RNA-sequencing and mathematical modeling identify suite of light-sensitive circadian genes in an orb-web weaving spider

Natalia Toporikova1,*, Wenduo Cheng2, Leyuan Qian2, Andrew Mah1,3, Thomas Clarke1, Thomas C. Jones4, Darrell Moore4, Nadia A. Ayoub1*

1Department of Biology, Washington and Lee University, Lexington, VA, 24450, USA
2Division of Natural and Applied Sciences, Duke Kunshan University, Kunshan, Jiangsu, 215300, People's Republic of China
3Center for Neural Science, New York University, New York, New York, 10003, USA
4Department of Biological Sciences, East Tennessee State University, Box 70703, Johnson City, TN, 37604, USA

*Authors for correspondence:
Natalia Toporikova toporikovan@wlu.edu
Nadia A. Ayoub ayoubn@wlu.edu
Abstract

Background

Most organisms rely on a molecular circadian clock to orchestrate a wide range of physiological processes to match the 24-hour day. These molecular clocks are typically based on a negative feedback loop among a small set of proteins that govern the circadian output. Light or other environmental conditions can reset the circadian clock, but true circadian behaviors continue to cycle even in constant darkness, with an intrinsic period called the free-running period (FRP).

Spiders have unusual FRPs, with some species having extremely short FRP (e.g. 18 hours for trashline orb weaver), and many having highly variable FRPs (intraspecific variation of up to 10 hours). In the absence of any genetic model of circadian rhythms in spiders, we developed a mathematical model to optimize experimental conditions for identifying circadian genes that also respond to light cues.

Results

Our mathematical model involved a single gene that encodes a protein that inhibits its own transcription. In our model, light degrades the circadian transcript, which allows a broad range of FRPs to be entrained to a 24-hour day. Our model predicted that exposing spiders to a pulse of light in the middle of the night would cause a pattern of expression between two later time points that was opposite the pattern exhibited by spiders who did not receive a pulse of light. RNA-sequencing of four groups of adult female orb weaving spiders, *Metazygia wittfeldae*, under these experimental conditions resulted in 528 significantly differentially expressed (DE) transcripts between the two collection times or between the light pulse and no light pulse.
Consistent with our model, we found a cluster of transcripts with the flipped pattern of expression between the two collection times, dependent on the application of light.

**Conclusions**

Our DE transcripts represent the first genetic evidence for circadian output in spiders. Furthermore, those transcripts with a flipped pattern of expression represent prime candidates for light-sensitive circadian genes, which may be involved in entraining the circadian clock to light. Functions of these genes varied from growth and development to reproduction to gene regulation, consistent with other circadian systems.

**Keywords**

Circadian rhythms, *Metazygia wittfeldae*, Araneidae, entrainment, RNA sequencing, *de novo* transcriptomics, mathematical modeling

**Background**

Most organisms rely on a molecular circadian clock to orchestrate a wide range of physiological processes [1]. Interactions among several key proteins form multiple positive and negative feedback loops resulting in a 24-hour rhythm in expression of circadian transcripts [2]. Even under constant darkness (DD), oscillations in circadian gene expression generate daily periodic behaviors such as locomotion, hormone release, sleep-wake cycles, and many others. Such regular oscillations under DD conditions (typically within 2 hours of 24 hours), called the free running period (FRP), has been documented in a diverse range of organisms [3–5]. Circadian oscillations are also synchronized by external stimuli (zeitgebers), of which light is the
strongest. In alternating light-dark environments (LD), the light-sensing system modifies expression of one or more circadian genes, adjusting (or entraining) the FRP to a LD period [6].

In the well-studied circadian model, the fruit fly *Drosophila melanogaster*, the circadian “clock” is controlled via a negative feedback loop, which contains transcriptional factors *timeless* (*tim*), *period* (*per*), *cycle* (*cyc*) and *Clock* (*clk*). Proteins CLK and CYC form a complex that acts as a transcriptional activator for *per* and *tim* genes. In early evening, concentrations of PER and TIM in the cytoplasm increase, and these two proteins form a PER-TIM complex, which enters the nucleus and depresses their own promoter regions by interacting with CYC-CLK, thereby completing the negative feedback loop. When the PER-TIM heterodimer is degraded through double-time (DBT) protein-dependent phosphorylation, CYC-CLK activates gene transcription once again [7]. The fruit fly can receive information about light through the visual system via activation of cryptochrome 1 (CRY1), a blue-light sensitive photoreceptor. At the end of the night and in early morning, there is a lower concentration of TIM, and during the light phase CRY1 binds with TIM, preventing PER-TIM formation and so causing a delay in accumulation of that complex. As CRY1 degrades TIM, PER-TIM complex cannot be formed, thus relieving promoter depression, stimulating CYC-CLK activity and *tim*, *per* gene transcription. Newly synthesized TIM resets clocks in the beginning of the day or returns it to the initial state without light (Kistenpfennig et al. 2012). In mammals and some of the arthropods, the PER/TIM heterodimer indirectly blocks the function of CLK/CYC through an additional protein, cryptochrome 2 (CRY2) [8, 9].

Despite the theoretical and experimental expectations for FRP to closely match the 24-hour day, some spider species have remarkably short and variable FRPs of locomotor activity in constant dark. A standout example is the trashline orb-weaving spider *Cyclosa turbinata*
(Araneidae), which has a mean FRP of 18.7 hours in DD [10]. This is the shortest known naturally occurring FRP, even shorter than ones caused by lab-induced mutations, such as the hamster tau mutant (20 hours [11]) and the fruit fly per^6 mutant (19 hours [12]). Other species of spiders have been found to have extreme variation in FRP among individuals. For example, the ranges of FRPs exhibited by three cobweb weaving spider species (Theridiidae) were 19-23.5 h in the common house spider *P. tepidariorum* (mean FRP = 21.7 h), 20-29 h in the subsocial spider *A. studiosus* (mean FRP = 23.1), and 20-30.1 h (mean FRP = 24.5) in the southern black widow *Latrodectus mactans* [13]. Even the orb weaver *Metazygia wittfeldae* (Araneidae), which possesses an almost “typical” FRP (mean = 22.7 h) varied from 19 – 24 hours [14]. The three theridiid and two araneid species discussed here have FRPs that vary an order of magnitude more than those of most animals examined thus far [13].

The high levels of FRP variance within each of these spider species sharply contrast with the low levels of variance in other organisms, where it is suggested that the endogenous period of the circadian clock is under tight genetic control [11, 15–18]. The combination of high variation in FRP and divergence of FRP from 24 hours suggests that spiders have evolved novel elements of circadian control. Possible new mechanisms include a highly robust circadian response to light stimuli, which allows rapid entrainment, or a weakly oscillating system that could be easily perturbed by environmental stimuli, among others [13].

Thus far, the genetic control of circadian rhythms in spiders has not been investigated. To identify genes potentially involved in circadian entrainment in spiders, we subjected the spider species, *M. wittfeldae*, to two experimental conditions. In brief, we compared gene expression between spiders kept in constant darkness to those exposed to a pulse of light. We used a mathematical model to predict two time points with the greatest difference in gene expression
levels between the constant dark group and the group exposed to a light pulse. Our simulations further predicted that the direction of differential expression would switch between the two time points. RNA-sequencing of four experimental groups of spiders identified a cluster of transcripts that fit the prediction of a switched pattern of gene expression.

Results

Mathematical model predicts optimal experimental protocol

To optimize the search for circadian genes, we utilized a simplified mathematical model. Most of the circadian clocks studied so far are based on a negative feedback loop, where a group of circadian genes inhibit their own expression [2]. We therefore assumed that the circadian clock in spiders has at its core at least one circadian gene (\textit{gene X}). Translation of the circadian gene’s mRNA (\textit{mRNA X}) induces production of the circadian protein (\textit{protein X}), which is transferred back to the nucleus and inhibits its own transcription (Fig 1A). We simulated light in the model by a rapid degradation of \textit{mRNA X}. The model reproduces a wide range of free-running periods in the dark as well as entrainment to LD 12:12.

Our experimental goal was to identify the light-sensitive subset of circadian genes. To achieve this, we first entrained our model to LD 12:12 and then simulated a brief (1 hour) pulse of light in progressively increasing time intervals after the end of the last light phase (\textit{pulse} group), while keeping an identical model in constant darkness (\textit{no pulse} group) (Fig 1B). Based on these simulations, we predicted that applying the light pulse 5 hours after the last light phase (experimental time (et) 5 – et 6) would produce the greatest change in magnitude of gene expression of the model’s circadian \textit{mRNA X} (Fig 1B). To further improve the signal to noise ratio, we also used the model to identify collection times that would provide the largest
difference between the pulse and no pulse groups. By comparing the ratio of mRNA X with and without pulse at every hour after the light, we determined that collection times 7 and 16 hours after the last light phase (et7 and et16), produces the largest difference between pulse and no pulse conditions (Fig 1C).

Figure 1: Modeling prediction leading to experimental design. (A) Mathematical model of circadian clock is based on hypothetical circadian mRNA inhibited by its own protein. (B) Model simulation of 36 hours of LD 12:12 condition followed by DD conditions with (top) and without (bottom) application of a brief (1 hr.) light pulse. The sampling times (red arrows in top graph, blue arrows in bottom graph) were selected to optimize the differences between mRNA levels with and without the light pulse. (C) Experimental protocol following the optimal collection times predicted to yield the greatest differences in mRNA levels between the two experimental conditions.
Model predicts pattern of differential expression in response to light pulse and collection times

Since light degrades circadian mRNA, the model predicts that one hour after light application (et7), the group receiving a light pulse (pulse) will have significantly lower expression level than the group receiving no light pulse (no pulse) (Fig 2A, filled bars). In other words, light induces a phase advance (Fig 1B, compare top and bottom traces), where expression of circadian mRNA in the pulse group lags behind the no pulse group. As a result of this phase advance, at the time of the second collection (et16), the pattern of mRNA expression reverses, with the pulse group having a much higher expression level than the no pulse group (Fig 2A, open bars). The reason for the flipped mRNA expression at et16 is that the no pulse group reaches its trough while the lagging pulse group is almost at its peak.

Since we used our model to predict the pattern of differential expression of circadian light-sensitive genes, we can compare the ratio (also called Fold Change) of pulse vs. no pulse groups at two collection time points (Fig 2C). At et7, expression in the pulse group is lower than in the no pulse group, resulting in negative log Fold Change (log FC). But at et16, expression in the pulse group exceeds the no pulse group, resulting in a positive log FC. We therefore expect the circadian genes in our experiment to reverse the log FC values.

Our model also predicts similar reverse pattern of differential expression in response to collection times when comparing the pulse to no pulse groups. In the pulse group, mRNA expression is much higher at et16 than et7 (Fig 2B, red bars), generating a negative log FC (et7/et16) (Fig 2 D, left). In the no pulse group, which did not experience mRNA suppression, the relationship is opposite, with mRNA level higher at et7 than et16 (Fig 2B, blue boxes). In other words, the differential expression with respect to collection time in no pulse group...
generates a positive log FC (et7/et16) (Fig 2D). Therefore, our model predicts that differential expression of circadian genes will be reversed by both pulse and collection time conditions.

**Figure 2:** Expected outcomes of experimental protocol predicted by the model for light-sensitive genes. (A) Absolute gene expression is expected to be increased by the 1-hour pulse of light (*pulse*) in et7, but decreased at et16. (B) Absolute gene expression with the pulse is higher at et16 compared with et11. Without the pulse of light (*no pulse*), the trend is opposite, with higher expression at et16. (C) Differential expression, measured as log of ratio (log FC) between *pulse* and *no pulse* condition is negative in et7 and positive at et16. (D) Differential expression, measured as log FC between et7 and et16 times is negative in *pulse* condition and positive in *no pulse* condition.

**Light affects circadian gene expression more than time**
We used our model to optimize the experimental timeline for spider entrainment, application of the light pulse, and tissue collection from the cephalothoraxes of adult female *M. wittfeldae*. These experimental spiders included 10 in the *pulse* and 10 in the *no pulse* groups. The *pulse* group received a 1-hour light pulse five hours (et5-6) after lights-off at the end of the 5th day of LD 12:12 but no light thereafter. The *no pulse* group received no light after the 5th day of LD 12:12. In each group, 5 spiders were sacrificed in the dark 7 hours (et7) and 16 hours (et 16) after the last LD 12:12 lights-off transition (see Figure 1C).

We *de novo* assembled 266,300 transcripts from the 20 individual RNA-seq libraries, representing 184,077 Trinity-defined “genes”. Annotation of the “genes” resulted in 9 categories of sequences, 4 of which were unlikely to encode proteins and were thus removed prior to further analyses (Table 1). Of the retained translated genes, 25,489 were further annotated by significant alignment to a protein in SwissProt and/or PFAM, and 25,127 were assigned GO Terms based on the SwissProt, PFAM, and/or *D. melanogaster* alignments. Comparison with 1066 Benchmarking Universal Single Copy Orthologs (BUSCO v3.0.2, [19]) from arthropods suggests that our transcriptome is of high quality, with 98.4% (91% as single copy and 7.4% as duplicates) of BUSCO genes represented completely and 0.3% represented by fragmented sequences.

| Table 1. Annotation of the *de novo* assembly of *Metazygia wittfeldae*. |
|---------------------------------------------------------------|
| **BEST**: has a reciprocal best BLASTx alignment to a protein in a reference database | 26114 |
| **GOOD**: has BLASTx alignment to a protein in one of the reference databases but another Trinity gene has a better BLAST alignment | 9655 |
**LONGORF**: predicted to encode a protein at least 50 amino acid long, but does not have a significant BLASTx alignment to any of the reference proteins

**LOW_EXP**: same as GOOD, but with < 1 RSEM estimated TPM in any library

**LOW_EXP_LONGORF**: same as LONGORF, but with < 1 RSEM estimated TPM in any library

**CHIMERA**: potential assembly of two different protein coding genes because of a very good alignment to different proteins in the reference databases

**LOW_COV**: same as GOOD, except that it only aligns to < 20% of the database protein

**RNA**: aligns to ribosomal or transfer RNA

**NO_HIT**: has no BLAST alignment and is not predicted to encode a protein > 50 amino acids

Counts are based on the Trinity-defined “gene”. The bottom 4 categories were retained for read mapping and estimation of expression levels with RSEM, but were removed prior to differential expression analyses (see details in Methods). TPM = RSEM [20] estimated Transcripts Per Million mapped transcripts.

We conducted four separate differential expression analyses at the “gene” level (see Methods and Additional File 1: Table S1). MDS (Multidimensional scaling) plots (Additional File 2: Figure S1) illustrate the relative clustering of samples in different conditions. A light pulse in the middle of the night (et5-et6) resulted in individuals having the most similar gene expression profiles of any of the four experimental conditions (Additional File 2: Figure S1). The pulse of light also resulted in many more differentially expressed genes 1-hour after the light pulse than
10 hours later (282 vs. 57, Table 2, Figure 3). Consistent with a large effect of light on gene expression, more genes were differentially expressed between the two collection times in the group that received a light pulse than the one that did not (227 vs. 95, Table 2, Figure 3). The light pulse caused many more genes to be downregulated (70%) than upregulated (30%) relative to no light pulse, when collected at et7 (Additional File 2: Figure S2); however, 10 hours later, approximately equal number of DE genes were upregulated and downregulated (Table 2 and Additional File 2: Figure S2). These findings are consistent with the prediction of our model that light would degrade circadian transcripts, but that expression would subsequently increase over time (Figure 1b).

| Comparison            | # DE genes (FDR < 0.05) | # genes examined |
|-----------------------|-------------------------|-----------------|
| pulse vs. no pulse at et7 | 282 (211 down, 71 up)   | 90341           |
| pulse vs. no pulse at et16 | 57 (30 down, 27 up)     | 102817          |
| et7 vs. et16 with pulse | 227 (165 down, 62 up)   | 100103          |
| et7 vs. et16 with no pulse | 95 (52 down, 43 up)     | 94062           |

Four comparisons were performed (Additional file 1: Table S1) to identify differentially expressed genes (false discovery rate, FDR < 0.05, |log fold change|>1) with edgeR (Robinson and Oshlack 2010), focusing only on genes with > 0.1 CPM in at least two libraries. Spiders received a 1-hour pulse of light 5 hours after lights off (pulse) or did not (no pulse) and were collected 1 hour (et7) or 10 hours (et16) after the end of the light pulse.
Despite a large effect of the 1-hour light pulse on differential expression patterns (Table 2), we found limited overlap in the identities of differentially expressed genes among comparisons (Figure 3). The greatest overlap was between pulse vs. no pulse at et7 and et7 vs. et16 with pulse. The 93 genes that were differentially expressed in both comparisons are strong candidates for light responsive circadian genes.

**Figure 3:** Venn diagram shows unique and shared DE genes in response to light pulse and collection time. The intersections between colors show the number of DE genes in common to different treatments. et7: Differentially expressed transcripts with light pulse vs without light pulse at et7; et16: Differentially expressed transcripts with light pulse vs without light pulse at et16; Pulse: Differentially expressed transcripts at et7 vs et16 with light pulse; No Pulse: Differentially expressed transcripts at et7 vs et16 without light pulse.

**Canonical clock genes are lowly and weakly differentially expressed**

We also investigated the expression patterns of 7 genes known to control the circadian clock in the fruit fly and other arthropods. We identified orthologs of each of these 7 canonical clock genes (*tim, per, cyc, Clk, dbt, cry1, cry2*) in our *M. wittfeldae* transcriptome using a phylogenetic approach (see Methods for details). Of the 7, only *cry2* was found to be significantly differentially expressed (FDR < 0.01). Specifically, *cry2* was more highly expressed at the second collection time (et16) than the first (et7) when not subject to a light pulse (Figure 4a). *Clk*
had a similar pattern of expression as cry2, although none of the comparisons met the FDR threshold for significance (Figure 4a). In addition, clk, cry2, and cyc have significant positively correlated pattern of expression when considering expression levels of all four conditions (Figure 4b). Cyc was also significantly correlated with dbt (Figure 4b). However, cyc and tim were very lowly expressed (<1 TPM in all individuals, Additional File 2, Figure S3).

Despite minimal differential expression of clock genes among our conditions, each clock gene was significantly correlated with numerous DE genes (Cyc: 51, Clk: 71, Dbt: 81, Per: 39, Tim: 32, Cry1: 54, Cry2: 75; Additional File 3, Table S2). The functions of DE genes that significantly correlated with at least one clock gene include transcriptional regulators, metabolism regulation, protein modifications, and embryonic development, among other functions (Additional File 3, Table S2).

**Figure 4:** Expression patterns of seven canonical clock genes. (A) Log FC values of the seven canonical clock genes with corresponding P-values (from edgeR [21]) rounded to two significant digits in response to collection time and light pulse. (B) Correlation coefficients among the seven canonical clock genes. Positive correlation coefficients are colored red while negative values are
colored blue. The numbers in the parentheses represent p values rounded to two significant digits.

**Light elicits a variety of functional changes**

To determine the functional significance of transcripts differentially expressed in our experimental conditions, we compared Gene Ontology (GO) terms associated with the DE genes to GO terms assigned to the entire transcriptome (see Methods). Once we identified GO terms enriched in DE genes (Additional File 4: Table S3), we used the REVIGO reduction analysis tool to visualize their functions (e.g. Figure 5). The pulse of light produced more enriched GO Terms between time points than the constant dark condition (Figure 5, compare C and D). When comparing pulse to no pulse conditions, the earlier collection time led to more enriched GO terms (Figure 5, compare A and B). Comparison of the pulse to no pulse conditions at the early collection time, et 7 (Figure 5A), yielded enrichment of GO-terms in Biological Process (BP) category related to cell death and response to stress. The enriched GO terms suggest that the light pulse affected developmental processes including vesicle-mediated transport, embryo development, and cell division (Additional File 4: Table S3). Also, some regulatory processes were likely modified based on the enriched GO terms membrane organization, vesicle-mediated transport, protein targeting and cytoskeleton organization. Analysis of enriched GO terms in the Cellular Component (CC) category demonstrates that gene expression changes were distributed within the cell, because components included lysosome, Golgi apparatus, nuclear chromosome, and extracellular region. Finally, analysis of Molecular Function (MF) category resulted in enriched GO-terms in structural molecule activity and GTPase activity (Additional File 4: Table S3).
In contrast, comparing the *pulse* to *no pulse* conditions at the late collection point, et16 (Figure 4B), there are only 2 enriched GO-terms in the BP category: cell adhesion and response to stress. The only molecular function enriched was in structural molecule activity (Additional File 4: Table S3).

When comparing the two collection times, the *pulse* condition yielded the largest pool of enriched GO terms (Figure 5C). Our analysis indicates that a wide range of biological processes were affected. Notably, response to stress and cell death were among the enriched GO terms. Also, GO terms related to proliferations were enriched, such as growth, embryo development, and cell differentiation, among others. In addition, enriched GO-terms in this category included some of the regulatory processes such as translation, secondary metabolic processes, etc. Those processes were not localized, but widely distributed within and outside the cell, since cellular component enriched terms included cell, organelle, nuclear chromosome and extracellular space. However, only three molecular functions were enriched: GTPase activity, oxidoreductase activity, unfolded protein binding (Additional File 4: Table S3).

Finally, when comparing the two collection times, the *no pulse* condition yielded only a few enriched terms (Figure 5D). Biological processes included carbohydrate metabolic processes, cell population proliferation, and aging. However, this effect was likely localized to extracellular space because it was the only enriched group in the CC category. In the molecular function category, terms involved ion binding, enzyme regulator activity, and hydrolase activity acting on glycosyl bonds were enriched (Additional File 4: Table S3).
Figure 5: REVIGO scatterplots of the enriched GO cluster representatives related to biological processes. The resulting lists of GO terms (Additional file 4: Table S3) along with their p-values were summarized and visualized by REVIGO reduction analysis tool. (A) Significantly enriched GO terms related to biological processes with light pulse vs without light pulse at et7. (B) Significantly enriched GO terms related to biological processes with light pulse vs without light pulse at et16. (C) Significantly enriched GO terms related to biological processes at et7 vs et16 with light pulse. (D) Significantly enriched GO terms related to biological processes at et7 vs et16 without light pulse. GO terms are represented by bubbles and are clustered according to their semantic similarities to other GO terms in the gene ontology (adjoining circles are most
closely related). Bubble color indicates log10-P-value for the enrichment, and the two ends of the colors are red and blue, depicting lower- and higher p-values respectively. Bubble size indicates the frequency of the GO term in the underlying reference EBI GOA database [22] (bubbles of more general terms are larger).

Flipped patterns of expression predicted by the model identifies numerous circadian genes

Our model predicts that circadian genes follow a well-defined pattern of expression in response to light pulse and collection time. Specifically, at et7 circadian genes should be downregulated by a pulse of light compared to the no pulse condition (Figure 2A, left). Circadian gene expression is expected to be flipped at et16, where they become upregulated by light (Figure 2A, right). As a result, we expected the ratio of mRNA expression between pulse and no pulse conditions (log FC) to reverse between the two collection times, with negative log FC at et7 and positive log FC at et16 (Figure 2C). To visualize this prediction, we introduced a new coordinate system where log FC at et7 is plotted on the horizontal axis and log FC at et16 is plotted on the vertical axis. Our model predicts that log FC at et7 would be negative, thus circadian genes would be located at the left plane. Another prediction of the model is that log FC at et17 would be positive and, therefore, circadian genes would be located on the upper half plane. In summary, our model predicts that, in this coordinate system, circadian light-sensitive genes will be in the upper left quadrant. When we plotted our experimentally determined DE transcripts, we found a large cluster of DE transcripts in the upper left quadrant, in the location predicted by our model (Figure 6A; Additional File 5: Table S4). Therefore, we hypothesize that the transcripts in the upper left quadrant of figure 6A are involved in the circadian response to light.
Similarly, our model predicts that circadian genes should be downregulated at et7 compared to et16 in the pulse condition (Figure 2B, left). Circadian gene expression is expected to be flipped in the no pulse condition (Figure 2B, right). As a result, we expect a negative log FC with a pulse and a positive log FC with no pulse (Figure 2D). In our coordinate system that plots log FC in pulse condition on the horizontal axis, downregulated circadian transcripts would be on the left-half plane. If log FC in no pulse condition is plotted on the vertical axis, the upregulated transcripts would be on the upper half-plane. Like collection time analysis, the circadian light-sensitive genes should be located at the upper left quadrant. Our experimentally determined DE genes in fact show a large cluster of transcripts in the upper left quadrant, where the model predicts our circadian light-sensitive genes would be located (Figure 6B, Additional File 5: Table S4). Therefore, we hypothesize that the transcripts in the third quadrant of figure 5B are involved in circadian response to a pulse of light.

Our proposed light-sensitive circadian genes (upper left quadrants of Figure 6A&B) are enriched for a number of GO terms identified in the 4 separate comparisons (Figure 4) as well as additional GO terms (Table 3). The similar pattern of expression found for these proposed light-sensitive circadian genes is further supported by the fact that the majority of transcripts in each quadrant are significantly correlated with each other when considering all four experimental conditions (56% of the 140 genes in upper left quadrant of Figure 6A; 68% of the 170 genes in upper left quadrant of Figure 6B).
Figure 6: Cluster of transcripts in response to collection time and light pulse. Circles in scatterplots indicate log FC values of differentially expressed genes (FDR < 0.05, |log FC|>1).

(A) Top left corner contains the cluster of genes whose log of ratio (log FC) values between pulse and no pulse condition are negative in et7 and positive at et16 as predicted by the model (Figure 2 A & C). Transcripts differentially expressed between pulse and no pulse at only et7 are red filled dots while genes differentially expressed at only et16 are colored blue. One gene is
differentially expressed at both et7 and et16, colored purple. (B) Top left corner contains the
cluster of genes whose log FC values between et7 and et16 times are negative in pulse condition
and positive in no pulse condition as predicted by the model (Figure 2 B & D). Transcripts
differentially expressed only with light pulse are open circles while transcripts differentially
expressed only without light pulse are filled circles. One gene is differentially expressed with or
without light pulse, colored purple. The larger circles indicate genes identified in both A & B.

Table 3. GO terms enriched in the circadian light-responsive transcripts.
| Cellular Component                  | • extracellular region       | • Golgi apparatus             |
|-----------------------------------|-----------------------------|-------------------------------|
|                                   | • extracellular space       | • plasma membrane             |
|                                   | • lysosome                   | • endoplasmic reticulum       |

| Biological Process                | • signal transduction        | • cell population proliferation |
|-----------------------------------|-----------------------------|---------------------------------|
|                                   | • anatomical structure development | • plasma membrane organization |
|                                   | • cell division              | • membrane organization         |
|                                   | • cell morphogenesis         | • secondary metabolic process   |
|                                   | • transport                  | • cell differentiation          |
|                                   | • cell junction organization | • cell death                    |
|                                   | • reproduction               | • growth                        |
|                                   | • vacuolar transport         | • protein folding               |
|                                   | • anatomical structure formation | • vesicle-mediated transport    |
|                                   | involved in morphogenesis    | • response to stress            |
|                                   | • small molecule metabolic process | • embryo development         |
|                                   | • cell motility              | • cytoskeleton organization     |

| Molecular Function                | • unfolded protein binding   |
|-----------------------------------|-----------------------------|
|                                   | • structural molecule activity |
|                                   | • GTPase activity            |

Light-responsive circadian transcripts are those with a flipped pattern of expression between time points depending on the light experiment (upper left quadrants of Fig. 6 A&B). The GO terms in this table are enriched in these transcripts relative to the entire transcriptome.

**Discussion**

To our knowledge, this study is the first to investigate genetic mechanisms of the circadian clock in spiders. By comparing differential gene expression between two time points under two
light conditions, we were able to identify potential cycling genes, as well as candidate genes for circadian light entrainment.

The total number of DE transcripts between our two collection times (227 for pulse, 95 for no pulse, 2 overlap, Fig. 3) is consistent with previously reported sets of time-dependent D. melanogaster genes. However, our experimental application of a 1-hour light pulse in the middle of the night is unusual. For example, a cDNA chip experiment of 13,500 genes in D. melanogaster identified 120 cycling transcripts by sampling every 4 hours for 2 days in a 12:12 LD cycle; approximately 50% of these also cycled in constant darkness [23]. A GeneChip (13,232 unique probes) analysis identified many more genes, 712, that showed a daily fluctuation in mRNA levels under light-dark conditions; 115 of which also cycled in DD [24]. High-density oligonucleotide arrays identified ~400 oscillating transcripts in DD [25]. A meta-analysis of these 3 studies and 2 additional microarray based studies found 218 significantly cycling transcripts, combining LD and DD conditions [26]. We identified homologs of 80 of these D. melanogaster cycling transcripts in our M. wittfeldae transcriptome but the spider homologs were not differentially expressed between our collection times or between the light pulse and no pulse conditions. The lack of overlap in differentially expressed genes between D. melanogaster and M. wittfeldae suggests that the circadian mechanism, or at least the circadian output, in spiders differs significantly from the fruit fly. The idea of distinct set of circadian genes in spiders was also supported by the finding that a large fraction of DE genes in our data set did not have homologs outside of Araneae (75% of the 385 genes DE between time points).

Our comparison of two time points in constant darkness is most similar to the D. melanogaster experiments described above. Only 95 M. wittfeldae transcripts were differentially expressed under these conditions, which is similar, slightly lower, or considerably lower than the
number reported for *D. melanogaster* in constant dark, depending on the study (see previous paragraph). Several potential explanations can be offered to justify the smaller number of DE transcripts in our experiment. Although we used a mathematical model to select collection time points to amplify the circadian signal, the lack of experimental data on kinetic constants for circadian mRNA and protein might result in some missing transcripts. For example, the transcripts in the middle of the rising and falling phases would have a similar level of expression in our experimental conditions (Fig. 1). In the future, more frequent sampling (at least every 4 hours) should be conducted to identify all cycling transcripts. Nevertheless, we identified a sizable cluster of DE transcripts with the flipped pattern of expression predicted by our model (Figure 6). We thus feel confident that we have captured many light-sensitive circadian genes, even if we have not captured all the cycling transcripts.

An alternative explanation of the small number of DE genes between time points might be an overall lower level of circadian gene expression in spiders. It is possible that the circadian oscillator in spiders has a lower amplitude than other species. Based on simulations with our model, and prior evidence [27–30], we hypothesized that such weaker circadian clock could explain an unusually wide range of free-running periods observed in spiders [13]. It is therefore possible that some evolutionary adaptation in spiders leads to lower level of circadian gene expression.

Another line of evidence that could support the weak oscillator model is the very low expression levels of the canonical clock genes, *tim* and *cyc*, in all conditions. The degree of difference between time points for 6 of the 7 canonical clock genes examined was also low (Figure 4A), consistent with low amplitude oscillations. Only *cry2* met our criteria for significantly different expression levels, and only between the two time points in constant
darkness. Cry2 is a member of the photo-lyase gene family, but its protein product, cryptochrome-2 is not light-sensitive [8]. In multiple arthropod species, CRY2 inhibits CLOCK:CYCLE-mediated transcription [8, 9]. We would thus predict a negative correlation with per and tim (e.g. as in mosquitoes, [31]). We found a weak negative correlation between cry2 and per but a weak positive correlation with tim. On the other hand, cry2 was significantly positively correlated with clk and cyc (Figure 4B), consistent with a possible interaction among their protein products. We also found that the clock genes were significantly correlated with genes encoding functions often under circadian control. It is thus possible that despite low amplitude differences, and unexpected correlations with each other, the canonical clock genes are important for circadian control in spiders. This control, however, may differ dramatically from what has been previously described.

**Short term response to light pulse produces largest set of DE transcripts**

By comparing differential gene expression patterns between pulse and no pulse groups in our experiment, we identified transcripts involved in short- and long-term responses to light. We found a larger number of genes were DE at the earlier collection time than the latter, suggesting an acute short-term response to the light pulse. Interestingly, a sizable portion of those light-responsive transcripts encode proteins with a wide range of cellular functions. Those results are consistent with previous findings in other species. For example, [32] compared response to 1 and 3 hours of 1,200 lux light pulse in zebra fish. They identified 117 light-regulated genes, with the majority being induced and some repressed by light. The regulated genes cover a variety of functions including stress response and DNA repair, consistent with our findings.
Another study measured response of the *Arabidopsis* transcriptome to 1-hour light pulses given either in the middle of the subjective day or in the middle of the subjective night [33]. This study identified a total of 2,237 light-sensitive genes, which is much higher than in our study. A light stimulus during the night preferentially promoted the expression of certain key clock components, consistent with the general observation that light present at the beginning or end of the photoperiod adjusts the circadian clock to seasonal changes in day length.

We were unable to find gene expression studies exposing any arthropod species to a short pulse of light. The most comparable arthropod experiments we could find applied weak light at night. For example, fireflies exposed to weak light at night for 2 weeks experienced much higher mortality than fireflies not exposed to light at night [34]. Furthermore, weak light at night resulted in 1262 down regulated and 105 upregulated genes [34]. We similarly found more downregulated genes in the short-term response to a light pulse in *M. wittfeldae*. Enriched GO terms in the firefly DE genes included functions related to immune response, development, reproduction, and biological rhythm. Some of the immune response and development terms reflect GO terms we found with response to light. The DE genes related to biological rhythm were not the “main” clock genes but ones that interact with the “main” clock – so it is perhaps not surprising that we did not identify an effect of light on canonical clock genes.

Transcripts involved in the circadian response to light relate to immune function, growth, development, and reproduction

Our most striking finding is the cluster of transcripts with switched patterns of differential expression over time, dependent on light pulse, as predicted by our mathematical model (Fig. 6). These transcripts represent our best candidates for light-sensitive circadian genes. These
transcripts had multiple functions and were distributed throughout the cell (Table 3). GO-terms involved in immune response, growth, development, and reproduction were enriched in this candidate set of light-sensitive circadian genes (Table 3). It has been shown in *D. melanogaster* that the immune response is under circadian control [35, 36]. It is also has been found that exposure to night light leads to reductions in immune function in vertebrates and invertebrates [37, 38].

We also found enriched GO-terms for growth, development, and reproduction in our candidate light-sensitive circadian genes, such as “cell morphogenesis”, “embryo development”, “cell differentiation”, and “cell division” (Table 3). These results are consistent with prior evidence that exposure to light at night alters circadian rhythms, inhibits reproduction, accelerates development, and leads to increased mortality (e.g. vertebrates [37], fireflies [34], and *D. melanogaster* [39]). In an Australian nocturnal orb weaving spider, exposure to artificial light at night accelerated juvenile development, resulting in spiders progressing through fewer molts [40]. Light intensity also controls the eclosion circadian rhythm in *D. melanogaster* [41]. Our results for *M. wittfeldae* further support a role of light in affecting the circadian rhythms of immune function, development, and reproduction. Our transcripts with flipped patterns of expression between time points, dependent on light (Fig. 6) may even include ones important for entraining the circadian output to light.

**Conclusions**

To our knowledge, this work is the first to identify circadian and light-responsive transcripts in spiders. We used a mathematical model of the circadian clock in spiders to design a light application experiment, and the subsequent collection times to optimize differences in gene
expression. By comparing the levels of expression at two time points after an application of light, or not, we were able to identify distinct clusters of transcripts responding to time of day and light. These transcripts are our best candidate set for light-sensitive circadian genes. Many of these transcripts likely encode outputs of an entrained circadian clock, including immune response, reproduction, and development. However, it also is possible that some of these transcripts encode central components of the clock that are critical to entrainment. Given the unusual characteristics of spider free-running periods, and the fact that many of our candidate light-sensitive circadian genes have no homologs outside of spiders, spiders may have evolved unique genetic mechanisms of entrainment and intrinsic circadian cycling.

Methods

Mathematical Model

Briefly, the minimal model of spider circadian clock was based on a previously developed Neurospora model [42] that involves a feedback loop among (i) a specific mRNA, (ii) the cytosolic protein encoded by this mRNA, and (iii) the gradual phosphorylation of the cytosolic protein, which changes its structure, enabling it to enter the nucleus. As illustrated in Figure 1(a), the mRNA, produced from a gene through the process of transcription, is positively coupled to the cytosolic protein through the process of translation. After the cytosolic protein changes its structure and re-enters the nucleus, the resulting nuclear protein acts to suppress expression of the gene coding for the mRNA, representing a negative coupling. Natural degradation of the mRNA sequence and the cytosolic protein is also included in the model.

The coupled system of equations governing the circadian oscillator is as follows:

\[
\frac{dM}{dt} = v_s \frac{K^2_{1}}{K^2_{1} + P^2_N} - v_m \frac{M}{K_m + M} - L M
\]
\[
\frac{dP_c}{dt} = k_s M - v_d \frac{P_c}{K_d + P_c} - k_1 P_c + k_2 P_N
\]

\[
\frac{dP_N}{dt} = k_1 P_c - k_2 P_N
\]

The three variables \( M \), \( P_C \), and \( P_N \) represent the concentrations (with respect to cell volume) of the mRNA, the cytosolic protein, and the nuclear protein, respectively. The parameters of the are minimally modified from the original model [42]. The rate of mRNA production is inhibited by nuclear protein, \( P_N \), with second order rate parameters \( K_1 = 1 \, nM^{1/2}, v_s = 1.6 \, \frac{nM}{hr}. \) mRNA also degrades with a first order rate parameters \( v_m = 0.5 \, \frac{nM}{hr}, K_m = 0.5 \, nM \). Exposure to light induces additional, exponential rate of mRNA degradation with parameter \( L = hr^{-1} \). mRNA is translated into protein with effective rate \( k_s = 0.5 \, hr^{-1} \), undergoes first order degradation with parameters \( v_d = 1.4 \, \frac{nM}{hr}, K_d = 0.15 \, nM \), and is transported to nucleus and back with linear rates respectively \( k_1 = 0.5 \, hr^{-1}, k_2 = 0.6 \, hr^{-1}. \)

**Experimental Setup**

We used the model to predict the time to apply a 1-hour light pulse that would maximize the difference in gene expression levels between spiders receiving the light pulse and those that did not. We also selected two collection points to optimize the ratio of mRNA at two collection points for each experimental group. The final experimental protocol is shown in Figure 1 C.

Adult female *Metazygia wittfeldae* were collected on 16 July 2018 in Washington County, TN and placed in clear plastic deli jars on a lab bench receiving natural day, night lighting. All spiders were provided with a wet cotton ball and offered a small cricket. On 18 July 2018, each spider was moved to a clear glass or plastic tube (25 mm diameter x 15 cm length) and inserted
into a locomotor activity monitor (model LAM25, Trikinetics Inc., Waltham, Massachusetts, U.S.A.). The monitors were placed into two separate temperature-controlled (24 ± 0.5 °C) environmental chambers with the lights on (1400-1600 lux) at ~16:00 until 20:00, after which the spiders remained in the dark for 3.5 days (84 hours) to establish the free-running period of each individual. In order to synchronize the circadian rhythms of the individual spiders, they were then entrained to a 12:12 LD cycle for 5 days. On the evening of the 5th day the lights went out at 20:00 (et0), as per the prior 4 days. However, one chamber received 1 hour of light (1400-1600 lux) starting 5 hours after the lights were turned off (et5-et6, 01:00-02:00 27 July 2018) (pulse group). The other chamber remained dark (no pulse group). One hour after the end of the light pulse (et7, 03:00) 5 spiders were removed from each environmental chamber and snap frozen in liquid nitrogen under dim red light. Both chambers remained dark until the final collection time 9 hours later (et16, 12:00), when 5 spiders were again removed from each chamber and snap frozen in liquid nitrogen under infrared light.

**RNA-sequencing, transcriptome assembly, and annotation**

Total RNA was isolated from one half of a cephalothorax (the fused head-body) of each of the 20 experimental spiders by homogenizing tissue in TRIzol (Invitrogen) and further purifying with the RNeasy Mini Kit (Qiagen) with on-column DNaseI digestion to remove contaminating DNA. RNA quality and quantity were verified with the Agilent 2100 Bioanalyzer and 20 individually barcoded cDNA libraries were constructed with the TruSeq kit (Illumina) by the Genomics Research Laboratory at the Biocomplexity Institute, Virginia Tech. All 20 libraries were multiplexed and sequenced in a single lane with the NextSeq 500 (Illumina) using the High
Output mode with 300 cycles (150 base paired end reads). De-multiplexing and barcode removal were performed by Virginia Tech.

Transcripts were de novo assembled with Trinity v2.8.4 [43] from all 20 RNAseq libraries. Raw reads were trimmed of low quality base calls using Trimmomatic [44] and in silico normalization was used to reduce the number of reads entering the assembly phase, as per scripts in Trinity v2.8.4.

Annotation involved a multi-step process using the Transcriptome Trimming and Annotation Pipeline (TrTAP, [45]). In brief, initial transcript abundance was estimated using RSEM v.1.3.1 [20]. Trimming and annotation then started with any transcripts with a significant BLASTN match to the SILVA ribosomal database v132 [46] or a tRNAscan v.1.3.1 match [47] were identified. Second, all the transcripts were compared to a custom-made database of spider silk proteins encoded by the spidroin gene family [48], and subsequently to the proteomes of 5 arthropod species using BLASTx. These species included 3 spiders (listed from most closely related to least): Araneus ventricosus (GCA_013235015.1_Ave_3.0, Kono et al. 2019)), Trichonephila clavipes (MWRG00000000.1, [50]), and Stegodyphus mimosarum (GCA_000611955, [51]); another arachnid, Ixodes scapularis (GCF_016920785.1,Gulia-Nuss et al. 2016); and an insect, Drosophila melanogaster (FB2019_08 [53]) The Drosophila gene matches with a cutoff of 1e-40 were used to identify chimeric sequences with a python script described in [54].

For each Trinity “gene”, the transcript with the best BLAST alignment, alignment as determined by the highest bit-score, or with the longest open reading frame (ORF) if there were no BLAST alignments, was chosen to represent that gene. To further reduce redundancy and identify fragments of genes, the best matching gene of each protein in the curated silk gene
database plus the 5 species proteomes were also identified. These reciprocal best matches were always retained for downstream analyses (BEST, Table 1). If a Trinity gene was not the best match of a database protein it was considered “GOOD” if it aligned to >20% of the database protein and exceeded 1 TPM (transcript per million) in at least one RNAseq library. Genes that aligned to >20% of the database protein and did not meet the 1 TPM threshold were dubbed “LOW_EXP” (Table 1). Genes that aligned to <20% of the database proteins were dubbed “LOW_COV” (Table 1). For the genes with no BLAST alignments to a proteome, the gene was called “LONGORF” if the ORF exceeded 50 amino acids and had >1 TPM in at least one RNAseq library (Table 1). ORFs that exceeded 50 amino acids but had <1 TPM were called “LOW_EXP_LONGORF” (Table 1). A reduced set of probable protein-coding genes (BEST, GOOD, LONGORF, and LOWEXP) was then translated using the direction and frame of the matching BLASTX hit, or in cases without such hit, translated according to the longest open reading frame. These translated genes were further annotated by comparison to SwissProt with BLASTP, and to PFAM with HMMER v 3.2.1[55] with Gene Ontology (GO) terms [56, 57] assigned to each “gene” based on the best alignment to *Drosophila melanogaster*, SwissProt, and PFAM, with GO SLIM annotations obtained from GO SLIM viewer [58]. The completeness of the reduced transcriptome was assessed with the set of single copy orthologous genes in arthropods (BUSCO v3, [19]).

**Canonical circadian gene identification**

Our initial BLAST results identified multiple transcripts with significant alignments to the canonical circadian genes: *Clk, cyc, tim, per, dbt, cry1, cry2*. Since these genes are each part of larger multi-paralog gene families, we used a phylogenetic approach to determine which
transcript was an ortholog of the canonical clock gene in *D. melanogaster* or *Apis mellifera*. We first used Orthofinder v2. [59] to identify groups of homologous genes among 6 spider species, a tick, and two insects. In addition to our *M. wittfeldiae* transcriptome, we used the predicted proteomes from published transcriptomes or genomes for the following species: *Araneus ventricosus* (Araneidae, BGPR01000000.1), *Trichonephila clavipes* (Araneidae, MWRG01000000.1), *Latrodectus hesperus* (Theridiidae, GBJN01000000.1), *Parasteatoda tepidariorum* (Theridiidae, GCF_000365465.3), *Stegodyphus mimosarum* (Eresidae, GCA_000611955.2), *Ixodes scapularis* (Ixodidae, GCF_016920785.1), *D. melanogaster* (Drosophilidae, FB2019_08), *A. mellifera* (Apidae, GCF_003254395.1). We then identified the “orthogroups” containing the canonical circadian gene of *D. melanogaster* or *A. mellifera* (Table 4). For 4 of the genes, only a single *M. wittfeldiae* transcript was placed in the orthogroup; these were considered the ortholog of the circadian gene. For *Doubletime*, *dbt*, and the cryptochromes (*cry1* and *cry2*), we inferred phylogenetic trees for their orthogroups using maximum likelihood (RAxML,[60]. From the resulting trees, we chose the single *M. wittfeldiae* transcripts in the clade with the canonical circadian clock gene (Additional File 2: Figures S4-S5).

| ID  | Drosophila Gene(s) | Orthogroup ID | Metazygia Gene                       |
|-----|--------------------|---------------|--------------------------------------|
| Cycle | **FBpp0074693** | OG0006887 | **MW_TRINITY_DN59140_c0_g1** |
| Clock | FBpp0076500 FBpp0099478 FBpp0099480 FBpp0306709 FBpp0306710 | OG0003118 | **MW_TRINITY_DN2943_c0_g1** |
| Doubletime | FBpp0085104 FBpp0085105 **FBpp0085106** FBpp0306615 FBpp0422974 | OG0001802 | **MW_TRINITY_DN15176_c0_g1** **MW_TRINITY_DN4945_c0_g1** |
| Period | FBpp0070455 FBpp0304590 | OG0004797 | **MW_TRINITY_DN146_c0_g1** |
The circadian genes (bold) were identified for *D. melanogaster* from FlyBase v2019_08. In cases where more than one *M. wittfeldae* transcript was in an orthogroup with a circadian gene, phylogenetic trees were examined for orthology (Additional File 2: Figures S4-S5). The ortholog is bolded.

### Differential gene expression analysis

Differential gene expression among experimental groups was based on gene-level expression estimates derived from transcript abundance estimates as recommended by [61]. In brief, the expected read counts of each Trinity-assembled transcript were calculated by RSEM v.1.3.1 [20], which takes into account read-mapping ambiguity due to multiple isoforms or even alleles having been assembled. The gene-level counts were then calculated as the sum of the included transcripts, weighted by their length as described in Soneson et al. [61], and implemented through Trinity v2.8.4. Because RSEM can distribute one read among multiple transcripts, some counts were not integer values. These values were rounded down for input into differential expression analyses, which require integer values for read counts. Prior to further expression analyses, genes unlikely to encode proteins were removed (RNA, CHIMERA, LOW_COV, and NO_HITS, Table 1). For comparing expression levels among circadian genes, we used the TPM normalized for differences in sequencing depth among libraries using the Trimmed Mean of M (TMM) values calculated with Trinity.
Four different pairwise gene expression analyses were performed on the read count data using the ‘EdgeR’ Bioconductor package [62]: 1) comparing samples with light pulse to samples without light pulse at et7; 2) comparing samples with light pulse to samples without light pulse at et16; 3) comparing samples collected at et7 to samples collected at et16 with a light pulse; 4) comparing samples collected at et7 to samples collected at et16 without a light pulse. A gene is kept in the analysis if it is sufficiently expressed (CPM > 0.1) in at least two samples. The raw library sizes were normalized using the TMM method from EdgeR. The significantly differentially expressed genes in the pair-wise comparisons were identified using false discovery rate (FDR) < 0.05 and absolute value of log FC > 1 (Additional file 1: Table S1).

**Gene Ontology analysis**

GO analysis was conducted to determine the functions enriched in DE genes relative to all annotated genes using the ‘GOseq’ Bioconductor package [63] based on Wallenius non-central hyper-geometric distribution, with an over represented p-value of 0.1. The lists of GO terms (Additional file 4: Table S3) along with their p-values generated from ‘Goseq’ were summarized and visualized by REViGO online tool [64]. The semantically similar terms of the identified GO terms were removed using SimRel semantic clustering. Scatterplots of the remaining terms after the redundancy reduction were generated using the source code from REViGO, which plotted the terms in a two-dimensional space by applying multidimensional scaling to a matrix of the GO terms' semantic similarities (Figure 4) [64].

**Data visualization and graphical output generation**
In order to check the dissimilarities among the samples, MDS plots (Additional File 2, Figure S1) were generated using the ‘limma’ Bioconductor R package [65]. The distance between each pair of samples is calculated as the leading fold change, defined as the root-mean-square of the largest 500 log2-fold changes between that pair of samples.

Venn diagrams (Figure 3) were generated to identify shared and unique DE genes between various contrast groups based on the DE analysis results (Additional File 1: Table S1) using the freely available VENNY 2.1 online tool [66].

Bar plots of up- and down- regulated DEGs, volcano plots of DEGs as well as scatterplots of logFC values (Additional File 2: Figure S2; Figure 5) were generated using the ‘ggplot2’ R package [67].

List of abbreviations

BP: Biological Process
CC: Cellular Component
Clk: Clock
CPM: Counts Per Million
Cry1: Cryptochrome 1
Cry2: Cryptochrome 2
Cyc: Cycle
Dbt: Doubletime
DD: Constant darkness
DE: Differentially expressed
Et7: The group collected 7 hours after the last light phase
**Et16**: The group collected 16 hours after the last light phase

**FC**: Fold Change

**FDR**: False Discovery Rate

**FRP**: Free-running period

**GO**: Gene Ontology

**LD**: Light-dark environments

**MDS**: Multidimensional scaling

**MF**: Molecular Function

**No pulse**: The group receiving no light pulse

**ORF**: Open reading frame

**PDEA**: pairwise differential expression analysis

**Per**: Period

**Pulse**: The group receiving a light pulse

**Tim**: Timeless

**TPM**: Transcripts Per Million

**TMM**: Trimmed mean of M values

---

**Declarations**

- Ethics approval and consent to participate

The experimental animals in this study, *Metazygia wittfeldae*, are not regulated.

- Consent for publication
Not applicable.

- Availability of data and materials

The datasets supporting the conclusions of this article are available within the article (and its additional files) and in NCBI under the Short Read Archive (SRR16156593- SRR16156612) and Transcriptome Shotgun Assembly (TSA: GJLH0000000).

- Competing interests

The authors declare that they have no competing interests.

- Funding

This work was supported by Washington and Lee University through Summer Lenfest Research Grants to NT and NAA and Summer Research Scholarships to AM and CSAL, the National Science Foundation [Award number IOS-1755142 to NAA], and Eastern Tennessee State University through a Research and Development Committee Major Grant.

- Authors' contributions

NT, AM, TCJ, DM, and NAA conceived and designed the experiments. AM, TCJ, DM, and NAA collected the data. NT and AM and developed the mathematical model. NT, WC, LQ, TC, and NAA analyzed the data. All authors read and approved the final manuscript.

- Acknowledgements
We thank Hyun Song and C.S. Adrian Lam for assistance with early analyses. We also thank Madeleine Miller for help collecting spiders.

- Authors' information (optional)

Bibliography

1. Edery I. Circadian rhythms in a nutshell. Physiol Genomics. 2000;3:59–74.

2. Wager-Smith K, Kay SA. Circadian rhythm genetics: from flies to mice to humans. Nat Genet. 2000;26:23–7.

3. DeCoursey PJ. The behavioral ecology and evolution of biological timing systems. In: Chronobiology: Biological timekeeping. Sunderland, MA, US: Sinauer Associates; 2004. p. 27–65.

4. Johnson CH. Precise circadian clocks in prokaryotic cyanobacteria. Curr Issues Mol Biol. 2004;6:103–10.

5. Refinetti R. Comparison of light, food, and temperature as environmental synchronizers of the circadian rhythm of activity in mice. J Physiol Sci. 2015;65:359–66.

6. Golombek DA, Rosenstein RE. Physiology of circadian entrainment. Physiol Rev. 2010;90:1063–102.

7. Zeng H, Qian Z, Myers MP, Rosbash M. A light-entrainment mechanism for the Drosophila circadian clock. Nature. 1996;380:129–35.

8. Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Mol Biol Evol. 2007;24:948–55.

9. Zhu H, Sauman I, Yuan Q, Casselman A, Emery-Le M, Emery P, et al. Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. PLoS Biol. 2008;6. doi:10.1371/journal.pbio.0060004.

10. Moore D, Watts JC, Herrig A, Jones TC. Exceptionally short-period circadian clock in Cyclosa turbinata: regulation of locomotor and web-building behavior in an orb-weaving spider. J Arachnol. 2016;44:388–96.

11. Ralph MR, Menaker M. A mutation of the circadian system in golden hamsters. Science. 1988;241:1225–7.
12. Konopka RJ, Benzer S. Clock mutants of *Drosophila melanogaster*. Proc Natl Acad Sci. 1971;68:2112–6.

13. Mah A, Ayoub N, Toporikova N, Jones TC, Moore D. Locomotor activity patterns in three spider species suggest relaxed selection on endogenous circadian period and novel features of chronotype. J Comp Physiol A. 2020;206:499–515.

14. Jones TC, Wilson RJ, Moore D. Circadian rhythms of locomotor activity in *Metazygia wittfeldae* (Araneae: Araneidae). J Arachnol. 2018;46:26–30.

15. Czeisler CA, Duffy JF, Shanahan TL, Brown EN, Mitchell JF, Rimmer DW, et al. Stability, precision, and near-24-Hour period of the human circadian pacemaker. Science. 1999;284:2177–81.

16. Ebihara S, Tsuji K, Kondo K. Strain differences of the mouse’s free-running circadian rhythm in continuous darkness. Physiol Behav. 1978;20:795–9.

17. Helm B, Visser ME. Heritable circadian period length in a wild bird population. Proc R Soc B Biol Sci. 2010;277:3335–42.

18. Shimizu T, Masaki S. Geographical and species variation in circadian rhythm parameters in nemobine crickets. Physiol Entomol. 1997;22:83–93.

19. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO Applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018;35:543–8.

20. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

21. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

22. Huntley RP, Sawford T, Mutowu-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, et al. The GOA database: Gene Ontology annotation updates for 2015. Nucleic Acids Res. 2015;43:D1057–63.

23. Ceriani MF, Hogenesch JB, Yanovsky M, Panda S, Straume M, Kay SA. Genome-wide expression analysis in *Drosophila* reveals genes controlling circadian behavior. J Neurosci. 2002;22:9305–19.

24. Ueda HR, Matsumoto A, Kawamura M, Iino M, Tanimura T, Hashimoto S. Genome-wide transcriptional orchestration of circadian rhythms in *Drosophila*. J Biol Chem. 2002;277:14048–52.

25. Claridge-Chang A, Wijnen H, Naef F, Boothroyd C, Rajewsky N, Young MW. Circadian regulation of gene expression systems in the *Drosophila* head. 2001;32:657–71.
26. Keegan KP, Pradhan S, Wang J-P, Allada R. Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes. PLOS Comput Biol. 2007;3:e208.

27. Abraham U, Granada AE, Westermark PO, Heine M, Kramer A, Herzl H. Coupling governs entrainment range of circadian clocks. Mol Syst Biol. 2010;6:438.

28. Brown SA, Kunz D, Dumas A, Westermark PO, Vanselow K, Tilmann-Wahnschaffe A, et al. Molecular insights into human daily behavior. Proc Natl Acad Sci U S A. 2008;105:1602–7.

29. Kurosawa G, Goldbeter A. Amplitude of circadian oscillations entrained by 24-h light–dark cycles. J Theor Biol. 2006;242:478–88.

30. Vaze KM, Helfrich-Förster C. Drosophila ezoana uses an hour-glass or highly damped circadian clock for measuring night length and inducing diapause. Physiol Entomol. 2016;41:378–89.

31. Gentile C, Rivas GBS, Meireles-Filho ACA, Lima JBP, Peixoto AA. Circadian expression of clock genes in two mosquito disease vectors: cry2 is different. J Biol Rhythms. 2009;24:444–51.

32. Weger BD, Sahinbas M, Otto GW, Mracek P, Armant O, Dolle D, et al. The light responsive transcriptome of the zebrafish: function and regulation. PLoS ONE. 2011;6:e17080.

33. Rugnone ML, Faigon Soverna A, Sanchez SE, Schlaen RG, Hernando CE, Seymour DK, et al. LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator. Proc Natl Acad Sci. 2013;110:12120–5.

34. Chen Y-R, Wei W-L, Tzeng DTW, Owens ACS, Tang H-C, Wu C-S, et al. Effects of artificial light at night (ALAN) on gene expression of Aquatica ficta firefly larvae. Environ Pollut. 2021;281:116944.

35. Lee J-E, Edery I. Circadian regulation in the ability of Drosophila to combat pathogenic infections. Curr Biol CB. 2008;18:195–9.

36. McDonald MJ, Rosbash M. Microarray analysis and organization of circadian gene expression in Drosophila. Cell. 2001;107:567–78.

37. Bedrosian TA, Fonken LK, Walton JC, Nelson RJ. Chronic exposure to dim light at night suppresses immune responses in Siberian hamsters. Biol Lett. 2011;7:468–71.

38. Durrant J, Michaelides EB, Rupasinghe T, Tull D, Green MP, Jones TM. Constant illumination reduces circulating melatonin and impairs immune function in the cricket Teleogryllus commodus. PeerJ. 2015;3:e1075.

39. McLay LK, Nagarajan-Radha V, Green MP, Jones TM. Dim artificial light at night affects mating, reproductive output, and reactive oxygen species in Drosophila melanogaster. J Exp Zool Part Ecol Integr Physiol. 2018;329:419–28.
40. Willmott NJ, Henneken J, Selleck CJ, Jones TM. Artificial light at night alters life history in a nocturnal orb-web spider. PeerJ Inc.; 2018. doi:10.7287/peerj.preprints.26943v1.

41. Myers EM, Yu J, Sehgal A. Circadian control of eclosion. Curr Biol. 2003;13:526–33.

42. Leloup J-C, Gonze D, Goldbeter A. Limit cycle models for circadian rhythms based on transcriptional regulation in Drosophila and Neurospora. J Biol Rhythms. 1999;14:433–48.

43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011;29:644–52.

44. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

45. Clarke TH. TrTAP: Transcriptome Trimming and Annotation Pipeline. https://github.com/thclarke/TrTAP.

46. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41 Database issue:D590–6.

47. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol Biol Clifton NJ. 2019;1962:1–14.

48. Ayoub NA, Friend K, Clarke T, Baker R, Correa-Garhwal S, Crean A, et al. Protein composition and associated material properties of cobweb spiders’ gumfoot glue droplets. Integr Comp Biol. 2021. doi:10.1093/icb/icab086.

49. Kono N, Nakamura H, Ohtoshi R, Moran DAP, Shinohara A, Yoshida Y, et al. Orb-weaving spider Araneus ventricosus genome elucidates the spidroin gene catalogue. Sci Rep. 2019;9:8380.

50. Babb PL, Lahens NF, Correa-Garhwal SM, Nicholson DN, Kim EJ, Hogenesch JB, et al. The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. Nat Genet. 2017;49:895–903.

51. Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlund TF, Gupta V, et al. Spider genomes provide insight into composition and evolution of venom and silk. Nat Commun. 2014;5:3765.

52. Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. Nat Commun. 2016;7. doi:10.1038/ncomms10507.

53. Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, et al. FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res. 2021;49:D899–907.
54. Clarke TH, Garb JE, Hayashi CY, Haney RA, Lancaster AK, Corbett S, et al. Multi-tissue transcriptomics of the black widow spider reveals expansions, co-options, and functional processes of the silk gland gene toolkit. BMC Genomics. 2014;15:365.

55. Eddy S. HMMER: biosequence analysis using profile hidden Markov models. www.hmmer.org.

56. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

57. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021;49:D325–34.

58. McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, et al. AgBase: a functional genomics resource for agriculture. BMC Genomics. 2006;7:229.

59. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.

60. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

61. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research. 2016;4:1521.

62. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25.

63. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11:R14.

64. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLOS ONE. 2011;6:e21800.

65. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47–e47.

66. Oliveros J. Venny 2.1.0. https://bioinfogp.cnb.csic.es/tools/venny/. Accessed 9 Oct 2021.

67. Hadley Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. https://ggplot2.tidyverse.org.