Nondestructive Estimation of Circadian Time in Harvested Green Perilla Leaves Using Hyperspectral Data

Shogo Nagano1,2, Yusuke Tanigaki1 and Hirokazu Fukuda1

1 Department of Mechanical Engineering, Graduate School of Engineering, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan
2 Research Fellow of the Japan Society for the Promotion of Science, Tokyo 102-0083, Japan

(REceived January 28, 2020; Accepted March 6, 2020)

The circadian clock, an internal oscillator with a period of approximately 24 hours, plays an important role in the regulation of biological processes, and an understanding of circadian rhythms can be employed to improve the quality of plant production. Many studies have measured the circadian rhythms of plants and estimate their circadian times. However, the circadian time estimation methods used in previous studies are difficult to apply to commercial crops because they require extraction of plant contents such as RNA, which involves destroying plant tissues. In this study, we sought to develop a nondestructive method for estimating circadian time in harvested leaves of green perilla (Perilla frutescens var. crispa f. viridis). The results of RNA sequencing (RNA-Seq) show that the gene expression of perillyl alcohol depend on the circadian time. A hyperspectral camera captured the light reflectance of 141 wavebands from 350 to 1,050 nm on leaves, and machine learning using the reflectance data successfully estimated the circadian time corresponding to the harvest time. The study results demonstrate the potential for the nondestructive use of hyperspectral reflectance data in circadian time estimation and its applicability to improving the quality of plant production.

Keywords: artificial neural network, circadian rhythm, hyperspectral camera, metabolome analysis, Perilla frutescens var. crispa f. viridis, RNA-Seq

INTRODUCTION

Vegetables that are harvested early in the morning or late in the afternoon are valued in plant production based on the diurnal variation of metabolism (Clarkson et al., 2005), which is regulated by a circadian clock with a period of approximately 24 hours. The circadian clock plays an important role in the regulation of biological processes, such as growth, photosynthesis, and flower induction (Barak et al., 2000; Dodd et al., 2005). Therefore, information on circadian time, that is, the internal body time denoted by the circadian clock, is useful in improving the quality of plant production. Recently, a statistical oscillatory analysis of time-series RNA sequencing (RNA-Seq) data, referred to as a molecular timetable method (MTM) (Ueda et al., 2004), was used to estimate the circadian time of various types of plant leaves, including lettuce and tomato (Higashi et al., 2016; Takeoka et al., 2018). A few hundred genes identified as time-indicating genes (TiGs) were found to exhibit circadian rhythms in their expressions, and the overall profile of their phases was found to represent circadian time. Although this MTM successfully estimates circadian time precisely, it requires extraction of RNA, which involves destroying plant tissues, and it also requires considerable time for sequencing.

In generally, optical features of plant tissues have been used in nondestructive real-time methods. Near-infrared spectroscopy can be used to estimate the soluble solids content of vegetables and fruits (Khurayti and Matsuoka, 2004), and a combination of visible and near-infrared wavelengths can be used to estimate the amount of chlorophyll (Markwell et al., 1995). Multispectral imaging (MSI) with a few dozens of different spectral bands has been shown to have an excellent ability to determine the spatial-spectral signature of plants, especially in characterizing a variety of chemical compositions and assessing the physiological status of plants. A recent study reported that MSI can be used nondestructively to detect circadian rhythms of chlorophyll concentration in soybean leaves (Pan et al., 2015). Therefore, it believed that circadian time can be estimated by means of multispectral analysis, although this has not been previously confirmed.

In this study, we used a hyperspectral camera, which is a device with an exceptionally high resolution (more than 100 bands) from visible to near-infrared wavelengths, because optical indices for estimation of circadian time have not been identified. In addition, it is expected that rich information is required for precise estimation of circadian time like as many TiGs in MTM. We also used a machine learning method based on an artificial neural network (ANN) to address nonlinearity of the relationships among wavelengths. As a starting point, we focused on the circadian time at harvest, which is a critical time for production quality. Our experiments were carried out using green perilla (Perilla frutescens var. crispa f. viridis), a...
type of basil mainly cultivated in East Asia that has a high nutritional content, including high levels of vitamin E and rosmarinic acid (Takano et al., 2004; Asif, 2012). This plant exhibits a very clear circadian rhythm in leaf movement during cultivation and a significant dependence of production quality on harvest time. To confirm that the circadian oscillation in functional substances is related to production quality, RNA-Seq and metabolome analyses were also conducted. Based on the results, we attempted to establish a precise and nondestructive method for estimating the circadian time in plants.

MATERIALS AND METHODS

Plant materials and growing conditions

Seeds of green perilla (Perilla frutescens var. crispa f. viridis) were sown under red light-emitting diodes (LEDs, $\lambda = 660$ nm) with a photosynthetic photon flux density of 80 $\mu$mol m$^{-2}$ s$^{-1}$ under 12-hour light-12-hour dark (12L:12D) conditions and a temperature of 25°C for 6 days. Then, the seedlings were transferred to water-laden sponge blocks in a covered tray (200 mm$\times$250 mm$\times$560 mm). This tray of seedlings was placed under 12L:12D using a white fluorescent lamp (130 $\mu$mol m$^{-2}$ s$^{-1}$; TBL-14/5N, Ohm Electric Inc., Saitama, Japan) for 14 days in a growth chamber (MLR-351HNB; Sanyo Electric Co., Ltd., Osaka, Japan). Then, the seedlings were placed in a deep-flow hydroponic system under a white fluorescent lamp (305 $\mu$mol m$^{-2}$ s$^{-1}$, 12L:12D, FLK206ED985; Prince Electric Co., Ltd., Yokohama, Japan) at 22°C for 14 days in a nursery. The plants were then transferred to another hydroponic cultivation system (1,200 mm$\times$650 mm$\times$2,000 mm; Mansei Corp., Osaka, Japan) under a different white fluorescent lamp (164 $\mu$mol m$^{-2}$ s$^{-1}$, 16L:8D, FHF32EX-D-HX-S; NEC Lighting, Ltd., Tokyo, Japan) and grown at 23±1°C for 52 days. The light and dark periods were set to 0:00-16:00 and 16:00-24:00 (0:00), respectively. Circadian time is generally defined as a normalized time scale with 24 hours period, in which 0:00 and 24:00 are set at the time of light-on and next light-on time, respectively. The nutrient solution was composed of tap water and fertilizer (N:P2O5:K2O:CaO:MgO = 10:8:27:0:4; Otsuka House et al., 2012). Hyperspectral reflectance image data were calculated using the HSD “Analyssiser” software (HSD “Analyssiser” software, EBA Japan Co., Ltd., Tokyo, Japan). All of the hyperspectral reflectance data were spatially averaged for each band—that is, only one representative value of each band for each leaf was used for circadian time estimation.

Estimation of circadian time using machine learning

The modeling was performed using the machine learning software NeuroWorks “Predict” (NeuralWare Inc., Carnegie, PA, USA). “Predict” is an artificial neural network (ANN) software that has been used in various studies (Francis and Stein, 2012; Ootomo et al., 2012; Gosukonda et al., 2015; Moriyuki and Fukuda, 2016; Gosukonda et al., 2017). ANN is effective at solving problems when the input data and output data have an ambiguous and complex relationship (Hopfield, 1988).

For our ANN model, a dataset was constructed with the hyperspectral reflectance and the circadian time as input and output data, respectively. Since the circadian time is cyclic (i.e., 24:00 means 0:00), six circadian times (0:00, 4:00, 8:00, 12:00, 16:00 and 20:00) were transferred in two-dimensional rectangular coordinates as $(x, y) = (0, 1), (0.866, 0.5), (0.866, -0.5), (0, -1), (-0.866, -0.5)$, and $(-0.866, 0.5)$. Thus, our ANN model consisted of 141 inputs (141 wave bands), one hidden layer, and two outputs (x, y). This hidden layer in “Predict” was optimized automatically in the software, blindfolded to the user (Gosukonda et al., 2015). The ratio of training data to test data was set to 33:15. To augment the data, 100 datasets were generated in which the combination between training data and test data was changed randomly. Figure 1A shows an example of the output for the reflectance data for the 0:00 harvest, the plot of which consists of approximately 250 output data points (y values for test; x values for the test data for the test; x values for the test data for the test; x values for the test data for the test; x values for the test data for the test; x values for the test data for the test). To extract only angle information, conversion from rectangular coordinates (x, y) to polar coordinates (r, $\theta$) was performed. Then, all of the output data were projected onto a unit circle (1, $\theta$) as shown in Fig. 1B. The final output $R$ in Fig. 1B was defined as the vector from the zero point to...
the center of gravity for all plots on the unit circle. The angle and length of $R$ represent the estimated circadian time and its likelihood in our trained model, respectively. $R$ also represents a Rayleigh test, which is a method used in biological-time studies, was used to examine statistically the synchrony among events, to determine whether a given distribution was significantly different from uniform (Duffield and Ebling, 1998; Refinetti et al., 2007).

RNA-Seq analysis, metabolome analysis, and aroma quality test

For the RNA-Seq analysis, two leaves was harvested at each 0:00, 1:00, 4:00, 8:00, 12:00, 15:00, 16:00, 17:00, 20:00 and 23:00, and stored immediately in a freezer at $-80^\circ$C. The total RNA was extracted from each sample using an Agilent Plant RNA Isolation Mini Kit, according to the manufacturer’s instructions (Agilent Technologies, Santa Clara, CA, USA). The RNA quantity was determined with a “bioanalyzer” (Agilent Technologies, Santa Clara, CA, USA). The RNA-Seq library preparation method is described in Wang et al. (2011) and Nagano et al. (2015). Transcriptome sequence analyses were carried out by the Beijing Genomics Institute (BGI, Shenzhen, China). A HiSeq 2000 sequencer (single-end, 50 bp; Illumina, San Diego, CA, USA) was used to read the files, and the construction of contigs was performed by BGI. A Trinity assembler (version 2011-11-26) was used for de novo short-read assembly (maximum contig length of 76,620 bp; minimum contig length of 201 bp). We estimated genetic information from genetic data on the model plant *A. thaliana* and published information on green perilla (Sato-Masumoto and Ito, 2014). This estimated information for 12,807 genes of perilla was linked to that of *A. thaliana* with genetic similarity (Tanigaki et al., 2017).

For the metabolome analysis, two leaves were harvested at each 0:00, 4:00, 8:00, 12:00, 16:00 and 20:00 and stored immediately in a freezer at $-80^\circ$C. The extraction process was performed by Inplanta Innovations, Inc. (Kanagawa, Japan). Liquid chromatography-mass spectrometry was performed with an Agilent 1200 series system (Agilent Technologies UK Ltd., Stockport, UK). Reversed-phase chromatography was performed with a phase C18 column (TSKgel ODS-100V 5 μm 3x50 mm; TOSOH, Tokyo, Japan). The mobile phase was a mixture of solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid with acetonitrile), according to a gradient. The injection volume was 5 μL. Mass was measured precisely with a mass spectrometer (LTQ ORBITRAP XL; Thermo Fisher Scientific, MA, USA). The ionization method was electrospray ionization in positive mode. Candidate signals of functional substances were identified by amounts of C, H and O.

For the aroma quality test, leaves were harvested at 0:00, 4:00, 8:00, 12:00, 16:00 and 20:00, and immediately placed in a refrigerator (SJ-WA35T, 7 ± 1°C; Sharp Corp., Osaka, Japan). Approximately 1 day later, the leaves were cut into strips for testing. Aroma quality testing was conducted using research participants (seven men in their 20s and two women in their 30s and 40s). The research participants graded the aroma on a five-point scale with a maximum of 2 for the best aroma and a minimum of −2 for the worst aroma.

RESULTS AND DISCUSSION

RNA-Seq and metabolome analyses were performed to confirm that the circadian oscillation in functional substances is related to production quality. Figure 2 shows the relationship between time and the expression of aldo-keto reductase gene (Fig. 2A), which converts limonene into perillyl alcohol; the peak area of perilla alcohol determined from metabolome analysis (Fig. 2B), which is a precursor of perillaldehyde; the aroma evaluation score (Fig. 2C); and the metabolic pathway of these aroma components in green perilla (Fig. 2D). The value of the peak area in Fig. 2B indicates the substance amount calculated from a 5-μL injection volume.
The gene expression level of aldo-keto reductase exhibited a large circadian oscillation, with a peak value at the middle of the light period (8:00). In contrast, the value of the peak area of perillyl alcohol, which is generated from aldo-keto reductase, remained almost constant throughout the day. The results of the aroma quality tests exhibited a diurnal variation, with the aroma quality being higher around the light-on time (i.e. morning) (Fig. 2C). Since this aroma quality is based on human sensation, it is strongly suppressed under low temperature; clock gene expressions of a perennial plant Arabidopsis halleri and cyanobacteria stop under 7°C and 19°C, respectively (Nagano et al., 2019; Murayama et al., 2017). This suppression of circadian clock might contribute to remaining the circadian time at harvest.

Next, we addressed the relationship between the hyperspectral reflectance data and the circadian time at harvest. Figure 3A and 3B show the spectral intensity of 500 nm and 745 nm, respectively, for each circadian time. The spectral intensity of 745 nm was found to depend on the circadian time, while that of 500 nm was independent of the circadian time. This indicates that the 745-nm spectral intensity included richer information about the circadian time than did the 500-nm spectral intensity. To obtain an overview of the contribution of each band to the circadian time estimation, we calculated the correlation ratio for all 141 wavelengths (bands between 350 and 1,050 nm) individually. The correlation ratio is a statistical measure of the correlation between quantitative and categorical data. In this study, the reflectance data and circadian time were treated as quantitative and categorical data, respectively. The value of the correlation ratio $\eta^2$ was determined in this study using the following equations, based on those presented by Huitema (2011):

$$\eta^2 = \frac{S_3}{S_0 + S_2}$$  \hspace{1cm} (1)

$$S_2 = n_1 (\bar{X}_1 - X)^2 + n_2 (\bar{X}_2 - X)^2 + \cdots$$  \hspace{1cm} (2)

$$S_3 = S_1 + S_2 + \cdots$$  \hspace{1cm} (3)

where $\bar{X}_n$ and $X$ are the means of category $n$ and all samples, respectively; $n_1$ and $S_1$ are the number of plots in category 1 and the sum of squared deviations of category 1, respectively; and $S_0$ and $S_2$ are between-class and within-class variations, respectively. Figure 3C shows the relationship between the correlation ratio and the wavelength. Wavelengths from 520 to 745 nm exhibited relatively high correlation ratios, whereas wavelength regions both 400–500 nm and 770–1,000 nm exhibited low correlation ratios. The 520-nm and 745-nm wavelengths are assumed to be related to anthocyanin and phycotochrome, respectively (McDonald, 2003). In addition, the 745-nm wavelength is not usually discussed with respect to chlorophyll, but previous studies have reported that the near-infrared range of the spectrum is correlated to the total chlorophyll content of leaves (Imanishi et al., 2010; Pan et al., 2015).

In Fig. 3C, the correlation ratio at 500 nm was 0.047, indicating that there was no contribution to the circadian time estimation. Otherwise, the correlation ratio at 745 nm was 0.357; this relatively high value reflects the diurnal variation of spectral intensity, which increased from 12:00 to 0:00 and decreased from 0:00 to 8:00 as shown in Fig. 3B. However, the highest correlation ratio was less than 0.357, which is regarded as a weak correlation. This may be because analysis of individual bands is not sufficiently precise. Thus, a model that combines multiple bands, such as ANN, was judged to be required for circadian time estimation from hyperspectral reflectance data.
Figure 4 shows the results of the circadian time estimation using our machine-learning model. Each unit circle describes a 24-hour clock and the angle of $R$ represents the estimated time at each harvest. The estimated times were 0:25, 2:58, 6:31, 12:18, 16:25 and 19:57 for the actual times of 0:00, 4:00, 8:00, 12:00, 16:00 and 20:00, respectively. The error in the estimation had a minimum of 3 minutes at 20:00 and a maximum of 89 minutes at 8:00. These results indicate that the model used may be able to estimate the circadian time within a range of less than 1.5 hours.

In addition, we successfully reduced the sampling efforts, using a small number of leaves (only 48) in the learning dataset. Similarly, in a previous study, Pan et al. (2015) used only 150 leaves. Figure 4G shows a sample of the experiment results for data for $R$ in the estimated time range of ±15 minutes, converging rapidly to the actual time. Thus, our model can perform time estimation with high convergence. Moreover, although the length of $R$ was small value at both 0:00 and 16:00 which were times when the light conditions changed as light-on or light-off, the time estimation error remained small in a range less than ±1.5 hours. The influence of light condition changes has been successfully removed in our model.

CONCLUSIONS

In this study, we used hyperspectral reflectance data and machine learning to estimate the circadian time in harvested green perilla leaves nondestructively. The results of this study reveal that reflectance data have potential for use in estimating the circadian time and that machine learning is an effective method for improving the accuracy of the estimation. Using a circular model with unit circle for the creation of continuous time information and artificially amplifying the amount of data, we succeeded in estimating the circadian time from a small amount of training data. However, intriguing questions remain, given that the data were not analyzed spatially, and the biological mechanism regulating the reflectance of leaves remains unclear. More research is needed to further explore the challenging topic of how to apply knowledge concerning the circadian clock to plant production.

ACKNOWLEDGMENTS

This study was supported, in part, by Grants-in-Aid for Scientific Research (18J14808 to S.N., 16H05011 to H.F.) and Precursory Research for Embryonic Science and Technology (No. JPMJPR1504) of the Japan Science and Technology Agency (to H.F.).

REFERENCES

Asif, M. 2012. Biological importance and health effect of Perilla frutescens plant. Indonesian J. Pharm. 23: 84–92.
Barak, S., Tobin, E. M., Andronis, C., Sugano, S., Green, R. M. 2000. All in good time: The Arabidopsis circadian clock. Trends Plant Sci. 5: 517–522.
Clarkson, G. J. J., Rothwell, S. D., Taylor, G. 2005. End of day
Fig. 4  Circadian time estimation at (A) 0:00, (B) 4:00, (C) 8:00, (D) 12:00, (E) 16:00, (F) 20:00, N = 260. The white and gray backgrounds indicate the times during which the light was on and off, respectively. The plots around the circles show the estimated time for each trial (approximately 250 plots for each circle). The thin vector shows the actual time, and the thick vector approximately 250 plots for each circle). The thin vector shows the trajectory of the arrowhead of R, adding a new plot of trial data. This trajectory started from one point of the circle and ended at the tip of the newest R. Panel (G) shows the locus of the error of estimated time for panel (A).
Sato-Masumoto, N., Ito, M. 2014. Two types of alcohol dehydrogenase from Perilla can form citral and perillaldehyde. Phytochemistry 104: 12–20.
Takano, H., Osakabe, N., Sanbongi, C., Yanagisawa, R., Inoue, K., Yasuda, A., Natsume, M., Baba, S., Ichihashi, E., Yoshikawa, T. 2004. Extract of Perilla frutescens enriched for rosmarinic acid, a polyphenolic phytochemical, inhibits seasonal allergic rhinoconjunctivitis in humans. Exp. Biol. Med. 229: 247–254.
Takeoka, M., Higashi, T., Honjo, M. N., Tezuka, A., Nagano, A. J., Tanigaki, Y., Fukuda, H. 2018. Estimation of the circadian phase by oscillatory analysis of the transcriptome in plants. Environ. Control Biol. 56: 67–72.
Tanigaki, Y., Higashi, T., Nagano, A. J., Honjo, M. N., Fukuda, H. 2017. Transcriptome analysis of a cultivar of green perilla (Perilla frutescens). Environ. Control Biol. 55: 77–83.
Using Genetic Similarity with Other Plants via Public Databases
Ueda, H., Chen, W., Minami, Y., Honma, K., Honma, S., Iino, M., Hashimoto, S. 2004. Molecular-timetable methods for detection of body time and rhythm disorders from single-time-point genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 101: 11227–11232.
Wang, L., Si, Y., Dedow, L. K., Shao, Y., Liu, P., Brutnell, T. P. 2011. A low-cost library construction protocol and data analysis pipeline for illumina-based strand-specific multiplex RNA-Seq. PLoS ONE. 6: e26426(1–12).