Estimation of the Circadian Phase by Oscillatory Analysis of the Transcriptome in Plants

Mari TAKEOKA, Takanobu HIGASHI, Mie N. HONJO, Ayumi TEZUKA, Atsushi J. NAGANO, Yusuke TANIGAKI, and Hirokazu FUKUDA

1 Graduate School of Engineering, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan
2 Graduate School of Life and Environmental Science, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan
3 Center for Ecological Research, Kyoto University, Higano 509-3-2 Otsu, Shiga 520-2113, Japan
4 Faculty of Agriculture, Ryouoku University, Otsu, Shiga 520-2194, Japan
5 Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

(Received July 10, 2017; Accepted January 4, 2018)

The phase of the circadian clock (body time) is related to various aspects of physiology and metabolism in plants. Therefore, it is important to evaluate body time to control plant quality and growth. The molecular timetable method, a statistical analysis of the transcriptome, is useful for plant body time estimation. However, it is necessary to verify some assumptions in this method for high resolution estimation in plants under various conditions. To investigate whether the assumptions are valid, we analyzed the period length and expression pattern of genes. We used Arabidopsis thaliana as a model plant, and both lettuce and tomato as model crops. A. thaliana and lettuce were cultivated under conditions of constant light in a closed plant factory, and tomato was cultivated in a sunlight-type plant factory. We found that estimation of the period length of the plant circadian rhythm was possible and that the gene expression pattern correlates well with test-fit cosine curves. Thus, we successfully demonstrated high-resolution estimation of body time in plants using a molecular timetable method.

Keywords: Arabidopsis thaliana, circadian clock, Lactuca sativa L., molecular timetable method, Solanum lycopersicum, RNA-Seq

INTRODUCTION

Plants have a circadian clock, a biological oscillator with an approximately 24 h period, which is known to dominate various physiologically active rhythms (Harmer et al., 2000). Photoperiodicity induction of flowering, stomatal opening, and concentration of nutrients (e.g. ascorbic acid) are also regulated by the circadian clock (Kotchoni et al., 2008). Recent research has clarified that the internal body time of the circadian clock (the phase of the circadian clock) is also related to pest resistance and metabolism of reactive oxygen species (Lai et al., 2012; Goodspeed et al., 2013). Therefore, knowledge of the phase of circadian clock is thought to be important in control of growth and crop quality.

The circadian clock works via clock genes, whose expression varies periodically. For the model plant Arabidopsis thaliana, a set of genes with periodic variation in expression level, called clock genes or clock-related genes, such as LHY (LATE ELONGATED HYPOCOTYL), CCA1 (CIRCADIAN CLOCK ASSOCIATED 1), and TOC1 (TIMING OF CAB EXPRESSION 1) and PRRs (PSEUDO-RESPONSE REGULATORS) are known factors related to the central oscillator of the clock (Nakamichi et al., 2012; Nagel et al., 2015; Kamioka et al., 2016; Liu et al., 2016; Ezer et al., 2017). Such genes generate a self-sustained circadian rhythm by negative feedback loops (Nohales and Kay, 2016). On the other hand, based on global gene expression analysis through microarray analysis, about 20% of plant genes are under the control of the circadian clock (Li et al., 2017). Recently, for crop species such as rice, tomato and lettuce, the circadian rhythm of the transcriptome in the field and in the plant factory has been clarified (Nagano et al., 2012; Matsuizaki et al., 2015; Tanigaki et al., 2015; Higashi et al., 2016a; 2016b).

It is also possible to estimate the phase of the circadian clock by analyzing the periodicity of the transcriptome (Matsushika et al., 2000; Ueda et al., 2004). A method for estimating the phase of the circadian clock, called the molecular timetable method, has been constructed for mammals (Ueda et al., 2004). It is possible to estimate body time from transcriptome data with high accuracy within about 2 h. In the standard application of this method, the period of the circadian rhythm for each expressed gene is regarded as a constant value, 24 h. In addition, the waveform of the circadian rhythm is assumed to be a simple form, that is, a cosine curve. However, in plants, the period and the waveform of the circadian rhythm varies depending on species, environment of cultivation, and other factors (Ninomiya, 1984; Higashi et al., 2014).

In this study, the model plant A. thaliana, and lettuce
and tomato were used as test plants. *A. thaliana* and lettuce were cultivated under constant light conditions in a closed plant factory, and tomato was cultivated in a sunlight-type plant factory. For all three plants, we inferred the period and the waveform of the gene expression rhythm from time-course transcriptome data using the molecular timetable method.

MATERIALS AND METHODS

Plant materials and growing systems

We used *A. thaliana* and lettuce (*Lactuca sativa* L. cv. Frill Ice), both of which were cultivated in a controlled environment with illumination by light-emitting diodes (LEDs), and tomato (*Solanum lycopersicum* cv. Taiankichijsitu, Nanto Seed Co., Ltd., Nara, Japan), which was cultivated in a sunlight-type plant factory. *A. thaliana* was sown on an aseptic 40 mm diameter dishes containing 4 mL of Murashige and Skoog plant salt medium (2% (w/v) sucrose), solidified with 2 g L\(^{-1}\) gellan gum. Lettuce was sowed in a water-laden urethane sponge in a tray (400 mm \(\times\) 280 mm \(\times\) 70 mm). We raised seedlings in an incubator for 2 weeks (*A. thaliana*) or 1 week (lettuce) at 22°C with a photosynthetic photon flux density (PPFD) of 100 \(\mu\)mol m\(^{-2}\) s\(^{-1}\) fluorescent white light (TBL-145N tubes, TES Lighting Co., Ltd., Japan) under 12-h light and 12-h dark (12L/12D) cycles. *A. thaliana* and lettuce seedlings were then moved to a clean room at 22°C, 50% relative humidity, CO\(_2\) component 1,000 ppm, and illumination with a PPFD of 180 \(\mu\)mol m\(^{-2}\) s\(^{-1}\) using red (\(\lambda_s=660\) nm), green (\(\lambda_s=520\) nm) and blue (\(\lambda_s=450\) nm) LEDs (NE02-00160; Shibasaki Inc., Japan) at a ratio of 120: 40: 40. Lettuce was cultivated using a cultivation medium composed of tap water and fertilizers (N: P\(_2\)O\(_5\): K\(_2\)O: CaO: MgO at ratios of 10:8:27:0:4 and pH 6.0 and EC 2.0). The lighting conditions were 12L/12D until 15 d after transplanting, then were changed to continuous light (LL) conditions. The samples for RNA-Seq analysis were obtained from the whole aerial part of *A. thaliana* and the leaf with the largest area for lettuce. Sampling for analysis was carried out every 2 h for 4 d, starting 36 h after the change to LL conditions. All samples were immediately frozen in liquid nitrogen and kept in a deep freezer. For both plant species, samples were obtained in one cultivation trial.

For tomato experiments, plants were cultivated in a sunlight-type plant factory (44.8 m [W] \(\times\) 23 m [D] \(\times\) 5 m [H]) in the Faculty of Agriculture, Ehime University, Japan. Individual plants are usually cultivated for a year; in this experiment, tomato seedlings were grown by Berg Earth Co., Ltd. (Ehime, Japan) and transplanted into Grodan Delta rockwool cubes (10 cm [W] \(\times\) 10 cm [D] \(\times\) 6.5 cm [H]); Grodan, Roermond, Netherlands) in August 2013. Rockwool cubes were placed on Grotop Expert rockwool slabs (100 cm [W] \(\times\) 20 cm [D] \(\times\) 7.5 cm [H]; Grodan) at four cubes per slab spaced at 25 cm intervals and watered using nutrient solution (Sonnewald, 1985). There were 20 slabs set in a line, with 28 lines per greenhouse. We sampled young leaves from the tomato canopy in January 2014. In the sunlight-type plant factory, the lighting conditions, relative humidity, and CO\(_2\) concentration were dependent on the ambient atmosphere. We sampled fifth leaves every 2 h for 2 d, starting at 14:00 on 6 January 2014 and ending at 14:00 on 8 January 2014. The environmental conditions are shown in supplemental (Fig. S1). We clipped leaf segments and stored them at 0°C with RNAlater solution (Thermo Fisher Scientific, Valencia, CA, USA), an aqueous nontoxic tissue storage reagent that rapidly permeates tissue to protect the integrity of RNA. Samples of tomato were obtained in one cultivation trial.

RNA-Seq analysis

Total RNA was extracted with an RNeasy Plant Mini Kit (Qiagen, Venlo, Netherlands), the quality was checked with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and RNA quantity was measured using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). RNA-Seq libraries were constructed as previously described (Wang et al., 2011; Nagano et al., 2015). The sequence data, for a single-end 50 bp read length obtained by a HiSeq 2000 sequencer (Illumina, San Diego, CA, USA) for *A. thaliana*, lettuce, and tomato, have been submitted to the DDBJ Sequence Read Archive (http://trace.ddbj.nig.ac.jp/DRASeach) under the accession numbers DRA005748, DRA004542 and DRA003530, respectively. After sequencing, all reads of each sample were quality checked using FastQC. *A. thaliana* reads were mapped using RSEM with Bowtie 2 software to the predicted transcriptome models of TAIR10 in the TAIR database (https://www.arabidopsis.org). Lettuce and tomato reads were mapped against the respective reference sequence in the NCBI database (https://www.ncbi.nlm.nih.gov). Finally, expression levels were normalized by RPKM (Higashi et al., 2016b).

In this study, for all three species, gene expression levels during 48 h (25 time points) were used for analysis. We defined the sampling time (measured in hours) as elapsed time from the start of sampling. Therefore, the sampling time 0 h indicates a time of 36 h after changing to LL conditions for both *A. thaliana* and lettuce, and 14:00 in the sunlight-type plant factory for tomato. For these three species, our time-course data from RNA-Seq analysis addressed a similar time duration from the end of daytime.

Ljung-box test

Using a Ljung-Box test, we tested whether the time series gene expression level had any autocorrelation. If a gene shows a high autocorrelation value, it is considered to be periodically expressed. The calculated statistic was compared with the \(\chi^2\) distribution, and the significance level was set to 5%.

Analysis of period length

We used autocorrelation analysis to estimate the period length of the periodic genes selected by statistical gene expression variation analysis. Because the sampling interval was 2 h, the autocorrelation value \(r\) was calculated with 2 h time steps. A correlogram was plotted from these autocorrelation results and used to estimate the period.
ESTIMATING THE CIRCADIAN PHASE

length of each gene. The autocorrelation value representing each gene was determined as the maximum in the range of 20 to 28 h. In addition, if the maximum correlation value was less than 0.2, it was regarded as showing no correlation.

When the period length of gene expression is estimated using autocorrelation analysis, the precision of the estimation (the time resolution) depends on the sampling interval. Therefore, we estimated period length using cross-correlative analysis to test cosine curves. For genes with \( r \geq 0.2 \), test cosine curves were prepared with a period length every 0.5 h from 20 h to 28 h and the cross-correlation value for each gene was calculated. For each gene, the period length of the best-fit cosine curve was defined as the period length of the gene. Using this method, a representative value for \( T \), that is, the natural period (free-running period) of the circadian clock, was estimated.

### Molecular timetable method

Genes that showed periodicity of expression and a high amplitude in the time-course transcriptome were defined as time-indicating genes (TiGs) (Ueda et al., 2004). To analyze their periodicity, test cosine curves with peaks from 0 to \( T \) in increments of 1 min were prepared. The correlation values \( r \) between the respective time series data and the prepared cosine curves were calculated. The cosine curve with the maximum \( r \) was defined as the best-fit curve. The peak time of expression of each gene was estimated by this best-fit curve. In addition, the amplitude, \( a \), was obtained from the standard deviation divided by the average of the time series expression level. We calculated the values \( r \) and \( a \) for each gene independently.

Next, we plotted the expression profiles of TiGs. The number of TiGs was determined by the cut-off values of \( r \) and \( a \). We normalized the expression level of each TiG using its average and standard deviation (Ueda et al., 2004). We then plotted the expression profiles that described the normalized expression level as a function of the expression peak time for every sampling time. Finally, the phase of the circadian clock was determined by the peak of the best-fit cosine curve.

### RESULTS AND DISCUSSION

#### Selection of periodic genes

As a result of the Ljung-Box test, the number of genes that showed statistically significant autocorrelation was 25,912 in *A. thaliana*, 37,041 in lettuce, and 22,384 in tomato. These genes comprised 77.1% of all genes in *A. thaliana*, 71.1% in lettuce and 81.6% in tomato.

#### Analysis of period length

According to the procedure used for analysis (Fig. 1), 1,972 genes in *A. thaliana*, 4,256 genes in lettuce and 4,833 genes in tomato were selected as showing a circadian rhythm. In *A. thaliana*, some well-known clock genes constituting the central oscillator, for example, *CCA1*, *TOC1*, *LHY*, *PRR5*, *PRR7* and *PRR9*, were included in these 1,972 periodic genes. The distribution of the peak time of each correlogram is shown in Fig. 2. In all three species, the median value of the period length was 24.0 h, as expected from a functioning circadian rhythm. Although autocorrelation analysis roughly estimates the period length with a 2 h time resolution, the cross-correlation analysis with cosine curves can provide a higher resolution of period length. The distribution of period length is shown in Table 1.

![Fig. 1](image-url)  
*Fig. 1* Analytical scheme for circadian gene expression. The left side of this figure shows what analysis was done. The right side shows how many genes were selected by each analysis. For the first stage, the number of whole genes extracted by RNA-Seq is shown. In the second stage, the genes were divided into ones that showed autocorrelation by a Ljung-Box test and those that did not. In the third stage, based on autocorrelation analysis, the genes were divided into two categories, those with a period length between 20 h and 28 h (a circadian rhythm) and others (a non-circadian rhythm). For the fourth stage, the period length of plants as estimated by cross-correlation analysis is shown. The fifth row shows the number of TiGs used for the molecular timetable method.
length for all genes with $r_{\epsilon} \geq 0.2$ was obtained, as shown in Fig. 3. The median of these distributions was 23.5 h in *A. thaliana*, 24.0 h in lettuce, and 24.0 h in tomato. *A. thaliana* had a shorter natural period length than 24.0 h.

Because the tomato plants were cultivated in a sunlight-type plant factory, the period length of the circadian rhythm can be considered entrained to the environment, which has a period of 24.0 h (Hotta et al., 2007). In lettuce, the period length was estimated as 24.0 h based on analyzing the largest leaf per plant in this study. The period length of gene expression varies depending on the leaf age (Kim et al., 2016). Therefore, the period length estimate may differ when analyzing leaves of different ages or the entire aerial part of the plant. In *A. thaliana*, for which the entire aerial part was analyzed, and the samples were thought to contain many young leaves, the estimated period length was shorter than 24.0 h, as also observed by Fukuda et al. (2008). In addition, Takahashi et al. (2015) reported that, depending on the analyzed organs and cultivation conditions, gene expression may oscillate with a non-24.0 h period.

In previous research, a transgenic plant, *AtCCA1::LUC*, was used to estimate the period length of gene expression (Higashi et al., 2014). Using our method, however, it may be possible to estimate the period length of many plant species under actual cultivation conditions.
ESTIMATING THE CIRCADIAN PHASE

Evaluation of waveform of expression rhythm

Figure 4 shows the distribution of the cross-correlation value, \( r \), in the three species. All genes of \( A. \) thaliana, lettuce and tomato had a high correlation value; the median value was over 0.6. Thus, the expression pattern of oscillating genes can be approximated by a cosine curve. This suggests that a cosine curve is appropriate for use as a fitting curve for determining period length.

Estimation of body time

The TiGs were selected using the threshold of \( a \geq 0.15, r \geq 0.8 \). This threshold was also used in previous studies (Ueda et al., 2004; Higashi et al., 2016a). As a result, 173 \( A. \) thaliana genes, 885 lettuce genes, and 1,131 tomato genes were selected as TiGs. Using these TiGs, expression profiles were constructed by the molecular timetable method (Fig. 5).

Based on these expression profiles, the circadian phase (body time) was estimated at each sampling time. Figure 6 shows how much the estimated body time (\( y \) value) differs from the sampling time (\( x \) value). Therefore, if \( y = x \), there is no difference between the estimated body time and the sampling time. The coefficient of determination for the \( y = x \) plot was calculated to be \( R^2 = 0.9625 \) in \( A. \) thaliana, 0.9932 in lettuce and 0.9951 in tomato, all very close to 1, indicating that estimation error was small in all three species. These results suggest that the natural periodicity is expected in our analysis and that it is possible to estimate the body time using the estimated natural period length.

While the number of experiments was limited in this study, the circadian clock of plants under different conditions or in different species will be clarified using our method. Estimating the circadian rhythm accurately will advance the methodology for controlling the circadian clock, leading to optimized cultivation.
CONCLUSION

By investigating the distribution of the period length of all genes showing periodic oscillation of expression and taking the median value, we estimated the period length of the circadian rhythm in individual plants. The waveform of oscillating gene expression also showed a high correlation for test-fit cosine curves. Finally, the circadian phase was estimated using the molecular timetable method by introducing the natural period length obtained by our suggested method.

ACKNOWLEDGEMENTS

We are grateful to Prof. Kotaro Takayama and Dr. Noriko Takahashi for tomato cultivation and Prof. Hiroshi Kudoh for RNA-Seq analysis. This study was supported by the Joint Usage/Research Program of the Center for Ecological Research, Kyoto University. This study was also partly supported by JST PRESTO (No. JPMJPR1504 to HF), JST CREST (No. JPMJCR1502 to AJN), and JSPS KAKENHI (No. JP16H05011 to HF, Nos. JP16H06171 and JP16H01473 to AJN).

REFERENCES


