Transcriptome Analysis of a Cultivar of Green Perilla (Perilla frutescens) 
Using Genetic Similarity with Other Plants via Public Databases

Yusuke Tanigaki1, Takanobu Higashi2, Atsushi J. Nagano3,4, Mie N. Honjo1 and Hirokazu Fukuda5,6

1 Department of Mechanical Engineering, Graduate School of Engineering, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan
2 Department of Applied Life Sciences, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan
3 Faculty of Agriculture, Ryukoku University, 1-5 Yokatani, Seto, Otsu, Shiga 520-2194, Japan
4 JST CREST, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan
5 Center for Ecological Research, Kyoto University, 2-509-3 Hirano, Otsu, Shiga 520-213, Japan
6 JST PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

(Received September 1, 2016; Accepted December 29, 2016)

Improving the yield and quality of crops is imperative within agriculture. Both traits depend on gene expression and metabolic activities that are affected by environmental factors and are therefore complex and variable. Transcriptome analysis is a helpful method to understand gene expression profiles in vivo; however, most crops are in fact cultivars for which there is little genetic information. In this study, we determined gene expression profiles of Perilla frutescens var. crispa f. viridis, a cultivar of green perilla. We compared the transcriptome gene sequence of P. frutescens in leaves and roots after 7, 14, and 35 d of growth with the gene sequences of other plants in public databases. Green perilla showed the highest similarity to Mimulus guttatus (Lamiales). Genetic information for approximately 13,000 genes was evaluated, and many of these genes have been classified into the organism’s biological processes using Gene Ontology analysis. De novo-based gene expression levels and other plants-based gene expression levels were similar in 90% of genes. Results suggest that information from public databases can assist in analyzing the genetic information of cultivars. This method will be a platform for providing rapid and cost-effective options for use in commercial agriculture.

Keywords: agriculture, cultivar, Perilla frutescens, transcriptome analysis

INTRODUCTION

Ensuring stable and increasing crop yields are important problems within agriculture. In many cases, studies on the cultivation of crops have a long history; however, few studies have focused on the changes occurring in crops at the genetic level. Gene expression profiles in vivo are thought to assist in understanding plant conditions, so as to stabilize and increase yield. Recently, comprehensive analyses such as genomics, proteomics, and metabolomics have advanced the understanding of information in vivo and provided a lot of information through one-time analysis.

Transcriptome analysis by RNA-seq, an omics analysis technique, is widely used in studies of animals, plants, and insects (Scherf et al., 2000; Rifkin et al., 2003; Lister et al., 2008). Bioinformatics approaches are used to elucidate biological implications. There is abundant genetic information available for model plants such as Arabidopsis thaliana, as well as a reference sequence (RefSeq) supporting highly accurate analysis (Fiehn et al., 2001). Software applications for visualizing the analyzed data and genetic information have been built into public databases, offering the ability to assess large quantities of data. However, there is little information for cultivars. On the other hand, basic mechanisms and genes involved in plant processes, such as those involved in growth metabolism, are conserved in many species. Therefore, the use of information available in databases is very helpful in building and improving the cultivation environment of cultivars. The use of model-plant genetic analysis techniques within cultivars has a strong potential to increase yield and improve crop quality.

Although species differ, the functions of genes tend to be similar as they are evolutionarily conserved (Tanigaki et al., 2014). In plants, the amino acid or nucleotide sequences of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) are evolutionarily conserved (Tabita et al., 2007). In some cases, this similarity is low because of differences in rates of molecular evolution (Xiang et al., 2004). However, species-specific genes are difficult to analyze by expression estimation and mapping to metabolic pathways. Using omics data for species-specific genes via genetic functional analysis is time-consuming. Moreover, genes specific to a particular cultivar can facilitate species-specific de novo genetic analysis (Hong et al., 2015). Conventional farmers seeking methods to improve crop growth require data rapidly (within a single planting sea-
son) and may not be willing to pay high costs for de novo analysis, which requires a sufficient analytical environment and expert knowledge. By contrast, transcriptome analysis is a relatively simpler method employing a user-friendly analysis environment, and can be analyzed using a personal computer. Smart agriculture fusing Internet of things (IoT) technology to agriculture is rapidly developing. However, there is insufficient genetic information for cultivars. Thus, we attempted to build a platform for genetic analysis to provide rapid and cost-effective choice for use in commercial agriculture. The aim of the present study is to present the possibility of building a new platform of genetic analysis techniques suited to agriculture.

In the present study, we focused on the cultivar *Perilla frutescens* var. *crispa* f. *viridis* (green perilla), for which there is little genetic information. Perilla is a herb of the mint family and has two forms—green and red—differing primarily in the accumulation of anthocyanins (Saito et al., 1999; He et al., 2015). Green perilla is cultivated throughout Asia and has an aroma derived from perillaldehyde accumulated in the glandular trichomes (Sato-Masumoto and Ito, 2014). Green perilla is a good source of several nutrients, including calcium, vitamin B, and carotene, and is also used as a herbal medicine and health supplement (Takano et al., 2004; Takagi et al., 2005). We cultivated green perilla in a plant factory using light-emitting diode (LED) lighting and performed transcriptome analysis of samples harvested at 7, 14, and 35 d after planting, using genetic similarity and public databases.

### MATERIALS AND METHODS

**Plant material**

Green perilla (*P. frutescens*) seeds were soaked in water and incubated to develop roots for 7 d in a growth chamber. The plants were cultivated in a closed plant factory system using a Deep Flow Technique (DFT) hydroponic system (Vinolmangkang et al., 2010) and the solution was circulated using a pump (flow rate 10.0 L min⁻¹, NW20-PTN, Aichi Tokei Denki Co., Ltd., Aichi, Japan). The cultivation liquid medium used was Otsuka House No. 1 and 2 diluted with tap water (N: P:O₅: K:O: CaO: MgO = 10: 8: 27: 0: 4 and N: P:O₅: K:O: CaO: MgO = 11: 0: 23: 0, OAT Agrio Co., Ltd., Tokyo, Japan) at pH 6.0 and electrical conductivity (EC) 2.0 (following OAT Agrio standard solution formulations). The cultures were grown at 22°C, 50% relative humidity (RH), and 1,000 μmol mol⁻¹ CO₂ concentration under red, blue, green (RBG)-LED (R: G: B = approximately 120: 40: 40 μmol m⁻² s⁻¹, DP-10403-01, SHIBASAKI Inc., Saitama, Japan) light on a 12 h cycle (10:00 AM lights on, 10:00 PM lights off).

**Extraction of RNA**

All leaves (complete and seed leaves) and roots (main and lateral roots) were sampled at 7, 14, and 35 d after planting, and samples were stored at −80°C. 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, California, USA). RNA quantity was determined with a Bioanalyzer (Agilent Technologies). The RNA-seq library preparation method has been previously described (Wang et al., 2011; Nagano et al., 2015). Transcriptome sequence analyses were performed by Beijing Genomics Institute (BIG, Shenzhen, China). A HiSeq 2000 sequencer (single end, 50 bp; Illumina) was used to read files. The construction of contigs was performed by BGI. A Trinity assembler (2011–11–26 version) was used for de novo short read assembly (Max contig length: 76,620 bp; Min contig length: 201 bp). These de novo data were published on our website (http://www.me.osakafu-u.ac.jp/bioproduction/news/denovo).

**Transcriptome analysis**

Quality control was performed using the FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Clean reads were analyzed for mapping against RNA-seq data from 46 plants in the Phytozone database (http://www.phytozone.net/) and de novo contigs of *P. frutescens*. Data from these plants are designated *Arabidopsis thaliana* (TAIR database ID). The mapping data were quantified using the RNA-seq by Expectation Maximization (RSEM) software (http://deweylab.biostat.wisc.edu/rsem/) with Bowtie2 software (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) (Li and Dewey, 2011). RSEM software calculates values based on the lengths of target contigs and numbers of mismatches, and isoform- or gene-level estimates were obtained using expectation-maximization (EM) algorithms and read counts for each contig. Accordingly, accurate estimates of expression data were provided, even for data that were mapped to multiple sites. The mapping parameter in Bowtie2 was set to default. Based on initial analysis, the mismatch rate was set to 0.2. The expression data were normalized to the RPKM (Reads Per Kilobase of exon per Million mapped reads) value. Expression data were registered with the DNA Data Bank of Japan (DDBJ) database (http://trace.ddbj.nig.ac.jp/DRASeach, Accession numbers: DRA004233, DRA004232). Histogram analyses with RPKM values of < 0.1 were set to 0.1, and RPKM values were log, transformed and analyzed with the ggplot2, reshape, and reshape2 packages of the R software.

**Gene ontology analysis**

The normalized expression data which have the *A. thaliana* ID (TAIR database ID) were categorized based on Gene Ontology (GO) terms using Gene Ontology Consortium (https://geneontology.org/).

**Expression level comparison analysis**

The normalized expression data used the average value between all cultivation periods and leaves and roots. The two expression values mapped to other plants and de novo contigs of *P. frutescens* have been adjusted (average 1). Basic Local Alignment Search Tool (BLAST) software was used to link the expression values mapped to other plants and de novo contigs of *P. frutescens* (BLAST+, ver.2.4.0).
RESULTS

Obtaining information from genes using similarity

Transcriptome analyses of gene expression in leaves and roots were obtained three times (7, 14, and 35 d) after planting; therefore the analysis result does not depend on the cultivation period. As a result, we estimated the genetic information of 12,807 genes of leaves and roots. All clean reads showed quality scores >Q30. The transcriptome data were each mapped to a reference sequence of 46 plants, including phytoplankton, in the Joint Genome Institute (JGI) database. The maximum map ratio of green perilla transcriptome data for the reference sequences in 46 plants was 58% of Mimulus guttatus (mismatch rate 0.2 in Bowtie2 mapping), and the minimum map ratio was 0.07% of Ostreococcus lucimarinus (Fig. 1). Compared with model plants, the map rates to A. thaliana and Oryza sativa were 6.69% and 5.17%, respectively. Green perilla is closely associated to M. guttatus; thus, its expression data were based on M. guttatus. For expression data that could not be obtained from the result of mapping to M. guttatus, non-overlapping expression data obtained from the result of mapping the whole sequence to 45 plants was used. We excluded genes that were not expressed in more than three of the six samples (three times × two sites (leaves and roots)). As a result, there were 36 expression-level data from 36 plants other than M. guttatus.

Genes with GO annotation

We categorized all genes based on their GO annotation in molecular function and biological processes. Using the PANTHER classification system, we showed that 12,453 of 12,807 genes have GO annotation (GO Ontology database ver.2016–05–20). We expressed the results as Fold enrichment, a value obtained by dividing the expected value (the number of the subject genes of analysis expected for each category based on the reference list) by the number of genes classified into each category. Molecular function was significantly enriched in 17 parent categories (excluding Unclassified) (P<0.05) (Fig. 2). The following were the major large parent categories of molecular function: metallopeptidase activity (Fold enrichment = 2.15, P = 1.03E-2), microtubule binding (Fold enrichment = 2.05, P = 2.60E-3), and ATP-dependent helicase activity (Fold enrichment = 2.05, P = 6.47E-3). In contrast, the following were the major small parent categories of molecular function categories: protein-disulfide reductase activity (Fold enrichment < 0.2, P = 1.69E-10) and ATP-dependent helicase activity (Fold enrichment = 0.34, P = 4.47E-4). Figure 3 shows representative biological process categories. Significantly enriched biological process contained 76 parent categories (excluding the unclassified categories) (P<0.05). The major large parent categories included DNA-dependent DNA replication (Fold enrichment = 2.40, P = 0.002), response to carbohydrate (Fold enrichment = 2.33, P = 3.00E-4), and Golgi vesicle transport (Fold enrichment = 2.32, P = 3.00E-4). The aromatic compound biosynthetic process formed a small cluster (Fold enrichment = 1.19, P = 0.013). Killing cells of other organisms was the smallest cluster (Fold enrichment < 0.20, P = 9.00E-16). Figure 4 shows representative cellular component parent categories. Significantly enriched biological process contained 28 parent categories (excluding the unclassified categories) (P<0.05). The major large parent categories included nuclear pore (Fold enrichment = 2.49, P = 2.47E-02) and phragmoplast (Fold enrichment = 2.37, P = 1.11E-05). The extracellular region formed a small cluster (Fold enrichment = 0.70, P = 6.74E-19).

Expression level of the comparison by the mapping to the de-novo assembly and RNA-seq data of other plants

We obtained expression levels and annotation by transcriptome data of green perilla in sequence data of all plant species in the existing database. In addition to this acquired information, we believed that green perilla-specific genes exist; thus, we compared gene expression level with reference to other plants (ROP) and gene expression level referring to de novo. The linkage to the ROP...
contigs and de novo contigs was performed by BLAST+. The expression level was evaluated using the difference between the log2 value of the ROP and de novo contigs. As a result, 11,895 genes had a difference between 1 and 1 in leaves, and these genes accounted for approximately 93% of the total genes (Fig. 5A). Two genes had three or more differences; 39 genes had one or more and two or less differences; and 871 genes had zero or more and one or less differences. In root samples, 11,616 genes had a difference of between 1 and 1, and these genes accounted for approximately 91% of the total genes (Fig. 5B). One gene had three or more differences; 46 genes had one or more and two or less differences; 1,144 genes had zero or more and one or less difference.

**DISCUSSION**

In the present study, we performed transcriptome analysis of the cultivar green perilla (P. frutescens) based on genetic similarity using a public database. We estimated information for 12,807 genes possessing a link to the model plant (A. thaliana) information using public databases. By comparing the transcriptome data of green perilla with RNA-seq data of ROP and de-novo contigs at the expression level, the difference of 90% or more of the genes in leaves and roots was one or less (log2 value).

Transcriptome data of other varieties have been mapped to many genes of M. guttatus. M. guttatus falls within the order Lamiales (the same order as green perilla) (The Angiosperm Phylogeny Group, 2016). Similarly, Solanum lycopersicum (Solanales) and Vitis vinifera (Vitales) have relatively high map rates and are closely related species. By contrast, the map rate of Oryza sativa (Poales), which showed a remote relationship, was low. These results suggest that green perilla and M. guttatus have similar genetic backgrounds.
CULTIVAR TRANSCRIPTOME ANALYSIS

Analysis of GO biological processes indicate that many of the genes, such as those governing DNA replication and cell division, were classified into categories related to the basic life activities of the organism. By contrast, there were few genes classified in killing of cells of other organisms. Genes that induce cell death have been classified in this category. Additionally, categorized genes in hormone-mediated signaling pathways were small compared with other genes. Hormone-mediated signaling is often observed in disease response. In addition, pathogens which infect plants are often species-specific; thus, a plant may possess a response mechanism that corresponds to the pathogen. Unlike green perilla, *M. guttatus* is adapted to live in riparian and wetland environments (Peterson et al., 2016). Given these differences, green perilla and *M. guttatus* could possess different response mechanisms against disease and wounding.

More than 90% of the genes in both the leaves and roots had a similar expression level. Many of the genes have been classified into the basic processes of the organism using GO analysis. Similar to the results of the mapping analysis, these results suggest that green perilla and *M. guttatus* have similar genetic backgrounds. On the other hand, 912 genes in leaves and 1,191 genes in roots had a difference of more than twofold. These differences could have been influenced by the quality of transcriptome data and analysis methods, such as mapping. However, the technique and materials described here will not apply to the level of all genes.

Methods (e.g., BLAST) to estimate gene function using nucleic acid and amino acid sequences are now becoming common. Methods such as genome-wide annotation to consider the similarity of the sequence in the global analysis cyclopedically are also established (Promponas et al., 2015). During transcriptome analysis, a first reference sequence is built from *de-novo* assembly data to obtain an accurate expression level (Haas et al., 2013). The analysis then estimates gene functions using similarity of the sequence. These procedures to obtain accurate data are easily applied to model plants, but difficult to apply to highly
diverse cultivars. By mapping directly to reference sequences of other plants in public databases, our study method expedited the analysis. Although our method may result in lower accuracy compared with the procedures commonly applied to model plants, the method can obtain over 90% of the expression level data of the estimated genes. Because the plants possess species-specific sequences, species-specific mechanisms, and orthologous and paralogous genes, we did not expect to obtain all of the information using this novel analysis. Although it is currently difficult to analyze unique genes of green perilla, there is a possibility that this method can be used for analysis if specific elucidation of the gene is advanced. Therefore, it is difficult to analyze the unique mechanism of green perilla using this method. On the other hand, as shown in Fig. 3, this method enabled the estimation of the genetic information relating to the basic metabolism of plants such as photosynthesis or development. Consequently, we believe that it will become possible to analyze the combined basic growth rate of plants corresponding to the outside environment. The present study demonstrates a successful method for using the omics platform to analyze cultivars, and with new and updated genetic data of various plants, we anticipate continual improvements in the technique that will further its application in commercial agriculture.

ACKNOWLEDGEMENTS

We are grateful to Hiroshi Kudoh for RNAseq analysis. The present study was supported by the Joint Usage/Research program of Center for Ecological Research, Kyoto University. The funders had roles in data analysis. This work was partially supported by the Canon Foundation and PRESTO, JST.
REFERENCES

Fiehn, O., Kopka, J., Dormann, P., Altman, T., Trethewey, R., Willmitzer, L. 2001. Metabolite profiling for plant functional genomics. Nat. Biotechnol. 18: 1157–1161.

The Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linnean Soc. 181: 1–20.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N., Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq reference generation and analysis with trinity. Nat. Protoc. 8: 1494–1512.

He, Y. K., Yao, Y. Y., Chang, Y. N. 2015. Characterization of anthocyanins in Perilla frutescens var. acuta extract by advanced UPLC-ESI-IT-TOF-MSn method and their anticancer bioactivity. Molecules 20: 9155–9169.

Hong, Y., Tang, X., Huang, H., Zhang, Y., Dai, S. 2015. Transcriptomic analyses reveal species-specific light-induced anthocyanin biosynthesis in chrysanthemum. BMC Genomics 16: 202.

Li, B., Dewey, C. N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323.

Lister, R., O’Malley, R., Tonti-Filippini, J., Gregory, B., Berry, C., Millar, A., Ecker, J. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.

Nagano, A. J., Honjo, M. N., Mihara, M., Sato, M., Kudoh, H. 2015. Detection of plant viruses in natural environments by using RNA-Seq. Methods Mol. Biol. 1236: 89–98.

Peterson, M. L., Kay, K. M., Angert, A. L. 2016. The scale of local adaptation in Mimulus guttatus: comparing life history races, ecotypes, and populations. New Phytol. 211: 345–356.

Pronopas, V. J., Iliopoulos, I., Ouzounis, C. A. 2015. Annotation inconsistencies beyond sequence similarity-based function prediction: phylogeny and genome structure. Stand. Genomic Sci. 10: 108.

Rifkin, S. A., Kim, J., White, K. P. 2003. Evolution of gene expression in the Drosophila melanogaster subgroup. Nat. Genet. 33: 138–144.

Saito, K., Kobayashi, M., Gong, Z., Tanaka, Y., Yamazaki, M. 1999. Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red form of Perilla frutescens. Plant J. 17: 181–189.

Sato-Masumoto, N., Ito, M. 2014. Two types of alcohol dehydrogenase from Perilla can form citral and perillaldehyde. Phytochemistry 104: 12–20.

Scherf, U., Ross, D., Waltham, M., Smith, L., Lee, J., Tanabe, L., Kohn, R., Reinhold, W., Myers, T., Andrews, D., Scudiero, D., Eisen, M., Sausville, E., Pommier, Y., Botstein, D., Brown, P., Weinstein, J. 2000. A gene expression database for the molecular pharmacology of cancer. Nat. Genet. 24: 236–244.

Tabita, F. R., Hanson, T. E., Li, H., Satagopan, S., Singh, J., Chan, S. 2007. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. Microbiol. Mol. Biol. Rev. 71: 576–599.

Takeda, S., Goto, H., Shimada, Y., Nakagomi, K., Sadakane, Y., Hatanaka, Y., Terasawa, K. 2005. Vasodilative effect of perillaldehyde on isolated rat aorta. Phytomedicine 12: 333–337.

Takano, H., Osakabe, N., Sanbongi, C., Yanagisawa, R., Inoue, K., Yasuda, A., Natsume, M., Baba, S., Ichishi, E., Yoshikawa, T. 2004. Extract of Perilla frutescens enriched for rosmarinic acid, a polyphenolic phytochemical, inhibits seasonal allergic rhinoconjunctivitis in humans. Exp. Biol. Med. (Maywood) 229: 247–254.

Taniyaki, Y., Ito, K., Obuchi, Y., Kosaka, A., Yamato, K., Okanami, M., Lehtonen, M., Valkonen, J., Aikta, M. 2014. Physcomitrella patens has Kinase-LRR R gene homologs and interacting proteins. PLoS One 9: e95118.

Vimolmangkang, S., Sitthithaworn, W., Vannavanich, D., Keattikunpairoj, S., Chittasupho, C. 2010. Productivity and quality of volatile oil extracted from Mentha spicata and M. arvensis var. piperascens grown by a hydroponic system using the deep flow technique. J. Nat. Med. 64: 31–35.

Wang, L., Si, Y., Dedow, L. K., Shao, Y., Liu, P., Brutnell, T. P. 2015. A low-cost library construction protocol and data analysis pipeline for illumina-based strand-specific multiplex RNA-Seq. PLoS One 10: e26426.

Xiang, Q., Zhang, W., Ricklefs, R., Qian, H., Chen, Z., Wen, J., Li, J. 2004. Regional differences in rates of plant speciation and molecular evolution: A comparison between eastern Asia and eastern North America. Evolution 58: 2175–2184.