Circadian regulation of the transcriptome in a complex polyploid crop

Hannah Rees[1], Rachel Rusholme-Pilcher[1], Paul Bailey[2], Joshua Colmer[1], Benjamen White[1], Connor Reynolds[1], Calum A. Graham[3,4], Luíza Lane de Barros Dantas[3], Antony N. Dodd[3], Anthony Hall[1]

[1] Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK
[2] Royal Botanic Gardens Kew, Richmond, Surrey, TW9 3AE, UK
[3] John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK
[4] School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK.

Abstract

The circadian clock is a finely balanced time-keeping mechanism that coordinates programmes of gene expression. In polyploids, this regulation must be coordinated over multiple subgenomes. Here, we generate and analyse a high-resolution time-course dataset to investigate the circadian balance between sets of three homoeologous genes (triads) from hexaploid bread wheat. We find a large proportion of circadian triads exhibit unbalanced rhythmic expression patterns, with no specific subgenome favoured. In wheat, period lengths of rhythmic transcripts are found to be longer and have a higher level of variance than in other plant species. Biological processes under circadian control are largely conserved between wheat and Arabidopsis, however striking differences are seen in agriculturally critical processes such as starch metabolism. Together, this work highlights the ongoing selection for balance versus diversification in circadian homoeologs, and identifies clock-controlled pathways that might provide important targets for future wheat breeding.

Introduction

Circadian clock homologs have been both inadvertently selected during crop domestication and identified as crop improvement targets[1–4]. Understanding circadian regulation of the transcriptome in crops such as bread wheat (*Triticum aestivum*) may provide useful insights for the future. Wheat also provides an excellent model system to explore how the circadian clock and its outputs are co-ordinated in a recently formed, complex allopolyploid. In *Arabidopsis*, circadian transcription factors act in a dose-dependent manner, with both knock-

out and over-expression mutants resulting in altered function of the circadian oscillator[5–8]. It is not yet understood how rhythmic gene expression is balanced in species with multiple copies of the same gene. *T. aestivum* is a hexaploid (AABBDD) formed through interspecific hybridisation of three diploid ancestors around 10,000 years ago[9,10]. 51.7% of high-confidence wheat genes still exist in triads; sets of three homoeologous genes present on each of the A, B and D genomes[11]. As these homoeologs evolved independently for several million years prior to hybridization, it is plausible that these independent species might have been subject to different selective pressures on their clocks (Fig. 1a).

The circadian network in *Arabidopsis* comprises a series of interlocking negative transcriptional feedback loops connected by key activators[12]. Although monocots such as wheat diverged from their dicot relatives over 140 million years ago[13], many circadian oscillator components seem to have been conserved, particularly those forming the core loop network. Orthologs of *TIMING OF CAB EXPRESSION 1 (TOC1*) and other PSEUDO-RESPONSE REGULATOR (PRR) genes have been identified in wheat, rice and barley, and several loci within these genes have been associated with altered flowering times, most notably *(ppd-1*) within *TaPRR3/7*[14–16]. Likewise, mutants of orthologs of *LATE ELONGATED HYPOCOTYL (LHY)*, *GIGANTEA* (*GI*), *EARLY FLOWERING 3 (ELF3)*, and *LUX ARRYTHMO (LUX)* have been identified that alter heading dates, pathogen susceptibility, plant height or lower grain yields[17–21].

Circadian control of carbon fixation and starch metabolism are thought to form part of the selective advantage conferred by the clock[22,23]. This is apparent in the short-period *lhy⁻/cca1⁻* double mutant, where night-time starch levels reach exhaustion earlier compared to wild-type, triggering early onset starvation responses that reduce plant productivity[23]. Similarly, genes encoding photosynthesis-related proteins are well-established targets of the circadian clock and include the *LIGHT HARVESTING CHLOROPHYLL A/B BINDING PROTEIN* genes (*LHCB* also known as *CAB* genes) and photosystem I and II reaction centre genes[24,25].

Here, we investigate circadian balance within wheat triads to understand how circadian control is co-ordinated in a polyploid crop with three subgenomes. Second, we examine similarities and differences between the circadian transcriptome in wheat and its distant dicot relative, *Arabidopsis*, at a global level and at the level of genes encoding key pathways such as primary metabolism and photosynthesis.

Results

Global analysis of the circadian transcriptome in wheat

We generated a circadian RNA-seq time-course and compared it with a recently published dataset from *Arabidopsis*[26] over 24h - 68h following transfer to constant light. Rhythmicity was assessed using Metacycle Benjamini-Hochberg (BH) *q*-values. Of the 86,567 genes expressed in wheat, 33.0% were rhythmically expressed with a BH $q < 0.05$ and 21.5% with a BH $q < 0.01$ (Supplementary Note 1, Supplementary Table 1). This was significantly lower than the proportions of rhythmically expressed genes in the *Arabidopsis* dataset (50.7% BH $q < 0.05$, 39.1% BH $q < 0.01$) using the same criteria (X^2 (1) = 2727.1, $p < 0.001$, one-tailed, two-proportions z-test). Circadian waveform characteristics of the rhythmically expressed genes (BH $q < 0.01$) in the wheat and *Arabidopsis* datasets were quantified using algorithms in Metacycle (JTK, ARSER, LS and meta2d) and Biodare2 (FFT-NLLS and MESA). Period, phase, and amplitude estimates from FFT-NLLS and meta2d were well-correlated for individual genes (Supplementary Fig. 1). All models reported that mean period length in wheat was approximately 3h longer than in *Arabidopsis* (wheat = 25.9 - 27.5h, *Arabidopsis* = 22.6 - 24.4h; t(36067) = 101.58, $p<0.001$, Welch's two sample t-test; Supplementary Fig. 2). There was no significant difference between mean periods across the three wheat sub-genomes (Fig. 1c, F(2, 28,276) = 0.179, $p=0.836$, One-way ANOVA).

We used meta2d to compare period means and distributions from four previously published circadian datasets, and found that period lengths in wheat were longer and had higher standard deviation than period distributions from *Arabidopsis*, *Brassica rapa*, *Brachypodium distachyon* and *Glycine max* (Fig 1b; Supplementary Table 2).

We investigated how wheat periods changed over the course of the three-day experiment and found that periods were longer immediately after transfer to constant light (28.61h, SD=3.421h), and progressively shortened over the following days (Supplementary Note 2).

For both *Arabidopsis* and wheat, we recalculated phases of rhythmic transcripts relative to endogenous period (circadian time; CT). Across all algorithms, most transcripts in *Arabidopsis* peaked during the subjective night (around CT12-24; Supplementary Fig. 3). In wheat, the greatest numbers of rhythmic genes peaked during the subjective day (around CT6-8) with a second, smaller group being expressed in the night (~CT20). When we grouped transcripts into 2h period bins, we found that transcripts with short periods contained

proportionally more dawn-peaking transcripts, whereas those with longer periods contained proportionally more dusk-peaking transcripts (Supplementary Note 3, Supplementary Fig. 4).

Balance of circadian regulation within triads

In wheat, over 72% of triads are estimated to have "balanced" expression, with similar relative abundance of transcripts from each of the three homoeologs[27]. Due to the importance of the clock in coordinating dosage of gene expression, our hypothesis was that many circadian triads would have balanced regulation. We defined unbalanced circadian regulation as triads harbouring differences in rhythmicity (i.e., BH *q*-values), period lengths, phases, and relative amplitudes.

Of the 16,359 expressed triads in our dataset, 9901 (60.52%) had at least one rhythmic homoeolog, and 3448 (21.08%) had three rhythmically expressed genes (BH $q < 0.05$), with the latter hereafter termed "rhythmic triads" (purple segment, Fig. 1d). 6453 triads lacked rhythmicity in either one or two expressed homoeolog(s) (green and blue segments, Fig. 1d). In both cases, there was no bias for absence of rhythmicity in the A, B or D copy ($\chi^2(2)= 6.8415$, $p = 0.40$ where one gene is arrhythmic, $\chi^2(2)= 6.8415$, $p = 0.03$ where two genes are arrhythmic). We found cases where high-confidence rhythmic homoeologs (BH $q < 0.01$) occurred alongside arrhythmic homoeologs (BH $q > 0.05$) represented by light-shaded outer-ring segments in Fig. 1d. In total there were 3450 unbalanced-rhythmicity triads, contrary to our original expectation that most circadian regulated triads would be balanced (Fig. 1h,i). To explore other forms of circadian imbalance, we assessed whether phase, period and relative amplitude were conserved between homoeologs within the rhythmic triad set (purple segment, Fig. 1d). Differences in phases were quantified by a cross-correlation analysis to assess whether the correlation between homoeologs was improved with a time lag of 4, 8 or 12 hours. We identified 464 triads with unbalanced phases with an optimum lag of >0h between homoeologs (Fig. 1e,j). 1,139 triads had unbalanced periods with more than 2h difference in period between homoeologs (Fig. 1f,k). 701 triads had unbalanced relative amplitudes with more than two-fold difference in relative amplitude (Fig. 1g,l). Within this last group, the homoeolog with the lowest amplitude was still rhythmic, as observed when data are mean-normalized (Fig. 1l,m).

Overall, there was a total of 5,082 circadian unbalanced triads where at least one homoeolog had different circadian regulation from the other homoeologs. The largest cause of unbalanced circadian expression within triads was absence of rhythmicity (67.89%) with differences in period (22.41%), relative amplitude (13.79%) and phase (9.13%) occurring

more infrequently and with some overlap between categories. This compares with 1816 circadian balanced triads where all three homoeologs are rhythmic and have similar, phases, periods, and amplitudes giving us a ratio of almost 3:1 for unbalanced to balanced circadian triads.

One potential explanation for unbalanced rhythmicity is that arrhythmic homoeologs are silenced. In support of this, we found that the rhythmic homoeologs in unbalanced triads were expressed at a significantly higher baseline level than their arrhythmic homoeologs (Fig. 1n; $F(16, 35{,}148) = 6.94$, $p<0.001$, Two-level, nested ANOVA on Log10 transformed data). We also found that triads with balanced rhythmicity were expressed at a uniformly higher level than the most highly expressed homoeolog(s) in the unbalanced triads (Fig. 1n; $F(7, 35{,}148) = 570.909$, $p<0.001$). Therefore, in unbalanced rhythmicity triads, the rhythmic homoeolog does not appear to compensate for reduced expression of the other homoeolog(s), so the overall expression across the triad is reduced. This is supported by research in diploid *Brassica rapa*, which found that circadian regulated paralogs are expressed at a higher level than single copy genes[28].

To investigate whether certain biological processes were associated with circadian balance, we compared GO-slim terms enriched in the 1816 circadian balanced triads, the 5082 differently circadian regulated triads and the 6458 arrhythmic triads, identifying significant terms unique to each group (Table 2). Some terms were enriched only in circadian balanced triads (p-value <0.0001, Fisher's exact test e.g., "photosynthesis", "generation of precursor metabolites and energy", "gene-expression" and "translation"). In contrast, GO-slim terms: "developmental process involved in reproduction", "and "system development" were significantly enriched in triads with differently regulated homoeologs (p-value <0.0001) but not in triads with circadian balanced homoeologs (p-value >0.5). A possible explanation for this enrichment could be that unbalanced circadian triads are more likely to be dynamically expressed over developmental stages or show local dominance of a subgenome in a particular tissue type.

Transcription factor (TF) triads were as likely to be circadian balanced/unbalanced as non-transcription factors (χ^2 (2, N = 13,356) = 3.03, p = 0.08, chi-square test). Previously validated wheat TFs with unbalanced circadian expression included *WCBF2* (aka *TaCBF1*) and *TaPCF5*, both of which regulate abiotic stress responses[29,30] (Supplementary Note 4, Supplementary Fig. 5).

Clustering of gene expression and GO-term enrichment

To establish whether similar phased transcripts in wheat and *Arabidopsis* had similar biological roles, we clustered rhythmic transcripts (BH $q < 0.01$) into 9 expression modules for each species and identified GO-slim terms enriched in each module ($p < 0.01$). Circadian characteristics of module eigengenes are shown in Supplementary Table 4. We compared the correlation and cross-correlation of pairwise modules in the two species to find modules that correlated with a peak lag of 0 (synchronous phase) or with a peak lag of 4, 8, or 12h (asynchronous phase). Overall, modules with synchronous phases in wheat and *Arabidopsis* shared more GO-slim terms than modules with asynchronous phases ($F(3,77) = 4.79$, $p < 0.01$, One-Way ANOVA), indicating that these modules in wheat and *Arabidopsis* contain genes with similar functions (Supplementary Fig. 6). We focused on four of these synchronous module pairs, broadly peaking at dawn, midday, dusk, and night for further analysis (Fig. 2). Eigengenes for dawn peaking modules A9 and W9 were highly correlated ($r>0.9$) and shared 14 overlapping enriched GO-slim terms ($p < 0.01$; Fig. 2a,b,f). These included terms for translation and gene expression as well as terms related to protein, amide, nitrogen and organonitrogen biosynthetic and metabolic processes (full lists in Supplementary Table 5). Co-expressed genes in the dawn-expressed modules included several orthologs involved in light, heat, and biological defence, as well as 45 ribosomal protein orthologs (Supplementary Note 5). Transcripts for ribosomal proteins in mouse liver have also been reported to oscillate, suggesting a conserved role for the circadian clock in co-ordinating ribosome biogenesis[31].

In addition to enriched GO-slim terms, we investigated enrichment for transcription factor (TF) superfamilies and transcription factor binding sites in each wheat module. In late-night/dawn modules W8 and W9, transcripts encoding MYB transcription factors were significantly enriched and included putative TFs involved in leaf morphogenesis, plant growth, regulation of flavonoid biosynthesis, and developmental transition to flowering (Supplementary Note 5, Supplementary Fig. 7).

Eigengenes for modules peaking in the day, W3 and A2, had a relatively low correlation (r=0.491) but peaked with similar CT values (CT 6.34h and 6.19) given the longer period length in wheat. 5 out of 15 of the GO-slim terms enriched in the W3 module were also found in the A2 module ($p < 0.01$; yellow triangles in Fig. 2a,c,f). These included terms relating to “photosynthesis”, “response to radiation” and “generation of precursor metabolites and energy”. Co-expressed genes peaking in day-time modules included light-harvesting and

light signalling genes as well as *CYP709B3*, which protects the plant from transpiration-triggered salinity stress during the day[32,33].

In dusk-peaking modules A5 and W4, 8 significantly enriched GO-slim terms were shared ($p < 0.01$, green triangles in Fig. 2a,d,f). Several genes co-expressed in these dusk modules were involved in auxin transport and signalling including the endosomal sorting complex protein *CHMP1A* which ensures proper sorting of auxin carriers (Supplementary Note 5)[34]. There was also a significant enrichment for expression of transcripts encoding AP2-EREBP (ethylene responsive) and ARF (auxin responsive) transcription factor superfamilies within the W4 module (Supplementary Fig. 7). Interestingly, this was followed two hours later by the expression of genes with AP2-EREBP transcription factor binding sites in their promoter region (W5, Supplementary Fig 8).

Finally, two evening-phased modules W5 and A6 (r=0.80) were enriched for GO-slim terms concerning several metabolic processes (blue triangles in Fig. 2f). Co-expressed orthologs in these two modules included *SEVEN IN ABSENTIA2* that regulates ABA-mediated stomatal closure and drought tolerance in *Arabidopsis*[35], and *HYDROPEROXIDE LYASE1* that contributes to responses to insect attack and mechanical wounding[36].

Components of the core circadian clock in *Arabidopsis* and wheat

We next compared the dynamics of circadian oscillator components in wheat and *Arabidopsis*. Clock gene orthologs belonging to large gene families were detected by phylogenetic analysis (Supplementary Fig. 9–13; Supplementary Table 6). Overall, wheat circadian clock genes were expressed rhythmically and with a similar phase to their *Arabidopsis* counterparts (Fig. 3). However, the free-running rhythms of clock transcripts in wheat had a mean circadian period that was approximately 3.49h longer than in *Arabidopsis* (27.23h and 23.74h, respectively).

TaLHY and *TaTOC1* peaked sharply at dawn and dusk, respectively, during the first cycle in constant light, and maintained an >8h difference in phase throughout the experiment (Fig. 3a,b). This is consistent with their phasing in *Arabidopsis.* All three homoeologs for *TaGI* were robustly rhythmic (BH $q < 0.01$) and peaked at CT7 (Fig. 3c). *TaPRR73* transcripts peaked approximately 5h before *TaPRR37* transcripts, consistent with the phase divergence of *Arabidopsis PRR7* and *PRR3* (Fig. 3d,e). However, wheat homoeologs *TaPRR59* and *TaPRR95* had similar expression profiles (R^2=0.68) peaking marginally apart at CT8 and CT10, in between the peak phases of *Arabidopsis PRR9* (CT5) and *PRR5* (CT11) (Fig. 3f,g). Therefore, the *PRR* gene family in wheat peaks in the order of; *TaPRR73*, [*TaPRR37*,

TaPRR95, TaPRR59] in quick succession, and finally *TaTOC1*. This sequential pattern matches the expression of PRR homologs in rice[37].

Transcripts encoding evening complex components, *LUX, ELF3* and *ELF4,* are circadian regulated in *Arabidopsis* and peak simultaneously at dusk. Three wheat triads for LUX-like genes were identified; one with higher identity to *LUX/BOA* and two similar to other *LUX*-like *Arabidopsis* genes (Supplementary Fig. 12). Transcripts from all three of these triads accumulated rhythmically and peaked from midday to dusk, *TaLUX-Lb* at CT7, *TaLUX/BOA* at CT10 and *TaLUX-La* at CT12 (Fig. 3h,i; Supplementary Fig. 14). Five wheat transcripts with homology to *Arabidopsis ELF4* and *ELF4-L1-4* accumulated with a mean circadian phase of 12.6h; similar to *ELF4,* but with lower relative amplitudes (Fig. 3j,k; Supplementary Fig. 14). *TaELF3* transcripts were all arrhythmic (BH $q > 0.36$), and the *TaELF3-1D* homoeolog was expressed with a particularly low baseline level of 0.73 TPM in comparison to the other two homoeologs (8.7-9.7 TPM) (Fig. 3l). This is consistent with previous findings linking a deletion in *TaELF3* to the *eps* QTL on chromosome 1D in Cadenza[18].

We next assessed the balance of circadian expression between triad homoeologs in the core clock network. *TaLHY, TaGI,* and *TaPRR59* had notably similar expression patterns in terms of phase, period, relative amplitudes, and baseline expression over all timepoints (Fig. 3a,c,f,g), suggesting that unbalance in these triads is strongly selected against. Homoeologs of *TaTOC1*, *TaPRR73* and *TaPRR37(Ppd)* had similar phases, but all had reduced expression in the A-genome homoeolog (Fig. 3b,d,e). *TaLUX/BOA-3A*, *TaLUX-La-3B* and *TaLUX-Lb-1D* had marginally shorter periods (>2h) and delayed phases (>2h) compared to their respective homoeologs (Supplementary Table 7).

The REVEILLE family are CCA1/LHY-like MYB-domain transcription factors that are predominantly activators of evening expressed genes[38,39]. The wheat *RVE* genes could be split into a *LHY* clade (containing the *TaLHY* triad described above), a *RVE6/8*-like clade containing three wheat triads and a *RVE1/2/7*-like clade also containing three triads (Supplementary Fig. 9). All *TaRVE6/8* transcripts peaked at CT0-4 concurrently with *TaLHY* (Supplementary Fig. 14). The *TaRVE2/7* transcripts peaked with distinct phases; *TaRVE27b* in phase with *TaLHY*, *TaRVE27c* 4h before *TaLHY* and *TaRVE27a* approximately 12h before *TaLHY* (Supplementary Fig. 14). Based on their phylogenetic relationships, it is probable that several *RVE2/7* clade paralogs in wheat and *Arabidopsis* arose independently after their evolutionary divergence, and it is therefore interesting that they both show distinct phases of expression, suggesting homoplastic circadian functions.

Expression of orthologs for additional transcripts involved in circadian regulation (*FKF1*, *ZTL*, *LKP2*, *LNK1/2*, *CHE* and *LWD*) are reviewed in Supplementary Note 6 and Supplementary Fig. 14.

Circadian control of photosystem and light signalling gene expression is largely conserved between *Arabidopsis* and wheat

A further GO-slim analysis across all rhythmically-expressed genes in *Arabidopsis* and wheat identified enrichment of similar GO-slim processes including "photosynthesis" ($p<1x10-14$), "rhythmic process" ($p<1x10-6$), "response to abiotic stimulus" ($p<1x10-13$) and "cellular macromolecule biosynthetic process" ($p<1x10-5$, Fisher's exact test, Supplementary Table 9). We used genes associated with some of these GO-slim terms as case-studies to highlight similarities and differences in circadian control between the two species. Expression data and Metacycle statistics for all transcripts in this analysis are in Supplementary Table 9.

In considering photosynthesis, we examined specifically nuclear genome-encoded photosystem (PS) proteins. Transcripts encoding the PSI components *LHCA1-6*, the PSI reaction centre subunits *PSAD* and *PSAE* and the *PSII* subunits *LHCB1-7* were rhythmically expressed in both species and had conserved phases (Supplementary Fig. 15). *LHCA1-4* peaked towards the end of the subjective day and *LHCA5* and *6* peaked during the subjective night. *PSAD* and *PSAE* peaked concurrently with *LHCA1-4*. In both species, *LHCB7* transcripts had lower relative amplitudes compared to other LHCB transcripts. PSB27 is a protein associated transiently with the PSII complex involved in adaption to fluctuating light intensities[40]. Transcripts for this protein peaked during the subjective day in *Arabidopsis* and during the subjective night in wheat.

In considering the GO-slim term "response to abiotic stimulus", we next investigated expression of transcripts for photoreceptors and light signalling proteins due to their pervasive influence upon development, metabolism, and circadian timing. Although transcripts for the UV-B photoreceptor *UVR8* accumulated with a circadian rhythm in both species, only one PHYTOCHROME ortholog (*PHYA)* and three *CRYPTOCHROME* orthologs *(CRY1)* were rhythmic in wheat out of 18 orthologs identified (Supplementary Fig. 16). This contrasts with *Arabidopsis*, where *PHYA-C*, *CRY1* and *CRY2* accumulated with a circadian rhythm.

Downstream light signalling proteins COP1 and SPA form complexes that degrade positive regulators of photomorphogenesis (e.g. *HFR1* and *HY5*) under dark conditions[41]. Transcripts

for *COP1*, *SPA4*, *HFR1* and *HY5* accumulated rhythmically and with conserved phases in both species (Supplementary Fig. 16). *COP1/SPA4* peaked synchronously around the end of the subjective night. Surprisingly, given the similar role HFR1 and HY5 proteins have in preventing hypocotyl elongation in low light, *HFR1* and *HY5* transcripts were expressed anti-phase to each other. HY5 and HFR1 act synergistically to coordinate the photomorphogenesis response, although it has been suggested that their activation is regulated through independent pathways[42].

Wheat triads with identity to *Arabidopsis PIN1*, *PIN4*, *PIN5*, and *PIF4/5* were rhythmically expressed, alongside two triads with high similarity to rice *OsPIL11* and *OsPIL13* ([43]; Supplementary Fig. 16). Overall, we observe that the arrhythmic accumulation of most of the wheat PHY and CRY transcripts is not reflected in the rhythmic expression of several downstream light signalling transcripts. This supports the notion that regulatory signals from photoreceptors might occur at the level of protein stability and localisation rather than at the level of transcript accumulation, as occurs for *ZTL* or *HY5* in *Arabidopsis*.

A set of proteins that link light signalling, circadian regulation and chloroplasts are the sigma factors[44]. These light-responsive nuclear-encoded regulators of chloroplast transcription guide promoter recognition and transcription initiation by plastid encoded RNA-polymerase (PEP) on the chloroplast genome[45–48]. In *Arabidopsis*, *SIG1, 3, 4, 5* and *6* were rhythmically transcribed (Supplementary Fig. 16). In wheat, all homoeologs in triads orthologous to *SIG1*, *SIG3* and *SIG5* were also rhythmic (BH $q < 0.01$). Whilst the dawn phase of *TaSIG5* transcripts were similar to *AtSIG5*, *TaSIG1* transcripts were expressed over 10h earlier than *AtSIG1* (Supplementary Fig. 16). Previous research has shown that activity of *AtSIG1* can be regulated through redox-dependent phosphorylation[49], and activity of all sigma factors are likely to be subject to multiple layers of regulation in addition to circadian control of transcript expression.

Similarities and differences in circadian control of primary metabolism genes in *Arabidopsis* and wheat

Expression profiles of genes with key roles in primary metabolism were compared in *Arabidopsis* and wheat with a focus on enzymes that regulate trehalose 6 phosphate (Tre6P) and starch metabolism (Fig. 4). Tre6P synthase (TPS) and Tre6P phosphatase (TPP) participate in the synthesis and dephosphorylation of Tre6P, respectively. Tre6P is an

important signalling metabolite associated with both sucrose regulation and circadian regulation in *Arabidopsis*[50–52]. Tre6P also affects grain yield and drought resilience in wheat, maize, and rice[53]. Transcripts for *TPS1*, *2*, *6*, *8*, *9*, *10* and *11* and *TPPA*, *E*, *F*, *G* and *H* were expressed rhythmically in *Arabidopsis* (Supplementary Fig. 17). Wheat transcripts for *TPS1* (the most well-characterised of the T6P synthases) were arrhythmic, however rhythmic transcripts were found in triads more closely related to TPS11, 6 and 7 (Supplementary Fig. 17). We identified three rhythmic TPP triads in wheat, two of which were orthologous to *Arabidopsis TPPA*, *F* and *G*. The third TPP triad was part of a monocot-specific clade identified by Paul et al. (2018), which also included *Zm00001d032298*, a crop improvement target in maize[53].

Ribulose bisphosphate carboxylase (Rubisco) comprises eight small (RbcS) and 8 large (RbcL) subunits, which are encoded by the nuclear and chloroplast genomes, respectively[54]. Rubisco requires activation by Rubisco activase (RCA) to release its activity from inhibitory substrates[55]. In our wheat expression data, 22 putative wheat orthologs for the small subunit of Rubisco were rhythmic, peaking during the subjective night, as *RBCS1A*, *RBCS1B*, *RBCS2B* and *RBCS3B* do in *Arabidopsis* (Supplementary Fig. 17). Two triads with identity to Rubisco activase were identified, one of which accumulated rhythmically (peaking at CT0, as with *Arabidopsis RCA*).

Circadian regulation has a pervasive influence on starch metabolism in *Arabidopsis*, particularly the nocturnal rate of transitory starch degradation[23,56]. Chloroplast phospho-glucose isomerase 1 (*PGI1*) and chloroplast phosphoglucose mutase (*PGM1*) are essential enzymes that link the Calvin-Benson cycle with starch biosynthetic pathway[57–59]. In *Arabidopsis*, these transcripts accumulated with a circadian rhythm (BH $q < 1x10^{-4}$); *PGM1* peaked just after dusk (CT14), and *PGI1* slightly later at CT20. In contrast, only one wheat *TaPGI1* homoeolog was rhythmic (BH $q < 0.01$), which had a low relative amplitude (0.16) and a peak phase of CT8. No homoeologs for *TaPGM1* were rhythmically expressed (BH q > 0.01, Supplementary Fig. 17).

ADP-glucose pyrophosphorylase (AGPase) mediates the first irreversible and rate-limiting step in starch biosynthesis through the formation of ADP-Glc. In *Arabidopsis*, transcripts encoding the small and large subunits of AGPase (*APL1*, *APL2*, *APL3*, *APS1*) were rhythmic, peaking at night around CT20. In comparison, in wheat only two of the eleven transcripts with homology to *APL1*, *APL2* and *APS1* were rhythmic (BH $q < 0.01$), with the remaining transcripts lacking a discernible rhythm (BH $q > 0.05$) (Supplementary Fig. 17).

Starch synthases (SS) represent another group of metabolically important enzymes that use the glucose from ADP-Glc to elongate glucan chains. In *Arabidopsis*, there are five types: SSI, SSII, SSIII, SSIV and granule bound GBSSI. SSI-IV are responsible for synthesis of amylopectin, with SSIII and IV determining starch granule number and morphology[60]. *GBSSI* is a known dawn-expressed gene, regulated directly by CCA1/LHY, specialised for amylose synthesis[61]. In wheat, *GBSSI* orthologs are called *TaWaxy* and cultivars with three null alleles produce amylose-free starch in their grain[62]. Comparison of starch synthase expression in *Arabidopsis* and wheat revealed several differences between the phases and relative amplitudes of these transcripts (Supplementary Fig. 18). In *Arabidopsis, GBSSI* transcripts had by far the greatest relative amplitude (1.26) with peak expression at dawn. The next greatest amplitudes were of *SSIV* transcripts, which peaked at CT17. *SSII* and *SSIII* peaked together at CT21 and *SS1* peaked at CT8 with a much smaller amplitude (0.12). In contrast, in wheat, an *SSIII* triad (*TaSSIIIb*) had the largest relative amplitude rhythms of the wheat starch synthases identified (0.64 - 0.73). Wheat transcripts for *SSI* and *SSIV* also peaked in the morning, whereas wheat *SSII* peaked instead in the subjective night (~CT15). In our data, *TaWaxy (GBSSI)* transcripts were present at a very low baseline level (<0.01 TPM) and without any circadian oscillation. However, another wheat triad, *TaGBSSII,* shared >62% identity with *Arabidopsis GBSSI,* and the B and D homoeologs had rhythmic expression which peaked at dawn. *TaWaxy* and *TaGBSSII* are specific to endosperm and leaf tissues, respectively[63], which might explain the distribution of transcript accumulation seen here. We can conclude that the circadian clock regulates the expression of SS transcripts in both *Arabidopsis* and wheat, although there might be an emphasis on different types of SS in each species.

The *Arabidopsis* circadian clock regulates the rate of starch degradation so that starch reserves are depleted precisely at subjective dawn[23]. Many transcripts encoding starch-degrading enzymes in *Arabidopsis* had synchronized dusk peaks: Isoamylase□type starch debranching enzyme *ISA3*; alpha-amylase *AMY3*; plastidial phosphorylase *PHS1-2*; disproportionating enzymes *DPE1-2*; glucan, water dikinases *GWD1* and *PWD* and glucan phosphatase *SEX4. Arabidopsis* transcripts for *BAM3*, *BAM5* and *PU1* also oscillated with a circadian rhythm, peaking later in the subjective night. Strikingly, wheat orthologs for several of these genes were not rhythmic, including *AMY3*, *DPE1*, *PWD*, *PHS1*, *PU1* and *BAM1*. Wheat orthologs for *ISA3* and *DPE2* were expressed rhythmically, but peaked approximately 8-12 h ahead of their *Arabidopsis* counterparts. Some starch degradation enzymes had conserved circadian expression patterns in the two species, such as *SEX4*, *GWD1*, *BAM3* and

BAM5 transcripts. GWD catalyses glucan phosphorylation and *SEX4* encodes a phosphoglucan phosphatase, both of which facilitate hydrolytic attack by β-amylases (BAM) in the early steps of starch degradation[56,64].

Discussion

Conservation of circadian regulation between homoeologous genes

We found a large proportion of unbalanced circadian triads in our dataset. Most of these were unbalanced due to arrhythmicity in one or two homoeologs expressed at a lower mean level than the rhythmic homeologs. The reduction of expression could be due to constitutive epigenetic silencing or changes to promoter regions, allowing differential binding of transcription factors[65–67]. We found additional circadian unbalance in the form of altered phase, period, and relative amplitudes. It is possible that some of these differences are due to retention of circadian regulation from the ancestral genome of each homeolog (Fig. 1a), although it is likely that other differences reflect more recent diversification in expression as a step towards neo-functionalisation. It has been previously suggested that functional divergence is a likely fate for duplicated genes in a sufficiently large population[68]. In *B. rapa*, 42% of circadian controlled paralogs had differential expression patterns[28], however these paralogs arose through whole genome duplication events around 13-43 million years ago, so have been exposed to longer periods of time during which selection could act upon these duplicate genes[69]. In comparison, specialisation of circadian homeologs in wheat could be comparatively lower due to the relative infancy of its polyploidisation.

Differences between periods of rhythmic transcripts in *Arabidopsis* and wheat

The mean period of circadian regulated genes in wheat was over three hours longer than in *Arabidopsis*. Period length is affected by a range of exogenous conditions (e.g. light and temperature), and varies between tissues and plant age[70]. There is also evidence that longer periods have been selected for during cultivation of crops at higher latitudes[1,2,71], potentially due to enhanced seasonal tracking capability enabling precision timing of growth and flowering[72]. Compared to other plant circadian transcriptome data sets, rhythmic wheat transcripts also had higher period variance (Fig 1b). The broad period distribution in wheat might be due to inclusion of all aerial material in our sampling strategy. Variation in free-

running periods could occur at the organ-, tissue- or cellular-level, and transcripts which are highly expressed in those regions may reflect those period differences[73,74]. An alternative possibility is that period variation is due to uncoupling of multiple circadian oscillators within the same cell which control expression of subsets of transcripts[75–77]. Future research could examine the relationship between period distributions of circadian transcriptomes and the effects of domestication, latitudinal adaption, monocot-dicot divergence, or polyploidy.

Similarities and differences in circadian regulation between wheat and *Arabidopsis*

Our analysis revealed extensive conservation of time-of-day specific GO-slim processes and co-expressed genes between *Arabidopsis* and wheat. These included genes involved in photosynthesis (e.g., photosystem proteins), light signalling (e.g., *HFR1, HY5, PINs* etc*),* translation (e.g., ribosome proteins) and auxin and ethylene responsive transcription factors. However, we also identified several interesting differences between the two species, including absence of rhythmicity in wheat *PHY* and *CRY* transcripts and antiphase expression of the wheat sigma factor *SIG1*. Furthermore, we found differences in rhythmic expression of many transcripts involved in regulating Tre6P and starch metabolism.

In our data, putative wheat homeologs of *TPS1* were arrhythmic. Instead, rhythmic *TPS* transcripts in wheat had similarity to *Arabidopsis TPS11*, *6* and *7* (Supplementary Fig. 18). In *Arabidopsis*, TPS1 is the most catalytically active and best characterised TPS, and feeds back into the entrainment of the circadian clock[51,78]. If the lack of rhythmicity in wheat *TPS1* transcripts is reflected at the level of protein activity, it may indicate that Tre6P synthesis is not regulated as tightly by the circadian clock in wheat as in *Arabidopsis*. On the other hand, circadian control of other TPS triads may have implications for biotic or abiotic defence in wheat. TPS5-11 have been previously implicated in control of stomatal aperture[79], thermotolerance[80], and defence against fungal, bacterial and aphid attack[81,82]. In rice, *OsTPS8* influences drought resistance through suberin deposition[83], and wheat *TaTPS11* participates in a cold stress response[84].

In wheat, transcripts for starch degradation enzymes (*PHS1, DPE1, BAM1, PU1, AMY3, PWD*) and starch biosynthesis enzymes (*PGI1*, *PGM1, ISA1* and *ATPase*) had either arrhythmic expression or low relative amplitudes compared with the robust rhythms of many of these transcripts in *Arabidopsis.* Additionally, *ISA2*, *ISA3* and the starch synthases (*SSI-IV*) had differing circadian phases between the two species. While it is possible that rhythmic expression of a reduced number of genes (e.g.: *SEX4*, *GWD1, BAM3* and *BAM5)* is sufficient to mediate circadian control of starch degradation in wheat, these data suggest that the

455 circadian clock has a less pervasive influence upon transcriptional control of starch
456 metabolism in wheat compared to *Arabidopsis*.
457
458 Conclusions
459 Our data reveal the influence of circadian regulation on the wheat transcriptome and highlight
460 several intriguing differences between rhythmically expressed transcripts in *Arabidopsis* and
461 wheat. It explores the added complexity of co-ordinating circadian expression across multiple
462 sub-genomes in a hexaploid species. Given the circadian clock has been under selection
463 during domestication and presents multiple targets for crop improvement, it is likely that this
464 new insight into the clock in wheat will be important in the development of new sustainable
465 and resilient cultivars. It is our hope that these data provide a resource for identifying target
466 genes regulated by the circadian clock, allowing the relationships between chronotype, yield
467 and resilience to be explored in future studies.
468

469 Methods

470 Plant Materials and Growth Conditions

471 Wheat: Wheat seeds cv. Cadenza were imbibed for three days on damp filter paper on a Petri
472 dish at 4°C. Plates were moved at dawn (06.00 = ZT0), to a growth cabinet set to 22°C under
473 12:12 light: dark cycles (approximately 200 µmol m−2 s−1). After two days, only seedlings
474 with fully emerged radicles were sown, 3 cm deep in Petersfield cereal mix in 9cm pots.
475 Seedlings were grown under 12hlight:12hdark conditions for 14 days. After 14 days, at dawn
476 (ZT0) seedlings were transferred to constant light conditions (L: L), tissue was sampled every
477 4h for 3 days (18 samples in total). At each timepoint, we sampled the entire aerial
478 tissue from 3 replicate plants, which was frozen immediately in liquid nitrogen before storage
479 at -80°C. Total RNA was extracted using Qiagen RNeasy plant mini kits (cat. no. 74904)
480 with on-column DNAse treatment (RNAse-Free DNase Set (cat. no. 79254). RNA
481 concentration and integrity were quantified using a Nanodrop Spectrophotometer and Perkin
482 Elmer LabChip GX Nucleic acid analyser before sequencing.
483 Details of growth conditions for Arabidopsis[26], Brassica rapa[28], Brachypodium distachyon[85]
484 and Glycine max[86] datasets can be viewed in their source manuscripts. Briefly, all circadian
485 data were measured under constant light and temperature following 12h:12h light:dark
486 entrainment other than Glycine max[86] which was entrained under 16h:8h light:dark cycles.

487 Wheat mRNA sequencing, read alignment and quantification

488 Library preparation was carried out following the Illumina TruSeq protocol and reads were
489 sequenced on a NovaSeq S2 flow cell at the Earlham Institute. 150bp paired-end reads were
490 generated from each library to an average depth of 84M reads per replicate. Reads were
491 filtered for quality and any remaining adaptor sequence was trimmed with Trimmomatic[87].
492 Surviving reads were aligned to the Chinese Spring RefSeq v1.1 wheat genome[11] using
493 HISAT2[88] with default parameters. Uniquely mapping reads were then quantified using
494 StringTie[89] and TPM values were extracted for each gene per sample.

495 Processing and quantification of previously published datasets

496 Raw reads from previously published circadian datasets were downloaded for Arabidopsis[26],
497 Brassica rapa[28], and Brachypodium distachyon[85]. These reads were filtered for quality, and
498 any remaining adaptor sequence trimmed with Trimmomatic[87]. Surviving reads were aligned
499 using HISAT2[88] to A. thaliana genome (TAIR 10), B. rapa genome (v1.0) and the B.
500 distachyon genome (v3.0) respectively. For the Arabidopsis alignment, maximum intron
501 length was set to 5000nt consistent with pre-processing in[26,90]. StringTie[89] was used to
502 quantify uniquely mapping reads before TPM value extraction at gene level. For Glycine
503 max[86], FPKM normalised reads were downloaded from the Glycine Max RNA-seq
504 Database[91] (accession GSE94228) and were converted from FPKM to TPM prior to analysis.
505

506 Homolog identification of circadian clock and circadian controlled genes

507 Wheat homologs of *Arabidopsis* core circadian clock genes were identified in the wheat
508 genome by detecting similarity to the following conserved protein family domains that are
509 present in the proteins encoded by these genes: MYB1R, a subtype of MYB domain that
510 contains a distinctive 'SHAQKY' sequence motif (present in the CCA1, LHY and RVE[1-8])
511 or a distinctive 'SHLQKY' sequence motif (present in LUX), PAS (present in ZTL), PRR
512 (present in TOC1 and PRR[3579]) and ELF4 (present in ELF4). A hidden Markov model
513 (HMM) for each domain was used in HMMER 3.2.1 HMMSEARCH[92] to search for
514 members of the domain family in the following proteome datasets: Araport11 (*Arabidopsis*
515 *thaliana*), RGAP7 (*Oryza sativa*), JGI Phytozome version 12 (*Brachypodium distachyon*),
516 IBSC (*Hordeum vulgare*), SpudDB PGSC v4.03 (*Solanum tuberosum*) and IWGSC Refseq
517 v1.1 (*Triticum aestivum*). The HMMs provided by Pfam (https://pfam.xfam.org/) were used

for the PAS domain (PAS_9, PF13426), the PRR domain (Response_reg, PF00072) and the ELF4 domain (PF07011). For the MYB domain, an HMM was built for the MYB1R subfamily using HMMER3 HMMBUILD[92] with an alignment of protein sequences from *Arabidopsis* and rice, previously established as being members of this subfamily. The sequences found from these genomes were re-aligned to the original alignment using HMMER 3.2.1 HMMALIGN[92]. Amino acids with non-match states in the HMM were removed from the alignment and alignment columns with <70% occupancy were also removed. The longest splice variant of each protein was selected to estimate a phylogenetic tree with bootstrap support using RAxML 8.2.12[93] with the following method parameters set: -f a, -x 12345, -p 12345, -# 100, -m PROTCATJTT. The trees were mid-point rooted and images created using the Interactive Tree of Life (iToL) tool[94]. For the larger MYB and PRR families, proteins from the tree clades containing known clock gene(s) were re-aligned across their full-length and a "nested" phylogenetic tree was re-estimated with RAxML as described above. The tree was visualised in the Interactive Tree Of Life (iTOL) website alongside the corresponding alignment. This view provided increased detail about the relationships within the clade and enabled orthologous sequences to be inferred. Wheat homologues for *ELF3*, *GI*, *LWD1/2*, *CHE*, and *LNK1/2* were identified by BLASTP searches using previously identified wheat and *Brachypodium* predicted proteins confirmed by reciprocal BLAST searches against *Arabidopsis*. IDs and source references can be viewed in Supplementary Table 6.

Putative wheat orthologs for *Arabidopsis* circadian controlled pathway genes involved in photosynthesis, light-signalling and primary metabolism were first extracted using Biomart v0.7[95] available from Ensembl Plants and taken forward if they had >40% identity in the DNA sequence. Orthologs were then verified using BLASTP using *Arabidopsis* protein sequences as a query against the wheat protein database to confirm the wheat gene IDs. Complete lists of wheat gene IDs used in the pathway analysis can be viewed in Supplementary Table 9.

Circadian quantification using Metacycle and Biodare2

To estimate proportions of rhythmic genes expressed in *Arabidopsis* and wheat, we removed only genes with 0 TPM at all timepoints. This approach has been used in several previous studies[26,96,97] and allows detection of low-expression rhythmic transcripts. An analysis of how

filtering for low-expression genes affects the estimates of proportions of rhythmically expressed genes is discussed in Supplementary Note 1 and Supplementary Table 1.

The R package MetaCycle[98] was used to identify rhythmically expressed transcripts (Benjamini-Hochberg *q*-values) and to quantify period lengths (hours), absolute phase (hours), baseline expression (TPM), amplitudes (TPM) and relative amplitudes of circadian waveforms. Relative amplitude is the ratio between amplitude and baseline TPM if the baseline is greater than 1. Metacycle integrates results from three independent algorithms (ARSER, JTK_CYCLE and Lomb-Scargle) to produce summary "meta2d" statistics that combine the outcome from these algorithms. Metacycle was run using the following parameters; minper = 12, maxper = 35, adjustPhase = "predictedPer". Transcripts were defined as rhythmic if they had *q*-values < 0.05 and high confidence rhythmic transcripts if they have *q*-values < 0.01. To calculate circadian phase (CT; relative to period length=24), meta2d phase estimates were multiplied by 24 and then divided by the period estimates for each transcript. Circular phase means were calculated using the package 'circular' implemented in R[99].

There are many different algorithms available for quantification of rhythmicity within time-series data, some of which perform better on datasets with higher levels of noise, non-24h periods or various sampling frequencies. To validate the meta2d results we also used the FFT-NLLS and MESA algorithms implemented in Biodare2 to verify our observations about period, phase and rhythmicity[100]. FFT-NLLS also provides relative amplitude error (RAE) statistics which represent a useful metric for assessing rhythmic robustness. FFT-NLLS and MESA were run using the BH $q < 0.01$ filtered transcripts categorized in Metacycle, and with the following parameters: no dtr, min-max, p(12.0-35.0).

To enable as close a comparison with the *Arabidopsis* dataset as possible, the wheat time-course was cropped to a data window of 24-68h for approximation of period, phase and relative amplitude unless specified otherwise. This data-window also ensures that measurements are being made under circadian conditions following transfer to constant light. For the triad analysis, meta2d estimates were measured over the full time-course (0-68h) as differentiation of homeolog behaviour was the main interest, including the response to transfer to L:L.

Clustering of rhythmic genes into expression modules

Gene co-expression analysis was carried out using the R package WGCNA (Langfelder and Horvath, 2008; R version 3.6.0.).

Arabidopsis: The 10,317 genes identified by MetaCycle as significantly rhythmic (q-value < 0.01) were filtered and genes with greater than 0.5 TPM average expression at more than three timepoints were retained for further analysis. The average expression at each timepoint for the remaining 10,129 genes was used to construct signed hybrid networks on a replicate basis using the blockwiseModules() function. The soft power threshold was calculated as 18, and the following parameters were used; minModuleSize = 30, corType = bicor, maxPOutliers = 0.05, mergeCutHeight = 0.15. Highly connected hub genes were identified for each of the 9 co-expression modules using the function chooseTopHubInEachModule() and eigengenes were identified for each module using the moduleEigengenes() function.

Wheat: The 18,633 genes identified by MetaCycle as significantly rhythmic across 12 timepoints ZT24 - ZT68 (q-value < 0.01) were filtered and genes with greater than 0.5 TPM average expression at more than three timepoints were retained for further analysis. The average expression at each timepoint for the remaining 16,327 genes was used to construct signed hybrid networks using the blockwiseModules() function. A soft power threshold of 18 was used, together with the following parameters; minModuleSize = 30, corType = bicor, maxPOutliers = 0.05, mergeCutHeight = 0.15. Eigengenes were identified for each module using the moduleEigengenes() function. Modules with closely correlated eigengenes were merged using the mergeCloseModules() function, with the parameters; cutHeight = 0.25, iterate = F) and new module eigengenes were calculated for the resulting 9 modules.

Cross-correlation analysis

A cross-correlation analysis was used to find the shift in time (lag) which produced the highest (peak) correlation between two rhythms. This approach was used to identify modules which peaked synchronously (had a peak lag of 0h) or asynchronously (had a peak lag of 4, 8 or 12h) by correlating eigengenes for each module (Supplementary Fig. 6). We also used cross-correlation to identify unbalanced phases within rhythmic triads (Fig. 1E). Before calculating the cross-correlation between two expression rhythms, we first scaled both expression patterns using their means and standard deviations, so the output reflects a time-dependent Pearson correlation coefficient ranging between -1 and 1:

$$Z_A = \frac{X_A - \bar{X}_A}{S_A}, Z_B = \frac{X_B - \bar{X}_B}{S_R}$$

Where Z_i, X_i, $\bar{X}_i$ and S_i represent the standardised expression level, tpm expression level, mean expression level, and standard deviation of gene A and B respectively. Once both expression patterns have been scaled, the discrete cross-correlation between the two expression patterns is calculated using the np.correlate function and is divided by the number of time points in the expression signal returning the Pearson correlation coefficient at different lags. The index of the array with the largest Pearson correlation coefficient score corresponds to the lag that maximises the phase similarity between the two temporal expression patterns.

Mean-normalised data for oscillation plots

Oscillation plots in Supplementary Fig. 15-17 were mean normalised to aid visualisation of period and phase differences between transcripts. Data was adjusted by dividing the TPM values at each timepoint by the mean across all timepoints for each gene so that the baseline expression was equal to 1.

Gene ontology term enrichment

Functional enrichment of differentially expressed genes for biological processes within each module was performed using the gene ontology enrichment analysis package, topGO[101] in R (version 3.6.0, with the following parameters: nodeSize = 10, algorithm = "parentchild", classicFisher test $p < 0.05$). Enrichment of terms in all rhythmic genes in *Arabidopsis* and wheat was compared against a background 'gene universe' of all expressed genes in each dataset (26,392 genes for *Arabidopsis* and 86,567 for wheat). This gene universe was also used in the GO-slim analysis for enrichment in circadian balanced versus unbalanced triads. Enrichment of terms in expression modules was compared against a background of all rhythmically expressed genes (BH $q < 0.01$) which clustered into modules in each dataset (10,129 genes for *Arabidopsis* and 16,327 for wheat). GO-slim terms refer to ontology terms for biological processes unless otherwise specified and were obtained from Ensembl Plants 51[11], using the BioMart tool. The bubble plot was plotted using ggplot in R adapting code from (De Vega et al., 2021).

641 Enrichment analysis of transcription factor superfamilies in wheat co-expression modules

642 Genes annotated as members of transcription factor superfamilies[27] were identified in each
643 co-expression module and the frequency of each TF superfamily compared to the frequency
644 observed in the 16,327 genes submitted to WGCNA. TF families were classed as either
645 significantly under or overrepresented in each module using Fisher's exact test ($p <= 0.05$).
646

647 Enrichment analysis of transcription factor binding sites in wheat co-expression modules

648 1.5 kb of sequence upstream of the transcription start site (TSS) was extracted for each of the
649 16,327 genes submitted to WGCNA. FIMO, from the MEME tool suite (v 4.11.1) was used
650 to predict TFBS in these regions based on similarity with previously DAP-seq validated
651 TFBS identified in *Arabidopsis*[102]. FIMO was run as reported in Ramírez-González et al.,
652 2018 (p-value threshold of <1e-04 (default), --motifpseudo set to 1e-08 as recommended for
653 use with PWMs and a --max-stored-scores of 1,000,000). The background model was
654 generated from the 16,327 promoter sequences using MEME fasta-get-markov. As the
655 significance of multiple matches of a single TFBS family in the putative promoter region for
656 each gene is unknown, we derived a non-redundant (nr) list of matched TFBS motifs for each
657 gene within each of the nine modules and for the complete set of 16,327 genes, where
658 multiple occurrences of a TFBS superfamily in a single promoter sequence were only
659 counted once. The frequency of these nrTFBS motifs for each co-expression module was
660 compared to the frequency of nrTFBS seen across all 16,327 genes and families significantly
661 under or overrepresented in each module were identified using Fisher's exact test ($p <= 0.05$).
662

663 Statistical analysis

664 Statistical tests including Welch's two sample t-test, Two-proportions z-test, One-way
665 ANOVA, Two-level, nested ANOVA and Chi-square tests of independence were all
666 conducted in the R 'stats' package (version 4.0.0) with default parameters.
667

668 Data availability

669 Fastq data from the RNA-seq circadian time course are available to view from the Grassroots
670 Data Repository: https://opendata.earlham.ac.uk/opendata/data/wheat_circadian_Rees_2021.

[Data will be uploaded to the European Nucleotide Archive (ENA) during the review process]

References

1. Greenham, K. *et al*. Geographic variation of plant circadian clock function in natural and agricultural settings. *J. Biol. Rhythms* **32**, 26–34 (2017).

2. Müller, N. A. *et al*. Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Genet*. **48**, 89–93 (2016).

3. McClung, C. R. Circadian Clock Components Offer Targets for Crop Domestication and Improvement. *Genes (Basel)*. **12**, 374 (2021).

4. Steed, G., Ramirez, D. C., Hannah, M. A. & Webb, A. A. R. Chronoculture, harnessing the circadian clock to improve crop yield and sustainability. *Science (80-.).* **372**, (2021).

5. Hazen, S. P. *et al*. LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10387–10392 (2005).

6. Nakamichi, N., Kita, M., Ito, S., Yamashino, T. & Mizuno, T. PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, Together play essential roles close to the circadian clock of Arabidopsis thaliana. *Plant Cell Physiol*. **46**, 686–698 (2005).

7. Somers, D. E., Webb, A. A. R., Pearson, M. & Kay, S. A. The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in Arabidopsis thaliana. *Development* **125**, 485–494 (1998).

8. Wang, Z. Y. & Tobin, E. M. Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**, 1207–1217 (1998).

9. Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R. & Martin, W. Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics* vol. 3 429–441 (2002).

10. Araus, J. L., Slafer, G. A., Romagosa, I. & Molist, M. FOCUS: Estimated wheat yields during the emergence of agriculture based on the carbon isotope discrimination of grains: Evidence from a 10th millenuim BP site on the euphrates. *J. Archaeol. Sci*. **28**, 341–350 (2001).

11. IWGSC *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science (80-.).* **361**, (2018).

12. Hernando, C. E., Romanowski, A. & Yanovsky, M. J. Transcriptional and post-transcriptional control of the plant circadian gene regulatory network. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860**, 84–94 (2017).

13. Chaw, S. M., Chang, C. C., Chen, H. L. & Li, W. H. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol*. **58**, 424–441 (2004).

14. Turner, A., Beales, J., Faure, S., Dunford, R. P. & Laurie, D. A. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science (80-.).* **310**, 1031–1034 (2005).

15. Shaw, L. M., Turner, A. S. & Laurie, D. A. The impact of photoperiod insensitive Ppd-1a mutations on the photoperiod pathway across the three genomes of hexaploid wheat (Triticum aestivum). *Plant J.* **71**, 71–84 (2012).

16. Laurie, D. A. Comparative genetics of flowering time. *Plant Mol. Biol.* **35**, 167–177 (1997).

17. Xiang, Y. Z., Mao, S. L., Jia, R. L., Chun, M. G. & Xian, S. Z. The wheat TaGI1, involved in photoperiodic flowering, encodes an Arabidopsis GI ortholog. *Plant Mol. Biol.* **58**, 53–64 (2005).

18. Zikhali, M., Wingen, L. U. & Griffiths, S. Delimitation of the Earliness per se D1 (Eps-D1) flowering gene to a subtelomeric chromosomal deletion in bread wheat (Triticum aestivum). *J. Exp. Bot.* **67**, 287–299 (2016).

19. Wang, J. *et al*. TaELF3-1DL, a homolog of ELF3, is associated with heading date in bread wheat. *Mol. Breed.* **36**, 1–9 (2016).

20. Mizuno, N., Nitta, M., Sato, K. & Nasuda, S. A wheat homologue of PHYTOCLOCK 1 is a candidate gene conferring the early heading phenotype to einkorn wheat. *Genes Genet. Syst.* **87**, 357–367 (2012).

21. Zhang, W. *et al*. Functional studies of heading date-related gene TaPRR73, a paralog of ppd1 in common wheat. *Front. Plant Sci.* **7**, 772 (2016).

22. Dodd, A. N. *et al*. Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science (80-.).* **309**, 630–633 (2005).

23. Graf, A., Schlereth, A., Stitt, M. & Smith, A. M. Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9458–9463 (2010).

24. Harmer, S. L. *et al*. Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science (80-.).* **290**, 2110–3 (2000).

25. Millar, A. J. & Kay, S. A. Circadian Control of cab Gene Transcription and mRNA Accumulation in Arabidopsis. *Plant Cell* **3**, 541–550 (1991).

26. Romanowski, A., Schlaen, R. G., Perez-Santangelo, S., Mancini, E. & Yanovsky, M. J. Global transcriptome analysis reveals circadian control of splicing events in Arabidopsis thaliana. *Plant J.* **103**, 889–902 (2020).

27. Ramírez-González, R. H. *et al*. The transcriptional landscape of polyploid wheat. *Science (80-.).* **361**, (2018).

28. Greenham, K. *et al*. Expansion of the circadian transcriptome in Brassica rapa and genome-wide diversification of paralog expression patterns. *Elife* **9**, 1–68 (2020).

29. Zhao, J. *et al*. Genome-Wide Identification and Expression Profiling of the TCP Family Genes in Spike and Grain Development of Wheat (Triticum aestivum L.). *Front. Plant Sci.* **9**, (2018).

30. Takumi, S., Shimamura, C. & Kobayashi, F. Increased freezing tolerance through up-regulation of downstream genes via the wheat CBF gene in transgenic tobacco. *Plant Physiol. Biochem.* **46**, 205–211 (2008).

31. Jouffe, C. *et al*. The Circadian Clock Coordinates Ribosome Biogenesis. *PLoS Biol.* **11**, 1001455 (2013).

32. Mao, G., Seebeck, T., Schrenker, D. & Yu, O. CYP709B3, a cytochrome P450 monooxygenase gene involved in salt tolerance in Arabidopsis thaliana. *BMC Plant Biol.* **13**, 1–13 (2013).

33. Pan, Y. *et al*. Cytochrome P450 monooxygenases as reporters for circadian-regulated pathways. *Plant Physiol.* **150**, 858–878 (2009).

34. Spitzer, C. *et al*. The ESCRT-Related CHMP1A and B Proteins Mediate Multivesicular Body Sorting of Auxin Carriers in Arabidopsis and Are Required for Plant Development. *Plant Cell* **21**, 749–766 (2009).

35. Bao, Y. *et al*. The tumor necrosis factor receptor-associated factor (TRAF)-like family protein SEVEN IN ABSENTIA 2 (SINA2) promotes drought tolerance in an ABA-dependent manner in Arabidopsis. *New Phytol.* **202**, 174–187 (2014).

36. Noordermeer, M. A., Veldink, G. A. & Vliegenthart, J. F. G. Fatty acid hydroperoxide lyase: A plant cytochrome P450 enzyme involved in wound healing and pest resistance. *ChemBioChem* **2**, 494–504 (2001).

37. Murakami, M., Ashikari, M., Miura, K., Yamashino, T. & Mizuno, T. The Evolutionarily Conserved OsPRR Quintet: Rice Pseudo-Response Regulators Implicated in Circadian Rhythm. *Plant Cell Physiol.* **44**, 1229–1236 (2003).

38. Shalit-Kaneh, A., Kumimoto, R. W., Filkov, V. & Harmer, S. L. Multiple feedback loops of the Arabidopsis circadian clock provide rhythmic robustness across environmental conditions. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7147–7152 (2018).

39. Hsu, P. Y., Devisetty, U. K. & Harmer, S. L. Accurate timekeeping is controlled by a cycling activator in Arabidopsis. *Elife* **2013**, 473 (2013).

40. Hou, X., Fu, A., Garcia, V. J., Buchanan, B. B. & Luan, S. PSB27: A thylakoid protein enabling Arabidopsis to adapt to changing light intensity. *Proc. Natl. Acad. Sci.* **112**, 1613–1618 (2015).

41. Ponnu, J. & Hoecker, U. Illuminating the COP1/SPA Ubiquitin Ligase: Fresh Insights Into Its Structure and Functions During Plant Photomorphogenesis. *Front. Plant Sci.* **0**, 486 (2021).

42. Jang, I.-C., Henriques, R. & Chua, N.-H. Three Transcription Factors, HFR1, LAF1 and HY5, Regulate Largely Independent Signaling Pathways Downstream of Phytochrome A. *Plant Cell Physiol.* **54**, 907–916 (2013).

43. Nakamura, Y., Kato, T., Yamashino, T., Murakami, M. & Mizuno, T. Characterization of a set of phytochrome-interacting factor-like bHLH proteins in Oryza sativa. *Biosci. Biotechnol. Biochem.* **71**, 1183–1191 (2007).

44. Noordally, Z. B. *et al.* Circadian Control of Chloroplast Transcription by a Nuclear-Encoded Timing Signal. *Science (80-.).* **339**, 1316–1319 (2013).

45. Tanaka, K. *et al.* Characterization of three cDNA species encoding plastid RNA polymerase sigma factors in Arabidopsis thaliana: Evidence for the sigma factor heterogeneity in higher plant plastids. *FEBS Lett.* **413**, 309–313 (1997).

46. Kanamaru, K. *et al*. Plastidic RNA polymerase σ factors in Arabidopsis. *Plant Cell Physiol.* **40**, 832–842 (1999).

47. Allison, L. A. The role of sigma factors in plastid transcription. *Biochimie* vol. 82 537–548 (2000).

48. Kanamaru, K. & Tanaka, K. Roles of chloroplast RNA polymerase sigma factors in chloroplast development and stress response in higher plants. *Biosci. Biotechnol. Biochem.* **68**, 2215–2223 (2004).

49. Shimizu, M. *et al.* Sigma factor phosphorylation in the photosynthetic control of photosystem stoichiometry. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10760–10764 (2010).

50. Figueroa, C. M. & Lunn, J. E. A tale of two sugars: Trehalose 6-phosphate and sucrose. *Plant Physiology* vol. 172 7–27 (2016).

786 51. Frank, A. *et al*. Circadian Entrainment in Arabidopsis by the Sugar-Responsive Transcription Factor
787 bZIP63. *Curr. Biol*. **28**, 2597-2606.e6 (2018).
788 52. Fichtner, F. & Lunn, J. E. The Role of Trehalose 6-Phosphate (Tre6P) in Plant Metabolism and
789 Development. *Annu. Rev. Plant Biol.* **72**, 737–760 (2021).
790 53. Paul, M. J., Gonzalez-Uriarte, A., Griffiths, C. A. & Hassani-Pak, K. The Role of Trehalose 6-
791 Phosphate in Crop Yield and Resilience. *Plant Physiol.* **177**, 12–23 (2018).
792 54. Gruber, A. V. & Feiz, L. Rubisco assembly in the chloroplast. *Frontiers in Molecular Biosciences* vol.
793 5 24 (2018).
794 55. Bhat, J. Y., Thieulin-Pardo, G., Hartl, F. U. & Hayer-Hartl, M. Rubisco activases: AAA+ chaperones
795 adapted to enzyme repair. *Frontiers in Molecular Biosciences* vol. 4 20 (2017).
796 56. Smith, S. M. *et al.* Diurnal changes in the transcriptome encoding enzymes of starch metabolism
797 provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in arabidopsis
798 leaves. *Plant Physiol*. **136**, 2687–2699 (2004).
799 57. Periappuram, C. *et al*. The Plastidic Phosphoglucomutase from Arabidopsis. A Reversible Enzyme
800 Reaction with an Important Role in Metabolic Control. *Plant Physiol.* **122**, 1193 (2000).
801 58. Bahaji, A. *et al*. Plastidic phosphoglucose isomerase is an important determinant of starch
802 accumulation in mesophyll cells, growth, photosynthetic capacity, and biosynthesis of plastidic cytokinins in
803 Arabidopsis. *PLoS One* **10**, e0119641 (2015).
804 59. Lee, S. K. *et al*. Plastidic phosphoglucomutase and ADP-glucose pyrophosphorylase mutants impair
805 starch synthesis in rice pollen grains and cause male sterility. *J. Exp. Bot.* **67**, 5557–5569 (2016).
806 60. Nielsen, M. M. *et al*. Crystal structures of the catalytic domain of Arabidopsis thaliana starch synthase
807 IV, of granule bound starch synthase from CLg1 and of granule bound starch synthase I of Cyanophora
808 paradoxa illustrate substrate recognition in starch synthases. *Front. Plant Sci*. **9**, 1138 (2018).
809 61. Tenorio, G., Orea, A., Romero, J. M. & Mérida, Á. Oscillation of mRNA level and activity of granule-
810 bound starch synthase I in Arabidopsis leaves during the day/night cycle. *Plant Mol. Biol.* **51**, 949–958 (2003).
811 62. Yamamori, M., Nakamura, T., Endo, T. R. & Nagamine, T. Waxy protein deficiency and chromosomal
812 location of coding genes in common wheat. *Theor. Appl. Genet.* **89**, 179–184 (1994).
813 63. Vrinten, P. L. & Nakamura, T. Wheat granule-bound starch synthase I and II are encoded by separate
814 genes that are expressed in different tissues. *Plant Physiol*. **122**, 255–263 (2000).
815 64. Kaplan, F., Sung, D. Y. & Guy, C. L. Roles of β-amylase and starch breakdown during temperatures
816 stress. *Physiol. Plant.* **126**, 120–128 (2006).
817 65. Otto, S. P. The Evolutionary Consequences of Polyploidy. *Cell* **131**, 452–462 (2007).
818 66. Gardiner, L. J. *et al.* A genome-wide survey of DNA methylation in hexaploid wheat. *Genome Biol.*
819 **16**, 1–15 (2015).
820 67. Chen, Z. J. & Mas, P. Interactive roles of chromatin regulation and circadian clock function in plants.
821 *Genome Biol*. **20**, 1–12 (2019).
822 68. Walsh, J. B. How often do duplicated genes evolve new functions? *Genetics* **139**, 421–428 (1995).
823 69. Lou, P. *et al*. Preferential retention of circadian clock genes during diploidization following whole
824 genome triplication in Brassica rapa. *Plant Cell* **24**, 2415–2426 (2012).

70. Rees, H. *et al.* A high-throughput delayed fluorescence method reveals underlying differences in the control of circadian rhythms in Triticum aestivum and Brassica napus. *Plant Methods* **15**, 51 (2019).

71. Yarkhunova, Y. *et al.* Selection during crop diversification involves correlated evolution of the circadian clock and ecophysiological traits in Brassica rapa. *New Phytol.* **210**, 133–144 (2016).

72. Michael, T. P. *et al.* Enhanced Fitness Conferred by Naturally Occurring Variation in the Circadian Clock. *Science (80-.).* **302**, 1049–1053 (2003).

73. Gould, P. D. *et al.* Coordination of robust single cell rhythms in the Arabidopsis circadian clock via spatial waves of gene expression. *Elife* **7**, (2018).

74. Takahashi, N., Hirata, Y., Aihara, K. & Mas, P. A Hierarchical Multi-oscillator Network Orchestrates the Arabidopsis Circadian System. *Cell* **163**, 148–159 (2015).

75. Locke, J. C. W. *et al.* Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. *Mol. Syst. Biol.* **2**, 59 (2006).

76. Hall, A., Kozma-Bognár, L., Bastow, R. M., Nagy, F. & Millar, A. J. Distinct regulation of CAB and PHYB gene expression by similar circadian clocks. *Plant J.* **32**, 529–537 (2002).

77. Hotta, C. T. *et al.* Are there multiple circadian clocks in plants? *Plant Signal. Behav.* **3**, 342 (2008).

78. Delorge, I., Figueroa, C. M., Feil, R., Lunn, J. E. & Van Dijck, P. Trehalose-6-phosphate synthase 1 is not the only active TPS in Arabidopsis thaliana. *Biochem. J.* **466**, 283–290 (2015).

79. Tian, L. *et al.* The trehalose-6-phosphate synthase TPS5 negatively regulates ABA signaling in Arabidopsis thaliana. *Plant Cell Rep.* **38**, 869–882 (2019).

80. Suzuki, N., Bajad, S., Shuman, J., Shulaev, V. & Mittler, R. The Transcriptional Co-activator MBF1c Is a Key Regulator of Thermotolerance in Arabidopsis thaliana. *J. Biol. Chem.* **283**, 9269–9275 (2008).

81. Wang, X., Du, Y. & Yu, D. Trehalose phosphate synthase 5-dependent trehalose metabolism modulates basal defense responses in Arabidopsis thaliana. *J. Integr. Plant Biol.* **61**, 509–527 (2019).

82. V, S. *et al.* TREHALOSE PHOSPHATE SYNTHASE11-dependent trehalose metabolism promotes Arabidopsis thaliana defense against the phloem-feeding insect Myzus persicae. *Plant J.* **67**, 94–104 (2011).

83. Vishal, B., Krishnamurthy, P., Ramamoorthy, R. & Kumar, P. P. OsTPS8 controls yield-related traits and confers salt stress tolerance in rice by enhancing suberin deposition. *New Phytol.* **221**, 1369–1386 (2019).

84. Liu, X. *et al.* Overexpression of the wheat trehalose 6-phosphate synthase 11 gene enhances cold tolerance in Arabidopsis thaliana. *Gene* **710**, 210–217 (2019).

85. MacKinnon, K. J. M. *et al.* Changes in ambient temperature are the prevailing cue in determining Brachypodium distachyon diurnal gene regulation. *New Phytol.* **227**, 1709–1724 (2020).

86. Li, M. *et al.* Comprehensive mapping of abiotic stress inputs into the soybean circadian clock. *Proc. Natl. Acad. Sci.* **116**, 23840–23849 (2019).

87. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, (2014).

88. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, (2019).

89. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, (2016).

90. Gardiner, L. J. *et al.* Interpreting machine learning models to investigate circadian regulation and facilitate exploration of clock function. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 2021 (2021).

91. Zhang, H. *et al.* A Comprehensive Online Database for Exploring ~20,000 Public Arabidopsis RNA-Seq Libraries. *Mol. Plant* **13**, 1231–1233 (2020).

92. Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).

93. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

94. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).

95. Smedley, D. *et al.* BioMart – biological queries made easy. *BMC Genomics 2009 101* **10**, 1–12 (2009).

96. Kim, J. A. *et al.* Transcriptome analysis of diurnal gene expression in Chinese cabbage. *Genes (Basel).* **10**, (2019).

97. Yang, Y., Li, Y., Sancar, A. & Oztas, O. The circadian clock shapes the Arabidopsis transcriptome by regulating alternative splicing and alternative polyadenylation. *J. Biol. Chem.* **295**, 7608–7619 (2020).

98. Wu, G., Anafi, R. C., Hughes, M. E., Kornacker, K. & Hogenesch, J. B. MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351–3353 (2016).

99. Agostinelli, C. & Lund, U. R package 'circular': Circular Statistics. (2017).

100. Zielinski, T., Moore, A. M., Troup, E., Halliday, K. J. & Millar, A. J. Strengths and Limitations of Period Estimation Methods for Circadian Data. *PLoS One* **9**, e96462 (2014).

101. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.44.0. *Alexa* (2021).

102. O'Malley, R. C. *et al.* Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**, (2016).

103. Marcussen, T. *et al.* Ancient hybridizations among the ancestral genomes of bread wheat. *Science (80-.).* **345**, 6194 (2014).

104. Zeeman, S. C., Smith, S. M. & Smith, A. M. The diurnal metabolism of leaf starch. *Biochem. J.* **401**, 13–28 (2007).

105. López-González, C., Juárez-Colunga, S., Morales-Elías, N. C. & Tiessen, A. Exploring regulatory networks in plants: Transcription factors of starch metabolism. *PeerJ* **2019**, e6841 (2019).

106. Skryhan, K., Gurrieri, L., Sparla, F., Trost, P. & Blennow, A. Redox regulation of starch metabolism. *Front. Plant Sci.* **9**, 1344 (2018).

Acknowledgements

H.R., R.R.P. and A.H. were funded by the BBSRC Core Strategic Programme Grant (Genomes to Food Security) BB/CSP1720/1 and its constituent work package, BBS/E/T/000PR9819 (WP2 Regulatory interactions and Complex Phenotypes). B.W., R.R.P. and A.H. was supported by the BBSRC Designing Future Wheat grant BB/P016855/1; BBS/E/T/000PR9783 (DFW WP4 Data Access and Analysis). C.R. by a BBSRC grant BB/V509267/1 and Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the "AI for Science" theme within that grant & The Alan Turing Institute. J.C. by the BBSRC funded Norwich Research Park Biosciences Doctoral Training Partnership grant BB/M011216/1. We would also like to acknowledge BBS/E/T/000PR9816 (NC1 □ Supporting EI's ISPs and the UK Community with Genomics and Single Cell Analysis) for data generation and BB/CCG1720/1 for the physical HPC infrastructure and data centre delivered via the NBI Computing infrastructure for Science (CiS) group. P.B. was supported by a BBSRC TRDF grant BB/N023145/1. A.N.D., L.L.B.D. and C.A.G. are funded by BBSRC ISP Genes in the Environment (BB/P013511/1). C.A.G. was also funded by UK BBSRC SWBIO DTP (BB/M009122/1). We thank Dr. Susan Duncan for her contribution to preliminary experiments prior to the start of this project.

Author contributions

H.R. and A.H. were involved with project concept and designed experiments. H.R. performed time-course experiments. R.R.P. processed data, quantified read counts and conducted clustering of rhythmic transcripts. P.B. conducted the phylogenetic analysis of core circadian protein families. J.C. conducted the cross-correlation analysis. C.R. processed previously published circadian datasets. H.R., L.L.B.D., C.A.G., B.W., R.R.P. A.H., and A.N.D. analysed and interpreted the RNA-seq data. H.R. wrote the initial manuscript and all authors contributed to subsequent drafts.

Corresponding author

Correspondence to Anthony Hall.

Ethics declarations

Competing interests: The authors declare no competing interests.

Main Figures and Tables

Fig. 1. Circadian regulation of homoeolog expression of wheat triads.

a, Schematic of the origins of hexaploid wheat, showing circadian clocks evolving independently in the ancestors of the A, B and D subgenomes following divergence from a common ancestor approximately 6.5 million years ago**.** Colours of clock icons represent theoretical differences in clock regulation integrated in the tetraploid and hexaploid hybrids either through circadian balance or through dominance of a particular homoeolog copy. Speciation and hybridisation event dates are based on estimates from[103]. **b,** Density plot showing the distribution of period lengths across rhythmic transcripts (BH q < 0.01) in *Arabidopsis*, *Brassica rapa*, *Brachypodium distachyon*, *Glycine max* (Soybean) and wheat based on meta2d estimates on 24-68h data following transfer to constant light. **c,** Histogram showing distribution of period lengths in wheat split between the A, B and D subgenomes. Dotted line indicates the mean period for the A, B and D subgenomes. **d,** Proportions of triads with either zero (red segment), one (green segment), two (blue segment) or three (purple segment) rhythmic gene(s) out of the 16,359 expressed triads in this dataset. Lighter shading in the outer segments represents cases where one/two homeolog(s) have high confidence rhythmicity (BH q < 0.01) alongside an arrhythmic homeolog (BH q >0.05). We term these genes "unbalanced rhythmicity" triads. **e,** Of the 3448 triads with three rhythmic genes, 464 had homoeologs which peaked with an optimum lag of 4, 8 or 12h following cross-correlation analysis. **f,** 1,139 had homoeologs with period differences of more than 2h. **g,** 701 had homoeologs with a more than two-fold difference in relative amplitude. **h,i,** Example triads for unbalanced rhythmicity, where either one or two homoeologs are rhythmic respectively. **j,** Example triad where the D genome homeolog lags by 8h. **k,** Example of a triad where the A genome homoeolog has a period estimate 4h longer than the D genome homoeolog. **l,** Example triad where the relative amplitude of the D-genome homoeolog is more than four times that of the A-genome homoeolog. **m,** The rhythmicity of all three homoeologs in **l,** is evident when the expression is mean normalized. **n,** Mean expression of transcripts across all timepoints in the A, B and D subgenomes within unbalanced rhythmicity triads compared with circadian balanced and arrhythmic triads. Error bars represent standard error.

Circadian statistics are meta2d estimates from data 0-68h after transfer to L:L. Data represent the mean of three biological replicates with transcript expression collapsed to gene level. Genes in example triads are: [Triad 1664: TraesCS3A02G177600, TraesCS3B02G207400, TraesCS3D02G183200], [Triad 408: TraesCS3A02G533700, TraesCS3B02G610500, TraesCS3D02G539000], [Triad 13405: TraesCS6A02G269100, TraesCS6B02G296400, TraesCS6D02G245800], [Triad 10854: TraesCS6A02G166500, TraesCS6B02G194000, TraesCS6D02G155100] and [Triad 2454: TraesCS2A02G333000, TraesCS2B02G348800, TraesCS2D02G329900].

974 **Fig. 2. Overlapping GO-slim terms shared between *Arabidopsis* and wheat modules**
975 **expressed at similar times in the day**

976 **a,** Pearson correlation coefficient (r) between eigengenes for wheat and *Arabidopsis*
977 expression modules ordered by circadian phase. Coloured triangles and axes labels
978 correspond to module expression profiles and columns in bubble-plot. **b-e,** Expression
979 profiles of *Arabidopsis* and wheat modules compared in the main text normalised to their
980 mean. Solid and dashed black lines represent the module eigengene for wheat and
981 *Arabidopsis* modules respectively. **f,** GO-slim terms associated with *Arabidopsis* and Wheat
982 modules. Modules are ordered by predicted CT phase for each species. Only terms with -
983 Log10*p* >3 are shown. Wheat W6 and *Arabidopsis* A4 contained no terms above the
984 significance cut-off and so are not shown. Bubble color indicates the -Log10*p*-value
985 significance from Fisher's exact test and size indicates the frequency of the GO-slim term in
986 the underlying EBI Gene Ontology Annotation database (larger bubbles indicate more
987 general terms).
988
989

Fig. 3. Free-running expression of core circadian clock genes in wheat and their homologs in *Arabidopsis.* a-l, Wheat circadian clock genes were identified through alignment of phylogenetic protein family trees or BLASTP to known clock gene homologs. Gene IDs for each gene set are in Supplementary Table 6. Wheat homoeologs are coloured according to their identity to either the A genome (orange), B genome (yellow) or D genome (blue) and grey and white blocks indicate subjective dark and light time periods under constant conditions. Data represent the mean of three biological replicates and transcript expression is collapsed to gene level. Expression profiles for additional core circadian clock genes are in Supplementary Fig. 14.

m, phases of core clock genes in *Arabidopsis* and wheat (meta2d estimates from data 24-68h after transfer to L:L). Genes were not plotted if B.H *q*-values were > 0.01. Wheat values represent circular mean circadian phases (CT) across homoeologs calculated in Supplementary Table 7.

Fig. 4. Similarities and differences in circadian control of transcript accumulation in key genes involved in primary metabolism and signalling. Circles represent metabolites involved in the breakdown and biosynthesis of starch. Starch synthesis occurs during the day and breakdown occurs at night as indicated by the yellow to grey shading gradient. The dotted line encloses processes which take place in the chloroplast. Abbreviations: HP: Hexose-phosphate, T6P: Trehalose-6-phosphate, TP: Triose phosphate, 3-PGA: Glycerate 3-phosphate, Fru6P: Fructose-6-phosphate, Glc6P: Glucose-6-phosphate, Glc1P: Glucose-1-phosphate, ATP: Adensoine tri phosphate, ADP-Glc: ADP-glucose, TPS: Trehalose phosphate synthase, TPP: Trehalose phosphate phosphatase, PGK1: Phosphoglycerate kinase 1, PGI1: Glucose-6-phosphate isomerase, PGM1: Phosphoglucomutase-1, PHS1 and 2: ALPHA-GLUCAN PHOSPHORYLASE 1 and 2, AGPase: ADP-Glc pyrophosphorylase, BAM1,3,5: β-amylase 1,3,5, ISA1,2,3: Isoamylase 1,2,3, DPE1,2: Disproportionating enzyme1 and 2, SBEI,II: Starch branching enzyme I, II, PU1: Pullulanase 1, PWD: Phosphoglucan, water dikinase, GWD: α-glucan, water dikinase, SEX4: starch excess 4, AMY3: α-amylase, GBSS: Granule bound Starch synthase, SSI-IV: Starch synthase I-IV. Pathway references:[104–106].

Table 1: Numbers of rhythmic genes at (BH $q < 0.05$ or BH $q < 0.01$) in *Arabidopsis* and wheat identified using Metacycle Benjamini Hochberg *q*-values. Periods, relative amplitudes, and *q*-values are estimates from meta2d. Data windows reflect hours relative to transfer to constant light from entrained 12:12h light conditions. A repeat of this table with pre-filtering to remove low-expression genes is provided in Supplementary Figure 1, and the effects on proportions of rhythmic genes are discussed in Supplementary Note 1.

	Wheat data from this study		*Arabidopsis* data from Romanowski et al.
	24-68 data window	0-68 data window	24-68 data window
Total number of expressed genes	86,567	86,567	26,392
Total rhythmic genes (BH $q < 0.05$)	28,594	28,530	13,392
Total rhythmic genes (BH $q < 0.01$)	18,633	21,059	10,317
Mean Period (h) (BH $q < 0.05$)	26.60h (SD 3.62)	26.75h (SD 2.82)	23.50 (SD 2.52)
Mean Period (h) (BH $q < 0.01$)	26.82h (SD 3.21)	26.83h (SD 2.42)	23.62 (SD 2.04)
Mean relative Amplitude (BH $q < 0.05$)	0.24 (SD 0.19)	0.26 (SD 0.20)	0.28 (SD 0.20)
Mean relative Amplitude (BH $q < 0.01$)	0.27 (SD 0.19)	0.29 (SD 0.21)	0.30 (SD 0.20)

Table 2: GO-slim terms for biological processes associated with circadian balanced, circadian unbalanced, and arrhythmic wheat triads. Only enriched terms which were highly enriched (Fisher's exact test $p<0.01$) in one category and non-significantly expressed ($p>0.05$) in other categories is displayed.

	GO ID	Terms	*p*-value in circadian balanced triads	*p*-value in circadian unbalanced triads	*p*-value in non-rhythmic triads
CIRCADIAN BALANCED	GO:0009628	response to abiotic stimulus	0.00	0.22	0.58
	GO:0015979	photosynthesis	0.00	1.00	1.00
	GO:0006091	generation of precursor metabolites and energy	0.00	1.00	1.00
	GO:0006518	peptide metabolic process	0.00	1.00	0.92
	GO:1901566	organonitrogen compound biosynthetic process	0.00	1.00	0.92
	GO:0009059	macromolecule biosynthetic process	0.00	1.00	0.99
	GO:0006412	translation	0.00	1.00	0.98
	GO:0034645	cellular macromolecule biosynthetic process	0.00	1.00	1.00
	GO:0010467	gene expression	0.00	1.00	0.90
	GO:0019725	cellular homeostasis	0.00	0.54	0.96
	GO:0065008	regulation of biological quality	0.00	0.71	1.00
CIRCADIAN UNBALANCED	GO:0003006	developmental process involved in reproduction	0.85	0.00	0.87
	GO:0090567	reproductive shoot system development	0.92	0.01	1.00
	GO:0009719	response to endogenous stimulus	0.97	0.01	0.10
	GO:0048731	system development	0.98	0.00	0.97
	GO:0048608	reproductive structure development	0.98	0.00	0.97
	GO:0043412	macromolecule modification	1.00	0.00	0.28
	GO:0022414	reproductive process	1.00	0.00	0.47
NON-RHYTHMIC	GO:0044237	cellular metabolic process	0.08	1.00	0.00
	GO:0009605	response to external stimulus	0.41	0.39	0.00
	GO:0009607	response to biotic stimulus	0.58	0.60	0.01
	GO:0060255	regulation of macromolecule metabolic process	0.60	0.76	0.00
	GO:0009056	catabolic process	0.65	0.93	0.00
	GO:0048869	cellular developmental process	0.71	0.19	0.00
	GO:0019222	regulation of metabolic process	0.78	0.85	0.00
	GO:0044238	primary metabolic process	0.81	0.24	0.00
	GO:0071704	organic substance metabolic process	0.81	0.26	0.00
	GO:0008219	cell death	0.82	0.94	0.00
	GO:0010468	regulation of gene expression	0.88	0.15	0.00

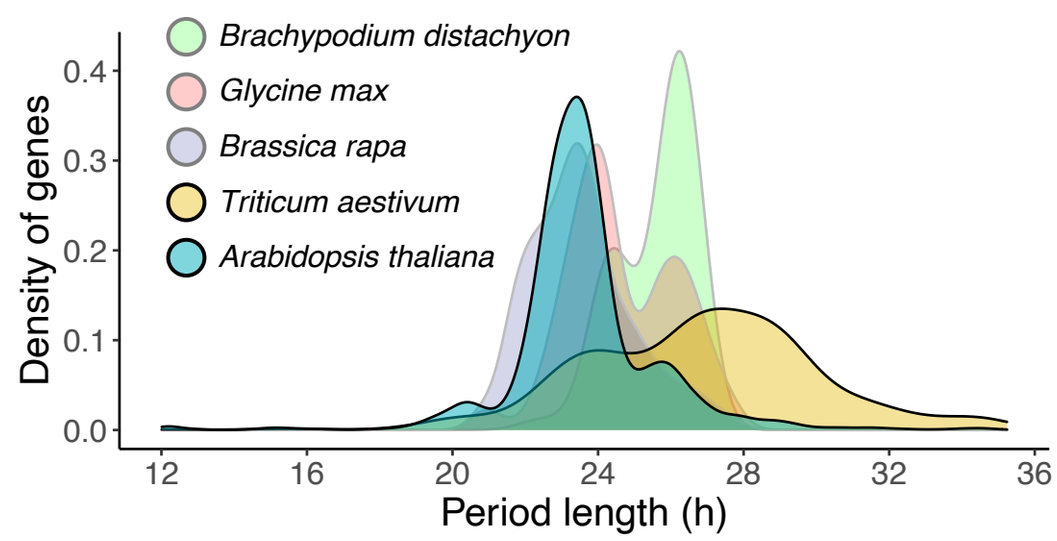
	GO:0009790	embryo development	0.95	0.98	0.00
	GO:0007049	cell cycle	0.97	1.00	0.00
	GO:0065009	regulation of molecular function	1.00	1.00	0.00
	GO:0006807	nitrogen compound metabolic process	1.00	1.00	0.00

1034

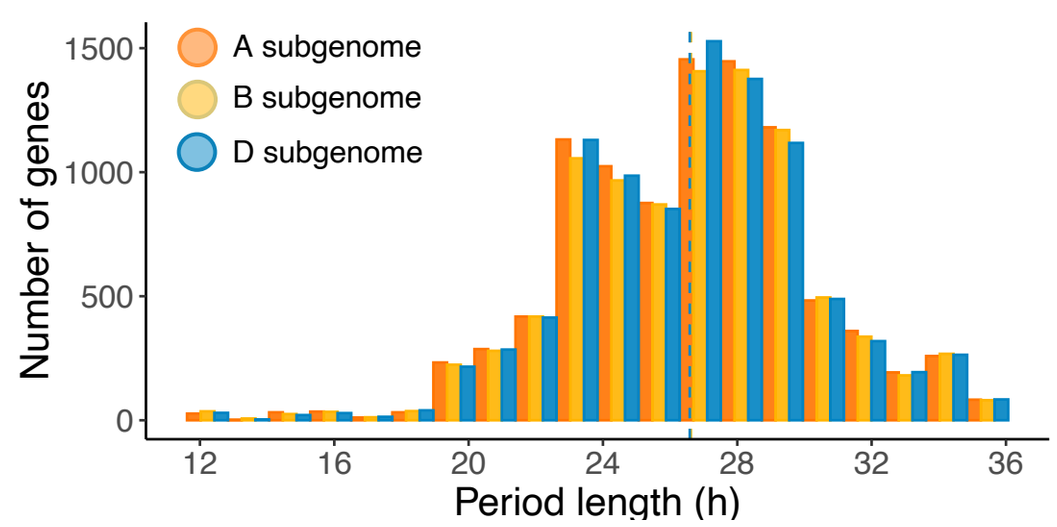
a

Hybridization events
Speciation
Triticeae common ancestor
~6.5 MYA
AA
BB
DD
~5.5 MYA
T. tauschii
T. urartu
<0.8 MYA
AABB
T. turgidum
Relative of Ae. speltoides
~10,000 YA
AABBDD
Hexaploid wheat *T. Aestivum*

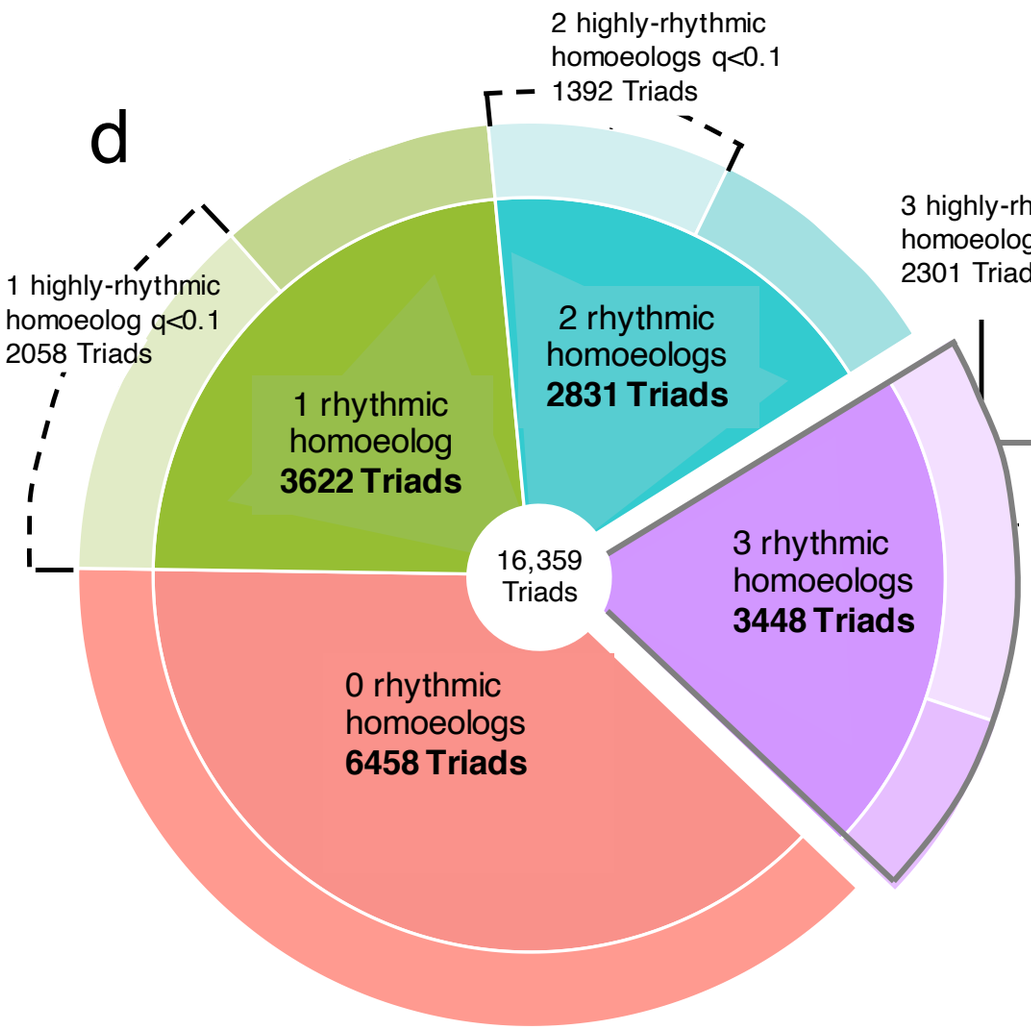
b



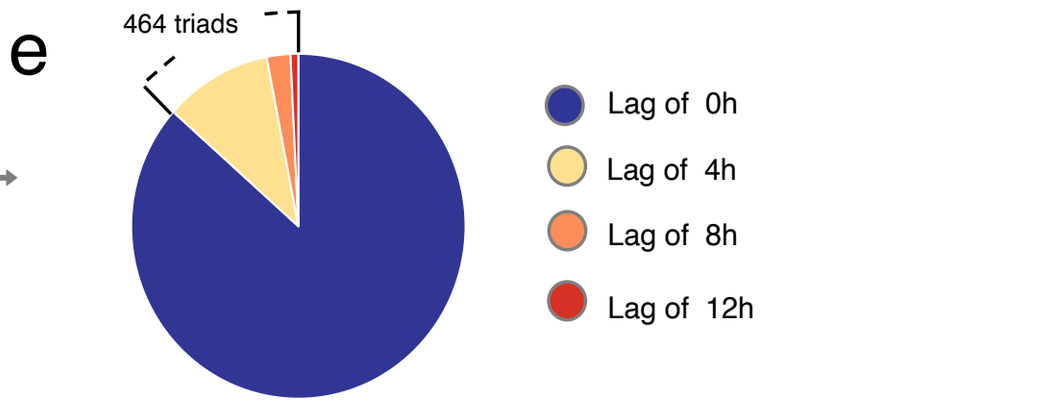
c



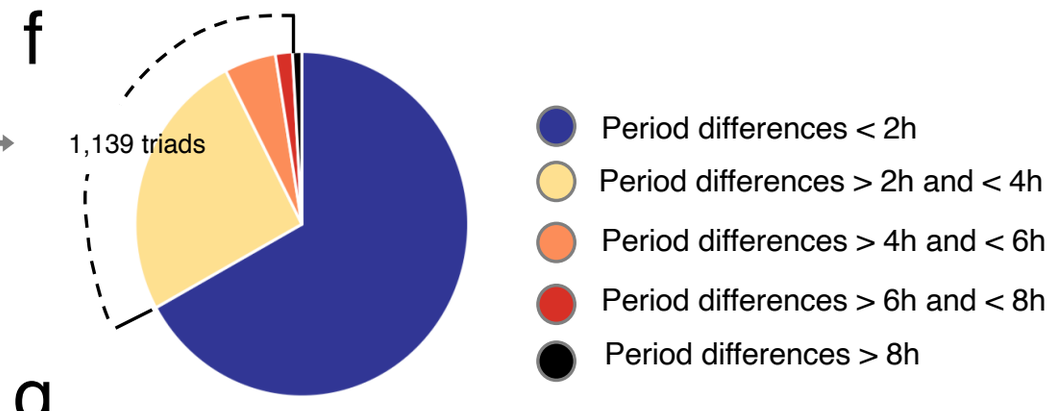
d



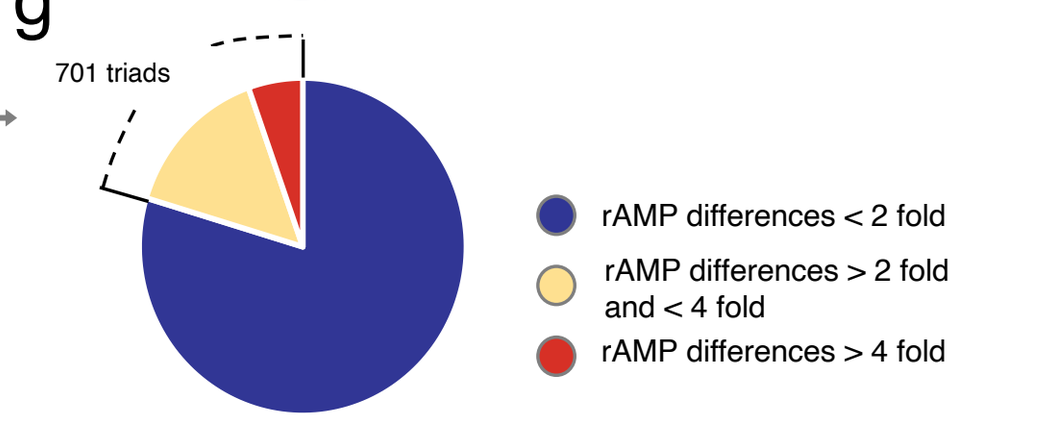
e



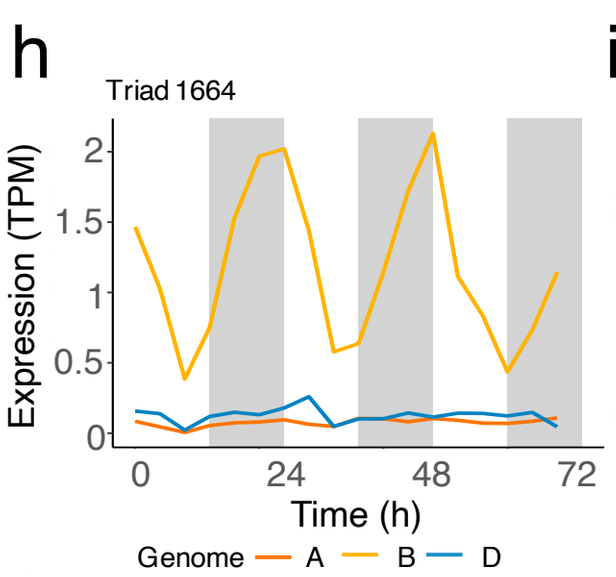
f



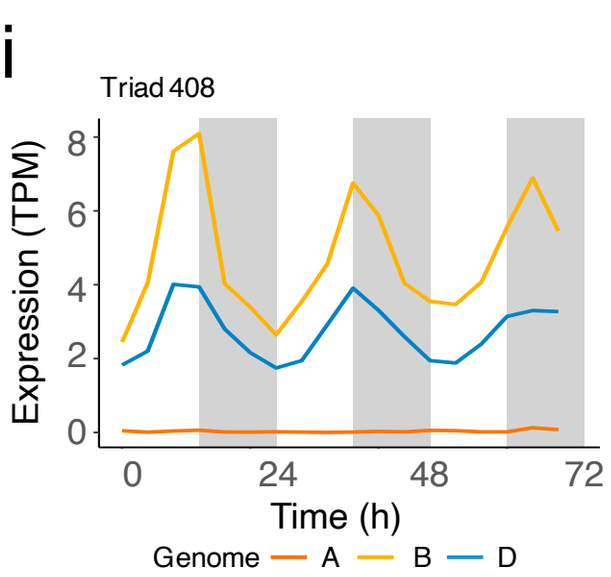
g



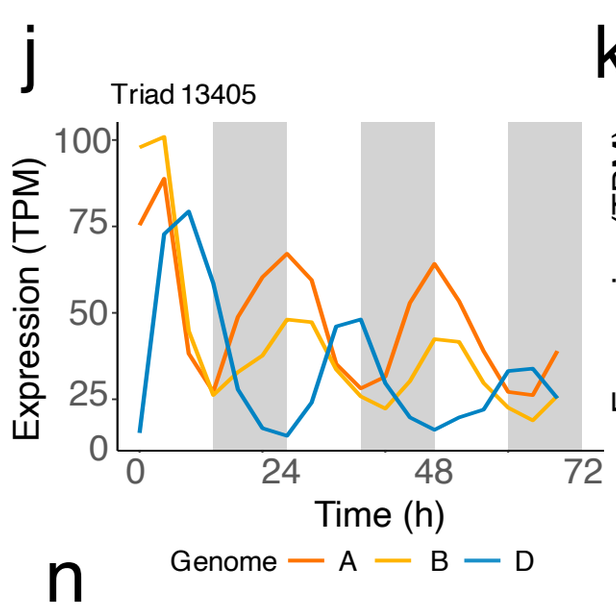
h



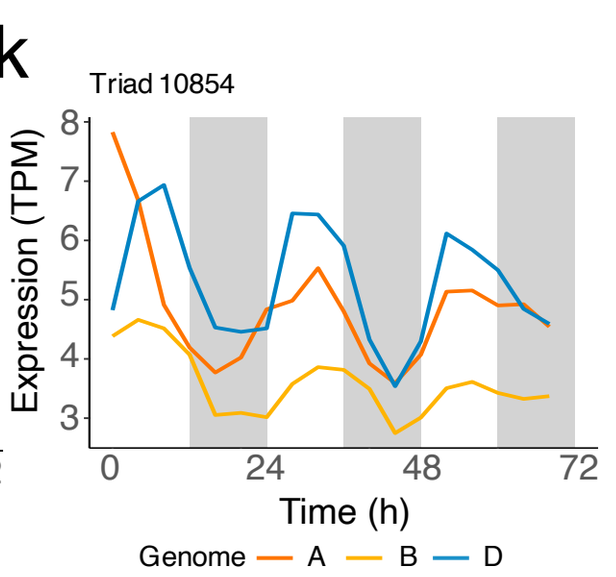
i



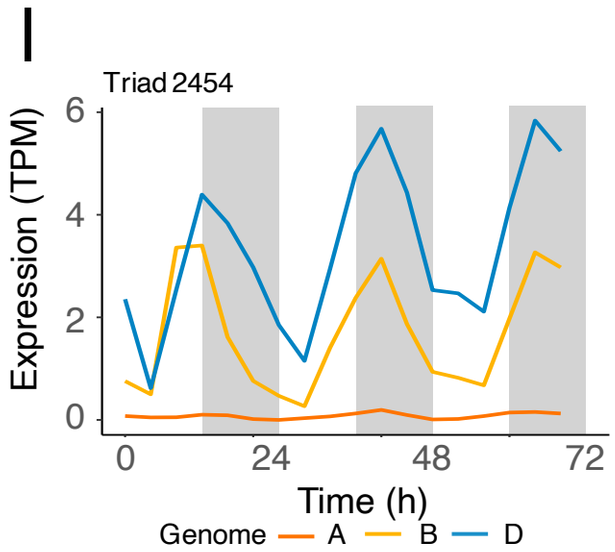
j



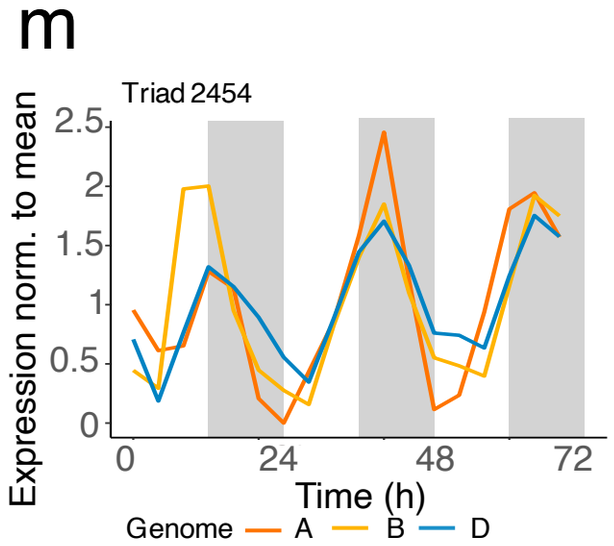
k



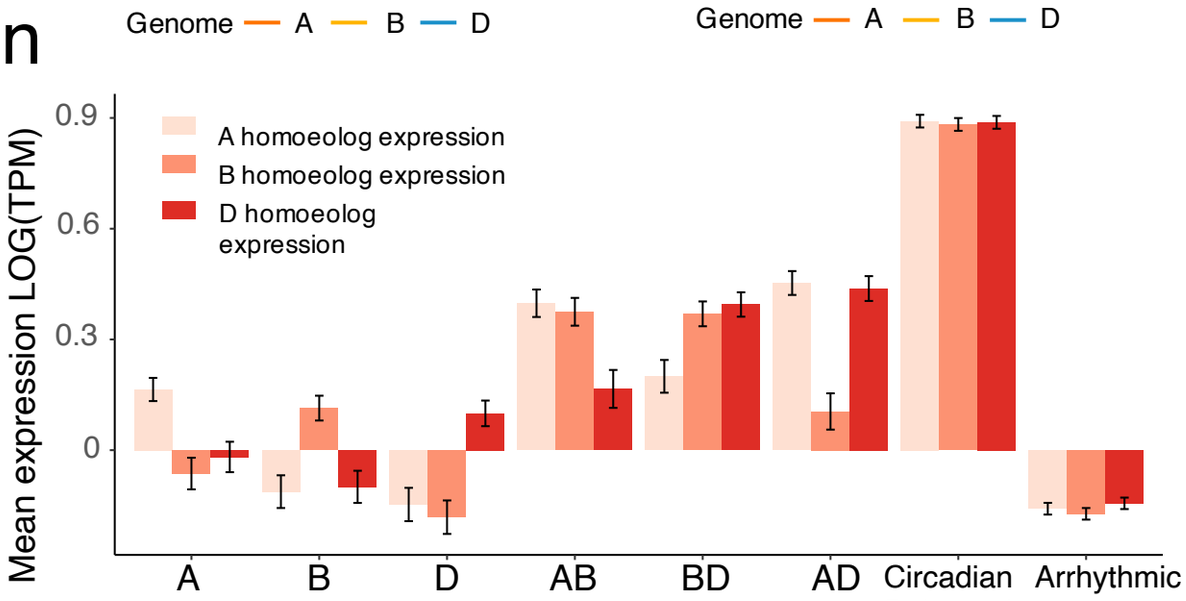
l

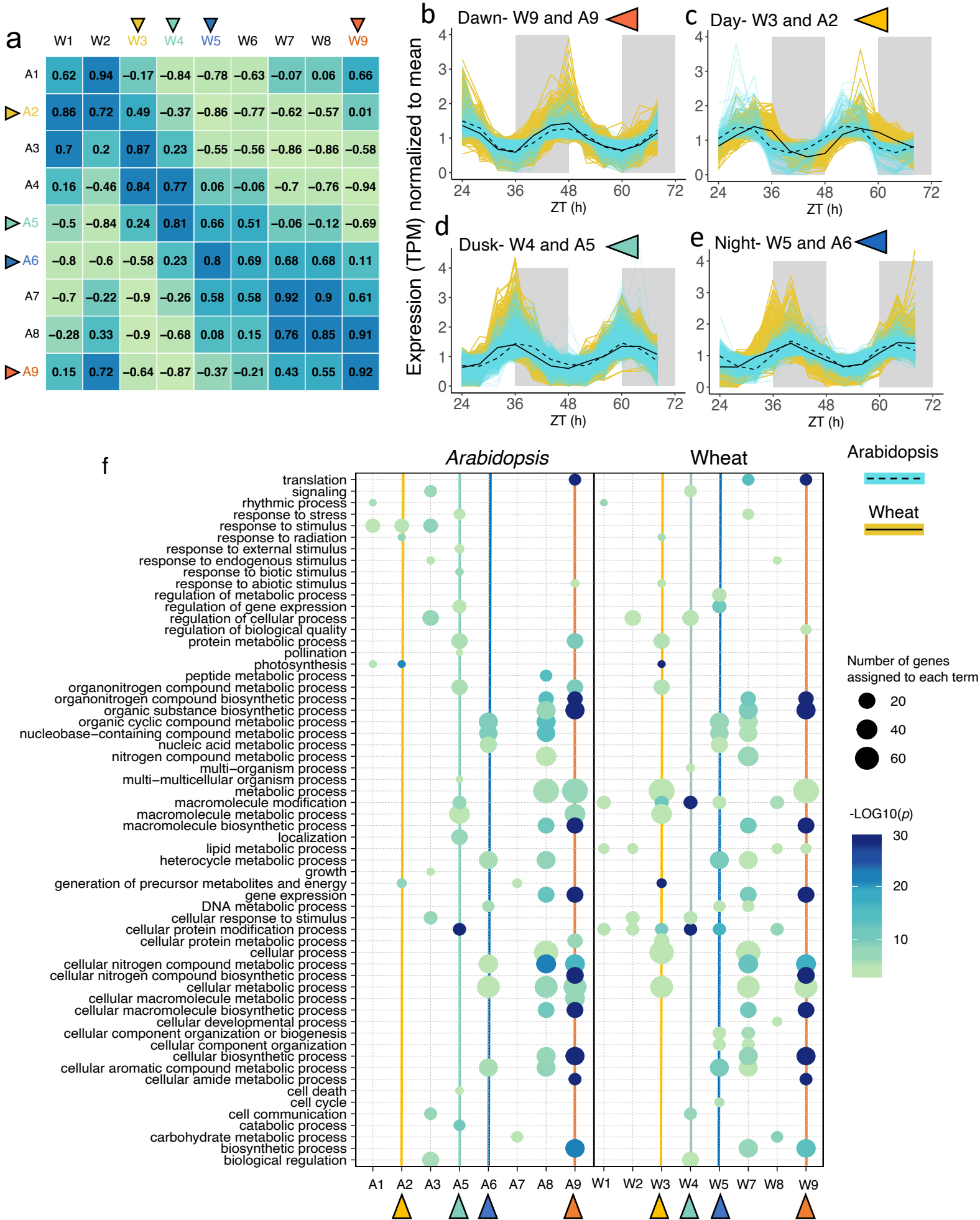


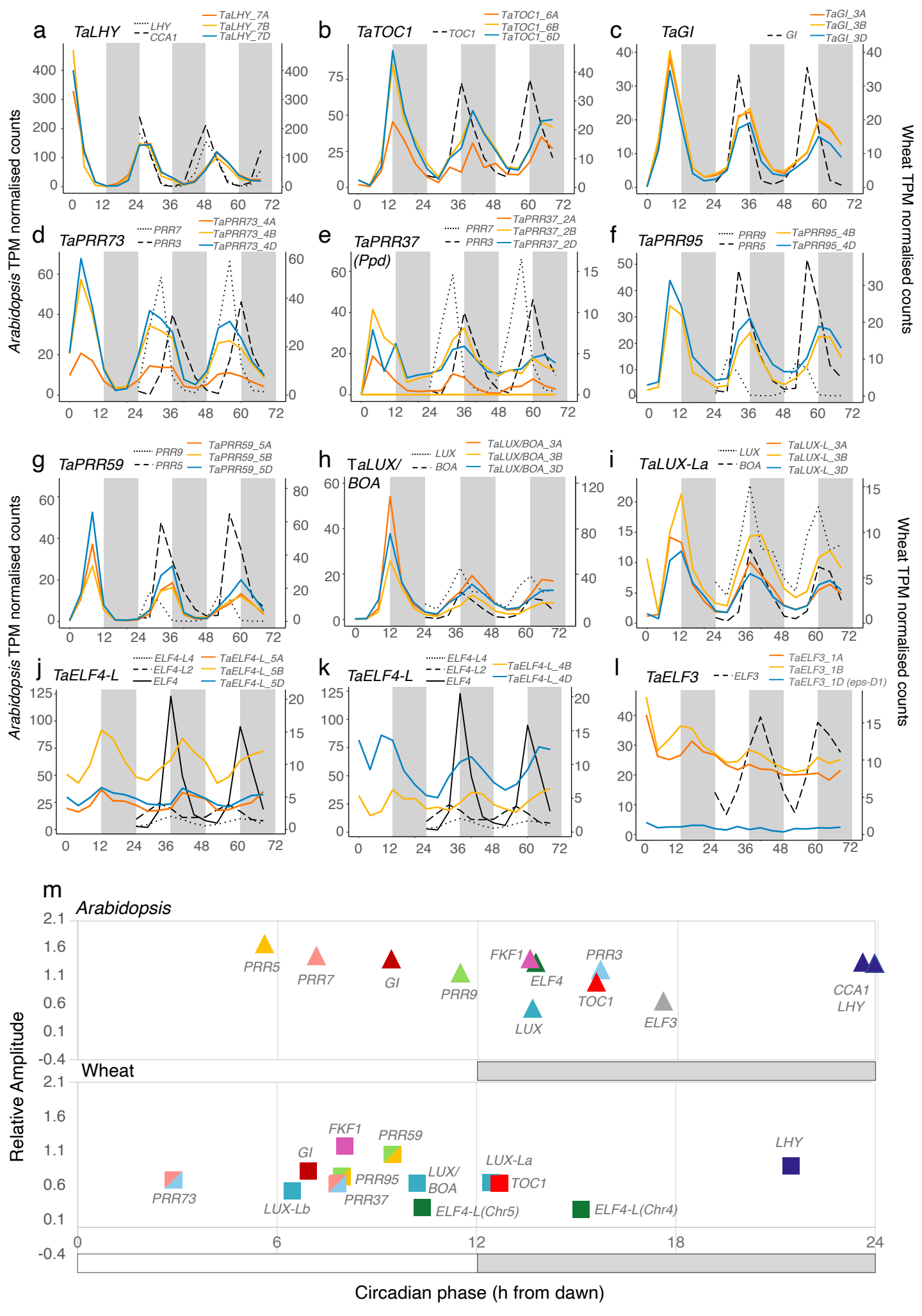
m

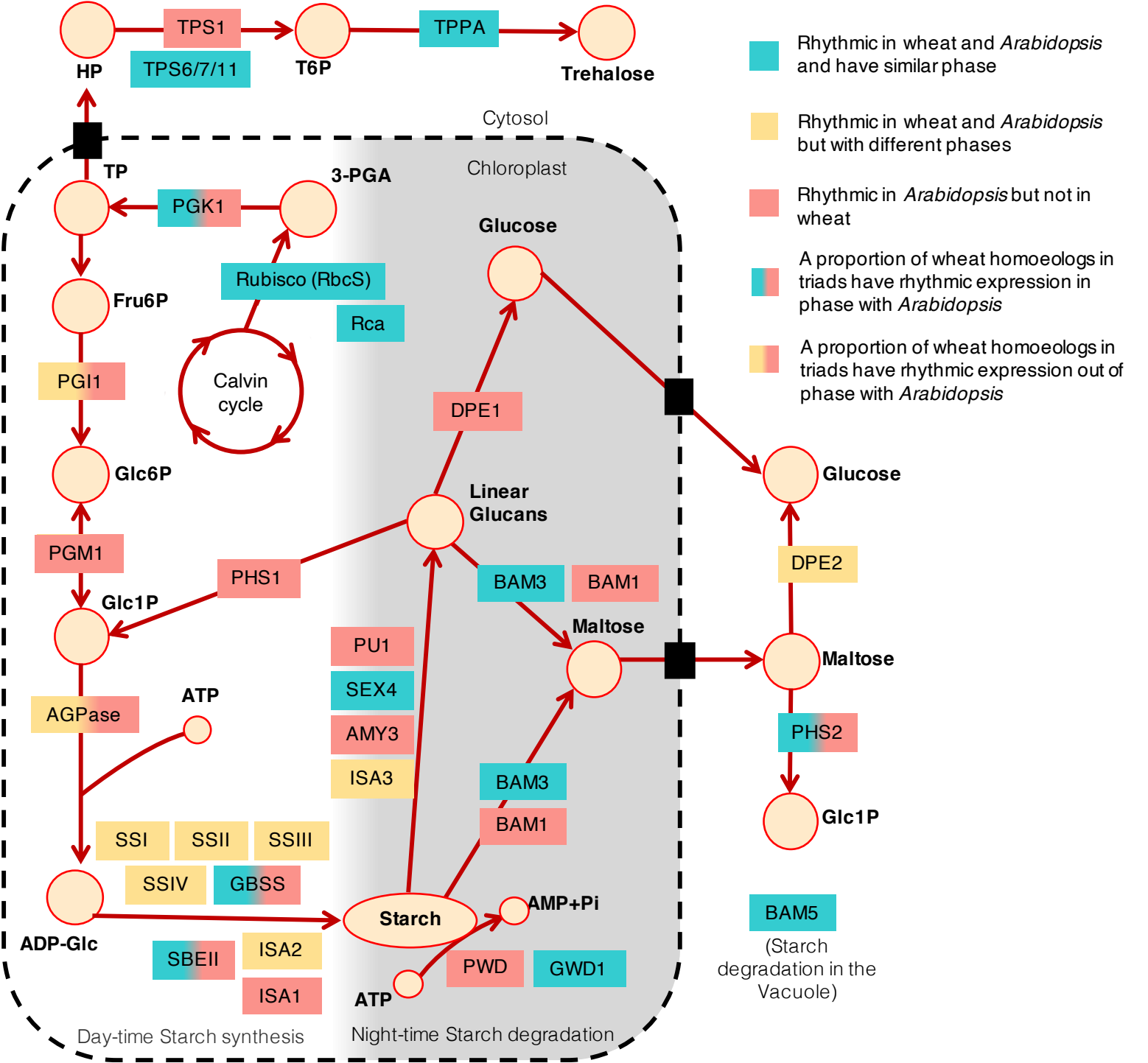


n




a
W1 W2 W3 W4 W5 W6 W7 W8 W9
A1 0.62 0.94 −0.17 −0.84 −0.78 −0.63 −0.07 0.06 0.66
A2 0.86 0.72 0.49 −0.37 −0.86 −0.77 −0.62 −0.57 0.01
A3 0.7 0.2 0.87 0.23 −0.55 −0.56 −0.86 −0.86 −0.58
A4 0.16 −0.46 0.84 0.77 0.06 −0.06 −0.7 −0.76 −0.94
A5 −0.5 −0.84 0.24 0.81 0.66 0.51 −0.06 −0.12 −0.69
A6 −0.8 −0.6 −0.58 0.23 0.8 0.69 0.68 0.68 0.11
A7 −0.7 −0.22 −0.9 −0.26 0.58 0.58 0.92 0.9 0.61
A8 −0.28 0.33 −0.9 −0.68 0.08 0.15 0.76 0.85 0.91
A9 0.15 0.72 −0.64 −0.87 −0.37 −0.21 0.43 0.55 0.92
b
Dawn- W9 and A9
c
Day- W3 and A2
d
Dusk- W4 and A5
e
Night- W5 and A6
Expression (TPM) normalized to mean
ZT (h)
f
Arabidopsis
Wheat
translation
signaling
rhythmic process
response to stress
response to stimulus
response to radiation
response to external stimulus
response to endogenous stimulus
response to biotic stimulus
response to abiotic stimulus
regulation of metabolic process
regulation of gene expression
regulation of cellular process
regulation of biological quality
protein metabolic process
pollination
photosynthesis
peptide metabolic process
organonitrogen compound metabolic process
organonitrogen compound biosynthetic process
organic substance biosynthetic process
organic cyclic compound metabolic process
nucleobase−containing compound metabolic process
nucleic acid metabolic process
nitrogen compound metabolic process
multi−organism process
multi−multicellular organism process
metabolic process
macromolecule modification
macromolecule metabolic process
macromolecule biosynthetic process
localization
lipid metabolic process
heterocycle metabolic process
growth
generation of precursor metabolites and energy
gene expression
DNA metabolic process
cellular response to stimulus
cellular protein modification process
cellular protein metabolic process
cellular process
cellular nitrogen compound metabolic process
cellular nitrogen compound biosynthetic process
cellular metabolic process
cellular macromolecule metabolic process
cellular macromolecule biosynthetic process
cellular developmental process
cellular component organization or biogenesis
cellular component organization
cellular biosynthetic process
cellular aromatic compound metabolic process
cellular amide metabolic process
cell death
cell cycle
cell communication
catabolic process
carbohydrate metabolic process
biosynthetic process
biological regulation
A1 A2 A3 A5 A6 A7 A8 A9 W1 W2 W3 W4 W5 W7 W8 W9
Arabidopsis
Wheat
Number of genes assigned to each term
20
40
60
-LOG10(p)
30
20
10


a
TaLHY
LHY
CCA1
TaLHY_7A
TaLHY_7B
TaLHY_7D
b
TaTOC1
TOC1
TaTOC1_6A
TaTOC1_6B
TaTOC1_6D
c
TaGI
GI
TaGI_3A
TaGI_3B
TaGI_3D
d
TaPRR73
PRR7
PRR3
TaPRR73_4A
TaPRR73_4B
TaPRR73_4D
e
TaPRR37
(Ppd)
PRR7
PRR3
TaPRR37_2A
TaPRR37_2B
TaPRR37_2D
f
TaPRR95
PRR9
PRR5
TaPRR95_4B
TaPRR95_4D
g
TaPRR59
PRR9
PRR5
TaPRR59_5A
TaPRR59_5B
TaPRR59_5D
h
TaLUX/
BOA
LUX
BOA
TaLUX/BOA_3A
TaLUX/BOA_3B
TaLUX/BOA_3D
i
TaLUX-La
LUX
BOA
TaLUX-L_3A
TaLUX-L_3B
TaLUX-L_3D
j
TaELF4-L
ELF4-L4
ELF4-L2
ELF4
TaELF4-L_5A
TaELF4-L_5B
TaELF4-L_5D
k
TaELF4-L
ELF4-L4
ELF4-L2
ELF4
TaELF4-L_4B
TaELF4-L_4D
l
TaELF3
ELF3
TaELF3_1A
TaELF3_1B
TaELF3_1D (eps-D1)
Arabidopsis TPM normalised counts
Wheat TPM normalised counts
m
Arabidopsis
PRR5
PRR7
GI
PRR9
FKF1
ELF4
LUX
PRR3
TOC1
ELF3
CCA1
LHY
Wheat
PRR73
GI
FKF1
PRR59
PRR95
PRR37
LUX-Lb
LUX/
BOA
ELF4-L(Chr5)
LUX-La
TOC1
ELF4-L(Chr4)
LHY
Relative Amplitude
Circadian phase (h from dawn)


HP
TPS1
TPS6/7/11
T6P
TPPA
Trehalose
Cytosol
Chloroplast
TP
3-PGA
PGK1
Rubisco (RbcS)
Rca
Calvin cycle
Fru6P
PGI1
Glc6P
PGM1
Glc1P
AGPase
ATP
ADP-Glc
SSI
SSII
SSIII
SSIV
GBSS
SBEII
ISA2
ISA1
Day-time Starch synthesis
Glucose
DPE1
Linear Glucans
PHS1
BAM3
BAM1
Maltose
PU1
SEX4
AMY3
ISA3
BAM3
BAM1
Starch
AMP+Pi
PWD
GWD1
ATP
Night-time Starch degradation
Glucose
DPE2
Maltose
PHS2
Glc1P
BAM5
(Starch degradation in the Vacuole)
Rhythmic in wheat and Arabidopsis and have similar phase
Rhythmic in wheat and Arabidopsis but with different phases
Rhythmic in Arabidopsis but not in wheat
A proportion of wheat homoeologs in triads have rhythmic expression in phase with Arabidopsis
A proportion of wheat homoeologs in triads have rhythmic expression out of phase with Arabidopsis