J. Plant Biol. (2014) 57:312-320
DOI 10.1007/s12374-014-0263-2

ORIGINAL ARTICLE

Analysis of Expressed Sequence Tags from a Normalized cDNA Library of Perilla (*Perilla frutescens*)

Sang-Choon Lee[1,†], Junki Lee[1,†], Nam-Hoon Kim[1], Jee Young Park[1], Hyun Uk Kim[2], Hyun Oh Lee[1], Beom-Soon Choi[3], Kyunghee Kim[1,3], Jonghoon Lee[1], Myoung-Hee Lee[4], Kalaiselvi Senthil[5], Xu Jiang[6], Chao Sun[6], Tae-Ho Kim[7] and Tae-Jin Yang[1,*]

[1]*Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute for Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea*
[2]*Department of Agricultural Biotechnology, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-707, Korea*
[3]*Phyzen Genomics Institute, 501-1, Gwanak Century Tower, Kwanak-gu, Seoul 151-836, Korea*
[4]*Department of Functional Crop, National Institute of Crop Science, Rural Development Administration, Miryang 627-803, Korea*
[5]*Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam University for Women, Coimbatore 641043, India*
[6]*Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China*
[7]*Genomics Division, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-707, Korea*

Received: June 11, 2014 / Accepted: August 3, 2014

Abstract Perilla (*Perilla frutescens*), an herbal plant belonging to the Lamiaceae family, has long been cultivated in Asia. Perilla is notable as an aroma-rich leaf vegetable and as the oilseed crop richest in omega-3 fatty acids. However, molecular analysis of this herbal plant is lacking due to insufficient genetic resources. Here, we constructed a normalized cDNA library from whole young perilla plants and analyzed expressed sequence tags (ESTs). A total of 4,582 uniESTs were generated from analysis of 5,435 ESTs. Among these, 307 uniESTs (6.7%) were identified as unique in perilla and 3,625 uniESTs were assigned at least one GO term through similarity searches in public databases. The most frequent GO terms were related to abiotic and biotic stress responses. In addition, we identified 141 uniESTs involved in lipid metabolism, of which four genes encoded fatty acid desaturases. We found one new candidate omega-3 fatty acid desaturase gene, in addition to the four that were previously reported. This analysis of uniESTs from perilla provides a valuable genetic resource to elucidate the molecular underpinnings of lipid metabolism and for molecular breeding of perilla species.

Keywords: cDNA library, Expressed sequence tag, Fatty acid desaturase, Perilla

Introduction

Perilla (*Perilla frutescens* (L.) Britton) is an annual herbal plant species that belongs to the Lamiaceae family and encompasses various natural varieties. It is cultivated mainly in Asian countries including Korea, China, and Japan, where the plant has been used as an ingredient in culinary dishes and as a traditional herbal medicine (Gwak et al. 2003; Yu et al. 2004). Two main varieties are cultivated and used in Korea and Japan; one variety, *P. frutescens* var. frutescens, commonly named "Deul-ggae" in Korea and "Egoma" in Japan, is cultivated as an oilseed crop. The other variety, *P. frutescens* var. crispa, called "Cha-jo-ki" in Korea and "Shiso" in Japan, is cultivated as a Chinese medicine or spicy vegetable crop (Makino 1961; Nitta and Ohnishi 1999; Lee and Ohnishi 2001; Nitta 2001; Lee et al. 2002; Lee and Ohnishi 2003). In addition to these varieties, *P. frutescens* var. acuta (purple perilla) has also been cultivated as a medicinal herb to take advantage of its high antioxidant activity due to high levels of phenolic compounds (Meng et al. 2009; Ha et al. 2012; Jun et al. 2014). Perilla seeds contain large amounts of unsaturated fatty acids, including 14-23% oleic acid (18:1), 11-16% linoleic acid (18:2) and

[†]Equal contribution

*Corresponding author; Tae-Jin Yang
Tel : +82-2-880-4547
E-mail : tjyang@snu.ac.kr

54-64% linolenic acid (18:3; Asif 2011). This high unsaturated fatty acid content confers various benefits for human health (Onogi et al. 1996; Sadiet al. 1996; Deng et al. 2007; Asif 2011; Igarashi and Miyazaki 2013). Perilla seed oil is used mainly as an ingredient for cooking and for industrial uses in paint, varnish and ink manufacturing. In addition, perilla seed cakes are used for animal and bird feeds (Asif 2011).

Most perilla studies to date have been focused on characterization of metabolites and their biological activities, although the genes involved in the biosynthesis of anthocyanins, flavones, and monoterpenoids in perilla have been described (Yonekura-Sakakibara et al. 2000; Kitada et al. 2001; Sompornpailin et al. 2002; Yamazaki et al. 2002; Yamazaki and Saito 2006; Mau et al. 2010; Yamazaki and Saito 2011). Whereas the fatty acid composition in developing seed of perilla has been well studied (Ichihara and Suda 2003), molecular analysis of lipid metabolism has not yet been intensively performed in this plant, probably due to a dearth of genetic resources. Only a few genes encoding enzymes involved in fatty acid biosynthesis have been characterized in perilla (Chung et al. 1999; Hwang et al. 2000; Hwang and Hwang 2000).

Only about 100 expressed sequence tags (ESTs) generated from perilla have been deposited in GenBank, while more than 10,000 ESTs have been deposited from other species in the Lamiaceae family such as Chinese salvia (*Salvia miltiorrhiza*) and sweet basil (*Ocimum basilicum*). Furthermore, comparison with another oilseed crop, sesame (*Sesamum indicum*), for which more than 40,000 ESTs have been reported and genome sequencing has been completed (Ke et al. 2011; Wang et al. 2014), underscores the fact that molecular genetic information is currently lacking for perilla. At present, whole-genome and transcriptome sequencing using next generation sequencing (NGS) technology has become one of the best approaches to acquire large quantities of genetic information. However, *de novo* sequencing of a cultivated crop without a reference genome remains challenging in terms of assembly and annotation of expressed genes (Varshney et al. 2009). As an alternative approach, we performed EST analysis, which is useful as a less expensive method to collect information about expressed genes. ESTs are short sub-sequences of cDNA sequences representing expressed genes, and have been widely used to provide global information about genes expressed in a given tissue or organism (Rudd 2003; Zhang et al. 2011).

Considering the importance of perilla with respect to nutritional and industrial applications, molecular study of this plant species is essential to understand and enhance its relevant traits, such as its high content of unsaturated fatty acids. To this end, we constructed a cDNA library from whole tissues of perilla plants and analyzed more than 5,400 ESTs to provide large-scale molecular genetic resources and to facilitate genetic study of this species.

Results

Construction of cDNA Library

A normalized cDNA library was constructed using total RNA from three-week-old plants of perilla cultivar Youngho Deul-ggae. The average insert size of the cDNA library was determined by colony PCR from 96 randomly selected cDNA clones. As shown in Table 1 and Fig. S1, most of inserts were longer than 0.5 kb, with an average insert size of 1.l kb. Inserts longer than 0.8 kb were found in ~ 90% of the clones. The recombination rate of the library was 97.9%.

Generation and Assembly of ESTs

Nucleotide sequences at the 5'-end of the inserts were obtained by sequencing in 5,760 clones. After quality filtering and removing vector contamination, a total of 4,241,946 bp from 5,435 ESTs, with an average of 780 bp per sequence, was obtained (Table 2). This result indicated that the total sequencing success rate was 94.4%. The 5,435

Table 1. Distribution of insert size in the perilla normalized cDNA library

Insert Length (kb)[1)]	Number of clones[1)]	Frequency (%)
> 2.0	2	2.1
1.5-2.0	2	2.1
1.0-1.5	30	31.2
0.5-1.0	60	62.5
< 0.5	0	0.0
no insert	2	2.1
Total	96	100

[1)]Insert size distribution in 96 randomly selected clones was determined by colony PCR as shown in Fig. S1.

Table 2. Summary of filtered ESTs and uniESTs for perilla

Filtered ESTs	
Number of ESTs	5,435
Total length	4,241,946 bp
Average length	780 bp
UniESTs	
Number of contigs	659
Total length	596,052 bp
Average length	904 bp
Number of singletons	3,923
Total length	3,059,637 bp
Average length	780 bp
Total uniESTs	4,582

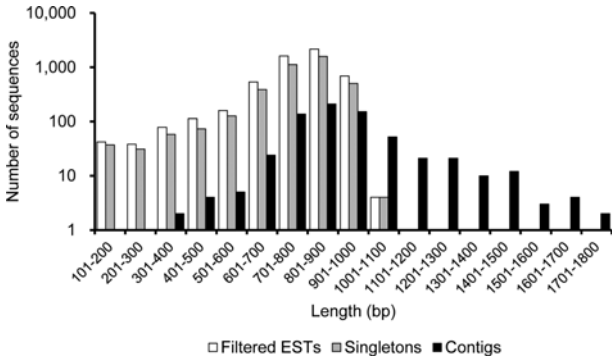

Fig. 1. Length distribution of perilla ESTs. The number of filtered ESTs, singletons and contigs within each length category on the x-axis is shown on the y-axis.

filtered ESTs were deposited in GenBank under accession numbers JZ578419 to JZ583853.

The filtered ESTs were assembled into 4,582 uniESTs consisting of 659 contigs and 3,923 singletons with 904 bp and 780 bp average length, respectively (Table 2, Table S1). While both filtered ESTs and singletons ranged from 101 to 1,019 bp in length, assembled contigs ranged from 320 to 1,780 bp (Fig. 1). Between 2 and 6 ESTs were assembled into each contig, 76.8% of which were generated with 2 ESTs, 18.7% with 3 ESTs, and the remaining 4.5% with 4 to 6 ESTs. This result suggested that the cDNA library was well normalized for generating uniESTs. When coding sequences of uniESTs were investigated, 1,219 (26.6%) of uniESTs contained a full-length open reading frame (ORF) with start and stop codons, while 1,859 (40.6%) and 491 (10.7%) had 5' or 3' partial ORFs, respectively. The remaining 1,013 (22.1%) had partial ORFs at the both 5' and 3' ends or sequence too short to predict ORFs.

UniEST Annotation

Through homology searches against the Genbank non-redundant protein (Nr) database using BLASTX, we identified 2,471

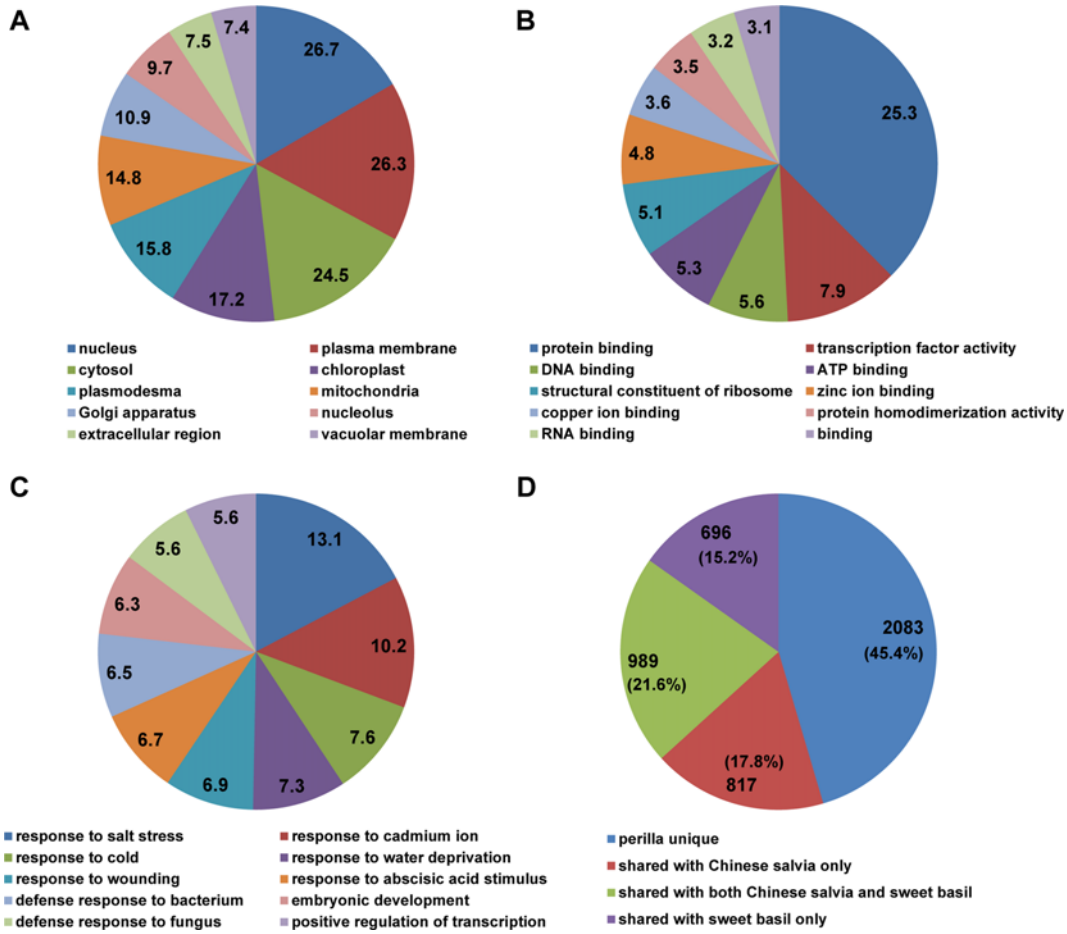

Fig. 2. GO analysis of perilla uniESTs and comparison with ESTs of other Lamiaceae species. Out of 4,582 total uniESTs, 3,625 were assigned to GO terms by the Blast2GO program using default parameters. Top 10 GO terms that large number of uniESTs were assigned were selected and their proportion to total assigned uniEST number was calculated for three GO categories: cellular component (A), molecular function (B), and biological process (C). (D) Perilla uniESTs were compared with those of Chinese salvia and sweet basil retrieved from GenBank using BLASTN with E-value cut-off of $1E^{-5}$.

uniESTs (53.9% of the total uniESTs) with similarity to known genes and 1,804 uniESTs (39.3%) with similarity to genes of unknown function at the amino acid level. The remaining 307 uniESTs (6.7%) did not show similarity to protein sequences deposited in the Nr database using an E-value cut-off of $1E^{-5}$. When the uniESTs were searched against the TAIR 10 database, 4,140 uniESTs (90.4%) showed similarity to Arabidopsis genes, whereas the remaining 442 uniESTs (9.6%) were not similar to any genes in the database. Among uniESTs, 9 (0.2%) and 109 (2.4%) encoded photosystem-related proteins and ribosomal proteins, respectively, confirming the successfully normalized coverage of the cDNA library.

In Gene Ontology (GO) analysis, a total of 3,625 uniESTs were assigned to at least one GO term, of which 3,423, 2,864, and 3,170 sequences were assigned terms for GO categories cellular component, molecular function, and biological process, respectively. The 10 GO terms to which the largest numbers of uniESTs were assigned were further investigated (Figs. 2A, 2B, and 2C). Among the top 10 GO terms, nucleus, plasma membrane, and cytosol for cellular component, and protein binding for molecular function were accumulated compared to the other terms. Interestingly, the most abundant GO terms for biological process were related to abiotic and biotic stress responses, which accounted for 70.2% of the assigned uniESTs (Fig. 2C).

The perilla uniESTs were compared with those of two other Lamiaceae species, Chinese salvia and sweet basil. For this analysis, 10,288 (4,595,480 bp) and 23,260 (14,847,174 bp) ESTs were retrieved from GenBank for Chinese salvia and sweet basil, respectively, and assembled to generate 7,191 and 11,729 uniESTs for the two species, respectively. When compared at the nucleotide level using the BLAST program, 817 (17.8%), 695 (15.2%), and 989 (21.6%) of the total perilla uniESTs were homologous to those in Chinese salvia, sweet basil, and both species, respectively (Fig. 2D). The remaining 2,081 (45.4%) were unique to perilla. These results suggest that the perilla uniESTs we generated contain a large number of perilla-specific genes.

UniESTs Homologous to Major Regulatory Genes

By homology searches against Nr and TAIR databases, we could identified 300 uniESTs showing similarity to major regulatory genes, which consisted of 106, 84, 83, and 27 uniESTs related to transcription factors, protein kinase, transporters, and E3 ubiquitin ligases, respectively (Table 3, Table S2). Among those 106 uniESTs related to transcription factors, 23 (21.7%) encoded AP2/EREBP domain containing proteins, 11 (10.4%) encoded basic helix-loop-helix (bHLH) proteins, and 10 (9.4%) encoded MYB proteins. Among those 84 uniESTs related to protein kinases, large number of uniESTs encoded receptor-like kinases and serine/threonine protein kinases, whose number were 32 (38.1%) and 13 (15.5%), respectively. In addition, 133 of the 300 uniESTs were only found in perilla, when compared with Chinese salvia and sweet basil ESTs (Table 3).

Table 3. Number of uniESTs homologous to major regulatory genes

Gene group	No. of uniESTs (%[1)])	No. of perilla unique uniESTs (%[1)])
Transcription factor	106 (2.3)	48 (1.0)
Protein kinase	84 (1.8)	33 (0.7)
Transporter	83 (1.8)	42 (0.9)
E3 ubiquitin ligase	27 (0.6)	10 (0.2)

[1)]Percentage to 4,582 perilla uniESTs in this study

Table 4. Number of perilla uniESTs involved in lipid metabolism

Pathway in lipid metabolism	No. of uniESTs (%)
Fatty acid elongation & wax biosynthesis	31 (22.0)
Oxylipin metabolism	18 (12.8)
Triacylglycerol & fatty acid degradation	12 (8.5)
Phospholipid signaling	11 (7.8)
Sphingolipid biosynthesis	10 (7.1)
Triacylglycerol biosynthesis	10 (7.1)
Mitochondrial fatty acid & lipoic acid synthesis	9 (6.4)
Prokaryotic galactolipid, sulfolipid, & phospholipid synthesis	9 (6.4)
Fatty acid synthesis	7 (5.0)
Suberin synthesis & transport	7 (5.0)
Other	17 (12.1)
Total	141 (100.0)

UniESTs Involved in Lipid Metabolism

Among the 4,582 uniESTs, 141 (3.1%) were homologous to Arabidopsis genes involved in lipid metabolism. Of those 141 uniESTs, 31 (22.0%) were related to fatty acid elongation and wax biosynthesis, 18 (12.8%) were related to oxylipin metabolism and 12 (8.5%) to degradation of triacylglycerol and fatty acids (Table 4, Table S3).

Four of the uniESTs, namely Contig445, and three singletons (IDs EPP167KIAA11S000071, EPP167KIAA11S000807, and EPP167KIAA11S00S4973; hereafter, the "EPP167KIAA-11S00" in the singleton and EST IDs is shortened to "S" and "E", respectively, for simplicity), showed high similarity to known fatty acid desaturases (FADs) at the amino acid level (Figs. S2-S9). For further investigation of these uniESTs, full nucleotide sequences of the relevant cDNA clones were determined. The cDNA clones for all four uniESTs had incomplete 5' sequences. The cDNA sequence for S0807 was 814 bp, with a deduced protein showing 84% similarity

to Arabidopsis FAD5 (AT3G15850), which is responsible for the biosynthesis of 16:1 fatty acids from 16:0 in galactolipids and sulpholipids in the chloroplast (Figs. S2 and S3). The cDNA sequence for S0071 was 1,176 bp, and encoded a protein with 81% similarity to Arabidopsis FAD2 (AT3G12120), an endoplasmic reticulum-localized omega-6 fatty acid desaturase for biosynthesis of 18:2 fatty acids in glycerolipids (Figs. S4 and S5). The cDNA sequence for S4973 was 1,460 bp, and the deduced protein showed 80% similarity to Arabidopsis FAD6 (AT4G30950), an omega-6 fatty acid desaturase for biosynthesis of 16:2 and 18:2 fatty acids from galactolipids and sulpholipids in the chloroplast (Figs. S6 and S7). Lastly, Contig445 was assembled from ESTs E3999 and E4399 (GenBank accession nos. JZ582185 and JZ582566, respectively), the cDNA clones for which had identical sequences of 1,405 bp (Fig. S8). The deduced protein encoded by this cDNA sequence showed 73%, 72%, and 67% similarity to chloroplast-localized omega-3 fatty acid desaturases encoded by Arabidopsis genes *FAD8* (AT5G05580), *FAD7* (AT3G11170) and *FAD3* (AT2G29980), respectively (Fig. S9). In addition, it shared 64% to 84% similarity with genes annotated as omega-3 fatty acid desaturases in sesame (AAA70334) and perilla [AGJ70389, AAB39387, AAD15744 (Chung et al. 1999), and AGT37060; Fig. 3, Fig. S9]. All four uniESTs could be assigned to the unsaturated fatty acid biosynthesis pathway for chloroplast and ER membrane lipids (Fig. 4). Furthermore, no identical perilla sequences were found in public databases for any of the four cDNA sequences. These four sequences encoding FADs thus provide useful information for characterization of unsaturated fatty acid biosynthesis in perilla.

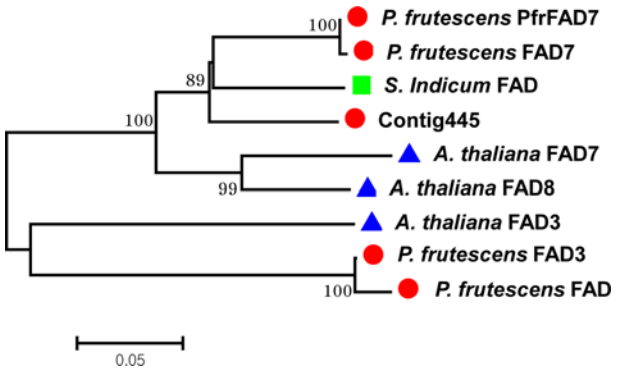

Fig. 3. Phylogenetic analysis of perilla Contig445 and omega-3 fatty acid desaturase proteins. The deduced protein sequence encoded by the cDNA sequence of Contig445 was aligned with those of reported omega-3 fatty acid desaturase genes using ClustalW and the well-aligned region of 350 amino acid residues was extracted from each gene for phylogenetic analysis. The phylogenetic tree was generated using Poisson correction and the neighbor-joining (NJ) method in MEGA5. Bootstrap percentages calculated for 1000 replicates are shown on the branches; values less than 50% are omitted. *Perilla frutescens* FAD7 (AGJ70389), PfrFAD7 (AAB39387), FAD3 (AAD15744; Chung et al. 1999) and FAD (AGT37060); *S. indicum* FAD (AAA70334); *A. thaliana* FAD3 (AT2G29980), FAD7 (AT3G11170), and FAD8 (AT5G05580).

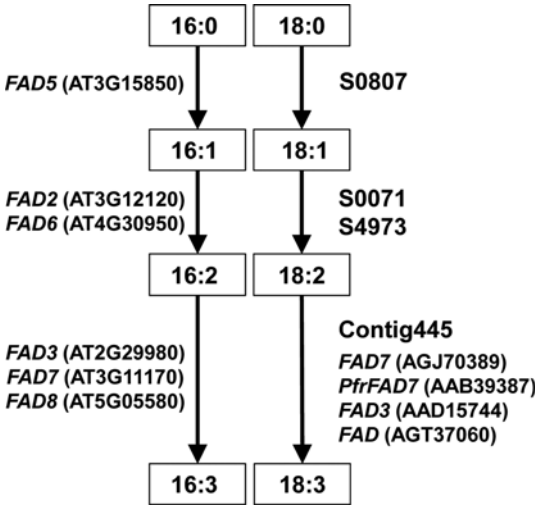

Fig. 4. Perilla uniESTs assigned to the unsaturated fatty acid biosynthesis pathway. Four uniESTs encoding fatty acid desaturases (FADs) were identified based on similarity to Arabidopsis genes involved in lipid metabolism. Through searches in TAIR and KEGG databases, four perilla FAD homologues, S0807 (GenBank accession no. KJ775813), S0071 (KJ781205), S4973 (KJ775814), and Contig445 (KJ775812), could be assigned to a predicted unsaturated fatty acid biosynthesis pathway. Reported omega-3 fatty acid desaturase genes in *P. frutescens, FAD7* (AGJ70389), *PfrFAD7* (AAB39387), *FAD3* (AAD15744; Chung et al. 1999) and *FAD* (AGT37060), are also shown. Arabidopsis counterpart genes are indicated on the left of the pathway.

Among the 141 uniESTs related to lipid metabolism, three encoded putative transcription factors. One uniEST, S5466, showed 59% similarity at the amino acid level to Arabidopsis transcription factor MYB96 (AT5G62470), which regulates cuticle wax biosynthesis (Seo et al. 2011). The other two, S2508 and S3437, showed more than 40% similarity at the amino acid level to Arabidopsis VP1/ABI3 type transcription factor HSI2/VAL1 (AT2G30470), involved in triacylglycerol biosynthesis (Li-Beisson et al. 2013).

Discussion

Genetic Resources for Perilla

Despite perilla having been cultivated and used widely as an ingredient and as a traditional herbal medicine in Asian countries, there are limited genomic and genetic resources for perilla as well as for other plants of the Lamiaceae family. The *Perilla* genus has chromosomes of 2n = 40 (Diao et al. 2009). Using the 'Youngho Deul-ggae' cultivar analyzed in this study, we estimated the haploid genome equivalent of *P. frutescence* to be approximately 1.4 to 1.5

Gbp by measuring and calculating the C value using flow cytometry compared to two known reference genomes (Fig. S10).

In this work, we constructed a cDNA library and analyzed a large number of ESTs to produce genetic resources and encourage further molecular study of perilla. The cDNA library contained an average insert size of 1.1 kb with few ESTs per contig assembly and low numbers of uniESTs for photosystem and ribosomal proteins, demonstrating that the library was well normalized. Finally, 4,582 uniEST sequences were generated through analysis of 5,435 EST sequences; this represents a significant number of uniESTs compared with the number of perilla genes in deposited in GenBank before this study. Accordingly, the cDNA library and uniEST sequences generated in this study provide important information about the transcriptome of perilla, although the data generated do not provide sufficient information to study the entire transcriptome of perilla. Further transcriptome analysis should be performed with various samples including developing seeds.

Perilla-unique Genes

Through annotation and comparative analysis, 307 uniESTs were found not to show similarity to reported protein sequences, implying that these uniEST sequences might be unique to perilla. In addition, 2,081 sequences including 133 encoding major regulatory genes had no similarity to reported ESTs of two other Lamiaceae species, Chinese salvia and sweet basil (Fig. 2D and Table 3). Whereas these 2,081 sequences are likely not all unique to perilla because the ESTs used for comparison do not represent full coverage of expressed genes of Chinese salvia and sweet basil, some of the putative unique genes should be useful for study of the distinctive traits of perilla that are absent in other Lamiaceae species.

A large number of the perilla uniESTs were related to responses to abiotic and biotic stresses according to GO analysis for the biological process category (Fig. 2C). Similarly, many abundant uniESTs encoded stress-responsive proteins, although highly abundant ESTs were normalized in the cDNA library. This indicates that many genes were highly expressed in response to environmental stress in the perilla plants used in this study. These genes might play protective roles against cellular damages caused by environment stresses.

Genes Related to Lipid Metabolism

As for other oilseed crops, understanding fatty acid biosynthesis is important in perilla, but molecular study of this process has not been actively performed in this species. We identified 141 uniESTs that might be involved in lipid metabolism (Table 4), of which four encode FADs probably involved in unsaturated fatty acid biosynthesis (Fig. 4). Among these four sequences, Contig445 appears to encode a chloroplast omega-3 fatty acid desaturase different from ones previously reported in perilla (Fig. 3). Omega-3 fatty acid desaturase produces hexadecatrienoic acid (16:3) and linolenic acid (18:3) from 16:2 fatty acid and linoleic acid (18:2), respectively in the chloroplast and endoplasmic reticulum (Ohlrogge and Browse 1995; Li-Beisson et al. 2013). Furthermore, this enzyme plays an important role in tolerance against abiotic stresses in many plants (Los and Murata 1998; Upchurch 2008). The gene corresponding to Contig445 could be a potential biological target to elucidate why perilla accumulates high amounts of linolenic acid in leaf or seed. In addition to Contig445, three other sequences encode FADs that could also play roles in unsaturated fatty acid biosynthesis. Collectively, these four sequences will be useful to understand the mechanism of unsaturated fatty acid biosynthesis in perilla.

Interestingly, 3.1% of the total perilla uniESTs were related to lipid metabolism, which is a higher proportion than that reported in a sesame EST analysis that identified 496 (1.4%) sequences for lipid metabolism among 32,421 uniESTs (Ke et al. 2011). Since the sesame EST analysis was performed on immature seeds after pollination, unlike this study in which we used young plants, the difference in proportion of lipid metabolism genes could be due to different sample type used. However, considering that sesame oil contains fatty acids such as palmitic acid (11%), stearic acid (7%), oleic acid (43%), and linoleic acid (35%) (Bedigian et al. 1985), which are quite different from those of perilla oil (Asif 2011), it is also possible that perilla expresses a set of genes for lipid metabolism that are not expressed in sesame.

In conclusion, we constructed a cDNA library and reported EST sequences for perilla, which is used as an oilseed crop, a culinary ingredient and an herbal medicine. The EST information provided in this study will be a valuable resource to facilitate further molecular study and breeding of perilla.

Materials and Methods

Plant Material and cDNA Library Construction

Seeds of *P. frutescens* var. frutescens cv. Youngho Deul-ggae provided from the National Institute of Crop Science in Korea were germinated in soil and plants were cultivated in a greenhouse for three weeks. For total RNA isolation, whole plants were collected and immediately frozen using liquid nitrogen. Total RNA was extracted

with Trizol reagent (Invitrogen, USA), and its quantity and quality were examined using spectrophotometry. A normalized cDNA library was constructed based on PCR-based oligo capping method according to the manufacturer's instructions (Core Bio System, Korea). Briefly, total RNA was ligated with RNA oligonucleotide linkers and reverse transcribed with oligo-dT primers to synthesize first-strand cDNA. Then, the second cDNA was synthesized by PCR amplification with the RNA oligonucleotide linkers. Subsequently, double-stranded cDNA was inserted into the unidirectional cloning vector pCNS-D2 (Oh et al. 2004). The library was constructed by transformation of *E. coli* with the recombinant vectors. Thereafter, the insert sizes of cDNA clones were examined by colony PCR with T7 (5'-TAATA-CGACTCACTATAGGG-3') and SP6 (5'-ATTTAGGTGACA-CTATAG-3') primers or digestion of isolated plasmids with *Eco*RI and *Not*I. Normalization processes were conducted with the constructed cDNA library. Single-stranded tracer DNA was induced from the cDNA library infected with helper phage and used as template to synthesize driver DNA by PCR amplification. Tracer DNA was then hybridized with driver DNA to form duplex DNA. The remaining non-redundant single-stranded DNA was isolated from hybridized duplex DNA, which was converted into doubled-strand DNA and cloned into pCNS-D2 vector. Insert sizes were again confirmed by colony PCR with T7 and SP6 primers.

EST Sequencing, Assembly and Annotation

Plasmids were isolated from clones randomly selected from the normalized cDNA library. DNA sequences were determined from the 5' end of the cDNA inserts with T7 promoter primers by the Sanger method using an ABI 3730xl capillary DNA sequencer. High-quality nucleotide bases were called using the Phred program with score 13, and these sequences were filtered to remove vector, repeat, and contaminating sequences, using the PESTAS web server for EST analysis and sequence mining (http://pestas.kribb.re.kr/pestas.jsp, Nam et al. 2009). Finally, sequences shorter than 100 bp and those including only homo-polymer tracks of more than 20 nucleotides such as poly$(A)_{20}$ and poly$(T)_{20}$ were removed. To generate the uniEST set, the filtered ESTs were assembled using the CAP3 program (http://seq.cs.iastate.edu/) with parameters of 98% sequence identity and 40-bp minimum match length. The uniESTs containing full-length ORFs were identified using the Transcript Decoder program (http://transdecoder.sourceforge.net/) with default parameters (minimum protein length of 100). To annotate uniESTs, BLASTX searches were performed with the NCBI BLAST program (Altschul et al. 1990) against the Nr database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and TAIR database (http://www.arabidopsis.org/Blast/index.jsp) with E-value cut-off of $1E^{-5}$. GO annotation of the uniEST set was performed using the Blast2GO program (Conesa et al. 2005) with default parameters, after BLASTX searches against the Nr database at GenBank. For comparison of perilla ESTs with those from other Lamiaceae species, ESTs of Chinese salvia and sweet basil were retrieved from GenBank. UniESTs for these two species were generated using the same methods as described above and compared with those of perilla at the nucleotide level using the BLAST program (E-value cut-off of $1E^{-5}$).

Identification of uniESTs Related to Lipid Metabolism

UniESTs with homology to genes involved in lipid metabolism were identified by BLASTX searches against genes for lipid metabolism listed at the Arabidopsis acyl-lipid metabolism website (ARALIP, http://aralip.plantbiology.msu.edu/, Li-Beisson et al. 2013). UniESTs encoding enzymes related to the fatty acid unsaturation process in lipid metabolism were further analyzed by full-length sequencing of relevant cDNA clones with mCMVp-forward (5'-ATGTCGTA-ACAACTCCGCC-3'), SP6, and gene-specific primers, and then searching public database and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/).

Acknowledgements

We thank all members of the Laboratory of Functional Crop Genomics and Biotechnology, Seoul National University for their technical assistance. This work was carried out with the support of the "Cooperative Research Program for Agriculture Science & Technology Development (Project title: Study on abiotic stress-tolerance related gene family in *Panax ginseng* and *Brassica rapa* using comparative genomics, Project No. PJ009076)", Rural Development Administration, Republic of Korea.

Author's Contributions

SCL, JL, NHK, JYP, HOL, BSC, KK, JL, MHL and TJY designed the experimental plan, grew perilla plants, isolated total RNAs, and analyzed EST data and cDNA clones; SCL, JL, HUK, MHL, KS, XJ, CS, THK and TJY wrote and revised the manuscript. All authors agreed on the contents of the manuscript and declared no conflicting interest.

Supporting Information

Table S1. List of 4,582 uniESTs of *Perilla frutescens*
Table S2. List of uniESTs homologous to major regulatory genes
Table S3. List of 141 uniESTs involved in lipid metabolism
Fig. S1. Estimation of insert size of cDNA clones.
Fig. S2. Nucleotide and deduced protein sequence of cDNA sequence for uniEST S0807.
Fig. S3. Sequence comparison of S0807 with FAD5.
Fig. S4. Nucleotide and deduced protein sequence of cDNA sequence for uniEST S0071
Fig. S5. Sequence comparison of S0071 with FAD2.
Fig. S6. Nucleotide and deduced protein sequence of cDNA sequence for uniEST S4973.
Fig. S7. Sequence comparison of S4973 with FAD6.
Fig. S8. Nucleotide and deduced protein sequence of cDNA sequence for Contig445.
Fig. S9. Sequence comparison of Contig445 with omega-3 fatty acid desaturases.
Fig. S10. Histograms of flow cytometric analysis to estimate the genome size of perilla.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Asif M (2011) Health effects of omega-3,6,9 fatty acids: *Perilla frutescens* is a good example of plant oils. Orient Pharm Exp Med 11:51–59

Bedigian D, Seigler DS, Harlan JR (1985) Sesamin, sesamolin and the origin of sesame. Biochem Syst Ecol 13:133–139

Chung CH, Kim JL, Lee YC, Choi YL (1999) Cloning and characterization of a seed-specific omega-3 fatty acid desaturase cDNA from *Perilla frutescens*. Plant Cell Physiol 40:114–118

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M

(2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Deng YM, Xie QM, Zhang SJ, Yao HY, Zhang H (2007) Anti-asthmatic effects of perilla seed oil in the guinea pig *in vitro* and *in vivo*. Planta Med 73:53–58.

Diao Y, Miao Y, Lin X, Liao C, Guo F, Hu Z (2009) Comparative analysis of five varieties in *Perilla frutescens* (L.) Britton by 45S rDNA FISH and 5S rDNA sequences. Genetika 45:506–510

Feng LJ, Yu CH, Ying KJ, Hua J, Dai XY (2011) Hypolipidemic and antioxidant effects of total flavonoids of *Perilla frutescens* leaves in hyperlipidenia rat induced by high-fat diet. Food Research International 44:404–409

Gwak YH, Kwon JK, Kim SD, Kim HK, Park CB, Bae SB, Bae SD, Oh KO, Yun ES, Lee DC, Lee MH, Jeong CS, Choi YH, Hong YK (2003) Dlggae. Yeongnam Agricultural Research Institute, National Institute of Crop Science (NICS), Rural Development Administration, Republic of Korea

Ha TH, Lee JH, Lee MH, Lee BW, Kwon HS, Park CH, Shim KB, Kim HT, Baek IY, Jang DS (2012) Isolation and identification of phenolic compounds from the seeds of *Perilla frutescens* (L.) and their inhibitory activities against α-glucosidase and aldose reductase. Food Chemistry 135:1317–1403

Hwang SK, Hwang YS (2000) Molecular cloning and functional expression of *Perilla frutescens* 3-ketoacyl-(acyl carrier protein) synthase III. Mol Cells 10:375–381

Hwang SK, Kim KH, Hwang YS (2000) Molecular cloning and expression analysis of 3-ketoacyl-ACP synthases in the immature seeds of *Perilla frutescens*. Mol Cells 10:533–539

Ichihara K, Suda Y (2003) Lipid biosynthesis in developing perilla seeds. Phytochemistry 63:139–143

Igarashi M, Miyazaki Y (2013) A review on bioactivities of perilla: progress in research on the functions of perilla as medicine and food. Evidence-based complementary and alternative medicine. 2013:Article ID 925342http://dx.doi.org/10.1155/2013/925342

Jun HI, Kim BT, Song GS, Kim YS (2014) Structural characterization of phenolic antioxidants from purple perilla (*Perilla frutescens* var. acuta) leaves. Food Chemistry 148:367–372

Kang R, Helms R, Stout MJ, Jaber H, Chen Z, Nakatsu T (1992) Antimicrobial activity of the volatile constituents of *Perilla frutescens* and its sysnergistic effects with polygodial. J Agric Food Chem 40:2326–2330

Ke T, Dong C, Mao H, Zhao Y, Chen H, Liu H, Dong X, Tong C, Liu S. (2011) Analysis of expression sequence tags from a full-length-enriched cDNA library of developing sesame seeds (*Sesamum indicum*). BMC Plant Biol 11:180

Kitada C, Gong Z, Tanaka Y, Yamazaki M, Saito K (2001) Differential expression of two cytochrome P450s involved in the biosynthesis of flavones and anthocyanins in chemo-varietal forms of *Perilla frutescens*. Plant Cell Physiol 42:1338–1344

Lee JK, Nitta M, Kim NS, Park CH, Yoon KM, Shin YB, Ohnishi O (2002) Genetic diversity of perilla and related weedy types in Korea determined by AFLP analyses. Crop Sci 42:2161–2166

Lee JK, Ohnishi O (2001) Geographic differentiation of morphological characters among perilla crops and their weedy types in East Asia. Breed Sci 51:247–255

Lee JK, Ohnishi O (2003) Genetic relationships among cultivated types of *Perilla frutescens* and their weedy types in East Asia revealed by AFLP markers. Genet Resour Crop Evol 50:65–74

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, Franke RB, Graham IA, Katayama K, Kelly AA, Larson T, Markham JE, Miquel M, Molina I, Nishida I, Rowland O, Samuels L, Schmid KM, Wada H, Welti R, Xu C, Zallot R, Ohlrogge J (2013) Acyl-lipid metabolism. Arabidopsis Book. 11:e0161

Lin ES, Chou HJ, Kuo PL, Huang YC (2010) Antioxidant and antiproliferative activities of methanolic extracts of *Perilla frutescens*. Journal of Medicinal Plants Research 46:477–483

Los DM, Murata N (1998) Structure and expression of fatty acid desaturases. Biochim Biophys Acta 1394:3–15

Makino T (1961) Makino's new illustrated flora of Japan. Hokuryukan, Tokyo (in Japanese)

Makino T, Furuta Y, Wakushima H, Fujii H, Saito K, Kano Y (2003) Anti-allergic effect of *Perilla frutescens* its active constituent. Phytother Res 17:240–243

Mau CJ, Karp F, Ito M, Honda G, Croteau RB (2010) A candidate cDNA clone for (−)-limonene-7-hydroxylase from *Perilla frutescens*. Phytochemistry 71:373–379

Meng L, Loazno YF, Gaydou EM, Li B (2009) Antioxidant activities of polyphenols extracted from *Perilla frutescens* varieties. Molecules 14:133–140

Nam SH, Kim DW, Jung TS, Choi YS, Kim DW, Choi HS, Choi SH, Park HS (2009) PESTAS: a web server for EST analysis and sequence mining. Bioinformatics 25:1846–1848

Nitta M (2001) Origin of Perilla crops and their weedy type. Ph.D. Thesis, Kyoto University, Kyoto

Nitta M, Ohnishi O (1999) Genetic relationships among two Perilla crops, shiso and egoma, and the weedy type revealed by RAPD markers. Jpn J Genet 74:43–48

Oh JH, Sohn HY, Kim JM, Kim YS, Kim NS (2004) Construction of multi-purpose vectors, pCNS and pCNS-D2, are suitable for collection and functional study of large-scale cDNAs. Plasmid 51:217–226

Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell 7:957–970

Onogi N, Okuno M, Komaki C, Moriwaki H, Kawamori T, Tanaka T, Mori H, Muto T (1996) Suppressing effect of perilla oil on azoxymethane-induced foci of colonic aberrant crypts in rats. Carcinogenesis 17:1291–1296

Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 8:321–329

Sadi AM, Toda T, Oku H, Hokama S (1996) Dietary effects of corn oil, oleic acid, Perilla oil, and primrose oil on plasma and hepatic lipid level and atherosclerosis in Japanese quail. Exp Anim 45:55–62

Saita E, Kishimoto Y, Tani M, IizukaM, Toyozaki, M, Sugihara, N, Kondo K (2012) Antioxidant activities of *Perilla frutescens* against low-density lipoprotein oxidation in vitro and inhuman subjects. J Oleo Sci 61:113–120

Seo PJ, Lee SB, Suh MC, Park MJ, Go YS, Park CM (2011). The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in Arabidopsis. Plant Cell 23:1138–1152

Sompornpailin K, Makita Y, Yamazaki M, Saito K (2002) A WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis in *Perilla frutescens*. Plant Mol Biol 50:485–495

Upchurch RG (2008) Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. Biotechnol Lett 30:967–977

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Yamazaki M, Saito K (2006) Isolation and characterization of anthocyanin 5-O-glucosyltransferase in *Perilla frutescens* var. crispa by differential display. Methods Mol Biol 317:255–266

Yamazaki M, Saito K (2011) Molecular genetic study on the anthocyanin chemotypes of *Perilla frutescens* var. crispa. Nat Prod Commun 6:423–427

Yamazaki M, Yamagishi E, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Yamaguchi M, Saito K (2002) Two flavonoid glucosyltransferases from Petunia hybrida: molecular cloning, biochemical properties and developmentally regulated expression.

Plant Mol Biol4 8:401–411
Yonekura-Sakakibara K, Tanaka Y, Fukuchi-Mizutani M, Fujiwara H, Fukui Y, Ashikari T, Murakami Y, Yamaguchi M, Kusumi T (2000) Molecular and biochemical characterization of a novel hydroxycinnamoyl-CoA: anthocyanin 3-O-glucoside-6"-O-acyltransferase from *Perilla frutescens*. Plant Cell Physiol 41:495–502
Yu HC, Kosuna K, Haga M (2004) Perilla: The Genus Perilla. Taylor & Francis 11 New Fetter Lane, London
Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, Zhang Y, Zhang X, Wang Y, Hua W, Li D, Li D, Li F, Yu J, Xu C, Han X, Huang S, Tai S, Wang J, Xu X, Li Y, Liu S, Varshney RK, Wang J, Zhang X (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15:R39
Zhang T, Li L, Hu F, Zhao X, Fu B, Yang D (2012) Analysis of ESTs from a normalized cDNA library of the rhizome tip of *Oryza longistaminata*. J Plant Biol 55:33–42