Tormey et al. BMC Genomics (2015) 16:754
DOI 10.1186/s12864-015-1937-y

BMC Genomics

RESEARCH ARTICLE Open Access

Evolutionary divergence of core and post-translational circadian clock genes in the pitcher-plant mosquito, *Wyeomyia smithii*

Duncan Tormey[1,2], John K. Colbourne[3,4], Keithanne Mockaitis[3,5], Jeong-Hyeon Choi[3,6], Jacqueline Lopez[3,7], Joshua Burkhart[1,8], William Bradshaw[1*] and Christina Holzapfel[1]

Abstract

Background: Internal circadian (*circa*, about; *dies*, day) clocks enable organisms to maintain adaptive timing of their daily behavioral activities and physiological functions. Eukaryotic clocks consist of core transcription-translation feedback loops that generate a cycle and post-translational modifiers that maintain that cycle at about 24 h. We use the pitcher-plant mosquito, *Wyeomyia smithii* (subfamily Culicini, tribe Sabethini), to test whether evolutionary divergence of the circadian clock genes in this species, relative to other insects, has involved primarily genes in the core feedback loops or the post-translational modifiers. Heretofore, there is no reference transcriptome or genome sequence for any mosquito in the tribe Sabethini, which includes over 375 mainly circumtropical species.

Methods: We sequenced, assembled and annotated the transcriptome of *W. smithii* containing nearly 95 % of conserved single-copy orthologs in animal genomes. We used the translated contigs and singletons to determine the average rates of circadian clock-gene divergence in *W. smithii* relative to three other mosquito genera, to *Drosophila*, to the butterfly, *Danaus*, and to the wasp, *Nasonia*.

Results: Over 1.08 million cDNA sequence reads were obtained consisting of 432.5 million nucleotides. Their assembly produced 25,904 contigs and 54,418 singletons of which 62 % and 28 % are annotated as protein-coding genes, respectively, sharing homology with other animal proteomes.

Discussion: The *W. smithii* transcriptome includes all nine circadian transcription-translation feedback-loop genes and all eight post-translational modifier genes we sought to identify (Fig. 1). After aligning translated *W. smithii* contigs and singletons from this transcriptome with other insects, we determined that there was no significant difference in the average divergence of *W. smithii* from the six other taxa between the core feedback-loop genes and post-translational modifiers.

Conclusions: The characterized transcriptome is sufficiently complete and of sufficient quality to have uncovered all of the insect circadian clock genes we sought to identify (Fig. 1). Relative divergence does not differ between core feedback-loop genes and post-translational modifiers of those genes in a Sabethine species (*W. smithii*) that has experienced a continual northward dispersal into temperate regions of progressively longer summer day lengths as compared with six other insect taxa. An associated microarray platform derived from this work will enable the investigation of functional genomics of circadian rhythmicity, photoperiodic time measurement, and diapause along a photic and seasonal geographic gradient.

Keywords: Transcriptome, Transcription-translation feedback, Post-translational modifiers, Gene sequencing, Gene alignment, Biological clocks, Culicidae, Sabethini

* Correspondence: mosquito@uoregon.edu
[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA
Full list of author information is available at the end of the article

BioMed Central

© 2015 Tormey et al. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

The rotation of the earth about its axis generates a daily cycle of light, temperature, moisture and resources that ultimately affect the microclimate and fitness of organisms [1–5]. A general property of Eukaryotes is that they possess an internal, self-sustaining circadian (*circa*, about; *dies*, day) clock that results in the anticipation and preparation for daily changes in both their external and internal environments [6–9]. Circadian rhythms "are inherent in and pervade the living system to the extent that they are fundamental features of its organization; and to an extent that if deranged, they impair it" ([6], p. 159). Indeed, studies from prokaryotes to mammals have shown that impairment of the circadian clock or imposition of daily environmental cycles that deviate from the innate duration or period of the circadian clock results in reduced fitness [6, 7, 10]. Even if the period of the clock is exactly 24 h, the clock will be able to track the daily cycle of light and dark if the oscillator driving the rhythm varies in its responsiveness to light through the daily cycle [11, 12]. Hence, life in a 24-h world should impose stabilizing selection for a biological clock with an innate period of about 24 h.

At the core of all eukaryotic circadian clocks are transcriptional-translational feedback loops (TTFL, pink in Fig. 1) [13, 14]. The concept of the TTFL existed before any clock genes were known [15] and has been described as comprising the "core" or canonical clock genes. Very quickly, it was recognized in *Drosophila* that the TTFL consisted of positive-acting elements (CLK/CYC) and negative-acting elements (PER/TIM) with input of light through CRY1 (aka dCRY) and its interaction with TIM and SGG. Subsequently, the PDP1, VRI, KAYα, and CWO feedback loops have been shown to interact with and regulate transcription in the CLK/CYC – PER/TIM cycle. We also included in our analyses CRY2 (aka mCRY) because, unlike in *Drosophila*, it is known to be a transcriptional regulator of TTFL genes in mosquitoes, Lepidoptera, Hemiptera, Orthoptera, and Hymenoptera, as well as mice [16–20].

Straightforward kinetics of the TTFL estimate that, unmodified, the TTFL would complete its cycle in a few hours [14, 21–24] and therefore be poor at orchestrating daily events. This observation elevated the appreciation of post-translational modifiers (PTM, blue in Fig. 1) that act as modulators (governors), delaying this cycle and thereby producing a rhythm of about 24 h [21–36]. It is the quality and quantity of phosphorylation by the PTMs that determine the kinetics of the negative-acting loop and, hence, the period of the circadian clock and ultimate degradation of the TTFL proteins [21, 24, 31–36]. Hence, it has been proposed that the post-translational or the post-transcriptional modifiers are more responsible for maintenance of a biological clock with a period

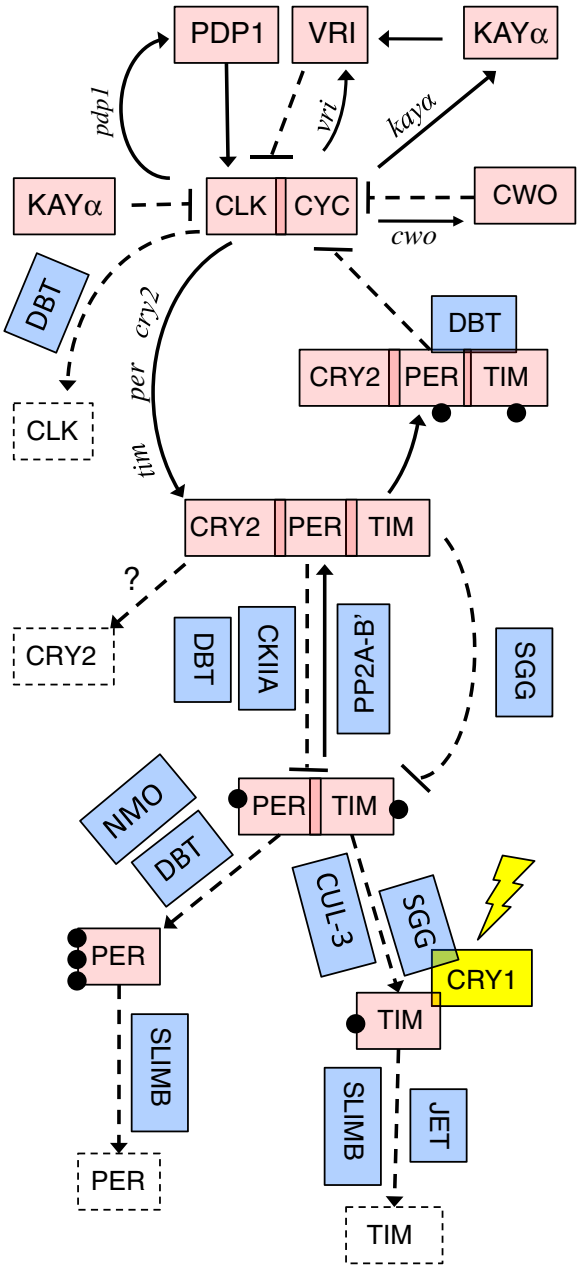

Fig. 1 Functional clockworks of the genes listed in Table 2. Pink: TTFL genes, the core transcription-translation feedback loop consists of positive-acting CLK and CYC and negative-acting CRY2, PER, and TIM; their cycling is affected by "stabilizing" loops involving CWO, KAYα, VRI, and PDP1. Blue: PTM genes, the duration of the circadian cycle is then altered by a number of post-translational modifiers, mainly kinases and phosphatases. Yellow: Entrainment of the circadian clock by external day and night is achieved via the blue-light receptor CRY1. Clear dashed boxes: phosphorylation or ubiquitination leading to ultimate protein degradation. Solid arrows: enhancing transcription or PP2A-B' reversing phosphorylation of PER. Dashed lines: inhibiting transcription or promoting phosphorylation. Upper case Roman, proteins; lower case Italic, transcripts promoted by CLK and CYC. Solid black circles: phosphate groups (compiled from [17, 25, 26, 30, 123, 125, 127])

of about 24 h than is the TTFL [21–23, 36–42]. This proposition would predict that PTM genes should be evolutionarily more conservative than TTFL genes.

Herein, we investigate the relative evolutionary rates of TTFL and PTM genes using the mosquito *Wyeomyia smithii*. The roles of post-transcriptional control [37, 40–42], micro-RNAs [38, 39, 43], *O*-GlcNAcylation [44], and histone acetylation and methylation [22, 45, 46] in circadian time-keeping are still emerging areas of research, especially in insects. Hence, We focused on the TTFL and phosphorylation-related PTM as the two best documented groups of genes involved in circadian rhythmicity that possessed both distinct roles in the circadian clock (TTFL vs. PTM) and distinct biochemical mechanisms (regulation of gene transcription vs. modification of protein stability).

The genus *Wyeomyia* is a member of the tribe Sabethini, which includes some 429 circumtropical species [47], only one of which, *W. smithii*, has invaded temperate North America, likely from tropical South America [48]. *Wyeomyia smithii* completes its pre-adult development only in the water-filled leaves of the carnivorous plant *Sarracenia purpurea* and has dispersed northwards from the Gulf of Mexico to northern and western Canada [49–51]. Over a similar south to north geographic range, the oviposition rhythm of *Drosophila melanogaster* has shown a decline in amplitude, and the eclosion rhythms of *D. subobscura* and *D. littoralis* have shown a decline in both amplitude and period [52]. This latitudinal gradient in period and amplitude of the circadian clock has been attributed to summer day length, which increases with latitude, thereby imposing selection for an increasingly robust oscillator, although evidence supporting this proposition remains equivocal [52, 53]. Regardless of the ultimate causality of the latitudinal gradient in *Drosophila*, *W. smithii* has encountered the same gradient in summer day lengths and we ask whether there has been greater rates of divergence in PTM or TTFL genes in a northern, derived population of *W. smithii* relative to other insects. We focus on a northern population of *W. smithii* first because we were able to use the recently collected F2 of field-collected larvae that reflect the genomics of a natural population. Second, we have over 30 years experience working with the genetics, evolution, physiology, and population biology of *W. smithii* from the Gulf of Mexico to northern Canada, including this particular population (http://www.uoregon.edu/~mosquito). We are therefore able to place our ongoing genomics experiments into a broader context relating to the bionomics of the focal species. Finally, this population represents a more polar population than any other Sabethine mosquito; the only other temperate Sabethine (*Trypteroides bambusa*) occurs in East Asia and does not reach the latitude (46 °N) of the focal population [54]. Hence, this population represents a more northern and, therefore, is more likely to parallel *Drosophila* in the northern, post-glacial divergence of its circadian clock than any other Sabethine species.

At present, there are no sequenced genomes or transcriptomes available for any member of the circumtropical mosquito tribe Sabethini, among which several Neotropical species, including members of the genus *Wyeomyia*, but not including *W. smithii*, have been implicated in the transmission of arboviruses [55, 56]. We therefore produced the first Sabethine transcriptome sequence, assembly and gene annotation. We compared amino acid substitutions from translated *W. smithii* sequences with annotated circadian clock genes in other insects and compared the sequence divergence between *W. smithii* and six other taxa of increasing phylogenetic divergence : mosquitoes in the same subfamily but different tribes (*Aedes* and *Culex*), a mosquito in a different subfamily (*Anopheles*), another Diptera in a different sub-order (*Drosophila*), and progressively more distant orders (Lepidoptera, *Danaus*; Hymenoptera, *Nasonia*). We compared evolutionary rates using nine genes of the TTFL with eight key genes of the PTM (Fig. 1). All six species we considered exhibit circadian rhythmicity under daily and constant conditions [16, 25, 27, 52, 57–60]. Finally, we estimated evolutionary divergence from branch lengths of the generated maximum-likelihood tree for each gene. Our goal was to present the Sabethine transcriptome, a concise application of that transcriptome, and to emphasize concepts rather than present a discussion of the genome-wide details of the transcriptome.

We made four basic assumptions: First, during its dispersal northwards in North America, *W. smithii* has undergone analogous directional selection on its circadian clock as reflected in circadian-based behaviors in *Drosophila melanogaster, D. subobscura, and D. littoralis*. Second, directional selection and drift will erode genetic variation in clock genes as it has in other protein-coding loci in *W. smithii* [50]; consequently, genes under stronger selection will exhibit, on average, shorter branch lengths between this northern population of *W. smithii* and the other taxa. Third, the sequence reads from the *W. smithii* transcriptome represent random samples of their respective genes. This third assumption bears the caveat that, from incomplete cDNA contigs in the assembly, we cannot estimate evolutionary rates of individual genes, since different domains and even different codons within a domain, may evolve at different rates [31]. Since we are aligning *W. smithii* sequences of varying completeness to identify orthologs across disparate taxa, there is an inherent bias towards enriching for more conserved segments of the clock genes. Since we are concerned with the comparative evolutionary rates of functional groups of genes in taxa that are separated

by 100-400my, this temporal separation means that we have to use more conservative portions of the genes involved in order to obtain a clear signal of protein divergence. Nonetheless, if conservative segments are randomly distributed among clock genes, average divergence of TTFL or PTM genes provides a composite estimate of those two functional components of the *W. smithii* circadian clock. Fourth, we assume that TTFL and PTM genes identified in *Drosophila* serve analogous functions in the other insect taxa we consider. The number and function of circadian clock genes is better documented in *Drosophila*, which has set the historical landmarks for comparison with other insects and mammals [12, 14, 21, 22, 61–63] When looked for, the TTFL genes that are rhythmically expressed in *Drosophila* are also found to be rhythmically expressed in *Danaus* [18] and *Nasonia* [16] as well as mosquitoes [58–60, 64, 65] (including *tim* in *W. smithii* [66]). Functionally, RNAi targeted against *Cry2* [16–19], *tim* [64, 67–69], *per* [70–72], *Clk* [73, 74], *cyc* [75, 76] all disrupted circadian rhythmicity in non-*Drosophila* insects ranging from other Diptera to apterygote Thysanura. At least *Cry2* and TTFL orthologs of *tim*, *per*, *Clk*, and *cyc* in *Drosophila* are involved in circadian clock function across a variety of insects.

Methods

Collection, maintenance, and experimental treatment of *Wyeomyia smithii*

Wyeomyia smithii were collected in spring, 2010, as overwintering larvae from Maine (46 °N, 68 °W, 270 m elevation; population KC of earlier studies from this lab). Populations were maintained at the University of Oregon under standard rearing conditions and run through two generations to minimize maternal and field effects [53]. In the F2 laboratory generation, larvae were reared on short days (L:D = 8:16) at 23 °C to induce larval diapause in the third instar. After the initiation of diapause, a group of larvae continued on short days while another group was directly transferred to long days (L:D = 18:6) in order to initiate development, both at 23 °C.

RNA isolation and cDNA library construction, transcriptome sequencing and assembly

RNA was extracted from 12 samples of 30 individuals each. The 12 samples represented diapausing larvae on short days (L:D = 8:16), diapausing larvae exposed to 10 diapause-terminating long days (L:D = 18:6), pupae on long days and adults on long days. Each stage of development was sampled at three times of day (Table 1). All samples were prepared in 500uL TRIzol (Ambion Life Technologies, 5791 Van Allen Way, Carlsbad, California 92008) according to manufacturer's protocol. RNA was resuspended in 20uL DEPC-treated water and stored at -70 °C until shipment on dry ice to the Center for Genomics and Bioinformatics at Indiana University.

Table 1 Equimolar sources of cDNA to generate the *W. smithii* transcriptome

Sample	Stage	Daylength[a]	Hours after lights-on
1	larvae	Short	5
2	larvae	Short	13
3	larvae	Short	21
4	larvae	Long	5
5	larvae	Long	13
6	larvae	Long	21
7	pupae	Long	5
8	pupae	Long	13
9	pupae	Long	21
10	adult	Long	5
11	adult	Long	13
12	adult	Long	21

[a]Short: L:D = 8:16; long: L:D = 18:6

The overall quality of RNA samples was evaluated in terms of purity and integrity of RNA by means of a NanoDrop ND-1000 UV–VIS spectrometer (Thermo Fisher Scientific, 81 Wyman St, Waltham, MA 02451), Bioanalyzer (Agilent Technologies, 5301 Stevens Creek Blvd., Santa Clara, CA 95051) and agarose gel electrophoresis. RNA sample quality was verified regarding high RNA concentration, absorbance ratios A260/A280 in the range 2.0 - 2.2, and A260/A230 above 1.8. Samples with lower absorbance ratio were ethanol-precipitated in order to improve the quality. Equivalent amounts of RNA mass per test condition were pooled together, with a total of 10 μg RNA from all samples of *W. smithii*. Normalized 454-sequencing libraries were constructed from an equal-molar pool of RNA obtained from the unique exposure samples described above using the procedures optimized for Roche/454 Titanium sequencing modified from Meyer et al. [77]. After the final purification step, the library was stored at −20 °C until sequencing. This library was sequenced using one full-plate sequencing run in a 454 Roche GS FLX pyrosequencing instrument with Titanium chemistry (454 Life Sciences Corporation, 15 Commercial St., Branford, CT 06405), following manufacturer's protocol and methods previously described [78]. After 454 sequencing, the generated sequence reads were cleaned using ESTclean [79] and assembled using Newbler v.2.5.3 (454 Life Sciences Corporation, 15 Commercial St., Branford, CT 06405) in *de novo* mode and default parameters.

Transcriptome annotation

Transcriptome annotation was performed through the ISGA transcriptome analysis pipeline [80]. First, sequence

homology to known metazoan proteins was obtained by submitting contigs and singletons to BLASTx searches against NCBI's non-redundant database and dbEST [81]. Moreover, protein domains were identified among the six frame translations of the assembled sequences using Pfam, TIGRfam and HMMER3 searches [82]. Open reading frames were determined with the ORFpredictor software on the proteomics server of Youngstown State University [83]. Finally, orthology and paralogy was assigned by BLASTx against orthoMCL databases [84].

Defining orthologous groups and ortholog sequence acquisition

Flybase was used to identify each individual circadian gene in *Drosophila melanogaster* [85]. The *D. melanogaster* Flybase gene numbers and peptide sequences were used to identify the pre-computed orthologous genes for all insects using OrthoDB7 [86] and specifically extracting amino acid sequences for five comparative species: *Aedes aegypti, Culex pipiens, Anopheles gambiae, Danaus plexippus* and *Nasonia vitripennis* (Fig. 2). Geneious [87] was

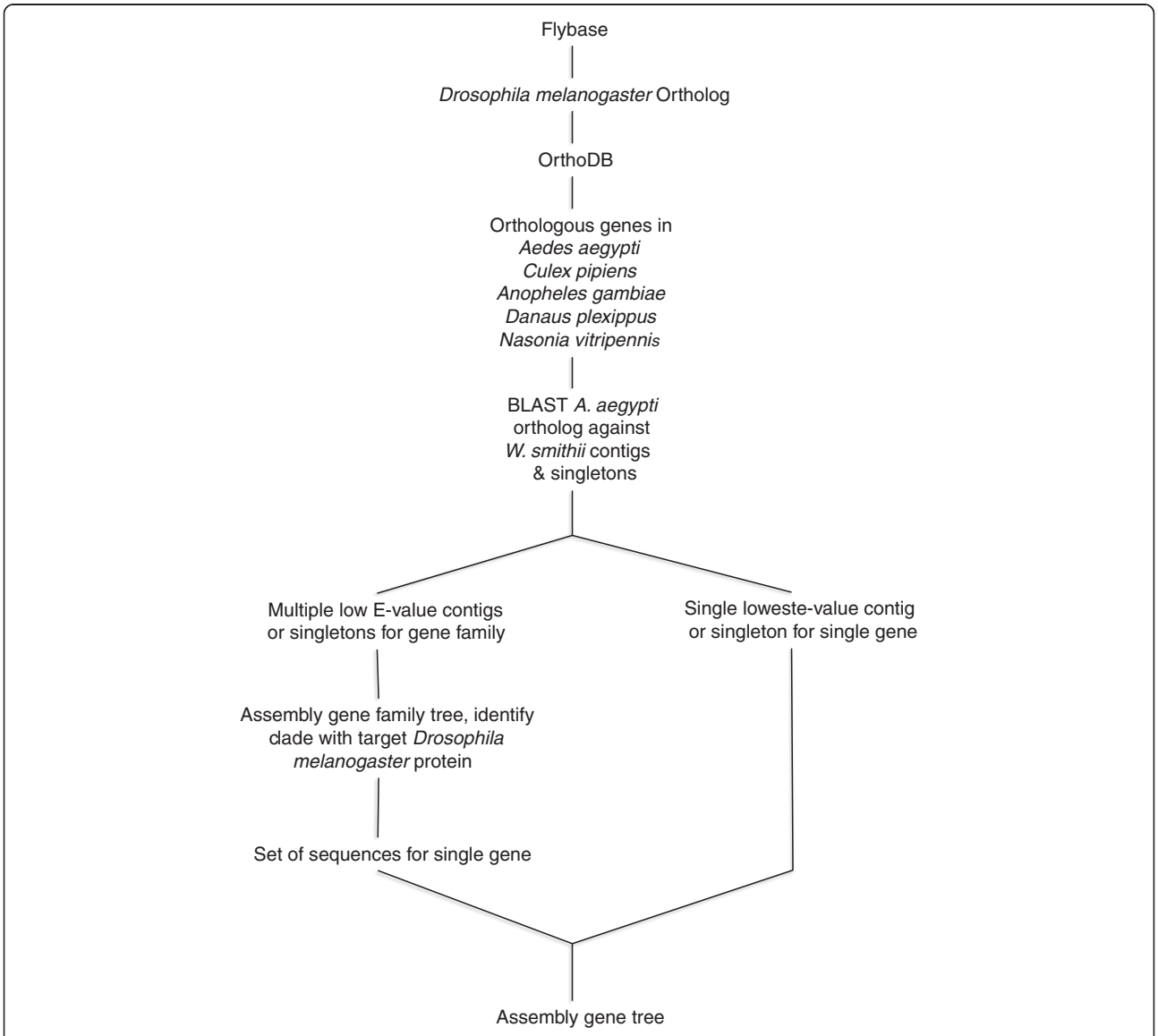

Fig. 2 Flow diagram of assigning contigs or singletons to specific circadian clock genes. The functional circadian clock gene was identified in *Drosophila melanogaster* through Flybase. The *Drosophila melanogaster* protein sequence was blasted against OrthoDB7 using the most recent common ancestor of all seven species as the search node. The orthologous genes were then taken from the resulting OrthoDB group, with the ortholog of *A. aegypti, W. smithii*'s most closely related species, and used in a local BLAST against the contigs and singletons from the *W. smithii* transcriptome. If the lowest E-value from that BLAST identified a single contig or singleton, that contig or singleton was assigned to the respective *D. melanogaster* gene function in the OrthoDB group. If the lowest E-value from the BLAST identified a multi-gene family, maximum likelihood trees were used to identify the orthologs of various genes in that family (Figs. 3 and 4)

used to perform a local BLAST (tblastx) of each *Aedes aegypti* ortholog against the entire *W. smithii* transcriptome. This procedure identified all possible homologous genes as contigs/singletons coding for *W. smithii* clock proteins, except for five groups of related genes that required additional analysis (Table 2): *Clk vs. cyc, tim vs. tim2* (*timeout*), *cry1 vs. cry2 vs. phr6-4*, and *dbt vs. Ck1α*. The contigs/singletons were then each evaluated as representing full gene transcripts, partial gene transcripts (including split genes), orthologs or paralogs, based on the alignments of their translated amino acids to those from the other six species, including their relative positions within the resulting phylogenetic gene trees (Figs. 3 and 4; Additional file 1 and Additional file 2).

Alignment processing and gene tree assembly

The orthologous groups of amino acid sequences were gathered into their respective gene families for each clock gene, including the translated *W. smithii* representative sequences. The 5' and 3' UTRs of each *W. smithii* amino acid sequence were removed based on start and stop codon positions. Each gene family was then aligned using MUSCLE [88] (Additional file 3) The protein alignments were then subjected to Gblock editing in order to identify conserved regions for phylogenetic analysis [89, 90] (Additional file 4 and Additional file 5). In order to be processed by ProtTest and Phyml, the alignments were converted into Phylip format. This conversion involved truncation of the identifiers for certain species' sequences. The identifiers were truncated in such a way to preserve the associated gene number, while changing the organism text identifier (Additional file 6). The best fit models of amino acid replacement for the Gblock edited alignments were determined using ProtTest [91, 92]. Maximum likelihood gene trees were then assembled using the phylogenetic software Phyml [93] and the best model of amino acid substitution according to the ProtTest results (Additional file 2).

Results

Transcriptome

Quality filtering of the reads was performed before assembly by applying default parameters using methods

Table 2 Circadian clock gene acronyms, names, and function

Abbreviation	Gene name	Clock function (See Fig. 1)	Reference
Ckla	*casein kinase Ia*	paralog of *dbt*	Fig. 4b
CkIIa	*casein kinase IIa*	promotes nuclear localization of PER-DBT-TIM via phosphorylation of TIM	[25–27]
Clk	*Clock*	transcription factor promoting transcription of *per, tim, cry2, vri, PDP1, cwo*	[25–27, 118]
Cul3	*Cullin 3*	ubiquitinates TIM leading to its degradation	[30]
cwo	*clockwork orange*	acts synergisticly with PER to inhibit CLK-mediated gene activation	[25, 119, 120]
cry	*cryptochrome 1, cry1, dcry*	photoreceptor; complexes with SGG & TIM to mediate light-input pathway into the clock	[25, 26]
cry2	*cryptochrome 2, mcry*	transcription regulator of *Clk* & *cyc* in insects other than Drosophila	[25, 27]
cyc	*cycle*	transcription factor promoting transcription of per, tim, cry2, *Pdp1*, vri	[25–27]
dbt	*doubltime = dco, discs overgrown*	major regulator of PER & CLK through phosphorylation	[25–27, 31]
jet	*jetlag*	promotes light-induced proteosomal degradation of TIM & CRY1	[27, 33, 121]
kaya	*kayaka*	Inhibits VRI suppression of *Clk* promoter; represses CLK activity	[127]
nmo	*nemo*	phosphorylates PER, enhances action of DBT; phosphorylates CLK	[25, 29, 31, 123]
Pdp1	*PAR-domain protein 1*	Likely in combination with VRI enhances *Clk* transcription and clk mRNA amplitude	[25–27, 118]
per	*period*	negative transcription regulator of *Clk* & *cyc* transcription after transport to nucleus as phosphorylated PER-TIM-DBT complex	[25, 27]
phr6-4	*(6–4)-photolyase*	paralog of *cry1* & *cry2*	[17], Fig. 3
PP2A-B'	*protein phosphatase 2A, regulatory B subunit*	regulates phosphorylation of PER; counter-balances PER & TIM phosphorylation by DBT	[25–27, 31]
sgg	*shaggy*	phosphorylates TIM, in concert with CRY1 & light; regulates PER phosphorylation-dephosphorylation; promotes PER nuclear localization	[25–27, 31]
slimb	*supernumerary limbs*	phosphorylates PER & TIM, leading to their degradation	[25–27, 124]
tim	*timeless*	binds to & facilitates transport of heterodimeric PER into the nucleus; interacts with JET, SGG & CRY to regulate the input of light; increases CKIIα phosphorylation of PER	[25, 125]
tim2	*timeout*	paralog of TIM; necessary for laval survivorshiop; promotes chromosomal stability; enhancs photoreception.	[122, 126]
vri	*vrille*	transcription inhibitor of *Clk*	[25–27, 118]

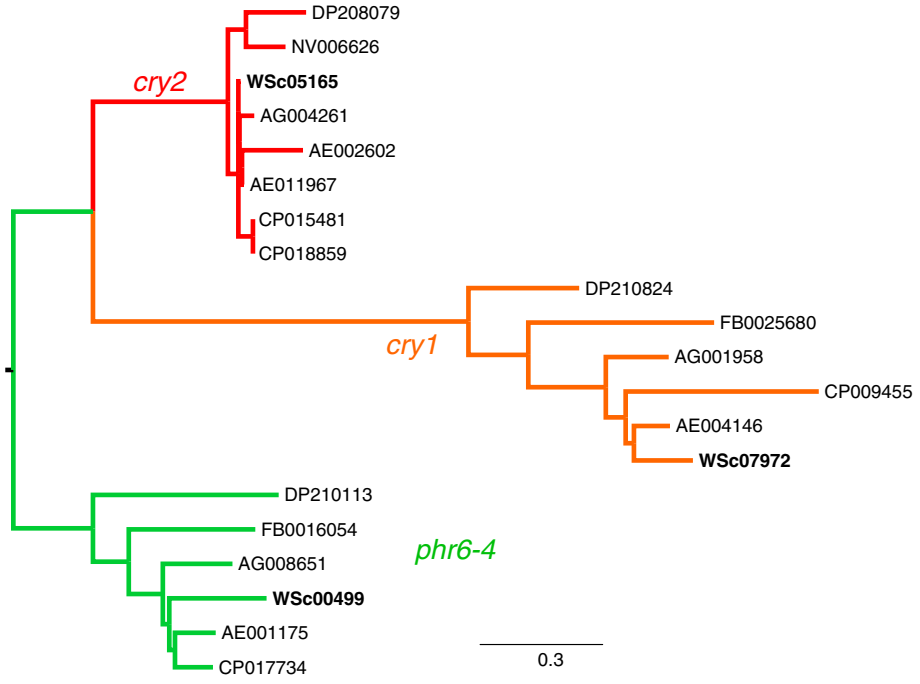

Fig. 3 Assigning *W. smithii* orthologs to *cry1*, *cry2*, and *phr6-4* (*6–4 photolyase*). The maximum likelihood tree identified a single *W. smithii* contig (bold) within each of the three monophyletic clades in the tree. Gene number abbreviations: AE, *Aedes aegypti*; AG, *Anopheles gambiae*; CP, *Culex pipiens*; DP, *Danaus plexippus*; FB, *Drosophila melanogaster*; NV, *Nasonia vitripennis*; WSc, *W. smithii* contigs

described by Vera et al. [94]. The assembly was performed on 1,081,284 quality-controlled reads summing up 94 % of raw sequence data (432,542,060 bases), after trimming of adaptor sequences (Table 3). Newbler aligned 92 % and assembled 87 % of quality-controlled reads, resulting in 25,904 contigs with lengths >50 bases, 14,459 contigs with lengths >500 bases, and 54,418 singletons (Table 3). The N50 for contig length >500 bases was 1373 bp. The Newbler assembler considers alternative splicing that resulted in the integration of contigs into 21,233 isotigs representing candidate transcripts. The N50 for isotigs was 1953 bp and the average size for isotigs was 1515 bp long (Table 3).

Assembly quality was tested by retrieving BLASTx hits against the *Drosophila* orthologs in the CEGMA core eukaryotic genes dataset [95] (Additional file 7). The contigs alone represent 493 of 523 genes known to exist as single copies, indicating that the transcriptome is >94 % complete.

From the total number of contigs and singletons, a significant BLASTx match was obtained for 13,470 (52 %) and 15,048 (28 %) of transcripts respectively (Additional file 8). This result implies that between 48 % (contigs) and 72 % (singletons) of the sequences do not show homology to any other sequence present in the investigated databases. However, among the 47,837 orphan transcripts, 33,183 have identifiable functional protein domains plus an additional 7486 have detectable open reading frames, indicating that they represent protein-coding genes as well as non-coding transcripts.

Gene Ontology (GO) terms were assigned to 10 % of singletons and 42 % of contigs; overall 11,342 sequences were mapped. Finally, orthoMCL [96] and OrthoDB [97] analyses of gene orthology revealed that 12,653 contigs and 11,787 singletons show orthology to one or more organisms in the two gene-orthology databases (Additional file 8).

Defining orthologous groups

The *Wyeomyia smithii* transcriptome included all 17 circadian clock genes we sought to identify (Fig. 1). The clock genes were represented by 15 contigs and two singletons, ranging from 450 to 3000 nucleotides (Table 4). As expected, *cry2* is absent in *D. melanogaster* and both *tim* and *cry1* are absent in *Nasonia vitripennis* [17, 25, 27, 98, 99]. The local BLAST (tblastx) of each *Aedes aegypti* ortholog against the *W. smithii* transcriptome identified a single best contig or singleton to represent 11 of the clock genes. Six other clock genes belonged to broader gene families and required additional analysis.

Clk vs. *cyc*: *Drosophila melanogaster Clk* and *cyc* are represented in two different EOG7 orthologous groups. Local BLASTS of the *A. aegypti* orthologs of *Clk* (AAEL012562) and *cyc* (AAEL002049) against the *W. smithii* transcriptome identified a *W. smithii* singleton

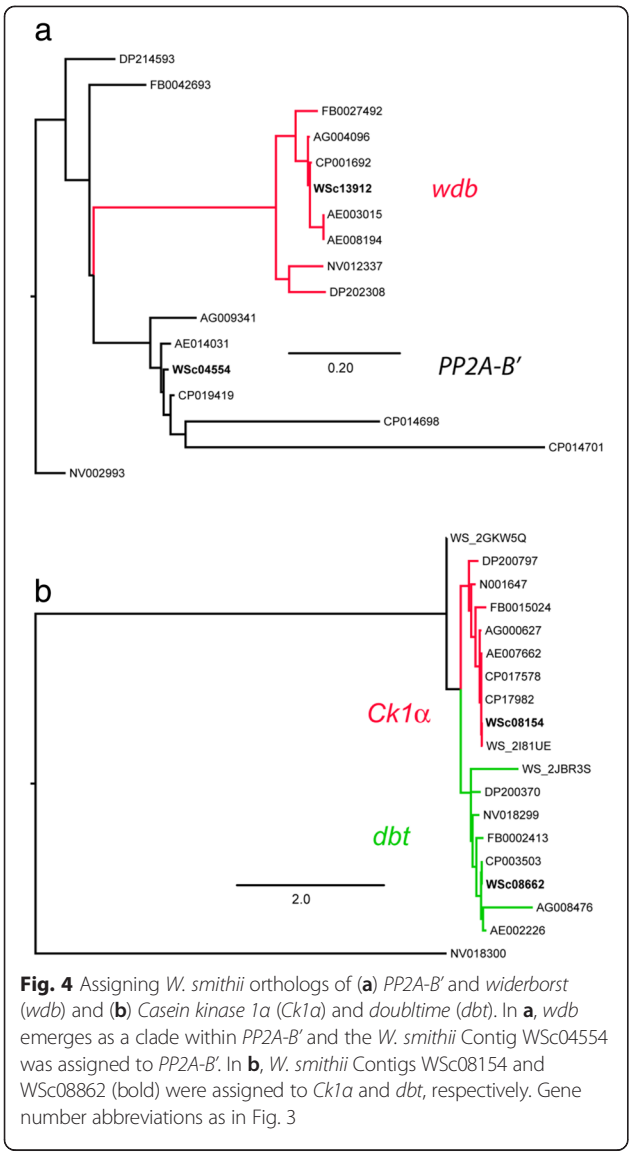

Fig. 4 Assigning *W. smithii* orthologs of (**a**) *PP2A-B'* and *widerborst* (*wdb*) and (**b**) *Casein kinase 1α* (*Ck1α*) and *doubltime* (*dbt*). In **a**, *wdb* emerges as a clade within *PP2A-B'* and the *W. smithii* Contig WSc04554 was assigned to *PP2A-B'*. In **b**, *W. smithii* Contigs WSc08154 and WSc08862 (bold) were assigned to *Ck1α* and *dbt*, respectively. Gene number abbreviations as in Fig. 3

(2GK8YT) and a *W. smithii* contig (Contig 17314), that best represented *clk* and cyc, respectively (Table 4).

tim vs. *tim2* (*timeout*): *Drosophila melanogaster Clk* and *cyc* are represented in two different EOG7 orthologous groups. Local BLASTS of the *A. aegypti* orthologs of *tim* (AAEL006411) and *tim2* (AAEL009518) against the *W. smithii* transcriptome identified two distinct *W. smithii* contigs that distinguished *tim* (Contig 06527) from *tim2* (Contig 18589), respectively (Tables 4 and 5).

cry1 vs. *cry2* vs. *phr6-4*: Protein sequences belonging to the cryptochrome family were identified and combined from the *cry1*, *cry2*, and *phr6-4* orthologous groups, EOG79SRM2 and EOG7P64PH. Local BLASTS of the *A. aegypti* sequences from the two OrthoDB groups against the *W. smithii* transcriptome identified three *W. smithii* candidate contigs. After trimming the 5' and 3' UTRs, the contigs were Gblock edited, and, in combination with the other six species, tested for the appropriate amino acid substitution model (see Methods). A maximum likelihood tree rooted with *phr6-4* [17, 98, 100] separated sequences into three distinct clades representing *cry1*, *cry2*, and *phr6-4* (Fig. 3). The three *W. smithii* candidate contigs were each placed in a separate clade. We therefore concluded that Contig 07972 is the ortholog of *cry1*, Contig 05165 is the ortholog of *cry2*, and Contig 00499 is the ortholog of *phr6-4*.

dbt vs.*Ck1α*: The OrthoDB group for *D. melanogaster doubletime* (EOG72CGPS) contained two gene families, *doubletime* (*discs overgrown*) and *Casein kinase 1α*. A gene tree was then assembled using the same protocols described for the cryptochromes, above. When rooted with the *N. vitripennis* sequence NV018300, the remaining orthologs separated into two distinct clades one including *D. melanogaster dbt* and *W. smithii* Contig 08662, the other one including *D. melanogaster Ck1α* and *W. smithii* Contig 08154 sequence (Fig. 4a). We therefore concluded that Contig 08662 is the ortholog of *dbt* and Contig 08154 is the ortholog of *Ck1α*.

PP2A-B' vs. *wdb*: The OrthoDB group for the *D. melanogaster* ortholog of *PP2A-B'* (EOG7S57VZ) contained two gene families, *PP2A-B'* and *widerborst*. To

Table 3 Sequencing results and assembly statistics

Sequencing results	Total	Aligned	Assembled
Number of reads[a]	1,081,284	999,150	936,525
Number of bases	432,542,060	401,519,734	380,523,665
Assembly statistics	Singletons	Number of Contigs (Contigs >500 bp)	Isotigs
Number of sequences	54,418	25,904 (14,459)	21,233
Number of bases	20,109,580	19,120,348 (16,952,075)	32,183,643
Average sequence size (bp)	355	738 (1,172)	1,515
Length of N50 sequence (bp)[b]	451	1,249 (1,373)	1,373

[a]After quality filtering steps and removal of outliers such as adaptor sequences and repeats
[b]N50 is a weighted median statistic, such that 50% of all bases are contained in sequences ≥N50 length

Table 4 Circadian clock genes, their role in the clock, properties of their *Wyeomyia smithii* transcripts and their relationship to homologs in *Drosophila melangaster* and *Aedes aegypti*

Gene[a]	Transcript[b]	*A. aegypti* AAEL- [c]	E-value[d]	ProtTest Model	FBgn-	Ortho DB EOG7-[e]	Nucleo- tides	Relative Rate
Cklla	15735	012094	0	Dayhoff+G+F	0264492	4RCF1	451	0.102
Clk	***2GK8YT***	**012562**	**0**	**JTT+G+F**	**0023076**	**NSNR1**	**1,865**	**0.679**
cry2	**05165**	**002602**	**4E-173**	**LG+G**	**—**	**P64PH**	**634**	**0.298**
Cul3	07535	006291 007187	0	LG+G	0261268	TN9FJ	1,771	0.348
cwo	**15437**	**010513**	**0**	**LG+G+F**	**0259938**	**XDP58**	**2,313**	**1.792**
cyc	**17314**	**002049**	**7 E-86**	**JTT+G**	**0023094**	**7TBTD**	**851**	**0.390**
dbt	08662	002226	1E-144	JTT+G	0002413	2CGPS	527	0.866
jet	03197	012126	1 E-93	LG+G+F	0031652	NSNRX	1,004	1.572
kaya	**13013**	**008953**	**8 E-145**	**LG+G+F**	**0001297**	**06BFN**	**1,162**	**1.625**
nmo	17870	004797	0	JTT+G	0011817	JF2MQ	2,130	0.254
Pdp1	**02480**	**005255**	**8 E-145**	**LG+I**	**0016694**	**2RZNP**	**3,083**	**0.602**
per	***1A331S***	**008141**	**1 E-76**	**JTT+G**	**0003068**	**MHBMZ**	**2,670**	**1.500**
PP2A-B'	04554	014031	0	LG+G	0042693	S57VZ	721	0.463
sgg	17080	005238	0	Dayhoff + G	0003371	D2S3R	2641	0.455
slimb	03120	003371	0	JTT+G	0023423	WHV0D	2,355	0.559
tim	**06527**	**006411**	**1 E-61**	**JTT+G+F**	**0014396**	**DG81S**	**2,752**	**0.793**
vri	**08372**	**011371**	**7 E-133**	**JTT+I+G+F**	**0173452**	**HXQF6**	**2,465**	**1.234**

[a]Bold font, TTFL; regular font, PTM
[b]Contigs in Roman; singletons in italics, omitting the prefix F5BTJ3O0-
[c]*Aedes aegypti* BLAST to *W. smithii* transcriptome
[d]E-values from reciprocal BLAST of contig or singleton to *A. aegypti*
[e]Rooted at most-recent common ancestor of Hymenoptera & Diptera

distinguish these genes in *W. smithii*, the three *A. aegypti* sequences in the same OrthoDB group, were aligned locally against the *W. smithii* transcriptome using BLAST. A gene tree was assembled using the same protocols described for *cry2* showing that the *widerborst* gene family occupied its own monophyletic clade within the *PP2A-B'* gene tree (Fig. 4b). The *wdb* clade included *W. smithii* Contig 13912. A separate branch included *D. melanogaster PP2A-B'* and *W. smithii* Contig 04554. We therefore concluded that Contig 13912 is the ortholog of *wdb*, and Contig 04554 the ortholog of *PP2A-B'*.

Evolutionary divergence

Divergence of *W. smithii* genes involved in the circadian clock was determined from relative, cumulative branch lengths from other taxa (Table 4) using maximum likelihood for phylogenetic inference. Among the 17 circadian genes (Fig. 1) ProtTest returned six different best-fit models for amino acid substitution for nine TTFL genes and five different models for eight PTM genes (Table 4). The frequency of different models did not differ between the two categories of genes (two-sided Fisher's exact test $P = 1.000$).

A distance matrix (Additional file 1) was generated for each maximum likelihood gene tree, showing the relative branch lengths between each organism for each particular protein. In order to measure rates of evolution for *W. smithii's* clock proteins relative to the other organisms in each gene family, relative rate for each clock protein was calculated from the distance matrices:

Table 5 Genes closely related to clock genes or in the same gene family[a]

Related gene	*W. smithii* Contig	*A. aegypti* AAEL-	E-value	Clock gene	FBgn	Ortho DB EOG7-	Nucleotides
Ckla	08154	007662	0	*dco*	0015024	2CGPS	483
cry1	07972	004146	0	*cry2, phr6-4*	0025680	9SRM2	1,228
phr6-4	00499	001175	0	*cry1, cry2*	0016054	P64PH	1,134
tgo	06945	010343	0	*cyc*	0264075	Q2ZVC	940
tim2	18589	009518	7 E-45	*tim*	0038118	2RZMT	589
wbt	13912	003015 008194	1 E-165	*PP2A-B'*	0027492	S57VZ	4,552

[a]Column headings as in Table 4

Relative rate = (Average branch length for *W. smithii* across all taxa for an individual gene) ÷ (average branch length for all seven taxa across all genes). For a given protein, when this ratio is greater than 1.0, it indicates that the protein is evolving faster in *W. smithii* relative to other organisms; when this ratio less than 1.0, it indicates that the protein is evolving more slowly in *W. smithii* relative to other organisms.

Relative divergence of *W. smithii* TTFL genes did not differ from 1.0 but divergence of PTM genes was significantly less than 1.0 (Table 6). However, relative divergence of TTFL and PTM genes did not differ significantly from each other (Fig. 5a). There was a marginally non-significant negative correlation between relative rate of combined gene divergence and number of nucleotides in their respective contigs or singletons (Fig. 5b). To account for the possibility of Type II error, ANCOVA of core vs. PTM genes with number of nucleotides as the covariate revealed no significant treatment effect ($t = 0.621$, $P = 0.545$) and ANOVA of the residuals from regression of divergence on number of nucleotides also revealed no significant difference between core and PTM genes (Fig. 5c).

Discussion

Using the "black sheep" counting technique of universal, single copy genes to determine the completeness of the *Wyeomyia smithii* transcriptome, we estimated that the *W. smithii* transcriptome encompasses >94 % of its transcribed genome. This result is not surprising since we used as the basis for the transcriptome both developing and diapausing larvae, pupae, and adults sampled at different times of the day and night and under long and short days (Table 1). In fact, the *W. smithii* transcriptome includes all of the 17 circadian clock genes we sought to identify (Fig. 1); its contigs and singletons translate into peptides of sufficient length to estimate comparative rates of evolution of both TTFL and PTM genes between the *W. smithii* and the six comparison taxa.

Even if historical directional selection on the circadian clock has occurred among populations dispersing along a latitudinal gradient, stabilizing selection at any locality along that gradient is still important in maintaining daily time-keeping in concert with a 24-h world. Concordance between the circadian clock and the external 24-h world is an important component of fitness in organisms from prokaryotes to mammals [6, 7, 10], including *W. smithii* [101]. The motivation for our study was to compare the relative rates of evolutionary divergence of TTFL and PTM genes between a northern population of *W. smithii* that has experienced a continual northward dispersal into temperate regions of progressively longer summer day lengths, with both closely related mosquitoes and more distantly related insects, including *Drosophila melanogaster*, *Danaus plexippus*, and *Nasonia vitripennis* (Fig. 6). Overall, we found that *W. smithii* clock genes are not evolving faster than expected from other insects (Table 6) and the rate of evolution of TTFL genes does not differ from PTM genes (Fig. 5).

Table 6 Relative divergence of W. smithii TTFL and PTM clock genes from other insects

Category[a]	Mean	StErr	*t*	*df*	*P = 1.0*[b]
TTFL & PTM combined	0.8	0.13	1.55	16	0.14
TTFL only	0.99	0.2	0.49	8	0.96
PTM only	0.58	0.16	2.6	7	0.04

[a]TTFL, transcription-translation feedback loops; PTM: post-translational modifiers
[b]Probability Mean differs from 1.0

The best models for amino acid substitution do not differ between TTFL and PTM genes, although six different models provided the best fit within TTFL genes and five different models within PTM genes, (Table 4). Clearly, neither the TTFL nor the PTM proteins represent a uniform group in terms of their evolution. Consequently, no single substitution model would be appropriate for phylogenetic inference of circadian clock genes within either functional group or within the two groups combined. There is, however, greater retention of PTM than TTFL genes among the six insect taxa we considered. *cry2* is dispensable in *Drosophila* (although present in lower Diptera) and *tim* and *cry1* are dispensable in Hymenoptera [17, 25, 27, 98, 99]. By contrast, all eight of the PTM genes are conserved in all six taxa. This observation indicates that natural selection within and between orders of insects has acted to conserve PTM genes more than TTFL genes.

What importance then are the TTFL genes? To be a functional time-keeper of overt circadian expression, the circadian clockworks cannot work in isolation but must communicate circadian time to downstream clock-controlled genes. "Much is known about how information is relayed to the *Drosophila* [*melanogaster*] clock and how the central clock itself functions, but less is understood about how information from the clock is relayed to the rest of the organism" ([102], p. 352) [103, 104]. Since all of the TTFL genes are transcription factors or transcription regulators of gene expression, it is not surprising that the TTFL genes likely provide this communication to clock-controlled behavioral and physiological processes [37, 104–114]. The TTFL genes provide a cyclical expression of genes and a pleiotropic, time-specific signal to the rest of the organism; the PTMs maintain this cycle with a period of about 24 h. It is the genetic co-adaptation, i.e., the co-evolution within and between these functional groups that enables different organisms to maintain biochemical, physiological, and behavioral activities in concert with the external daily environment.

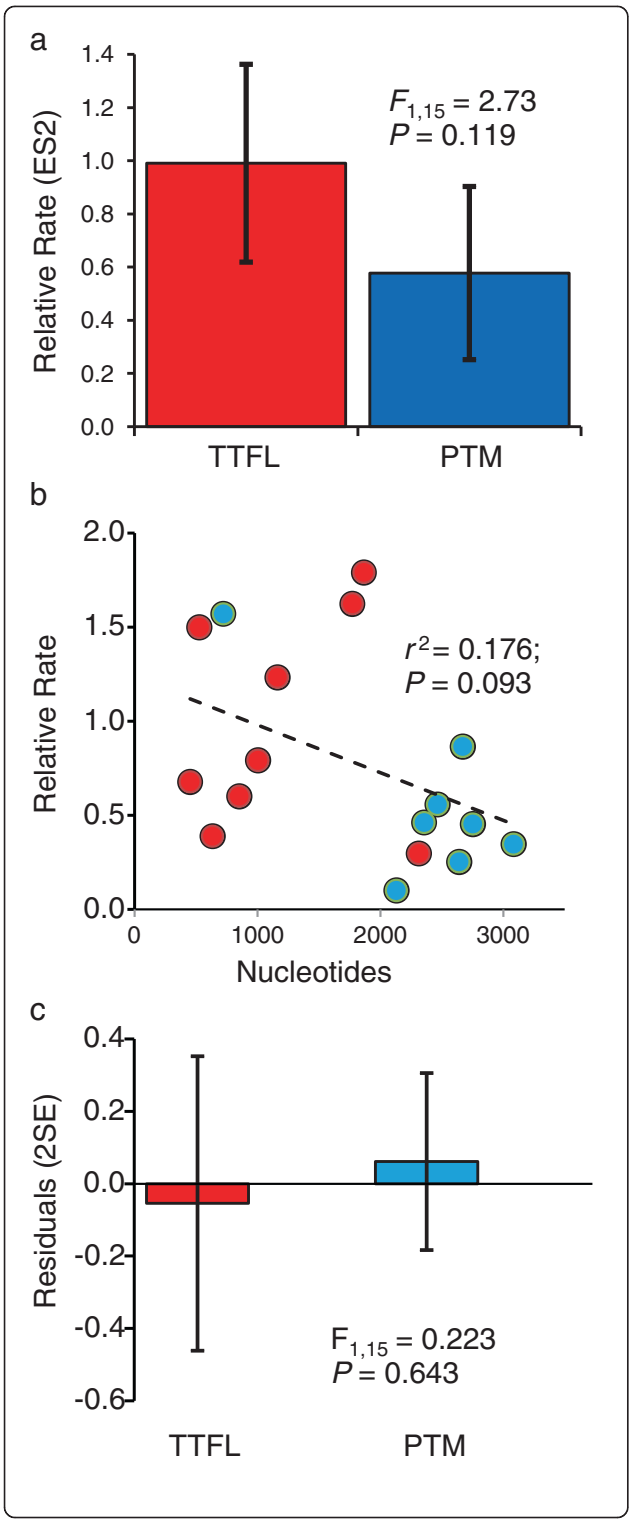

Fig. 5 Rates of amino acid divergence in circadian clock genes of *Wyeomyia smithii* relative to other insects (Table 4). **a** Relative rates (±2SE) of divergence in the core transcription-translation feedback loop (TTFL) and of post-transcriptional modifiers (PTM). **b** relationship between relative rates of amino acid divergence and the number of nucleotides in the contigs or singletons upon which the rates were based. TTFL (red) and the PTM (blue). **c** deviations from regression (residuals) in **5b**. The residuals essentially factor out any differences in relative rates due to the number of nucleotides upon which amino acid divergence was based

Conclusion

We report the first genome or transcriptome of any member of the mosquito tribe Sabethini (subfamily Culicinae). This transcriptome serves as a point of departure for annotating a future scaffolding genome of *W. smithii*. As an application of the transcriptome, we compared rates of

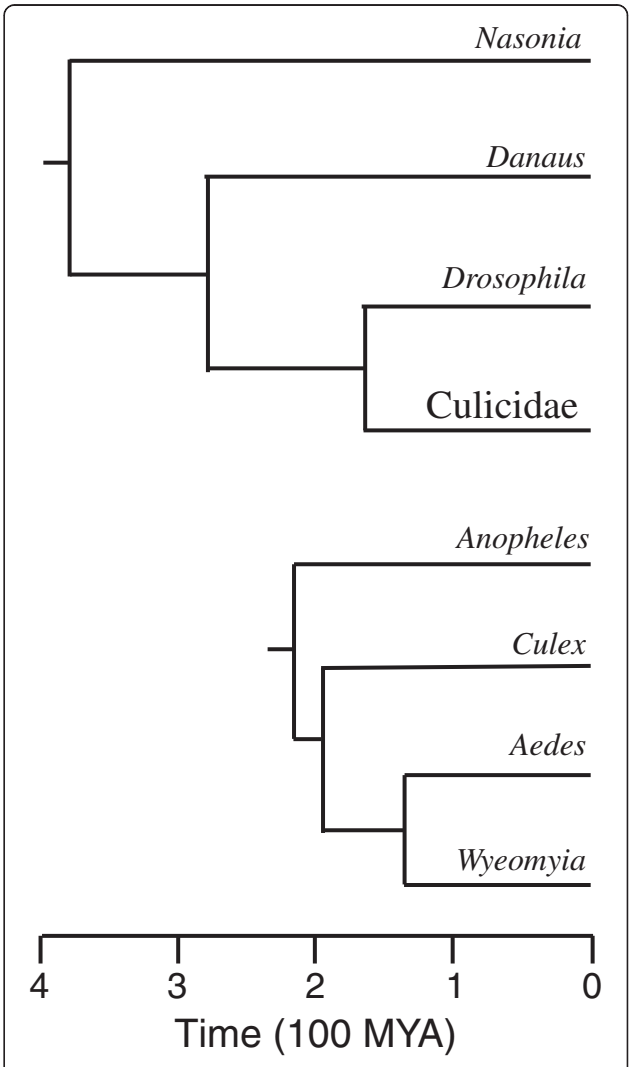

Fig. 6 Phylogenetic relationships of insects used in this study. The nodes indicate approximate time since the most recent common ancestor of a given branch. Orders and families (top) based on [128]; genera within the family Culicidae (bottom) based on [129]

evolutionary divergence of *W. smithii* circadian clock genes from six other insect taxa. We found no significant difference in rates of evolutionary divergence between genes involved in the central transcription-translation feedback loop and genes involved in post-translational modifiers. All of the species we considered exhibit circadian rhythmicity under constant conditions and include all the PTM genes in Fig. 1. By contrast, the representation of TTFL genes varies among taxa, including suborders of Diptera. This contrast means that there has to be genetic coadaptation both within the TTFLs to maintain a rhythmic circadian output and between the TTFL and their PTMs to maintain that rhythmic output with a period of about 24 h in concert with the 24-h variation in the external environment.

Additional files

Additional file 1: Distance matrices (XLSX 20 kb)

Additional file 2: Maximum likelihood trees in Newick format; visualize with FigTree [116]. (DOC 28 kb)

Additional file 3: Muscle protein alignments; visualize with MEGA [117]. (ZIP 46 kb)

Additional file 4: Gblock edited alignments, FA-GB files in .txt format. (ZIP 20 kb)

Additional file 5: Gblock edited alignments in .pdf format. (ZIP 1181 kb)

Additional file 6: Gblock edited alignments, truncated for PhyML, .phy files in .txt format. (ZIP 24 kb)

Additional file 7: Blastx of *W. smithii* contigs against *D. melanogaster* orthologs in the CEGMA database. (XLSX 116 kb)

Additional file 8: Contig and singleton annotations. (XLSX 8403 kb)

Abbreviations
PTM: Post-translational modifiers; TTFL: Transcription-translation feedback loops.

Competing interests
The authors declare that they have no competing interests.

Authors' contribution
WB, CH and JKC conceived and designed the study; KM and JL ran the experiments; DT, JKC, J-HC, WB, and JB analyzed the data; CH, DT, WB and JKC wrote the manuscript. All authors read and approved the final manuscript.

Authors' information
Not applicable.

Availability of supporting data
Both the EST clean [79] and the "Transcriptome sequence, assembly and annotation of *Wyeomyia smithii*" are available through NCBI [115].

Acknowledgements
Allison Burns, Taylor Pearce, and Jack Benda aided in the gene alignments used in this manuscript.

Funding
This research was financially supported by NSF Grants IOS-083998 and IOS1255628.

Author details
[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA. [2]Stowers Institute for Medical Research, Kansas City, MO, USA. [3]Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN, USA. [4]School of Biosciences, University of Birmingham, Birmingham, UK. [5]Pervasive Technology Institute, Indiana University, Bloomington, IN, USA. [6]GRU Cancer Center, Georgia Regents University, Augusta, GA, USA. [7]Department of Biological Sciences, Notre Dame University, Notre Dame, IN, USA. [8]Burke E. Porter Machinery, Grand Rapids, MI, USA.

Received: 27 May 2015 Accepted: 19 September 2015
Published online: 06 October 2015

References
1. Hopkins AD. Bioclimatics: a science of life and climate relations. USDA Miscellaneous Publication No. 280. Washington (DC): United States Government Printing Office; 1938.
2. MacArthur RH. Geographical ecology. Princeton (NJ): Princeton University Press; 1984.
3. Hoffmann AA, Parsons PA. Extreme environmental change and evolution. Cambridge (UK): Cambridge University Press; 1997.
4. Angilletta Jr MJ. Thermal adaptation: a theoretical and empirical synthesis. Oxford (UK): Oxford University Press; 2009.
5. Gilbert SF, Epel D. Ecological developmental biology: integrating epigenetics, medicine, and evolution. Sunderland (MA): Sinauer Associates; 2009.
6. Pittendrigh CS. Circadian rhythms and the circadian organization of living systems. Cold Spring Harbor Symp Quant Biol. 1960;25:159–84.
7. Bünning E. The physiological clock. New York: Springer; 1973.
8. Majercak J, Sidote D, Hardin PE, Edery I. How a circadian clock adapts to seasonal decreases in temperature and day length. Neuron. 1999;24:219–30.
9. Vanin S, Bhutani S, Montelli S, Menegazzi P, Green EW, Pegoraro M, et al. Unexpected features of *Drosophila* circadian behavioral rhythms under natural conditions. Nature. 2012;484:371–6.
10. Yerushalmi S, Green RM. Evidence for the adaptive significance of circadian rhythms. Ecol Let. 2009;12:970–81.
11. Pittendrigh CS. Temporal organization: reflections of a Darwinian clock-watcher. Annu Rev Physiol. 1993;55:17–54.
12. Golombek DA, Rosenstein RE. Physiology of circadian entrainment. Physiol Rev. 2010;90:1063–102.
13. Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, et al. Circadian rhythms from multiple oscillators: Lessons from diverse organisms. Nat Rev Genet. 2005;6:544–56.
14. Brown SA, Kowalska E, Dallmann R. (Re)inventing the circadian feedback loop. Dev Cell. 2012;22:477–87.
15. Hastings JW. Biochemical aspects of rhythms: phase shifting by chemicals. Cold Spring Harbor Symp Quant Biol. 1960;25:131–43.
16. Bertossa RC, van de Zande L, Beukeboom LW, Beersma DGM. Phylogeny and oscillating expression of period and cryptochrome in short and long photoperiods suggest a conserved function in *Nasonia vitripennis*. Chronobiol Intl. 2014;31:749–60.
17. Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect clocks. Mol Biol Evol. 2007;24:948–55.
18. 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:e4.
19. Ikeno T, Katagiri C, Numata H, Goto SG. Causal involvement of mammalian-type cryptochrome in the circadian cuticle deposition rhythm in the bean bug *Riptortus pedestris*. Insect Mol Biol. 2011;20:409–15.
20. Okamura H, Myake S, Sumi Y, Yamaguchi S, Yasui A, Miujtjens M, et al. Photic induction of *mPer1* and *mPer2* in *Cry*-deficient mice lacking a biological clock. Science. 1999;286:2531–34.
21. Gallego M, Virshup DM. Post-translational modifications regulate the ticking of the circadian clock. Nat Rev Mol Cell Biol. 2007;8:139–48.
22. Mehra A, Baker CL, Loros JJ, Dunlap JC. Post-translational modifications in circadian rhythms. Trends Biochem Sci. 2009;34:483–90.
23. Mahesh G, Jeon E-H, Ng FS, Liu Y, Gunawardhana K, Houl JH, et al. Phosphorylation of the transcription activator CLOCK regulates progression through a ~24-h feedback loop to influence the circadian period in *Drosophila*. J Biol Chem. 2014;289:19681–93.

24. Larrondo LF, Olivares-Yañez C, Baker CL, Loros JJ, Dunlap JC. Decoupling circadian clock protein turnover from circadian period determination. Science. 2015;347:518. doi:10.1126/Science1257277.
25. Hardin PE. Molecular genetic analysis of circadian timekeeping in *Drosophila*. Adv Genetics. 2011;74:141–73.
26. Allada R, Chung BY. Circadian organization of behavior and physiology in *Drosophila*. Annu Rev Physiol. 2010;72:605–24.
27. Zhan S, Merlin C, Boore JL, Reppert SM. The monarch butterfly genome yields insights into long-distance migration. Cell. 2011;147:1171–85.
28. Lee H-M, Chen R, Kim H, Etchegaray J-P, Weaver DR, Lee C. The period of the circadian oscillator is primarily determined by the balance between casein kinase 1 and protein phosphatase 1. Proc Natl Acad Sci U S A. 2011;108:16451–6.
29. Yu W, Houl JH, Hardin PE. NEMO kinase contributes to core period determination by slowing the pace of the *Drosophila* circadian oscillator. Curr Biol. 2011;21:756–61.
30. Grima B, Dognon A, Lamouroux A, Chélot E, Rouyer F. CULLIN-3 controls TIMELESS oscillations in the *Drosophila* circadian clock. PLoS Biol. 2012;10:e1001367.
31. Garbe DS, Fang Y, Zheng X, Sowcik M, Anjum R, Gygi SP, et al. Cooperative interaction between phosphorylation sites on PERIOD maintains circadian period in *Drosophila*. PLoS Genet. 2013;9:e1003749.
32. Muratani M, Tansey WP. How the ubiquitin-proteosome system controls transcription. Nat Rev Mol Cell Biol. 2003;4:192–201.
33. Koh K, Zheng X, Sehgal A. JETLAG resets the *Drosophila* circadian clock by promoting light-induced degradation of TIMELESS. Science. 2006;312:1809–12.
34. Fu Y-H. Oscillating *Per*-cision. PLoS Biol. 2008;6:e192.
35. Syed S, Saez L, Young MW. Kinetics of Doubletime kinase-dependent degradation of the *Drosophila* period protein. J Biol Chem. 2011;286:27654–62.
36. Robles MS, Cox J, Mann M. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. PLoS Genet. 2014;10:1004047.
37. Taghert PH, Shafer OT. Mechanisms of clock output in the *Drosophila* circadian pacemaker system. J Biol Rhythms. 2006;21:445–57.
38. Kadener S, Menet JS, Sugino K, Horwich MD, Weissbein U, Nawathean P, et al. A role for microRNAs in the *Drosophila* circadian clock. Genes Devel. 2009;23:2179–91.
39. Yang M, Lee J-E, Padgett RW, Edery I. Circadian regulation of a limited set of conserved microRNAs in *Drosophila*. BMC Genomics. 2008;9:83.
40. Lim C, Allada R. Emerging roles for post-transcriptional regulation in circadian clocks. Nat Neurosci. 2013;16:1544–50.
41. Kojima S, Shingle DL, Green CB. Post-transcriptional control of circadian rhythms. J Cell Sci. 2011;124:311–20.
42. Lim C, Allada R. ATAXIN-2 activates PERIOD translation to sustain circadian rhythms in *Drosophila*. Science. 2013;340:875–79.
43. Chen W, Liu Z, Li T, Zhang R, Xue Y, Zhong Y, et al. Regulation of *Drosophila* circadian rhythms by miRNA let-7 is mediated by a regulatory cycle. Nat Commun. 2015;5:5549.
44. Kim EY, Jeong EH, Park S, Jeong H-J, Edery I, Cho JW. A role of *O*-GcNAcylation in setting circadian clock speed. Genes Dev. 2012;26:490–502.
45. Ripperger JA, Merrow M. Perfect timing: Epigenetic regulation of the circadian clock. FEBS Lett. 2011;585:1406–11.
46. Adewoye AB, Kyriacou CP, Tauber E. Identification and functional analysis of early gene expression induced by circadian light-resetting in *Drosophila*. BMC Genomics. 2015;16:570.
47. Harbach RE. Mosquito Taxonomic Inventory. 2013. http://mosquito-taxonomic-inventory.info/. Accessed 25 Feb 2015.
48. Ross HH. The colonization of temperate North America by mosquitoes and man. Mosq News. 1964;24:103–18.
49. Bradshaw WE, Lounibos LP. Evolution of dormancy and its photoperiodic control in pitcher-plant mosquitoes. Evolution. 1977;31:546–67.
50. Armbruster P, Bradshaw WE, Holzapfel CM. Effects of post-glacial range expansion on allozyme and quantitative genetic variation of the pitcher-plant mosquito, *Wyeomyia smithii*. Evolution. 1998;52:1697–704.
51. Merz C, Catchen JM, Hanson-Smith V, Emerson KJ, Bradshaw WE, Holzapfel CM. Replicate phylogenies and post-glacial range expansion of the pitcher-plant mosquito, *Wyeomyia smithii*, in North America. PLoS One. 2013;8:e72262.
52. Hut RA, Paolucci S, Dor R, Kyriacou CP, Daan S. Latitudinal clines: an evolutionary view on biological rhythms. Proc R Soc B. 2013;280:2013433.
53. Bradshaw WE, Quebodeux MC, Holzapfel CM. Circadian rhythmicity and photoperiodism in the pitcher-plant mosquito: adaptive response to the photic environment or correlated response to the seasonal environment? Am Nat. 2003;161:735–48.
54. Tanaka K, Mizusawa K, Saugstad EB. A revision of the adult and larval mosquitoes of Japan (including the Ryukyu Archipelago and Ogasawara Islands) and Korea (Diptera: Culicidae). Contrib Am Entomol Inst. 1979;16:1–987.
55. Forattini OP. Culicini: *Haemagogus*, *Mansonia*, *Culiseta*. Sabethini, Toxorhynchitini. Arboviruses. Filariose bancroftiana. Genética. Entomologia Médica Vol 3. São Paulo: Universidade de São Paulo; 1965.
56. Judd DD. Review of the systematics and phylogenetic relationships of the Sabethini (Diptera: Culicidae). Syst Entomol. 1996;21:129–50.
57. Reppert SM. The ancestral circadian clock of monarch butterflies: role in time-compensated sun compass orientation. Cold Spring Harbor Symp Quant Biol. 2007;72:113–8.
58. Rund SSC, Hou TY, Ward SM, Collins FH, Duffield GE. Genome-wide profiling of diel and circadian gene expression in the malaria vector *Anopheles gambiae*. Proc Natl Acad Sci U S A. 2011;108:E421–30.
59. Meuti ME, Stone M, Ikeno T, Denlinger DL. Functional circadian clock genes are essential for overwintering diapause of the Northern house mosquito, *Culex pipiens*. J Exp Biol. 2015;218:412–22.
60. Leming MT, Rund SSC, Behura SK, Duffield GE, O'Tousa J. A database of circadian and diel rhythmic gene expression in the yellow fever mosquito *Aedes aegypti*. BMC Genomics. 2014;15:1128.
61. Weber F, Zorn D, Rademacher C, Hung H-C. Post-translational timing mechanism of the *Drosophila* circadian clock. FEBS Lett. 2011;585:1443–49.
62. Pegoraro M, Tauber E. Animal clocks: a multitude of molecular mechanisms for circadian timekeeping. WIREs RNA. 2011;2:312–20.
63. Panda S, Hogenesch JB, Kay SA. Circadian rhythms from flies to humans. Nature. 2002;417:329–35.
64. Gentile C, Rivas BS, Lima JBP, Bruna RV, Peixoto AA. Circadian clock of *Aedes aegypti*: effects of blood feeding, insemination and RNA interference. Mem Inst Oswaldo Cruz. 2013;108(Suppl I):80–7.
65. 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.
66. Mathias D, Jacky L, Bradshaw WE, Holzapfel CM. Geographic and developmental variation in expression of the circadian rhythm gene, *timeless*, in the pitcher-plant mosquito, *Wyeomyia smithii*. J Insect Physiol. 2005;51:661–67.
67. *Timeless* is an essential component of the circadian clock in a primitive insect, the firebrat *Thermobia domestica*. J Biol Rhythms 2012;27:126–134.
68. Kobelková A, Bajgar A, Dolezel D. Functional analysis of a circadian gene *timeless* promoter from the drosophilid fly *Chymomyza costata*. J Biol Rhythms. 2010;25:399–409.
69. Danbara Y, Sakamoto T, Uryu O, Tomioka K. RNA interference of *timeless* gene does not disrupt circadian locomotor rhythms in the cricket *Gryllus bimaculatus*. J Insect Physiol. 2010;56:1738–45.
70. Sauman I, Reppert S. Molecular characterization of prothoracicotropic hormone (PTTH) from the giant silkmoth *Antheraea pernyi*: developmental appearance of PTTH-expressing cells and relationship to circadian clock cells in central brain. Dev Biol. 1996;178:418–29.
71. Moriyama Y, Sakamoto T, Karpova SG, Matsumoto A, Noji S, Tomioka K. RNA interference of the clock gene *period* disrupts circadian rhythms in the cricket *Gryllus bimaculatus*. J Biol Rhythms. 2008;23:308–18.
72. Takekata H, Matsuura Y, Goto SG, Satoh A, Numata H. RNAi of the circadian clock gene *period* disrupts the circadian rhythm but not the circatidal rhythm in the mangrove cricket. Biol Lett. 2012;8:488–91.
73. Takekata H, Numata H, Shiga S, Goto SG. Silencing of the circadian clock gene *Clock* using RNAi reveals dissociation of the circatidal clock from the circadian clock in the mangrove cricket. J Insect Physiol. 2014;68:16–22.
74. Moriyama Y, Kamae Y, Uryu O, Tomioka K. *Gb'Clock* is expressed in the optic lobe and is required for the circadian clock in the cricket *Gryllus bimaculatus*. J Biol Rhythms. 2012;27:467–77.
75. Ikeno T, Tanaka SI, Numata H, Goto SG. Photoperiodic diapause under the control of circadian clock genes in an insect. BMC Biol. 2010;8:116.
76. Uryu O, Karpova SG, Tomioka K. The clock gene *cycle* plays an important role in the circadian clock of the cricket *Gryllus bimaculatus*. J Insect Physiol. 2013;59:697–704.
77. Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, et al. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. BMC Genomics. 2009;10:219.

78. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high density picoliter reactors. Nature. 2005;437:376–80.
79. Tae H, Ryu D, Sureschandra S, Choi J-H. ESTclean: a cleaning tool for next-gen transcriptome shotgun sequencing. BMC Bioinformatics. 2012;13:247.
80. Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q. An Ergatis-based prokaryotic genome annotation web server. Bioinformatics. 2010;26:1122–4.
81. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
82. http://blast.jcvi.org/web-hmm/. Accessed Jul 2011.
83. Min XJ, Butler G, Storms R, Tsang A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 2005;33:W677–80.
84. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
85. Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, et al. The FlyBase Consortium. FlyBase: enhancing *Drosophila* Gene Ontology annotations. Nucleic Acids Res. 2009;37:D555–9.
86. Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucl Acids Res. 2013;41:D358–65.
87. http://www.Geneious.com. Accessed Nov 2014
88. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–97.
89. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52.
90. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56:564–77.
91. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011;27:1164–5.
92. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52:696–704.
93. https://github.com/stephaneguindon/phyml/ Accessed Dec 2014.
94. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, et al. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol. 2008;17:1636–47.
95. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23:1061–7.
96. Chen F, Mackey AJ, Stoeckert CJ, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006;34:D363–8.
97. Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV. Ortho DB: The hierarchical catalog of eukaryotic orthologs in 2011. Nucl Acids Res. 2011;39:D283–8.
98. Cashmore AR, Jarillo JA, Wu Y-J, Liu D. Cryptochromes: blue light receptors for plants and animals. Science. 1999;284:760–5.
99. Ingram KK, Kutowoi A, Wurm Y, Shoemaker D, Meir R, Block G. The molecular clockwork of the fire ant *Solenopsis invicta*. PLoS One. 2012;7:e45715.
100. Haug MF, Gesemann M, Lazović V, Neuhauss SCF. Eumetazoan cryptochrome phylogeny and evolution. Genome Biol Evol. 2015;7:601–19.
101. Emerson KE, Bradshaw WE, Holzapfel CM. Concordance of the circadian clock with the environment is necessary to maximize fitness in natural populations. Evolution. 2008;62:979–83.
102. Boothroyd CE, Young MW. The in(put)s and out(put)s of the *Drosophila* circadian clock. 2008. Ann NY Acad Sci. 2008;1129:350–7.
103. Luo W, Sehgal A. Regulation of circadian behavioral output via a microRNA-JAK/STAT circuit. Cell. 2012;148:765–79.
104. Park JH, Helfrich-Förster C, Lee G, Liu L, Rosbash M, Hall JC. Differential regulation of circadian pacemaker output by separate clock genes in *Drosophila*. Proc Natl Acad Sci U S A. 2000;97:3608–13.
105. Jaramillo AM, Zheng X, Zhou Y, Amado DA, Sheldon A, Sehgal A, et al. Pattern of distribution and cycling of SLOB, Slowpoke channel binding protein, in *Drosophila*. BMC Neurosci. 2004;5:3.
106. Benito J, Zheng H, Hardin PE. PDP1ε functions downstream of the circadian oscillator to mediate behavioral rhythms. J Neurosci. 2007;27:2539–47.
107. Reddy AB, Maywood ES, Karp NA, King VM, Inoue Y, Gonzalez FJ, et al. Glucocorticoid signaling synchronizes the liver transcriptome. Hepatology. 2007;45:1478–88.
108. Nader N, Chrousos GP, Tomoshige K. Circadian rhythm transcription factor CLOCK regulates the transcriptional activity of the glucocorticoid receptor by acetylating its hinge region lysine cluster: potential physiological implications. FASEB J. 2009;23:1572–83.
109. Smith KM, Sancar G, Dekhang R, Sullivan CM, Li S, Tag AG, et al. Transcription factors in light and circadian clock signaling networks revealed by genomewide mapping of direct targets of Neurospora white collar complex. Eukar Cell. 2010;9:1549–56.
110. Jung C-H, Kim EM, Park JK, Park JK, Hwang S-G, Moon S-K, et al. Bmal1 supresses cancer cell invasion by blocking the phosphoinositide 3-kinase-Akt-MMP-2 signaling pathway. Oncol Rpt. 2013;29:2109–13.
111. Li Y, Guo F, Shen J, Rosbash M. PDF and cAMP enhance PER stability in *Drosophila* clock Neurons. Proc Natl Acad Sci U S A. 2014;111:E1284–90.
112. Mei Q, Dvornyk V. Evolution of PAS domains and PAS-containing genes in eukaryotes. Chromosoma. 2014;123:385–405.
113. Shafer OT, Yao Z. Pigment dispersing factor signaling and circadian rhythms in insect locomotor activity. Curr Opin Insect Sci. 2014;1:73–80.
114. Zhang Y, Fang B, Emmett MJ, Damle M, Sun Z, Feng D, et al. Discrete functions of nuclear receptor Rev-erbα couple metabolism to the clock. Science. 2015;348:1488–92.
115. http://www.ncbi.nlm.nih.gov/under Bioproject, enter 259209. Accession: PRJNA259209. Entered 22 Aug 2014.
116. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 24 May 2015.
117. http://www.megasoftware.net/. Accessed 19 Jul 2015.
118. Cyran SA, Buchsbaum AM, Reddy KL, Lin M-C, Glossop NRJ, Hardin PE, et al. Blau J. *vrille, Pdp1*, and *dClock* form a second feedback loop in the *Drosophila* circadian clock. Cell. 2003;112:329–41.
119. Kadener S, Stoleru D, McDonald M, Nawthean P, Rosbash M. *Clockwork Orange* is a transcriptional repressor and new *Drosophila* circadian pacemaker component. Genes Devel. 2009;21:1675–86.
120. Richier B, Michard-Vanhée C, Lamouroux CP, Royer F. The clockwork orange *Drosophila* protein functions as both an activator and repressor of clock gene expression. J Biol Rhythms. 2008;23:103–16.
121. Pesche N, Ko FC, Szabo G, Stanewsky R. Light-dependent interaction between the *Drosophila* circadian clock factors Cryptochrome, Jetlag and Timeless. Curr Biol. 2009;19:241–7.
122. Rosato E, Kyriacou CP. Flies, clocks and evolution. Phil Trans R Soc Lond B. 2001;356:1769–78.
123. Chiu JC, Ko HW, Edery I. NEMO/NLK phosphorylates PERIOD to initiate a time-delay phosphorylation circuit that sets circadian clock speed. Cell. 2011;145:357–70.
124. Grima B, Lamouroux A, Chélot E, Papin C, Limbourg-Bouchon B, Rouyer F. The F-box protein Slimb controls the levels of clock proteins Period and Timeless. Nature. 2002;420:178–82.
125. Meissner R-A, Kilman VL, Lin J-M, Allada R. TIMELESS is an important mediator of CK2 effects on circadian clock function *in vivo*. J Neurosci. 2008;28:9732–40.
126. Benna C, Bonaccorsi S, Wübeck C, Helfrich-Förster C, Gatti M, Kyriacou CP, et al. *Drosophila timeless2* is required for chromosome stability and circadian photoreception. Curr Biol. 2010;20:346–52.
127. Ling J, Dubruille R, Emery P. KAYAK-α modulates circadian transcriptional feedback loops in *Drosophila* pacemaker neurons. J Neurosci. 2013;32:16959–70.
128. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014;346:763–7.
129. Reidenbach KR, Cook S, Bertone MA, Harbach RE, Wiegmann BM, Besansky NJ. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol Biol. 2009;9:298.