Artif Life Robotics (2015) 20:347–352
DOI 10.1007/s10015-015-0237-6

ORIGINAL ARTICLE

Detection of periodic patterns in microarray data reveals novel oscillating transcripts of biological rhythms in *Ciona intestinalis*

Hiromi Matsumae[1,2] · Ryosuke Ishiwata[1,3] · Toshifumi Minamoto[4] · Norio Ishida[5,6] · Soichi Ogishima[7] · Hiroshi Tanaka[1,7]

Received: 7 April 2015 / Accepted: 20 October 2015 / Published online: 17 November 2015

Abstract A circadian rhythm is a roughly 24-h cycle in biological processes and physiological phenomena such as sleep, feeding, and photosynthesis for many organisms on Earth. The circadian patterns are coordinated by rhythmical gene expression of clock genes. Time-course transcriptomic analyses involving statistical methods have shown coordination of periodic gene expression in many organisms. Here we applied the cosine fitting method *COSOPT* to identify novel oscillating genes in microarray data for the chordate *Ciona intestinalis*. This organism showed rhythmic oxygen consumption in our previous study, but there were few homologous clock genes showing rhythmic mRNA expression. To understand circadian behavior at the transcriptomic level, we analyzed the 817 of 21,938 probes showing a 23- to 25-h period by means of COSOPT. Coupling the analysis of period detection with functional annotations indicated that previously unknown rhythmic mRNA expression might exist in *C. intestinalis*. In addition, we are releasing our implementation of COSOPT by means of R and C. All source code and supplementary information are available from https://github.com/mhiromi/cosopt.

Keywords *Ciona intestinalis* · Circadian rhythms · COSOPT · Microarray · Time-series data analysis · Transcriptome analysis

This work was presented in part at the 20th International Symposium on Artificial Life and Robotics, Beppu, Oita, January 21–23, 2015.

Electronic supplementary material The online version of this article (doi:10.1007/s10015-015-0237-6) contains supplementary material, which is available to authorized users.

✉ Hiromi Matsumae
hiromi.matsumae@gmail.com

1 Department of Bioinformatics, Tokyo Medical and Dental University, Tokyo, Japan

2 Present Address: Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland

3 Present Address: Department of Complex Systems Science, Graduate School of Information Science, Nagoya University, Nagoya 4648601, Japan

4 Graduate School of Human Development and Environment, Kobe University, Kobe, Japan

5 Ishida Gr. of Clock Gene, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

6 Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

7 Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan

1 Introduction

A circadian rhythm is an approximately 24-h cycle in a biological process. A circadian behavior is coordinated by a transcriptional network of clock genes, such as *Clock*, *Per1*, *BMAL1*, and *Cry2* and a transcriptional enhancer element *E-box* in mammals [1]. In terms of evolution, the molecular mechanism of circadian rhythms is one of the physiological responses conserved between insects and mammals. A statistical analysis that was based on time-series expression data uncovered a periodic change in the expression of clock genes, their regulatory genes at the transcriptomic level, and evolutionarily conserved components of circadian rhythms [2, 3]. In vertebrates, oscillation of the expression of clock genes in the central nervous system (CNS) drives the peripheral clocks of the body [4].

Considering the evolutionary development of the CNS, the chordate *Ciona intestinalis* is a well-studied model

organism in evolutionary biology, and its genome was sequenced in the early postgenomic era [5]. This organism is a sister clade of the vertebrates and has a primitive CNS called the *neural complex* [5]. The *C. intestinalis* genome, however, does not contain many clock genes [6, 7]. In a previous study, we measured oxygen consumption, analyzed microarray data, and demonstrated that *C. intestinalis* has circadian rhythms at the physiological and transcriptomic levels [8]. The study revealed oscillation of the expression of a few orthologs of clock genes and unannotated probes by means of the *moving window* analysis [4]. Although this method can detect a correlation between two time points, it does not test a specific period of the observed data.

To assess periodic gene expression using another statistical method, we utilized a well-known statistical algorithm: COSOPT. COSOPT estimates the best-fitting period of gene expression for each probe [2, 3, 9]. Because the original source code of COSOPT was not made public, we implemented COSOPT ourselves for this study. In addition, new and high-quality genome assembly and gene models of *C. intestinalis* were released in 2008 [10]. In the present work, we identified possible genes of the circadian rhythm in *C. intestinalis* by combining the published gene expression profiles obtained by means of COSOPT and the new genomic resources.

2 Materials and methods

2.1 The dataset

We used time-course gene expression profiles of cDNA microarrays, which we sampled previously at 6-h intervals [8]. The microarray data analyses (except for COSOPT) were carried out using R. The cDNA microarray data were normalized using local polynomial regression fitting the *loess* function of the Bioconductor package *marray* [11]. Then, the signal intensity for a probe at each time point was expressed as a ratio of circadian time 2 (CT2) on Day 1 to each time point (CT8, CT14, CT20, CT26, CT32, CT38, and CT44).

2.2 COSOPT

COSOPT detects periods of rhythmic gene expression as follows [2, 3, 9]: (1) this method generates many patterns of cosine waves to fit the time-series data from each probe/gene. First, test cosine functions were generated for 1000 periods that range from $^1/_{12}$ to $^1/_{12,000}$ frequency. For each period, 101 test cosine functions were generated, whose phases varied across a range of phase values from plus one-half the period to minus one-half the period in increments of 1 %. In total, 101,000 test functions were obtained. (2) Improving fitness of a cosine wave to the experimental data statistically, we also generated 1000 surrogates of the time-series gene expression signals for each probe with the addition of white noise generated from standard deviations of the signal. (3) Using the least squares method, we fitted 1001 gene expression data points from each probe (1 real data point + 1000 surrogates) to 101,000 cosine waves. Then, the method determined the period and phase for each probe with the minimum distance and within 5 % of one side. Finally, we utilized the period and phase of a probe if all 101 results showed the same period and phase.

The original COSOPT uses the standard deviation within replicates of each probe, but our data did not include replicates. Therefore, we set standard deviations to 0.1 in our analysis. We confirmed adequate performance of our COSOPT on the cluster computing system in our laboratory by means of published time-series gene expression profiles from circadian research. The data are available at the NCBI Gene Expression Omnibus (GEO accession number: GDS2232) [12]. We detected circadian periods of clock genes in these data (Online Resource 1–3). COSOPT for this study was implemented by means of C and R. The source code and all supplementary information are available online https://github.com/mhiromi/cosopt.

2.3 Gene annotations

Functional annotations of the probes were retrieved from the Kyoto Hoya Model (KH model) [10] and ANISEED database [13]. We extracted 1000-kbp upstream sequences corresponding to each detected gene from the KH models and computationally searched for E-box motifs in *Ciona* sequences (CANNTG) [14] by means of in-house scripts.

3 Results and discussion

Using COSOPT, we found that 10,681 of the 21,938 probes in the gene expression profiles were periodic (Fig. 1, Online Resource 4–6). Nevertheless, we restricted the periods

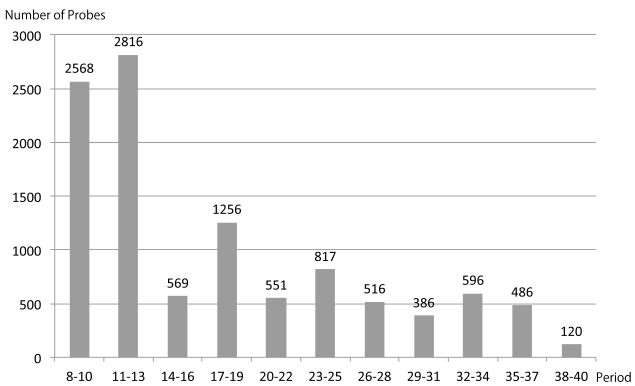

Fig. 1 The number of probes in each period range

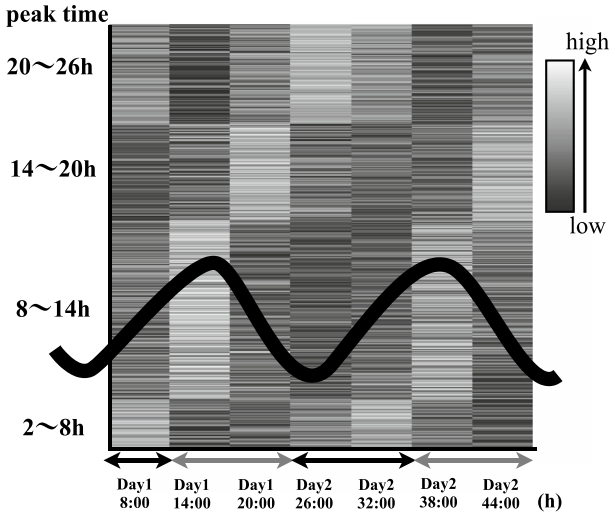

Fig. 2 A heatmap of expression of genes with 23- to 25-h periods. The *x*-axis indicates sampling time points. The *y*-axis represents clusters of phases, which showed circadian time (CT) with the highest signal intensity among all CTs. A *gray scale bar* on the *right* denotes signal intensity of probes from the highest (*white*) to lowest (*black*) on the *grayscale*

detected by COSOPT to 8–40 h because we used data at 6-h intervals to find a 24-h period, and because a sampling interval must be shorter than a half of the period we want to extract (the 12-h interval in our case), according to the so-called sampling theorem. Among these periodic probes, 817 probes clearly showed circadian-like oscillation with a period ranging from 23 to 25 h (Figs. 1, 2).

By comparing these data with our previous data [8], we constructed a correspondence table among KH models, probe IDs, and functional annotations (Table 1). Only *Hlf/Pdp1*—which contains three E-box motifs in the upstream sequence—oscillated according to both the previous and present datasets. We found that five of the eight oscillating genes (which were not orthologs of clock genes) showed rhythmic patterns among them (Online Resource 5). Two oscillating genes (*CT8-2* and *Amp4*; Table 1) were not annotated functionally in the current database but showed ~24-h periods in our analysis.

Then, we searched for other candidates for circadian genes on the basis of the functional annotations of the oscillating genes in our analysis (Table 2). The list included orthologs of presumably circadian rhythm-related genes, basic-helix-loop-helix (bHLH)-domain containing genes, and genes related to photoreceptors.

An ortholog of a well-known clock gene, casein kinase I (whose clone was isolated from the neural complex), oscillated with a longer period according to our analysis (Table 2).

In flies and mammals, most clock genes coordinate circadian rhythms as transcription factors. *Clock* and *Bmal1/Arntl* (which *cycle* in flies) contain the bHLH-Per ARNT

Table 1 Comparison of the oscillating genes between the previous [8] and this study

Name		KH accession	Number of E-box	Probe ID	Period	
					in 2010	This study
Orthologues of clock genes reported in 2010						
Mammal	Fly					
Hlf	*pdp1*	KH2012:KH.S1032.1	3	ciad045n22	Oscillated	12.0
Rxr		KH2012:KH.C9.892	2	citb010c11	ND	8.6
Fwd1	*slimb*	KH2012:KH.C5.76	2	cieg001c06	ND	ND
Ck1 epsilon	*double–time*	KH2012:KH.C7.188	3	cieg027g08	ND	35.9
Rev-erb alpha	*E75*	KH2012:KH.C13.51	6	cilv001k03	ND	ND
E4bp4	*vrille*	KH2012:KH.C5.255	3	ciad100b09	Weakly oscillated	ND
Ror alpha	*HR3*	KH2012:KH.C3.17	3	cilv010m12	ND	ND
Other genes in 2010						
Name in 2010	Name in ANISEED database					
CT8-2	–	KH2012:KH.C2.273	0	cibd034k24	Oscillated	23.4
Amp2	–	KH2012:KH.C9.458	4	citb064b04	Oscillated	ND
Amp4	–	KH2012:KH.C2.1070	4	ciad033o23	Oscillated	24.4
Amp30	HMCN1 (Ci-meta2)	KH2012:KH.C6.197	1	ciad002j17	Oscillated	30.8
				ciad002j17	Oscillated	31.1
Amp31	–	KH2012:KH.C7.152	3	cieg032o11	Oscillated	ND
Amp37	–	KH2012:KH.C1.602	6	citb005m02	Oscillated	ND
Amp56	–	KH2012:KH.C1.805	3	cigd012m22	Oscillated	9.1

Table 2 Representative oscillating genes in this study

Name	KH accession	Number of E-box	Probe ID	Clone library		Period
				Tissues	Developmental stage	
Known circadian genes						
TIMELESS	KH2012:KH.C1.487	3	cign004g15	WO	Gastrula and neurula	12.0
			cicl007g14	WO	Cleavage	15.1
TIPIN	KH2012:KH.C4.657	3	cieg096d24	WO	Egg	19.1
			cicl103k03	WO	Cleavage	20.4
CK1(CSNKIE)	KH2012:KH.C7.188	3	cieg027g08	WO	Egg	35.9
CK2(CSNK2B)	KH2012:KH.C1.412	4	cinc014g17	NC	Mature adult	8.8
			cinc034k20			18.0
			cinc014g17			18.5
bHLH-PAS domain including genes						
ARNT	KH2012:KH.C5.213	1	ciad002e16	WO	Young adult	16.5
NPAS3	KH2012:KH.L154.23	3	citb035h10	WO	Tailbud	23.8
bHLH domain including genes						
AP4	KH2012:KH.C14.448	8	cilv006k04	WO	Swimming larva	30.4
MAD2	KH2012:KH.C14.279	1	cieg055o16	WO	Egg	15.9
Max	KH2012:KH.L20.34	3	cieg011p17	WO	Egg	18.1
Photoreception						
COP9 homologue (CSN)	KH2012:KH.L96.66	4	cicl044e06	WO	Cleavage	9.3
	KH2012:KH.C12.484	1	cicl030a12	WO	Cleavage	8.6
	KH2012:KH.C2.895	4	cicl014d11	WO	Cleavage	12.0
	N/A		cicl056m14	WO	Cleavage	23.2
	N/A		cieg021l23	WO	Cleavage	28.4
CRALBP(RLBP1)	KH2012:KH.C13.36	5	cicl006l14	WO	Cleavage	12.0
	KH2012:KH.C11.439	1	cilv021j03	WO	Swimming larva	18.3
						24.0
PDE6D(PDED)	KH2012:KH.C12.604	7	cicl042m09	WO	Cleavage	18.1

WO and NC mean "whole organism" and "neural complex", respectively

Sim (bHLH-PAS) domains in mammals and insects [6, 15]. Because the bHLH domain is a major motif among transcription factors and is also found in *Ciona* [7], we compiled the data on the periods of several bHLH genes (Table 2). Only orthologs of *NPAS3* and *ARNT* in the bHLH-PAS family of genes showed periodic expression in *C. intestinalis*; *NPAS3* plays an important role in neurogenesis and schizophrenia but not in circadian rhythms in mammals [16, 17]. A clone of the *NPAS3* ortholog in *C. intestinalis* was originally found in a whole body specimen at the tail bud stage (regulating formation of the CNS) in an ascidian, but not in neural complex extracted from an adult organism [7, 18]. Even though clock genes encoding the bHLH and PAS domains are absent in *Ciona*, the *Ciona NPAS3* ortholog showed a clear-cut circadian period (23.8 h).

In the regulation of the circadian rhythms, phototransduction is a key mediator and synchronizes the body with the environment (e.g., jetlag and seasonal daylight changes). Unexpectedly, we observed cyclic expression of some genes related to phototransduction, even though the RNA was extracted from a whole body specimen of *C. intestinalis* in this experiment. For example, one probe of *CRALBP* (which is expressed in *Ciona* photoreceptor cells [19]) showed a 24.0-h period. A recent study showed that *CRALBP* is a direct downstream factor of *Pax6* and is coexpressed with *Pax6* in several regions of the CNS [20]. In addition, *Pax6* regulates *CLOCK* sleep–wake cycles in *Drosophila* neurons [21]. These pieces of evidence are suggestive of an unknown function of *CRALBP* in circadian rhythms in a broad range of animals. Another probe of *COP9* (*CSN*) showing a light response in flies and the vampire bat [22, 23] has a 23.2-h period.

If *C. intestinalis* has circadian rhythms, then our results suggest that these orthologs of nonclock genes may have replaced clock genes in this animal. The rhythmic behavior of oxygen consumption is supported by good physiological evidence, but it is still unclear what kinds of

genes are involved in the pathways of oxygen consumption in *C. intestinalis*. Therefore, analysis of the metabolic mechanisms or identification of other evidence of circadian behavior in this animal would be necessary to understand the biological meaning of the rhythmic gene expression.

In addition, analysis of tissue-specific gene expression should help to identify the "clock genes" and the core organ controlling circadian rhythms in *C. intestinalis*. The pacemaker of circadian rhythms in vertebrates is the suprachiasmatic nucleus (SCN) in the brain. The neural complex of ascidians performs several physiological functions, just as the vertebrate brain does, e.g., the light response [24] and hormone release [25, 26]. The neural complex shows expression of neuron-specific genes [27, 28]. Therefore, analysis of oscillating genes such as *NPAS3* and *CRALBP* in the neural complex will be a good subject of a future study designed to identify circadian rhythms at the transcriptomic level in *C. intestinalis*.

In this study, we focused on the transcripts linked to circadian rhythms (according to functional annotations) and on the transcripts with 23- to 25-h periods. Our COSOPT detected numerous oscillating transcripts that are outside the range of circadian behavior (Online Resource 4). The transcripts with specific suitable periods may shed light on the molecular mechanisms behind other novel biological rhythms such as the tidal rhythm.

Approximately a half of the periodic probes have short periods: from 8 to 13 h (Fig. 1 and Online Resource 4). We believe that this phenomenon can be attributed to the detection limits of COSOPT and the sampling interval used. The gene expression data match test cosine waves by both the specific period and a twofold period according to COSOPT. For example, the periods of two probes corresponding to cinc014g17 are 8.8 and 18.5 h, even though the observed data are similar to one another (Table 2 and Online Resource 6). Transcriptomic analyses by means of microarrays and next-generation sequencing became cheaper when we started the work on this study; therefore, it is better to obtain samples at shorter intervals to overcome this problem in future projects.

References

1. Gallego M, Virshup D (2007) Post-translational modifications regulate the ticking of the circadian clock. Nat Rev Mol Cell Biol 8:139–148
2. Panda S, Antoch MP, Miller BH et al (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. Cell 109:307–320
3. Ceriani MF, Hogenesch JB, Yanovsky M et al (2002) Genome-wide expression analysis in Drosophila reveals genes controlling circadian behavior. J Neurosci 22:9305–9319
4. Akhtar RA, Reddy AB, Maywood ES et al (2002) Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. Curr Biol 12:540–550
5. Dehal P, Satou Y, Campbell RK et al (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298(80):2157–2167
6. Rubin E, Shemesh Y, Cohen M et al (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
7. Imai KS, Hino K, Yagi K et al (2004) Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development 131:4047–4058. doi:10.1242/dev.01270
8. Minamoto T, Hanai S, Kadota K et al (2010) Circadian clock in *Ciona intestinalis* revealed by microarray analysis and oxygen consumption. J Biochem 147:175–184
9. Straume M (2004) DNA microarray time series analysis : automated statistical assessment of circadian rhythms in gene expression patterning introduction. Methods Enzymol 383:149–166
10. Satou Y, Mineta K, Ogasawara M et al (2008) Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. Genome Biol 9:R152
11. Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80. doi:10.1186/gb-2004-5-10-r80
12. Oster H (2006) Transcriptional profiling in the adrenal gland reveals circadian regulation of hormone biosynthesis genes and nucleosome assembly genes. J Biol Rhythms 21:350–361
13. Tassy O, Dauga D, Daian F et al (2010) The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. Genome Res 20:1459–1468. doi:10.1101/gr.108175.110
14. Erives A, Corbo JC, Levine M (1998) Lineage-specific regulation of the Ciona snail gene in the embryonic mesoderm and neuroectoderm. Dev Biol 225:213–225
15. Panda S, Hogenesch JB, Kay SA (2002) Circadian rhythms from flies to human. Nature 417:329–335. doi:10.1038/417329a
16. Pieper AA, Wu X, Han TW et al (2005) The neuronal PAS domain protein 3 transcription factor controls FGF-mediated adult hippocampal neurogenesis in mice. Proc Natl Acad Sci 102:14052–14057
17. Pickard BS, Christoforou A, Thomson PA et al (2008) Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. Mol Psychiatry. doi:10.1038/mp.2008.24
18. Sasakura Y, Mita K, Ogura Y, Horie T (2012) Ascidians as excellent chordate models for studying the development of the nervous system during embryogenesis and metamorphosis. Dev Growth Differ 54:420–437. doi:10.1111/j.1440-169X.2012.01343.x
19. Takimoto N, Kusakabe T, Tsuda M (2007) Origin of the vertebrate visual cycle. Photochem Photobiol 83:242–247. doi:10.1562/2006-06-30-IR-957
20. Boppana S, Scheglov A, Geffers R, Tarabykin V (2012) Cellular retinaldehyde-binding protein (CRALBP) is a direct downstream target of transcription factor Pax6. Biochim Biophys Acta Gen Subj. doi:10.1016/j.bbagen.2011.09.015
21. Glossop N, Gummadova J, Ghangrekar I, Hardin P, Coutts G (2014) Effects of TWIN-OF-EYELESS on clock gene expression and central-pacemaker neuron development in drosophila. J Biol Rhythms 29:151–166
22. Knowles A, Koh K, Wu J-T et al (2009) The COP9 signalosome is required for light-dependent timeless degradation and Drosophila clock resetting. J Neurosci 29:1152–1162. doi:10.1523/JNEUROSCI.0429-08.2009

23. Rajan KE, Rajkumar R, Liao C-C et al (2010) Light-induced COP9 signalosome expression in the Indian false vampire bat Megaderma lyra. J Physiol Sci 60:43–49. doi:10.1007/s12576-009-0064-4
24. Tsutsui H, Oka Y (2000) Light-sensitive voltage responses in the neurons of the cerebral ganglion of *Ciona savignyi* (Chordata: Ascidiacea). Biol Bull 198:26–28
25. Tsutsui H, Yamamoto N, Ito H, Oka Y (1998) GnRH-immunoreactive neuronal system in the presumptive ancestral chordate, *Ciona intestinalis* (Ascidian). Gen Comp Endocrinol 112:426–432
26. Kawada T, Aoyama M, Okada I et al (2009) A novel inhibitory gonadotropin-releasing hormone-related neuropeptide in the ascidian, *Ciona intestinalis*. Peptides 30:2200–2205
27. Shoguchi E, Hamada M, Fujie M, Satoh N (2011) Direct examination of chromosomal clustering of organ-specific genes in the chordate *Ciona intestinalis*. Genesis 49:662–672
28. Matsumae H, Hamada M, Fujie M et al (2013) A methodical microarray design enables surveying of expression of a broader range of genes in *Ciona intestinalis*. Gene 519:82–90. doi:10.1016/j.gene.2013.01.042