Construction of a cDNA library and preliminary analysis of expressed sequence tags in *Piper hainanense*

R. Fan[1,2], P. Ling[1,4], C.Y. Hao[1,3], F.P. Li[1,3], L.F. Huang[1,2], B.D. Wu[1,3] and H.S. Wu[1,2]

[1]Spice and Beverage Research Institute,
Chinese Academy of Tropical Agricultural Sciences, Wanning, Hainan, China
[2]Key Laboratory of Genetic Resources Utilization of Spice and Beverage Crops,
Ministry of Agriculture, Wanning, Hainan, China
[3]Hainan Provincial Key Laboratory of Genetic Improvement and Quality Regulation
for Tropical Spice and Beverage Crops, Wanning, Hainan, China
[4]Citrus Research and Education Center, University of Florida, Lake Alfred, FL, America

Corresponding author: H.S. Wu
E-mail: 13807622912@163.com

Genet. Mol. Res. 14 (4): 12733-12745 (2015)
Received June 22, 2015
Accepted August 12, 2015
Published October 19, 2015
DOI http://dx.doi.org/10.4238/2015.October.19.17

ABSTRACT. Black pepper is a perennial climbing vine. It is widely cultivated because its berries can be utilized not only as a spice in food but also for medicinal use. This study aimed to construct a standardized, high-quality cDNA library to facilitated identification of new *Piper hainanense* transcripts. For this, 262 unigenes were used to generate raw reads. The average length of these 262 unigenes was 774.8 bp. Of these, 94 genes (35.9%) were newly identified, according to the NCBI protein database. Thus, identification of new genes may broaden the molecular knowledge of *P. hainanense* on the basis of Clusters of Orthologous Groups and Gene Ontology categories. In addition, certain basic genes linked to physiological processes, which can contribute to disease resistance and thereby to the breeding of black pepper. A total of 26 unigenes were found to be SSR markers. Dinucleotide SSR was the main repeat motif, accounting

for 61.54%, followed by trinucleotide SSR (23.07%). Eight primer pairs successfully amplified DNA fragments and detected significant amounts of polymorphism among twenty-one piper germplasm. These results present a novel sequence information of *P. hainanense*, which can serve as the foundation for further genetic research on this species.

Key words: EST; Expressing; *Piper hainanense*

INTRODUCTION

As a significant member of *Piperaceae* (Tian et al., 2006; Ahmad et al., 2010), *Piper nigrum* is world-renowned as the “King of Spices” due to its trade in the international market. Interestingly, it is also called “black pepper” because of its peppercorn’s color. With around 2 000 species, *Piperaceae* (genus *Piper*) is the largest family of the basal angiosperms (Quijano et al., 2006). Through our long-term resources introduction and evaluation, we screened out some good quality attributes such as *Piper flaviflorum*, *Piper hainanense*, *Piper sinense*. As a common species, *P. hainanense* can mostly be found in provinces such as Guangxi, Hainan, and Guangdong in China (Hao et al., 2011). The part of *P. hainanense* above the ground has been utilized in folk medicine to alleviate pain, treat bacterial infections, and to relieve symptoms of asthma. As *P. hainanense* has more merits than *P. nigrum*, including more fruit branches, better disease resistance, and longer ear length, which reduced the need for fungicides and pesticides during the production of black peppers with a high quality.

With a single-pass sequence that is shorter than mRNAs, expressed sequence tags (ESTs) represent the genes expressed in a particular tissue and/or during a particular period of development (Adams et al., 1991; Mekhedov et al., 2000). EST sequencing is an effective approach to obtain data at the functional genomic level for non-model organisms. Extensive EST sequencing would be of great value for the discovery and annotation of genes, comparative genomics, molecular marker development, and population genomic studies of genetic variations that are relevant to adaptive traits (Bouck et al., 2007; Namroud et al., 2008). Besides, molecular markers contribute greatly to plant breeding in various respects, including identification of genes governing desirable traits. Simple sequence repeats (SSRs) offer several advantages over other molecular markers, including extensive genomic coverage, simplicity, hypervariability, effectiveness, reproducibility, abundance, and codominant inheritance (Powellet al., 1996). Based on the original sequences used for their identification, SSRs can be classified into two categories: EST-SSRs and genomic SSRs. The currently available approaches for the identification and isolation of genomic SSRs are expensive, labor intensive, and time consuming (Zane et al., 2002; Squirrell et al., 2003). On the contrary, as EST-SSRs are developed from the expressed sequences that are more evolutionally conserved than the noncoding sequences, the transferability of EST-SSR markers is higher. The number of EST datasets that can be applied to both model and nonmodel organisms is steadily growing, although there are few EST datasets that can be applied to black pepper.

This paper reports the construction of a high-quality cDNA library using the leaves of *P. hainanense*, the expression pattern of the genes in the leaves, and the primary analysis of the ESTs. The obtained EST datasets would be a valuable resource for marker-assisted selective breeding and discovery of genes of black peppers.

MATERIAL AND METHODS

Total RNA isolation

P. hainanense was provided by the Ministry of Agriculture in Wanning City, from its Black Pepper (*Piper* spp.) Germplasm Repository. Young leaves in good condition were chosen to extract RNA for the construction of cDNA library. They were frozen in liquid nitrogen and stored at -80°C until RNA extraction, which was conducted using the TRIzol reagent (OMEGA Laboratories Inc.,2696 Kempston Drive Woodbury MN55125, USA) in accordance with the manufacturer's instructions. The quality of extracted RNAs was determined according to the A260/A280 and agarose gel ratio.

cDNA library constructions

Double-stranded cDNA was synthesized using the Advantage 2 PCR Kit and In-Fusion SMARTerTM Directional cDNA Library Construction Kit (Clontech Laboratories Inc., CA, USA) in the BIO-RAD CFX96 (Bio-Rad Laboratories, Inc., California, USA), in accordance with the manufacturer's instructions. The first-strand of the cDNA was synthesized using the primers CDS-3 M (5'-AAGCAGTGGTATCAACGCAGAGTGGCCGAGGCGGCC (T) 20VN-3') and SMART Oligo IV (5'-AAGCAGTG GTATCAACGCAGAGTGGCCATTACGGCCGGG-3'). For the production of ds-cDNA, 15 cycles of polymerase chain reaction (PCR) with a long range and a 5'-PCR primer (5'-AAGCAGTGGTATCAACGCAGAGT- 3') were used. The obtained PCR products were subjected to agarose gel electrophoresis to determine the quality of the ds-cDNA, and ds-cDNAs shorter than 1000 bp were extracted using the OMEGA Gel Extraction Kit (OMEGA Laboratories Inc., 2696 Kempston Drive Woodbury MN55125,USA). The cDNAs longer than 1000 bp were screened out and subjected to ligation with the Psmart2ifd linearized vector at 50°C for 15 min. The resulting plasmids were used to transform DH-10B competent cells (Invitrogen, Carlsbad, CA, USA).

Primary library titration

cDNAs longer than 1000 bp were cloned into Psmart2ifd linearized vectors. Vectors containing the inserts were used to transform the XL1-Blue strain of *Escherichia coli*, and the transformants were cultured on LB/MgSO4 /IPTG /X-Gal (with 30 μg/mL ampicillin) plates for blue-white screening. Using the quantity of clones obtained, the library titer can be calculated according to the formula: PFU/mL = number of plaques x dilution factor x 10^3 μL/mL (μL of diluted phage plated). Fifty clones were selected randomly for determination of insertion rate and insert size, and the cDNA library obtained following amplification were placed in 384-well plates and stored at -80°C.

Sequencing and analysis

5'-End single-pass sequencing of 612 randomly chosen cDNA clones was conducted using the ABI 3130 XL Genetic Analyzer (Applied Biosystems). Low-quality cDNAs, those shorter than 100 bp were excluded, and the raw single-pass sequence data were separated from the vector sequence data. Sequence assembly and clustering were conducted using the CAP3 sequence assembly program. The cDNA sequences were then compared with the protein and nucleotide sequences in the Swiss-Prot (http://www. Ecstasy.Ch/sprot/), and Interpro (http://www. ebi.ac.uk/interpro/) databases, as well as using Blastx (http://www.ncbi.nlm.nih.gov/BLAST/). The functions

of ESTs were identified using the Gene Ontology (GO, http:// www. geneontology.org/) and the Clusters of Orthologous Groups of proteins (COGs) (http://www.ncbi.nlm.nih.gov/COG/) databases (E value < 10-5). Moreover, the reaction networks and molecular interactions of the sequences in the library were explored using Pathways of Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/).

Primer design and EST-SSR detection

Possible SSR markers were screened from the 262 unigenes using the MISA tool (http:// pgrc.ipk-gatersleben.de/misa). In order to achieve the best hexanucleotide and di-, tetra-, tri-, and penta- motif identification, the parameters were adjusted by setting the number of repeats to 4, 6, 4, 5, and 4, respectively. Because it is difficult to differentiate simon-pure mononucleotide repeats from the single nucleotide stretch errors and polyadenylation products, which are generated during sequencing, mononucleotide repeats were ignored. Primer pairs were designed using BatchPrimer3 (Kortt et al., 1991). Furthermore, the main parameters for the design of primer pairs were as follows: primer length, 18-23 bp (optimal, 20 bp); amplicon size, 100-400bp (optimal, 200 bp); GC content, 40-70% (optimal, 50%); and annealing temperature, 50-60°C (optimal, 55°C). In order to detect germplasm polymorphism in black pepper the synthesis and design of 50 primer pairs were carried out on the basis of these parameters.

Detection of EST-SSR polymorphism

To investigate polymorphisms in EST-SSRs, 21 accessions of black peppers, including cultivars, foreign collections, and Chinese landraces, were chosen. Total DNA was extracted from the seedlings of sesame using the CTAB method (Porebski et al., 1997). Polymerase chain reaction (PCR) amplifications were performed under the following conditions: final volume of reaction mixture, 10 µL (50 ng template DNA, 1X PCR buffer, 2.0 mM $MgCl_2$, 2.5 mM dNTPs, 4 µM each primer, and 0.8 U Taq polymerase [Fermentas, Burlington, Ontario,USA]). The PCR cycle was as follows: 4 min at 94°C, 30 cycles for 40 s at 94°C, 40 s at 55°C, 1 min at 72°C, and final extension for 10 min at 72°C. The amplicons were separated on an 8% polyacrylamide gel and a DNA marker of 50 bp (Promega,Madison,USA) was used to determine the total length of the EST-SSR amplicons. Finally, the PCR products were mixed with half the volume of the loading buffer.

RESULTS

LD-PCR and total RNA isolation

The total RNA extracted from the leaves was 1.05 µg/µl, and the ratio of A260 to A280 was 2.10, indicating that the RNA was suitable for the construction of cDNA library. A 1.2% agarose/EtBr gel electrophoresis of the separated RNA revealed the expected 18S and 28S bands at 2.0 kb and 5.0 kb, respectively (Figure 1-A). When the a double-stranded cDNA (5 µL) was analyzed using the 1.2% agarose/EtBr gel, a smear of 0.4-5 kb was observed, confirming successful synthesis of double-stranded cDNA (Figure 1-B).

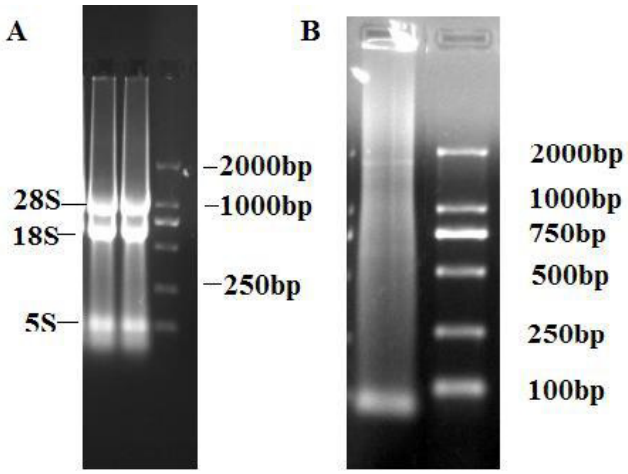

Figure 1. Electrophoretograms of the amplified products through LD-PCR with agarose gel electrophoresis and the total RNA separated from the leaves of Piper hainanense (A). **A**-1,2: total RNA separated from the leaves; **B**-1: double-stranded cDNA; and M; DNA Marker.

cDNA library construction and characterization

The original titer of the cDNA library was 2.0 x 10^6 cfu/ml. As 83% clones have inserts, 50 clones were randomly chosen for the identification of insertion rate and insert size. In order to determine the insert size, PCR amplification of the inserts, using specific primers, was performed. The amplification results revealed that all 50 clones had inserts of 800-2000 bp (Figure 2), confirming that the cDNA library of pepper comprised large inserts as well as a high titer and recombination rate.

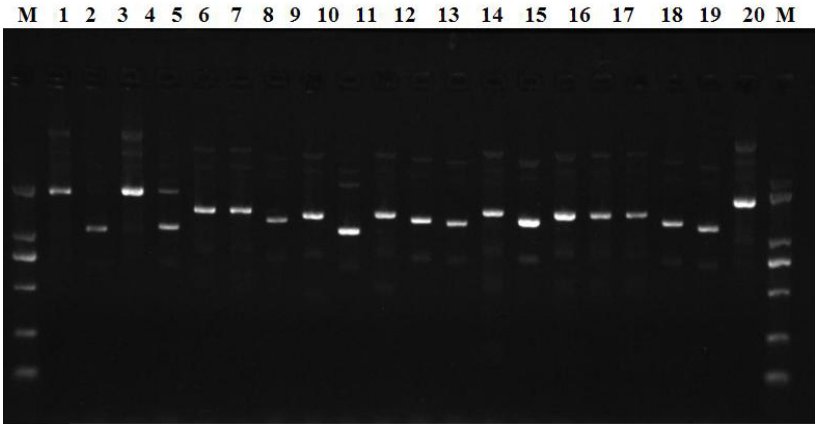

Figure 2. Amplified cDNA inserts from the constructed cDNA library. *Lanes 1-20*: cDNA inserts in phage plaques; M: DNA size marker.

Normalized cDNA library characterization

Using 5'-terminus sequencing, 612 clones from the cDNA library of pepper were selected randomly and assembled. Apart from the 15 ESTs of insufficient quality and 51 vector ESTs, 546 ESTs of high quality were identified, and their average length was 774.8 bp. In addition, 29 singlets and 233 contigs, representing 262 unparalleled genes, were obtained following compiling and assembly of the original sequences. The proportion of unigenes in the cDNA library was 47.9% (Table 1). However, only the expression of 2 unigenes was regarded as efficient, because each contig had more than 20 ESTs, and the largest unigene contained 29 ESTs (Figure. 3).

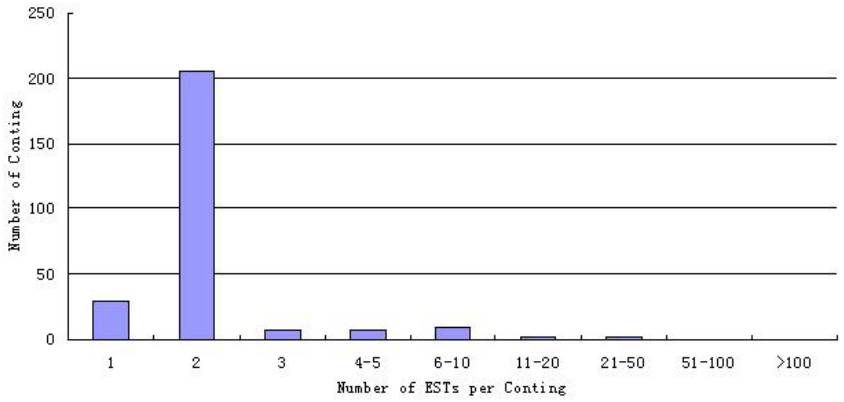

Figure 3. Characteristics of the assembled contigs in the cDNA library.

Table 1. Summary of ESTs obtained from the cDNA library of *Piper hainanense.*

Items	Number
Total sequence cDNA	612
Total number of ESTs	546
Average length of ESTs (bp)	774.8
Unique genes	262
Average length of unigenes	1316.31
Contings	233
singletons	29
With ORF	212
Average length of unigene ORF (bp)	253.45
Redundancy (%)	52.01%

Sequence annotation and analyses

Based on an E-value less than 10^{-5} in the Blastx analysis, the unigenes were compared with nucleotide sequences, as well as with those in the database of non-redundant proteins in NCBI. According to the Blastx analysis, 168 (64.1%) of the 262 sequences closely corresponded to the database of non-redundant protein. By comparing these 262 unigenes with Swiss-Prot database, it was found that 86 (32.8%) unigenes matched those in a certain particular species. Furthermore, further functional analysis revealed that 11.1% of the unigenes matched some known genes in the COG Database (Table 2).

Table 2. Annotation of unigenes in the cDNA library.

Database	Total	Annotaed	Percent (%)	Unannotated	Percent (%)
Nr	262	168	64.1	94	35.9
Swiss-prot	262	86	32.8	176	67.2
COG	262	29	11.1	233	88.9
Kegg	262	139	53.1	123	46.9

According to the functions of these unigenes, they were divided into 16 categories, including the prediction of general function, translation, posttranslational modification, ribosomal biogenesis and structure, energy production, replication and conversion, as well as repair and recombination, which can be panoramically viewed in Figure 4 and Table 2. Of these unigenes, 139/262 (53.1%) were found to match the metabolic pathways found in KEGG Pathway Database, which allows for environmental information processing, amino acid metabolism, cellular processes, and genetic information processing.

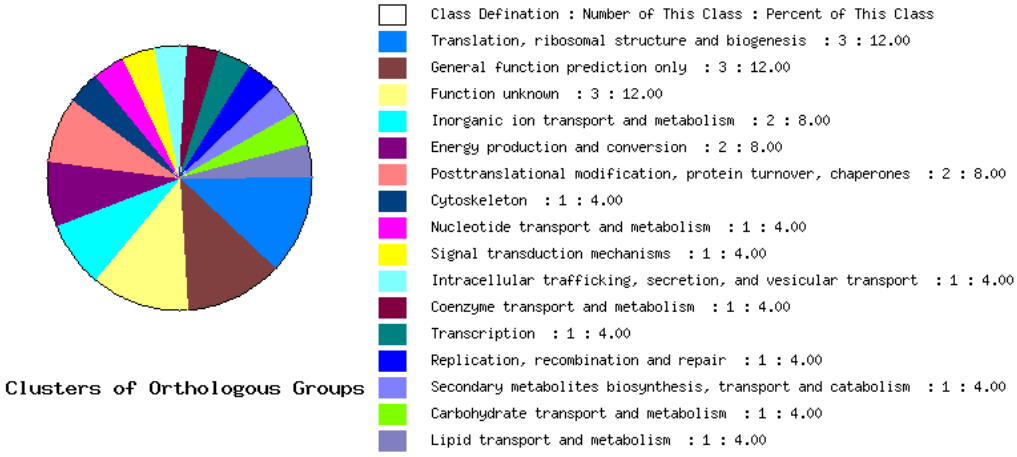

Figure 4. Classification of unique genes in P. hainanense according to COG.

Using the GO database, the sequences that matched the known genes of other species were analyzed; it was found that the unparalleled sequences matched a number of GO annotations. The unparalleled sequences that were matched were classified into three categories according to their functions: biological process (Figure 5), cellular component, and molecular function (Figure 6), which can be panoramically viewed in Figure 7.

Blastx analysis revealed that the previously assumed transcription-coupled DNA repair protein, protein, and polyprotein were in fact the most sufficient ESTs, which also showed the highest expression in the leaves of *P. hainanense*. In addition, up to 144 ESTs were found to possess over 2 copies and 25 single copies of ESTs, and 88 ESTs were homologous to the assumed protein in other plants.

Figure 5. Gene Ontology (GO) categories of biological process of genes.

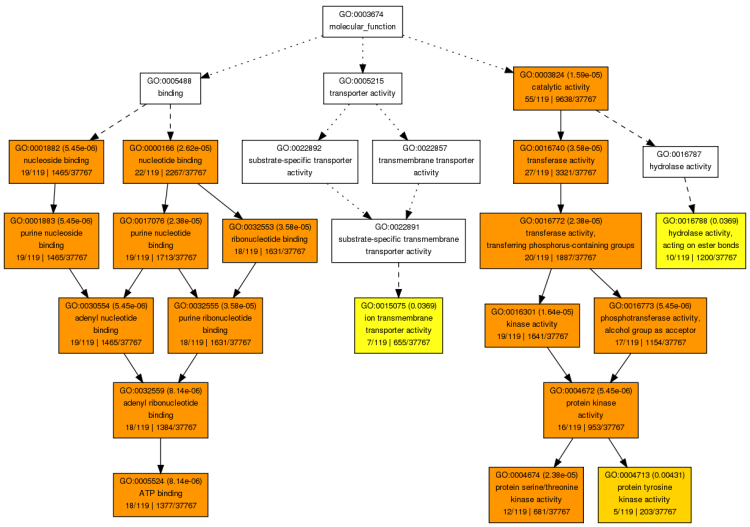

Figure 6. Gene Ontology (GO) categories of molecular function of genes.

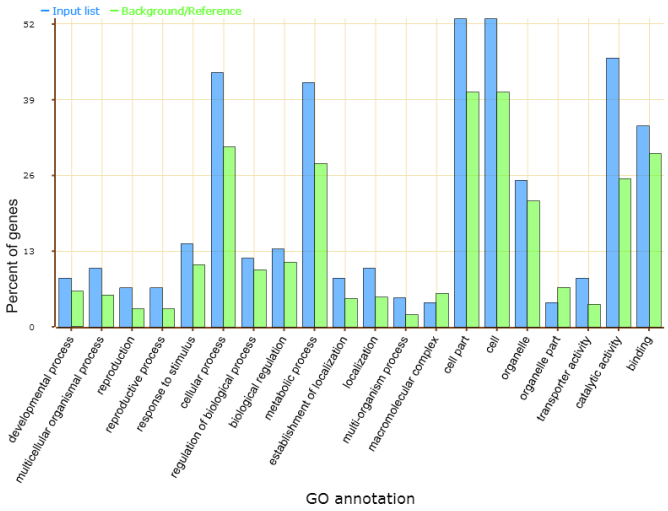

Figure 7. Gene Ontology (GO) categories of genes derived from cDNA library of P. hainanense.

Distribution and frequency of EST-SSRs in black pepper transcriptome

From the 262 sequences, 26 EST-SSRs were identified. The distribution density of EST-SSRs in black pepper transcriptome was 16.2 kb, with a frequency of 2.61%. Dinucleotide was the most abundant repeat motif, accounting for 61.54%, followed by trinucleotide, accounting for 23.07% (Table 3). In addition, SSRs with 9 tandem repeats were in the majority, accounting for 30.77%, followed by those with 5 tandem repeats (26.92%), those with more than 10 tandem repeats (15.38%), those with 10 tandem repeats (15.38%), and those with 7 tandem repeats (7.69%). AG/CT was the predominant repeat motif in the EST-SSRs, accounting for 38.46%, followed by AC/GT (23.07%), AAG/CTT (7.69%), and ACC/GGT (7.69%) (Table 4). Ten pairs of primers were chosen randomly from these 26 EST-SSR primers for evaluation of polymorphism and application across 21 accessions of piper (Figure 8). Of these, 8 pairs successfully amplified the fragments; one generated PCR fragments longer than the anticipated value, and 7 generated amplicons with the anticipated size.

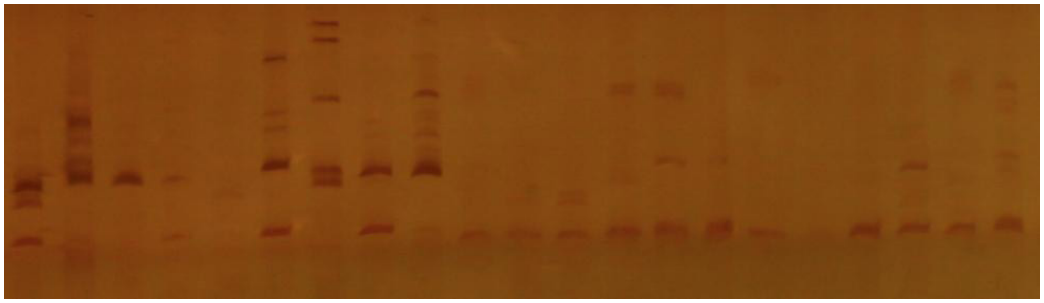

Figure 8. Fingerprinting of black pepper using EST-SSR markers.

Table 3. Frequency of EST-SSRs in piper.

Motif Length	Repeat number							Total	%
	5	6	7	8	9	10	>10		
Di					8	4	4	16	61.54
Tri	6							6	23.07
Tetra		1	1					2	7.69
Penta			1					1	3.85
Hexa	1							1	3.85

Table 4. Frequency of di- and tri-nucleotide EST-SSR motifs in piper.

Repeat Motif	Repeat numbers							Total	%
	5	6	7	8	9	10	>10		
AG/CT					6	1	3	10	38.46
AC/GT					2	3	1	6	23.07
AAG/CTT	2							2	7.69
ACC/GGT	2							2	7.69
AAT/AAT	1							1	3.85
CCG/CGG	1							1	3.85

DISCUSSION

Characterization of cDNA library of *P. hainanense*

Because cDNA library analysis can provide vast amounts of useful information on various procedures and genomic mechanisms, it has been regarded as an essential approach for functional genomic analysis in the last few years (Shao et al., 2009). The major characteristic of cDNA construction by SMART technique could improve the ratio of full-length cDNA sequences. Conventionally generated cDNA libraries contain a high percentage of 5'-truncated clones.. Nevertheless, the SMART technique is an efficient approach to generate cDNA libraries with a full range and high quality while maintaining the 5' terminal sequence of the mRNA (Chen et al., 2005).

As the elementary cDNA library is not stable enough and has a limited volume, there is a need to increase its size and stability. The quality of a cDNA library can be determined based on 3 main factors. According to the formula of Clareke-Carbon, a minimum of 1.7×10^5 independent clones should exist in a cDNA library for it to obtain 99% mRNA with low abundance (Clarke et al., 1976). Another factor that can prove that the library has sufficient quality is high recombination efficiency (Li et al., 2009), and the third factor is that the inserted cDNA should have an average length that is more than or equivalent to 1.0 kb to ensure that the cDNA is complete. The full-range *P. hainanense* cDNA library in this study meets the preconditions for a normative library.

PCR amplification was used to confirm insertion and determine insert size. Furthermore, it was found that several cDNA clones had inserts less than or equivalent to 1 500 bp long, which indicates that they can be the best choice for further research on gene expression and cloning. According to the previous studies on the generation of extensive ESTs, an efficient library should have minimum 50% new genes, various transcripts, and uninformative sequences less than or equivalent to 20% (Adams et al., 1995). The cDNA library constructed in this study met these requirements.

EST analysis and generation

At the initial stage, the collection of ESTs were built by single-pass sequencing of random cDNAs to discover new genes at only fraction of the cost of genomic sequencing while facilitating the identification of coding regions in genomic sequences (Lievens et al., 2001). It has been demonstrated that EST analysis, generation, and high-throughput single-run partial sequencing can be a high-efficiency and quick approach to obtain information on the expression patterns of mRNAs (Ablett et al., 2000; Wu et al., 2002; Bausher et al., 2003). Thus, a complete view of the biological processes in *P. hainanense* can be achieved by analyzing the expression of various genes and identifying their functions.

A total of 66% of the ESTs constructed in this study were gene sequences having functions that have been predicted or known previously, while the rest were undiscovered gene sequences or proteins that did not match any in the databases. The latter can be utilized in the future detection of new genes using functional analysis.

According to the function catalogues, which was constructed for *Arabidopsis* and yeast, the genes were classified on the basis of COG and the presumptive notes on the functions obtained through Blastx analysis (Mewes 1997; Initiative 2000). Among these ESTs, Blast X analysis revealed that 88.9% did not match any protein sequence in COG databases. These ESTs were tentatively allocated to the prediction of the general functions, translation, as well as ribosomal biogenesis and structure, and the categories with undiscovered functions were in the majority, accounting for 25%. Nevertheless, 2 ESTs belonged to the category associated with energy, representing 16.67% of the total functional genes on the basis of COG. Nonetheless, compared with the results of studies on citrus (34%) and grape (39.56%), this result is far lower.

A comprehensive understanding of piper can be achieved through the ESTs of *P. hainanense* that match genes with assumed or known functions. It was found that Clone 2373 matched an intriguing gene associated with a plant disease resistance protein, i.e., NBS-LRR proteins. According to some reports, NBS-LRR proteins control constitutive and negative regulation of its signaling potential (Shirano et al., 2002; Bendahmane et al., 2002; Hwang and Williamson, 2003). Furthermore, a sequence (i.e., clone 2478) was found to be similar to the protein NPR1. Systemic acquired resistance (SAR) is a kind of resistance with a broad spectrum in plants, which has been linked to the upregulation of a battery of pathogenesis-related (PR) genes. NPR1 is an essential regulator in the transduction pathways of signals that generate SAR. Several ESTs greatly matched the possible, underlying or assumed proteins, and many were identified in by genome sequencing studies on diverse organisms (e.g., *Drosophila*, human). Nonetheless, it was beyond the scope of this study to analyze the assumed genes recognized in this research. Further, genetic, biochemical and sequence analyses are needed for further understanding of the functions of these genes.

EST-SSR marker characterization and identification

EST-SSR markers are important for comparative genomics, genetic diversity assessment, marker-assisted selection breeding, and genetic map development. To date, as only few hundreds of EST-SSR markers have been identified (An et al., 2009; Triwitayakorn et al., 2011), they are rarely utilized for black peppers. However, it is easy to identify EST-SSR markers for black peppers owing to the abundant sequences supplied by the cDNA library constructed in this study. A total of 26 possible EST-SSRs were recognized in 262 unigenes. The most common type of SSR motif was the dinucleotide repeat. These results are in accordance with those for peanut, sugar beet,

canola, cabbage, sunflower, soybean, pea, sweet potato, grape, and *Arabidopsis* (Kumpatla and Mukhopadyay, 2005). In contrast, trinucleotide repeat is the major type of SSR found in cereals, including barley, rice, and wheat (La et al., 2005). Furthermore, the AAG/CTT motif (7.69%) and AG/CT motif (38.46%) were the most common in the dataset. These results too are in line with those reported for other plant species (Morgante et al., 2002; La et al., 2005; Kumpatla and Mukhopadyay, 2005). CTT repeats and TC repeats are found in the transcribed regions in the plants, with a high occurrence rate in 5'-UTRs; antisense transcription might involve the CT microsatellites in 5'-UTRs, which can contribute to gene regulation (Martienssen et al., 2001). Among the 10 pairs of primers chosen randomly for PCR confirmation, 8 (80%) successfully generated amplicons, and 7 of these 8 generated amplicons of the anticipated size; the failure of one primer that could not generate amplicons of the expected size can be attributed to the existence of introns (Saha et al., 2004; Varshney et al., 2005), assembly errors, the variations in the repeat number, large insertions, or insufficient specificity. These results show that the assembled unigenes were of high quality and the identified EST-SSRs in this dataset can contribute to future studies.

In conclusion, this study demonstrates that sequencing of the cDNA library as well as EST analysis and identification can not only reveal the expression profile of an mRNA but also present a cost-effective approach for the identification of the novel functional genes, which is also an efficient strategy for functional genomic studies, especially for plants that lack genomic information, such as *P. hainanense*. Nevertheless, as this kind of analysis can only uncover the type of function of a gene, large amounts of biological and biochemical data are still required to obtain a clear picture of genes and their functions under various conditions.

ACKNOWLEDGMENTS

Financial support from the Ministry of Agriculture of the People's Republic of China (# 15RZZY-13) and the Program for the Top Young Talents in Chinese Academy of Tropical Agricultural Sciences (#1630052015049) are gratefully acknowledged.

REFERENCES

Ablett E, Seaton G, Scott K, Shelton D, et al. (2000). Analysis of grape ESTs: global gene expression patterns in leaf and berry. *Plant Sci.* 159: 87-95

Adams MD, Kelley JM, Gocayne JD, Dubnick M, et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252: 1651-1656.

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, et al. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377: 3-174

Ahmad N, Fazal H, Abbasi BH, Rashid M, et al. (2010). Efficient regeneration and antioxidant potential in regenerated-tissues of Piper nigrum L. *Plant Cell, Tissue Organ Cult*. 102: 129-134.

An ZW, Zhao YH, Cheng H, Li WG, et al. (2009). Development and application of EST-SSR markers in Hevea brasiliensis Muell. *Arg. Hereditas* 31: 311-319.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis* thaliana. *Nature* 408: 796-815.

Bausher M, Shatters R, Chaparro J, Dang P, et al. (2003). An expressed sequence tag (EST) set from Citrus sinensis L. Osbeck whole seedling and the implications of further perennial source investigations. *Plant Sci.* 165: 415-422.

Bendahmane A, Farnham G, Moffett P and Baulcombe DC (2002). Constitutive gain-of-function mutants in a nucleotide binding site–leucine rich repeat protein encoded at the Rx locus of potato. *Plant J.* 32: 195-204.

Bouck A and Vision T (2007). The molecular ecologist's guide to expressed sequence tags. *Mol. Ecol.* 16: 907-924.

Chen XH, Chen Z, Yao HP, Chen F, et al. (2005). Construction and characterization of a cDNA library from human liver tissue with chronic hepatitis B. *J. Zhejiang Univ. Sci. B*. 6: 288-294.

Clarke L and Carbon J (1976). A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome. *Cell* 9: 91-99.

Hao CH, Tan LH, Fan R, Cheng HP, et al. (2011). Predicting potential geographical distributions of medicinal plant Piper hainanense using maximum entropy. *Chin. J. Trop. Crops.* 32: 1561-1566.

Hwang CF and Williamson VM (2003). Leucine-rich repeat-mediated intramolecular interactions in nematode recognition and cell death signaling by the tomato resistance protein Mi. *Plant J.* 34: 585-593.

Kortt AA, Caldwell JB, Lilley GG and Higgins TJV (1991). Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (Helianthus annuus L.). *Eur. J. Biochem.* 195: 329-334.

Kumpatla SP and Mukhopadhyay S (2005). Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48: 985-998.

La Rota M, Kantety R, Yu J-K and Sorrells M (2005). Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genom.* 6: 23.

Li YP, Xia RX, Wang H, Li XS, et al. (2009). Construction of a full-length cDNA Library from Chinese oak silkworm pupa and identification of a KK-42-binding protein gene in relation to pupa-diapause termination. *Int. J. Biol. Sci.* 5: 451-457.

Lievens S, Goormachting S and Holsters M (2001). A critical evaluation of differential display as a tool to identify genes involved in legume nodulation: looking back and looking forward. *Nucleic Acids Res*. 29: 3459-3468.

Martienssen RA and Colot V (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* 293: 1070-1074.

Mekhedov S, Martinez O and Ohlrogge J (2000). Towards a functional catalog of the plant genome: a survey of genes for lipid biosynthesis. *Plant Physiol.* 122: 389-401.

Mewes W (1997). Overview of the yeast genome. *Nature* 387: 7-84.

Morgante M, Hanafey M and Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* 30: 194-200.

Namroud MC, Beaulieu J, Juge N, Laroche J, et al. (2008). Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* 17: 3599-3613.

Porebski S, Bailey L and Baum B (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15: 8-15.

Powell W, Morgante M, Andre C, Hanafey M, et al. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225-238.

Quijano-Abril MA, Callejas-Posada R and Miranda-Esquivel DR (2006). Areas of endemism and distribution patterns for Neotropical Piper species (Piperaceae). *J. Biogeogr.* 33: 1266-1278.

Saha M, Mian M, Eujayl I, Zwonitzer J, et al. (2004). Tall fescue EST-SSR markers with transferability across several grass species. *Theor. Appl. Genet.* 109: 783-791.

Shao ZT, Cong X, Yuan JD, Yang GW, et al. (2009). Construction and characterization of a cDNA library from head kidney of Japanese sea bass (Lateolabrax japonicus). *Mol. Biol. Rep.* 36: 2031-2037.

Shirano Y, Kachroo P, Shah J and Klessig D F (2002). A gain-of-function mutation in an *Arabidopsis* toll interleukin 1 receptor–nucleotide binding site-leucine-rich repeat type R gene triggers defense responses and results in enhanced disease resistance. *Plant Cell.* 14: 3149-3162.

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, et al. (2003). How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* 12: 1339-1348.

Tian B, Lin ZB, Ding Y and Ma QH (2006). Cloning and characterization of a cDNA encoding Ran binding protein from wheat. *DNA Seq*. 17: 136-142.

Triwitayakorn K, Chatkulkawin P, Kanjanawattanawong S, Sraphet S, et al. (2011). Transcriptome sequencing of Hevea brasiliensis for development of microsatellite markers and construction of a genetic linkage map. *DNA Res.* 18: 471-482.

Varshney RK, Graner A and Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23: 48-55.

Wu J, Maehara T, Shimokawa T, Yamamoto S, et al. (2002). A comprehensive rice transcript map containing 6591 Expressed Sequence Tag sites, *Plant Cell.* 14: 525-535.

Zane L, Bargelloni L and Patarnello T (2002). Strategies for microsatellite isolation: a review. *Mol. Ecol.* 11: 1-16.