`

Circadian Rhythm in *Diaphorina citri*

Max Reynolds[1], Thomson Paris[2], Tom D'Elia[1], Teresa Shippy[3], Surya Saha[4]

[1]Indian River State College, Fort Pierce, FL 34981; [2]USDA ARS, Fort Pierce, FL 34981; [3]Division of Biology, Kansas State University, Manhattan, KS 66506; [4]Boyce Thompson Institute, Ithaca, NY 14853

Introduction:

The circadian rhythm is a process that involves multiple genes and allows a cell to maintain a roughly 24-hour cycle (Glossop et al. 2003). This cycle allows the cell to regulate gene expression that can control daily physiology and behavior (Glossop et al. 2003). Light is the main stimulus controlling the synchronization of the circadian rhythm to a day-night cycle and may possibly influence an organism's ability to keep a seasonal rhythm due to the change in day length over various seasons (Lincoln et al. 2003). The basic mechanism for circadian clocks in animals is highly conserved and is best characterized in mice and *Drosophila* (Cortes et al. 2010).

There are two main feedback loops in the *Drosophila melanogaster* circadian rhythm which are regulated by six transcription factors (Cryan et al. 2003). The first loop involves the expression of *period* (*per*) and *timeless* (*tim*) genes, which encode the transcription factors PER and TIM (Cortes et al. 2010). The expression of *per* and *tim* is promoted by CLOCK and CYCLE, which form a CLK-CYC heterodimer and bind to the E-box promoters of *tim* and *per* (Cortes et al. 2010). The levels of the *per* and *tim* transcripts peak at dusk, but production of PER and TIM protein is delayed, causing the proteins to accumulate at night (Cortes et al. 2010). The newly formed PER-TIM complex is then imported into the nucleus, where it interacts with the CLK-CYC heterodimer, causing transcriptional inhibition of the *tim* and *per* genes, resulting in mRNA levels being minimum at dawn (Hardin 2005). In the second feedback loop, CLK-CYC heterodimers activate

the dusk transcription of *vrille* and *pdp1*, which control the expression of *clock*, causing the peak of CLOCK to be opposite that of PER and other CLK-CYC activated genes (Cortes et al. 2010).

Light is involved in the circadian rhythm through *cryptochrome 1* (*cry1*), a photoreceptor, which when activated by light degrades TIM (Cortes et al. 2010). By degrading TIM, CRY1 prevents the PER-TIM heterodimer from forming, allowing the continued activity of the CLK-CYC complex (Lin et. al 2001). In nature there are multiple types of CRY; CRY1 functions as a blue light receptor, while CRY2 functions as a circadian transcription repressor (Emery et al. 2000, Zhu et al. 2005). The presence of two different *cry* genes in insects gives possible insight into the evolution and function of the circadian rhythm in insects (Yuan et al. 2007). There are currently three different CRY feedback loop models which are based on the presence of CRY 1 in *Drosophila*, CRY 1 and CRY 2 in the butterfly *Danaus plexippus,* and CRY 2 in the honeybee *Apis mellifera* and the beetle *Tribolium castaneum* (fig 1) (Zhu et al. 2005; Yuan et al. 2007).Because of the presence of *cry2* in non-dipterans, the *D. melanogaster* model cannot be completely generalized when considering the first feedback loop (Yuan et al. 2007).

Huanglongbing, also known as citrus greening disease, is a severe citrus disease caused by the non-culturable bacteria *Candidatus* Liberibacter asiaticus (Garnier et al, 2000). The bacteria inhabit the citrus phloem causing infected plants to have many symptoms including inedible fruit (da Graca 1991). The Asian citrus psyllid (*Diaphorina citri*) transmits citrus greening disease causing crop damage (Halbert and Manjunath, 2004). Due to the economic impact caused by *D. citri*'s spreading of citrus greening, many methods are being investigated to combat the spread of the disease. Better understanding of the *D. citri* genome would lead to better molecular therapeutics. The genes involved in the main circadian rhythm pathway loops, along with many ancillary circadian rhythm genes, have been manually annotated in the *D. citri* genome (table 1). These newly annotated genes could be applied in molecular therapeutics to interrupt the circadian rhythm in *D. citri*, leading to altered behavior, which could possibly stop or slow the spread of the disease.

Materials and Methods:

The *D. citri* genes in the circadian rhythm pathway were annotated following the guidelines for community annotation described previously (Hosmani et al 2019). Orthologous sequences for the circadian rhythm pathway were collected from Genebank on NCBI and Uniprot. Sequences from Hemiptera were first collected, but if no sequences were found in the databases, *D. melanogaster* sequences were used. Collected sequences were then blasted against the Citrus Greening Solutions *D. citri* MCOT model database (Flores et al. 2017). Gene models were identified that shared sequence similarity with the orthologous sequence. The identified gene models were reciprocal blasted using NCBI BLASTp program, and if Blast results for *D. citri* matched orthologous sequences from other insects with the same domains, the model was used for annotation. Web Apollo was used for manual annotation and evidence tracks were used to support a model. RNAseq data was used to evaluate exon boundaries. Iso-Seq data was used to validate each gene model. Orthologous tracks provided sequences from other Hemiptera to help compare *D. citri* gene models to gene models in closely related organisms (Table 1). Following annotation, the resulting model was blasted against the NCBI non-redundant protein database to confirm the accuracy of the model. The gene copy numbers of circadian rhythm genes from other insects were gathered from NCBI, Uniprot, Flybase, and published literature (Table 2).

Results and Discussion:

Gene models were annotated in the *D. citri* genome for the first circadian rhythm feedback loop based on the *D. melanogaster* model, which consists of: *clock* (*clk*), *cycle* (*cyc*), *period* (*per*), and *timeless* (*tim*) (Peschel and Helfrich-Forster 2011). All four genes were found as complete models with good evidence supporting their presence in the genome (table 1). Only one copy of each gene was found in the *D. citri* genome, which is consistent with gene copy numbers of other insects (Table 2) (Cortes et al. 2010).

We also annotated *timeless* and the timeless paralog *timeout* (also known as *timeless 2*) (Rubin et al. 2006). To confirm that the models made were two different genes and not the result of a false duplication in the genome assembly, we compared them to orthologs from other insects. The protein sequences for Timeless from *D. melanogaster* (NP_722914.3) and in *Acyrthosiphon pisum* (ARM65416.1) were gathered and the conserved domains were compared. In both the *D. melanogaster* and *A. pisum* sequences, the only conserved domain is the Timeless (pfam04821) domain near the N-terminus. The *D. citri* model reflects this conservation by containing only the Timeless domain near the N-terminus. Timeout is also found in both *D. melanogaster* (NP_524341.3) and *A. pisum* (XP_016662949.1), and contains two domains, the Timeless domain near the N terminus, and the Timeless C domain located near the C terminus end. The *D. citri* sequence for *timeout* has both the Timeless domain at the N terminus, and the Timeless C domain located at the C terminus, similar to the *D. melanogaster* and *A. pisum* protein sequences, showing support for the *timeout* model not being a duplication of *timeless*. While initial BLASTp results and a comparison of the domain differences between the two protein sequences provides good evidence that the annotated *timeless* and *timeout* genes annotated are not duplicates, there is an issue with the *timeout* sequence causing it to be split between different models (Table 1).

Vrille (*vri*) and *par domain protein one* (*pdpd1*) are regulated by CLK-CYC heterodimers and make up the second feedback loop and were identified in *D. citri* (table 2) (Peschel and Helfrich-Forster 2011). The annotation of the *vri* model may prove important since a nonfunctional *vri* leads to embryo mortality in *D. melanogaster* (George and Terracol 1997). This may make *vri* a useful target for trying to control *D. citri* populations.

Additional genes involved in the circadian rhythm, but not directly classified in the two feedback loops, were also annotated (Table 1). Of these genes, *takeout* is of interest as potentially a good molecular target to control *D. citri*. *takeout* in *Drosophila* has been shown to regulate starvation response, and if the gene or protein loses function, the organism dies faster in starvation conditions when compared to the wild

type (Sarov-Blat et al. 2000). *Takeout* and *takeout like* in *D. citri* have low gene copy numbers when compared to other non-drosophilid insects (Table 2). Two possible reasons for this are that the gene models reported in these other insects are computer predicted, and that many of the sequences are short, meaning they may be fragments of larger genes, which would cause an inaccurate count of copies present in other insects.

Doubletime (*dbt*) (also known as *discs overgrown* (*dco*)), which encodes a circadian rhythm regulatory protein that affects the stability of PER (Price et al. 1998), was also annotated. The mammalian homolog of *dbt* is *casein kinase 1 epsilon* (Kloss et al. 1998). The *Homo sapiens* (ABM64212.1) and *D. melanogaster* (NP_733414.1) homologs both have the STKc CK1 delta epsilon domain (accession number: cd14125) at the same position, amino acids 8-282. The *D. citri* model reflects the conservation between species by also exhibiting a CK1 delta epsilon domain located at the exact same position.

The presence of two different *cry* genes in insects gives possible insight into the evolution and function of the circadian rhythm in insects (Yuan et al. 2007). In *D. citri*, we found both *cry1* and *cry2* genes. A pairwise alignment between the two annotated models resulted in 41% amino acid identity with 93% query coverage. Despite low identity, the two models both have the DNA photolyase (pfam00875) and FAD binding 7 (pfam03441) domains. DNA photolyase (pfam00875) encodes for a light harvesting cofactor, and FAD binding 7 domain (pfam03441) is the FAD binding domain of DNA photolyase (Tamada et al. 1997). Phylogenetic analysis confirms that the two models were for different genes and not duplications (Fig 2). The presence of *cry1* and *cry2* genes in *D. citri* reinforces the data that non-drosophilid insects possess CRY 2 (Yuan et al. 2007). The presence of both genes also suggests that *D. citri* follows an ancestral [butterfly] clockwork model (fig 1) (Yuan et al. 2007). *D. citri* only has one *cry2* gene, and *A. pisum* has two *cry2* genes; this supports the assertion that *A. pisum* core clock genes may be evolving at a faster rate than in other hemipterans (Cortes et al. 2010).

The circadian rhythm is an important pathway for an organism's ability to regulate their biological systems in conjunction with the external environment (Glossop et al. 2003). The circadian rhythm genes in *D. citri* have been annotated to model complete feedback loops. Annotation of these genes provides a better understanding of the circadian pathway in *D. citri* and adds to the evidence supporting how hemipteran circadian rhythm pathways function.

Works Cited

Bell-Pedersen D, Cassone V, Earnest D, Golden S, Hardin P, Thomas T, Zoran M. 2005 Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nat Rev Genet 6: 544–556.

Cortes T, Ortiz-Rivas B, Martinez-Torres D. 2010. Identification and characterization of circadian clock genes in the Pea aphid Acyrthrosiphon pisum. Insect Molecular Biology. 19(2):123-139

Cyran S, Buchsbaum A, Reddy K, Lin M, Glossop R.J., Hardin P, Young M, Storti R, Blau J. 2003. Vrille, Pdp1, and Clock form a second feedback loop in the Drosophila circadian clock. Cell. 112:329-341

da Graca JV, 1991. Citrus Greening disease. Annual Review of Phytopathology 29, 109–136

Emery P, Stanewsky R, Helfrich-Forster C, Emery-Le M, Hall JC, Robash M. 2000. *Drosophila* CRY is a deep brain circadian photoreceptor. Neuron. 26: 493-504

Flores M, Saha S, Hosmani PS, Brown S, A Mueller L. Community portal for the Citrus greening disease resources [Internet]. Figshare; 2017. doi:10.6084/m9.figshare.5307391.v1

Garnier M, Jagoueix-Eveillard S, Cronje PR, Roux HF, Bové JM, 2000. Genomic characterization of a liberobacter present in an ornamental rutaceous tree, Calodendrum capense, in the Western Cape

Province of South Africa. Proposal of 'Candidatus Liberibacter africanus subsp. capensis'. International Journal of Systematic and Evolutionary Microbiology 50, 2119–25.

George H, Terracol R. 1997. The vrille gene of *Drosophila* is a maternal enhancer of decapentaplegic and encodes a new member of the bZip family of transcription factors. Genetics. 146:1345-1363

Glossop N, Houl J, Zheng H, Ng F, Dudek S, Hardin P. 2003. VRILLE feeds back to control circadian transcription of clock in the Drosophila circadian oscillator. Neuron. 37: 249-261

Hall, D.G. (2008) Biology, history and world status of *Diaphorina citri*. Proceedings of the International Workshop on Huanglongbing and Asian Citrus Psyllid, North American Plant Protection Organization. 1-11.

Halbert, S. E. and K. L. Manjunath. 2004. Asian citrus psyllids (Sternorrhyncha: Psyllidae) and greening disease of citrus: a literature review and assessment of risk in Florida. Fla. Entomol. 87 (3): 330-353.

Hosmani PS, Shippy T, Miller S, Benoit JB, Munoz-Torres M, Flores-Gonzalez M, et al. (2019) A quick guide for student-driven community genome annotation. PLoS Comput Biol 15(4): e1006682. https://doi.org/10.1371/journal.pcbi.1006682

Hung, T. H., S. C. Hung, C. N. Chen, M. H. Hsu and H. J. Su. 2004. Detection by PCR of Candidatus Liberibacter asiaticus, the bacterium causing citrus huanglongbing in vector psyllids: application to the study of vector-pathogen relationships. Plant Pathology. 53: 96-102

Jergic S, Ozawa K, Williams N, Su X, Scott D, Hamdan S, Crowther J, Otting G, Dixon N. 2007. The unstructured c-terminus of the tau subunit of *Escherichia coli* DNA polymerase subunit III holoenzyme is the site of interaction with the alpha subunit. Nucleic acids Research. 35(9):2813-2824

Kloss B, Price J, Saez L, Blau J, Rothenfluh A, Wesley C, Young M. 1998. The *Drosophila* clock gene *double-time* encodes a protein closely related to human casein kinase 1 epsilon. Cell. 94: 97-107

Krishnan B, Levine JD, Lynch MK, Dowse HB, Funes P, Hall JC, Hardin PE, Dryer SE. 2001. A new role for cryptochrome in a *Drosophila* circadian oscillator. Nature. 411: 313-317.

Lincoln G, Andersson H, Loudon A. 2003. Clock genes in calendar cells as the basis of annual timekeeping in mammals- a unifying hypothesis. Journal of Endocrinology. 179 (1):1-13

Peschel N, Helfrich-Forster C. 2011. Setting the clock- by nature: Circadian rhythm in the fruit fly *Drosophila melanogaster*. FEBS letters. 585:1435-1442

Price J, Blau J, Rothenfluh A, Abodeely M, Kloss B, Young M. 1998. *Double-time* is a novel *Drosophila* clock gene that regulated PERIOD protein accumulation. Cell. 94:83-95

Rubin E.B, Shemesh Y, Cohen M, Elgavish S, Robertson H.M., Bloch G. 2006. Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock. Genome Res. 16:1352-1365

Sarov-Blat L, Venus So W, Liu L, Rosbash M. 2000. The *Drosophila* takeout gene is a novel molecular link between circadian rhythms and feeding behavior. Cell. 101: 647-656

Tamada T, Kitadokoro K, Inaka K, Yasui A, de Ruiter PE, Eker AP, Miki K. 1997. Crystal structure of DNA photolyase from *Anacystis nidulans*. Nature Structural Biology. 4 (11):887-891

Yuan Q, Metterville D, Briscoe A, Reppert S. 2007. Insect Cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Molecular Biology and Evolution. 24(4):948-955

Zhu H, Yuan Q, Froy O, Casselman A, Reppert S. 2005. The two CRYs of the butterfly. Current Biology. 15 (23):953-954

Table 1: Manually annotated circadian rhythm pathway genes in Diaphorina citri. There are 33 models in total. Each model has been assigned an identifier and denoted as complete or partial based on available evidence. MCOT evidence means a de novo Oases or Trinity model from an independent transcriptome was identified and the sequence from that transcript was used to validate or modify our model. IsoSeq means single, long RNAseq reads generated with Pacific Biosciences technology were available and were used to help validate the exon structure of the model. RNASeq means that individually mapped Illumina RNASeq reads were used to help validate or modify our model. Ortholog means ortholog sequences from other insects and information about conserved motifs and domains had to be used to help determine the final annotation.

Gene	Gene Model		Evidence Supporting annotation			
	Complete	Partial	MCOT	IsoSeq	RNASeq	Ortholog
protein phosphatase PP2A 55 kDa regulatory subunit	X		X	X	X	X
serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon	X			X	X	X
serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon	X		X		X	
Take out Like	X				X	
Take out	X		X	X	X	X
Take out	X		X	X	X	X
CK2 beta subunit	X			X	X	X
CK2 alpha subunit	X				X	
E75	X		X	X	X	
Clock work Orange Isoform 1	X		X	X	X	
Clock work Orange Isoform 2	X		X	X	X	
Double Time	X		X	X	X	X
Vestigial	X		X	X	X	
Cry 1	X		X	X	X	X
Cry 2	X		X		X	X
Par Domain Protein 1 (Pdp1)	X		X	X	X	
Vrille	X		X	X		
Cycle	X		X	X	X	X
Ecdysone receptor	X		X	X	X	
HR3 Homolog	X		X	X	X	X
HR3/Rora	X		X	X	X	X
Vitellogenin 2	X		X	X	X	
Vitellogenin 1	X		X	X	X	X

Ovary Maturing Parsin	X		X	X	X	
Insulin Related Peptide	X		X		x	
Timeless 2/ Timeout fragment 1		X			X	
Timeless 2/ Timeout fragment 2		X		X	X	
Timeless 2/ Timeout fragment 3		X		X	X	
Timeless	X		X	X	X	
Period	X			X	X	X
Clock	X		X	X	X	X
Takeout like	X			X	X	
Lark	X		X	X	X	X

*Table 2. Gene copy numbers for D. citri compared to other insects. Gene copy numbers were obtained from NCBI, with the exception of D. melanogaster, whose copy numbers were obtained from flybase.org. * denotes gene copies retrieved from Cortes et al. 2010.*

Gene name	D. citri	A. pisum	B. tabaci	C. lectularius	H. halys	D. melanogaster	A. aegypti
Timeless	1	1*	1	1	1	1	1
Period	1	1*	1	1	1	1	1
Vestigial	1	0	1	1	1	1	1
Hr3	1	0	1	1	2	1	1
Cycle	1	1*	1	1	1	1	1
Clock	1	1*	0	1	0	1	0
Lirp Cant find sequence?	1	0	1		1	0	1
Vitellogenin 1	1	1	1	1	2	0	2
Vitellogenin 2	1	0	0	0	0	0	0
Ecdysone Receptor	1	0	1	1	1	1	1
Vrille	1	1*	0	0	0	1	0
Pdp1	1	1*	1	1	0	1	0
Doubletime (Ck1 epsilon)	1	1*	1	1	0	0	0
Clock work orange	1	0	0	0	1	1	1
Casein Kinase 2 alpha subunit	1	1*	1	0	2	1	1
Casein Kinase 2 beta subunit	1	1*	1	1	1	1	1
Timeless 2/ timeout/ timeless homolog	1	0	1	1	1	1	1
Ovary maturing Parsin	1	0	0	0	0	0	0
Hr3 homolog	1	0	0	0	0	0	0
E75	1	1	1	1	1	1	1
Cryptochrome 1	1	1*	1	0	0	1	3
Cryptochrome 2	1	2*	1	0	0	0	1
Takeout	2	3	0	1	11	1	8
Takeout like	1	7	12	19	23	0	1
Lark	1	1	1	0	1	1	1
PP2A 55 kDa subunit	1	1	1	1	2	0	1
serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon	2	1	1	1	1	0	1

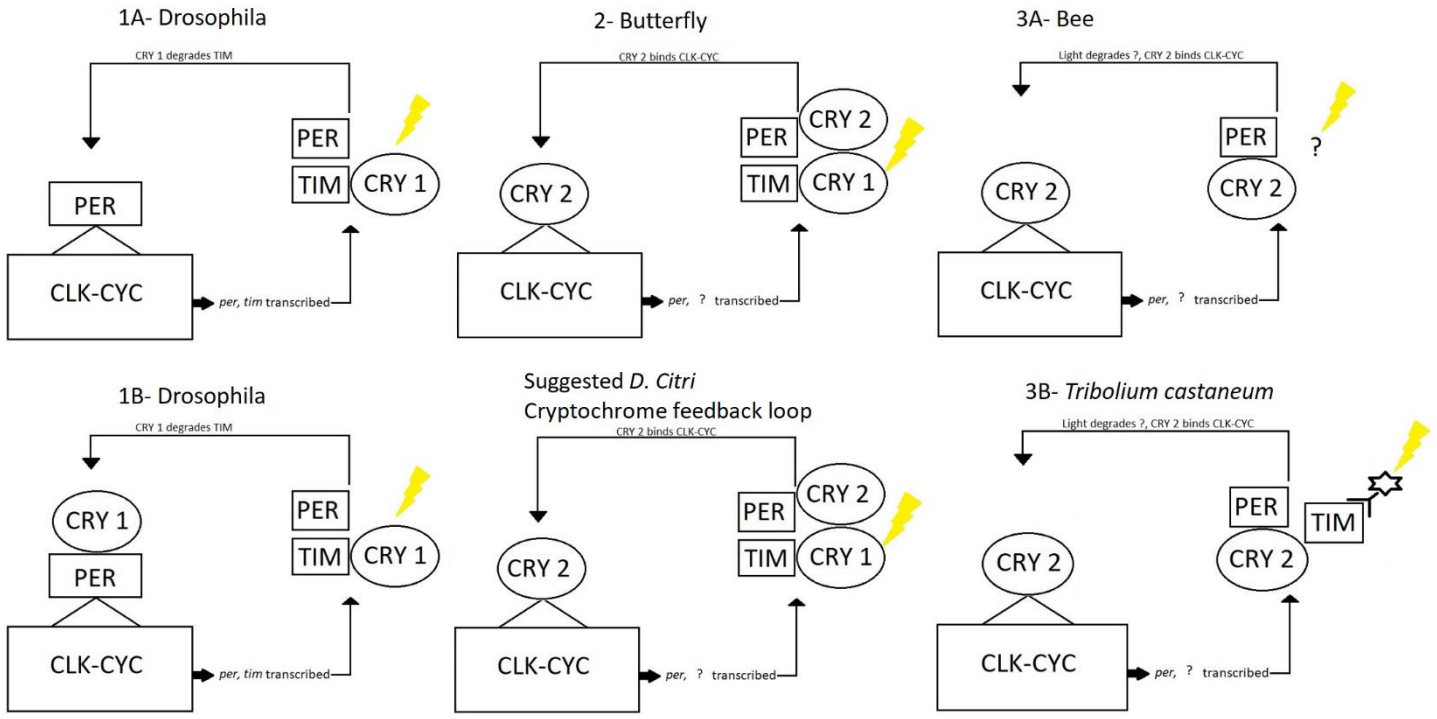
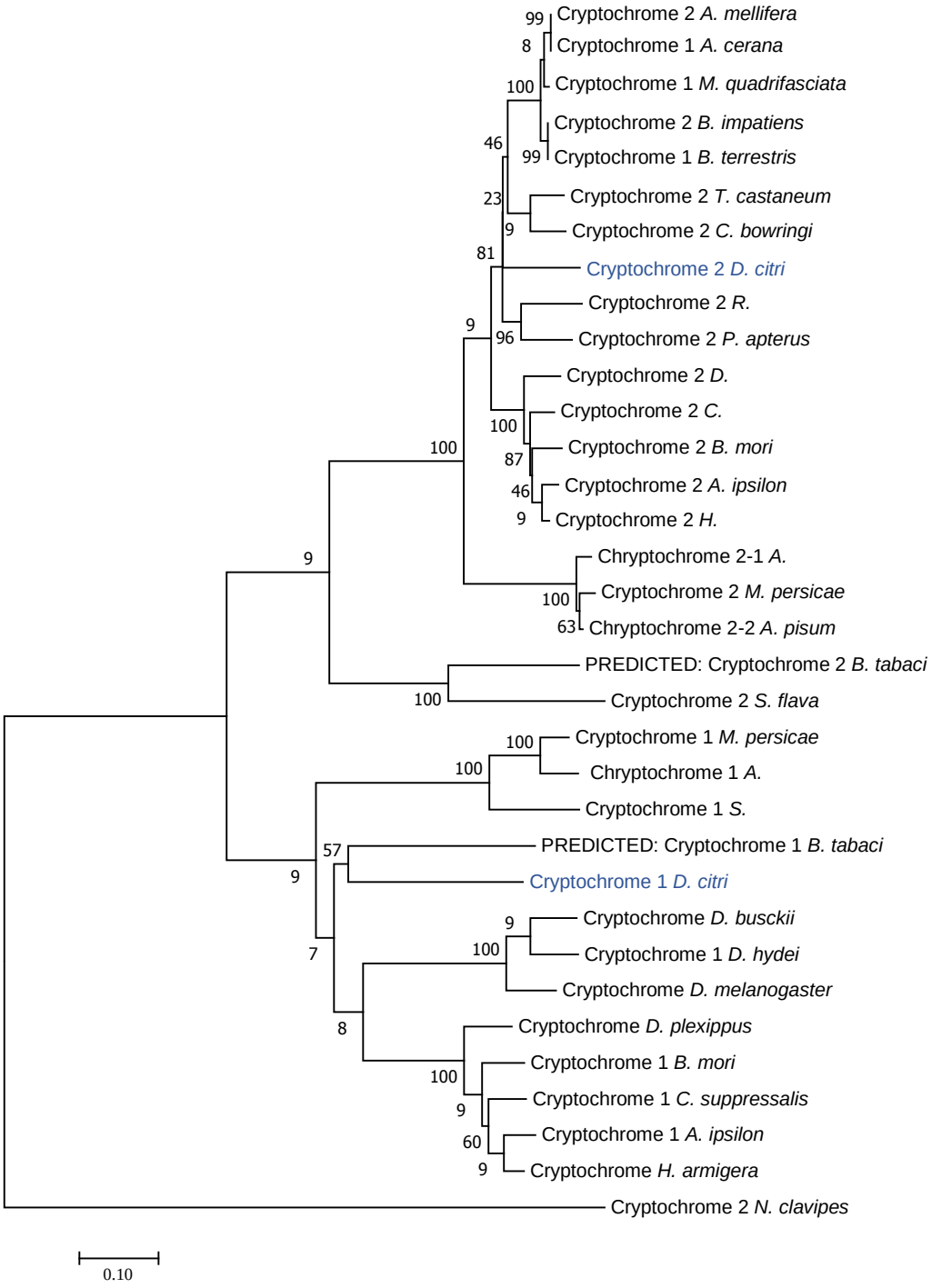


Figure 1. Insect clock work models, the existence of 2 functionally different CRY genes show a possible 3 different pathways possible (Yuan et al. 2007).Drosophila has 2 proposed models, panel 1a shows CRY 1 degrading time, and only PER binding the CLK-CYC complex (Emery et al. 2000). The second Drosophila model proposes CRY 1 degrades TIM then binds to PER (panel 1b)(Krishnan et al. 2001). The butterfly model (also known as the ancestral form) has both CRY 1 and CRY 2 performing different functions (Yuan et al. 2007). The bee and T. castaneum models have light entering the pathway from two different proteins since CRY 1 is not present, but CRY 2 serves the same function in both (Rubin et al. 2006).


Cryptochrome 2 A. mellifera
Cryptochrome 1 A. cerana
Cryptochrome 1 M. quadrifasciata
Cryptochrome 2 B. impatiens
Cryptochrome 1 B. terrestris
Cryptochrome 2 T. castaneum
Cryptochrome 2 C. bowringi
Cryptochrome 2 D. citri
Cryptochrome 2 R.
Cryptochrome 2 P. apterus
Cryptochrome 2 D.
Cryptochrome 2 C.
Cryptochrome 2 B. mori
Cryptochrome 2 A. ipsilon
Cryptochrome 2 H.
Chryptochrome 2-1 A.
Cryptochrome 2 M. persicae
Chryptochrome 2-2 A. pisum
PREDICTED: Cryptochrome 2 B. tabaci
Cryptochrome 2 S. flava
Cryptochrome 1 M. persicae
Chryptochrome 1 A.
Cryptochrome 1 S.
PREDICTED: Cryptochrome 1 B. tabaci
Cryptochrome 1 D. citri
Cryptochrome D. busckii
Cryptochrome 1 D. hydei
Cryptochrome D. melanogaster
Cryptochrome D. plexippus
Cryptochrome 1 B. mori
Cryptochrome 1 C. suppressalis
Cryptochrome 1 A. ipsilon
Cryptochrome H. armigera
Cryptochrome 2 N. clavipes
0.10

Organism	Cryptochrome 1	Cryptochrome 2
Acyrthosiphon pisum	D2T2J6	D2T2J8, D2T2J7
Agrotis ipsilon	AFJ22638.1	AFJ22639.1
Apis cerana	A0A2A3EC50	
Apis mellifera		NP_001077099.1
Bemisia tabaci	XP_018915448.1	XP_018904762.1
Bombus impatiens		NP_001267051.1
Bombus terrestris	XP_003398531.1	
Bombyx mori	NP_001182628.1	NP_001182627.1
Chilo suppressalis	CDK02014.2	AHL69753.1
Colaphellus bowringi		A0A1L5YQJ4
Danaus plexippus	A0A212EI23	Q0QWP3
Drosophila busckii	ALC47765.1	
Drosophila hydei	XP_023171160.1	
Drosophila melanogaster	NP_732407.1	
Helicoverpa armigera	ADN94464.1	ADN94465.2
Melipona quadrifasciata	KOX67887.1	AUN43314.1
Myzus persicae	AUN43313.1	
Nephila clavipes		A0A2P6LHB8
Pyrrhocoris apterus		M4WL45
Rhodnius prolixus		A0A0H2UI66
Sipha flava	A0A2S2Q297	A0A2S2R0G9
Tribolium castaneum		NP_001076794.1