

## Research Article

# Expressed Sequence Tags Analysis and Design of Simple Sequence Repeats Markers from a Full-Length cDNA Library in *Perilla frutescens* (L.)

Eun Soo Seong,<sup>1</sup> Ji Hye Yoo,<sup>2</sup> Jae Hoo Choi,<sup>2</sup> Chang Heum Kim,<sup>2</sup>  
Mi Ran Jeon,<sup>2</sup> Byeong Ju Kang,<sup>2</sup> Jae Geun Lee,<sup>3</sup> Seon Kang Choi,<sup>4</sup>  
Bimal Kumar Ghimire,<sup>5</sup> and Chang Yeon Yu<sup>2</sup>

<sup>1</sup>Bioherb Research Institute, Kangwon National University, Chuncheon 200-701, Republic of Korea

<sup>2</sup>Department of Bioconvergence Science and Technology, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon 200-701, Republic of Korea

<sup>3</sup>Hwajin Cosmetics, Hongcheon 250-807, Republic of Korea

<sup>4</sup>Department of Agricultural Life Sciences, Kangwon National University, Chuncheon 200-701, Republic of Korea

<sup>5</sup>Department of Applied Bioscience, Konkuk University, Seoul 143-701, Republic of Korea

Correspondence should be addressed to Chang Yeon Yu; [cyyu@kangwon.ac.kr](mailto:cyyu@kangwon.ac.kr)

Received 27 July 2015; Accepted 21 October 2015

Academic Editor: Xiaohan Yang

Copyright © 2015 Eun Soo Seong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Perilla frutescens* is valuable as a medicinal plant as well as a natural medicine and functional food. However, comparative genomics analyses of *P. frutescens* are limited due to a lack of gene annotations and characterization. A full-length cDNA library from *P. frutescens* leaves was constructed to identify functional gene clusters and probable EST-SSR markers via analysis of 1,056 expressed sequence tags. Unigene assembly was performed using basic local alignment search tool (BLAST) homology searches and annotated Gene Ontology (GO). A total of 18 simple sequence repeats (SSRs) were designed as primer pairs. This study is the first to report comparative genomics and EST-SSR markers from *P. frutescens* will help gene discovery and provide an important source for functional genomics and molecular genetic research in this interesting medicinal plant.

## 1. Introduction

*Perilla frutescens* (L.) is a self-compatible annual herb known as the beefsteak mint plant. It is cultivated in East Asian countries, including Japan, China, and Korea, and is an economical crop in the medicinal herb family, Lamiaceae [1]. Its seeds can be processed into foods and nutritional edible oils, and its leaves can be utilized as a traditional medicinal herb or flavor for vegetables [2, 3]. Perilla oil contains abundant polyunsaturated fatty acids (PUFAs), including linolenic (56.8%) and linoleic (17.6%) acids, which are used in salad oils or cooking [4, 5]. The flavor and odor of perilla are caused by the essential oils of monoterpenoids and sesquiterpenoids, including terpenoids, and they are commercially used as a natural fragrance or for flavoring [6]. The perilla leaf is

composed of a number of chemical variants of the volatile essential oil classified as PA-type (mainly in perillaldehyde), EK-type (elsholtziaketone), PK-type (perilla ketone), PL-type (perillene), PP-type (phenylpropanoids), and PT-type (piperitenone) [7]. Perilla has been described as an important pharmaceutical with anti-inflammatory, anti-allergic, and broad antioxidant functions [8, 9].

Expressed sequence tags (ESTs) are fragments of expressed genes occurring from single-pass sequencing of cDNA libraries [10]. EST databases are sources of SSRs that can be developed as ortholog-specific EST-SSR markers and are dependent on genotype applications in many plant species [11–18]. As a molecular tool, EST-SSRs are highly important for studies on genetic populations [19]. They can identify functional markers in the open reading frames (ORFs) or

5'- or 3'-untranslated regions (UTRs) as well as exerting a phenotypic effect [20]. One advantage of the EST-SSR is that it is more transferable across closely related genera compared with unknown SSRs in the UTRs or noncoding sequences. Therefore, EST-SSRs are easy to understand for studying polymorphisms and genetic diversity [21, 22]. EST-derived SSRs have been reported in various plant species, including *Arabidopsis thaliana*, cacao, and sugarcane [23–25]. EST-SSRs also provide a new source for genetic and evolutionary studies based on homology searches of putative SSR functions [26].

In this study, we developed a full-length enriched cDNA library from *P. frutescens* leaves. EST sequence analysis allowed for genome annotation and gene ontologies and the identification of EST-SSR markers for genomic tool development in this less-well-studied medicinal plant species. These results provide useful and multipurpose data for further studies on *P. frutescens*.

## 2. Materials and Methods

**2.1. Plant Materials.** Seeds of *P. frutescens* were obtained after harvest in attached farm of Kangwon National University (Republic of Korea) during each year of collected accessions and grown on pot supplemented with commercial soil (GFC, Hongseong, Republic of Korea) in a greenhouse for a photoperiod of 16-hour light/8-hour dark at 25°C under well-water conditions. Leaves were sampled for RNA isolation.

**2.2. RNA Extraction, cDNA Library Construction, Plasmid DNA Extraction, and Sequencing.** Leaves were ground with a pestle in the presence of liquid nitrogen and ground tissue was used to RNA isolation using the Trizol method [27]. Total RNA was stored at –80°C until use. The full-length cDNA library was constructed using the Creator SMART cDNA Construction Kit (Clontech Laboratories, CA, USA). Concentrations of isolated RNAs were evaluated using a nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and then used for first-strand cDNA synthesis. Second-strand cDNA was purified by QIAquick (Qiagen, Venlo, Netherlands) and ligated or transformed into the pTripleEX Vector. Plasmid DNA extraction was processed using a Multiscreen Plasmid Extraction Kit (Millipore) and purified. The cDNA library was amplified and using GeneAmp PCR System 9700–384 (Applied Biosystems, CA, USA) and DNA clones was sequenced with single-pass sequencing from the 5'-ends of the cDNA.

**2.3. Assembly Annotation.** Because genome and gene information were unavailable, assembly was performed without clustering. In the pretreatment process, PHRED was used to transfer peak information into the quality file and trim low-quality bases. Vector sequences were trimmed using Crossmatch (<http://www.macvector.com/Assembler/trimmingwith-crossmatch.html>). Chimeric clones, polyA-tails, and sequences less than 100 bp were removed with Seqclean. Assembly was performed with CAP3 software. Contigs were manually checked and, together with singlet reads, compiled to generate a final unigene file. Finally, 412 unigene sequences were obtained from 1,000 ESTs, which were composed of 69 contigs and 343 singletons.

Unigenes were searched against the NCBI nonredundant nucleotide (NT) and protein (NR) databases (<http://www.ncbi.nlm.nih.gov/>), the Uniprot\_sprot database (<http://www.ebi.ac.uk/>), and BLAST2GO (<https://www.blast2go.com/>) for functional annotation using the BLAST alignment tool. A sequence was considered as a significant match when the BLAST probability value (E-value) was less than 1e-5, and the match with the most significant E-value was recognized as the best annotation. A BLASTx search was also conducted against the UniProtKB/Swissprot database (<http://www.ebi.ac.uk/>) using default parameters. Unigenes were further annotated with GO terms (<http://geneontology.org/>).

**2.4. Primer Design for EST-SSR Markers.** A total of 1,000 ESTs of 1,056 samples obtained from a cDNA library were detected and analyzed by TRF version 4.07b online software (<http://tandem.bu.edu/trf/trf.html>). SSR sequences were then obtained. SSRs that fit the following criteria were considered for primer design: a minimum length of 18 bp with minimum repetitions for di-, tri-, tetra-, penta-, and hexa-4 and 4, respectively. Primers were designed using Primer 3 (<http://www.premierbiosoft.com/primerdesign/>) according to the following core criteria: a primer length ranging from 18 bp to 22 bp, with 20 bp as the optimum; product size ranging from 100 bp to 400 bp; melting temperature between 50°C and 62°C, with 60°C as the optimum; and GC content between 40% and 60%, with avoidance of mismatch, hairpin structures, and primer dimers that can cause nonspecific amplification.

## 3. Results

**3.1. cDNA Library Quality Check and Reads Assembly.** A full-length cDNA library was constructed from a mixture of *P. frutescens* samples. Library quality was evaluated after sequencing 96 randomly selected clones. On average, the insert size was greater than 1.2 kb. Forty-nine clones (51.04%) yielded sequencing reads above 700 bp; 11 clones (11.45%) were less than 500 bp. After confirming clone quality, a mass-scale sequencing approach was used. Construction of the full-length cDNA library was produced from *P. frutescens*. A total of 1,000 randomly selected clones from the cDNA library were subjected to single-orientation sequencing from the 5'-end using an ABI3730xl Platform (BGI). Read lengths ranged from 420 bp to 844 bp, with an average of 632 bp (Figure 1).

**3.2. GC Content by Assembly of cDNA Reads.** One thousand EST reads were obtained by trimming vector contaminants with Crossmatch and eliminating chimeric clones and short sequences (less than 100 bp). EST reads were then assembled by PHRAP and CAPS software [28, 29]. Results from the CAP3 assembly indicated that the GC content of unigenes varied from 29.46% to 61.32%. Ninety-one percent of the unigenes exhibited GC content between 37.93% and 52.87% (Figure 2).

**3.3. Sequence Annotation.** Annotation of the EST library was achieved through BLAST (Table 1). The NCBI nonredundant

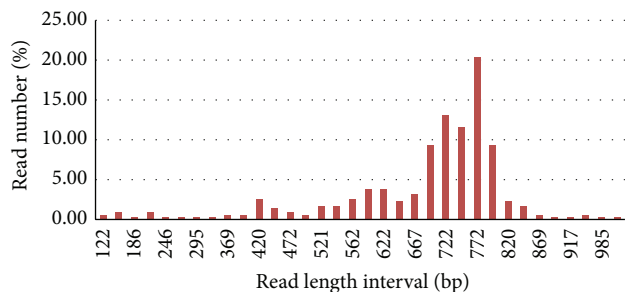


FIGURE 1: Reads length representation in EST (expressed sequence tags) sequencing of *Perilla frutescens*. Range of read length was indicated from 121 bps to 1051 bps.

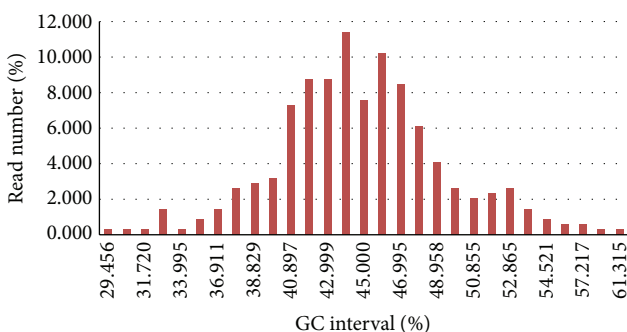


FIGURE 2: GC content division of unigenes. GC content of unigenes changed from 29.45% to 61.32%.

nucleotide (NT) (BLASTn) database resulted in 312 (90.96%) unigenes, whereas the protein (NR) database (BLASTx) produced 322 (93.88%) annotations. Uniprot/Swissprot (BLASTx) databases revealed the annotation of 317 (92.42%) unigenes. Moreover, the annotation data from COG (BLASTx) classification revealed 111 (32.36%) unigenes. Results from the NR database were determined to match that of the sequence homology with two species, *Sesamum indicum* (185 genes, 57.45%) and *Erythranthe guttata* (78 genes, 24.22%). The remaining genes exhibited low levels (less than 1.86%) of sequence homology (Table 2).

3.4. Classification of Annotated Genes by GO Analysis. Gene Ontology (GO) distribution using hierarchy level 2 of the GO program resulted in three major clusters: biological process, cellular component, and molecular function (Figure 3). First, the biological process group was separated into 13 subclasses: signaling (5 genes), response to stimulus (24 genes), growth (1 gene), developmental process (3 genes), multicellular organismal process (3 genes), cellular process (93 genes), biological regulation (16 genes), single-organism process (59 genes), metabolic process (97 genes), localization (15 genes), reproductive process (2 genes), multiorganism process (6 genes), and cellular component organization or biogenesis (21 genes). Organelle (68 genes), cell (94 genes), extracellular region (8 genes), membrane-enclosed lumen (6 genes), cell junction (1 gene), macromolecular complex (44 genes), symplast (1 gene), and membrane (47 genes) genes were distributed from the cellular component cluster. The major components of the molecular function subset consisted of binding (67 genes) and catalytic activity (60 genes) genes.

TABLE 1: Annotated unigenes from different databases by EST (expressed sequence tags) sequencing of *Perilla frutescens*.

Annotation DB (methods)	Hits	%	No hits	%
NT (BLASTn)	312	90.96%	31	9.04%
NR (BLASTx)	322	93.88%	21	6.12%
Uniprot + Swissprot (BLASTx)	317	92.42%	26	7.58%
COG (BLASTx)	111	32.36%	232	67.64%

TABLE 2: List of species containing sequence matches to *Perilla frutescens*.

species (total: 38)	genes (total: 322)
<i>Sesamum indicum</i>	185
<i>Erythranthe guttata</i>	78
<i>Salvia miltiorrhiza</i>	6
<i>Coffea canephora</i>	4
<i>Vitis vinifera</i>	4
<i>Nicotiana sylvestris</i>	4
<i>Genlisea aurea</i>	3
<i>Prunus persica</i>	2
<i>Nicotiana tomentosiformis</i>	2
<i>Ricinus communis</i>	2
<i>Brassica napus</i>	2
<i>Gossypium arboreum</i>	2
<i>Phoenix dactylifera</i>	2
<i>Perilla frutescens</i>	2
<i>Prunus mume</i>	1
<i>Medicago truncatula</i>	1
<i>Malus domestica</i>	1
<i>Solanum tuberosum</i>	1
<i>Schiedea haleakalensis</i>	1
<i>Mentha × piperita</i>	1
<i>Ajuga reptans</i>	1
<i>Codiaeum variegatum</i>	1
<i>Scutellaria baicalensis</i>	1
<i>Nicotiana tabacum</i>	1
<i>Morus notabilis</i>	1
<i>Citrus sinensis</i>	1
<i>Eucalyptus grandis</i>	1
<i>Eutrema salsugineum</i>	1
<i>Citrus clementina</i>	1
<i>Elaeis guineensis</i>	1
<i>Tarenaya hassleriana</i>	1
<i>Miscanthus sinensis</i>	1
<i>Arabidopsis thaliana</i>	1
<i>Arachis diogeni</i>	1
<i>Glycine max</i>	1
<i>Populus trichocarpa</i>	1
<i>Lolium perenne</i>	1
<i>Jatropha curcas</i>	1

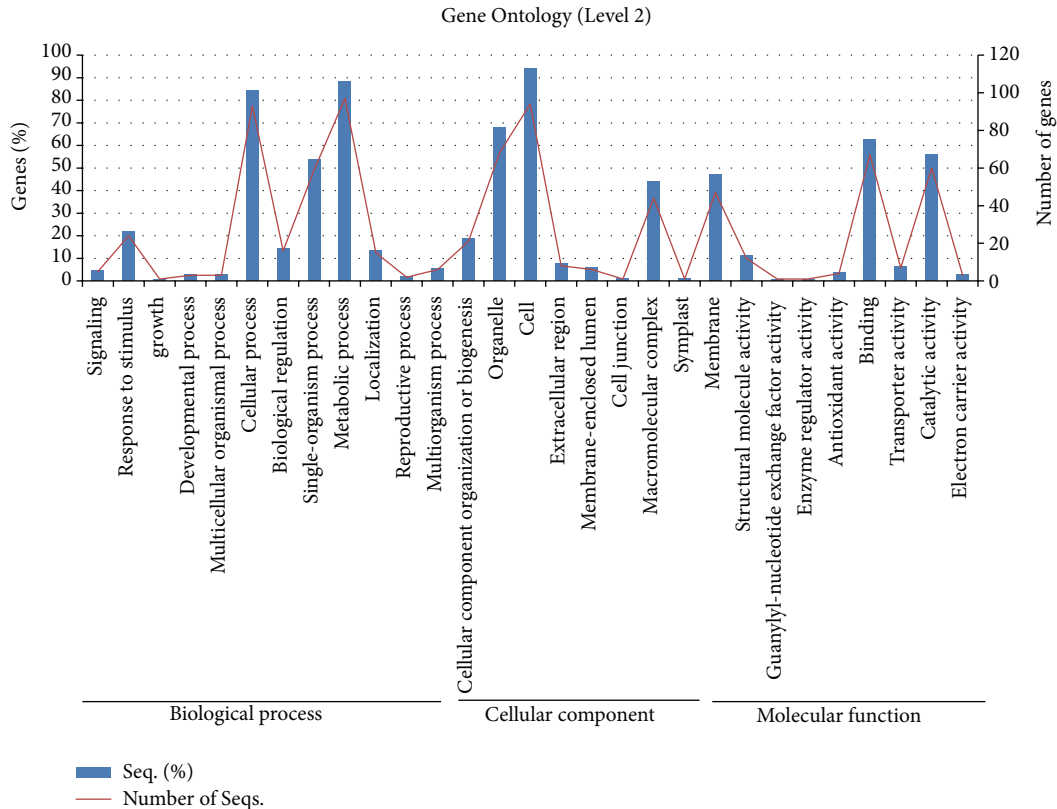


FIGURE 3: Classification of unigenes by Gene Ontology (GO) analysis. Three major clusters were displayed with annotated genes at hierarchy level 2 of GO analysis.

**3.5. EST-SSR Traits in *P. frutescens*.** A total of 343 unigene sequences were investigated. SSR sequences were obtained using TRF version 4.07b online software. Eighteen EST-SSR sequences were selected and analyzed following functional annotation (Table 3). Primer pairs were designed using the Primer 3 program. Expected product sizes ranged from 191 bp to 773 bp. In the future, we will perform additional classification studies through gene functions of *P. frutescens* using these EST-SSR primers.

#### 4. Discussion

The major outcomes of this study were the construction of a full-length cDNA library from the important *P. frutescens* L-type (with limonene component) and the preliminary 1,000 ESTs identified (average 632 bp in length). Genome segment quality was affected by many factors. GC content analysis revealed a distribution between 29.46% and 61.32%. Earlier study showed that thirty to fifty percent of GC content influenced genome sequence quality in *Medicago truncatula* and *Lotus japonicas* [30]. GC content increment was related to the ratio of segments with matching EST data [31], consistent with that from the human genome [32].

Gene Ontology (GO) was utilized to obtain functional information and descriptions of gene products by studying domain-specific ontologies [33]. Annotation results consisted

of biological data related to stress response genes, which were classified functionally using the GO hierarchy [34]. The corresponding classifications were processed to obtain additional information on the putative functionality for the subject accession number of pepper EST data from the GO databases [35]. GO “biological process” and “molecular function,” generated by level 3, were annotated and associated with the number of sequences from each term, which were normalized by labeling with a GO term [36].

We also established 18 EST-SSR primers from the full-length cDNA library of *P. frutescens*. In *Vitis vinifera*, *Artemisia tridentata*, *Panax ginseng*, and *S. multiorrhiza*, the EST-SSR motifs were generally di- and trinucleotide repeats [37–40]. However, this study revealed various penta-, hexa-, dodeca-, and tetradecanucleotide repeat motifs. This finding is in agreement with that for *Scutellaria baicalensis*, which contains penta- and hexanucleotide repeats [41]. Differences in repeat type may be attributed to the degree of the SSR search criteria for the EST database in various plant species. The development of EST-SSR markers has many advantages compared with other molecular markers and can be used to study genetic diversity, evolution, comparative genomics, and gene-based associations.

Construction of a full-length cDNA library is significant for comparative genomics, genome sequence validation, and design of EST-SSR primers that display entire transcription

TABLE 3: EST-SSR primer pairs produced in EST (expressed sequence tags) sequencing database of *Perilla frutescens*.

Number	Unigene ID	Repeat motifs	left primer sequence	Tim.	right primer sequence	Tim.	Product size (bp)	Annotation (NR DB)
1	Contig 17	ATCAT(8)	GAGAGTATAAACAATCCAAAACAGC	58.795	AGCCGGTATATCCAATTCCC	60.006	562	PREDICTED: protein CURVATURE THYLAKOID 1B, chloroplastic [Sesamum indicum]
2	Contig 67	A(29)	AGCAACTGCGGTAGCTAGA	60.176	CAATCCGACCACAGTTGATG	59.96	172	PREDICTED: photosystem I subunit O [Sesamum indicum]
3	Perilla-1-1a.p-Trip1Ex2-seq_C16	GA(16)	AGCGTACTGTTGAAAGCGGTG	59.148	CAGCAAACGTGCTCGAAATTA	60.014	247	PREDICTED: uncharacterized protein LOC101572991 isoform X2 [Sesamum indicum]
4	Perilla-1-1a.p-Trip1Ex2-seq_E18	CTT(9)	GCCAAATTTGAAGCTTTAGCC	58.969	GAATGTGAAAGTGGAAACGGCT	60.119	773	PREDICTED: GRF1-interacting factor 3 [Sesamum indicum]
5	Perilla-1-1a.p-Trip1Ex2-seq_M02	AGAAATG(4)	TGGAGCAAAGTGAAGCAAACAG	60.175	CCTTTTCAGTGAGGAGCCAG	59.982	191	PREDICTED: homeobox-leucine zipper protein HAT5-like [Sesamum indicum]
6	Perilla-1-2a.p-Trip1Ex2-seq_A02	TTTTTG(7)	AATGATGGGTGTGATGAGCA	59.925	AAAGAAATTTGAAGGGCCAGA	59.96	401	hypothetical protein MIMGU_mgvia015066mg [Erythranthe guttata]
7	Perilla-1-3a.p-Trip1Ex2-seq_B13	ATCAT(8)	GAGAGTATAAACAATCCAAAACAGC	58.795	CGGTATATCCAATTCCCACG	60.031	559	PREDICTED: uncharacterized protein LOC105160440 [Sesamum indicum]
8	Perilla-1-3a.p-Trip1Ex2-seq_L05	CT(14)	CCCAAAATTCACATCCACTGA	59.343	AACAACGTGACATGGGCTTCC	59.973	185	PREDICTED: annexin D5 [Sesamum indicum]
9	Perilla-1-3a.p-Trip1Ex2-seq_L15	TC(15)	CAGTTTAACTTCGCCCTCGC	60.018	CACTCGCAAAAAGGGGTAAG	59.741	619	hypothetical protein MIMGU_mgvia016040mg [Erythranthe guttata]
10	Perilla-2-1a.p-Trip1Ex2-seq_A19	GGA(8)	GCTCCTCGCAGTAACTTTGG	60.015	TCAITCTCTTGCTCTGTTTCCA	58.583	107	hypothetical protein MIMGU_mgvia020048mg [Erythranthe guttata]
11	Perilla-2-2a.p-Trip1Ex2-seq_A06	CT(12)	CAITGGCCCTTAAACTTCGGA	60.067	ATAAAITGGATGGATTGGGGCAA	60.016	341	hypothetical protein MIMGU_mgvia012334mg [Erythranthe guttata]
12	Perilla-2-2a.p-Trip1Ex2-seq_C18	AG(14)	GGGGATCAATTCAGTCTT	60.133	GTGCCCACTGGTTCCTTTGTT	60.012	404	hypothetical protein MIMGU_mgvia012334mg [Erythranthe guttata]
13	Perilla-3-2a.p-Trip1Ex2-seq_E14	GATGACGATGAT(2)	CTTTCCAAACCCTCCGAAATT	60.291	CGACGCCGTGTCATCTACTACA	60.008	522	NAAC transcription factor 1 [Salvia miltiorrhiza]
14	Perilla-3-2a.p-Trip1Ex2-seq_O22	GA(17)	GGGATATGTTATGTTGCTTGT	59.179	TCGCCGCTACTGATCCCCTAC	60.096	514	PREDICTED: uncharacterized protein LOC105169169 isoform XI [Sesamum indicum]
15	Perilla-3-3a.p-Trip1Ex2-seq_B03	CT(16)	CGAGTGTGTTCTGATGGGTG	60.025	AACGCCCTACGGAACAGAGAC	60.321	184	hypothetical protein MIMGU_mgvia014836mg [Erythranthe guttata]
16	Perilla-3-3a.p-Trip1Ex2-seq_B23	TCCTCTTCCTCTCC(2)	TAGTGTGGAAGCTCAATGGC	59.028	TGACCAGCATCAGCTTTCAC	59.992	662	PREDICTED: chloroplast stem-loop binding protein of 41kDa a, chloroplastic [Sesamum indicum]
17	Perilla-3-3a.p-Trip1Ex2-seq_F11	GAG(9)	GAAAGACTGGTTGGCTCTGG	59.844	ATCCAAAATTCGTCCTGTGC	59.939	381	hypothetical protein MIMGU_mgvia01207mg [Erythranthe guttata]
18	Perilla-3-3a.p-Trip1Ex2-seq_P03	GA(13)	AAAGCTGTTTGCCCTTGCTA	60.018	CTCAAAATGGAGTCACGCCAGA	59.984	284	hypothetical protein MIMGU_mgvia016830mg [Erythranthe guttata]



units rather than partial gene sequences [42]. One benefit of constructing a full-length cDNA library is that it allowed us to conduct proper gene modeling while comparing other cDNA sequences in *P. frutescens*. The full-length cDNA sequences will be useful for annotation of the plant genome. Another advantage of EST sequencing is the increased ratio of unigenes with definitive GO categories compared with other libraries. The library built by this method included a high proportion of full-length cDNAs [42], allowing us to have a database of this library available for *P. frutescens* genomics studies.

This full-length cDNA library provides a wealth of knowledge about the unique EST sequences available for the *P. frutescens* genome and, particularly, about the addition of 5'-end sequences that are more unique and valuable for gene identification. These EST tags will be useful for functional gene annotation, analysis of splice site variations, and gene homologies as additional whole-genome sequences become available in *P. frutescens*.

## 5. Conclusions

*Perilla frutescens* is valuable as a medicinal plant as well as a natural medicine and functional food. However, comparative genomics analyses of *P. frutescens* are limited due to a lack of gene annotations and characterization. A full-length cDNA library from *P. frutescens* leaves was constructed to identify functional gene clusters and probable EST-SSR markers through 1,056 examples of expressed sequence tag (EST) sequencing data. Unigene assembly was performed using basic local alignment search tool (BLAST) homology searches and annotated Gene Ontology (GO). A total of 18 simple sequence repeats (SSRs) were designed as primer pairs. This study is the first to report comparative genomics and EST-SSR markers from *P. frutescens* to ease gene discovery and provide an important source for functional genomics and molecular genetic research in this interesting medicinal plant.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Authors' Contribution

Eun Soo Seong and Ji Hye Yoo contributed equally to this work.

## Acknowledgments

This study was supported by the Ministry of Education (MOE) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation (no. 2014H1C1A1067085) and was partially supported by the Bioherb Research Institute and the Research Institute of Agricultural Science, Kangwon National University, Republic of Korea.

## References

- [1] Y.-J. Park, A. Dixit, K.-H. Ma et al., "Evaluation of genetic diversity and relationships within an on-farm collection of *Perilla frutescens* (L.) Britt. using microsatellite markers," *Genetic Resources and Crop Evolution*, vol. 55, no. 4, pp. 523–535, 2008.
- [2] A. K. Pandey and K. C. Bhatt, "Diversity distribution and collection of genetic resources of cultivated and weedy type in *Perilla frutescens* (L.) Britton var. *frutescens* and their uses in Indian Himalaya," *Genetic Resources and Crop Evolution*, vol. 55, no. 6, pp. 883–892, 2008.
- [3] C. X. You, K. Yang, Y. Wu et al., "Chemical composition and insecticidal activities of the essential oil of *Perilla frutescens* (L.) Britt. aerial parts against two stored product insects," *European Food Research and Technology*, vol. 239, pp. 481–490, 2014.
- [4] H.-S. Shin and S.-W. Kim, "Lipid composition of perilla seed," *Journal of the American Oil Chemists' Society*, vol. 71, no. 6, pp. 619–622, 1994.
- [5] T. Longvah and Y. G. Deosthale, "Chemical and nutritional studies on Hanshi (*Perilla frutescens*), a traditional oilseed from northeast India," *Journal of the American Oil Chemists Society*, vol. 68, no. 10, pp. 781–784, 1991.
- [6] S. J. Kim, E. Y. Kang, S. E. Won et al., "Chemical composition and comparison of essential oil contents of *Perilla frutescens* Britton var. *japonica* HARA leaves," *Korean Journal of Medicinal Crop Science*, vol. 16, pp. 242–254, 2008.
- [7] M. Nitta, H. Kobayashi, M. Ohnishi-Kameyama, T. Nagamine, and M. Yoshida, "Essential oil variation of cultivated and wild *Perilla* analyzed by GC/MS," *Biochemical Systematics and Ecology*, vol. 34, no. 1, pp. 25–37, 2006.
- [8] H. Ueda, C. Yamazaki, and M. Yamazaki, "Luteolin as an anti-inflammatory and anti-allergic constituent of *Perilla frutescens*," *Biological and Pharmaceutical Bulletin*, vol. 25, no. 9, pp. 1197–1202, 2002.
- [9] M.-K. Kim, H.-S. Lee, E.-J. Kim et al., "Protective effect of aqueous extract of *Perilla frutescens* on tert-butyl hydroperoxide-induced oxidative hepatotoxicity in rats," *Food and Chemical Toxicology*, vol. 45, no. 9, pp. 1738–1744, 2007.
- [10] M. D. Adams, M. B. Soares, A. R. Kerlavage, C. Fields, and J. C. Venter, "Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library," *Nature Genetics*, vol. 4, no. 4, pp. 373–380, 1993.
- [11] R. K. Varshney, A. Graner, and M. E. Sorrells, "Genic microsatellite markers in plants: features and applications," *Trends in Biotechnology*, vol. 23, no. 1, pp. 48–55, 2005.
- [12] R. K. Varshney, R. Sigmund, A. Börner et al., "Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice," *Plant Science*, vol. 168, no. 1, pp. 195–202, 2005.
- [13] R. Peakall, S. Gilmore, W. Keys, M. Morgante, and A. Rafalski, "Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants," *Molecular Biology and Evolution*, vol. 15, no. 10, pp. 1275–1287, 1998.
- [14] S. Temnykh, G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch, "Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential," *Genome Research*, vol. 11, no. 8, pp. 1441–1452, 2001.

- [15] T. Thiel, W. Michalek, R. K. Varshney, and A. Graner, "Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)," *Theoretical and Applied Genetics*, vol. 106, no. 3, pp. 411–422, 2003.
- [16] J.-K. Yu, T. M. Dake, S. Singh et al., "Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat," *Genome*, vol. 47, no. 5, pp. 805–818, 2004.
- [17] J.-K. Yu, M. La Rota, R. V. Kantety, and M. E. Sorrells, "EST derived SSR markers for comparative mapping in wheat and rice," *Molecular Genetics and Genomics*, vol. 271, no. 6, pp. 742–751, 2004.
- [18] A. Heesacker, V. K. Kishore, W. Gao et al., "SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility," *Theoretical and Applied Genetics*, vol. 117, no. 7, pp. 1021–1029, 2008.
- [19] J. R. Ellis and J. M. Burke, "EST-SSRs as a resource for population genetic analyses," *Heredity*, vol. 99, no. 2, pp. 125–132, 2007.
- [20] Y.-C. Li, A. B. Korol, T. Fahima, and E. Nevo, "Microsatellites within genes: structure, function, and evolution," *Molecular Biology and Evolution*, vol. 21, no. 6, pp. 991–1007, 2004.
- [21] C. H. Pashley, J. R. Ellis, D. E. McCauley, and J. M. Burke, "EST databases as a source for molecular markers: lessons from *Helianthus*," *Journal of Heredity*, vol. 97, no. 4, pp. 381–388, 2006.
- [22] M. A. Chapman, J. Hvala, J. Strever et al., "Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.)," *Theoretical and Applied Genetics*, vol. 120, no. 1, pp. 85–91, 2009.
- [23] A. Depeiges, C. Goubely, A. Lenoir et al., "Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci," *Theoretical and Applied Genetics*, vol. 91, no. 1, pp. 160–168, 1995.
- [24] G. M. Cordeiro, R. Casu, C. L. McIntyre, J. M. Manners, and R. J. Henry, "Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Erianthus* and sorghum," *Plant Science*, vol. 160, no. 6, pp. 1115–1123, 2001.
- [25] L. S. Lima, K. P. Gramacho, A. S. Gesteira et al., "Characterization of microsatellites from cacao-*Moniliophthora perniciosa* interaction expressed sequence tags," *Molecular Breeding*, vol. 22, no. 2, pp. 315–318, 2008.
- [26] E. De Keyser, J. de Riek, and E. van Bockstaele, "Discovery of species-wide EST-derived markers in *Rhododendron* by intron-flanking primer design," *Molecular Breeding*, vol. 23, no. 1, pp. 171–178, 2009.
- [27] E. S. Seong, J. H. Yoo, J. H. Choi et al., "Construction and classification of a cDNA Library from *Miscanthus sinensis* (Eulalia) treated with UV-B," *Plant Omics Journal*, vol. 8, pp. 264–269, 2015.
- [28] F. C. Peixoto and J. M. Ortega, "On the pursuit of optimal sequence trimming parameters for EST projects," in *Proceedings of the 1st Brazilian Symposium/Workshop on Bioinformatics (BSB/WOB '02)*, pp. 48–55, Gramado, Brazil, October 2002.
- [29] X. Huang and A. Madan, "CAP3: a DNA sequence assembly program," *Genome Research*, vol. 9, no. 9, pp. 868–877, 1999.
- [30] L. Shangguan, J. Han, E. Kayesh et al., "Evaluation of genome sequencing quality in selected plant species using expressed sequence tags," *PLoS ONE*, vol. 8, no. 7, Article ID e69890, 2013.
- [31] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, and M. C. Zody, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, pp. 860–921, 2001.
- [32] R. Versteeg, B. D. C. van Schaik, M. F. van Batenburg et al., "The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes," *Genome Research*, vol. 13, no. 9, pp. 1998–2004, 2003.
- [33] S. K. Arasan, J.-I. Park, N. U. Ahmed et al., "Gene ontology based characterization of expressed sequence tags (ESTs) of *Brassica rapa* cv. Osome," *Indian Journal of Experimental Biology*, vol. 51, no. 7, pp. 522–530, 2013.
- [34] The Gene Ontology Consortium, "Gene ontology annotations and resources," *Nucleic Acids Research*, vol. 41, pp. 530–535, 2013.
- [35] H. J. Kim, K. H. Baek, S. W. Lee et al., "Pepper EST database: comprehensive *in silico* tool for analyzing the chili pepper (*Capsicum annuum*) transcriptome," *BMC Plant Biology*, vol. 8, article 101, 7 pages, 2008.
- [36] A. Conesa and S. Götz, "Blast2GO: a comprehensive suite for functional analysis in plant genomics," *International Journal of Plant Genomics*, vol. 2008, Article ID 619832, 12 pages, 2008.
- [37] K. D. Scott, P. Eggler, G. Seaton et al., "Analysis of SSRs derived from grape ESTs," *Theoretical and Applied Genetics*, vol. 100, no. 5, pp. 723–726, 2000.
- [38] P. Bajgain, B. A. Richardson, J. C. Price, R. C. Cronn, and J. A. Udall, "Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*)," *BMC Genomics*, vol. 12, article 370, 2011.
- [39] T. S. Wu, C. Liang, H. B. Li, and Z. Y. Piao, "Development of unigene derived microsatellite (UGMS) markers in *Panax ginseng*," *Scientia Agricultura Sinica*, vol. 44, pp. 2650–2660, 2011.
- [40] K.-J. Deng, Y. Zhang, B.-Q. Xiong et al., "Identification, characterization and utilization of simple sequence repeat markers derived from *Salvia miltiorrhiza* expressed sequence tags," *Acta Pharmacologica Sinica*, vol. 44, no. 10, pp. 1165–1172, 2009.
- [41] Y. Yuan, P. Long, C. Jiang, M. Li, and L. Huang, "Development and characterization of simple sequence repeat (SSR) markers based on a full-length cDNA library of *Scutellaria baicalensis*," *Genomics*, vol. 105, no. 1, pp. 61–67, 2015.
- [42] M. Seki, M. Narusaka, A. Kamiya et al., "Functional annotation of a full-length *Arabidopsis* cDNA collection," *Science*, vol. 296, no. 5565, pp. 141–145, 2002.