTTTGCCTACATTGTGGAGCGCGTCG
Metabolism	Vde	XM_002178607.1	Phatr3_J51703	GGCGATGACGGATCATATCCCGTTCCCGCACCCGAAATCGTCGTACCCAAGTTTGACACCAAGTTCTTCGATGGACGACTCTACATCTCAGCCGGACAAA
Metabolism	Vdl1	XM_002180599.1	Phatr3_J36048	AAGGCTCTCTTTAGCAATCCGCGTGGAATCAAAGGTGTCTCCTGTTTGGGACGCTGCAAGGGCGAGCAATCGTGTGCGACGCGGTGTTTTGCCGAATTCG
Metabolism	Vdl2	XM_002180015.1	Phatr3_J45846	ACACACCGCCTTCGGCAATGAATATCCAATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCACGTACGAAACATCGGTAAAGTGTGTACCGGGTT
Metabolism	Vdr	XM_002177477.1	Phatr3_J43240	GTCGCTTGCGGTGCCAACGAAGCCTATGATAAGTTTCCTTCGCAGAACCAACTCTTCTATCCCGCTGCTCGCGGTCGAGATTTGTGGTACGATCCCGTTT
Metabolism	Zds	XM_002184417.1	Phatr3_J30514	CGGCGTCAACGAATCCATGCGGGACACCGCCAACGTACTCGCCCGTTTTAATTCACTCATTCTCGCCCGAGTCTTTGGACACGCCGACGTCACGGAACTC
Metabolism	Zep1	XM_002179586.1	Phatr3_J45485	TTCCCAATGAAGTTTTCTACCACGGTGTCATCGGCACTGTTCCTGATCGCGTCGGTATCGACCACGACTTCCTTCACACCCGTCCAGTCCTTTGGCGTGC
Metabolism	Zep3	XM_002178331.1	Phatr3_J10970	CCAAGGTTTTGACCAAGATGCCCACCATGGATGTTACCGTACTGGAACAAACGTCCGAATTCAAACGGTTCGGTGGACCCATCCAGCTCGCTAGTAACGC

Table S2: Calculated periods, phases, and amplitudes in Wt and OE-1 cells growing in 16L:8D photocycles.

Data	N	Phase	Ph.Std	Amp.	Amp.Std
Wt	3	13.21	0.3	6.80E-01	3.00E-02
OE-1	3	14.01	0.34	5.10E-01	5.00E-02

Table S3: Accession numbers of the proteins utilized in the bHLH-PAS phylogenetic analysis.

SPECIES	ABBREVIATION	ACCESSION	ACCESSION(Alt)
Amphiprora sp., Strain CCMP467	Amph	CAMPEP_0168730192	
Asterionellopsis glacialis, Strain CCMP134	Agla2	CAMPEP_0195290872	
Asterionellopsis glacialis, Strain CCMP1581	Agla3	CAMPEP_0197142484	
Astrosyne radiata, Strain 13vi08-1A	Arad1	CAMPEP_0116831938	
Astrosyne radiata, Strain 13vi08-1A	Arad2	CAMPEP_0116837632	
Attheya septentrionalis, Strain CCMP2084	Asep	CAMPEP_0198303966	
Capsaspora owczarzaki ATCC 30864	Cowc1	XP_004345696	
Capsaspora owczarzaki ATCC 30864	Cowc2	XP_004343694	
Chaetoceros debilis, Strain MM31A-1	Cdeb	CAMPEP_0194099558	
Chaetoceros dichaeta, Strain CCMP1751	Cdic	CAMPEP_0198277646	
Chaetoceros neogracile, Strain CCMP1317	Cneo2	CAMPEP_0195415676	
Chaetoceros neogracile, Strain CCMP1317	Cneo3	CAMPEP_0195453836	
Chaetoceros sp., Strain GSL56	Chae	CAMPEP_0176495412	

Corethron hystrix, Strain 308	Chys	CAMPEP_0113330172	
Corethron pennatum, Strain L29A3	Cpen	CAMPEP_0194280324	
Cyclophora tenuis, Strain ECT3854	Cten1	CAMPEP_0116540656	
Cyclophora tenuis, Strain ECT3854	Cten2	CAMPEP_0116579710	
Cyclotella meneghiniana, Strain CCMP 338	Cmen	CAMPEP_0172279412	
Cylindrotheca closterium	Cclo	CAMPEP_0113624952	
Dactyliosolen fragilissimus	Dfra	CAMPEP_0184871870	
Detonula confervacea, Strain CCMP 353	Dcon	CAMPEP_0172306268	
Ditylum brightwellii, Strain Pop2 (SS10)	Dbri	CAMPEP_0181012646	
Drosophila melanogaster	dmClock	AAC62234	FBpp0099478
Drosophila melanogaster	dmCycle	NP_524168	FBpp0074693
Drosophila melanogaster	dmSim	AAC64519	FBpp0082178
Drosophila melanogaster	dmTango	NP_731308	FBpp0081483
Drosophila melanogaster	dmTrh	AAA96754	FBpp0293065
Durinskia baltica, Strain CSIRO CS-38	Dbal	CAMPEP_0170381008	
Ectocarpus siliculosus	Esil1	Ec-02_004780 PAS domain (681) ;mRNA; f:5082087-5086119	
Eucampia antarctica, Strain CCMP1452	Eant	CAMPEP_0197830682	
Extubocellulus spinifer, Strain CCMP396	Espi1	CAMPEP_0178587872	
Extubocellulus spinifer, Strain CCMP396	Espi2	CAMPEP_0178723458	
Fragilariopsis kerguelensis, Strain L26-C5	Fkue1	CAMPEP_0170867208	
Fragilariopsis kerguelensis, Strain L26-C5	Fkue2	CAMPEP_0170907188	
Fragylariopsis cylindrus	FcbHLH1a	jgi|Fracy1|204663	
Fragylariopsis cylindrus	FcbHLH4	jgi|Fracy1|241747	
Fragylariopsis cylindrus	FcbHLH7	jgi|Fracy1|172104	
Galdieria sulphuraria	Gsul	XP_005709404	
Glenodinium foliaceum, Strain CCAP 1116/3	Gfol1	CAMPEP_0167841300	
Grammatophora oceanica, Strain CCMP 410	Goce	CAMPEP_0194036250	
Guillardia theta CCMP2712	Gthe1	XP_005835918	
Homo sapiens	hsArnt	AAA51777	
Homo sapiens	hsbMal1	NP_1284651	
Homo sapiens	hsClock	AAB83969	
Homo sapiens	hsNpas1	AAH39016	
Homo sapiens	hsSim2	NP_5060	
Kryptoperidinium foliaceum, Strain CCMP 1326	Kfol1	CAMPEP_0176099000	
Leptocylindrus danicus var. danicus, Strain B650	Ldan1	CAMPEP_0116039394	
Leptocylindrus danicus var. danicus, Strain B651	Ldan2	CAMPEP_0116053596	
Leptocylindrus danicus, Strain CCMP1856	Ldan3	CAMPEP_0196809694	
Licmophora paradoxa, Strain CCMP2313	Lpar	CAMPEP_0202478474	
Minutocellus polymorphus, Strain RCC2270	Mpol	CAMPEP_0185804732	
Monosiga brevicollis	Mbre2	jgi|Monbr1|26507|	
Mus musculus	mmArnt	AAA56717	
Mus musculus	mmbMal1b	BAD26600	
Mus musculus	mmbMal2	ABC50103	

Mus musculus	mmClock	AAC53200	
Mus musculus	mmNpas1	NP_32744	
Mus musculus	mmSim2	AAA91202	
Nannochloropsis gaditana CCMP526	Ngad_2	Naga_100016g14	
Nannochloropsis gaditana CCMP526	Ngad_3	Naga_100165g4	
Nannochloropsis oceanica CCMP1779	Noce	NannoCCMP1779_9983-mRNA-1	
Nannochloropsis oceanica CCMP1779	Noce_2	NannoCCMP1779_1600-mRNA-1	
Nitzschia punctata, Strain CCMP561	Npun	CAMPEP_0178827522	
Nitzschia sp.	Nitz	CAMPEP_0113465330	
Phaeodactylum tricornutum	PtbHLH1a	jgi|Phatr2|44962	Phatr3_J44962
Phaeodactylum tricornutum	PtbHLH1b	jgi|Phatr2|44963	Phatr3_J44963
Phaeodactylum tricornutum	PtbHLH2	jgi|Phatr2|54435	Phatr3_J54435
Proboscia alata, Strain PI-D3	Pala	CAMPEP_0194393728	
Pseudo-nitzschia arenysensis, Strain B593	Pare	CAMPEP_0116139930	
Pseudo-nitzschia australis, Strain 10249 10 AB	Paus1	CAMPEP_0168282354	
Pseudo-nitzschia australis, Strain 10249 10 AB	Paus2	CAMPEP_0168310394	
Pseudo-nitzschia delicatissima, Strain B596	Pdel1	CAMPEP_0116091770	
Pseudo-nitzschia delicatissima, Strain B596	Pdel2	CAMPEP_0116094064	
Pseudo-nitzschia fraudulenta, Strain WWA7	Pfra1	CAMPEP_0201229664	
Pseudo-nitzschia fraudulenta, Strain WWA7	Pfra2	CAMPEP_0201232896	
Pseudo-nitzschia heimii, Strain UNC1101	Phei1	CAMPEP_0197189528	
Pseudo-nitzschia heimii, Strain UNC1101	Phei2	CAMPEP_0197199168	
Pseudo-nitzschia multiseries	PmbHLH1a	jgi|Psemu1|26622	
Pseudo-nitzschia multiseries	PmbHLH7	jgi|Psemu1|228145	
Pseudo-nitzschia pungens, Strain cf. pungens	Ppun	CAMPEP_0172413298	
Rhizosolenia setigera, Strain CCMP 1694	Rset	CAMPEP_0178942232	
Skeletonema costatum, Strain 1716	Scos1	CAMPEP_0113408658	
Skeletonema dohrnii, Strain SkelB	Sdoh	CAMPEP_0195221452	
Skeletonema japonicum, Strain CCMP2506	Sjap	CAMPEP_0201725346	
Skeletonema marinoi, Strain FE7	Smar1	CAMPEP_0180846198	
Skeletonema marinoi, Strain UNC1201	Smar2	CAMPEP_0197231700	
Skeletonema menzelii, Strain CCMP793	Smen	CAMPEP_0183679140	
Stauroneis constricta, Strain CCMP1120	Scon1	CAMPEP_0119548954	
Stauroneis constricta, Strain CCMP1120	Scon2	CAMPEP_0119572288	
Staurosira complex sp., Strain CCMP2646	Stau	CAMPEP_0202487246	
Synedropsis recta cf, Strain CCMP1620	Srec1	CAMPEP_0119009808	
Synedropsis recta cf, Strain CCMP1620	Srec2	CAMPEP_0119029558	
Thalassionema frauenfeldii, Strain CCMP 1798	Tfra1	CAMPEP_0178899126	
Thalassionema frauenfeldii, Strain CCMP 1798	Tfra2	CAMPEP_0178923098	
Thalassionema nitzschioides, Strain L26-B	Tnit1	CAMPEP_0194218124	
Thalassionema nitzschioides, Strain L26-B	Tnit2	CAMPEP_0194240022	
Thalassiosira antarctica, Strain CCMP982	Tant	CAMPEP_0202006238	
Thalassiosira gravida, Strain GMp14c1	Tgra	CAMPEP_0201684800	

Thalassiosira miniscula, Strain CCMP1093	Tmin	CAMPEP_0183747574	
Thalassiosira oceanica	Toce2	EJK65393	
Thalassiosira pseudonana	TpbHLH1	jgi|Thaps3|24007	
Thalassiosira pseudonana	TpbHLH2	jgi|Thaps3|20899	
Thalassiosira pseudonana	TpbHLH7	jgi|Thaps3|23208	
Thalassiosira sp., Strain FW	Tfw	CAMPEP_0172354094	
Thalassiosira weissflogii, Strain CCMP1010	Twei	CAMPEP_0203515806	
Thalassiothrix antarctica, Strain L6-D1	Txnt	CAMPEP_0194143228	
Tiarina fusus, Strain LIS	Tfus1	CAMPEP_0117003500	
Tiarina fusus, Strain LIS	Tfus3	CAMPEP_0117046192	

Table S4: List of the oligonucleotides used in this work.

Gene	Phatr3 Accession	Oligo name	Sequence	Type
bHLH1a	Phatr3_J44962	bHLH1a-endogenous-QFw	TTATGTCTTTCGGCGACGGG	QPCR
		bHLH1a-endogenous-QRv	AGCAACGAATGCATGCAAGG	
		bHLH1a-total-QFw	ATTCTTGGTCCCACCCGGTA	QPCR
		bHLH1a-total-QRv	ACGCCACATTGAAAAACCGAG	
		Pt-bHLH1a-DraI-Fw	GATTTTAAAATGAATAAGCCAGGACAGCG	Full lenght cloning
		Pt-bHLH1a-XhoI-Rv	TTGCTCGAGCACAGCTTCGCTGCATCGTC	
bHLH1b	Phatr3_J44963	bHLH1b-QFw	TGCGATCTCAACGGCTAATA	QPCR
		bHLH1b-QRv	CGCAAAACGAGGCTAATTTC	
bHLH3	Phatr3_J42586	bHLH3-QFw	CACTCTCATCATGCGGGAAT	QPCR
		bHLH3-QRv	GCGCGTTGTCTTCCTCTATC	
bZIP7	Phatr3_J48800	bZIP7-QFw	CCTTATTGATATTCAAGATTCCAAGG	QPCR
		bZIP7-QRv	GTTTCGGAACCTGCATAGGA	
CDKA1	Phatr3_J20262	CDKA1-QFw	AGCGGTATCAAAGGATGGAAAAG	QPCR
		CDKA1-QRv	CTTCATCTTCGGCTTCAAGGC	
CDKD	Phatr3_J10160	CDKD-QFw	ATTACTTCTGCGGAGACCATTC	QPCR
		CDKD-QRv	GCGGTAAAGATTTCGTCAAAGG	
CYCA/B;1	Phatr3_J17135	CYCA/B1-QFw	TACACCGCCACTCCAAGAC	QPCR
		CYCA/B1-QRv	TCGGAGGACGGGATGGG	
CYCB1	Phatr3_J46095	CYCB1-QFw	TCCTGGTCCGCTACTTGAAAG	QPCR
		CYCB1-QRv	GCTGGCTGGGAAGATAACGC	
Por2	Phatr3_J13001	Por2-QFw	CCTGGTTGCATTGCCGAATC	QPCR
		Por2-QRv	TCTCCAACGTATCCTCCCGT	
Rps	Phatr3_J10847	Rps-QFw	CGAAGTCAACCAGGAAACCAA	QPCR
		Rps-QRv	GTGCAAGAGACCGGACATACC	
Tbp	Phatr3_J10199	Tbp-QFw	ACCGGAGTCAAGAGCACACAC	QPCR
		Tbp-QRv	CGGAATGCGCGTATACCAGT	
VDR	Phatr3_J43240	VDR-QFw	TTTCCTTCGCAGAACCAACT	QPCR
		VDR-QRv	TTGTCGAGCACTGAAAATCG	