

1 **A bHLH-PAS protein regulates light-dependent diurnal rhythmic processes in the**
2 **marine diatom *Phaeodactylum tricorutum***

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

6
7 ^aSorbonne Université, CNRS, Institut de Biologie Paris-Seine, Laboratory of Computational
8 and Quantitative Biology, F-75005, Paris, France.

9 ^bStazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy.

10 ^cImmunRise Technologies, 75006 Paris, France.

11 ^dGhent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark
12 927, B-9052 Ghent, Belgium.

13 ^eVIB Center for Plant Systems Biology, Technologiepark 927, B-9052 Ghent, Belgium.

14 ^fDepartment of Biology, NTNU Norwegian University of Science and Technology, 7491
15 Trondheim, Norway.

16 ^hSorbonne Université, UMR 7621, Laboratoire d'Océanographie Microbienne, Observatoire
17 Océanologique, Banyuls-sur-Mer, France.

18
19 ¹These authors contributed equally to this work.

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

23
24 ²Corresponding Author:

25 Angela Falciatore

26 Diatom Functional Genomics Team

27 Laboratory of Computational and Quantitative Biology (LCQB)

28 Institut de Biologie Paris-Seine (IBPS)

29 Jussieu Campus – Sorbonne Université

30 4, Place Jussieu, 75005 Paris, France

31 Tel: +33 0144278142

32 Email: angela.falciatore@upmc.fr

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

34 **ABSTRACT**

35 Periodic light-dark cycles govern the timing of basic biological processes in organisms
36 inhabiting land as well as the sea, where life evolved. Although prominent marine
37 phytoplanktonic organisms such as diatoms show robust diurnal rhythms in growth, cell cycle
38 and gene expression, the molecular foundations controlling these processes are still obscure.
39 By exploring the regulatory landscape of diatom diurnal rhythms, we unveil the function of a
40 *Phaeodactylum tricornutum* bHLH-PAS protein, *PtbHLH1a*, in the regulation of light-
41 dependent diurnal rhythms. Peak expression of *PtbHLH1a* mRNA occurs toward the end of
42 the light period and it adjusts to photoperiod changes. Ectopic over-expression of *PtbHLH1a*
43 results in lines showing a phase shift in diurnal cell fluorescence, compared to the wild-type
44 cells, and with altered cell cycle progression and gene expression. Reduced oscillations in
45 gene expression are also observed in overexpression lines compared to wild-type in
46 continuous darkness, showing that the regulation of rhythmicity by *PtbHLH1a* is not directly
47 dependent on light inputs and cell division. *PtbHLH1a* homologs are widespread in diatom
48 genomes which may indicate a common function in many species. This study adds new
49 elements to understand diatom biology and ecology and offers new perspectives to elucidate
50 timekeeping mechanisms in marine organisms belonging to a major, but underinvestigated
51 branch of the tree of life.

52

53 **SIGNIFICANCE STATEMENT**

54 Most organisms experience diurnal light-dark changes and show rhythms of basic biological
55 processes such that they occur at optimal times of the day. The ocean harbours a huge
56 diversity of organisms showing light-dependent rhythms, but their molecular foundations are
57 still largely unknown. In this study, we discover a novel protein, *PtbHLH1a* that regulates cell
58 division, gene expression and the diurnal timing of these events in the marine diatom
59 *Phaeodactylum tricornutum*. The identification of *PtbHLH1a*-like genes in many diatom
60 species suggests a conserved function in diurnal rhythm regulation in the most species-rich
61 group of algae in the ocean. This study unveils critical features of diatom biology and
62 advances the field of marine rhythms and their environmental regulation.

63

64

65 INTRODUCTION

66 The Earth's rotation means that life evolved under a 24h diurnal cycle of alternate light and
67 dark periods. Most living organisms have developed daily rhythms of many fundamental
68 biological processes, ranging from physiology to behaviour, such that they occur at optimal
69 times of the day (1) which can enhance fitness (2, 3). These rhythms are the product of the
70 coordinated action of signals from endogenous timekeepers, together with environmental and
71 metabolic inputs (4, 5). Robust diel rhythms in growth, cell cycle, gene expression, pigment
72 synthesis, phototactic movements and bioluminescence have been also observed in a variety
73 of phytoplanktonic organisms, including diatoms (6-13). Diatoms represent the most species-
74 rich group of algae in the ocean and populate a wide range of aquatic environments (14, 15).
75 These algae of the Stramenopile lineage show peculiar genomic, metabolic and cellular
76 features, and are evolutionarily distant from the most studied model organisms in the field of
77 biological rhythms (16-20). Diatoms have an impressive capacity to deal with environmental
78 changes thanks to sophisticated acclimation mechanisms (21-25). Recent genome-wide
79 analyses also showed that 25% of the diurnal transcriptome is influenced by light-dark cycles
80 in the centric diatom *Thalassiosira pseudonana* (26). Moreover, detailed diurnal studies in the
81 pennate diatom *Phaeodactylum tricornutum* highlighted a strict temporal separation of
82 transcriptional gene networks (27, 28), as previously observed in other algae (29). Tight
83 diurnal control of the *P. tricornutum* cell cycle has also been observed (27, 30, 31), with light
84 onset triggering cell cycle progression through the transcriptional activation of the diatom-
85 specific cyclin dsCYC2 by the blue light sensor Aureochrome-1a (30). Furthermore, the
86 transcription factor bZIP14 has recently been identified as a diurnal activator of the
87 tricarboxylic acid (TCA) cycle, a process restricted to the late afternoon in diatoms (32).
88 Together, these studies illustrate complex regulation of diurnal cellular activities in *P.*
89 *tricornutum*. However, the molecular mechanisms orchestrating diurnal processes are still
90 unknown in diatoms and many other phytoplanktonic organisms. Notably, no orthologs of the
91 circadian clock components discovered in bacteria, fungi, animals or plants have been found
92 in the diatom genomes except for cryptochromes (33). Nonetheless, a number of proteins
93 containing bHLH-PAS domains, which feature in genes involved in the regulation of
94 rhythmic processes in animals (34), have been identified in diatom genomes (35).

95 In this work, we integrate transcriptomic, physiological and functional analyses to explore
96 the regulatory landscape of *P. tricornutum* diurnal rhythms. We uncover the function of a
97 bHLH-PAS protein (*PtbHLH1a*) in the regulation of critical light-dependent rhythmic
98 processes, such as cell cycle and diurnal transcription. Phylogenetic analyses reveal that

99 bHLH1a homologs are widely distributed in diatoms, thus we speculate a common function in
100 many diatom species. These results open the way to new exploration of diatom genomes in
101 search of their elusive molecular timekeepers.

102

103 **RESULTS**

104 **Transcriptome profiling identifies potential regulators of diurnal rhythms in *P.*** 105 ***tricornutum***

106 To identify potential regulators of cellular rhythmicity in *P. tricornutum*, a publicly available
107 diurnal transcriptomic dataset (27) was analyzed. One hundred and four genes with robust diel
108 oscillating expression were selected, of which eight were photoreceptors (30, 33, 36, 37) and
109 66 were transcription factors (TFs) (35), which might be involved in the generation of light-
110 dependent rhythms. The remaining 30 genes selected were potential output genes implicated
111 in diel rhythmic processes (pigment synthesis, cell cycle regulation and photosynthesis)
112 (Table S1). The transcriptional dynamics of the selected genes were examined in a 16:8-h
113 light:dark (L:D) photocycle for 32h. Hierarchical clustering (HCL) analysis of the resulting
114 expression profiles revealed 4 clusters of co-expressed genes, termed A-D (Fig. 1A), with
115 peak expression at different times between dawn and dusk (Fig. 1B). Cluster A phased at
116 dawn, suggesting a transcriptional anticipation of the light onset (Fig. 1A-B). This cluster
117 comprised 18 genes including 14 TFs, mostly belonging to the Heat Shock Transcription
118 Factor family (HSF), two DNA repair enzymes CPD photolyases (*CPD2* and *CPD3*) and one
119 carotenoid synthesis enzyme (*PDS1*). Cluster B phased around 7h Zeitgeber Time (hours after
120 illumination, ZT) and encompassed 36 genes (Fig. 1A-B), including the *dsCYC2* gene
121 controlling the onset of cell division (30). Cluster B also contained 18 TFs, of which eight
122 were sigma factors putatively involved in the regulation of chloroplast transcription, three
123 genes implicated in photoprotection (*LHCX1*, *Vdl2* and *Zep1*) and the chlorophyll synthesis
124 *POR1* gene. Such active transcription of genes involved in chloroplast activity during the light
125 period has been shown previously (21, 38). The blue light sensors *Aurochrome1b* and the
126 cryptochromes *CPF1* and *CryP-like* also belonged to cluster B and show a strong expression
127 following light onset, in accordance with previous observations (39-41). Cluster C phased
128 around ZT9 (Fig. 1A-B) and comprises 9 TFs and 10 metabolism-related genes, including
129 genes encoding photosynthetic apparatus. Finally, cluster D phased before dusk and included
130 23 TF genes including the TCA cycle regulator *bZIP14* (27, 32), likely contributing to
131 preparing cells for light to dark transition (Fig. 1A-B). Cluster D also contained the *CPF4* and

132 the Far-Red light sensing phytochrome (*DPH1*) whose peak expression at the end of the light
133 period has been observed previously (36, 39).

134 Together these results underline the existence of tight transcriptional programs phasing at
135 discrete moments of the day which potentially control the timing of cellular activities along
136 the diurnal cycle.

137

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

139 Our analysis identified two TFs, *PtbHLH1a* (Phatr3_J44962) belonging to cluster D and
140 *PtbHLH1b* (Phatr3_J44963) belonging to cluster C, which each have a Per-ARNT-Sim (PAS)
141 domain in conjunction with a bHLH DNA-binding domain. Because bHLH-PAS proteins
142 have been shown to be involved in the regulation of rhythmic processes in animals (4, 34, 42),
143 the expression profiles of *PtbHLH1a* and *PtbHLH1b* were examined in *P. tricornutum* cells
144 growing under different photoperiods. *PtbHLH1a* expression peaked at ZT8 in the 12L:12D
145 photoperiod and at ZT12 in the 16L:8D photoperiod, 4 hours before the end of the light period
146 in both cases, then gradually decreased to below detection limits at ZT0 (Fig. 1C).
147 Transcription of *PtbHLH1b* appeared to start earlier than that of *PtbHLH1a*. In cells entrained
148 in 12L:12D cycles, *PtbHLH1b* expression peaked at ZT8, whereas it peaked between
149 ZT8/ZT12 in 16L:8D photoperiods (Fig. 1C). Thus, *PtbHLH1b* expression onset almost
150 coincided in the two photoperiods although transcription dramatically dropped after ZT8 in
151 12L:12D, while remained at maximum levels up to ZT12 in long days (Fig. 1C).

152 The robustness of *PtbHLH1a* and *PtbHLH1b* diurnal expression profiles was further
153 examined under stress conditions using recent transcriptome datasets from *P. tricornutum*
154 cells grown in 12L:12D cycles in iron replete and deplete conditions (28). Iron homeostasis is
155 diurnally regulated in phytoplankton (43) and it affects rhythmic processes such as cell cycle
156 progression and diurnal gene expression in *P. tricornutum* (28). Interestingly, *PtbHLH1a* and
157 *PtbHLH1b* expression profiles showed similar patterns in both control and iron depleted
158 conditions, with peaks of expression before dusk at ZT9 (Fig. S1), similar to our observations
159 (Fig. 1C).

160 Altogether these results demonstrate robust control of *PtbHLH1a* and *PtbHLH1b* diurnal
161 expression timing, which is adjusted in a photoperiod-dependent manner and unaffected by
162 iron depletion. The involvement of *PtbHLH1a* in the regulation of diurnal light-dependent
163 rhythmic processes was hypothesized considering a possible role in dusk anticipation.

164

165 ***PtbHLH1a* ectopic expression determines phase shifts in cellular rhythmicity**

166 To determine *PtbHLH1a*'s function, cell lines were generated expressing HA-tagged
167 bHLH1a under the regulation of the *Light harvesting complex protein family F2* promoter
168 (*Lhcf2p*) (Fig. 2A), which activates transcription 3h after light onset (44). Gene expression
169 analysis allowed the selection of three independent lines, hereafter named OE-1, OE-2 and
170 OE-3, showing over-expression of the *PtbHLH1a* gene (Fig. 2B) and earlier expression
171 timing compared to the wild type (Wt) strain (Fig. 4A and S2). Next, daily cellular
172 rhythmicity was analyzed using the flow cytometer channel FL3 (excitation 488nm, emission
173 655-730 nm) that estimates chlorophyll a cellular content (8, 38), over the course of three
174 days (Fig 2C). Cellular fluorescence displayed highly oscillating rhythms in 16L:8D grown
175 cultures with a periodicity of approximately 24h (Fig. 2C, Table S2). Cell fluorescence in Wt
176 cultures increased during daytime to peak around ZT13 (Fig. 2D and Table S2), and then
177 started to decrease before night onset. This fall in fluorescence was concomitant with an
178 increase in cell concentration ((38) and Fig. S4), likely reflecting chloroplast partitioning to
179 daughter cells during cell division (7). Fluorescence progressively declined during the night
180 period, reaching a trough in the early morning (Fig. 2C). Despite maintaining rhythmicity in
181 the cellular fluorescence dynamics, OE-1 (Fig 2C), as well as OE-2 and OE-3 lines (Fig. 2D),
182 displayed a remarkable phase shift of around 1-2 h in the maximum fluorescence timing
183 compared to the Wt (Table S2). Cellular fluorescence phase responses were further
184 investigated in resetting experiments. Wt and OE-1 cultures, that show the strongest phase
185 shift phenotype, were grown in 16L:8D photoperiods, then transferred to 8L:16D and
186 monitored for another 6 days. After the transfer to 8L:16D photocycles, the timing of
187 maximum cell fluorescence in Wt cells was maintained for two days and then re-synchronized
188 to the newly imposed photoperiod, peaking at $ZT7.99 \pm 1.88$ starting from the third day (Fig.
189 2E-F). In contrast, after 3 days of re-entrainment in 8L:16D, the OE-1 line showed a 3-hour
190 phase delay ($ZT 11.49 \pm 2.87$) compared to Wt (Fig. 2E-F). Together, these results indicate that
191 diatom cellular rhythmicity is entrained in a photoperiod-dependent manner and that
192 *PtbHLH1a* deregulation alters the capacity of cells to set diurnal phase pattern.

193

194 ***PtbHLH1a* regulates diurnal cell cycle progression**

195 The altered rhythm of fluorescence upon *PtbHLH1a* over-expression described above may
196 reflect delayed or asynchronous cell division dynamics. To get further insights into the effect
197 of *PtbHLH1a* overexpression on rhythmic processes, cell cycle dynamics in the WT and OE-1
198 lines were thoroughly analysed. Cell cultures were synchronized by 40h of dark treatment and

199 harvested on an hourly basis for 12h after re-illumination. At T0, total DNA content
200 measurements showed comparable proportions of cells in G1 phase in all samples indicating
201 effective synchronization of cell cultures (Fig. 3A). Starting after 3h of illumination, a
202 progressive reduction of G1 cell number was observed in the Wt, with a minimum number
203 reached after 10h of illumination. After 10h, the percentage of Wt cells in G1 increased with
204 the emergence of daughter cells (Fig. 3A). Interestingly, compared to Wt, OE-1 cells showed
205 a slower exit from the G1 phase, with the percentage G1 cells continuing to decrease over the
206 entire 12h of illumination studied (Fig. 3A).

207 To further characterize the cell cycle deregulation caused by *PtbHLH1a* over-expression, the
208 expression profiles of specific cell cycle phase marker genes (31) were analyzed in dark-
209 synchronized Wt and OE-1 lines illuminated for 12h. The G1 phase gene markers *CDKAI* and
210 *CDKDI* showed similar expression profiles in both lines until 4h from the onset of
211 illumination (Fig. 3B). Starting from this time point, transcript levels of both genes were
212 consistently higher in the OE-1 line compared to Wt except for at the end of the time course
213 when they converged. This illustrates that the G1 phase duration of the two cell lines is
214 different. Conversely, the G2/M marker *CYCB1* showed lower expression in OE-1 compared
215 to the Wt between 4 and 8h after the onset of illumination (Fig. 3B). The expression of
216 another G2/M phase marker, *CYCA/B1*, also resulted deregulated in OE-1, presenting reduced
217 amplitude and peaking 2h later compared to Wt. Together, these results suggest that the
218 deregulation of *PtbHLH1a* affects cell cycle, possibly by altering transition from G1 to S or
219 G2/M phases.

220

221 ***PtbHLH1a* regulates pace of diel gene expression**

222 The effect of *PtbHLH1a* de-regulation on gene expression was investigated since the
223 expression of many *P. tricornutum* genes phase diurnally (27). To this end, Wt and OE-1 lines
224 were grown in 16L:8D photocycles and sampled every 3 hours over 25 hours. For this
225 analysis, genes with strong diurnal transcription oscillation were selected, including TFs
226 (*bHLH1a*, *bHLH1b* and *bHLH3*) and rhythmic genes putatively involved in chlorophyll and
227 carotenoid synthesis (*NADPH:protochlorophyllide oxidoreductase 2*, *Por2*, and *Violaxanthin*
228 *de-epoxidase-related*, *Vdr*) (Fig. 1A, (27, 38)). Total *PtbHLH1a* transcript levels, including
229 endogenous and transgenic mRNAs, were shown to be higher in OE-1 cells compared to the
230 Wt, and the expression peaking at ZT7 in the OE-1 line and ZT10 in the Wt (Fig. 4A). A
231 decrease of endogenous *PtbHLH1a* transcripts was observed in the OE-1 line compared to the
232 Wt, possibly reflecting negative feed-back mechanism of *PtbHLH1a* regulating its own

233 transcription (Fig. 4A). A similar pattern was also observed for the *PtbHLH1b* gene,
234 suggesting that *PtbHLH1a* and *PtbHLH1b* transcription is controlled by the same regulatory
235 circuit. In addition, the *bHLH3* gene showed earlier phases of expression in OE-1 compared
236 to the Wt (Fig. 4A). Similar deregulations of *PtbHLH1a*, *PtbHLH1b* *PtbHLH3* were also
237 observed in the OE-2 and OE-3 lines at ZT10 (Fig. S3). Besides TFs, the Chlorophyll
238 biosynthesis gene *Por2* was also anticipated and the *Vdr* gene presented increased amplitude
239 of expression in the OE-1 line compared to the Wt (Fig. 4A).

240 Altered gene expression observed in *PtbHLH1a* overexpression cell lines could be the
241 consequence of the deregulation of cell cycle progression (Fig. 3). To test this hypothesis,
242 gene expression was analyzed in dark conditions, when the cell cycle is arrested ((31) and
243 (Fig. S5)). Because information about transcription dynamics in this condition was limited, an
244 initial survey of expression of the previously selected 104 *P. tricornutum* diurnal rhythmic
245 genes (Fig. 1A) was performed in cells exposed to continuous dark for 30 hours. Comparative
246 analysis of transcript profiles revealed that around 20% of the genes show persistent
247 oscillation of expression in D:D, although in some cases they displayed reduced amplitudes
248 and/or shifted phases of expression compared to the 16L:8D condition. In particular, 19 genes
249 were identified which showed the highest amplitude of expression in both L:D and D:D (for
250 details see Materials and Methods), consisting of 16 putative TFs and 4 pigment-related
251 enzymes (Fig. S6). Among the analyzed transcripts, genes that were severely affected by the
252 absence of light were also found, being strongly down-regulated or over-expressed when
253 compared to the L:D condition (Fig. S7). The expression of some of these genes was further
254 analyzed in constant darkness in Wt and *PtbHLH1a* OE-1 cells for a period of 24h. In the Wt,
255 the analyzed genes showed comparable transcript profiles in D:D and 16L:8D conditions (Fig.
256 4B, Fig. S8). Conversely, 10 out of 13 tested genes displayed reduced amplitudes and shifts in
257 the phase of expression in OE-1 compared to Wt in D:D (Fig. 4B, Fig. S8). It is worth
258 mentioning that two of the analysed genes, *HSF1d* and *bZIP5*, showed almost overlapping
259 profiles in OE-1 and Wt lines (Fig. S8), excluding global deregulation of transcription by
260 modulation of bHLH1a expression. Taken together, these results suggest that *PtbHLH1a*
261 contributes to define timing of diurnal gene expression and that its activity is independent of
262 direct light inputs and cell division.

263

264 ***PtbHLH1a*-like proteins are widely represented in the genome of marine algae**

265 bHLH-PAS proteins were thought to be restricted to the animal (Opisthokonta) lineage (45)
266 until genome and transcriptome sequencing projects revealed bHLH-PAS family members in

267 diatoms (35) and other microalgae (46). Interestingly, when compared to animal bHLH-PAS,
268 diatom proteins show peculiar features including a single predicted PAS domain (Fig. 5A),
269 whereas animal bHLH-PAS proteins have two, and a N-ter extension that is absent in the
270 animal counterparts. Available transcriptomic and genomic databases of marine algae and
271 animals were searched for bHLH-PAS proteins and ≈ 90 novel bHLH-PAS proteins were
272 discovered from Rhodophyta, Cryptophyta, Stramenopila, Alveolata and basal Opisthokonta
273 organisms (Table S3). With one exception, all the newly identified proteins showed a single
274 predicted PAS domain, short C-ter extensions and N-ter regions of variable length, similar to
275 the predicted structure of diatom bHLH-PAS proteins (Fig. 5A). Notably, the only bHLH-
276 PAS possessing two PAS domains like the animal proteins was identified in *Galdieria*
277 *sulphuraria* (Rhodophyta) and represents the first TF of this family identified in
278 Archaeplastida. All the identified sequences, including selected bHLH-PAS from Opisthokonta
279 lineages, were used to perform a detailed phylogenetic analysis of the protein family using the
280 bHLH and PAS domains. This analysis revealed at least three clades of algal bHLH-PAS
281 proteins clearly separated from their Opisthokonta counterparts (Fig. 5B). Interestingly, domain
282 organization and branching positions of proteins from basal Opisthokonta (*Monosiga*
283 *brevicollis*) and microalgae (*Guillardia theta* (Cryptophyta) and *Nannochloropsis*
284 (Stramenopila)) (Fig. 4B) support a possible common origin for this TF family, from an
285 ancestor featuring single bHLH and PAS domains. However, the possible contribution of
286 horizontal gene transfer and convergent evolution to the proliferation and diversification of
287 this family cannot be excluded, and may be supported by the features of the *G. sulphuraria*
288 bHLH-PAS protein, likely independently acquired by this alga. Based on our analysis, the
289 majority of microalgal bHLH-PAS proteins fall into three separate clades: the first containing
290 9 TFs from diatoms and *Ectocarpus siliculosus*, the second comprising *PtbHLH1a* together
291 with 35 proteins from diatoms and Alveolata (Dinoflagellata), and the third comprising 41
292 proteins from Alveolata (Ciliophora and Dinoflagellata) and diatoms, including *PtbHLH1b*
293 (Fig. 5B).

294 Our results highlight diversification and widespread distribution of bHLH-PAS family
295 members in different groups of algae. Moreover, the presence of bHLH1a-like genes in the
296 genome of dinoflagellate and diatom species suggests that these proteins may share similar
297 functions in microalgae. The similarities in diel transcript regulation and timing of expression
298 between *PtbHLH1a* and the centric diatom *Thalassiosira pseudonana* ortholog, *TpHLH1* (26),
299 further reinforce this hypothesis.

300

301 DISCUSSION

302

303 The diurnal cycle is characterised by profound periodic light and temperature changes
304 which have shaped the evolution of most ecosystems on Earth. In most organisms, biological
305 rhythms are controlled by interconnected transcriptional-translational feed-back loops
306 involving TFs and integrating signals from the environment (5). Although this regulatory
307 framework is conserved among eukaryotes, the regulators responsible for the timing of events
308 within biological rhythms seem to have emerged several times through evolution (47).
309 Therefore, our current understanding of diurnal and circadian regulation, largely based on the
310 study of terrestrial model organisms, is not always appropriate or relevant for evolutionarily
311 distant marine organisms. In this study, we have shed light on the unknown regulators of
312 diurnal patterns in diatoms, one of the most prominent phytoplanktonic groups in the Ocean.
313 In agreement with previous studies (27, 28), we showed the existence of organized
314 transcriptional programs defining cellular activities along the daily cycle in *P. tricornutum*
315 and identified a number of TFs phasing at different times during the 24h cycle, as novel
316 candidates for diatom diurnal regulation. By monitoring diurnal variations in chlorophyll
317 fluorescence, robust regulation of diatom physiological rhythms that can be re-entrained to
318 changing photoperiods was also unveiled. The ability to adjust the phase according to the
319 photoperiod length constitutes one defining criterion of circadian clock-regulated mechanisms
320 (48). Likewise, the strongly oscillating diel expression pattern of a *P. tricornutum* bHLH-PAS
321 gene, *PtbHLH1a*, responded to photoperiod length with peak expression 4 hours before night
322 onset in both 12h and 16h day photoperiods. The timing of *PtbHLH1a* expression is also
323 preserved in cells under iron deficiency, in contrast to the expression of many other *P.*
324 *tricornutum* genes observed previously, and despite the growth rate reductions caused by
325 nutrient depletion (28). Functional characterization of *PtbHLH1a* established its
326 involvement in the regulation of *P. tricornutum* diurnal rhythms. Transgenic lines over-
327 expressing *PtbHLH1a* using a promoter that is activated earlier in the light period than the
328 endogenous gene maintained cellular rhythms of ~24 h but show phase-alterations that are
329 even more accentuated in re-entrainment experiments. This phenotype may reflect a reduced
330 capacity of cells to synchronize to environmental light-dark cycles and adjust the phase to the
331 new photoperiod. The participation of bHLH1a in the regulation of *P. tricornutum* cell cycle
332 progression, reported in this study, could also explain the altered cellular fluorescence
333 rhythmicity observed in the mutants. Altered cell division timing could derive from a delayed
334 exit from the G1 phase in transgenic lines compared to Wt, as also supported by the altered

335 expression of the mitotic cyclins CYCB1 and CYCA/B1 in these lines (31). The *PtbHLH1a*
336 gene could participate in gating cell divisions at night time, therefore maximizing the
337 energetic budget, as observed in several unicellular algae (7, 8, 30, 49). Interestingly, a similar
338 regulation of the cell cycle occurs in mammalian cells, where the circadian clock controls the
339 expression of G2 cycle-related genes to gate cell division at specific times of the day (50).

340 Besides cell cycle, *PtbHLH1a* deregulation also affected diurnal rhythmicity of several
341 gene transcripts. This phenotype was uncoupled from cell cycle deregulation as it was
342 observed also in the absence of cell division, during darkness. Interestingly, the deregulation
343 of gene transcription was much more pronounced when analyzed in D:D compared to L:D
344 conditions. These results suggest on one hand that multiple regulatory inputs participate in the
345 regulation of diurnal rhythmic gene transcription (48), partially masking *PtbHLH1a*
346 contribution to this process in cyclic environments, and, on the other hand, support *PtbHLH1a*
347 involvement in the maintenance of rhythms in the absence of light inputs.

348 The evidence provided in this work support the hypothesis that *PtbHLH1a* is one
349 component of an uncharacterized endogenous circadian clock in diatoms, either as part of a
350 central oscillator or as a mediator of clock inputs or outputs. With the exception of CRY (33,
351 39), orthologs of plant and animal circadian clock genes are absent in diatom genomes.
352 However, *PtbHLH1a* contains bHLH and PAS protein domains that are also present in the
353 CLOCK and BMAL proteins, components of the mammalian central circadian oscillator (51,
354 52). Interestingly, previous studies showed that the *P. tricornutum* animal-like blue light
355 sensor Cpf1 can repress the transcriptional activity of these proteins in a heterologous
356 mammalian cell system (33), suggesting at least partial conservation in the regulatory
357 program generating rhythmicity in animals and diatoms. The downregulation of endogenous
358 *PtbHLH1a* and *PtbHLH1b* transcripts in the OE lines analysed in this work likely reflects a
359 negative feedback loop in the regulation of these genes, which is typical of circadian genetic
360 oscillators, and further supports a possible role for *PtbHLH1a* in diatom rhythm regulation.
361 Moreover, as observed for the clock components, we show that *PtbHLH1a* contributes to set
362 the phase of output processes in cycling environments and to the maintenance of gene
363 expression rhythmicity in constant darkness. Some TFs characterized in this study and
364 showing altered expression patterns in *PtbHLH1a* transgenic cells (*i.e.*, *bHLH1b*, *bHLH3*,
365 *bZIP5*, *bZIP7*, *HSF1d*, *HSF1g*, *HSF3.3a*, *HSF4.7b*) represent direct or indirect targets of
366 *PtbHLH1a* activity and possible additional components of the network participating in diel
367 rhythm regulation. *PtbHLH1a* might act downstream of signal transduction cascades activated
368 by the diatom photoreceptors analysed in this study and elsewhere. The presence of a PAS

369 domain in *PtbHLH1a* also suggests that this protein might have its own light-sensing ability
370 (53).

371 Although further analyses under prolonged free running conditions (for example
372 continuous darkness and continuous illumination) will be necessary to conclusively assess the
373 involvement of *PtbHLH1a* in circadian regulation, this protein constitutes a promising entry
374 point for the characterization of diatom molecular timekeepers. Finally, the discovery of the
375 wide distribution of bHLH-PAS domain-containing proteins in diatoms, as well as in other
376 algae, has the potential to shed new light on the evolution of biological rhythms. bHLH-PAS
377 proteins might have independently acquired their function in rhythm regulation by convergent
378 evolution. However, the existence of this function in an ancient heterotrophic marine ancestor
379 that subsequently acquired plastids via endosymbiosis events (54) and prior to colonization of
380 land cannot be excluded. Regulators of cellular rhythmicity such as *PtbHLH1a* may have
381 played a critical role for diatom prominence in marine ecosystems, by synchronizing cellular
382 activities in optimal temporal programs and maximizing diatoms' ability to anticipate and
383 adapt to cyclic environmental variations.

384 Biological rhythms are still poorly understood at molecular and mechanistic levels in
385 marine algae, despite their fundamental significance to these organisms' biology and ecology.
386 Further characterization of *PtbHLH1a* homologs in diatoms and other algae is expected to
387 provide new insights into biological rhythms in marine organisms.

388

389 **METHODS**

390 **Culture conditions**

391 Wild-type *P. tricornutum* (*Pt1* 8.6; CCMP2561) cells and transgenic lines were grown at
392 18°C with shaking at 100 rpm in F/2 medium (55) without silica and illuminated at 40 μmol
393 photons $\text{m}^{-2} \text{s}^{-1}$ of white light (Philips TL-D De Luxe Pro 950). Detailed information is in SI.

394

395 **Cell cycle analysis**

396 Cells were synchronized in the G1 phase by 40h of darkness. After re-illumination, samples
397 were collected every hour for 12h. Details are in SI.

398

399 **RNA extraction and gene expression analyses**

400 Total RNA was extracted and qRT-PCR performed as described in (36). Codeset information
401 and raw nCounter data are available from the GEO database ([Series GSE112268](#)). Detailed
402 information is in SI.

403

404 **Selection of rhythmic transcripts and clustering analysis**

405 To select genes with rhythmic expression in the light-dark cycle, we used microarray data
406 from (27). To select rhythmic transcripts in D:D, standard deviation from the average
407 expression were calculated and used as selective criteria. Detailed information is in SI.

408

409 **Generation of the *PtbHLH1a* overexpressing lines**

410 Diatom transgenic lines were obtained by co-transformation of the pDEST-C-HA-*PtbHLH1a*
411 plasmid together with Nourseothricin resistance plasmid. Details are described in SI.

412

413 **Data mining, protein sequence and phylogenetic analysis**

414 Detailed information about data mining, protein sequence and phylogenetic analysis is
415 provided in SI.

416

417 **ACKNOWLEDGEMENTS**

418 We thank M. Jaubert and L. De Veylder for critical suggestions, A. Manzotti for help with
419 characterization of OE lines, P. Oliveri for support with nCounter technology, D. Petroustos
420 and G. Finazzi for support in monitoring cell physiology. This work was funded by the HFSP
421 research grant (#RGY0082/2010), EMBRIC, EMBRC-FR and a grant from the Gordon and
422 Betty Moore Foundation (GBMF 4966) to A.F.

423

424

425 **REFERENCES**

- 426
- 427 1. Pittendrigh CS (1993) Temporal Organization: Reflections of a Darwinian Clock-Watcher. *Annual Review of Physiology* 55(1):17-54.
- 428
- 429 2. Dodd AN, et al. (2005) Plant circadian clocks increase photosynthesis, growth, survival, and
- 430 competitive advantage. *Science* 309(5734):630-633.
- 431 3. Woelfle MA, Ouyang Y, Phanvijhitsiri K, & Johnson CH (2004) The adaptive value of
- 432 circadian clocks: an experimental assessment in cyanobacteria. *Curr Biol* 14(16):1481-1486.
- 433 4. Dunlap JC (1999) Molecular bases for circadian clocks. *Cell* 96(2):271-290.
- 434 5. Bell-Pedersen D, et al. (2005) Circadian rhythms from multiple oscillators: lessons from
- 435 diverse organisms. *Nat Rev Genet* 6(7):544-556.
- 436 6. Chisholm SW (1981) Temporal Patterns of Cell-Division in Unicellular Algae. *Can B Fish*
- 437 *Aquat Sci* (210):150-181.
- 438 7. Moulager M, et al. (2007) Light-dependent regulation of cell division in *Ostreococcus*:
- 439 evidence for a major transcriptional input. *Plant physiology* 144(3):1360-1369.
- 440 8. Ragni M & D'Alcala MR (2007) Circadian variability in the photobiology of *Phaeodactylum*
- 441 *tricornutum*: pigment content. *J Plankton Res* 29(2):141-156.
- 442 9. Naylor E (2010) Chronobiology of marine organisms (Cambridge University Press,
- 443 Cambridge; New York) pp x, 242 p.
- 444 10. Tessmar-Raible K, Raible F, & Arboleda E (2011) Another place, another timer: Marine
- 445 species and the rhythms of life. *Bioessays* 33(3):165-172.
- 446 11. Ottesen EA, et al. (2013) Pattern and synchrony of gene expression among sympatric marine
- 447 microbial populations. *P Natl Acad Sci USA* 110(6):E488-E497.
- 448 12. Brierley AS (2014) Diel vertical migration. *Current Biology* 24(22):R1074-R1076.
- 449 13. Hafker NS, et al. (2017) Circadian Clock Involvement in Zooplankton Diel Vertical
- 450 Migration. *Current Biology* 27(14):2194-2201.
- 451 14. Field CB, Behrenfeld MJ, Randerson JT, & Falkowski P (1998) Primary production of the
- 452 biosphere: integrating terrestrial and oceanic components. *Science* 281(5374):237-240.
- 453 15. Malviya S, et al. (2016) Insights into global diatom distribution and diversity in the world's
- 454 ocean. *Proc Natl Acad Sci U S A* 113(11):E1516-1525.
- 455 16. Baldauf SL (2003) The Deep Roots of Eukaryotes. *Science* 300(5626):1703-1706.
- 456 17. Bowler C, et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom
- 457 genomes. *Nature* 456(7219):239-244.
- 458 18. Moustafa A, et al. (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms.
- 459 *Science* 324(5935):1724-1726.
- 460 19. Flori S, et al. (2017) Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nat*
- 461 *Commun* 8:15885.
- 462 20. Raible F & Falciatore A (2014) It's about time: Rhythms as a new dimension of molecular
- 463 marine research. *Mar Genomics* 14:1-2.
- 464 21. Bailleul B, et al. (2010) An atypical member of the light-harvesting complex stress-related
- 465 protein family modulates diatom responses to light. *P Natl Acad Sci USA* 107(42):18214-
- 466 18219.
- 467 22. Allen AE, et al. (2011) Evolution and metabolic significance of the urea cycle in
- 468 photosynthetic diatoms. *Nature* 473:203.
- 469 23. Goss R & Lepetit B (2015) Biodiversity of NPQ. *Journal of Plant Physiology* 172:13-32.
- 470 24. Morrissey J, et al. (2015) A novel protein, ubiquitous in marine phytoplankton, concentrates
- 471 iron at the cell surface and facilitates uptake. *Curr Biol* 25(3):364-371.
- 472 25. McQuaid JB, et al. (2018) Carbonate-sensitive phytoferritin controls high-affinity iron
- 473 uptake in diatoms. *Nature* 555(7697):534-537.
- 474 26. Ashworth J, et al. (2013) Genome-wide diel growth state transitions in the diatom
- 475 *Thalassiosira pseudonana*. *P Natl Acad Sci USA* 110(18):7518-7523.
- 476 27. Chauton MS, Winge P, Brembu T, Vadstein O, & Bones AM (2013) Gene regulation of
- 477 carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated
- 478 to light/dark cycles. *Plant physiology* 161(2):1034-1048.

- 479 28. Smith SR, et al. (2016) Transcriptional Orchestration of the Global Cellular Response of a
480 Model Pennate Diatom to Diel Light Cycling under Iron Limitation. *PLoS Genet*
481 12(12):e1006490.
- 482 29. Zones JM, Blaby IK, Merchant SS, & Umen JG (2015) High-Resolution Profiling of a
483 Synchronized Diurnal Transcriptome from *Chlamydomonas reinhardtii* Reveals Continuous
484 Cell and Metabolic Differentiation. *Plant Cell* 27(10):2743-2769.
- 485 30. Huysman MJ, et al. (2013) AUREOCHROME1a-mediated induction of the diatom-specific
486 cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*).
487 *Plant Cell* 25(1):215-228.
- 488 31. Huysman MJ, et al. (2010) Genome-wide analysis of the diatom cell cycle unveils a novel
489 type of cyclins involved in environmental signaling. *Genome Biol* 11(2):R17.
- 490 32. Matthijs M, et al. (2017) The transcription factor bZIP14 regulates the TCA cycle in the
491 diatom *Phaeodactylum tricornutum*. *The EMBO journal* 36(11):1559-1576.
- 492 33. Coesel S, et al. (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with
493 DNA repair and transcription regulation activity. *EMBO Rep* 10(6):655-661.
- 494 34. Kewley RJ, Whitelaw ML, & Chapman-Smith A (2004) The mammalian basic helix-loop-
495 helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol* 36(2):189-204.
- 496 35. Rayko E, Maumus F, Maheswari U, Jabbari K, & Bowler C (2010) Transcription factor
497 families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol*
498 188(1):52-66.
- 499 36. Fortunato AE, et al. (2016) Diatom Phytochromes Reveal the Existence of Far-Red-Light-
500 Based Sensing in the Ocean. *Plant Cell* 28(3):616-628.
- 501 37. Jaubert M, Bouly JP, Ribera d'Alcala M, & Falciatore A (2017) Light sensing and responses in
502 marine microalgae. *Curr Opin Plant Biol* 37:70-77.
- 503 38. Hunsperger HM, Ford CJ, Miller JS, & Cattolico RA (2016) Differential Regulation of
504 Duplicate Light-Dependent Protochlorophyllide Oxidoreductases in the Diatom
505 *Phaeodactylum tricornutum*. *PLoS One* 11(7):e0158614.
- 506 39. Fortunato AE, Annunziata R, Jaubert M, Bouly JP, & Falciatore A (2015) Dealing with light:
507 the widespread and multitasking cryptochrome/photolyase family in photosynthetic
508 organisms. *J Plant Physiol* 172:42-54.
- 509 40. Oliveri P, et al. (2014) The Cryptochrome/Photolyase Family in aquatic organisms. *Mar*
510 *Genomics* 14:23-37.
- 511 41. Banerjee A, et al. (2016) Allosteric communication between DNA-binding and light-
512 responsive domains of diatom class I aureochromes. *Nucleic Acids Res* 44(12):5957-5970.
- 513 42. Young MW & Kay SA (2001) Time zones: a comparative genetics of circadian clocks. *Nat*
514 *Rev Genet* 2(9):702-715.
- 515 43. Botebol H, et al. (2015) Central role for ferritin in the day/night regulation of iron homeostasis
516 in marine phytoplankton. *Proc Natl Acad Sci U S A* 112(47):14652-14657.
- 517 44. Russo MT, Annunziata R, Sanges R, Ferrante MI, & Falciatore A (2015) The upstream
518 regulatory sequence of the light harvesting complex Lhcf2 gene of the marine diatom
519 *Phaeodactylum tricornutum* enhances transcription in an orientation- and distance-independent
520 fashion. *Mar Genomics* 24 Pt 1:69-79.
- 521 45. Yan J, Ma Z, Xu X, & Guo AY (2014) Evolution, functional divergence and conserved exon-
522 intron structure of bHLH/PAS gene family. *Mol Genet Genomics* 289(1):25-36.
- 523 46. Thiriet-Rupert S, et al. (2016) Transcription factors in microalgae: genome-wide prediction
524 and comparative analysis. *BMC Genomics* 17:282.
- 525 47. Bhadra U, Thakkar N, Das P, & Bhadra MP (2017) Evolution of circadian rhythms: from
526 bacteria to human. *Sleep Med* 35:49-61.
- 527 48. Meroz M, et al. (2006) Entrainment of the *Neurospora* circadian clock. *Chronobiol Int* 23(1-
528 2):71-80.
- 529 49. Noordally ZB & Millar AJ (2015) Clocks in algae. *Biochemistry* 54(2):171-183.
- 530 50. Matsuo T, et al. (2003) Control mechanism of the circadian clock for timing of cell division in
531 vivo. *Science* 302(5643):255-259.
- 532 51. Bunker MK, et al. (2000) Mop3 is an essential component of the master circadian pacemaker
533 in mammals. *Cell* 103(7):1009-1017.

- 534 52. Gekakis N, et al. (1998) Role of the CLOCK protein in the mammalian circadian mechanism.
535 *Science* 280(5369):1564-1569.
- 536 53. Taylor BL & Zhulin IB (1999) PAS domains: internal sensors of oxygen, redox potential, and
537 light. *Microbiol Mol Biol Rev* 63(2):479-506.
- 538 54. Archibald JM (2015) Endosymbiosis and Eukaryotic Cell Evolution. *Current biology*
539 25(19):R911-921.
- 540 55. Guillard RRL (1975) Culture of Phytoplankton for Feeding Marine Invertebrates. In *Culture*
541 *of Marine Invertebrate Animals*, D.R. Smith and M.H. Chanley, eds. Springer US.
542
543
544
545

546 FIGURE LEGENDS

547 **Fig. 1. Diurnal expression analysis of selected rhythmic *P. tricornutum* genes.** A) nCounter expression analysis of 104 selected genes in cells grown under a 16L:8D photoperiod and sampled every 4 hours for 32 hours. Four major groups of co-regulated genes (A-D) are shown based on hierarchical clustering. Expression values were normalized using *PtRPS*, *PtTBP* and *PtACTIN12* as reference genes and represent the average of three biological replicates. 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. Plot petal length is proportional to the standard deviation of phases over biological replicates. In the table, N: number of genes within each cluster; Av. Phase: average periods and phases 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 analysed by qRT-PCR. Expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of three biological replicates \pm s.e.m (standard error of the mean; black bars). For each gene, the expression value is relative to its maximum expression (maximum expression=1). Light and dark periods are represented by white and grey regions respectively.

565 **Fig. 2. *PtbHLH1a* over-expression determines phase shifts in cellular rhythmicity.**

566 A) Schematic representation of the *Lhcf2p:bHLH1a-3xHA:Lhcf1t* construct used to generate *PtbHLH1a* over-expressing lines. *Lhcf2p*: Light Harvesting Complex F2 promoter; 3HA: triple hemagglutinin tag; *Lhcf1t*: *Lhcf1* terminator. B) Quantification of total *PtbHLH1a* transcripts in OE-lines and Wt grown in 16L:8D photocycles and sampled at the ZT10. qRT-PCR expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average of 3 biological replicates (n=3) \pm s.e.m (black bars). C) Diurnal oscillation of chlorophyll fluorescence (FL-3 parameter) in Wt and OE lines entrained under 16L:8D over three days. D) Diurnal phase time calculation of the FL-3 value in Wt and OE-1, OE-2 and OE-3 lines. Values represent the average of at least 4 biological replicates (n \leq 4) from at least 2 independent experiments \pm s.e.m. E) Phase re-entrainment analysis of fluorescence rhythms after photoperiod change from 16L:8D to 8L:16D in Wt (yellow) and OE-1 (red). The FL-3 parameter was monitored in cultures grown in 16L:8D photocycles and then transferred to 8L:16D for 6 days. Values represent the average of three biological

579 replicates \pm s.e.m. F) Bar plot representation of the re-entrained FL-3 phases in Wt and OE-1
580 cultures after three 8L:16D photocycles. Values represent the average of 3 biological
581 replicates \pm s.e.m. (black bars).

582 * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, t-test.

583

584 **Fig. 3. *PtbHLH1a* over-expression affects cell cycle progression.** A) Cell cycle progression
585 dynamics of Wt (yellow) and OE-1 (red) lines shown as the proportion of cells in the G1
586 phase measured by flow cytometry each hour over 12h of illumination following dark
587 synchronization. Results are representative of three biological replicates \pm s.e.m (black bars);
588 t-test significance is indicated by *: $P < 0.05$, **: $P < 0.01$, t-test. B) qRT-PCR expression
589 profiles of G1 (*PtCDKA1*, *PtCDKD1*) and G2/M phase marker genes (*PtCYCB1*,
590 *PtCYCA/B1*) in synchronized Wt and OE-1 cell lines over 12h of illumination. Expression
591 values were normalized using *PtRPS* and *PtTBP* as reference genes and represent the average
592 of two independent biological replicates. For each gene, the expression value is relative to its
593 maximum expression (maximum expression=1).

594

595 **Fig. 4 *PtbHLH1a* over-expression alters rhythmic diel gene expression.** A) qRT-PCR
596 diurnal expression analysis of *PtbHLH1a*, *PtbHLH1b*, *PtbHLH3*, *PtPor2* and *PtVdr*
597 transcripts in 16L:8D entrained Wt (yellow) and OE-1 (red) cultures. *bHLH1a-total*
598 represents the sum expression of the *bHLH1a* endogenous and the transgene transcripts;
599 *PtbHLH1a-endogenous* refers to the endogenous gene only. Expression values represent the
600 average of 2 biological replicates \pm s.e.m. (black bars). B) nCounter expression analysis of
601 *PtbHLH1a*, *PtbHLH1b*, *PtbHLH3*, *PtPor2* and *PtVdr* transcripts during 24 hours of dark free-
602 running period in the Wt (yellow) and OE-1 (red) lines. Cells were previously entrained in
603 16L:8D cycles. Data represent the average of 3 biological replicates \pm s.e.m (black bars).

604 Grey rectangles represent dark periods. Expression values were normalized using *PtRPS* and
605 *PtTBP* as reference genes. For each gene, the expression value is relative to its maximum
606 expression (maximum expression=1).

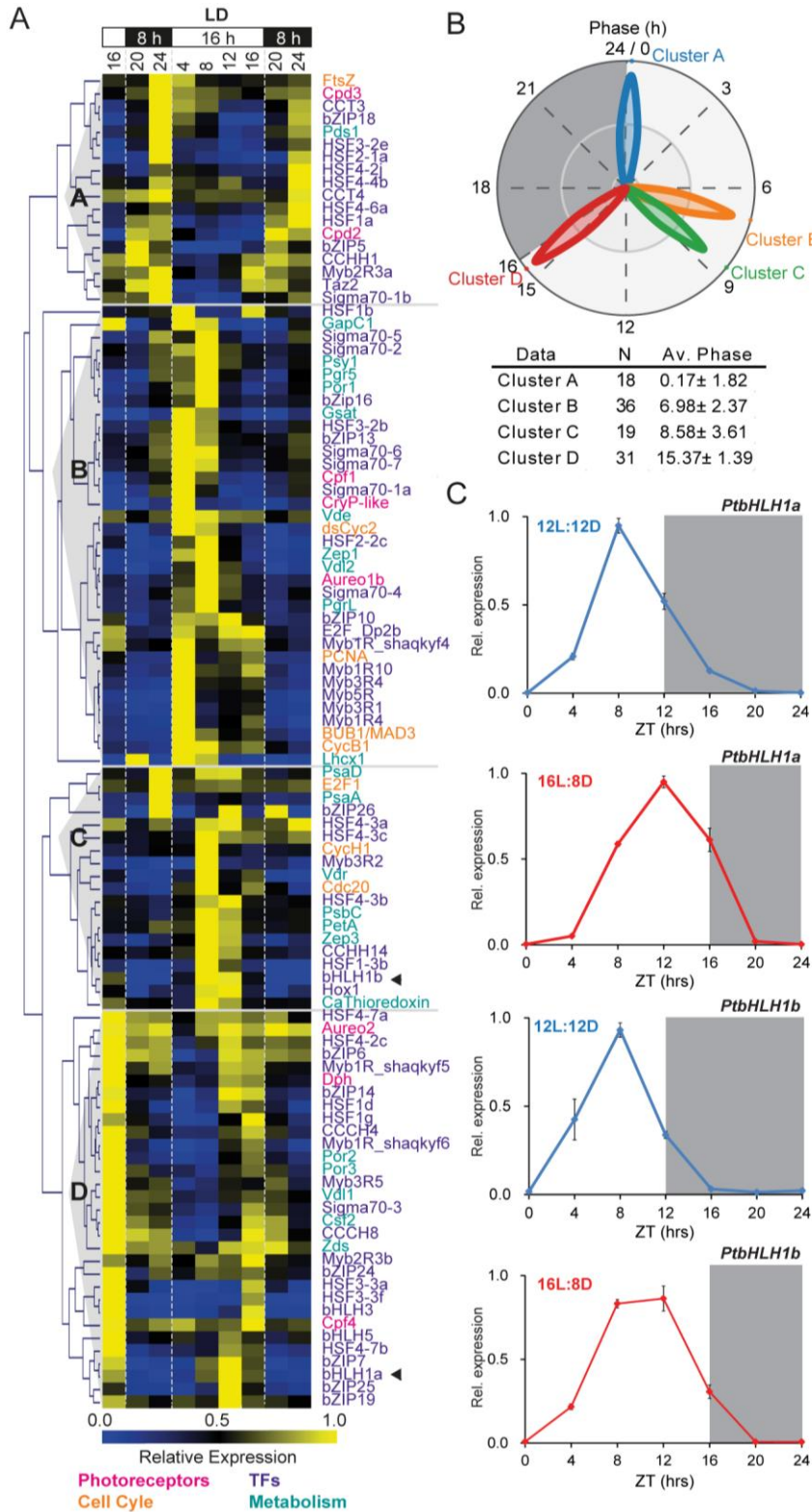
607

608 **Fig. 5. bHLH-PAS protein family structure and phylogeny.** A) Schematic representation
609 of bHLH-PAS protein domain architecture across Eukaryotes. Segmented line indicates
610 possible absence of the second PAS domain in some Opisthokonta species. The grey patterns
611 before the bHLH domain and after the PAS domain represent the variations in N-ter and C-ter
612 length in different organisms. B) Maximum Likelihood (ML) phylogenetic tree of the bHLH-

613 PAS family. The Opisthokonta clade is used as the outgroup. Numbers refer to bootstrap values
614 of the basal nodes using ML (RAxML, 1000 bootstraps) and Bayesian Inference (MrBayes,
615 2.5M generations, 25% burn-in). The asterisk, the arrows and the square indicate the position
616 of *Monosiga brevicollis* bHLH-PAS, *P. tricornutum* HLH1a and bHLH1b, and *Thalassiosira*
617 *pseudonana* bHLH1, respectively. The colour code indicates the lineage corresponding to
618 each bHLH-PAS protein, shown in Fig 5A.

619

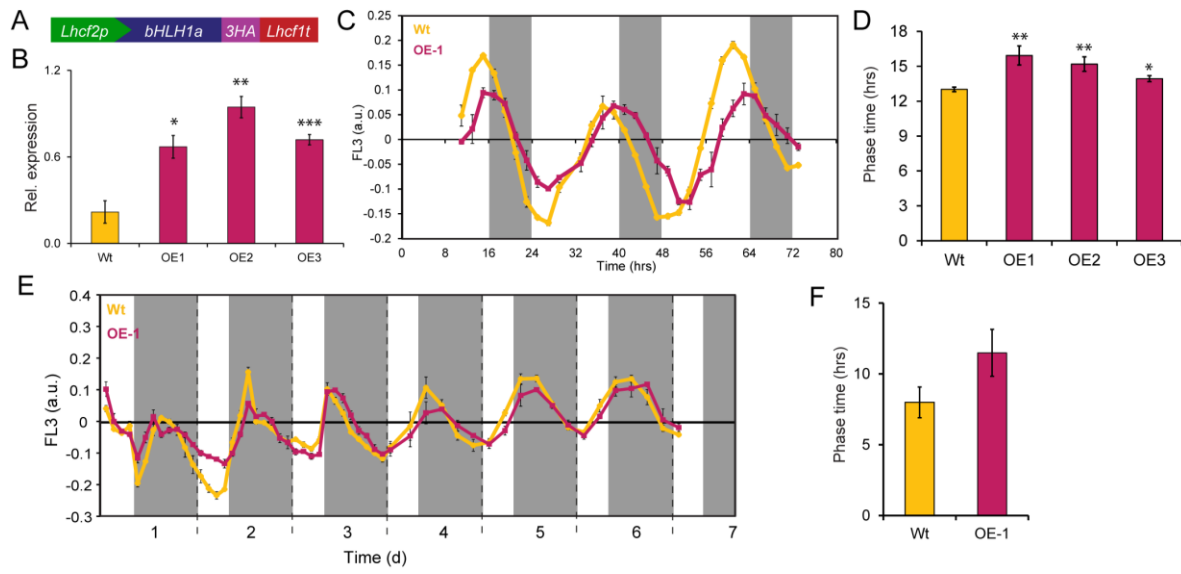
620 **Figure 1 - Annunziata et al.**



621

622

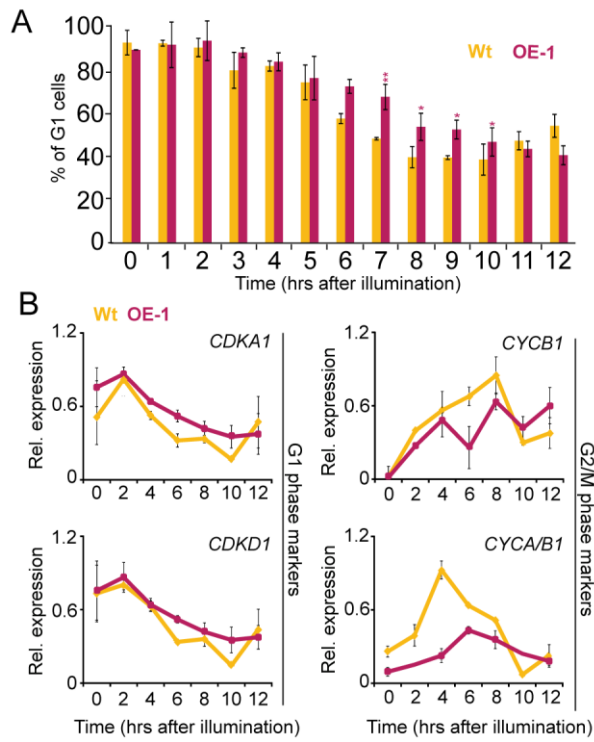
623 **Figure 2 - Annunziata et al.**



624

625

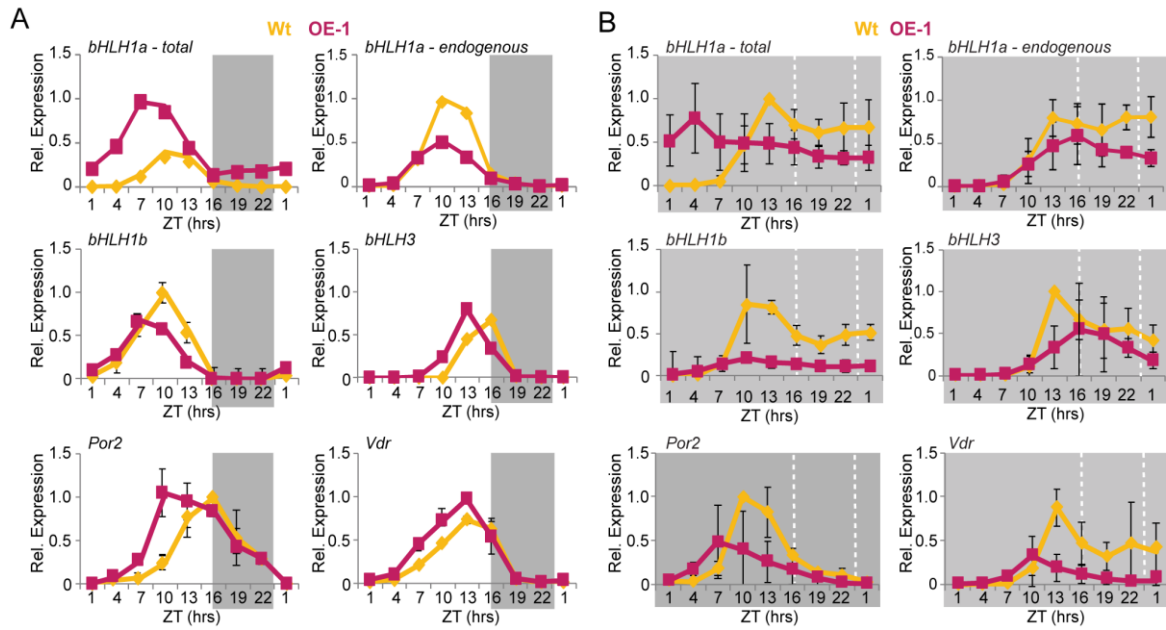
626 **Figure 3 - Annunziata et al.**



627

628

629 **Figure 4 - Annunziata et al.**

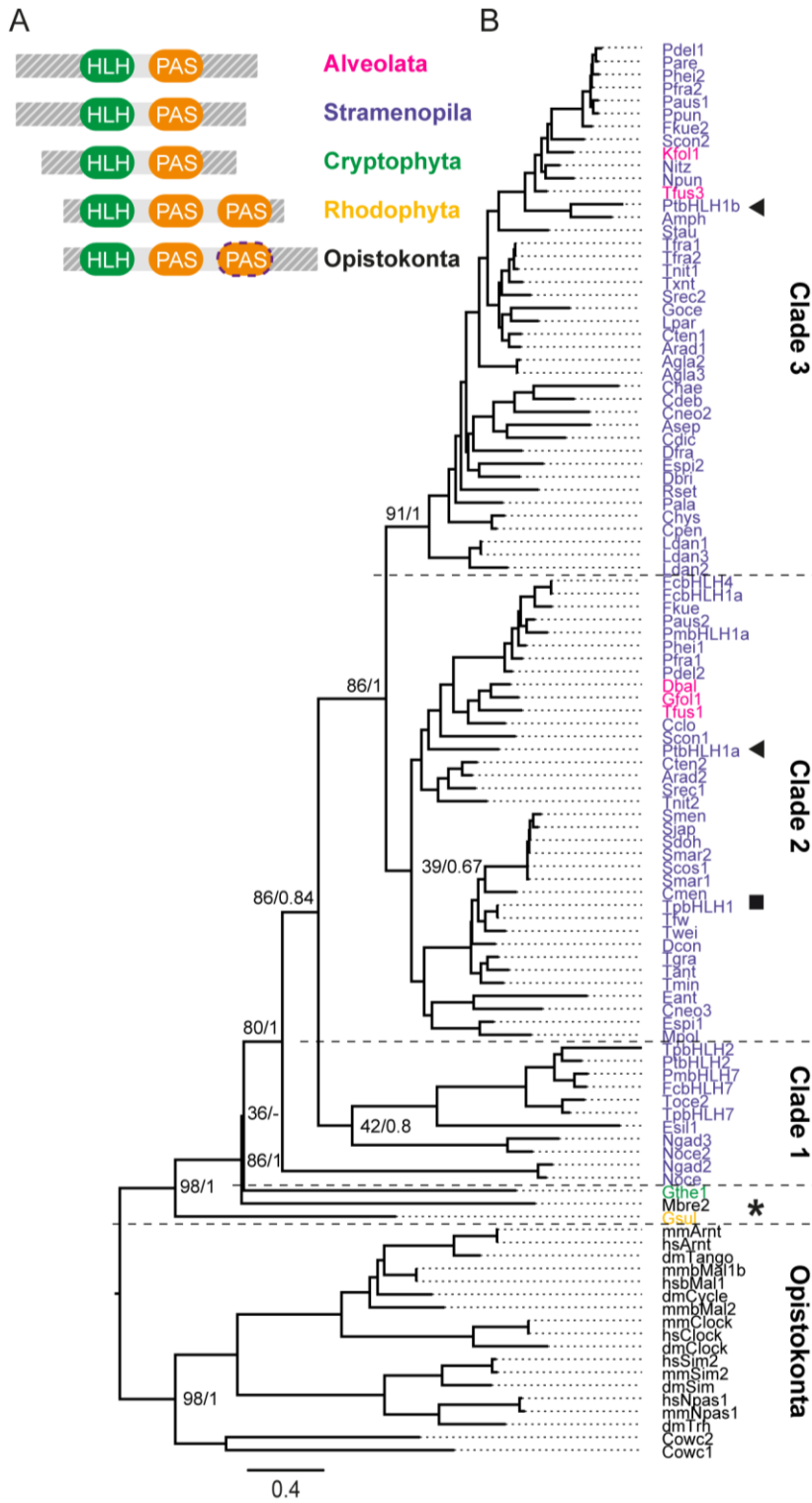


630

631

632

633 **Figure 5 - Annunziata et al.**



634
635
636
637

638 **SUPPORTING INFORMATION**

639

640 **Culture conditions**

641 For experiments in different photoperiods, cultures were pre-adapted to the different L:D
642 cycles for 2 weeks before starting the experiment. For experiments in continuous darkness
643 cells were pre-adapted in 16L:8D photocycles for 2 weeks, then transferred to D:D at the start
644 of the experiment. Growth measurements were performed using a MACSQuant Analyser flow
645 cytometer (Miltenyi Biotec, Germany) by counting the cells based on the R1-A (630nm
646 excitation, 670-700nm emission) versus the R1-H parameters. Phase rhythmicity assays were
647 carried out by measuring the Chlorophyll fluorescence using the flow cytometer FL-3
648 parameter (488 nm excitation, 670-700 nm emission). For the re-entrainment experiment,
649 cultures were initially entrained under 16L:8D photocycles at $40 \mu\text{mol m}^{-2} \text{s}^{-1}$ of white light,
650 then transferred to 8L:16D photocycles at $80 \mu\text{mol m}^{-2} \text{s}^{-1}$ of white light for 6 days. All phase
651 time and period calculations were performed using the MFourfit curve-fitting method using
652 the Biodare2 tool (biodare2.ed.ac.uk, (1)).

653

654 **Cell cycle analysis**

655 For cell cycle analysis cells were pelleted by centrifugation (4000 rpm, 15 minutes, 4°C),
656 fixed in 70% EtOH and stored in the dark at 4°C until processing. Fixed cells were then
657 washed three times with 1×PBS, stained with 4',6-diamidino-2-phenylindole (at a final
658 concentration of 1 ng/ml) on ice for 30', then washed and resuspended in 1xPBS. After
659 staining, samples were immediately analyzed with a MACsQuant Analyser flow cytometer
660 (Miltenyi Biotec, Germany). For each sample, 30,000 cells were analysed and G1 and G2
661 proportions were inferred by calculating the 2c and 4c peak areas at 450 nm (V1-A channel)
662 using the R software. A peak calling method was applied to the resulting histogram, based on
663 a 1st derivative approach (2). The locations of G1 and G2 peaks were first determined using
664 G1 and G2 reference samples and then used to identify G1 and G2 cells in the experimental
665 samples. The area under each peak was used as a proxy for the proportion of cells in each
666 population.

667

668 **RNA extraction and gene expression analyses**

669 For qRT-PCR analysis *PtRPS* and *PtTBP* were used as reference genes. Each independent
670 replica of the qRT-PCR data was normalized against the maximum expression value of each
671 gene (*i.e.*, gene expression range lies between 0 and 1 across the time series). Average
672 expression and standard error was then calculated and plotted. The full list of oligonucleotides
673 used in this work can be found in Table S4. For the nCounter analysis, gene specific probes
674 (Table S1) were designed and screened against the *P. tricornutum* annotated transcript
675 database (JGI, genome version 2, Phatr2) for potential cross-hybridization. Total RNA
676 extracts (100 ng) from three biological replicates were used for hybridization. Transcript
677 levels were measured using the nCounter analysis system (Nanostring Technologies) at the
678 UCL Nanostring Facility (London, UK) and at the Institut Curie technical platform (Paris,
679 France) as previously described (3). Expression values were first normalized against the
680 internal spike-in controls, then against the geometric mean of the 3 reference genes *PtRPS*,
681 *PtTBP* and *PtACTIN12*.

682 **Selection of rhythmic transcripts and clustering analysis**

683 For the selection of genes with rhythmic expression in the light-dark cycle, we used
684 microarray data from (4). First, we identified all the genes belonging to the TFs,
685 photoreceptors, cell cycle and metabolism-related categories (pigment synthesis and
686 photosynthesis). Then, transcripts were ranked based on a defined t-value for each time point
687 (mean gene expression of the replica/(1+s.d.)) and those showing t-value $>+0.7$ or <-0.7
688 across the time series, were retained. The nCounter expression data was normalized against
689 the maximum expression value of each gene, in a similar way to qRT-PCR expression data.
690 This normalization was applied to the 3 replicas independently and for each condition (L:D
691 and D:D) with the average expression value used for the clustering analysis. Hierarchical
692 clustering analysis was performed with MeV 4.9 (5) using Pearson correlation. Peak analysis
693 was performed using the MFourfit curve-fitting method defining average expression phases
694 for each cluster (biodare2.ed.ac.uk, (1)).

695 For the selection of rhythmic transcripts in D:D, expression values were normalized using as
696 reference genes *PtRPS*, *PtTBP* and *PtACTIN12* and the genes with the highest values of
697 standard deviation from the average expression (M value) over the two time courses (16L:8D
698 and D:D) were selected. A threshold equal to 1 was set using the published *P. tricornutum*
699 diurnal microarray dataset (4) as background. Gene expression profiles were further
700 empirically examined and false positives eliminated.

701 **Generation of the *PtbHLH1a* overexpressing lines**

702 Transformed cells were tested for the presence of the transgene by PCR and qRT-PCR
703 analysis (see Table S4 for oligonucleotide sequences). The full length *PtbHLH1a* coding
704 sequence was obtained by PCR amplification with the specific oligonucleotides *PtbHLH1a*-
705 *DraI-Fw* and *PtbHLH1a-XhoI-Rv* on cDNA template using the Phusion high fidelity DNA
706 polymerase (Thermo Fisher, USA). The PCR fragment was inserted into the pENTR1A
707 vector (Invitrogen, USA) using the *DraI/XhoI* restriction sites, and recombined with the
708 pDEST-C-HA vector (6).

709

710 **Data mining, protein sequence and phylogenetic analysis**

711 The *P. tricornutum* bHLH1a (Phatr3_J44962) protein sequence was used as the query for
712 blastP analyses on the JGI, NCBI and MMETSP public database (7). Searches of the pfam
713 database for proteins possessing both the HLH and PAS domains were also performed. The
714 retrieved sequences were analyzed using the batch search tool on the CDD (Conserved
715 Domain Database) NCBI server to retrieve proteins presenting at least one HLH and one PAS
716 domain only. We identified 100 HLH-PAS proteins from 71 marine algal species which were
717 aligned using MAFFT (8), along with 22 HLH-PAS proteins from relevant metazoan (*Homo*
718 *sapiens*, *Mus musculus* and *Drosophila melanogaster*) and unicellular Opisthokonta (*Monosiga*
719 *brevicollis* and *Capsaspora owczarzaki*). Preliminary phylogenies were produced with MEGA
720 7 (9) to eliminate ambiguously aligned sequences, refining the alignment to 107 sequences
721 and a final length of 198 aa (<5% gap per position). The best aminoacidic model to fit the
722 data was estimated with ProtTest 3.4.2 (10). Phylogenetic analyses were performed with
723 RAxML (1000 bootstraps) and MrBayes 3.2.6 (2.5 million generations, 2 runs, 25% burn-in)
724 on the CIPRESS gateway (11). The final tree was edited in FigureTree 1.4
725 ([http://tree.bio.ed.ac.uk/software/Figure tree/](http://tree.bio.ed.ac.uk/software/Figure%20tree/)). GenBank accession codes of the genes utilized
726 in the bHLH-PAS phylogenetic analysis are reported in Table S3.

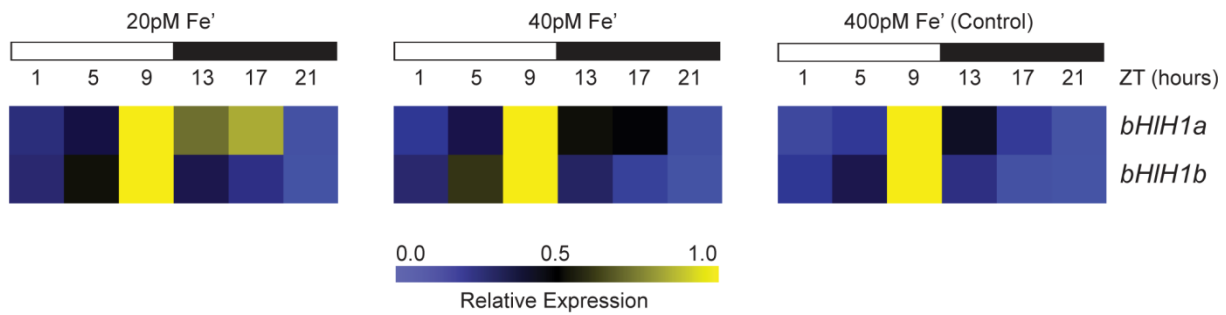
727

728

729

730

731 **Figure S1 – Annunziata et al.**



732

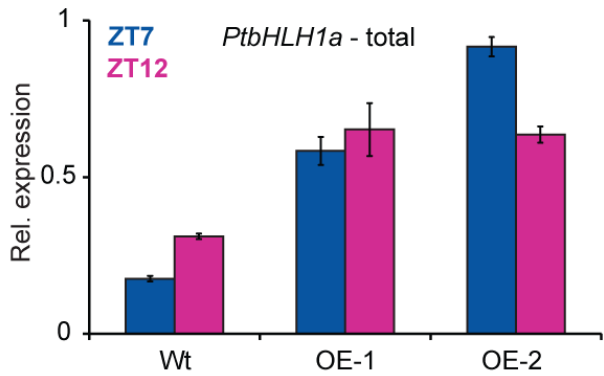
733

734 **Fig. S1. Diel expression patterns of *PtHHL1a* and *bHHL1b* genes under Fe-depletion conditions in**
735 **12L:12D photoperiods.** Diel expression patterns of *PtHHL1a* and *bHHL1b* in normal (400
736 pM Fe') and iron depletion conditions (40 and 20 pM Fe') were obtained using transcriptome
737 data extracted from (36). Light and dark periods are represented by white and black
738 rectangles. Expression values are given relative to the maximum expression for each gene,
739 where '1' represents the highest expression value of the time series.

740

741 **Figure S2 – Annunziata et al.**

742
743



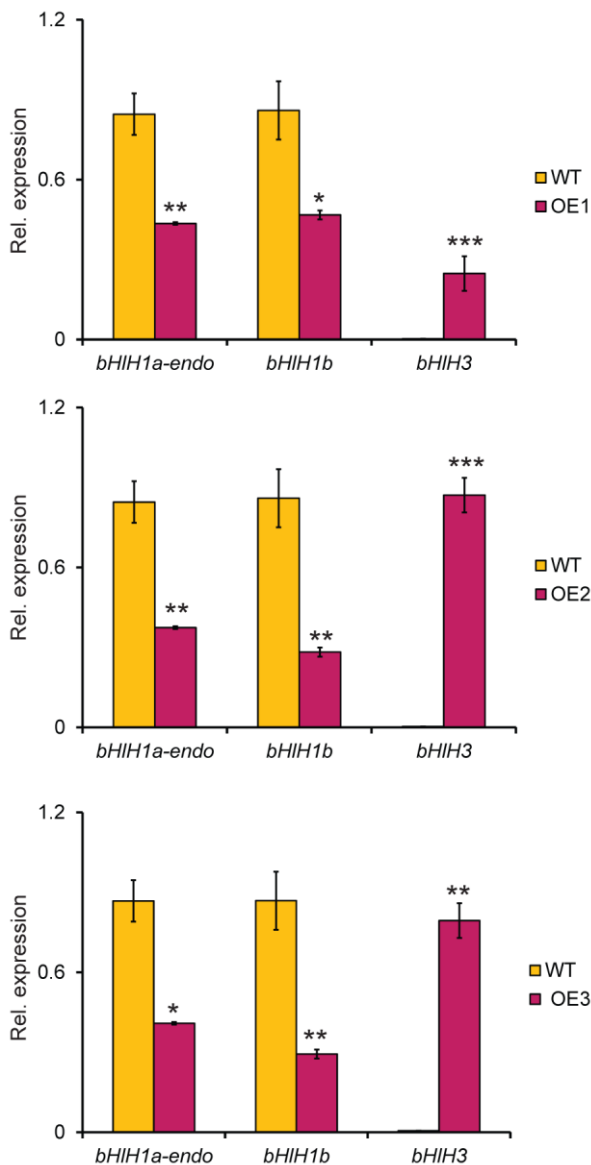
744
745

746 **Fig. S2. The *Lhcf2p:bHLH1a-3xHA:Lhcf1t* construct drives over-expression and**
747 **anticipation of *PtbHLH1a*.** Quantification of total *PtbHLH1a* transcripts in two independent
748 *PtbHLH1a* over-expressing lines (OE-1, OE-2) compared to the wild type (Wt) strain. Cells
749 were grown in 16L:8D photocycles and sampled at the ZT7 and ZT12 time points. qRT-PCR
750 expression values were normalized using *PtRPS* and *PtTBP* as reference genes and represent
751 the average of two biological replicates (n=2) \pm s.e.m (black bars).

752

753 **Figure S3 – Annunziata et al.**

754



755

756

757 **Fig. S3. *PtbHLH1a* over-expression alters rhythmic diel gene expression in three**

758 **independent lines.** Quantitative gene expression analysis by qRT-PCR of *PtbHLH1a-*

759 *endogenous*, *PtbHLH1b* and *PtbHLH3*, transcripts in 16L:8D entrained Wt, OE-1, OE-2 and

760 OE-3 cultures. Samples were harvested at ZT10. *PtbHLH1a-endo* refers to the transcript

761 levels of the endogenous gene only. Expression values represent the average of 3 biological

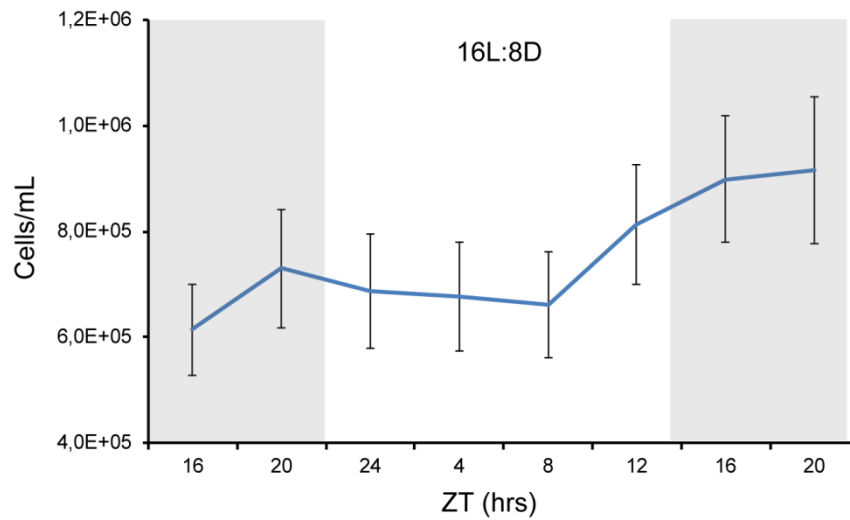
762 replicates \pm s.e.m. (black bars) and were normalized using *PtRPS* and *PtTBP* as reference

763 genes. For each gene, expression values are represented as relative to its maximum expression

764 corresponding in the graph to '1'. *P<0.05, **P<0.01, ***P<0.001, t-test.

765

766 **Figure S4 – Annunziata et al.**



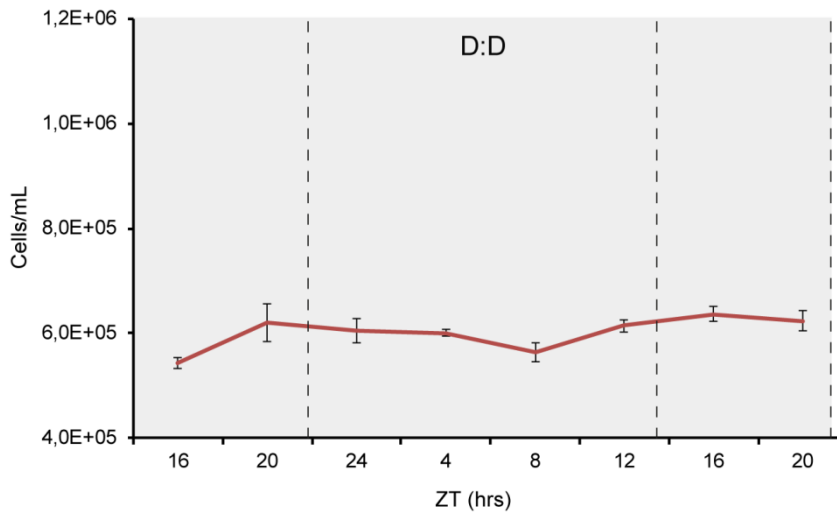
767

768

769 **Fig. S4. Diurnal cell growth dynamics in *P. tricornutum*.** Diel cell number measurements in
770 wild type cultures grown under 16L:8D photoperiods. Values represent the mean counts of
771 three independent biological replicates \pm s.e.m. (black bars).

772

773 **Figure S5 – Annunziata et al.**



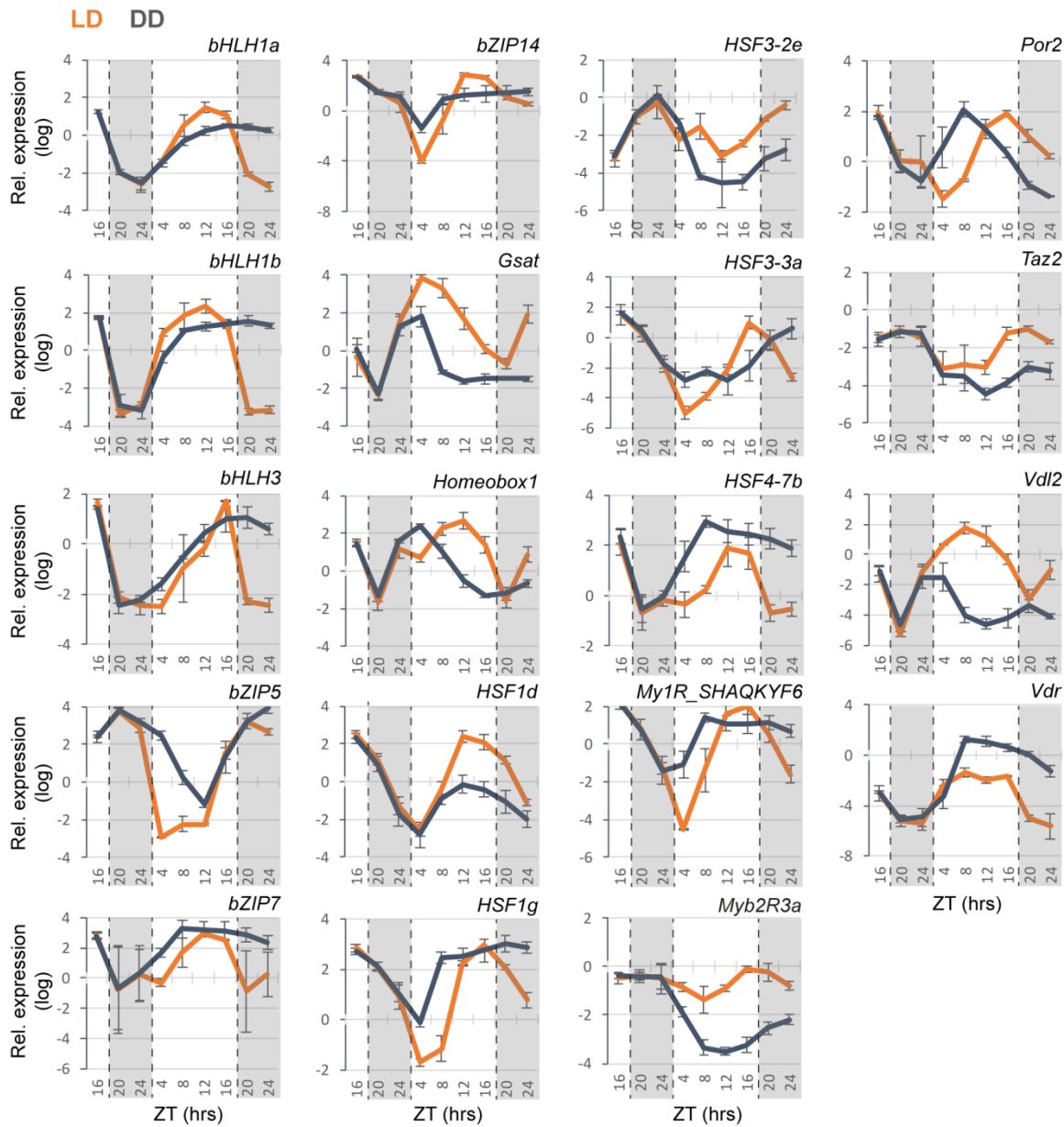
774

775 **Fig. S5. Cell division is arrested in continuous dark conditions in *P. tricornutum*.** Diurnal
776 cell growth measurements of the wild type cultures entrained in 16L:8D and then transferred
777 to continuous darkness condition (D:D). Values represent the mean counts of three
778 independent experiments \pm s.e.m. (black bars).

779

780 **Figure S6 – Annunziata et al.**

781



782

783

784 **Fig. S6. nCounter expression analysis of genes maintaining rhythmic expression in D:D**

785 **conditions and 16L:8D condition in Wt cells.** Data represent the average expression of

786 biological triplicates \pm SD and are normalized using the *PtRPS*, *PtTBP* and *PtACTIN12*

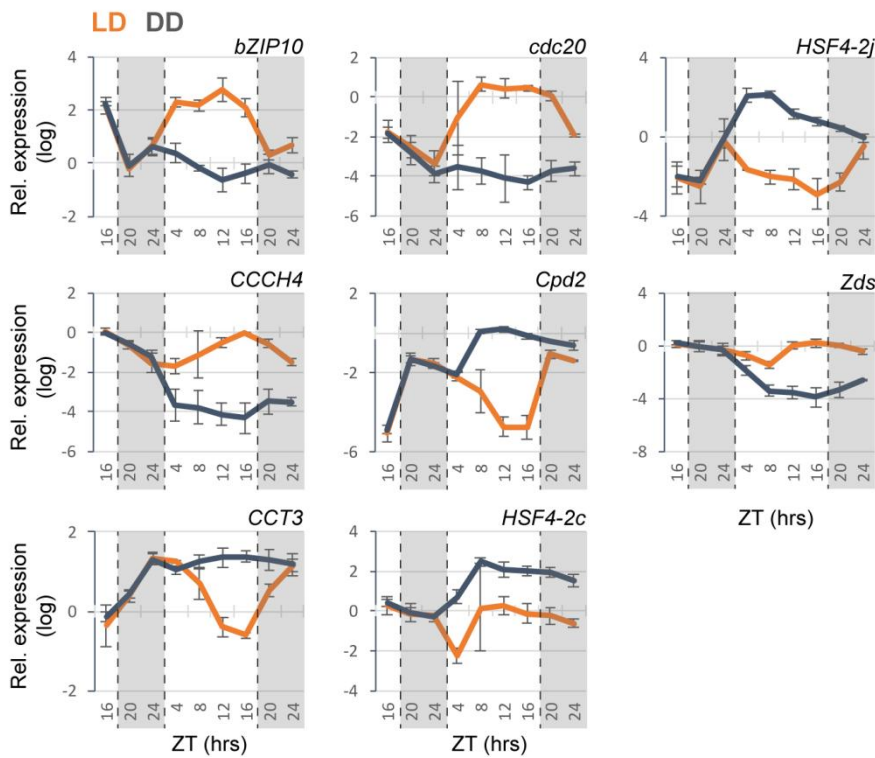
787 reference genes. For each gene, the expression value is relative to its maximum expression

788 corresponding in the graph to '1'. Results for cells grown in 16L:8D cycle are shown in

789 orange (L:D); Results for cells in constant darkness (following 16L:8D adaptation) are shown

790 in grey (D:D).

791 **Figure S7 – Annunziata et al.**



794 **Fig. S7. nCounter expression analysis of selected genes with altered rhythmic expression**
795 **in Wt cells in D:D conditions compared to 16L:8D condition.** Expression values represent
796 the average of three biological triplicates \pm SD and are normalized using the *PtRPS*, *PtTBP*
797 and *PtACTIN12* reference genes. For each gene, the expression value is relative to its
798 maximum expression corresponding in the graph to '1'. Results for cells grown in 16L:8D
799 cycle are shown in orange (L:D); Results for cells in constant darkness (following 16L:8D
800 adaptation) are shown in grey (D:D).

801

802

803

804

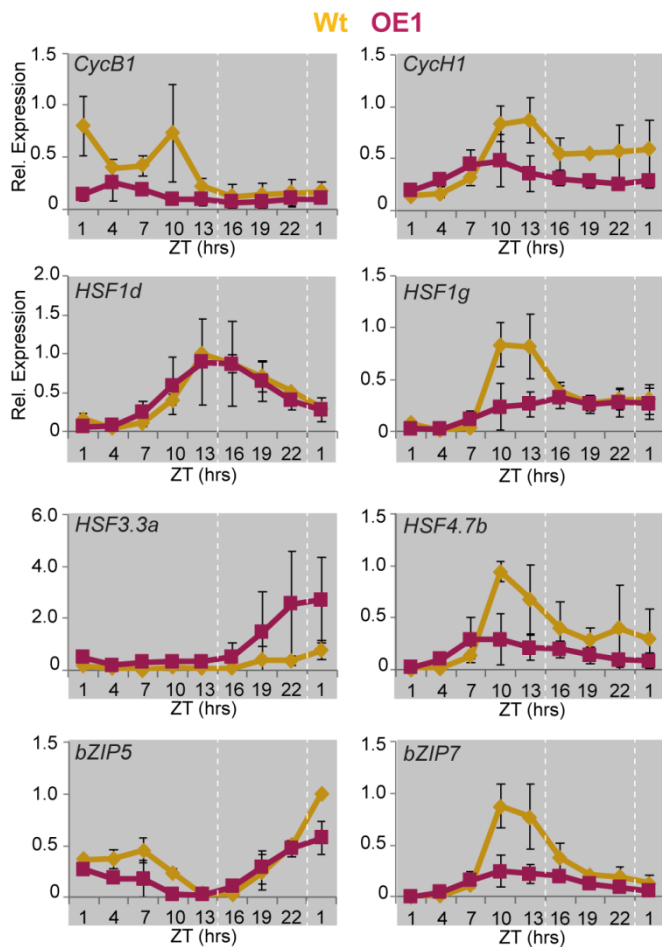
805

806

807

808 **Figure S8 – Annunziata et al.**

809



810

811

812 **Fig. S8. nCounter analysis of selected rhythmic gene expression profiles in continuous**
813 **darkness in Wt and *PtBHLH1a* OE-1 cells.** Cells were entrained in 16L:8D cycles, then
814 transferred to D:D and collected every 3 hours for 24h. Expression values represent the
815 average of three biological replicates (n=3) \pm s.e.m (black bars) and have been normalized
816 using *PtRPS* and *PtTBP*.

817

818

819

820

821

822 **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	ATGTAAGCCTGCAGCCACTGAGGACTTGTGCTGTAAACCTGATTCGATATTCAAAATGCGGCTCACTAAAAAGACATCGTAGTCCAGTGCAGTCCC
Photoreceptor	Aureo1b	#N/A	Phatr3_J49458	TTATTGCGCGAGTGTCTACCTCGAGTCCGTTGTTTGAAGAAATGGACGGCATAGATCAAACGAAAGGGCCAACTCGGAACGTGCCGATTTTCATTGAT
Photoreceptor	Aureo2	XM_002183279.1	Phatr3_J15468	AGACGTTACTGCTACTCTGAACACACTGCGCATGATACACCATTTTGGAAACAAGCTCTTTATTGCGGCATTGCGTGACGCGCAAAAATAACATTGTC
TF	bHLH1a_PAS	XM_002179032.1	Phatr3_J44962	TACTTTGGTCAATATCAAAGTCAGTTTGGTACGAACAGCTCAGCATAGCCCTCGGTTTTTCAATGTGGCGTTGGTGCATCAGACGATGCAGCGAAGCTG
TF	bHLH1b_PAS	XM_002183812.1	Phatr3_J44963	AGGAATTGACTACCGTCTGTTTTCAATCATTGTCATATGCCATGGGGTGGCTTCATTGGACGGAAGAACTAGTGTCTGCAACAGCTCTTCGAATCT
TF	bHLH3	XM_002176560.1	Phatr3_J42586	ATCACTAGCAGACATTCAGATTTCATCTGCGGGAGAAGGACGGTCTATTCCGTCAGTAGTTGGTACAAGTTAGTGGGTTCAAGTCTCGCTGGGAAGCTG
TF	bHLH5	XM_002177858.1	Phatr3_J43365	ACCCTCTGCAGCGCAATCATTGCTCCGCATACAGCATAGGGTGCATTTCTCAATGAACGCCATTCCTTTTCTCAGACCTTTGCTCTCGCTGGCGT
Cell cycle	BUB1/MAD3	XM_002178639.1	Phatr3_J10954	CGTGTGTCATGTATGCGGACAAAACCGATCGTCCACTGGAAGTTTTCCAGCATCTACATCAGCAAAGGATCGGAGCGATATTGCCGTGTTCTGGATG
TF	bZIP10	XM_002177776.1	Phatr3_J43744	AAGTGTCTCGGAATCGCTCGCGGAAGAAGTTATGATTTGAAGAATCTCAACGACGCGTATTTCTTCTCGCGTGCACGGAACCCCTCAAACA
TF	bZIP13	XM_002181635.1	Phatr3_J47278	CGCGCGGCTTCGCGTCTAAGATTTCGATGCGCCATCACCAGACTTGAACCGAGGTTAGTGGTTGGAAGGATAAATACACAAGCAATGGAACGCTTG
TF	bZIP14	XM_002179477.1	Phatr3_EG02108	ATGCAGAGCTTGGAGGTCGAGTACGGATCTCAAGCATGAGCAATTCGTTAAACAGATCATTAAACGAAAGAACCGCGCAACATTCGTTGGGGC
TF	bZIP16	XM_002181462.1	Phatr3_J47279	CTTTTGAAGCAGTCAACAATCCGCAACATGAACAGTTCACACGCCAGGAAGTCGTCGTCACCTGAACGAAGATCTCGGCTCGCTCGGAAGCGTCTG
TF	bZIP18	XM_002176556.1	Phatr3_J42577	ACTCTCTATAGCGGAATCCGAGCCGATTTCAAACATATAGCTCAGGCTGCTGTTTCAATCTGATCATGTGCGTGGAACTACCAAGTTGAATCTGG
TF	bZIP19	XM_002183101.1	Phatr3_J48701	CCTGCACCTTCTCAAATCAGTCTGCAACAGAGTGAACAAGAAGAAATATGGCTGCCAGGCATTTACCTACACCTTTTATAGCGTTTTGCGATTGTC
TF	bZIP24	XM_002184633.1	Phatr3_J49887	TTGTCACGGGCCAACAGACTATCTGGAAGATCACTTCTTGCCAAAGAGAAGCGAAAGAACTCGTTCGGCTTTGTCAAGCATCGCGCTTCGGGGAA
TF	bZIP25	XM_002185584.1	Phatr3_J46647	TCGTCGTTACGATGGCATCTGTACCACTCACCAGTAGGAAATATCTGTCATCCTGATACCAGTTATCAGCAGTTTCTGCTGCTATTACTGCAGCGG
TF	bZIP26	XM_002182641.1	Phatr3_EG02494	GATCTTTGACGACTCAAGGAGCGCAACTGTCTCAGCATCTCAGGAGCTTCGCGTGCAGCAGGAAAGTGAAGCCGCTTTGAATCGAAACCTC
TF	bZIP5_PAS	XM_002179145.1	Phatr3_J45142	CCCTGTTTCCGAAACAGCGCATTCGATCCGCAAAACAGCTAGCAACGCAACATCAGATCGCAATGATGATGCAAGACGTGGCTATTATGAGTG
TF	bZIP6_PAS	XM_002184884.1	Phatr3_J50039	CGCGCAATTAATACACAATATCAGTAGACAGCCCAACATCCGAAAATACGATGTGCGAGTGTAGACCGCAGCAGGGTTCAAATAATTCGGCGAAC
TF	bZIP7_PAS	XM_002183297.1	Phatr3_J48800	CGTGAGAGACCTGGCCATCGTAATCAAGGATTAGAGGCATTAGCGCGGATTGCGACAGCAAGCTGCTCTCAGGCGCACAGTTTATGACCAACCGAAT
Metabolism	CaThioredoxin	XM_002184396.1	Phatr3_J49634	CTGGAGAGCCCTACCGTGTAGTGTCCGTTATTTGAAAATGTCGTTGTTTGGCTTACGCTAGAAACATTCGACAGCTCCATTCGAGTCTGAA
TF	CCCH4	XM_002181104.1	Phatr3_J13664	CATGTGCGCCTTGGAGCTCGCACAATAACATGTCTACTGGCGAAAACATTGGAAGGACATTGATGGTCCGAAGCCCAAGCAATCTTAATACTCTA
TF	CCCH8	XM_002186091.1	Phatr3_J44042	TCCGAAAATATGTGCGAAGAGTCCCACTTGTCCGATGTTCTTGGCTCATAGAGTATTGGCCGACGGAAGTGAAGAGATTGAGTTTCCGGGTGAA
TF	CCHH1_CCCH20	XM_002179821.1	Phatr3_EG02317	TAACATAACAACGTTCTCAGCTAAACGGTTGAACCATCTGCCTTCAAGAACCTATCCGACTTACATCCCACTCATAGGATCAGGATCCAAAC
TF	CCHH14	#N/A	Phatr3_J34600	GCCCAGTTTTCGCGGAGGCATCCCGTTCAATCCGCGCATCTGAGTTACGAAACGCACTACATGGCTACCAAAAAGTTTACACTGATCCCGAGTTGTTT
TF	CCT3	XM_002185983.1	Phatr3_J43850	AAACGCTGCGCGCGTTTGAATAAGAAGATTCGCTACGGTTGCCGTAATAATTTAGCAGACCGCGGTTGCGCGTGAAGGGCGGATTCGTGAACGTT
TF	CCT4	XM_002178133.1	Phatr3_J44285	TATGAATGAGTGCAGCGGAAATGAGTTTCCATCAAACAACAGAGCTAAGTGGTGCAGCACAGCAGCGCCCTTCGCCCATCTTTTACAGTAGTG
Cell cycle	cdc20	XM_002180546.1	Phatr3_J12783	TTACGACGGAATGCTTTGAAGAAAATCAGAACAATTCATGGGCACACAGCCGAATATCGTCTGTTGGATGGAACCAACTGTTGAGTTCAGCGCA
Photoreceptor	Cpd2	XM_002179343.1	Phatr3_J54342	TCTGTTGCGCGCAATACGCGGACAGTGTTCGAGCTCTGTAATAATGCTTCTGAGATTGGCGGAGTGCAGCTATTGGAATCGCGAGATGACACCT
Photoreceptor	Cpd3	XM_002180035.1	Phatr3_J51952	CTCTGATTTTTCTCGATACCGGAATACCGTCAATGGATGGAATCCAGCCGTAACATATCTGGAGGAAGCAAGGTCGCCGTTTATCAGTTGATGCC
Photoreceptor	Cpf1	XM_002180059.1	Phatr3_J27429	GGATGCTGCCAAAATCCGAATCACTCTCCGATTACGTAGTCGACCTGAATTCCTTCGCGCAACTGCTGGTGGCGCGCGGTACAAATTCGT
Photoreceptor	Cpf4	XM_002184521.1	Phatr3_J55091	CTCGATCCCTTGTTCGCGCGCAACGAGACTGCATTTCCGTAATGGGTCTACCTAATGATTTTGTAGACTCCATTGTCGAGGCAGCGTTTGAAGCAGCTG
Photoreceptor	CrypP-like	XM_002178853.1	Phatr3_J34592	CACCTTATGGTGGCGAAACAGCGGACTCGCAGATGTTTGAAGGATTACTTCGAAACGCGCAACGGGATGCTGGGGCCGAACTTCGACCAATTC
TF	CSF2	XM_002185522.1	Phatr3_J41601	GAACAGAAAGAAAGATTTCGGGACGCTTATGCATTCACAGACGAAATCGACAGGCTCATCAACTATTGTCAATTTGGGATGGAGCTTCGAGACT
Cell cycle	CYCB1	XM_002180361.1	Phatr3_J46095	TTAGAGAAAATCAATCGCAATCTCGAGCTCCGAGCGGTGAATAAAAGTACAGCGGTCATCGATACGCGGAGTTGCTTCGACCTTCTGTTATTGA
Cell cycle	CYCH1	XM_002185709.1	Phatr3_J36892	TTGCGATGGCTCGCTTTGACGTTTCGTTTCCCATCCGACCCGAGGAGTTGCCGCAATCGTCAACCATCTCGTGATTGCTCACCTCAATGGTTTGGC
Photoreceptor	Dph	XM_002179026.1	Phatr3_J54330	CAACAATCCATTTACAACAAGAGCTGACGGAATGTGATCTGAGCCTGTGCATTTGATCGAAACGACAGGAGGTTACCGCCATTTGTTGTTCAAT
Cell cycle	dsCYC2	XM_002179247.1	Phatr3_J34956	GATATTGCGTTTACCGAACCGGAAATGACTGCTCTTCTTGGCACTTGCATCTTCAACCCGATTTGGCTTTTACGAATGACTGGTCTCTGCTGG
TF	E2F_Dp2b	XM_002181452.1	Phatr3_EG02016	TGATTTGAATAGAGCTGTGAGGAAATGAGGGTACAAAAGCGCCGATTTATGACATTACCAACGTTTTTGAAGGATTGGGCTGATTACCAAGGATAGC
Cell cycle	E2F1	XM_002176842.1	Phatr3_J43065	GTGTGGTCTCATCGAAAACGATCCAAAACACAGTCCGTTGGAAAGGAGTGAAGTCTTCTTTCGCTGCTCTTTTCAAGTCCGCGCAACGAGAAAT
Cell cycle	FtsZ	XM_002185088.1	Phatr3_J42361	GAAAAGGAGAAATGAAGCAACGATTCGCGGATGAGAGGTTGCGAGAGAATGGGATACGGTATCGTAGTTCACACGACAGACTGCTCGAGATTAT
Metabolism	GapC1	XM_002182255.1	Phatr3_J21222	TGAAGGATTCCTCGGATACCTCGACGAACCGTTGGTCTCCACCGACTTTGAAGGTGACTTGGCTCTCCATCTTTGATGCGGATCGCGTATCATGCT

Metabolism	Gsat	XM_002180931.1	Phatr3_J36347	ATTTCGTTGGTTCGCGCACACACGGTTCGGGAAAAGTTCATCAAGTTCGAAGGATGCTACCAACGGACACGCCGATTCATTTTGGTCCAAGCCGGATCCG
TF	Hox1	XM_002184935.1	Phatr3_EG02213	AACAAACATCGGTTCACATACTCGCTCATACCCGGACAAGTCCCGACAGTCGAAGATGTTACAGTCTTACCAATGTTCCCGCGGAACGCATTCGCGAAC
TF	HSF1.3b	XM_002179894.1	Phatr3_J35419	CTCCAGGGAACAACCTCGATACCAGAGTTCCTCTATCAGCTAACAAAGATGCTGACGGATAACAACCGAGATATTATGAATGGACAATGGCAAGATTGA
TF	HSF1a	XM_002177362.1	Phatr3_J43051	CCACCACAGTCGGATAGCTCCGACTGGGGTTCCTCAATCCAATCATCCAAGCCAGCTGTTGATCTCTTAACGAGGAGTATGCCCAATGGCAAAATATAC
TF	HSF1b	XM_002181395.1	Phatr3_J47181	TTGTAGTGGACACTTCGATGCTCGAAGAGGTCATCAACGAAACGACGCCACTCGTTGGAGACAAAGGTAGCTTCCCTCTTGACAATAGAGCACCAGT
TF	HSF1d	XM_002178673.1	Phatr3_J44750	GTCACCATTCAGCATCGGATCGGGCTCTTACGCCATTTCTCCGAAACCTCACCGAGAGACTTGACTCTGAGACACAACGAAATTTCCCACTTGTATG
TF	HSF1g	XM_002177016.1	Phatr3_J42514	ATAGAGAATCTGCAAGCACTAGGACCATGACGACTACTTATCGAAGCGTAGCATTACAGAGAGCGAAGATAGTTCCCTGTGACACTCAGCATCAGGAG
TF	HSF2.1a	XM_002181681.1	Phatr3_J47360	AGCCCTTGGCCCTGGAAATTTGGGCTCGTTGTCGCCCTCAAGTCAACCAACCCAGCCGCTGGATGATCAACACGCCCTGTCATCGCTCAAGCGTTAGTT
TF	HSF2.2c	XM_002185405.1	Phatr3_J50481	GGACTGGAATCTAAGCCGACATATTTCCGCGAGCCCTTCTAGCGATTTTGGAGCAATCATCCCTATCTGACATTTACTTGGCTTCCCATGAGCAG
TF	HSF3.2b	XM_002186124.1	Phatr3_J44099	TTGACTCTTCGCGACAGCAAGCACTTAGCGGTTTCAACAGTCCCTCGGTGATCCGAGAGTCAAGTCTATGAGCTTTTCCACCACAACCTCTC
TF	HSF3.2e	XM_002179404.1	Phatr3_J45206	TTGGAACGGGAACGCTTCCCTCAGATTAGCAATCCGCAACCTTCGCTACTCTCTCAAGTTCGACGATGTTAAATCTCGGACTTAGCAAGACTACTT
TF	HSF3.3a	XM_002179738.1	Phatr3_J45393	TCTGGCAATCTTTCGATGGAAGTTTGGCGACCGGAGTAGTCAGGAGTGGCGTAACCAAGGAGTTTGGCGCTCGAGACCAAGCGCTGTTA
TF	HSF3.3f	XM_002179735.1	Phatr3_J45389	CATGCCAATTGAACACCAATTCAGCCATACATGGAAGAGCTTAAGAGGGTCCATCTTGACGCGCGCGGGAAGTCTCGAAAATGCGCTTGAATTTTG
TF	HSF4.2c	XM_002182762.1	Phatr3_J48361	CGCAGTATATCAGCATTTCCGCAAGTCCGCGCCGACCGGATCCCATGATGATCGTCCGAAAAAGACTGGAGAGTCACTCAGCATTTCGCGAAA
TF	HSF4.2j	XM_002177548.1	Phatr3_J43363	GTCGCGGACGGCCCTTTATCATTCAAGCCTGCAAGTCTTTAAGGATATAGTCCGCTATATTTTCCGCACTCGCGCTGAGTTTCAACGCGC
TF	HSF4.3a	XM_002184368.1	Phatr3_J49594	CGCGAACACTCGGTTCCAAATGTCGCGGAAGAACCTTCAAGCGAACCTTTGTTTCCAGCATATTTACCATGACCATATGAACGACCTGAGCAGGTGATG
TF	HSF4.3b	XM_002183012.1	Phatr3_J48558	CTGCTTTGAGCCTATCGCAGCATGGCGCCGAATTTCTACGCTATGCCAATCTACTGAGGTGGAGACAAGACCAATTTATCTCGATCCAAAATACGTGG
TF	HSF4.3c	XM_002183081.1	Phatr3_J48667	AAACTGTGAGTACTGCTTGACCATGTAAGACCGCAGGATGGAGCAGCTTATTTCTTGGCCAGTCAAGCGCTGTTTATGACATCACAAACAGACC
TF	HSF4.4b	XM_002184371.1	Phatr3_J55070	CCCATTCGCGTAAACCCGCTTCCAACTTGGCGTATCCAGAGCCGCTAGTTTGGTAAGTTCGAACAGGCAATTCGCTGCTTATGCGGTTGTATGA
TF	HSF4.6a	XM_002184347.1	Phatr3_J49557	CTCAACACTGGCGATTGGCGTATCGCTCGTTCGCGCATCCAGTGCAGCCAACTCGTGAATCTTTCGCACTTTCGCTCTGTCTCCAAATACGGGCAC
TF	HSF4.7a	XM_002185211.1	Phatr3_EG00092	CGGACATCCCGGGCGGATTTGAGATCGACGCGATAGTGTCACTTTCGCTTCCAAAGATTTTCGCACTTACATTTATCGTGGTACGGTACTCAC
TF	HSF4.7b	XM_002184277.1	Phatr3_J49596	CTCAGGTTGGTGGGAACTCTTTTCCACTCAAGCTGCAGAAATGCTTCAAATGTCGTACAAGAGCGCTACGCTCAGATTTGATCGTGGCAGCTCA
Metabolism	Lhx1	XM_002179724.1	Phatr3_J27278	AATCCTTGAGAACTTCCAGGTTAAAGAGTGCATCCATCCAAAACCTGGAGTATGGCTGGTACATGCACAGAGTTAAACCAATCCCTTTTCAAGCGA
TF	Myb1R_SHAQKYF4	XM_002178425.1	Phatr3_J44331	CGTATCTGCGGAATCGTCTCTGGTCAAATCAAGTCCCAACCGCCAAAAGGCTCAAGCGCATCGATCAAGCGCAACAGCTTCCCGACGATCAGGGA
TF	Myb1R_SHAQKYF5	XM_002181623.1	Phatr3_J47256	ACGCTTCACTGGCCGAAGACTTACACCGGATTTTGTATCGGCAATTTTGTATGTCGGCTTGAACAGCTGTCACCGTCCACATTTTGGAAACCATG
TF	Myb1R_SHAQKYF6	XM_002180835.1	Phatr3_J46535	TAACGACGAAGCGCAAACTCTTACGCAAACTTCCGCTCCCGCAAAAAGGTAAGACACGAAAGTTCGCCCTTGGTACAGCGTAATACTTCAAGTTTC
TF	Myb1R10	XM_002181361.1	Phatr3_J37257	CATATCCCTCTTCGGGGCCCTCACTTCCCTGTACAAGTGAATCTTTGGGGTGAAGCGAATCGTAGAAAAGAACGTGCTAAGGACGAGACCTCGGTTCT
TF	Myb1R4	XM_002181410.1	Phatr3_J47205	AAAGTTCGTTGACAGTTTACAGGAAAGCTTACCGCATGTGCAATGGGAACAATCCGGCAGTAGGATTCACCATATGCGATTGAACTTCTGGAAGCCG
TF	Myb2R3a	XM_002181911.1	Phatr3_J47726	TCCATCTTTCTTTGAGCGACTGTGACACCTTGGAGGGAGAGTATTTTGTACTGTGAAGTTGTGCTACTCCGGCACTCGCAGGTAATACTACTC
TF	Myb2R3b	XM_002182076.1	Phatr3_EG02275	GCCCAACAAGCTTAAAGTATTCGGGATTCATTCACAACAGTCAATGGCTAATGGCTCCATCGTTCATGTTGACATTTGCGATCGGCCAACACT
TF	Myb3R1	XM_002182738.1	Phatr3_J15016	GTATGATCTCGAGTGTCTTTCAGTTCGGAAACCGCTGGGCGAAATGCAAGCGCTTACCGGGAAGACTGATAACGCTATTTAAAATCACTGGAA
TF	Myb3R2	XM_002180449.1	Phatr3_J6839	CGAATTCCTATAGGACACAGGACACAATGGAAATCGATGGGCTGAGATTGCCAAGCGCTGCCAGCAGTACGGCAATGCGATAAAGAACTCGTGGGA
TF	Myb3R4	XM_002180741.1	Phatr3_J36337	GACAGATGCCGAAGACGCAATTTAATGATGCGGTCAATTCAGCTCGGAGCAACCTTACTCGTTGGTCACTTGGCGCAACGCTTACCGGACGCT
TF	Myb3R5	XM_002184658.1	Phatr3_J7959	GACGCTGTATCTGAATTCGTTGAGCCAAACAATCCGAAAAGACGAATGGACGGAGCAAGAGAAGAAGCTTTTGTGATGACACGCGCGCATGGGAAA
TF	Myb5R	XM_002185245.1	Phatr3_J50365	TCAACACACATAAACGCTCGTTTCCAGCTTTGGATGTTAGTGAACGCGTTGGTTTTCATCGAGAACAATGACCAATGAAGCAGACGAAGATGT
Cell cycle	PCNA	XM_002182225.1	Phatr3_J29196	AATGCGGAACCGGTCGAACCTCACTTTGCCCTCGCTACCTCAACTTTTTTACCAAGCGACTCCGTTGAGTGACACGCTATTTATCTCCATGGCACCG
Metabolism	Pds1	XM_002180135.1	Phatr3_J35509	AGCATGGCCAAAGCCTCGATTTCAATGATCCGCAAAATGAGTATGACGGTCTTTTGGCCCATGATCGCTTTTGAACGAAGCAACGGCCCTCC
Metabolism	petA	GI:118411009	#N/A	TCTCTGGATCAGGACCAATAACAGGGAATAAATTAATCTTGTAGTCTTCTTCTGCAATTTGGTCCAATCTACTAAGATATTATCAAAATCACTACTATACGG
Metabolism	Pgr5	XM_002178672.1	Phatr3_J44748	GTAACGTTCCGCTTCTGATCGTACCCTGCGCGTTCGCGTGGCGGGGGAGCCACCGGGGTTCTTCTTTTGTGCTTTCCCGACGAACCCAGCATCG
Metabolism	PgrL	XM_002177043.1	Phatr3_J42543	CCACAACAAGTACGCTTCCAAATTTAGACAAGACTACGAACAACAGTGCAGTCAACTCTGTGATTTGGCTTTTGGTGTCTCGGCCGTGACGAT
Metabolism	Por1	XM_002179653.1	Phatr3_J12155	TCCGGTGGTGAAGATTTTGAAGATCAGCAGTCCGACGCACTCGGGATCTTCTACCAGCAAGAAATGTGAAACTGAGTAGGGAAGCAGCTGGTCTTT
Metabolism	Por2	XM_002180956.1	Phatr3_J13001	AATCACCACATATTCGTTGAAAAGCGCCCTGGTTCGCTAAATACTTTCCATCTTTATGAAATTTTATACGGAGGATACGTTGAGACGACGAAGCCGG
Metabolism	Por3	XM_002177432.1	Phatr3_J43164	CGCGTCTCAACAGGAAAAGCGGTGAGACTCATGATCGTATCCGACAGTAGTCTGATCGGAGCAACCGTACGCTGGGAGCAGTACAGCTTCCC
Metabolism	psaA	GI:118410964	#N/A	GCTAAGAAGAACGCCCACTAGTTCCAAATCCACCCAAATAAATGAGCTAAACCACTGCACGACCTGAGTAATGCTTAATGCACGAGTTGGATTG
Metabolism	psaD	GI:118410999	#N/A	TTAAATGACTGATTTGGTTCCCAATGCTAAACTGCTCTTTCGCGCTAAGTATAAATTTTCCACTACGATGATGCTGCCCAATG

Metabolism	psbC	HQ912250.1	#N/A	GGCCCAACAGTCTCTGAGGCTTCAAGCACAAGCATTACGCTTTTGTAGTTCGAGATCAACGTTTAGTGTCTAATGTTTCATCAGCACAGGTCACCACTG
Metabolism	Psy1	XM_002178740.1	Phatr3_EG02349	ACCTGTCGTCGGGAATTGCGATTGGAGCGTTTGTGGCAATACGGACAAGTACAGGATGTCCTTGACCTGTGCTGTAGATCTCGAATCCAGTATCC
Reference Gene	Rps	XM_002178225.1	Phatr3_J10847	CTCCGCGCATTTTTGCCCGGTTCCCAATTTGACCGGACAGTTGCCCGACGAAGATCTCATTGGAACAACCTTGACGCTTAAATTCCTCGAAGTCAACCAGG
TF	Sigma70.1a	XM_002182291.1	Phatr3_J14599	GACAATTAGGGTATCTCGGGATCGGATTTCGCTTGTGCGAAACCCGAGCACTGAACAAATTCGCGACCCGCAAAAATTAACAGCTGCAGTCTTACGT
TF	Sigma70.1b	XM_002182124.1	Phatr3_J3388	TTCTTGTCTATTGGAAGAAACGCGCAAACTAGGGATGTCCTCGGGATCGGATTTCGCTTGTGCGAAGCCGAGCACTGAATAAATTCGACGCTCCACAA
TF	Sigma70.2	XM_002178682.1	Phatr3_J5537	ACGTGGATTGCGATTTCGCCACCTACGCCACTTACTGGGTACCAATTCGCTACGCTTCTGCTTCCAAACCGCATCCACCGGATGTTTGGGGTACCGGTA
TF	Sigma70.3	XM_002185461.1	Phatr3_J17029	TATGATGGTCTCTGATCTAAGAAGCAAAATTCGCTCGCTCGCTTCGCGCAAGGCTGCCCTGACGGAATCAAATATTCGACTCGTGGTATCCATTGCGAAA
TF	Sigma70.4	XM_002177232.1	Phatr3_J9312	AACCCGAAAGATCTCGTTCAGGAAAGTACAGATTGTCTAGGGGACACAATCTCCGCCAGTAGTGTCTTTTTGACGAAAGCACACCCGAAAACGAGTCG
TF	Sigma70.5	XM_002182854.1	Phatr3_J14908	AAGAATCATACAGGAAGGAGTTGGGATGTTGCGCGCGCGCAATGTTTGTATCCCGAGCGAGGACTGCGATTACGACTTACCGGTAGTGTGGAT
TF	Sigma70.6	XM_002178042.1	Phatr3_J9855	CTGGAAAAATCAATATGGAAGCTATTTCGGCAGACCATTGAAGAAGACTGGAAGCCAAAAATCAGTAGTCACTTCGAATCTGCGGATGGTACAGAGTGT
TF	Sigma70.7	XM_002185047.1	Phatr3_J50183	AATGATGACGCTCTCTCGCATACGGTGAACCTCCGCTGATCAAGCAAATCGAAGCTGATCTTGTGCTTTCGATATCCATCTCCCGGAAACCCACTACT
TF	Taz2	XM_002178935.1	Phatr3_J44807	CCTAATGTAGAGACTGTAGTGGTAAAAGCAAGCGCGACGCTCGACTATTGGACTTTACTTTCATGCACAGAATTCGCTTTTGGGGACAGATGCCAC
Reference Gene	Tbp	XM_002186285.1	Phatr3_J10199	CTGATTTTTCCAGCGTAAAATGTGCGTGACCGGAGTCAAGAGCACACACAACGCCAATCTGGCGCGAAAAAGTTTGCTACATTGTGGAGCGCGCTCG
Metabolism	Vde	XM_002178607.1	Phatr3_J51703	GGCGATGACGGATCATATCCCGTCCCGCACCCGAAATCGCTACCCAAAGTTTGACACCAAGTTCTTCGATGGAGCACTTACATCTCAGCCGGACAAA
Metabolism	Vdl1	XM_002180599.1	Phatr3_J36048	AAGGCTCTCTTTAGCAATCCGCGTGAATCAAAGGTGTCTCTCTGTTGGGACGCTGCAAGGGCGAGCAATCGTGTGCGACGCGGTTTTCGCCAATTCG
Metabolism	Vdl2	XM_002180015.1	Phatr3_J45846	ACACACGCGCTTCGGCAATGAATATCCAATTTCCGGAGCCAGACGAATCACTCCATGTCTGGGACCAGTACGAAACATCGGTAAAGTGTGTACCGGTT
Metabolism	Vdr	XM_002177477.1	Phatr3_J43240	GTCGTTGCGGTGCCAACGAAGCCTATGATAAGTTTCTTCGCGAGAACCACTTCTTATCCCGCTGCTCGCGGTGAGATTGTGGTACGATCCCGTTT
Metabolism	Zds	XM_002184417.1	Phatr3_J30514	CGCGCTCAACGAATCAATCGCGACACCCGCAACGACTCTGCGCGTTTAAATTCACACTTCTTCGCGCGGATGTTTGGACACGCGGACGCTCAGGAATC
Metabolism	Zep1	XM_002179586.1	Phatr3_J45485	TTCCCAATGAAGTTTCTACCAGGTGTCACTGGCACTGTCTGATCGCGTGGTATCGACCAGCACTTCTTACACCCGCTCCAGTCTTTGGCGTGC
Metabolism	Zep3	XM_002178331.1	Phatr3_J10970	CCAAGGTTTTCACCAAGATGCCACCATTGGATGTTACCGTACTGGAACAAACGTCGAATTCAAAACGGTTCGGTGGACCCATCCAGCTCGTAGTAACGC

823

824

825 **Table S2:** Calculated phases, amplitudes and periods of the *Wt* and *PtbH1H1a* OE lines growing
 826 in 16L:8D. (P., Period; Ph., phase; Amp., amplitude; Std, standard deviation; t-test, t-test P-
 827 value between *Wt* and the indicated line).

Data	N	Period	P.Std	P. t-test	Phase	Ph.Std	Ph. t-test	Amp.	Amp.Std	Amp. t-test
<i>Wt</i>	8	23.7	0.32		13.02	0.54		8.30E-01	1.20E-01	
<i>OE-1</i>	5	23.37	0.51	0.168894675	15.93	1.83	0.001213017	5.60E-01	2.60E-01	0.02447126
<i>OE-2</i>	6	23.38	0.48	0.159824505	15.19	1.52	0.002642208	6.10E-01	7.00E-02	0.001606661
<i>OE-3</i>	4	23.89	0.22	0.323356278	13.94	0.51	0.017556508	8.40E-01	1.20E-01	0.841165049

828

829

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

SPECIES	ABBREVIATION	ACCESSION	ACCESSION (Alt)
<i>Amphiprora</i> sp., Strain CCMP467	Amph	CAMPEP_0168730192	
<i>Asterionellopsis glacialis</i> , Strain CCMP134	Agla2	CAMPEP_0195290872	
<i>Asterionellopsis glacialis</i> , Strain CCMP1581	Agla3	CAMPEP_0197142484	
<i>Astrosyne radiata</i> , 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 dichaeata, 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 Fracyl 204663	
Fragylariopsis cylindrus	FcbHLH4	jgi Fracyl 241747	
Fragylariopsis cylindrus	FcbHLH7	jgi Fracyl 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	hsbMall1	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 tricornerutum	PtbHLH1a	jgi Phatr2 44962	Phatr3_J44962
Phaeodactylum tricornerutum	PtbHLH1b	jgi Phatr2 44963	Phatr3_J44963
Phaeodactylum tricornerutum	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 multiseriata	PmbHLH1a	jgi Psemu1 26622	
Pseudo-nitzschia multiseriata	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	

831

832

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

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

834

835

836

837

838

839

840

841

842 **References**

- 843 1. Zielinski T, Moore AM, Troup E, Halliday KJ, & Millar AJ (2014) Strengths and limitations
844 of period estimation methods for circadian data. *PLoS One* 9(5):e96462.
- 845 2. Agier N & Fischer G (2016) A Versatile Procedure to Generate Genome-Wide Spatiotemporal
846 Program of Replication in Yeast Species. *Methods Mol Biol* 1361:247-264 .
- 847 3. Geiss GK, et al. (2008) Direct multiplexed measurement of gene expression with color-coded
848 probe pairs. *Nature biotechnology* 26(3):317-325.
- 849 4. Chauton MS, Winge P, Brembu T, Vadstein O, & Bones AM (2013) Gene regulation of
850 carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated
851 to light/dark cycles. *Plant physiology* 161(2):1034-1048.
- 852 5. Saeed AI, et al. (2003) TM4: a free, open-source system for microarray data management and
853 analysis. *Biotechniques* 34(2):374-378.
- 854 6. Siaux M, et al. (2007) Molecular toolbox for studying diatom biology in *Phaeodactylum*
855 *tricornutum*. *Gene* 406(1-2):23-35.
- 856 7. Keeling PJ, et al. (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project
857 (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through
858 Transcriptome Sequencing. *Plos Biol* 12(6).
- 859 8. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7:
860 improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.
- 861 9. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis
862 Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7):1870-1874.
- 863 10. Darriba D, Taboada GL, Doallo R, & Posada D (2011) ProtTest 3: fast selection of best-fit
864 models of protein evolution. *Bioinformatics* 27(8):1164-1165.
- 865 11. Miller MA, Pfeiffer, W., and Schwartz, T. (2010) Creating the CIPRES Science Gateway for
866 Inference of Large Phylogenetic Trees. In Proceedings of the Gateway Computing
867 Environments Workshop (GCE) (New Orleans, LA):1-8.
- 868