Circadian protein regulation in the green lineage I. A phospho-dawn of protein modification anticipates light onset in the picoeukaryote *O. tauri*

Running title: Algal phospho- and protein rhythms

Zeenat B. Noordally[1,2]*, Matthew M. Hindle[1]*, Sarah F. Martin[1,3], Daniel D. Seaton[1,4], T. Ian Simpson[5], Thierry Le Bihan[1], Andrew J. Millar[1]****

[1]SynthSys and School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK. [5]Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK.
* These authors contributed equally to this work.
** corresponding authors: tlebihan@gmail.com; andrew.millar@ed.ac.uk +44 131 651 3325

[2] Present address: Norfolk County Council, Community and Environmental Services, County Hall, Martineau Lane, Norwich NR1 2DH, United Kingdom.
[3] Present address: Office of the Chief Statistician and Strategic Analysis, Scottish Government, Edinburgh EH1 3DG, UK
[4] Present address: GlaxoSmithKline, Stevenage SG1 2NY, UK

Author; Email, ORCID:
Zeenat Noordally; zeenat.noordallyed@gmail.com, 0000-0003-2817-1330
Matthew Hindle; matthew.hindle@gmail.com, 0000-0002-6870-4069
Sarah F. Martin; sarahfriedemartin@gmail.com, -
Daniel Seaton; daniel.d.seaton@gmail.com, 0000-0002-5222-3893
Ian Simpson; Ian.Simpson@ed.ac.uk, 0000-0003-0495-7187
Thierry Lebihan; tlebihan@gmail.com, 0000-0003-0498-8063
Andrew Millar; andrew.millar@ed.ac.uk, 0000-0003-1756-3654
URL https://www.ed.ac.uk/biology/centre-engineering-biology

Submitted 13 December 2022; Main text: ~5000 words, excluding Methods 2730 words; 5 figures; 11 Supplementary Figures; 6 Supplementary Tables.

Table of Contents

Highlight (<30 words)

The phosphorylation of most protein sites was rhythmic under light-dark cycles, and suggested circadian control by particular kinases. Day-peaking, rhythmic proteins likely reflect light-stimulated protein synthesis in this microalga.

Abstract

Diel regulation of protein levels and protein modification had been less studied than transcript rhythms. Here, we compare transcriptome data under light-dark cycles to partial proteome and phosphoproteome data, assayed using shotgun mass-spectrometry, from the alga *Ostreococcus tauri*, the smallest free-living eukaryote. 10% of quantified proteins but two-thirds of phosphoproteins were rhythmic. Mathematical modelling showed that light-stimulated protein synthesis can account for the observed clustering of protein peaks in the daytime. Prompted by night-peaking and apparently dark-stable proteins, we also tested cultures under prolonged darkness, where the proteome changed less than under the diel cycle. The dark-stable, prasinophyte-specific proteins were also reported to accumulate when *O. tauri* formed lipid droplets. In the phosphoproteome, 39% of rhythmic phospho-sites reached peak levels just before dawn. This anticipatory phosphorylation suggests that a clock-regulated phospho-dawn prepares green cells for daytime functions. Acid-directed and proline-directed protein phosphorylation sites were regulated in antiphase, implicating the clock-related, casein kinases 1 and 2 in phase-specific regulation, alternating with the CMGC protein kinase family. Understanding the dynamic phosphoprotein network should be facilitated by the minimal kinome and proteome of *O. tauri*. The data are available from ProteomeXchange, with identifiers PXD001734, PXD001735 and PXD002909. This submission updates a previous version, posted on bioRxiv on 4th April 2018, as https://www.biorxiv.org/content/10.1101/287862v1

Keywords and Abbreviations

Keywords: Systems biology; light signalling; proteomics; phosphoproteomics; photoperiod; marine microalgae; photosynthetic pico-eukaryotes

Abbreviations: PM, phosphopeptide motif; LD, light-dark cycles; ZT, Zeitgeber Time; DA, dark adaptation; PC, principal component; CK1, casein kinase 1; CK2, casein kinase 2; GSK3, Glycogen Synthase Kinase 3; CMGC, Cyclin-dependent kinase, Mitogen-activated protein kinase, Glycogen synthase kinase, CDC-like kinase; CCA1, Circadian Clock Associated 1 protein.

Introduction

Responses to light are critical for organisms of the green lineage (Noordally and Millar, 2015; Paajanen *et al.*, 2021). The rapid effects of photosynthetic light harvesting, for example on redox state and sugar metabolism, are complemented by signalling photoreceptors (Whitelam and Halliday, 2007) and the slower, 24-hour regulation by the biological clock (Millar, 2016; Creux and Harmer, 2019). Circadian regulation allows organisms to anticipate the predictable, day-night transitions of the diel cycle, complementing the responses to faster changes in light levels (Troein *et al.*, 2011). Mehta et al. (2021) refer to these as 'anticipatory' and 'reactive' regulation. At the macromolecular level, the transcriptomes in the green lineage show widespread and overlapping regulation of mRNA abundance by both light and circadian signals (see below), whereas the diel regulation of proteins and their post-translational modifications had been less studied (Mehta *et al.*, 2021). We addressed that gap using a minimal biological system, focussing on protein phosphorylation.

Phosphorylation of an existing protein is energetically inexpensive, occurs rapidly and can then alter protein activity through conformational change or intermolecular recognition (Khoury *et al.*, 2011). These characteristics seem fitted to reactive regulation. Some plant photoreceptor proteins include protein kinases that initiate light signalling (Christie, 2007; Djouani-Tahri el *et al.*, 2011*a*).

Protein synthesis is not only far slower but also among the costliest macromolecular processes (Scott *et al.*, 2010; Karr *et al.*, 2012), seemingly more suited to anticipatory regulation. Rhythmic regulation might then provide a selective advantage, loosely summarised as making proteins when they are needed in the diel cycle (Laloum and Robinson-Rechavi, 2022). That reasoning helped to interpret the co-regulation of functional clusters of RNAs, when transcriptome studies demonstrated that over 50% of Arabidopsis RNAs can be rhythmic under diel, light-dark cycles (LD) (Smith *et al.*, 2004; Blasing *et al.*, 2005; Michael *et al.*, 2008). Most strikingly, almost the whole transcriptome of the marine unicellular alga *Ostreococcus tauri* was rhythmic in controlled conditions (Monnier *et al.*, 2010) and this was also the most-rhythmic taxon among the diverse plankton of a Pacific timeseries (Kolody *et al.*, 2019). The clock might also allow anticipation, to ensure that the proteins had been fully synthesised and assembled to their active state by the appropriate time.

Proteomic data, in contrast, revealed that most detected proteins had stable levels, with an average half-life >6 days in the model plant *Arabidopsis thaliana* (Li *et al.*, 2017), suggesting little scope for diel rhythmicity. Timeseries under constant light or a diel cycle found up to 6% of rhythmic proteins (Baerenfaller *et al.*, 2012, 2015)(Choudhary *et al.*, 2016; Uhrig *et al.*, 2021; Krahmer *et al.*, 2022). The most short-lived, regulatory proteins are harder to detect, but such proteins seem to be exceptions to the general protein stability, consistent with mammalian systems (Doherty *et al.*, 2009). Global regulation of protein synthesis is also clearly relevant in plants and algae (Piques *et al.*, 2009; Juntawong and Bailey-Serres, 2012; Pal *et al.*, 2013; Missra *et al.*, 2015; Ishihara *et al.*, 2015). In this context, circadian RNA regulation was proposed to offer a selective advantage through seasonal adaptation to day-length on a timescale of weeks (Seaton *et al.*, 2018).

More protein phosphorylation sites change over the diel cycle, compared to protein levels (Kusakina and Dodd, 2012; Mehta *et al.*, 2021). Protein phosphorylation in plants and algae is most directly light-regulated by the photoreceptor kinases (Christie, 2007; Djouani-Tahri el *et al.*, 2011*a*), though light also affects the broader phosphoproteome (Turkina *et al.*, 2006;

Boex-Fontvieille *et al.*, 2014; Schönberg *et al.*, 2017), for example affecting 25% of Arabidopsis phosphopeptides within 30 minutes (Uhrig *et al.*, 2021). Circadian studies in Arabidopsis under constant light found up to 23% rhythmic phosphopeptides (Choudhary *et al.*, 2015; Krahmer *et al.*, 2022). These studies suggest that light responses and the circadian clock in Arabidopsis each control five- to ten-fold more phosphopeptides than the diel rhythm of total protein level, so it is also important to understand which phospho-regulators mediate these effects.

The amino acid sequences of rhythmically-regulated phosphosites have implicated a range of protein kinases with overlapping contributions in Arabidopsis (Choudhary *et al.*, 2015; Uhrig *et al.*, 2021; Krahmer *et al.*, 2022). However, ~1000 protein kinases shape the phosphoproteome in Arabidopsis (Champion *et al.*, 2004) including several in plastids (Baginsky and Gruissem, 2009), compared to half that number in the human genome (Manning *et al.*, 2002). Of particular interest, the casein kinases (CK1, CK2) and Glycogen Synthase Kinase 3 (GSK3), affect the circadian timing of all organisms suitably studied (Mehra *et al.*, 2009). These kinases have central positions in the yeast kinase-target network (Breitkreutz *et al.*, 2010) and are highly conserved (Hindle *et al.*, 2014), in contrast to photoreceptor proteins or circadian transcription factors (Noordally and Millar, 2015; Dunlap and Loros, 2017).

Here, we compare the prevalence of proteomic and phosphoproteomic regulation under LD cycles, using *O. tauri* as a minimal model for the green lineage (Noordally and Millar, 2015). This alga not only has a ubiquitously-rhythmic transcriptome, but its genome is also reduced to 13Mbp (Blanc-Mathieu *et al.*, 2014), likely due to selection pressure to reduce cell size to 1-2µm (Courties *et al.*, 1994). Its 7699 protein-coding genes include just 133 protein kinases that represent the core families for eukaryotic signalling (Hindle *et al.*, 2014) and a minimal set of Arabidopsis clock gene homologues (Corellou *et al.*, 2009; Djouani-Tahri el *et al.*, 2011*b*; Troein *et al.*, 2011; Ocone *et al.*, 2013). CK1 and CK2 modulate circadian timing in the light, with widespread effects on the algal phosphoproteome (Le Bihan *et al.*, 2011, 2015; van Ooijen *et al.*, 2013). A non-transcriptional, 24-hour oscillator of unknown mechanism was also revealed under prolonged darkness, when transcription stops in this organism (O'Neill *et al.*, 2011; van Ooijen *et al.*, 2011; Edgar *et al.*, 2012; Bouget *et al.*, 2014; Feeney *et al.*, 2016). In cyanobacteria, the non-transcriptional clock is driven by rhythmic protein phosphorylation, so rhythmic protein kinase activities could also be relevant in *O. tauri* (van Ooijen and Millar, 2012; Wong and O'Neill, 2018).

Our results reveal widespread daily rhythms in both the proteome and phosphoproteome in *O. tauri*, including expected features such as the diel control of conserved, cell cycle phospho-regulators. Rather than rapid phosphorylation responses and slow, rhythmic anticipation in protein profiles, however, the level of many rhythmic proteins appears light-responsive, whereas much of the rhythmic phosphoproteome anticipates dawn. The phosphosite sequences strongly implicate phase-specific protein kinase classes. Moreover, we identify a set of rhythmic, algal-specific proteins that accumulate in prolonged darkness and were also identified in conditions that promote the formation of lipid droplets.

Materials and Methods (2730 words)

Materials

Chemicals were purchased from Sigma-Aldrich (now a subsidiary of Merck Life Science UK Ltd, Dorset, UK) unless otherwise stated. Main solvent, acetonitrile and water for liquid chromatography– dual mass spectrometry (LC-MSMS) and sample preparation were HPLC quality (Thermo Fisher Scientific, Loughborough, UK). Formic acid was Suprapure 98-100% (Merck) and trifluoroacetic acid (TFA) was 99% purity sequencing grade. Porcine trypsin TPCK treated was from Worthington (Lorne Laboratories, Reading, UK). All HPLC-MS connectors and fittings were from Upchurch Scientific (Hichrom, Theale, UK) or Valco (RESTEK, High Wycombe, UK). % are expressed in v/v.

O. tauri media and culturing

Ostreococcus tauri OTTH95 were cultured as previously described (van Ooijen *et al.*, 2012), supplemented with 0.22 µm filtered 50 µg ml^{-1} ampicillin, neomycin and kanamycin antibiotics in vented tissue culture flasks (Sarstedt, Leicester, UK). Cultures were maintained by splitting weekly at 1:50 dilution. In preparation for proteomics experiments, cultures were grown in growth media supplemented with 200 mM sorbitol and 0.4% glycerol prior to harvesting (O'Neill *et al.*, 2011). Cells were cultured under cycles of 12 hour light/ 12 hour dark (LD) at 20°C in a controlled environment chamber (MLR-350, Sanyo Gallenkamp, Loughborough, UK) at a light intensity of 17.5 μEm^{-2} s^{-1} white fluorescent light filtered by 724 Ocean Blue filter (LEE Filters Worldwide, Andover, UK).

O. tauri cell harvesting

Cells were grown for 7 days in LD and on the seventh day harvested, with five replications, at Zeitgeber Time (ZT) 0, 4, 8, 12, 16 and 20, where ZT0 corresponds to dawn. At ZT0 cells were harvested a few minutes before the lights went on and at ZT12, before the lights went off. 135 ml culture was harvested by centrifugation (4000 rpm, 10 min, 4°C) per sample replicate, each from a separate culture vessel. Pellets were resuspended in ice cold phosphate buffered saline solution (PBS). Cultures were centrifuged as before, pellets were air dried and then vortex-mixed in 250 µl 8M urea and stored at -80°C. For total cell lysate, cells were dissolved by sonication (Branson Ultrasonics) and diluted with 500 µl dH_2O.
Cells were grown for 7 days in LD and on the eighth day the Dark Adaptation (DA) experiment cell harvests were performed at ZT24, 48, 72 and 96 in constant darkness with five replications. The samples were harvested and prepared as for the LD experiment.

Protein digestion

Samples were analysed by Bradford Assay (Bio-Rad, Watford, UK) and 400 µg protein of each sample was used in the digestion. Samples were reduced in 10 mM dithiothreitol and 50 mM ammonium bicarbonate, and alkylated with 25 mM iodoacetamide. Samples were digested overnight with 10 µg (1:40 ratio) trypsin under agitation at room temperature at pH8 in a total volume of 1 ml. Samples were cleaned on SPE BondElut 25 mg columns (Agilent Technologies, Stockport, UK) following the vendor instruction. 50 µl (~20 µg) was removed and dried for LC-MS (Speedvac, Thermo Fisher Scientific). The remaining ~380 µg were also dried in preparation for phosphopeptide enrichment, and stored at -20°C.

Phosphopeptide enrichment

Dried peptide samples (~380 µg) were sonicated in 50 µl solution 0 (2.5% acetonitrile, 0.5% TFA) and 100 µl solution 2 (80% acetonitrile, 0.5% TFA, 100% lactic acid). Titansphere Phos-TiO Kit spin tip-columns (GL Sciences, Tokyo, Japan) were washed with 40 µl solution 1 (80% acetonitrile, 0.5% TFA). Samples were loaded on the spin tip-columns and passaged three times through a centrifuge; 5 min at 200 xg, 15 min incubation at room temperature and 10 min at 200 xg. Spin tip-columns were subsequently washed once with solution 1, twice with solution 2 and twice with solution 1for 2 min at 200x g. Phosphopeptides were eluted in two steps, first with 50 µl 5% ammonium hydroxide (5 min at 200 xg) and secondly, with 5% pyrrolidine solution. 20 µl 20% formic acid was added to lower the pH and samples were cleaned on Bond Elut OMIX C18 pipette tips (Agilent Technologies) following the manufacturer's instruction.

Protein and phosphoprotein quantification

15 µg protein from total *O. tauri* cell lysates were run on a Novex NuPAGE 4-12% Bis-Tris by SDS-PAGE with PeppermintStick Phosphoprotein Molecular Weight Standards and Spectra Multicolor Broad Range Protein Ladder (Thermo Fisher Scientific). The gel was fixed overnight (50% methanol, 40% ddH_2O, 10% glacial acetic acid), washed in ddH_2O and stained with Pro-Q Diamond Phosphoprotein Gel Stain (Invitrogen, now Theremo Fisher Scientific, Loughborough, UK) in the dark at 25°C following manufacturer's instructions. The gel was imaged on a Typhoon TRIO variable mode imager (GE Healthcare, Amersham, UK) at 532 nm excitation/ 580 nm emission, 450 PMT and 50 micron resolution. Images were processed using ImageQuant TL software (GE Healthcare, Amersham, UK). The gel was re-used for protein quantification using SYPRO Ruby Protein Gel Stain (Themo Fisher Scientific, Loughborough, UK) following manufacturer's instructions and imaged using a UV transilluminator (Ultra-Violet Products Ltd, Cambridge UK). Protein and phosphoprotein bands were quantified using Image Studio Lite v 4.0 (LI-COR Biosciences, Cambridge, UK).

Protein per cell quantification

Cells were grown (as described above) and independent, triplicate cultures were harvested at the times indicated. Cultures were monitored using spectrophotometry at 600nm. Total protein was quantified using the Quick Start Bradford Assay following manufacturer instructions (Bio-Rad, Watford, UK). Cell number was estimated either by counting four fields of view per culture in a haemocytometer after trypan blue staining (Abcam protocols, Cambridge, UK), or by fluorescence-activated cell sorting (FACS). For FACS, a 1/200 dilution of cells were transferred to fresh media containing 1X SYBR Green I Nucleic Acid Gel Stain (Invitrogen, now Theremo Fisher Scientific, Loughborough,UK) and FACS-counted (FACScan, BD Bioscience, Wokingham, UK) at a flow rate of 60µl per minute.

qPCR for transcriptional regulation during dark adaptation (DA)

Cells were cultured and harvested in the same experimental regime (described above) and harvested in biological triplicate at the times indicated for the LD and DA experiments. Total RNA was extracted from frozen cells using an RNeasy Plant Mini Kit and DNase treated (QIAGEN, Manchester, UK). First-strand cDNA was synthesised using 1 µg RNA and 500 ng μl^{-1} $Oligo(dT)_{15}$ primer (Promega, Southampton, UK), denatured at 65°C for 5 min, and reverse transcribed using SuperScript II (Invitrogen, now Theremo Fisher Scientific, Loughborough, UK) at 42 °C for 50 min and 70 °C for 10 min. 1/100 cDNA dilutions were analysed using a LightCycler®480 and LightCycler®480 SYBR Green I Master (Roche,

Welwyn Garden City, UK) following manufacturer's instructions and cycling conditions of pre-incubation 95°C for 5 min; 45x amplification cycles of 95°C for 10 s, 60°C for 10 s, 72°C for 10 s. The following 5' to 3' forward (F) and reverse (R) primers to *O. tauri* gene loci were used: ostta01g01560 GTTGCCATCAACGGTTTCGG (F), GATTGGTTCACGCACACGAC (R); ostta03g00220 AAGGCTGGTTTGGCACAGAT (F), GCGCTTGCTCGACGTTAAC (R); ostta03g04500 GCCGCGGAAGATTCTTTCAAG (F), TCATCCGCCGTGATGTTGTG (R); ostta04g02740 ATCACCTGAACGATCGTGCG (F), CCGACTTACCCTCCTTAAGCG (R); ostta10g02780 GGCGTTCTTGGAATCTCTCGT (F), TATCGTCGATGATCCCGCCC (R); ostta10g03200 GGTACGGAGGAAGAAGTGGC (F), ATGTCCATGAGCTTCGGCAA (R); ostta14g00065 GACAGCCGGTGGATCAGAAG (F), TCGAGGTAGCTCGGGAGATC (R); ostta16g01620 ACGGGTTGCAGCTCATCTAC (F), CCGCTTGGGTCCAGTACTTC (R); ostta18g01250 CTTGCAAATGTCCACGACGG (F), ATGATGTGGCACGTCTCACC (R); OtCpg00010 ACATGACTCACGCGCCTTTA (F), TGCCAAAGGTGCCCTACAAA (R). Primers to eukaryotic translation elongation/initiation factor (EF1a) ostta04g05410 GACGCGACGGTGGATCAA (F) and CGACTGCCATCGTTTTACC (R) were used as an endogenous control. Data were combined for biological and two technical replicates and relative quantification performed using LightCycler®480 1.5 software (Roche).

HPLC–MS analysis

Micro-HPLC-MS/MS analyses were performed using an on-line system consisting of a micro-pump 1200 binary HPLC system (Agilent Technologies) coupled to an hybrid LTQ-Orbitrap XL instrument (Thermo Fisher Scientific). The complete method has been described previously (Le Bihan *et al.*, 2010). For all measurements, 8µl of sample was injected using a micro-WPS auto sampler (Agilent Technologies) at 5µl /min. After sample loading, the flow rate across the column was reduced to approximately 100-200 nl/min using a vented column arrangement. Samples were analysed on a 140 min gradient for data dependant analysis.

HPLC-MS data analysis

To generate files compatible with public access databases PRIDE (Vizcaino *et al.*, 2016) and the former pep2pro (Hirsch-Hoffmann *et al.*, 2012), Mascot Generic Format (MGF) input files were generated using MSConvert from ProteoWizard (Kessner *et al.*, 2008). MSMS data was searched using MASCOT version 2.4 (Matrix Science Ltd, London, UK) against the *O. tauri* subset of the NCBI protein database (10114 sequences from NCBI version 2014 June 6th including common contaminants) using a maximum missed-cut value of 2, variable oxidation (M), N-terminal protein acetylation, phosphorylation (STY) and fixed carbamidomethylation (C); precursor mass tolerance was 7 ppm and MSMS tolerance 0.4 amu. The significance threshold (p) was set below 0.05 (MudPIT scoring). Global FDR was evaluated using decoy database search and removal of peptides ranked higher than 1 for a mascot score above 20 (~1% global FDR). Mass spectrometry proteomics data have been deposited in PRIDE ProteomeXchange Consortium (Vizcaino *et al.*, 2014) via the PRIDE partner repository with the dataset identifier LD global proteomics, PXD001735; LD phosphoproteomics, PXD001734; DA global proteomics, PXD002909. Data was converted into PRIDEXML using Pride converter 2.0.20 and submitted using proteome exchange tool pxsubmission tool 2.0.1. The LC-MS data were also publicly available in the former pep2pro database (Assemblies 'Ostreococcus tauri Light:dark cycle,LD global', 'Ostreococcus tauri Light:dark cycle,LD phospho', and 'Ostreococcus tauri dark adaptation,DA global'). Label-free quantification was performed using Progenesis version 4.1 (Nonlinear Dynamics, Newcastle, UK). Only MS peaks with a charge of 2+, 3+ or 4+ and the five most intense

spectra within each feature were included in the analysis. Peptide abundances were mean-normalised and ArcSinH transformed to generate normal datasets. Within-group means were calculated to determine fold changes. Neutral losses of phosphoric acid typical of serine and threonine phosphorylated were validated manually in all significantly differential phosphopeptides. Ambiguous sites were confirmed by cross-referencing (by sequence, charge, and quantity of residue modifications) with most probable site predictions from MaxQuant version 1.0.13.8 (Cox and Mann, 2008) in singlet mode, Mascot settings as above. Where multiple occurrences of residue phosphorylation events were quantified, abundances were summed, collating all charge states, missed cuts and further modifications.

Data analysis

Merging

For accurate and unique phosphopeptide quantification we addressed variant redundancy at different charge states, alternative modifications (e.g. oxidation and acetylation) and multiple sites of protease digestion. All unique phosphorylation events were retained, including multiple phosphorylation, at a given amino acid motif, while summing the quantification of these technical variants. The qpMerge (http://sourceforge.net/projects/ppmerge/) software was used to combine Progenesis and MaxQuant phospho-site predictions and produce a unique set of quantified phosphopeptide motifs (Hindle *et al.*, 2016).

Outlier identification and removal

To detect outliers we applied principal component analysis (PCA) of replicates and comparing each replicates r^2 to the respective median abundance at that ZT. A single replicate, 4E, was excluded based on extreme differences in peptide quantification as viewed in the wider distribution of the ratios to the sample median, which was confirmed with a Pearson's correlation against a defined criterion of a sample median of < 0.8 (Supplementary Figures S1).

P-value calculation and false discovery rate (FDR)

For analysing the significance of changing protein and peptide abundance over time, non-linear response of expression using polynomial regression was modelled using the R Stats Package. A third order polynomial was fitted, allowing for an expected peak and trough within a 24 h daily cycle. An arcsinh transformation of abundance was applied to meet the required assumption of normality (Burbidge *et al.*, 1988). FDR was calculated using the Benjamini and Hochberg (BH) method (Benjamini and Hochberg, 1995). More than 2 quantifying peptides were required to report protein abundance.

Equivalence testing

Using the R equivalence package, the statistical equivalence of mean abundance across time was tested as the highest *p*-value from exhaustive pairwise Two one-sided test approach (TOST) tests over all ZTs (Schuirmann, 1981; Westlake, 1981). We tested whether abundances had upper and lower differences of less than 0.3 within the equivalence margin (ε).

O. tauri gene identifiers

O. tauri genome version 1 gene IDs (Derelle *et al.*, 2006) for microarray data were converted to version 2 IDs (Blanc-Mathieu *et al.*, 2014) by finding exact sequence matches for the microarray probes (Accession GPL8644) (Monnier *et al.*, 2010) in the version 2 FASTA coding sequence file.

Principle component analysis (PCA)

PCA was used to investigate the main components of variation in the data using prcomp from the R Stats Package. The abundances were zero-centred per-feature. The PCA values for each feature were extracted and then used for Gene Ontology (GO) enrichment analysis.

Clustering

Hierarchical clustering was performed with hclust from the R Stats Package and applied on per-feature (protein or phosphopeptide motif) mean abundances over time, which were zero-centred and scaled. Pearson's correlation was used to calculate distance matrix and the Ward method (Ward, 1963) for linkage criteria. The hierarchical tree was divided into clusters using the dynamicTreeCut algorithm (Langfelder *et al.*, 2008). The hybrid cut tree method with a cut height of 100 and a minimum cluster size of 20 was used for both datasets. Clusters are displayed with 95% (black lines) and 99% (orange lines) confidence via multiscale bootstrap resampling (AU determined p-value).

Enrichment analysis for GO terms

TopGO was used to evaluate the enrichment of GO terms, for each ontology aspect, within clusters, peaks, troughs, and principal components. For clusters, peaks and troughs a Fisher's exact test was used by partitioning at 95% confidence on FDR corrected *p*-values, and with a fold change >1.5 in normalised abundance. For each test, we use a relevant background of non-significant observed features. To test for enrichment of GO terms for each PCA the Kolmogorov-Smirnov test was applied over the absolute PCA values for each gene. GO terms were predicted by InterProScan 5 (Jones *et al.*, 2014) on amino acids sequences for *O. tauri* coding sequences (NCBI version 140606 (Blanc-Mathieu *et al.*, 2014)).

Homology modelling

Structural homology models were generated using I-TASSER (Yang and Zhang, 2015) for prasinophyte-family specific proteins of unknown structure and function, including for ostta02g03680 compared to the human Bar-domain protein structure in PDB entry with DOI 10.2210/pdb2d4c/pdb. Other suggested homologies were more limited.

pLOGO and binomial statistics

Significantly over- and under-represented amino acid residues at different time-points were calculated using the binomial based pLogo tool (O'Shea *et al.*, 2013). The Motif-X tool (Chou and Schwartz, 2011) was used to discover novel motifs in the dataset. Binomial statistics were applied to calculate the enrichment of motifs and the combined probabilities of amino acids with similar properties in a phospho-motif (*e.g.* the acidic D/E positions in the CK2 motif).

Kinase target prediction

Computational prediction of protein kinase motifs associated with the identified phosphorylation sites was performed using Group-based Prediction System, GPS Version 3.0 (http://gps.biocuckoo.org/index.php) (Xue *et al.*, 2011).

***O. tauri* loci IDs mapping to *A. thaliana* loci IDs**

O. tauri and *A. thaliana* IDs were mapped using EggNOG4.1 (http://eggnogdb.embl.de). *O. tauri* proteins were downloaded from https://bioinformatics.psb.ugent.be/gdb/ostreococcusV2/LATEST/OsttaV2_PROT_20140522.fasta.gz (May 22nd, 2014). Viridiplantae (virNOG) hmms and their descriptions and annotations were transferred to *O. tauri* proteins using hmmr 3.1 (http://hmmer.janelia.org)

Mathematical simulations

Simulated protein rhythms

Protein dynamics (*P(t)*) were simulated according to the following model:

$$\frac{dP(t)}{dt} = \Big(\big(k_{syn} - 1\big)L + 1\Big)\, m(t) - k_{deg}P(t)$$

Where *L(t)* = 1 during the day (ZT <=12), and 0 otherwise. The rate of protein degradation (k_{deg}) was set to 0.1 h^{-1}, and the ratio of protein synthesis in the light compared to the dark (k_{syn}) was set to 4, based on (Martin *et al.*, 2012). The rhythmically expressed mRNA levels (*m(t)*) are given by:

$$m(t) = \cos\left(\frac{2\pi(t-\varphi)}{24}\right) + 1$$

The peak phase of expression is given by φ. To obtain the distributions of peak and trough protein levels, the peak phases (φ) of mRNA expression were uniformly distributed at 0.1 h intervals across the range [0,24]. For each phase of mRNA expression, the timing of peak and trough protein levels was determined by simulating the model dynamics in MATLAB using the ode15s ODE solver. The peaks and troughs were identified across a 24 h period, following 240 h simulation to allow the dynamics to reach a steady behaviour (i.e. with the same protein levels at ZT0 and ZT24).

Protein degradation rates and depletion during dark adaptation

Degradation rates were calculated from published proteomics data (Martin *et al.*, 2012), which characterised the dynamics of partial ^{15}N isotope incorporation. We assumed a labelling efficiency of 0.93 (=maximum labelled fraction achieved of any protein + 0.01), and fitted a simple kinetic model assuming: (1) constant labelling efficiency over time; (2) different proteins are labelled at the same efficiency; (3) heavy and light fractions are turned over at equal rates, similar to (Seaton *et al.*, 2018). One protein with a high degradation rate ~0.03 h^{-1} was excluded as an outlier, which increased the correlation from R = -0.48 to -0.7 when included.

Results

To understand the landscape of protein abundance and phosphorylation across the diel cycle, we harvested quintuplicate biological samples of *O. tauri* at six timepoints across a 12 h light/12 h dark (LD) cycle. Dawn samples (zeitgeber time 0, ZT0) were harvested just before lights-on, and samples at ZT12 before lights-off, to detect biological regulation that anticipated these transitions. The proteome and phosphoproteome were measured in whole-cell extracts from each sample, by label-free, liquid chromatography–mass spectrometry (Figure 1A). After removing a technical outlier (Supplementary Fig. S1), 855 proteins were quantified with 2 or more peptides (Supplementary Table S1). Phosphopeptides were enriched by metal-affinity chromatography prior to detection. For quantification, we combined the phosphopeptide species that shared phosphorylation on a particular amino acid, irrespective of other modifications (Hindle *et al.*, 2016). We refer to this set of phosphorylated species as a phosphopeptide motif (PM). 1472 phosphopeptide motifs were quantified, from 860 proteins (Supplementary Table 2). Serine and threonine residues were modified most; only 1% of PMs included phospho-tyrosine. The quantified proteins and phosphoproteins each represent ~11% of the total *O. tauri* proteome (Figure 1B). 29 out of 61 proteins encoded on the chloroplast genome (Robbens *et al.*, 2007) were quantified, with 6 PMs. 3 out of 43 mitochondrial-encoded proteins were quantified with no PMs, consistent with other studies (Ito *et al.*, 2009).

Diel rhythmicity of the transcriptome, proteome and phosphoproteome

To compare the patterns and prevalence of daily rhythms at different regulatory levels, we re-analysed published transcriptome data in parallel with these protein and phosphoprotein data, summarised in Figure 1C. Gene expression in *O. tauri* was strongly rhythmic under LD cycles, with 89% of transcripts scored rhythmic, as previously reported (Monnier *et al.*, 2010). 85 (9.5%) of the detected proteins were significantly rhythmic and changed by at least 1.5-fold, with only 11 of these proteins changing level by more than 5-fold. In contrast, 66%

of phosphoproteins or 58% of PMs (570 of 860 proteins; 850 of 1472 PMs) were rhythmic by these criteria and the levels of 35 PMs changed more than 20-fold. The overlap among all three datasets included only 110 genes. The most common pairwise overlaps involved genes with changing levels of RNA and/or PMs but not of protein (Figure 1C).

Protein levels nonetheless changed smoothly, with distinct waveforms. Of the twenty most highly-detected proteins, likely including the most abundant, 11 were significantly rhythmic but with low amplitudes (Supplementary Figure S2A), such that only ostta10g03200 exceeded the 1.5-fold change threshold (Table S1). 15 of the twenty most highly-detected PMs, in contrast, were rhythmic by both criteria (Supplementary Figure S2B). The more stringent, "equivalence" test revealed 49 proteins with significantly non-changing protein abundance but with significantly changing transcript and PMs, illustrated by the 10-fold change in PM abundance on the non-changing, chlorophyll-binding protein CP26, amongst others (Supplementary Figure S3).

To address our major question on the dominant patterns of regulation, we used undirected, principal component (PC) analysis (Fig. 1D-1I). Clustering (Fig. 1D-1I, Supplementary Figure S4) and analysis of peak distributions (Fig. 2A-C) informed more detailed hypotheses on upstream regulation and downstream, functional effects. The PC analysis represented most (83-86%) of the variance in the data sets but indicated a differing balance of molecular regulation (Supplementary Table S3). The transcriptome and phosphoproteome data clearly separated between dawn and dusk timepoints (in PC1), whereas the light and dark intervals were separated by the secondary PC2. This mapped the 13 transcriptome and 6 phosphoproteome timepoints into their respective, temporal sequences. Gene Ontology (GO) terms relating to translation, ribosome biogenesis and RNA processing were enriched among dawn-expressed RNAs, and mitotic processes (DNA replication and repair) among dusk-expressed transcripts (Supplementary Table S3), similar to past analysis (Monnier *et al.*, 2010). The functions of rhythmic phosphoproteins are discussed in more detail, below.

The regulation was strikingly reversed in the proteome (Fig. 1C, 1D), where the major separation (in PC1) was between samples from light and dark intervals. The early day (ZT4), when translation and chlorophyll biosynthesis GO terms were enriched, was separated from all other timepoints, most strongly from mid-night (ZT16 and 20). There was less separation (in PC2) of the late night (ZT0), when proteins involved in the TCA cycle and transport processes are prominent, from the late day (ZT8-12), when translation and chlorophyll biosynthesis were still enriched. In contrast, the ZT0 timepoint stood out in the phosphoproteome (Fig. 1E, 1F), when PMs enriched for transcription, glucose metabolism, K^+ and protein transport and ubiquitin-dependent proteolysis functions were clearly separated from the late-day timepoints (ZT8, ZT12). PC2 separated mid-day timepoints (ZT4, ZT8, with enrichment for regulation of gene expression, translation and transmembrane transport) from mid-night (ZT16, ZT20, when mitosis and Ca^{2+} transmembrane transport terms were enriched).

The even distribution of changing RNAs across all the transcriptomic timepoints was not reflected either in the proteome or the phosphoproteome data, where the early day (ZT4) or pre-dawn (ZT0) timepoints, respectively, stood out in the PC analysis. Hierarchical clustering grouped the protein and PM abundance profiles into 8 clusters (termed P1–P8 and PM1–PM8, respectively; Supplementary Figure S2C, S2D), which are coloured on the PC plots in Fig. 1F-1I. GO term enrichment for RNAs, proteins and PMs in the principal component, clustering and peak time analyses is presented in Supplementary Tables S3-S5, with a

summary for proteins and PMs in Supplementary Figure S5. We analysed the distribution of individual peak times (Figure 3) to understand these patterns, starting with the proteins.

Daytime peaks of protein abundance

Hundreds of transcripts reach peak abundance at every timepoint around the day/night cycle (Fig. 3A) (Monnier *et al.*, 2010). In contrast, most protein profiles peaked in the light interval (85% at ZT4-12; Fig. 2B), separating the day and night samples in line with the PC analysis. Metabolic labelling of *O. tauri* has shown ~5-fold higher protein synthesis rates in the day compared to the night (Martin *et al.*, 2012). Consistent with this, our analyses showed translation-related proteins were enriched among the rhythmic proteins with high abundance in the daytime, in PC1, protein cluster P1 and in profiles with daytime peak phase (Supplementary Tables S3-S5, Supplementary Figures S4, S5). We therefore tested whether this light-regulated synthesis alone could explain the observed distribution of protein peaks.

We simulated protein dynamics (Fig. 2D-2F; Supplementary Figure S6) using measured protein synthesis and degradation rates (Martin *et al.*, 2012), and an even temporal distribution of rhythmic mRNAs. The simulated distribution of protein profiles matched well with our experimental results (Fig. 2E; Supplementary Figures S6D-S6G), with a slightly stronger daytime preference than in the data. ostta03g04520 is an example of an RNA that peaks at ZT0 and its protein profile (Fig. 2G) was very similar to the predicted protein from such an RNA (Fig. 2D). The overall distribution of protein profiles substantially reflects the light-stimulated translation rate of this organism (see Discussion).

Unusual, night-time proteins suggest a 'dark state'

An intriguing pattern of protein regulation stood out from the common, daytime abundance. Protein cluster P6 (and P8) included the rare protein profiles that fell at ZT4 (Supplementary Figure S2C), associated with oxidative metabolism and protein transport GO terms (Supplementary Table S4). Four un-annotated, prasinophyte-specific proteins in cluster P6 not only peaked at night, but were also among the 11, highest-amplitude profiles of all the rhythmic proteins (Fig. 3A). Their dramatic fall in abundance at ZT4 suggested a destabilisation by light, so we tested whether such proteins would remain stable during several days of dark-adaptation (DA).

O. tauri cells are photo-autotrophic. Their division is entrained by the LD cycle (Farinas *et al.*, 2006) and they arrest transcription in prolonged darkness, when they can survive without growth or division (O'Neill *et al.*, 2011). Cell density (optical density at 600nm) in our cultures increased by ~25% after one LD cycle. Cellular protein content was consistent (18-20 pg cell^{-1}) in replicate measures at ZT0 and ZT24 (Fig. 3B). In cultures transferred to three further days of darkness, optical density remained constant but protein content per cell dropped by over 60% on the first day (ZT24 to ZT48) and was then stable to ZT96. This result was suggestive of an altered, but potentially stable, cellular 'dark state', which we tested in a further, proteomic timeseries, sampling in darkness at ZT24, 48, 72 and 96.

The proteomic landscape changed less during dark adaptation (DA) than under a standard LD cycle. 98 of the 865 proteins quantified by LC-MS changed levels more than the average and only 64 (7%) also changed more than 1.5-fold (Supplementary Table S6). The 35 significantly-increasing proteins in DA included five transmembrane transporters, a Lon-related protease and two superoxide dismutases, suggestive of nutrient acquisition, protein mobilisation and oxidative stress responses. The four prasinophyte-specific proteins noted

above were among the ten most-increasing proteins in DA, confirming their unusual regulation and suggesting a shared function in both standard night-time and the putative 'dark state'. The most-decreasing among 63 significantly-decreasing proteins in DA was a starch synthase (ostta06g02940). Its abundance declined in the night under LD cycles, as did all 10 of the DA-decreasing proteins that were also rhythmic in LD. The largest functional group of depleted proteins comprised 22 cytosolic ribosomal proteins and translation factors (Supplementary Table S6), suggesting that *O. tauri* selectively mobilised this protein pool in darkness.

The night-abundant, prasinophyte proteins that accumulated in DA, and night-depleted proteins that fell in DA (such as ostta06g02940, above; or PPDK ostta02g04360, Supplementary Figure S7C), suggested that prolonged darkness preserved a night-like state. An alternative explanation was that protein stability in general was altered in the putative dark state. We sought to test that notion, using the protein degradation rates that were previously measured by metabolic labelling in LD conditions (Martin *et al.*, 2012). Falling protein abundance under DA was significantly correlated with higher degradation rates in LD (Fig. 3C; R = -0.48, p=0.004, n=34), even among these abundant, stable proteins. We also tested RNA abundance for a subset of these proteins in DA by qRT-PCR, showing stable levels after one day of prolonged darkness (ZT48; Supplementary Figure S8A). The lack of RNA regulation seemed consistent with the lack of transcription in these conditions (O'Neill *et al.*, 2011). For example, a further prasinophyte-specific protein ostta03g4500 with a stable RNA level and slightly-increasing protein level in DA also had among the lowest protein degradation rates in LD (Fig. 3C), and was among the most-detected proteins in these conditions (Supplementary Figures S2A, S8B). The RNA data and protein degradation rates suggested that the prasinophyte-specific proteins accumulated due to a focussed, regulatory mechanism, rather than generalised refactoring of the proteome (see Discussion).

A phospho-dawn of protein modification

In contrast to the many daytime-peaking protein profiles, 39% of the changing phosphopeptide motifs (PMs) peaked in abundance at ZT0 (Fig. 2C), double the proportion of any other timepoint. The ZT0 samples were harvested before lights-on, so this 'phospho-dawn' anticipated the transition and was not due to light-stimulated translation. Other, high-amplitude PM profiles tracked the levels of their cognate proteins, with little evidence of regulated phosphorylation (Fig. 2G). We therefore tested the contribution of protein levels to PM profiles more broadly, among the 138 genes that were quantified in both protein and PM datasets (Supplementary Figures S7A-B). This subset of 261 protein-PM pairings included proteins peaking at all timepoints, and PM profiles that reflected the peak time distribution of the full dataset. 80% of the PMs peaked at different timepoints than their cognate protein (Supplementary Figure S7C; examples in Fig. 2H). The LHC linker protein CP29 (ostta01g04940) illustrates one pattern: its protein level rises in the light while a PM is de-phosphorylated (Supplementary Figure S7C) adjacent to a target site of chloroplast kinase STN7 in Arabidopsis (Schönberg *et al.*, 2017).

To test the phospho-dawn pattern by a different method, we estimated the bulk protein phosphorylation across the diel cycle using protein gel staining (Supplementary Figures S9A-B). The proportion of phosphorylated proteins was lowest in the daytime and increased during the night to peak at ZT0 (Supplementary Figures S9C). Total phosphorylation was therefore broadly consistent with the distribution of PM profiles (Fig. 2C). Taken together, these results indicate that a regulator other than light or protein abundance controls the *O.*

tauri phosphoproteome before dawn. Below, we report phosphosite sequences that suggested its identity.

Functions of proteins with rhythmic phospho-motifs

The LD datasets confirmed that protein phosphorylation profiles often diverged from protein abundance. Colour-coding in Fig. 1H shows that clustering of the phospho-motif (PM) profiles aligned with the PC analysis more clearly than for the lower-amplitude, protein profiles (Fig. 1F). The largest cluster PM1 reflected the profiles that peaked in the ZT0 timepoint, which PC analysis also highlighted (Supporting Figure S2D). Phosphopeptide enrichment allowed the detection of many regulatory proteins, including PMs on predicted CONSTANS-like B-box transcription factors (OtCOL) related to the plant clock protein TOC1 (Fig. 4), and on the RWP-RK mating-type factor ostta02g04300 (Blanc-Mathieu *et al.*, 2017). PM1 also includes the predicted CK2 target site pS10 in the clock protein CCA1 (ostta06g02340; Fig. 4), close to the homologous location of a CK2 site in Arabidopsis CCA1 (Lu *et al.*, 2011).

PMs in cluster PM3 peaked in the light, consistent with many protein profiles (examples in Fig. 2G). PMs on the photoreceptors phototropin and LOV-HK illustrate these daytime profiles (Fig. 4). Protein functions predicted to regulate transcription, metal ion transport and protein phosphorylation are enriched in this cluster (summarised in Supplementary Figure S2F; Supplementary Table S4) and in profiles with daytime peaks (Supplementary Figure S4B; Supplementary Table S5).

In contrast, the PM2, PM4, PM7 and PM8 clusters peaked at ZT16, with or without accumulation in daytime (Supplementary Figure S2D). These clusters are enriched for PMs on protein kinases including cell-cycle-related kinases (Supplementary Figures S2F, S4B; Supplementary Tables S4 and S5). We therefore analysed the phospho-regulators that might control these profiles, including potential contributions to non-transcriptional timing.

Phase-specific target sites

We first analysed motifs of amino acids that were enriched in rhythmic PMs, compared with all quantified phosphopeptides to avoid potential detection bias due to PM abundance. PMs that peaked at ZT16 were strikingly enriched for the proline-directed motif [pS/pT]P (Fig. 5B-C). This strongly implicates the CMGC family of protein kinases, including Cyclin-Dependent Kinases (CDKs) and GSK. Consistent with this, the profiles of PMs with predicted GSK target sequences also most often peaked at ZT16 (Supplementary Figure S10A-B). Levels of *GSK3* RNA and a PM on GSK3 peaked at ZT12 (Fig. 4), though the auto-phosphorylation site pY210 was not rhythmic (Supplementary Fig. S2B; Supplementary Table S2). More specific CDK target motifs [pS/pT]PXX[K/R] were enriched at ZT12, consistent with the known timing of cell division (Farinas *et al.*, 2006; Moulager *et al.*, 2007) and the peak level of the activation phospho-site of CDKB (Fig. 4). During the day (ZT4 and 8), enrichment of hydrophobic residues at positions -5 and +4 is suggestive of the SnRK consensus (Vlad *et al.*, 2008), the plant kinase most related to animal AMPK.

In contrast, acid([D/E])-directed target motifs were significantly enriched among the many rhythmic PMs that peaked at ZT0 and the proline-directed motifs were depleted (Fig. 5C). Conversely, these acid-directed motifs were depleted on PMs peaking at ZT16 or ZT4, suggesting a strong phase-specificity. Considering the more specific, predicted target sites for the clock-related protein kinases (Supplementary Figures S10A), predicted CK1 targets were most abundant, and most often peaked at ZT0. Predicted CK2 target sequences were even

more phase-specific, with at least 5-fold more peaking at ZT0 than at other times. Thus predicted targets of the clock-related kinases CK1 and CK2 both contribute to the phospho-dawn profiles, in antiphase to the evening peaks of proline-directed phospho-sites.

Rhythmic regulation of the kinome

The protein abundance of the three detected protein kinases and two phosphatases was not rhythmic (Supplementary Table S1). We therefore analysed the 68 rhythmic PMs on protein kinases and five PMs on protein phosphatases, as candidate mediators of rhythmic phosphorylation (Figs. 5A, 5D). The PMs on kinases represent 8% of the total, though protein kinase genes comprise ~1.5% of the genome. Indeed, the most heavily-phosphorylated protein with 14 PMs was the WITH NO LYSINE (WNK) kinase that might target clock proteins in Arabidopsis (Murakami-Kojima *et al.*, 2002) (Supplementary Table S2; Supplementary Figure S10C). The most-changing PM on a predicted protein phosphatase was pT175 in ostta11g02830, related to human Dual-specificity phosphatase DUSP12 (Fig. 5D).

Among the clock-related protein kinases, we note the dusk-peaking PM of GSK3 (above). CK2 subunits were not detected in our data and the PM on CK1 was not strongly rhythmic (Fig. 4). 21 other protein kinases bore rhythmic PMs that are predicted targets of these clock-related kinases (Supplementary Figures S10C).

Around mitosis at ZT12-16, significantly peaking PMs were detected on cell cycle regulators CDKA, CDKB and WEE1 (Fig. 5D). Kinase PMs peaking at ZT4-8 included Serine-Arginine Protein Kinases (SRPKs), MAPKs, CDKA and a site on Yet Another Kinase (YAK1). PMs that peaked at ZT0, coincident with the phospho-dawn, included RIO2, YAK1 and CDPK, all implicated in cell cycle regulation and progression (Garrett *et al.*, 1991; LaRonde-LeBlanc and Wlodawer, 2005). RIO's are among the few kinase families shared with the Archaea (Kennelly, 2014), making them candidate contributors to an ancient, non-transcriptional oscillator (Edgar *et al.*, 2012).

Discussion

The diel proteome and phosphoproteome

Our results contribute to understand the 'reactive' and 'anticipatory' components of protein regulation in the green lineage under diel (LD) cycles (Mehta *et al.*, 2021). A small fraction of the *O. tauri* proteins quantified here were rhythmic (just under 10%), compared to a majority (58%) of the phosphomotifs (PMs). Most protein profiles peaked in daytime, consistent with the 'reactive' effect predicted from the light-regulated translation that was previously measured in this organism (Martin *et al.*, 2012), and with enrichment of translation-related functions among daytime-peaking proteins. This result reinforces the dangers of using RNA profiles as a proxy for biological function. It further supports our prediction that "translational coincidence" should alter the *O. tauri* proteome in different day lengths, as some rhythmic RNAs will coincide with light-stimulated translation only in long days (Seaton *et al.*, 2018).

In contrast, the largest number of PM profiles peaked in the pre-dawn, ZT0 timepoint. This anticipatory 'phospho-dawn' might be controlled by the circadian clock. Circadian regulation would be expected to persist under constant conditions, which were not tested here. Studies in Arabidopsis under constant light, however, identified a high fraction of rhythmic phosphopeptides that peaked at subjective dawn (Choudhary *et al.*, 2015)(Krahmer *et al.*, 2018), suggesting a similar, circadian-regulated phospho-dawn in higher plants. Such

phospho-regulation might prepare green cells for daytime functions and/or end night-time activities, before light-stimulated translation facilitates new protein synthesis.

Acid-directed target sites were clearly enriched at ZT0, implicating the clock-related kinases CK1 and CK2 in regulating the phospho-dawn in *O. tauri*. Enrichment of proline-directed target sites occurs in antiphase, at ZT12-16, which implicates the 19 CMGC-class kinase proteins (Hindle *et al.*, 2014) including CDKs, MAPKs and GSK3. These phase-specific enrichments were clearer than in the Arabidopsis studies, suggesting that the minimal kinase-target network of *O. tauri* might be easier to resolve in future. Comparison to the rhythmic phosphoproteome in animals is limited, because the most-rhythmic kinase Akt (also known as Protein Kinase B) in mouse liver (Robles *et al.*, 2016) is absent from the green lineage (Hindle *et al.*, 2014).

The low overall rhythmicity (<10%) in the partial proteome quantified here is consistent with similar studies in Arabidopsis, which identified 0.1-1.5% rhythmic proteins from 7-9 % of the proteome in LD, using iTRAQ labelling with similar statistical criteria to ours (Baerenfaller *et al.*, 2015; Baerenfaller *et al.*, 2012), or 4-7% rhythmic proteins from 4% of the proteome under constant light using a gel-based approach (Choudhary *et al.*, 2016). Our results provide 11% coverage in the minimal *O. tauri* proteome, with a simpler experimental protocol. Broader coverage of this proteome was reported after our preprint was released (Kay *et al.*, 2021), in experiments that included a depletion of abundant proteins, among several technical differences. Their higher reported fraction of rhythmic proteins might reflect the detection of low-abundance proteins and/or analysis with no minimum amplitude threshold.

The 'dark state' is indirectly associated with lipid synthesis

Among the rhythmic proteins reported here, some of the most highly-regulated were four prasinophyte-specific sequences (unnamed proteins ostta02g03680, ostta03g04960, ostta07g00470, ostta09g00670; Figure 3A) along with ostta03g04500 (Supplementary Figure S2A). These proteins accumulated further in prolonged darkness (Figure 3A). We previously showed that *O. tauri* stop transcription and cell division in those conditions. Cultures resume gene expression and growth upon return to LD cycles, suggesting that dark adaptation induces a state of cellular quiescence. The ecological relevance of a quiescent 'dark state' for photo-autotrophic, surface-dwelling *O. tauri* is not immediately obvious. However, *Ostreococcus* relatives can persist under the Polar Night (Joli *et al.*, 2017), and quiescent forms in other phytoplankton (Roy *et al.*, 2014) can be ecologically important in benthic-pelagic coupling (Marcus and Boero, 1998). Cells near the deep chlorophyll maximum (Cardol *et al.*, 2008) could be moved into the dark, benthic zone by turbulence, to return later *via* upwelling (Collado-Fabbri *et al.*, 2011; Countway and Caron, 2006). Understanding the laboratory 'dark state' might therefore have ecological relevance.

Protein content dropped significantly between 12h and 36h of darkness (ZT24 to ZT48) but was then stable. Proteins associated with cytosolic translation were notably depleted (Supplementary Table 6), rather than abundant, chloroplast proteins involved in photosynthesis. Photosynthetic functions might be particularly important to recover from quiescence, similar to the rapid regrowth observed after nutrient starvation (Liefer *et al.*, 2018).

A preprint coincident with our first report showed that three of these night-expressed, prasinophyte-specific proteins accumulated strongly in *O. tauri* when the growth medium was depleted of nitrogen, under LD cycles (Smallwood *et al.*, 2018*a*). The most-depleted protein in their conditions was the same starch synthase (ostta06g02940) that fell most in

abundance under our prolonged dark treatment. Nitrogen depletion is commonly used to induce lipid synthesis in algae, in the context of third-generation biofuel production (Zienkiewicz *et al.*, 2016). Prolonged darkness and/or hypoxia can also induce lipid accumulation, and hypoxia can occur in dark-adapting algal cultures due to continued respiration (Hemschemeier *et al.*, 2013). Our culture conditions included sorbitol and glycerol in the growth medium (O'Neill *et al.*, 2011), which are required for viability in prolonged darkness. Our 'dark state' proteome might therefore reflect active lipid synthesis from these substrates.

O. tauri can form both intracellular lipid droplets and extracellular droplets in membrane-bound 'pea-pod' structures (Smallwood *et al.*, 2018*b*). Lipid droplets in other algae include major proteins that are restricted to limited taxonomic groups (Zienkiewicz *et al.*, 2016). Some of these are predicted to have all-alpha-helical structure, including the Major Lipid Droplet Protein Cre09.g405500 of *Chlamydomonas reinhardtii* or the Lipid Droplet Surface Protein of the diatom *Nannochloropsis oceanica*. Protein structure homology modelling aligned ostta02g03680 with a human BAR domain dimer, an all-helical protein domain that can sense and create membrane curvature (Simunovic *et al.*, 2015)(Supplementary Figure S11), suggesting that this *O. tauri* protein might also be involved in lipid droplets. *N. oceanica* lipid synthesis and LDSP accumulation is highly rhythmic but day-phased (Poliner *et al.*, 2015). The night-expressed proteins in *O. tauri* indirectly suggest a different regulation of lipid synthesis, that could have biotechnological relevance.

Supplementary Data Summary

Supplementary Figure S1. Identification of outlier phosphopeptide replicate 4E.

Supplementary Figure S2. Most-detected protein and PM profiles.

Supplementary Figure S3. Changing PMs on non-changing proteins.

Supplementary Figure S4. Clustered protein and PM profiles with enriched functions.

Supplementary Figure S5. Phase-specific GO term enrichment.

Supplementary Figure S6. Simulation of light-regulated translation.

Supplementary Figure S7. Loci identified in both LD protein and phosphopeptide motif datasets.

Supplementary Figure S8. Regulation of proteins tested under Dark Adaptation (DA).

Supplementary Figure S9. Protein and phospho-protein abundance in LD cycle.

Supplementary Figure S10. CK1, CK2 and GSK3 kinase targets and phosphorylation sites in rhythmic kinases.

Supplementary Figure S11. Structural homology of rhythmic, prasinophyte-specific protein.

Supplementary Table S1. Proteins quantified under LD.

Supplementary Table S2. Phosphopeptide Motifs (PMs) quantified under LD.

Supplementary Table S3. GO term enrichment among RNA, proteins and PMs contributing to PCA.

Supplementary Table S4. GO term enrichment among RNA, proteins and PMs in clusters. Individually-significant, rhythmic protein profiles are considered, to provide sufficient numbers for enrichment analysis. Only BH-corrected significant PM profiles with >1.5-fold changes are considered.

Supplementary Table S5. GO term enrichment among rhythmic proteins and PMs by peak/trough times. Only BH-corrected significant protein or PM profiles with >1.5-fold changes are considered.

Supplementary Table S6. Proteins quantified under DA.

Acknowledgements

We are very grateful to K. Kis, L. Imrie and D. Kelly for expert technical help, to B. Kolody and A. Dodd for helpful discussion, to M. Hirsch-Hoffmann and K. Baerenfaller for support on pep2pro, and to C.R. Smallwood, J.E. Evans and colleagues for clarifying the comparison to their work. The proteomics analyses were carried out by the EdinOmics research facility (RRID: SCR_021838) at the University of Edinburgh. Funded by the Biotechnology and Biological Sciences Research Council (UKRI-BBSRC award BB/J009423/1). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author Contributions

Z.B.N. and M.M.H. contributed equally to this study. Investigation and formal analysis, Z.B.N., S.F.M. and T.L.B; formal analysis (bioinformatics), M.M.H.; formal analysis (mathematical modelling), D.D.S.; conceptualisation and methodology, A.J.M., T.I.S., T.L.B., S.F.M., M.M.H. and Z.B.N.; funding acquisition and supervision, A.J.M., T.I.S. and T.L.B.; writing, Z.B.N., M.M.H., T.L.B. and A.J.M.

Conflicts of Interest

The authors declare no competing financial interests.

Data Availability

The OTTH95 strain is available from the CCAP (www.ccap.ac.uk) and RCC (roscoff-culture-collection.org) stock centres. Mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers: LD global proteomics, PXD001735; LD phosphoproteomics, PXD001734; DA global proteomics, PXD002909. The LC-MS data were also previously available in pep2pro at www.pep2pro.ethz.ch (Assemblies 'Ostreococcus tauri Light:dark cycle,LD global', 'Ostreococcus tauri Light:dark cycle,LD phospho', 'Ostreococcus tauri dark adaptation,DA global'). Processed data lists are provided in the Supplementary Information.

854

855 REFERENCES

856

857 **Baerenfaller K, Massonnet C, Hennig L, Russenberger D, Sulpice R, Walsh S, Stitt M, Granier C,**
858 **Gruissem W**. 2015. A long photoperiod relaxes energy management in Arabidopsis leaf six. Current
859 Plant Biology **2**, 34–45.

860 **Baerenfaller K, Massonnet C, Walsh S, *et al.*** 2012. Systems-based analysis of Arabidopsis leaf
861 growth reveals adaptation to water deficit. Molecular Systems Biology **8**, 606.

862 **Baginsky S, Gruissem W**. 2009. The chloroplast kinase network: new insights from large-scale
863 phosphoproteome profiling. Molecular Plant **2**, 1141–53.

864 **Benjamini Y, Hochberg Y**. 1995. Controlling the False Discovery Rate: A Practical and Powerful
865 Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) **57**,
866 289–300.

867 **Blanc-Mathieu R, Verhelst B, Derelle E, *et al.*** 2014. An improved genome of the model marine alga
868 Ostreococcus tauri unfolds by assessing Illumina de novo assemblies. BMC Genomics **15**, 1103.

869 **Blasing OE, Gibon Y, Gunther M, Hohne M, Morcuende R, Osuna D, Thimm O, Usadel B, Scheible**
870 **WR, Stitt M**. 2005. Sugars and circadian regulation make major contributions to the global regulation
871 of diurnal gene expression in Arabidopsis. The Plant Cell **17**, 3257–81.

872 **Boex-Fontvieille E, Davanture M, Jossier M, Zivy M, Hodges M, Tcherkez G**. 2014. Photosynthetic
873 activity influences cellulose biosynthesis and phosphorylation of proteins involved therein in
874 Arabidopsis leaves. Journal of Experimental Botany **65**, 4997–5010.

875 **Bouget FY, Lefranc M, Thommen Q, Pfeuty B, Lozano JC, Schatt P, Botebol H, Verge V**. 2014.
876 Transcriptional versus non-transcriptional clocks: a case study in Ostreococcus. Mar Genomics **14**,
877 17–22.

878 **Breitkreutz A, Choi H, Sharom JR, *et al.*** 2010. A global protein kinase and phosphatase interaction
879 network in yeast. Science **328**, 1043–6.

880 **Champion A, Kreis M, Mockaitis K, Picaud A, Henry Y**. 2004. Arabidopsis kinome: after the casting.
881 Funct Integr Genomics **4**, 163–87.

882 **Chou MF, Schwartz D**. 2011. Biological sequence motif discovery using motif-x. Curr Protoc
883 Bioinformatics **Chapter 13**, Unit 13 15-24.

884 **Choudhary MK, Nomura Y, Shi H, Nakagami H, Somers DE**. 2016. Circadian Profiling of the
885 Arabidopsis Proteome Using 2D-DIGE. Frontiers in Plant Science **7**, 1007.

886 **Choudhary MK, Nomura Y, Wang L, Nakagami H, Somers DE**. 2015. Quantitative Circadian
887 Phosphoproteomic Analysis of Arabidopsis Reveals Extensive Clock Control of Key Components in
888 Physiological, Metabolic, and Signaling Pathways*. Molecular & Cellular Proteomics **14**, 2243–2260.

889 **Christie JM**. 2007. Phototropin blue-light receptors. Annual Review of Plant Biology **58**, 21–45.

890 **Corellou F, Schwartz C, Motta J-P, Djouani-Tahri EB, Sanchez F, Bouget F-Y**. 2009. Clocks in the
891 Green Lineage: Comparative Functional Analysis of the Circadian Architecture of the Picoeukaryote
892 Ostreococcus. The Plant Cell **21**, 3436–3449.

893 **Courties C, Vaquer A, Troussellier M, Lautier J, Chrétiennot-Dinet MJ, Neveux J, Machado C,**
894 **Claustre H**. 1994. Smallest eukaryotic organism. Nature **370**, 255–255.

895 **Cox J, Mann M**. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-
896 range mass accuracies and proteome-wide protein quantification. Nat Biotechnol **26**, 1367–72.

897 **Creux N, Harmer S**. 2019. Circadian Rhythms in Plants. Cold Spring Harbor Perspectives in Biology **11**,
898 a034611.

899 **Derelle E, Ferraz C, Rombauts S, *et al.*** 2006. Genome analysis of the smallest free-living eukaryote
900 Ostreococcus tauri unveils many unique features. Proceedings of the National Academy of Sciences
901 of the U S A **103**, 11647–52.

902 **Djouani-Tahri el B, Christie JM, Sanchez-Ferandin S, Sanchez F, Bouget FY, Corellou F**. 2011*a*. A
903 eukaryotic LOV-histidine kinase with circadian clock function in the picoalga Ostreococcus. Plant
904 Journal **65**, 578–88.

905 **Djouani-Tahri el B, Sanchez F, Lozano JC, Bouget FY**. 2011*b*. A phosphate-regulated promoter for
906 fine-tuned and reversible overexpression in Ostreococcus: application to circadian clock functional
907 analysis. PLoS ONE **6**, e28471.

908 **Doherty MK, Hammond DE, Clague MJ, Gaskell SJ, Beynon RJ**. 2009. Turnover of the Human
909 Proteome: Determination of Protein Intracellular Stability by Dynamic SILAC. Journal of Proteome
910 Research **8**, 104–112.

911 **Dunlap JC, Loros JJ**. 2017. Making Time: Conservation of Biological Clocks from Fungi to Animals.
912 Microbiol Spectr **5**.

913 **Edgar RS, Green EW, Zhao Y, *et al.*** 2012. Peroxiredoxins are conserved markers of circadian
914 rhythms. Nature **485**, 459–64.

915 **Farinas B, Mary C, Manes CLD, Bhaud Y, Peaucellier G, Moreau H**. 2006. Natural synchronisation for
916 the study of cell division in the green unicellular alga Ostreococcus tauri. Plant Molecular Biology **60**,
917 277–292.

918 **Feeney KA, Hansen LL, Putker M, Olivares-Yañez C, Day J, Eades LJ, Larrondo LF, Hoyle NP, O'Neill**
919 **JS, van Ooijen G**. 2016. Daily magnesium fluxes regulate cellular timekeeping and energy balance.
920 Nature **532**, 375–379.

921 **Garrett S, Menold MM, Broach JR**. 1991. The Saccharomyces cerevisiae YAK1 gene encodes a
922 protein kinase that is induced by arrest early in the cell cycle. Molecular and Cellular Biology **11**,
923 4045–4052.

924 **Hemschemeier A, Casero D, Liu B, Benning C, Pellegrini M, Happe T, Merchant SS**. 2013. COPPER
925 RESPONSE REGULATOR1–Dependent and –Independent Responses of the Chlamydomonas
926 reinhardtii Transcriptome to Dark Anoxia. The Plant Cell **25**, 3186–3211.

927 **Hindle MM, Le Bihan T, Krahmer J, Martin SF, Noordally ZB, Simpson TI, Millar AJ**. 2016. qpMerge:
928 Merging different peptide isoforms using a motif centric strategy.

929 **Hindle MM, Martin SF, Noordally ZB, van Ooijen G, Barrios-Llerena ME, Simpson TI, Le Bihan T,**
930 **Millar AJ**. 2014. The reduced kinome of Ostreococcus tauri: core eukaryotic signalling components in
931 a tractable model species. BMC Genomics **15**, 640.

932 **Hirsch-Hoffmann M, Gruissem W, Baerenfaller K**. 2012. pep2pro: the high-throughput proteomics
933 data processing, analysis, and visualization tool. Frontiers in Plant Science **3**, 123.

934 **Ishihara H, Obata T, Sulpice R, Fernie AR, Stitt M**. 2015. Quantifying protein synthesis and
935 degradation in Arabidopsis by dynamic 13CO2 labeling and analysis of enrichment in individual
936 amino acids in their free pools and in protein. Plant Physiology **168**, 74–93.

937 **Ito J, Taylor NL, Castleden I, Weckwerth W, Millar AH, Heazlewood JL**. 2009. A survey of the
938 Arabidopsis thaliana mitochondrial phosphoproteome. Proteomics **9**, 4229–40.

939 **Jones P, Binns D, Chang HY, *et al.*** 2014. InterProScan 5: genome-scale protein function
940 classification. Bioinformatics **30**, 1236–1240.

941 **Juntawong P, Bailey-Serres J**. 2012. Dynamic Light Regulation of Translation Status in Arabidopsis
942 thaliana. Frontiers in Plant Science **3**, 66.

943 **Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI,**
944 **Covert MW**. 2012. A whole-cell computational model predicts phenotype from genotype. Cell **150**,
945 389–401.

946 **Kay H, Grunewald E, Feord HK, Gil S, Peak-Chew SY, Stangherlin A, O'Neill JS, van Ooijen G**. 2021.
947 Deep-coverage spatiotemporal proteome of the picoeukaryote Ostreococcus tauri reveals
948 differential effects of environmental and endogenous 24-hour rhythms. Commun Biol **4**, 1147.

949 **Kennelly PJ**. 2014. Protein Ser/Thr/Tyr phosphorylation in the Archaea. Journal of Biological
950 Chemistry **289**, 9480–7.

951 **Kessner D, Chambers M, Burke R, Agus D, Mallick P**. 2008. ProteoWizard: open source software for
952 rapid proteomics tools development. Bioinformatics **24**, 2534–6.

953 **Khoury GA, Baliban RC, Floudas CA**. 2011. Proteome-wide post-translational modification statistics:
954 frequency analysis and curation of the swiss-prot database. Scientific reports **1**, 90.

955 **Kolody BC, McCrow JP, Allen LZ, *et al.*** 2019. Diel transcriptional response of a California Current
956 plankton microbiome to light, low iron, and enduring viral infection. ISME J **13**, 2817–2833.

957 **Krahmer J, Hindle M, Perby LK, Mogensen HK, Nielsen TH, Halliday KJ, van Ooijen G, Le Bihan T,**
958 **Millar AJ**. 2022. The Circadian Clock Gene Circuit Controls Protein and Phosphoprotein Rhythms in
959 Arabidopsis thaliana. Mol Cell Proteomics **21**, 100172.

960 **Kusakina J, Dodd AN**. 2012. Phosphorylation in the plant circadian system. Trends in Plant Science
961 **17**, 575–83.

962 **Laloum D, Robinson-Rechavi M**. 2022. Rhythmicity is linked to expression cost at the protein level
963 but to expression precision at the mRNA level. PLOS Computational Biology **18**, e1010399.

964 **Langfelder P, Zhang B, Horvath S**. 2008. Defining clusters from a hierarchical cluster tree: the
965 Dynamic Tree Cut package for R. Bioinformatics **24**, 719–720.

966 **LaRonde-LeBlanc N, Wlodawer A**. 2005. The RIO kinases: An atypical protein kinase family required
967 for ribosome biogenesis and cell cycle progression. Biochimica et Biophysica Acta (BBA) - Proteins
968 and Proteomics **1754**, 14–24.

969 **Le Bihan T, Hindle M, Martin SF, Barrios-Llerena ME, Krahmer J, Kis K, Millar AJ, van Ooijen G**.
970 2015. Label-free quantitative analysis of the casein kinase 2-responsive phosphoproteome of the
971 marine minimal model species Ostreococcus tauri. Proteomics.

972 **Le Bihan T, Martin SF, Chirnside ES, van Ooijen G, Barrios-Llerena ME, O'Neill JS, Shliaha PV, Kerr**
973 **LE, Millar AJ**. 2011. Shotgun proteomic analysis of the unicellular alga Ostreococcus tauri. J
974 Proteomics **74**, 2060–70.

975 **Li L, Nelson CJ, Trosch J, Castleden I, Huang S, Millar AH**. 2017. Protein Degradation Rate in
976 Arabidopsis thaliana Leaf Growth and Development. The Plant Cell **29**, 207–228.

977 **Lu SX, Liu H, Knowles SM, Li J, Ma L, Tobin EM, Lin C**. 2011. A role for protein kinase casein kinase2
978 alpha-subunits in the Arabidopsis circadian clock. Plant Physiology **157**, 1537–45.

979 **Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S**. 2002. The protein kinase complement
980 of the human genome. Science **298**, 1912–34.

981 **Martin SF, Munagapati VS, Salvo-Chirnside E, Kerr LE, Le Bihan T**. 2012. Proteome Turnover in the
982 Green Alga Ostreococcus tauri by Time Course (15)N Metabolic Labeling Mass Spectrometry. J
983 Proteome Res **11**, 476–86.

984 **Mehra A, Baker CL, Loros JJ, Dunlap JC**. 2009. Post-translational modifications in circadian rhythms.
985 Trends Biochem Sci **34**, 483–90.

986 **Mehta D, Krahmer J, Uhrig RG**. 2021. Closing the protein gap in plant chronobiology. The Plant
987 Journal.

988 **Michael TP, Mockler TC, Breton G, *et al.*** 2008. Network discovery pipeline elucidates conserved
989 time-of-day-specific cis-regulatory modules. PLoS Genet **4**, e14.

990 **Millar AJ**. 2016. The intracellular dynamics of circadian clocks reach for the light of ecology and
991 evolution. Annual Review of Plant Biology **67**, 595–618.

992 **Missra A, Ernest B, Lohoff T, Jia Q, Satterlee J, Ke K, von Arnim AG**. 2015. The Circadian Clock
993 Modulates Global Daily Cycles of mRNA Ribosome Loading. The Plant Cell **27**, 2582–2599.

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

Moulager M, Monnier A, Jesson B, Bouvet R, Mosser J, Schwartz C, Garnier L, Corellou F, Bouget FY. 2007. Light-dependent regulation of cell division in Ostreococcus: evidence for a major transcriptional input. Plant Physiology **144**, 1360–9.

Noordally ZB, Millar AJ. 2015. Clocks in algae. Biochemistry **54**, 171–83.

Ocone A, Millar AJ, Sanguinetti G. 2013. Hybrid regulatory models: a statistically tractable approach to model regulatory network dynamics. Bioinformatics **29**, 910–6.

O'Neill JS, van Ooijen G, Dixon LE, Troein C, Corellou F, Bouget FY, Reddy AB, Millar AJ. 2011. Circadian rhythms persist without transcription in a eukaryote. Nature **469**, 554–8.

van Ooijen G, Dixon LE, Troein C, Millar AJ. 2011. Proteasome Function Is Required for Biological Timing throughout the Twenty-Four Hour Cycle. Current Biology **21**, 869–875.

van Ooijen G, Hindle M, Martin SF, Barrios-Llerena M, Sanchez F, Bouget FY, O'Neill JS, Le Bihan T, Millar AJ. 2013. Functional analysis of Casein Kinase 1 in a minimal circadian system. PLoS ONE **8**, e70021.

van Ooijen G, Knox K, Kis K, Bouget FY, Millar AJ. 2012. Genomic transformation of the picoeukaryote Ostreococcus tauri. J Vis Exp, e4074.

van Ooijen G, Millar AJ. 2012. Non-transcriptional oscillators in circadian timekeeping. Trends Biochem Sci **37**, 484–92.

O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D. 2013. pLogo: a probabilistic approach to visualizing sequence motifs. Nature Methods **10**, 1211–1212.

Paajanen P, Lane de Barros Dantas L, Dodd AN. 2021. Layers of crosstalk between circadian regulation and environmental signalling in plants. Current Biology **31**, R399–R413.

Pal SK, Liput M, Piques M, *et al.* 2013. Diurnal changes of polysome loading track sucrose content in the rosette of wild-type arabidopsis and the starchless pgm mutant. Plant Physiology **162**, 1246–65.

Piques M, Schulze WX, Hohne M, Usadel B, Gibon Y, Rohwer J, Stitt M. 2009. Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. Molecular Systems Biology **5**, 314.

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

Robbens S, Derelle E, Ferraz C, Wuyts J, Moreau H, Van de Peer Y. 2007. The complete chloroplast and mitochondrial DNA sequence of Ostreococcus tauri: organelle genomes of the smallest eukaryote are examples of compaction. Mol Biol Evol **24**, 956–68.

Schönberg A, Rödiger A, Mehwald W, Galonska J, Christ G, Helm S, Thieme D, Majovsky P, Hoehenwarter W, Baginsky S. 2017. Identification of STN7/STN8 kinase targets reveals connections between electron transport, metabolism and gene expression. The Plant Journal **90**, 1176–1186.

Schuirmann DL. 1981. On Hypothesis-Testing to Determine If the Mean of a Normal-Distribution Is Contained in a Known Interval. Biometrics **37**, 617–617.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. Science **330**, 1099–102.

Seaton DD, Graf A, Baerenfaller K, Stitt M, Millar AJ, Gruissem W. 2018. Photoperiodic control of the Arabidopsis proteome reveals a translational coincidence mechanism. Molecular Systems Biology **14**, e7962.

Smallwood CR, Chen J-H, Kumar N, *et al.* 2018*a*. Integrated systems biology and imaging of the smallest free-living eukaryote *Ostreococcus tauri*. bioRxiv.

Smallwood CR, Chrisler W, Chen J-H, Patello E, Thomas M, Boudreau R, Ekman A, Wang H, McDermott G, Evans JE. 2018*b*. *Ostreococcus tauri* is a high-lipid content green algae that extrudes clustered lipid droplets. bioRxiv.

Smith SM, Fulton DC, Chia T, Thorneycroft D, Chapple A, Dunstan H, Hylton C, Zeeman SC, Smith AM. 2004. Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in Arabidopsis leaves. Plant Physiology **136**, 2687–99.

Troein C, Corellou F, Dixon LE, van Ooijen G, O'Neill JS, Bouget F-Y, Millar AJ. 2011. Multiple light inputs to a simple clock circuit allow complex biological rhythms. The Plant Journal **66**, 375–385.

Turkina MV, Kargul J, Blanco-Rivero A, Villarejo A, Barber J, Vener AV. 2006. Environmentally modulated phosphoproteome of photosynthetic membranes in the green alga Chlamydomonas reinhardtii. Mol Cell Proteomics **5**, 1412–25.

Uhrig RG, Echevarría-Zomeño S, Schlapfer P, Grossmann J, Roschitzki B, Koerber N, Fiorani F, Gruissem W. 2021. Diurnal dynamics of the Arabidopsis rosette proteome and phosphoproteome. Plant, Cell & Environment **44**, 821–841.

Vizcaino JA, Csordas A, del-Toro N, *et al.* 2016. 2016 update of the PRIDE database and its related tools. Nucleic Acids Research **44**, D447-56.

Vizcaino JA, Deutsch EW, Wang R, *et al.* 2014. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nature Biotechnology **32**, 223–6.

Vlad F, Turk BE, Peynot P, Leung J, Merlot S. 2008. A versatile strategy to define the phosphorylation preferences of plant protein kinases and screen for putative substrates. The Plant Journal **55**, 104–117.

Ward JH. 1963. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association **58**, 236-.

Westlake WJ. 1981. Bioequivalence Testing - a Need to Rethink - Reply. Biometrics **37**, 591–593.

Whitelam GC, Halliday KJ. 2007. *Light and plant development*. Oxford ; Ames, Iowa: Blackwell Pub.

Wong DC, O'Neill JS. 2018. Non-transcriptional processes in circadian rhythm generation. Curr Opin Physiol **5**, 117–132.

Xue Y, Liu Z, Cao J, Ma Q, Gao X, Wang Q, Jin C, Zhou Y, Wen L, Ren J. 2011. GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. Protein Eng Des Sel **24**, 255–60.

Yang J, Zhang Y. 2015. Protein Structure and Function Prediction Using I-TASSER. Curr Protoc Bioinformatics **52**, 5 8 1-15.

Zienkiewicz K, Du Z-Y, Ma W, Vollheyde K, Benning C. 2016. Stress-induced neutral lipid biosynthesis in microalgae — Molecular, cellular and physiological insights. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1861**, 1269–1281.

FIGURE LEGENDS

Figure 1. Daily variation in transcripts, proteins, and phosphopeptide motifs. (A) Workflow for proteomics in *O. tauri* under LD. Overlap in (B) detected and quantified gene loci, (C) significantly changing (solid circles) or not significantly-changing (dashed circles) loci for transcripts (Monnier *et al.*, 2010), proteins and PMs; genomic loci excluded (square brackets). (D-I) Bi-plots of PCA for (D, E) transcript, (F, G) protein and (H, I) phosphomotif profiles. Proportion of the variance for each PC is indicated. Dot locations show the weighting of each RNA/protein/PM in each PC; colours show the assigned cluster (as in Supplementary Figure S4).

Figure 2. Distribution of rhythmic protein and phosphopeptide motif peaks, with examples. Temporal distribution of peaking profiles in (A) transcripts, (B) proteins and (C) PMs. (D, F) Simulated protein profiles from RNAs peaking at (D) ZT0 or (F) ZT16, with (red line) or without light-regulated translation (black line). (E) predicted distribution of protein peak times, with light-regulated translation. Examples of genes with (G) high-amplitude and similar protein (solid line) and PM profiles (coloured lines), or (H) PM profiles that differ from the protein profile. (G, H) protein and PM, left axis; RNA profile (dashed line), right axis. Error bars, S.E. Light/dark indicated by white/black bars.

Figure 3. Regulation of dark-accumulating proteins. Protein abundance profiles (A) of rhythmic prasinophyte-specific proteins in cluster P6 in LD and DA conditions. (B) Optical density (OD600; line, right axis) and total protein per cell (columns, left axis) under LD and DA conditions. (C) Correlation of protein degradation rates (Martin *et al.*, 2012) and relative protein levels after DA; chloroplast proteins (circles, chloroplast-encoded have solid outline); mitochondrial proteins (triangles, mitochondria-encoded outlined); PLP-enzymes (squares, marked in legend); prasinophyte-specific proteins (diamonds).

Figure 4. Protein and phosphopeptide motif regulation.
Phosphomotif (coloured lines) and RNA profiles (Monnier *et al.*, 2010)(dashed lines) of the photoreceptors, clock components, transcription factors and kinases indicated, under LD. Left axis range 2^6 (64-fold) except OtCCA1 (PM changes 150-fold) and OtCOL2 (PMs change up to 20-fold). Right (RNA) axis range 12, for log2 data (2^{12}=4096-fold in untransformed data). Error bars, S.E. Light/dark indicated by white/black bars. PHOT, phototropin photoreceptor; LOV-HK, LOV domain – histidine kinase photoreceptor; COL, CONSTANS-like transcription factor.

Figure 5. Motif enrichment and rhythmic protein kinases and phosphatases under LD. (A) Rhythmic PMs peaking at each timepoint on protein kinase (black) and phosphatase (grey) proteins (numbers). (B) Enrichment of proline-directed motifs, for kinases shown in the legend (dashed line, *p*-value = 0.05). (C) pLogo sequence motifs of rhythmic PMs peaking at each timepoint (foreground; fg), relative to all detected phosphopeptides (background; bg). ± 3.80 indicates *p*-value = 0.05, residues above and below axis are over- and under-represented, respectively. (D) Rhythmic PMs by kinase/phosphatase family, annotated with example proteins.

FIGURES

Figure 1

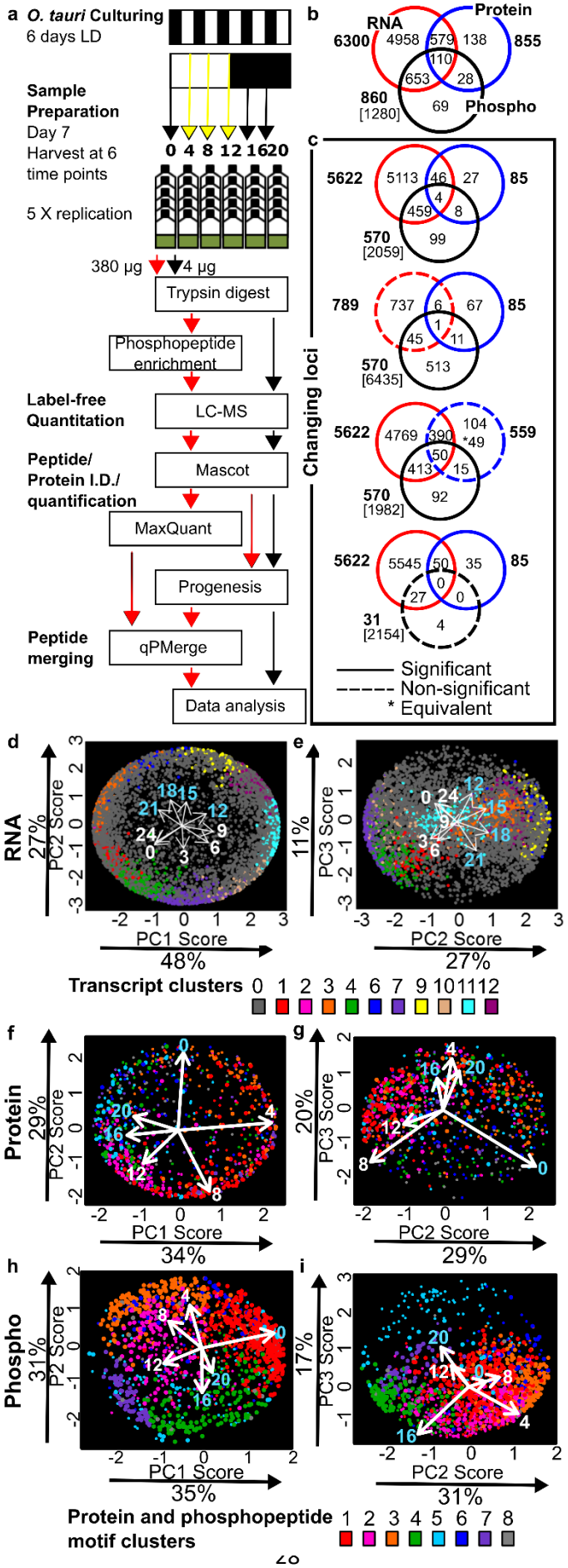


Figure 2

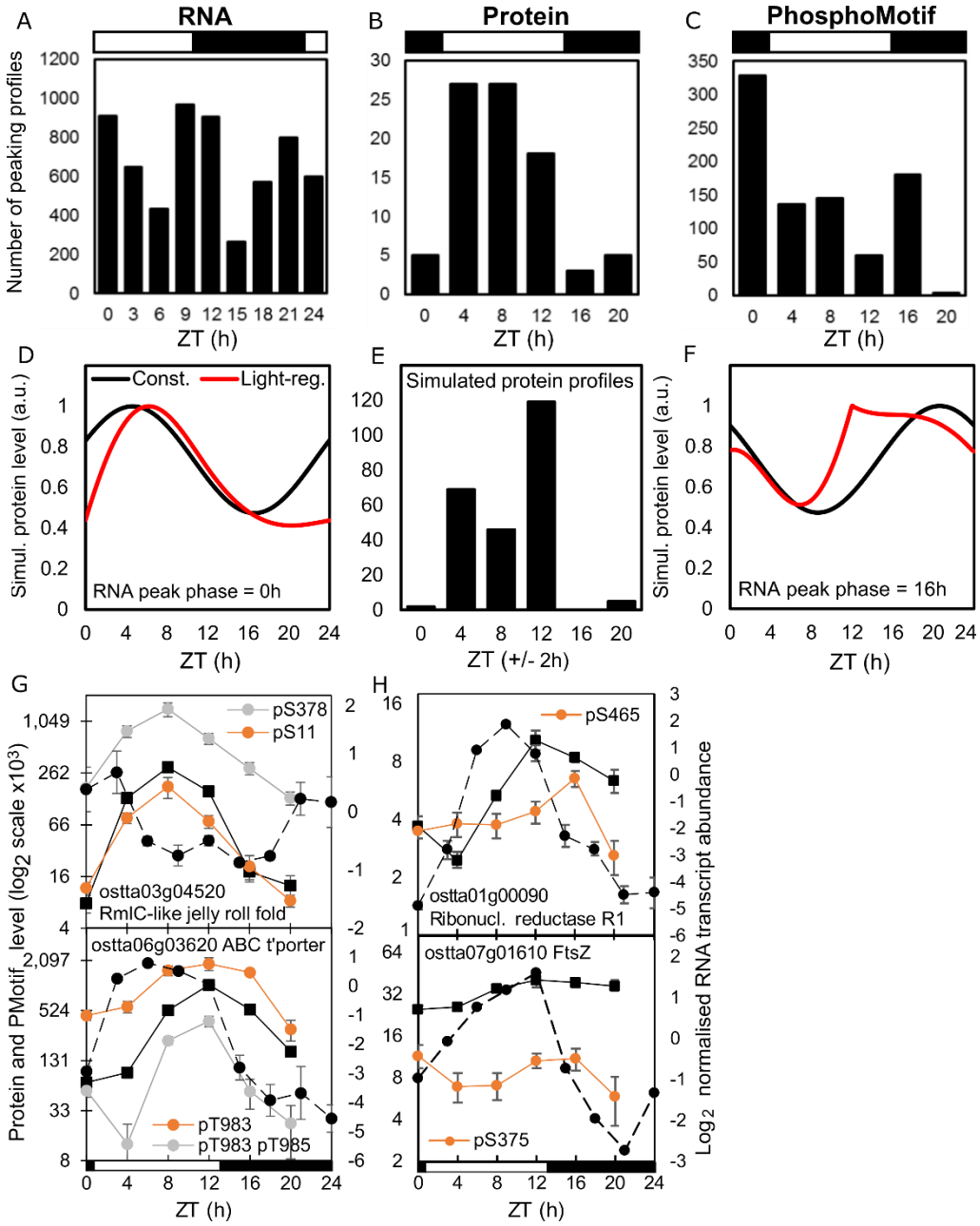


Figure 3

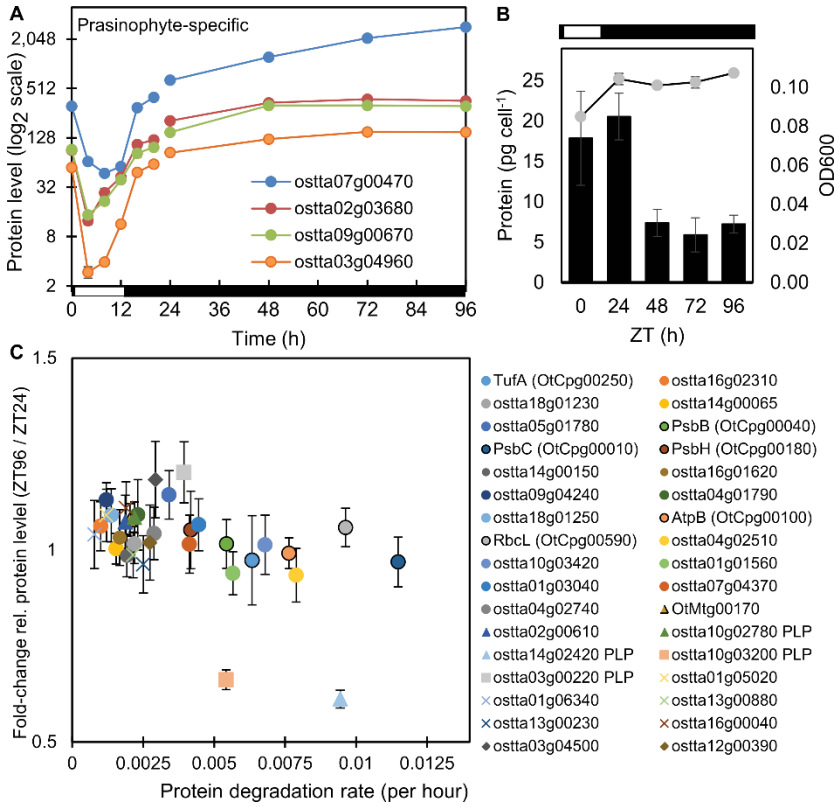


Figure 4

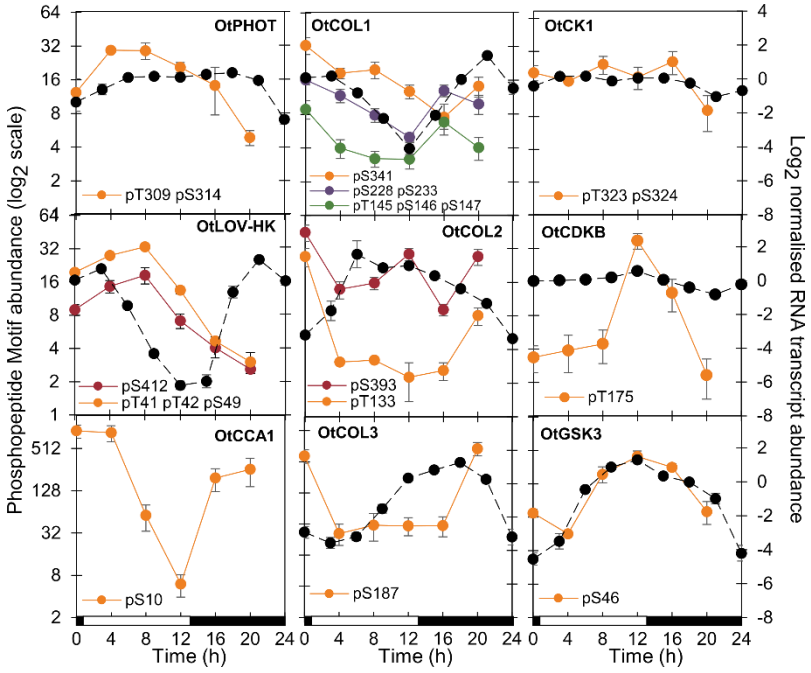
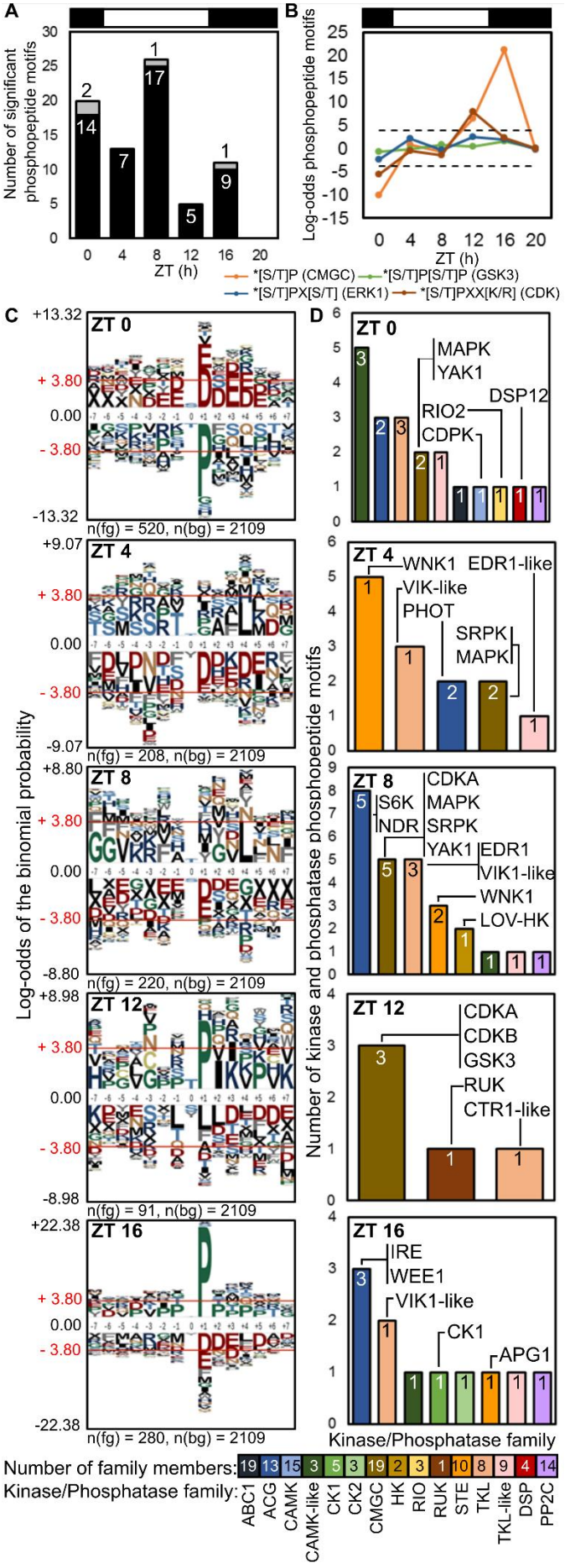


Figure 5



Supplementary Figure Legends

Supplementary Figure S1. Identification of outlier phosphopeptide replicate 4E. Pearson's correlation for (a) proteins and (b) phosphopeptide motifs and sample replicate r^2 respective to median abundance at a ZT for (c) proteins and (d) phosphopeptide motifs. Note differing scales in (a,b), (c,d).

Supplementary Figure S2. Most-detected protein and PM profiles. with comprehensive heat maps, clusters and enriched functions. Highly-abundant proteins (a) and PMs (b) under LD conditions (* marks rhythmic PMs). Error bars, S.E. Light/dark indicated by white/black bars, above.

Supplementary Figure S3. Changing PMs on non-changing proteins. Significantly non-changing proteins (black lines) determined by two one-sided tests (TOST; $\varepsilon = 0.3$), plotted with their rhythmic phosphopeptide motifs ± S.E., square brackets show phosphorylated residue. Light/dark indicated by white/black bars.

Supplementary Figure S4. Clustered protein and PM profiles with enriched functions. Heat maps of median-normalised (A) protein and (B) PM abundance, with insets top left showing the distribution of levels and colour scale. Clusters P1-8 or PM1-8 are shown, colours in 'cluster' track are as in Fig. 1; FDR track shows >1.5 fold-change and BH FDR adjusted *p*-value <0.05 (black line) or <0.01 (orange line); bars to right of each panel show the mean protein or PM abundance ($\log_{10}$ scale). Light/dark indicated by white/black bars, above. (C, D) Examples of significantly-changing proteins and PMs in each cluster (as noted in the main text).

Supplementary Figure S5. GO enrichments for peaks and troughs. GO Biological Process term enrichment for rhythmic (a) proteins and (b) phosphopeptide motifs, that was significant (Fisher's exact test p-value <0.05) in profiles with peak (no shading) or trough (pink shading) time at each timepoint. Light/dark indicated by white/black column. Grey bars represent proportion of significant terms identified with respect to total number of background annotated terms.

Supplementary Figure S6. Simulation of light-regulated translation. (a-c) Simulation of protein dynamics for an RNA with peak expression at ZT0 (a), ZT8 (b) and ZT16 (c), with observed, light-regulated translation rate (red lines) or with constant translation rate (black lines). Distribution of protein peaks (d,f) and troughs (e,g) for the model with light-regulated translation (d,e) compared to data (f,g). Distributions for constant translation would reflect the distribution of RNA profiles.

Supplementary Figure S7. Loci identified in both LD protein and phosphopeptide motif datasets. (a, b) Peak time is compared for genes identified in both LD protein and phosphopeptide motif datasets, with examples (c). (a) Mixed phase: multiple PMs, peaking at same and different times from cognate protein. Green shading in (b) follows number per bin. Plotting conventions in (c) follow Fig. 2c, 2d.

Supplementary Figure S8. Regulation of proteins tested under Dark Adaptation (DA). For ten proteins compared in the DA and metabolic labelling (Martin *et al.*, 2012) data (Fig. 4c), (a) RNA abundance under LD and DA conditions from qRT-PCR assays, and (b) protein profiles under LD. *, rhythmic proteins. Error bar, S.E.

Supplementary Figure S9. Protein and phospho-protein abundance in LD cycle. Stained gels showing changes in (a) protein and (b) phosphorylated protein abundance in LD, with (c) ratio of quantified, phosphorylated protein to total protein intensity.

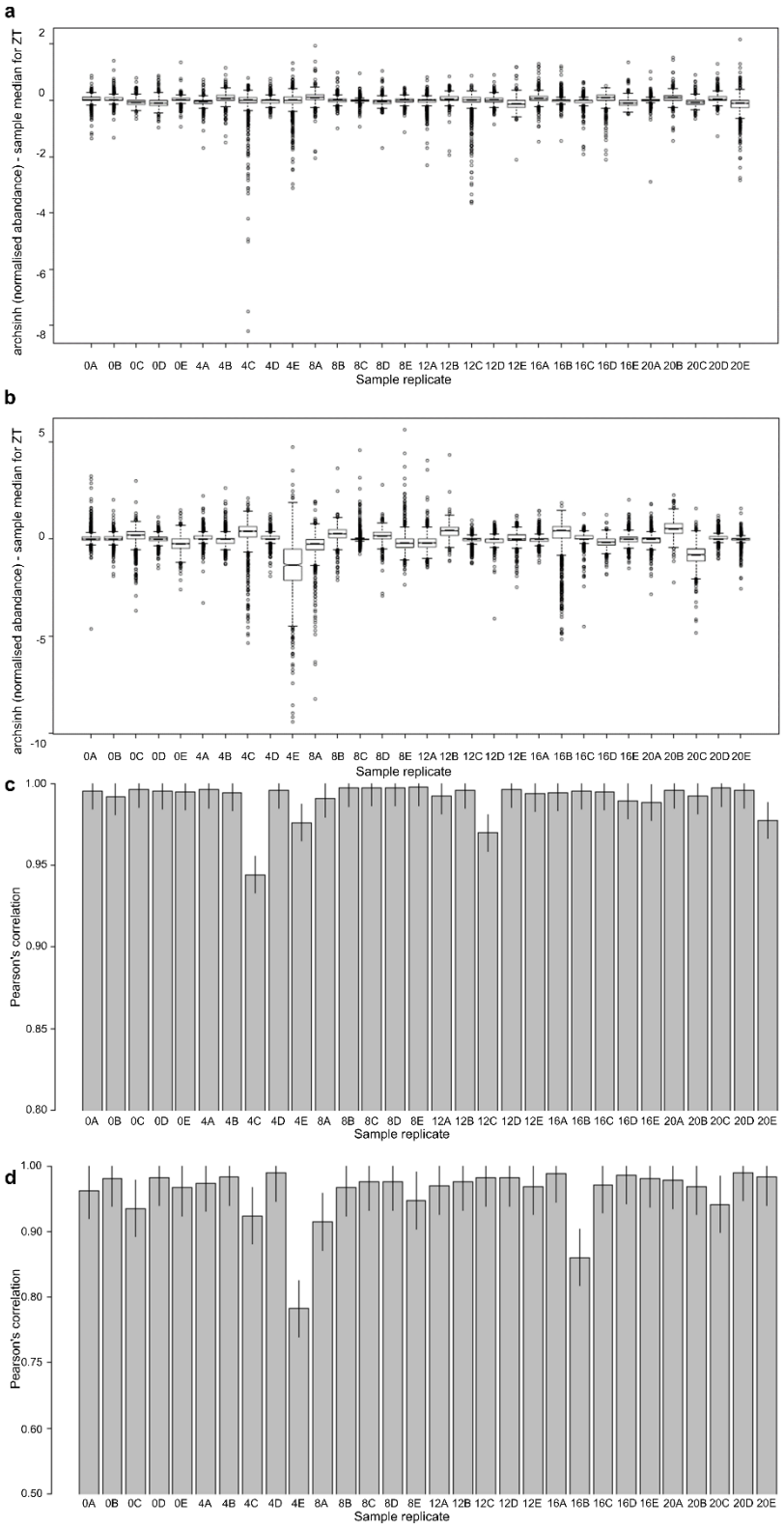
Supplementary Figure S10. CK1, CK2 and GSK3 kinase targets and phosphorylation sites in rhythmic kinases. Distribution of GPS3-predicted CK1 (black), CK2 (red) and GSK3 (black) targets among rhythmic phosphopeptide motifs, binned by peak (a) and trough (b) times. (c) Phosphosites on rhythmic protein kinases predicted to be phosphorylated by CK1, CK2 and GSK3, site location labels

coloured as in (a). * sites first reported here; †‡ sites observed previously (van Ooijen *et al.*, 2013). Protein kinase classes are coloured as in Figure 5.

Supplementary Figure S11. Structural homology of a rhythmic prasinophyte-specific protein. Structural homology models predicted using I-TASSER of (a) ostta02g03680 where the model is overlaid with (b) *H. sapiens* BAR domain structure (2d4c). Model α-helices (purple) and β-sheets (green) are numbered in black on the *O. tauri* model and in blue where structure is conserved with homologue protein overlay and in white where secondary structure is not conserved.

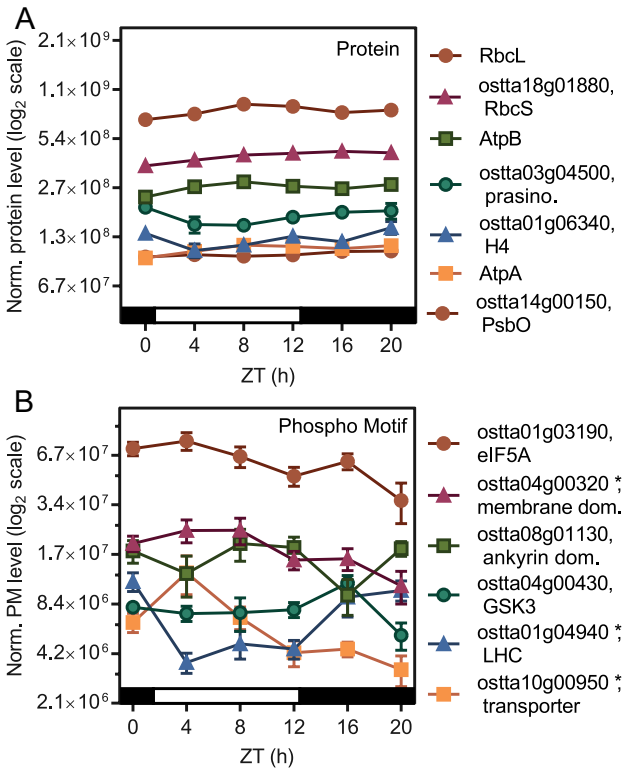
Supplementary Figures

Supplementary Figure S1

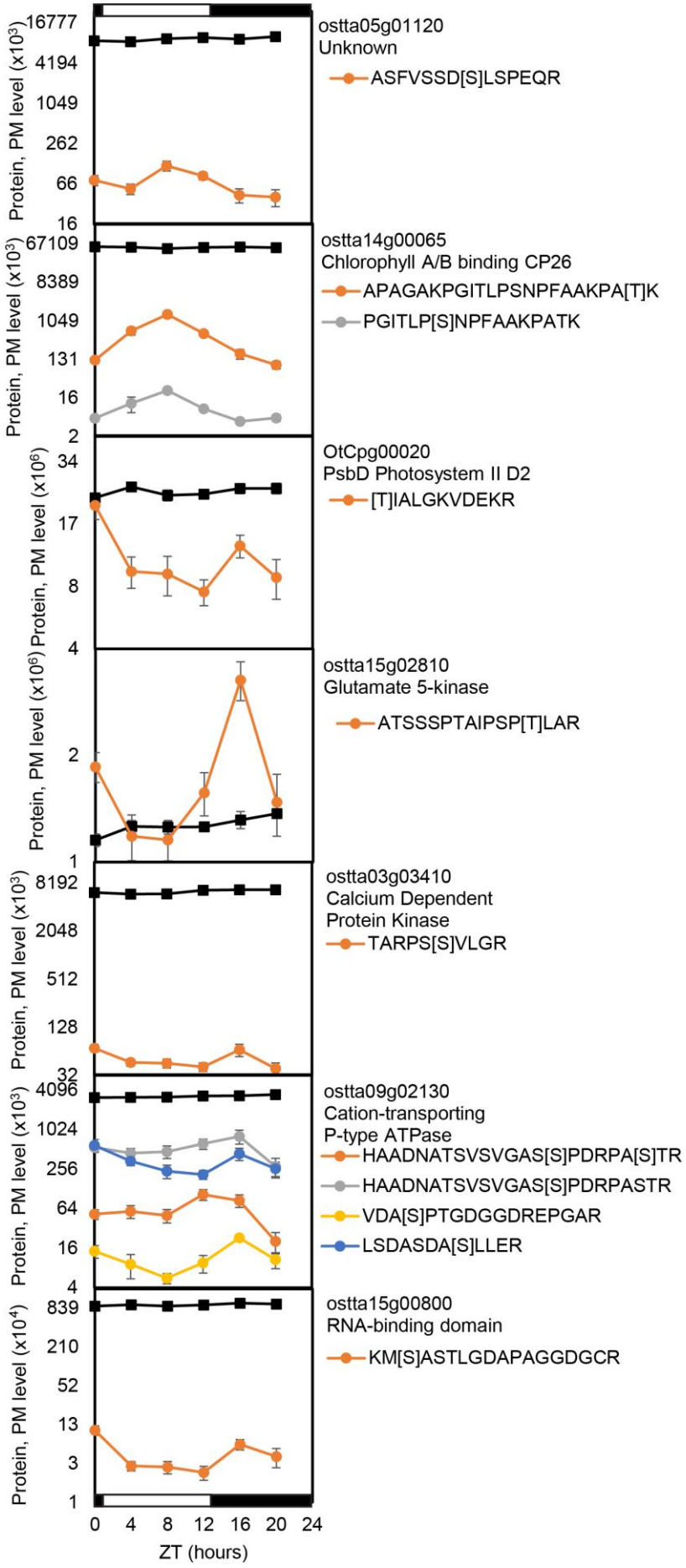


1207 **Supplementary Figure S2**

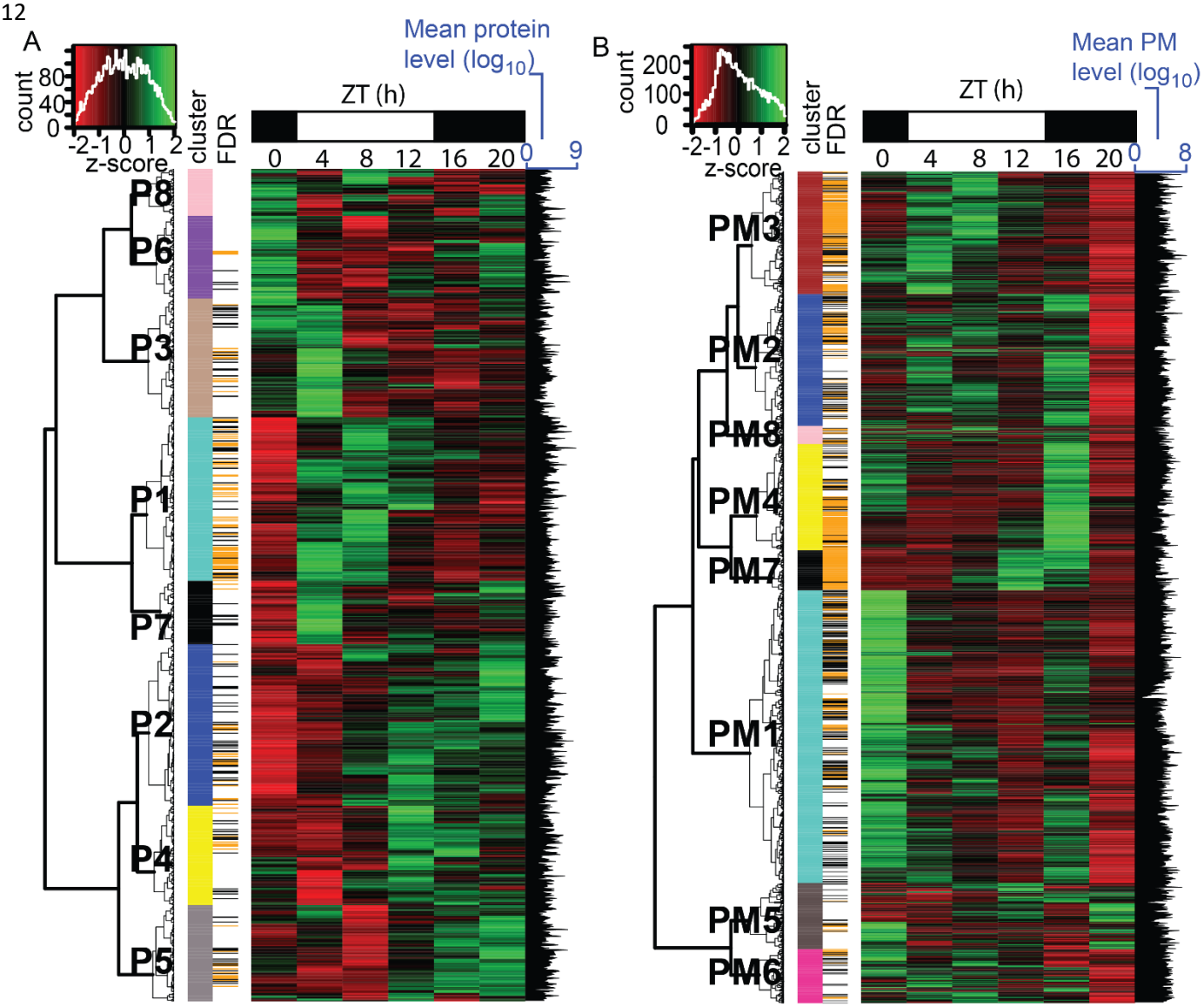
1208



Supplementary Figure S3



Supplementary Figure S4



C PROTEIN CLUSTERS

P1 Ribosomal subunits, translation initiation/elongation
P2 Thioredoxin-like, FNR, PGK, GAPDH, RuBisCO, PPDK, malate dehydrogenase CBM20, PLP transferase
P3 Chlorophyll A/B binding, PSII PsbM
P4 PLP transferases, SSII
P5 Thioredoxin-like, SQR/FR cytochrome b5, phosphogluconate dehydrogenase, transaldolase, aldehyde dehydrogenase
P6 Prasinophyte specific, NH_4+ transporter
P7 Cytochrome b6, ribosomal subunits, translation initiation/elongation

D PHOSPHOPEPTIDE MOTIF CLUSTERS

PM1 TFs, ZF, Myb; cold-shock protein; Zn, Na/solute, Mg2+, transport; AGC, YAK1, CAMK-like, RIO2, CDPK; PIPK; PsaM **PM2** TFs, ZF; eIFs; MAPK,APG1, NDR, AGC, CAMK-like; PP2C;6-phosphofructo-2-kinase; Ca ATPase **PM3** TFs, ZFs; eIFs; chlorophyll A/B binding, RubisCO activase; thioredoxin; K, Mg, ABC transport; FHA; AGC, SRPK, CDKA, CDKC ; PIPK **PM4** TFs ZF; PEP carboxylase proteosome; NO_3^- , Mg transport; WEE1, IRE, MAPK; PIPK **PM5** TFs, CO; GAPDH;chlorophyll A/B binding; Asn synthetase; T6P synthase; **PM6** eIFs, ribosome biogenesis; S6K; PIPK, dacylglycerol kinase **PM7** CDKA, CDKB, GSK3; PPDK; phospholipid ATPase **PM8** WNK1, EDR1, CDK regulatory subunit

Supplementary Figure S5

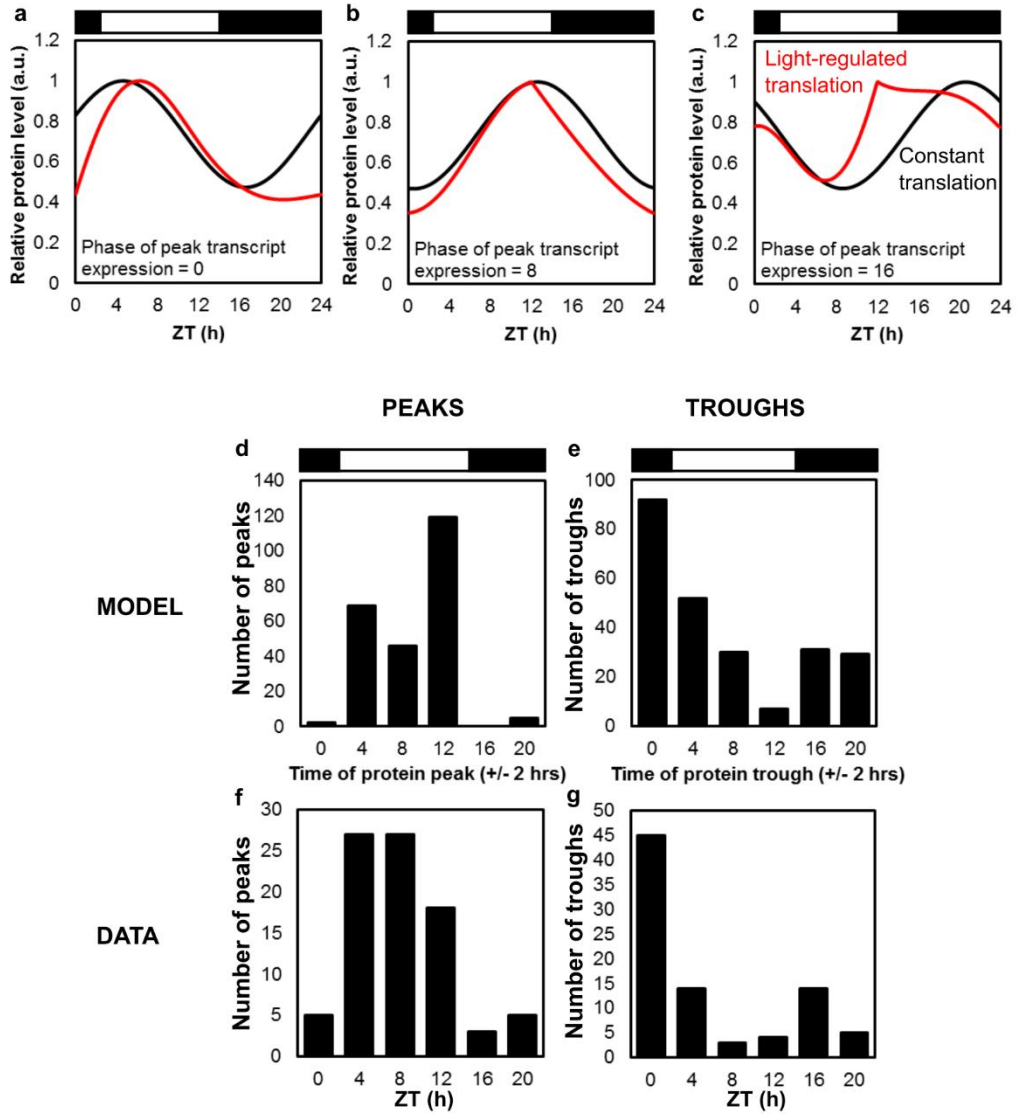
a Protein - Biological Process

ZT	GO.ID	Term	Annotated	Significant	Expected	p-value
0	GO:0098662	inorganic cation transmembrane transport	17	1	0.1	9.40E-02
	GO:0006412	translation	116	17	6.8	1.30E-04
4	GO:0006412	translation	116	16	4.82	1.30E-06
	GO:0006414	translational elongation	8	2	0.33	4.00E-02
	GO:0006260	DNA replication	3	3	0.04	1.40E-06
8	GO:0015995	chlorophyll biosynthetic process	7	4	0.24	3.10E-05
	GO:0006414	translational elongation	8	2	0.27	2.70E-02
	GO:0006412	translation	116	9	3.95	3.00E-02
	GO:0006418	tRNA aminoacylation for protein translation	15	1	0.03	2.80E-02
12	GO:0006260	DNA replication	3	3	0.07	1.20E-05
	GO:1901606	alpha-amino acid catabolic process	5	2	0.12	4.50E-02
	GO:0090150	establishment of protein localization to membrane	4	1	0.02	2.30E-02
20	GO:0015995	chlorophyll biosynthetic process	7	1	0.04	3.90E-02
	GO:0006414	translational elongation	8	1	0.05	4.50E-02

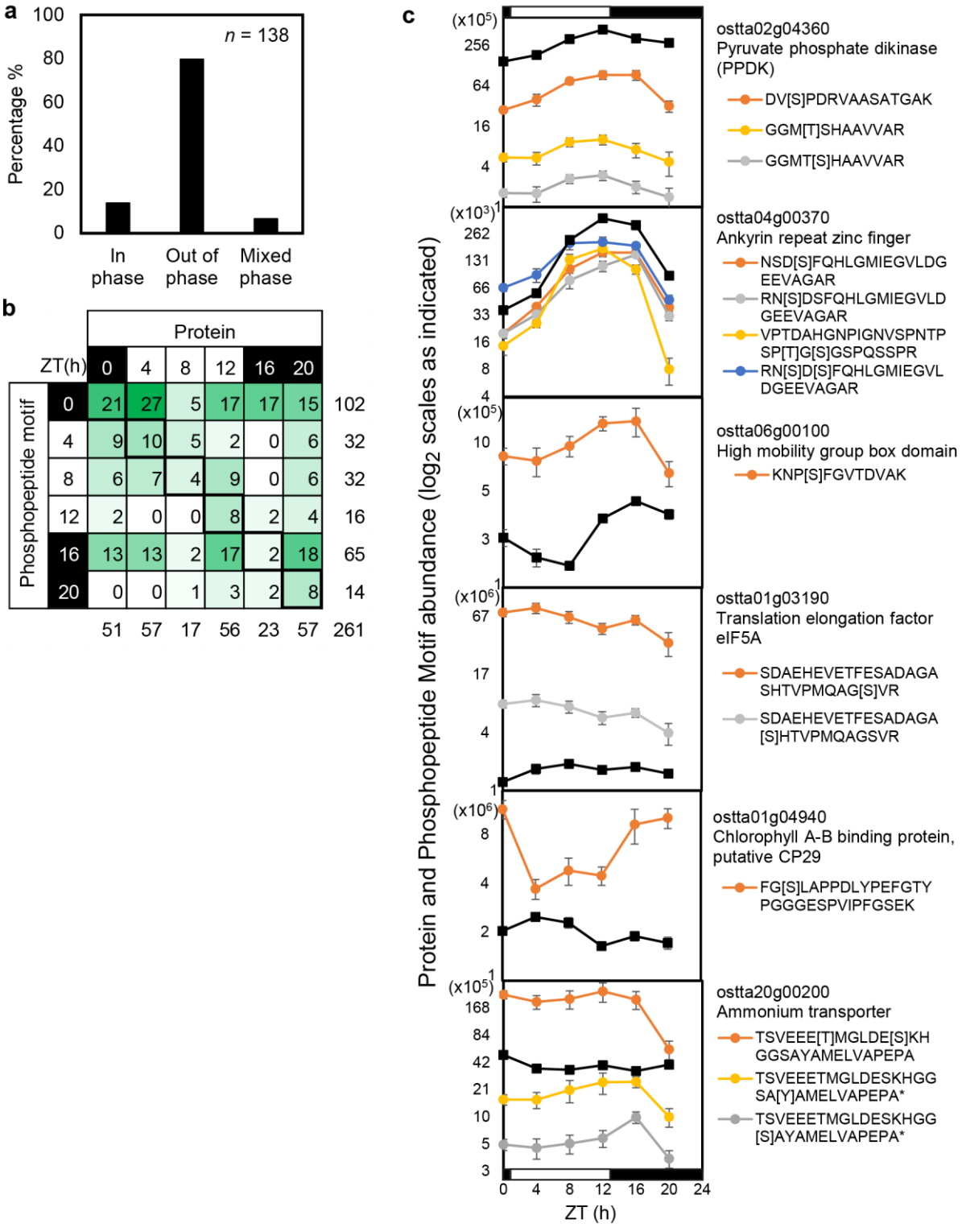
b Phosphopeptide motifs - Biological Process

ZT	GO.ID	Term	Annotated	Significant	Expected	p-value
0	GO:0035556	intracellular signal transduction	17	6	3.38	3.10E-03
	GO:0046488	phosphatidylinositol metabolic process	6	4	1.19	8.10E-03
	GO:0006457	protein folding	4	3	0.79	1.55E-02
	GO:0006413	translational initiation	4	3	0.79	1.55E-02
	GO:0006468	protein phosphorylation	48	14	9.53	1.62E-02
	GO:0050896	response to stimulus	34	11	6.75	2.49E-02
	GO:0034622	cellular macromolecular complex assembly	5	3	0.99	2.69E-02
	GO:0055085	transmembrane transport	38	12	7.55	3.23E-02
	GO:0015031	protein transport	22	7	4.37	3.29E-02
	GO:0006352	DNA-templated transcription, initiation	5	3	0.99	3.42E-02
	GO:0006468	protein phosphorylation	48	5	1.17	2.60E-03
4	GO:0055085	transmembrane transport	38	11	3.97	1.20E-05
	GO:0006355	regulation of transcription, DNA-templated	22	8	2.3	7.50E-05
	GO:0000160	phosphorelay signal transduction system	6	3	0.63	6.70E-03
	GO:0008152	metabolic process	231	29	24.15	1.20E-02
	GO:0072525	pyridine-containing compound biosynthetic process	3	2	0.31	1.57E-02
	GO:0044271	cellular nitrogen compound biosynthetic process	46	12	4.81	2.07E-02
	GO:0009108	coenzyme biosynthetic process	5	2	0.52	4.74E-02
	GO:0006355	regulation of transcription, DNA-templated	22	6	1.3	6.10E-04
	GO:0009168	purine ribonucleoside monophosphate biosynthetic proc	3	2	0.18	8.20E-03
	GO:0009152	purine ribonucleotide biosynthetic process	4	2	0.24	1.58E-02
	GO:0046129	purine ribonucleoside biosynthetic process	4	2	0.24	1.58E-02
	GO:0034654	nucleobase-containing compound biosynthetic process	40	10	2.37	4.11E-02
	GO:0006813	potassium ion transport	7	2	0.41	5.01E-02
8	GO:0006468	protein phosphorylation	48	18	4.93	8.40E-10
	GO:0032012	regulation of ARF protein signal transduction	3	2	0.31	1.80E-02
	GO:0005992	trehalose biosynthetic process	3	2	0.31	1.80E-02
	GO:0006355	regulation of transcription, DNA-templated	22	5	2.26	2.50E-02
	GO:0006511	ubiquitin-dependent protein catabolic process	3	2	0.13	4.10E-03
	GO:0007017	microtubule-based process	4	2	0.17	8.00E-03
	GO:1901362	organic cyclic compound biosynthetic process	46	3	2	2.01E-02
	GO:0044238	primary metabolic process	192	12	8.36	2.87E-02
	GO:0098662	inorganic cation transmembrane transport	9	2	0.39	4.29E-02
12	GO:0006261	DNA-dependent DNA replication	3	2	0.11	3.10E-03
	GO:0006468	protein phosphorylation	48	5	1.84	1.58E-02
	GO:0006396	RNA processing	7	2	0.27	1.98E-02
	GO:0006974	cellular response to DNA damage stimulus	10	2	0.38	3.19E-02
	GO:0019219	regulation of nucleobase-containing compound metabolic process	23	2	0.88	3.26E-02
	GO:0006457	protein folding	4	3	0.28	8.70E-04
	GO:0050896	response to stimulus	34	6	2.37	2.99E-02
	GO:0008152	metabolic process	231	27	16.1	4.15E-02
16	GO:0006281	DNA repair	9	5	1.08	8.20E-04
	GO:0006310	DNA recombination	4	3	0.48	3.65E-03
	GO:0009168	purine ribonucleoside monophosphate biosynthetic process	3	2	0.36	2.82E-02
	GO:0006468	protein phosphorylation	48	9	5.77	4.15E-02
	GO:0006355	regulation of transcription, DNA-templated	22	3	0.5	8.60E-03
	GO:0007165	signal transduction	19	2	0.43	3.91E-02
20	GO:0006412	translation	11	1	0.02	1.90E-02
	GO:0006468	protein phosphorylation	48	28	16.64	2.80E-07
	GO:0055085	transmembrane transport	38	20	13.17	7.50E-05
	GO:0006355	regulation of transcription, DNA-templated	22	13	7.63	5.60E-04
	GO:0008152	metabolic process	231	105	80.09	8.70E-04
	GO:0016310	phosphorylation	52	32	18.03	2.27E-03
	GO:0000160	phosphorelay signal transduction system	6	5	2.08	4.50E-03
	GO:0046488	phosphatidylinositol metabolic process	6	5	2.08	4.50E-03
	GO:0009058	biosynthetic process	83	45	28.78	8.60E-03
	GO:0006796	phosphate-containing compound metabolic process	76	42	26.35	9.00E-03
	GO:0044249	cellular biosynthetic process	76	40	26.35	9.53E-03
	GO:0035556	intracellular signal transduction	17	13	5.89	1.31E-02
	GO:0006812	cation transport	22	11	7.63	1.45E-02
	GO:0006352	DNA-templated transcription, initiation	5	4	1.73	1.54E-02
	GO:0009108	coenzyme biosynthetic process	5	4	1.73	1.54E-02
	GO:0072525	pyridine-containing compound biosynthetic process	3	3	1.04	1.55E-02
	GO:0032012	regulation of ARF protein signal transduction	3	3	1.04	1.55E-02
	GO:0030258	lipid modification	3	3	1.04	1.55E-02
	GO:0090407	organophosphate biosynthetic process	9	4	3.12	1.57E-02
	GO:0044711	single-organism biosynthetic process	28	14	9.71	4.57E-02

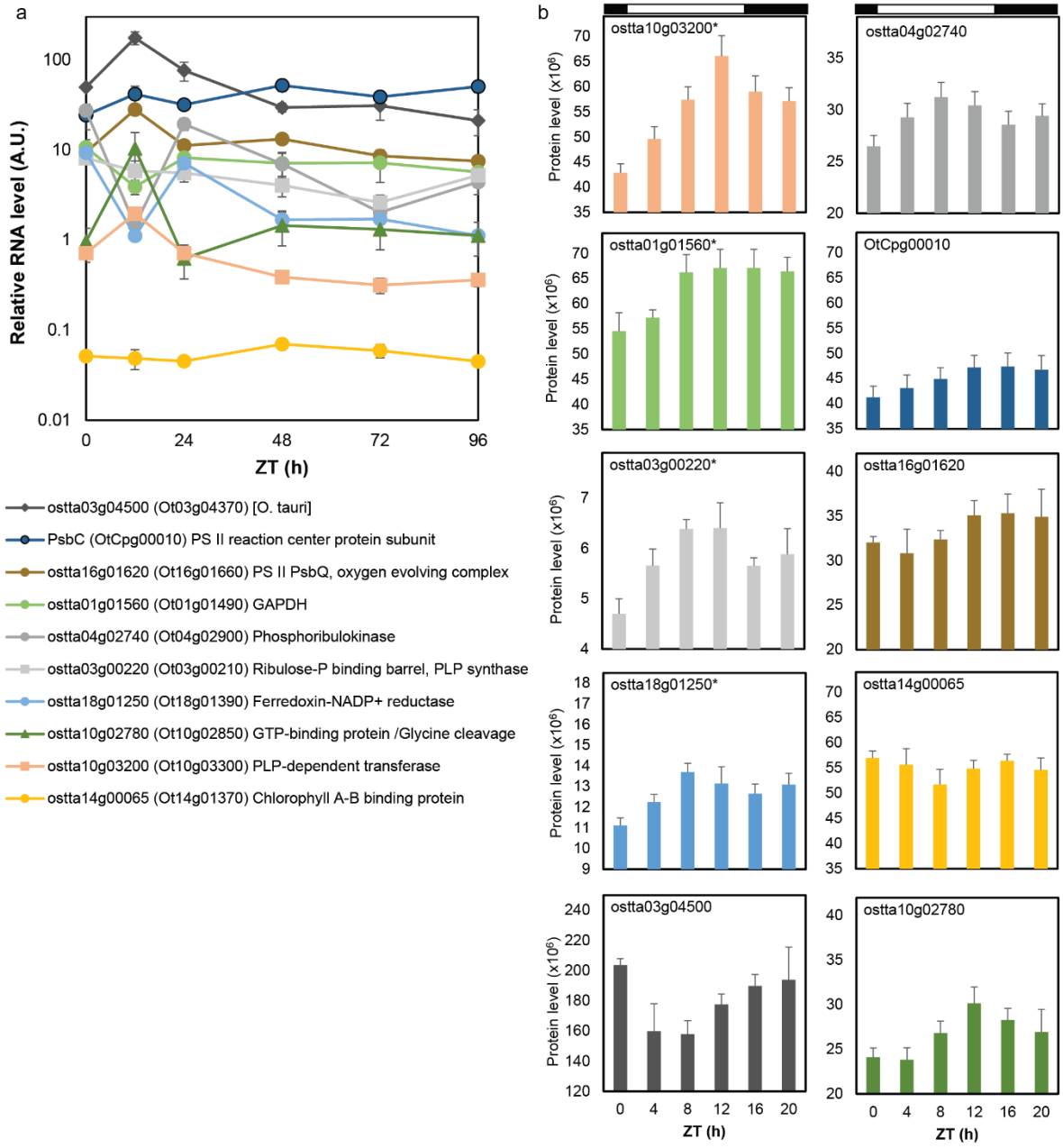
Supplementary Figure S6



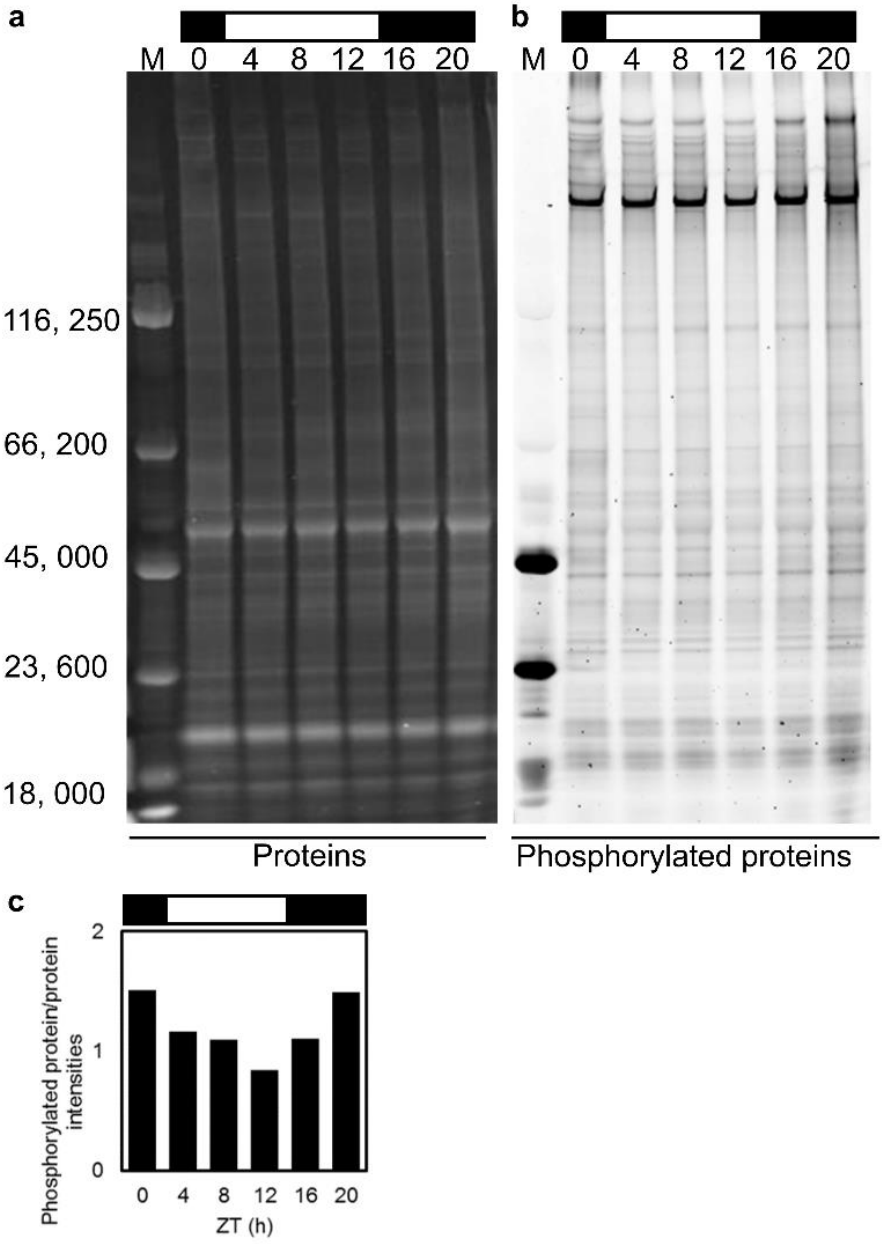
Supplementary Figure S7



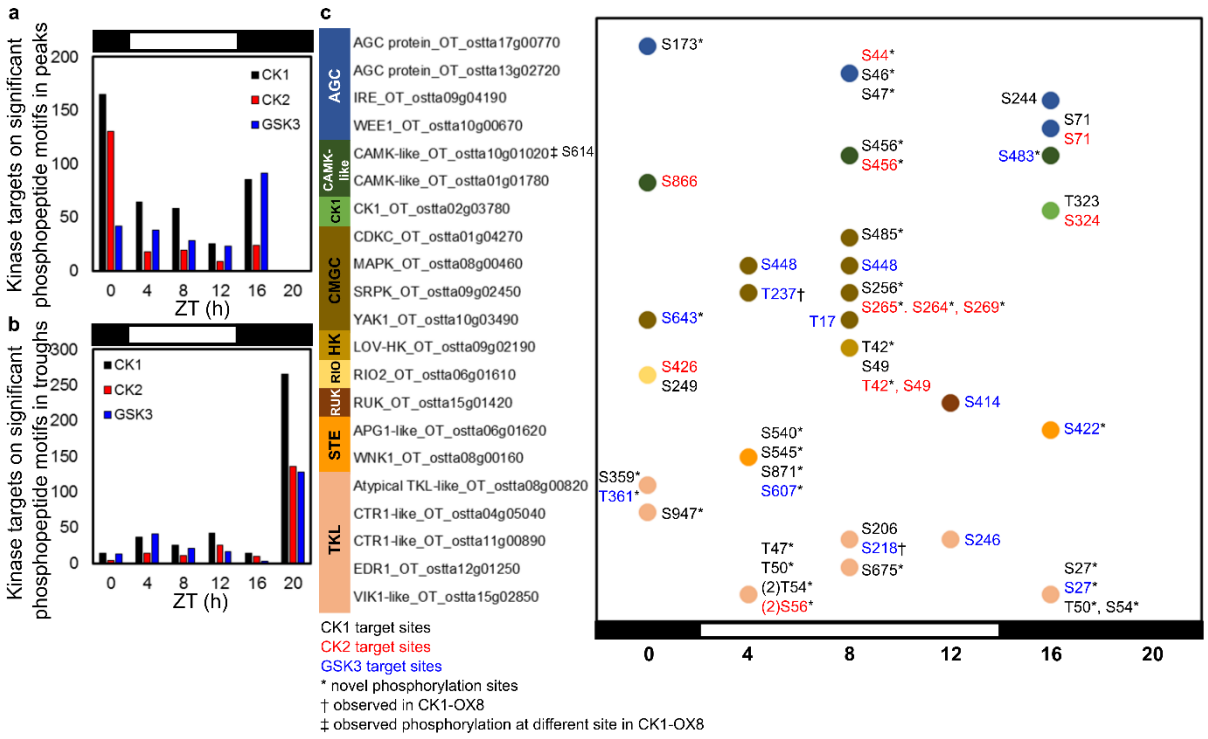
Supplementary Figure S8



Supplementary Figure S9



Supplementary Figure S10



Supplementary Figure S11