VOLUME 8 NOMOR 2 DECEMBER 2021 **ISSN 2548 – 611X**

JURNAL
BIOTEKNOLOGI & BIOSAINS INDONESIA

Homepage Jurnal: http://ejurnal.bppt.go.id/index.php/JBBI

SSR MARKERS CHARACTERIZATION FOR TEMU IRENG (*Curcuma aeruginosa* Roxb.) GENERATED FROM EST OF *Curcuma longa*

Karakterisasi Marka SSR untuk Temu Ireng (*Curcuma aeruginosa* Roxb.) dari EST *Curcuma longa*

Devit Purwoko[1]*, Siti Zulaeha[1], Teuku Tajuddin[1], Hayat Khairiyah[1], Reynaldi Zulfikar Fauzi[2], Priyanti[2]
[1]Laboratory for Biotechnology OR-PPT BRIN, 630 Building, Puspiptek Area, Tangerang Selatan, Banten 15314, Indonesia
[2]Program Studi Biologi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah, Jl. Ir. H. Juanda No 95, Ciputat Timur, Tangerang Selatan, Banten 15412, Indonesia
*Email: devit.purwoko@bppt.go.id

ABSTRACT
*Temu ireng (*Curcuma aeruginosa*) is used as a traditional herb medicine in Indonesia. Expressed Sequence Tags (EST) of* C. longa *was used because there were no genome data of* C. aeruginosa*. This study aimed to recognize the diversity of SSR characteristic and to develop a primer for* C. aeruginosa *genetic diversity. The EST was preprocessed and analyzed fogg7r the SSR motif. Primers were designed using SSR motifs excluding mononucleotide (≥ 20 bp) and analyzed based on gene ontology. The result showed that out of 12.675 EST used there were 3.005 contig and 548 perfect SSR motif with the motif frequency of 1/15,27 kb sequences achieved. AGG trinucleotide was mostly distributed (37,40%). There were 380 primers designed and 10 primers validated by PCR on three* C. aeruginosa *clones from Ponorogo, Cikarang, and Kebumen. The number of alleles detected was about 2 to 3 per locus. The polymorphic microsatellite markers produced from this study could be used for the analysis of genetic diversity of* C. aeruginosa *in Indonesia.*

Keywords: Curcuma aeruginosa, Curcuma longa*, EST, primer, SSR*

ABSTRAK
Temu ireng *(Curcuma aeruginosa)* digunakan sebagai bahan baku obat herbal tradisional di Indonesia. *Expressed Sequence Tags (EST)* dari *C. longa* digunakan karena sampai saat ini belum adanya data genome dari *C. aeruginosa*. Penelitian ini bertujuan untuk mengetahui keragaman karakteristik SSR dan mengembangkan primer untuk prastudi keragaman genetik *C. aeruginosa*. EST dilakukan prapemrosesan dan dianalisis motif SSR. Primer didesain menggunakan motif SSR selain mononukleotida (≥ 20 bp) dan dianalisis berdasarkan ontologi gen. Hasil yang diperoleh dari 12.678 EST yang digunakan dihasilkan 3.005 contig dan 548 motif perfect SSR dengan frekuensi motif 1/15,27 kb sekuen. AGG trinukleotida merupakan yang paling banyak terdistribusi (37,40%). Sebanyak 380 primer didesain dan 10 primer divalidasi melalui PCR pada 3 klon *C. aeruginosa* asal Ponorogo, Cikarang, dan Kebumen. Jumlah alel yang terdeteksi sekitar 2 sampai 3 per lokus. Marka mikrosatelit polimorfik yang dihasilkan dari penelitian ini dapat digunakan untuk analisis keragaman genetik *C. aeruginosa* di Indonesia.

Kata Kunci: *Curcuma aeruginosa*, *Curcuma longa*, EST, primer, SSR

Received: *20 April 2021* ***Accepted:*** *03 July 2021* ***Published:*** *01 December 2021*

INTRODUCTION

Temu ireng (*Curcuma aeruginosa* Roxb.) is commonly used as a raw material for traditional herbal medicines in Indonesia. The rhizome is often used in the herbal mixture which contains bioactive compounds such as flavonoids, steroids, and polyphenols (Jose and Thomas 2014, Khumaida et al. 2019) which are effective for cough medicine (Sofyan et al. 2013), dengue fever medicine (Moektiwardoyo et al. 2014), for disinfectant and antifungal (Atun et al. 2016), as well as anti-viral (Setiyono and Bermawie 2013). C. aeruginosa is known to have anti-cancer activity potential as a drug to prevent breast tumor growth (Suphrom et al. 2012, Rachmanita 2015). Septaningsih et al. (2018) identified 175 metabolites content and their biological activities of *C. aeruginosa* using LC-MS fingerprint in combination with chemometrics. Curzerenon was identified as one of the active compounds. Curzerenon found in *C. aeruginosa* can be used as a biomarker associated with SSR markers. This compound has a significant cytotoxicity effect on several cancer cell lines (Simoh and Zainal 2015).

As yet, there is very little known about molecular information regarding the genetic character of *C. aeruginosa* compared to its close relatives such as *C. longa* (Siju et al. 2010, Sahoo et al. 2017), *Zingiber officinale* (Awasthi et al. 2017), *C. zedoaria* (Lan et al. 2018) and *C. alismatifolia* (Taheri et al. 2019). Hence, expressed sequenced tags (EST) could be used since they contain candidate genes and could produce molecular markers associated with particular traits (Kalia et al. 2011). This study is the first step in an effort to clarify the optimum utilization of genetic diversity of *C. aeruginosa* in the future.

The EST sequence of *C. longa* is used as a reference because it has a high transferability rate between species up to 100% (Siju et al. 2010) and is from the same genus as *C. aeruginosa* so that it is expected to have most genes matches to each other. The EST sequences are used to perform in silico experiment on the diversity of characteristics of simple sequence repeat (SSR) markers. SSRs are short repetitive sequences consisting of tandemly repeated short motifs of 1-6 nucleotides which are abundant in the genome of eukaryotic organisms (Sen et al. 2017). In addition, SSR is also present in the EST so its use is very valuable, especially for plants that lack genome information. The advantages of SSR markers are that they have a high degree of polymorphism and codominant inheritance (Kalia et al. 2011). EST-SSR is frequently applied on other plant species for molecular mapping, genetic diversity, transferability etc. such as *Brassica oleracea* (Izzah et al. 2014), *Colocasia esculenta* (You et al. 2015), *Melilotus* (Yan et al. 2017), *Lolium multiflorum* (Pan et al. 2018), *Elettaria cardamomum* (Sakthipriya and Sabu 2018).

The in silico method is a study of biological sciences with bioinformatics principles used for computational-based data analysis (Bare et al. 2019). This method is known to be more cost-effective because it utilizes sequences that are already available in the database and can be applied in genetic diversity studies, one of which is to develop molecular markers. Molecular markers are used to identify specific DNA sequences in a group of organisms and conserve biodiversity (Selvakumari et al. 2017). The use of molecular markers is more efficient and accurate, especially for identification purposes (Terryana et al. 2018).

There are available nucleotides, protein and transcript (EST) databases on the online platform for many model plant species such as *Physcomitrella patens*, *Oryza sativa* and *Arabidopsis thaliana*. However, their availability is only few or none for the non-model plant, such as Curcuma. SSR markers are molecular markers that frequently used for plant identification. SSR markers from one plant can be used to explore the variation at the genus or family level because these markers have cross-gene transferability (Sen et al. 2017). Since generating SSR or microsatellites are labor intensive and can be expensive, so the in silico utilization of these sequences in model plant databases and their respective transfers to the non-model plant is a potentially fruitful strategy (Victoria et al. 2011).

This study comprised of in silico identification of EST-SSR markers, functional domain analysis of markers, characterization using *C. longa* references, and validation with *C. aeruginosa* in order to analyze the level of marker transferability. The purpose of our study was to generate the EST-SSR markers

to be applied to genetic diversity study of *C. aeruginosa* which was developed from the *C. longa* EST data bank, as an effort to conserve genetic resources of *C. aeruginosa* in Indonesia.

MATERIALS AND METHODS

Research location and time

This research was accomplished in January-June 2020 at the Center for Biotechnology, the National Research and Innovation Agency (BRIN).

Research materials

Our study used EST *C. longa* sequences which were resulted from cDNA construction derived from rhizome tissues, targeted for EST preprocessing and downloaded on March 7, 2020, by selecting the complete record file type and FASTA format, from the NCBI website (https://www.ncbi.nlm.nih.gov/nuccore/?term=EST+Curcuma+longa). The accession number was HO002075.1-HO002159.1 and DY382717.1-DY395309.1. Primers validation used three DNA genomes of *C. aeruginosa* from Ponorogo, Cikarang, and Kebumen.

EST sequence processing and contig assembly

The EST sequences that have been downloaded from the NCBI were then uploaded to the EGassembler website (https://www.genome.jp/tools/egassembler/). The aims of these steps were to clean sequences, remove vector contamination, and assemble contig sequences (Masoudi-Nejad et al. 2006). Sequence processing was performed using standard parameters as recommended by the website. The contig sequences resulted from previous assembly and processing were then downloaded in FASTA format for further SSR motif identification.

SSR sequence mining

The contig sequences were further analyzed using Geneious prime software version 2019.2.3 (https://www.geneious.com/prime/) with Phobos plugin version 3.3.12 (http://www.ruhr-unibochum.de/ecoevo/cm/cm_phobos.htm) for identification of SSR motifs. The identified SSR motif was a perfect repeating motif with a minimum repetition unit length of 1 and a maximum repetition unit length of 6, so that mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs could be obtained. The motifs results that have been identified were then saved in FASTA format. Further analyses were completed on the number of motifs, the distribution of motifs, and the frequency of motifs using the software of Microsoft Excel.

Primer designing for SSR

Primer was designed using Primer3 version 2.3.7 (Untergasser et al. 2012) which is a tool in Geneious Prime software version 2019.2.3 (https://www.geneious.com/prime/). The contig sequence that had previously identified its SSR motif was then used as a target for primer design by selecting a motif other than the mononucleotide motif which has a length of ≥20 bp. Parameters used in designing primers including GC content of 50-60%, melting temperature (Tm) 55-60 ºC, and primer product with the length of 18-20 bp. The primer product was then saved in FASTA format for further analysis of the functional gene ontology.

Functional analysis and gene ontology (GO)

The successfully designed primers were then analyzed for the three domains of biological processes (BP), molecular function (MF) and cellular components (CC), using the Blast2GO or new software, which has been incorporated into Omicsbox version 1.3.11. The primers BLAST was compared to the NCBI database using the BLASTx. The parameter used was the e-value 1.0E-3 using a non-redundant data type. The next step was GO mapping analysis aimed to provide annotation from the highest BLAST hit results (Gotz et al. 2008).

The parameters used for GO mapping followed the suggested standards by the software. After mapping the GO, then proceed with the GO annotation which aims to provide functional annotations to the query sequence (Gotz et al. 2008). The parameters used for GO annotations were Annotation Cut Off of 55, GO weight of 5, Hit-Filter e-value of 1.0E-6, and HSP-Hit Coverage Cut Off of 0. After completing all the process, the combine graph menu was assigned for viewing the distribution of biological processes, molecular

Table 1. *C. longa* EST sequence processing results

Parameter	Value
Total EST sequences used	12.678
EST sequence after processing	12.675
Analyzed EST sequence length (bp)	8.371.507
EST sequence length after cleaning (bp)	8.291.575
Number of contig	3.005
Number of singletons	3.008
Number of ESTs that generated contig	9.667 (76,27%)

Table 2. Distribution of SSR motifs on contig sequences

Motif Type	Total Number of Motifs	Frequency (%)	Most Abundant Motifs
Mononucleotide	118	21.53	A
Dinucleotide	58	10.58	AG
Trinucleotide	205	37.40	AGG
Tetranucleotide	40	7.29	AAAG
Pentanucleotide	32	5.83	AAAAG
Hexanucleotide	95	17.33	ACGGCG
Total	548	100	–

functions and cellular components. In order to find out the mapping of metabolic pathways based on GO, the Pathway analysis menu was selected and then choose load KEGG pathways (Kanehisa and Goto 2000).

Primer validation

SSR primers containing information on metabolic pathways of metabolites were then validated using the polymerase chain reaction (PCR) method. The PCR composition was set to a total volume of 25 μL/reaction, consisting of Go taq green master mix (12.5 μL), forward primer (2 μL), reverse primer (2 μL), DNA template (2 μL) and sterile H_2O up to a volume of 25 μL. The PCR program used was as follow: predenaturation at 95 ºC for 3 minutes, denaturation at 95 ºC for 30 seconds 35 cycles, annealing Tm-5 ºC for 30 seconds and extension at 72 ºC for 30 seconds, for 35 cycles each, and final extension 72 ºC for 60 seconds and finally hold at 4 ºC. The PCR product was then electrophoresed on an agarose at the concentration of 1.5% at 50 volts for 120 minutes.

RESULTS AND DISCUSSION

EST sequence processing and assembly

Identification of SSR diversity in EST *C. longa* for *C. aeruginosa* pre-study is important because it is the first step in an effort to study the genetic information of *C. aeruginosa*, which up until now does not have genomic data available in the NCBI database. The results obtained from a total of 12,675 EST sequences of *C. longa* that were successfully cleaned which bring about the contig sequences of 76.27% (Table 1). With a high percentage of the number of ESTs that produce contigs, it indicated that there was an overlapping area in the sequences. This result was in accordance with Sen et al. (2017) conclusion that the higher the EST value producing contigs, the more the sequences have overlapping areas. Sequences with the overlapping areas can be used as references for combining the sequences into a longer unit, that it would be possible to create a complete gene sequence (Hadiarto et al. 2015).

SSR sequences mining

The purpose of sequence processing was to clean up sequences that have low quality or complexity (Kaur and Vaidya 2015). The sequences were first cleaned before further analysis was implemented, so that the sequences that had poor criteria were not analyzed. The cleaned sequences were then assembled into contigs and singletons. Contig itself was a sequence assembled from processing results that have overlapping

areas, while singleton was a sequence assembled from processing results that did not have overlapping areas (Palmer et al. 2012). Then the 3,005 contigs generated from the processing stage were identified for their SSR sequence perfect motifs and the length of perfect repeat motifs with the value of 1 to 6 (mono, di-, tri-, tetra-, penta-, and hexa-), so that 548 motifs were obtained distributed over the contig sequences (Table 2). The most frequent motif distributed in the contig sequence was the trinucleotide motif with a frequency of 37.40% and the abundance of the AGG motif. Whereas the least distributed motif was the pentanucleotide motif with a frequency of 5.83% and the abundance of the AAAAG motif. This is in agreement with the results of Joshi et al. (2010) which stated that the trinucleotide motif was the most frequently distributed motif in the *C. longa* contig sequences.

The SSR retrieving frequencies in the *C. longa* sequence obtained from our study was 1/15.27 kb EST sequence, in other word, there was one SSR motif in every 15.27 kb EST sequence. This result was lower than the previous study where one SSR was obtained in every 14.73 kb (Joshi et al. 2010), nevertheless was comparatively higher than the results of Siju et al. (2010) where one SSR was retrieved in every 17.96 kb. The reason for this was because the number of EST sequences in NCBI was increasing every year, causing the results of the calculation of the motif frequency became larger. This result was still much higher when compared to plants from the same family such as ginger (*Zingiber officinale*) which has a frequency of 1/25.21 kb (Awasthi et al. 2017). The frequency of SSR motifs illustrated the number of motifs per kilobyte of sequence. The higher the frequency, the greater the number and variation of motifs in the sequence or it can be said that the sequence has polymorphic characters, and consequently was good for analysis.

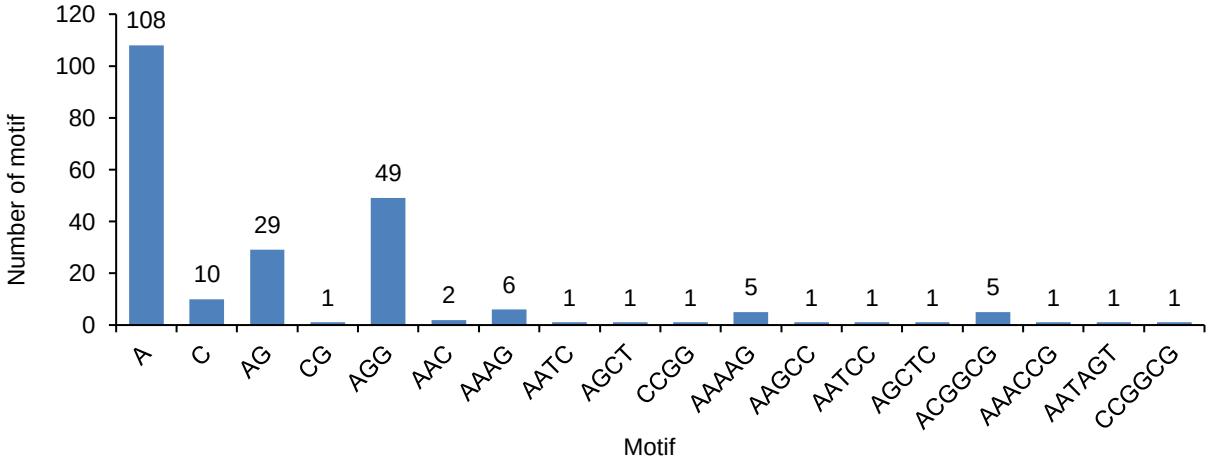

Figure 1. Distribution of the highest and lowest motifs based on the type of motif

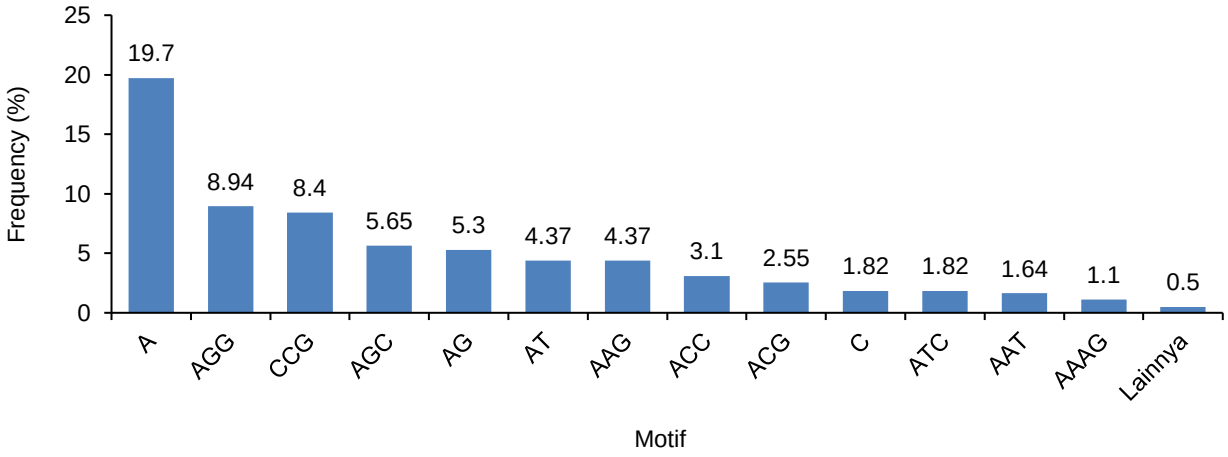

Figure 2. Distribution of SSR motifs based on the number of each motif

Figure 1 exhibits the distribution of the highest and lowest motifs in the contig sequence based on the type of motif. Compared with other types of motifs, the distribution of repetitive mononucleotide motif was the highest, with a total of 108 motifs. Motif C was the lowest distributed with a total of 10 motifs. Different results were obtained by Taheri et al. (2019) on *C. alismatifolia* which showed more distribution of trinucleotide motifs than other motifs. However, since the mononucleotide repeat motif only consists of 1 type of nitrogen base, it could cause failure in the PCR process. Zalapa et al. (2012) stated that the mononucleotide repeat motif has a shorter size and might result in mispairing.

The distribution of SSRs generated from the EST sequences based on repetitive motifs is presented in Figure 2. The motif A was the highest frequency motif with a frequency of 19.7%, followed by the AGG motif with a frequency of 8.94%. Moreover, the CCG, AGC, AG, AT, AAG, ACC, ACG, C, ATC, AAT, AAAG motifs with frequency of 8.4%, 5.65%, 5.3 %, 4.37%, 4.37%, 3.1%, 2.55%, 1.82%, 1.82%, 1.64%, 1.1%, respectively. Data on the frequency of motifs can be used as a reference, especially for primers design and according to Joshi et al. (2010) the data could be used in the future for the genetic diversity study.

SSR motifs can be divided into class I and class II SSR based on the length of the motif sequence. Class I SSR had a motif length of ≥ 20 bp, while SSR class II had a motif length of ≥ 12 bp and < 20 bp (Temnykh et al. 2001). The type of SSR used in our study was a perfect repeat motif. Its means that the sequence has a repetitive motif without being interrupted by any other motifs, for example AGAGAGAGAGAG or (AG)6 or (N1N2)x and CTGTGCTG or (CTG)3 or (N1N2N3)x (Wang et al. 2011). The perfect repeat motif was used because it tends to have a long size sequence. The longer and more perfect the SSR motif, the more varied (de Faria Muller et al. 2014), and the greater the chance to produce polymorphic alleles (Zalapa et al. 2012).

Primer design and analysis

SSR motifs, other than mononucleotides, were selected with a length of ≥20 bp for the primer design. The specifications of the designed primers were the minimum GC content value of 50%, the maximum GC content value of 60%, and the average GC content value of 56.31%. The minimum Tm value was 56.4 ºC, the maximum Tm value was 60 ºC, and the average Tm value was 59.25 ºC. According to Chuang et al. (2013), a suitable primer has a length of 16-28 bp with a difference between reverse and forward primers not more than 3 bp in length, GC content ranging from 40-60%, having a Tm ranging from 50-62 ºC, and lack of dimers or hairpins in its sequence. The size of the primers product that had been successfully designed was in the range of 100-303 bp (Table 3). According to Yustinadewi et al. (2018), primer products that are too short might cause misprimming or primers stick to undesirable areas so that it could affect the efficiency of the PCR process. Accordingly, the primers that were successfully synthesized in our study have met all requirements.

The primers synthesized in this study were sequences derived from motifs other than mononucleotides. Mononucleotide motifs were not used in primer design because they were neither informative nor polymorphic. Choi et al. (2011) emphasized that mononucleotide motifs could cause failure in the PCR process. The purpose of the primer design was to produce proper primers that can be used in the DNA amplification process using the PCR method in order to identify genetic diversity. It is expected that these primers could be applied to the genetic diversity test of *C. aeruginosa* in the next study. After analysis based on gene ontology (GO), 10 primers were predicted to be associated to enzymes involved in metabolic processes.

The results of the analysis showed that there were 6 enzymes, each of which was involved in metabolic pathways in plants. The 13S-lipoxygenase enzyme plays a role in the initiation of the formation of jasmonic acid in plants. This enzyme is produced by the pathogenic fungus *Fusarium sp*. during the invasive growth period to lyse plant cell membranes (Wennman et al. 2015). The major function of jasmonic acid is inhibiting the process of plant growth. It occurs due to the defense responses of injured plant tissue because of fungal attack (Borrego and Kolomiets 2016). This indicated that plants of

Table 3. Results of primer analysis of SSR contig sequences

Contig	Primer Code	Primer Sequences	Product Size	Enzyme
Contig135	192 F	CCTGCAACTCTGGGGTTCA	279	13S-lipoxygenase
	470 R	TCCTGCATCGCCTTGGAT		
Contig192	761 F	TGAACGCCACGCTGATGA	222	Phosphatase and Adenylpyrophosphatase
	982 R	TGTCGGTCCTGAAGATGCC		
Contig393	156 F	ACTCCGCACCCTTCTCTAGT	189	N-acetyltransferase
	344 R	TGGCAGGCATAGCAGGTT		
Contig1948	8 F	CGAAGGTGAGCAGGGCATAT	110	Isomerase
	117 R	GAAAAGGTTCGCTGGGCAG		
Contig2058	1208 F	TGTGGGCTGGTTCTTGTGT	203	Phosphatase
	1410 R	GCACAGAAGGGGGAAAAGG		
Contig2544	489 F	GCTCCGACCACCCATGAAT	198	Synthase
	686 R	TCTGCCCCACAAACCTTCC		
Contig2544	212 F	TGTTGTTCTCGGGAGCCAG	296	Synthase
	507 R	ATTCATGGGTGGTCGGAGC		
Contig2544	668 F	GGAAGGTTTGTGGGGCAGA	101	Synthase
	768 R	TGATTTCTTGGTCGGGGCA		
Contig2544	1551 F	CGCGAGTCTGGGTCTTGAA	141	Synthase
	1691 R	ACGTGAGGTATGGGGAGGA		
Contig2544	212 F	TGTTGTTCTCGGGAGCCA	177	Synthase
	388 R	AGGAGGAGGGAGAAGGCAA		

the genus Curcuma have antifungal activity. The N-acetyltransferase enzyme plays a role in gene activation and cellular activities. The components of polyphenolic bioactive compounds in the form of curcumin contained in the rhizomes of plants of the genus Curcuma work as selective inhibitors of the N-acetyltransferase enzyme that could inhibit the inflammatory response, suppress viral proliferation and the growth of B-cell lymphoma that can cause cancer in lymphatic system (Zhou et al. 2011).

The phosphatase enzyme performs in the process of cell differentiation, one of which is in the process of fetal development. Rhizome extract of plants of the genus Curcuma could reduce the activity of these enzymes, so that they can be used as contraceptive drugs for fertility-regulating methods after being tested on female rats (Yadav and Jain 2011). The adenylpyrophosphatase enzyme plays a role in the process of cellular activities, which regulates and maintains the electrochemical gradient used for secondary active transport across the plasma membrane and regulates plant physiological processes such as stomatal opening and closing mechanism (Elmore and Coaker 2011). The isomerase enzyme is involved in catalyzing the reaction of changes in the configuration of the substrate molecule, so that new molecules are produced with isomeric changes. One of them is protein disulfide isomerase (PDI) which can catalyze the formation of disulfide bonds on substrate proteins that have incorrect disulfide bonds (Fan et al. 2018). Synthase enzymes play a role in the process of biosynthesis of bioactive compounds for instance flavonoids and various polyphenols in plants (Resmi and Soniya 2012), this is related to the content such as curcumin contained in the rhizomes of plants of the genus Curcuma.

Functional analysis and gene ontology (GO)

A total of 190 primer sequences which generated from a total of 380 primers were subjected to BLASTx analysis using the Blast2GO software, the data were then associated with the NCBI database. As the results, 113 out of 190 primer sequences

Table 4. Distribution of contig sequences resulting from Blast2GO analysis

Sequences Analysis Criteria	No. of Sequences
Blast2Go analysis	190
Memiliki gen ontologi	90
Memiliki BLASTx Hit	113
Tanpa BLASTx Hit	77
Berhasil dipetakan	90
Memiliki hit spesies terbanyak *Musa balbisiana*	24

were successfully retrieved. While the rest, 77 primary sequences, did not match any similar sequences to the query in the database, this was presumably because the sequences only carried the specific characteristics of the plant. The resulting parameters include the highest BLAST Hit value of 20 and the lowest BLAST Hit value of 1, the minimum similarity value on 113 primer sequences was 54.73%, the maximum similarity value was 100% and the average similarity value was 89.01. In addition, from 113 primer sequences, 90 sequences have gene ontologies that can be mapped according to three categories, namely biological processes, molecular functions, and cellular components (Table 4).

The Blast method is used to align a sequence with a homologous sequence using a data-based search mainly for identification purposes (Yang et al. 2014). The results obtained from our study indicated a high level of sequence homology which is known from the average similarity value of 89.01%. The average value of sequence similarity which greater than 60% indicated a high level of homology (Wee and Roslan 2012). Additionally from the BLAST results, it was known the e-value, which forms a homologous sequence or Hits which describes the match between the sequences.

BLAST analysis resulted from 190 sequences to species in the NCBI database obtained the information on the suitability of several plant species (Figure 3) consisting of 11 monocot plants (*Musa balbisiana*, *M. acuminata* subsp. *Malaccensis*, *Elaesis guineensis*, *Ensente ventricosum*, *Carex littledalei*, *Apostasia shenzhenica*, *Phalaenopsis equestris*, *Triticum aestivum*, *Oryza sativa* Japonica Group, *Asparagus officinalis*, *Sorghum bicolor*, *Triticum turgidum* subsp. Durum) and 14 dicot plants (*Mucuna pruriens*, *Helianthus annuus*, *Hibiscus syriacus*, *Cynara cardunculus var. scolymus, Dimocarpus longan, Hevea brasiliensis, Momordica charantia, Mikania micrantha, Arabidopsis thaliana, Jatropha curcas, Syzygium oleosum, Ziziphus jujube, Sesamum indicum, Macleaya cordata*). Sequences of *C. longa* were most widely distributed in dicotyledonous plants. This finding was in accordance with the results of Rai et al. (2018) which found that the EST sequences of *C. longa* were most widely distributed in dicotyledonous plants (9 monocotyledons and 18 dicotyledonous plants). Based on sequence analysis, the highest hits were obtained in *Musa balbisiana* (24) and *M. accuminata* subsp. *Malaccensis*

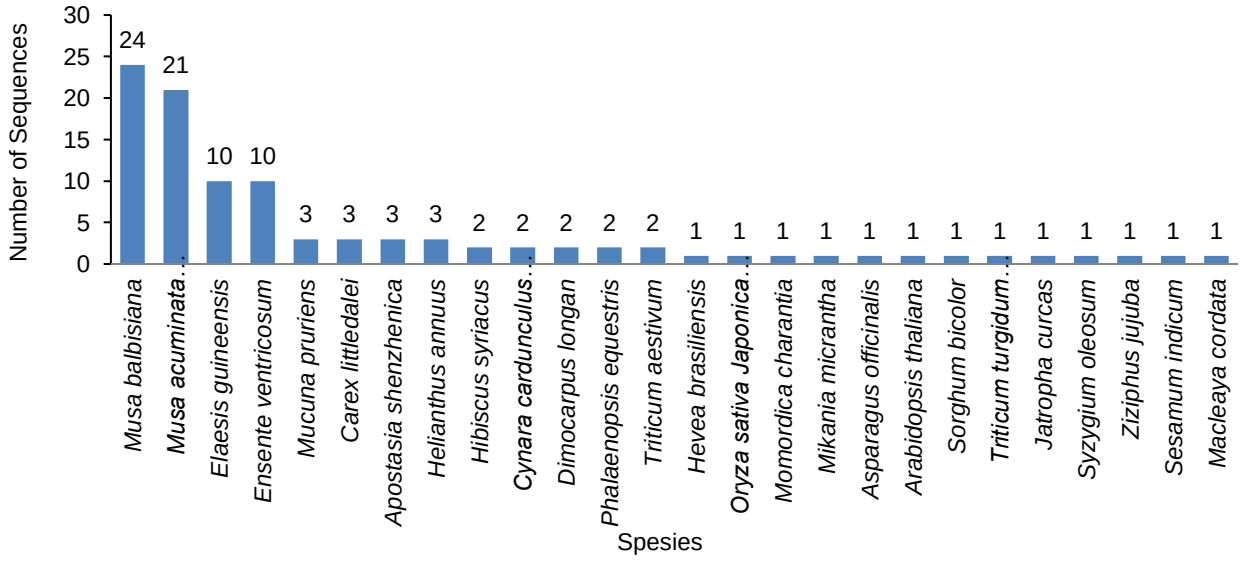

Figure 3. The frequency of 26 plant species with the most hits

(21) belonging to the order Zingiberales. This indicates that the primer synthesized from the EST sequence of *C. longa* can also be used in both plants.

The analysis results of gene ontology and functional categories obtained in our study is revealed in Figure 4. Based on the sequence annotation, it was known that the distribution of gene ontologies in three categories comprising of molecular function categories with 39 gene ontologies, biological processes with 58 gene ontologies and cellular components with 29 gene ontologies. According to Chang and Scharfenstein (2014) molecular function is a category that describes the activity of gene products at the molecular level, while biological processes is a category that describes molecular activity from beginning to end. Likewise, cellular components are category that indicates cell parts and their extracellular environment. The gene ontology analysis is important to determine the expressed genes in the

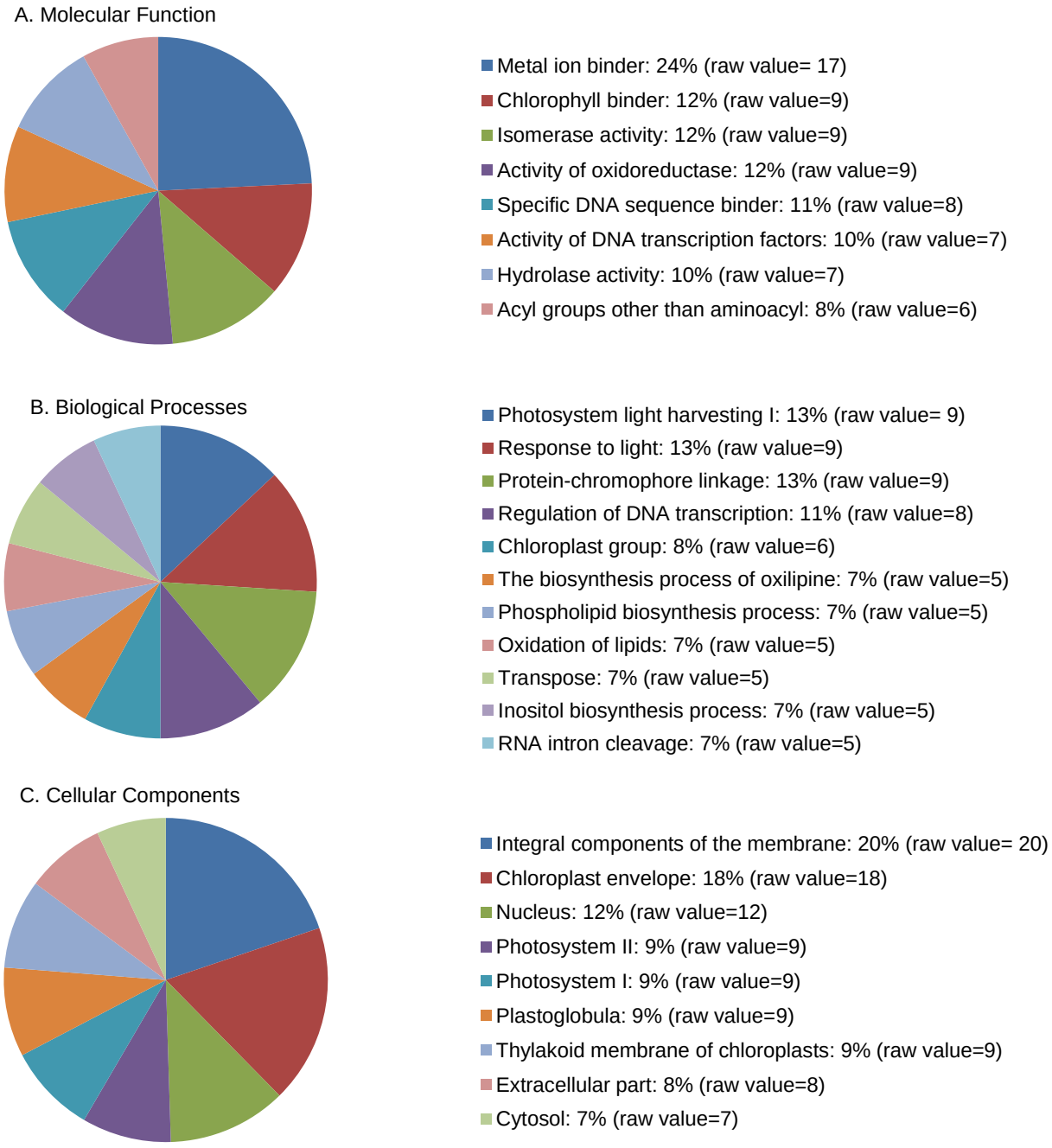

Figure 4. Functional grouping of gene ontology: (A). Molecular function; (B). Biological processes; and (C); Cellular

sequence and their presence in cell organelles.

Based on the functional classification of gene ontology, it was known that the molecular function category (Figure 4A) formed 8 functional groups in which genes associated to metal ion binding were mostly found (24%). Meanwhile, relatively few genes linked to hydrolase activity and acyl groups were found ≤10%. The binding of metal ions is the mechanism of action of antioxidant compounds to prevent the formation of reactive oxygen species (ROS), where antioxidants can be found in food, one of which is in the form of flavonoids (Silvia et al. 2016). Flavonoid in the form of curcumin is one of the compounds contained in the rhizome of plants from the genus Curcuma such as turmeric and temu ireng so that it can be seen that the rhizome of the Curcuma plant contains antioxidants.

In the category of biological processes (Figure 4B), 11 functional groups were formed which dominated by genes associated to the photosynthesis process and plant response to light (13%). In addition, genes linked to the biosynthetic process of oxylipin were also revealed. Oxylipin is a compound formed from unsaturated fatty acids that play a role in the mechanism of fungal

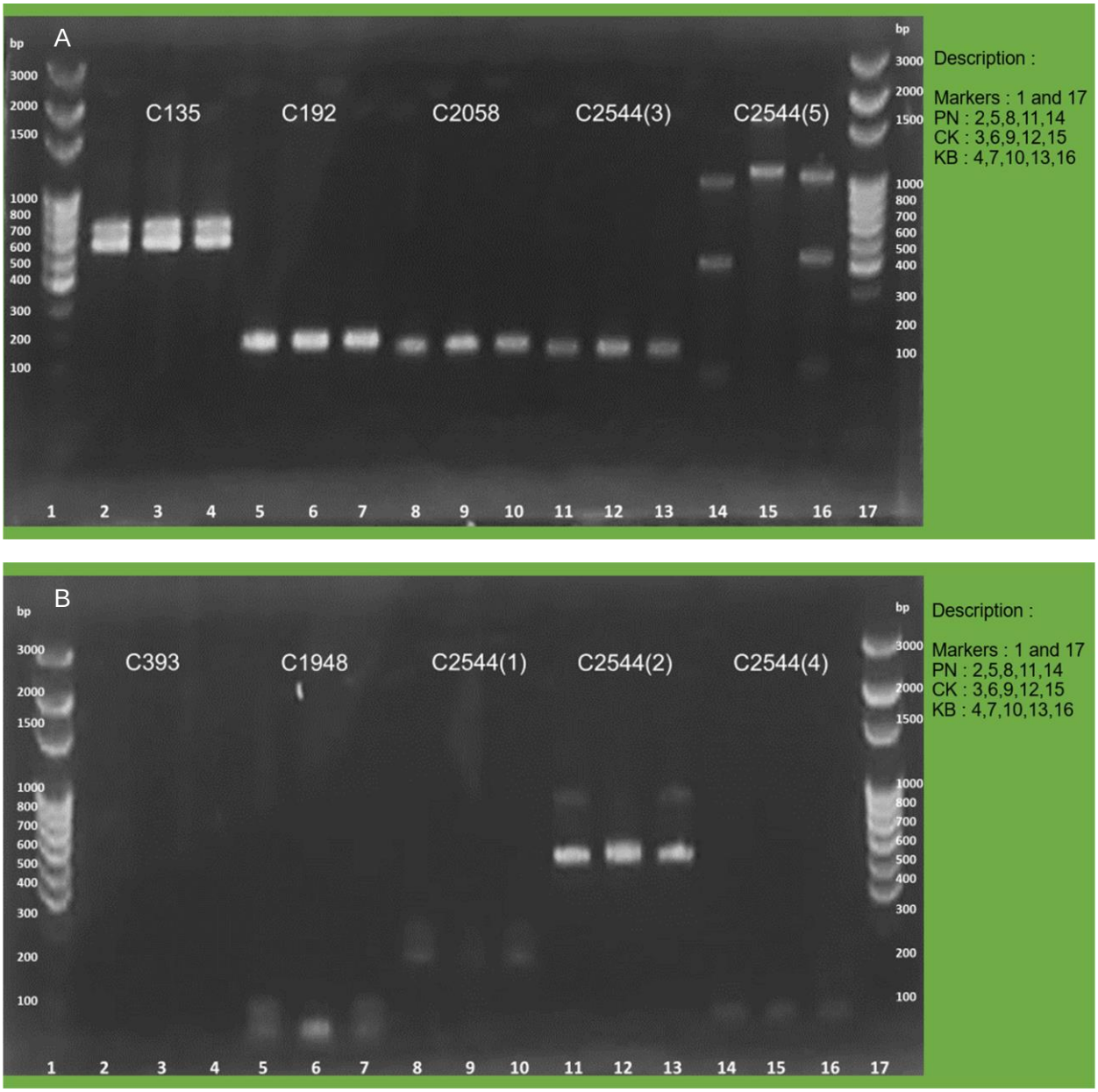

Figure 5. The results of the validation amplification of 10 primers: (A). Annealing temperature 53 ºC; and (B). Annealing temperature 55 ºC

pathogenicity in plants. One of its derivative compounds is jasmonic acid which acts as a regulator of defense mechanisms against pathogens and has antifungal properties (Wasternack and Feussner 2018). Based on this, it can be seen that *C. longa* has antifungal properties. Chen et al. (2018) confirmed that *C. longa* has antifungal abilities that could damage the cell membrane of the fungus *Fusarium graminearum*.

The category of cellular components (Figure 4C) consisted of 9 functional groups with membrane integral components (20%) being the most dominating. Several genes are also known to be present in the Photosystems I and II, plastoglobula and thylakoid membranes of chloroplasts (9%). A small number of genes were located in the extracellular and cytosol at ≤8%. Based on the results of the sequence analysis on the three functional categories, it showed that there were sequences that did not match to any sequences in the NCBI database. It is presumably that these sequences contain genes that are only found in *C. longa* and have not been identified yet. While sequences that match with BLAST Hit indicated that the two sequences have similar traits or gene matches to each other.

Primer validation

The SSR primer capability generated from the *C. longa* EST data needs to be validated. Primer validation was carried out to find out whether the designed primers could be amplified. The primer ability to reveal amplification products is influenced by several factors, namely internal stability, melting temperature, secondary structure or competition among primers (Sint et al. 2012). Validation of 10 SSR primers was conducted using 3 DNA samples of *C. longa* (Figure 5). Subsequently, all primers were successfully amplified by showing bands, except the C393 primer. Primers of C1948, C2544(2), and C2544(5) resulted in polymorphic bands while the other primers were monomorphic.

CONCLUSION

The development of SSR primers and markers can be carried out by utilizing the EST sequences from plants belonging to the same genus. In our study, there was SSR diversity in the EST sequence of *C. longa* which was dominated by the trinucleotide motif of AGG. The primers constructed have met the standards and it was known that as many as 10 pairs of primers were associated to several enzymes that contributed to metabolic processes. The C1948, C2544(2) and C2544(5) primers produced polymorphic bands, while the others were monomorphic. Sequence analysis using BLASTx and comparing to NCBI database revealed similarities with *Musa balbisiana* and *M. accuminata* and the sequences had genes expressing antioxidant and antifungal properties.

REFERENCES

Atun S, Arianingrum R, Aznam N, Malek SNA (2016) Isolation of sesquiterpenes lactone from *Curcuma aeruginosa* rhizome and the cytotoxic activity against human cancer cell lines, Int J Pharmacogn Phytochem Res 8: 1168–1172

Awasthi P, Singh A, Sheikh G, Mahajan V, Gupta AP, Gupta S, Bedi YS, Gandhi SG (2017) Mining and characterization of EST-SSR markers for *Zingiber officinale* Roscoe with transferability to other species of *Zingiberaceae*. Physiol Mol Biol Plants 23: 925–931. doi: 10.1007/s12298-017-0472-5

Bare Y, Maulidi A, Sari DRT, Tiring SSND (2019) Studi in silico prediksi potensi 6-gingerol sebagai inhibitor c-Jun N-terminal kinases (JNK). J Jejaring Mat Sains 1: 59–63. doi: 10.36873/jjms.v1i2.211

Borrego EJ, Kolomiets MV (2016) Synthesis and functions of jasmonates in maize. Plants 5: 41. doi: 10.3390/plants5040041

Chang P-K, Scharfenstein LL (2014) *Aspergillus flavus* Blast2GO gene ontology database: Elevated growth temperature alters amino acid metabolism. J Genet Genome Res 1: 005. doi: 10.23937/2378-3648/1410005

Chen C, Long L, Zhang F, Chen Q, Chen C, Yu X, Liu Q, Bao J, Long Z (2018) Antifungal activity, main active components and mechanism of *Curcuma longa* extract against *Fusarium graminearum*. PLoS One 13:

e0194284. doi: 10.1371/journal.pone.0194284

Choi H-l, Kim NH, Kim JH, Choi BS, Ahn IO, Lee JS, Yang TJ (2011) Development of reproducible EST-derived SSR markers and assessment of genetic diversity in *Panax ginseng* cultivars and related species. J Ginseng Res 35: 399–412. doi: 10.5142/jgr.2011.35.4.399

Chuang LY, Cheng YH, Yang CH (2013) Specific primer design for the polymerase chain reaction. Biotechnol Lett 35: 1541–1549. doi: 10.1007/s10529-013-1249-8

de Faria Muller BS, Sakamoto T, de Menezes IPP, Prado GS, Martins WS, Brondani C, de Barros EG, Vianello RP (2014) Analysis of BAC-end sequences in common bean (*Phaseolus vulagaris* L.) towards the development and characterization of long motifs SSRs. Plant Mol Biol 86: 455–470. doi: 10.1007/s11103-014-0240-7

Elmore JM, Coaker G (2011) The role of the plasma membrane H^+-ATPase in plant-microbe interactions. Mol Plant 4: 416–427. doi: 10.1093/mp/ssq083

Fan F, Zhang Y, Wang S, Han Y, Wang L, Lu D (2018) Characterization of the oxidative protein folding activity of a unique plant oxidoreductase, Arabidopsis protein disulfide isomerase-11. Biochem Biophys Res Commun 495: 1041–1047. doi: 10.1016/j.bbrc.2017.11.111

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36: 3420–3435. doi: 10.1093/nar/gkn176

Hadiarto T, Listanto E, Riyanti EI (2015) Identifikasi cDNA Gen RB pada tanaman kentang produk rekayasa genetika Katahdin SP951. J AgroBiogen 11: 59–64. doi: 10.21082/jbio.v11n2.2015.p59-64

Izzah NK, Lee J, Jayakodi M, Perumal S, Jin M, Park B-S, Ahn K, Yang T-J. 2014. Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. capitata L.) and construction of an EST-based genetic map. BMC Genomics 15: 149. doi: 10.1186/1471-2164-15-149

Jose S, Thomas TD (2014) Comparative phytochemical and anti-bacterial studies of two indigenous medicinal plants *Curcuma caesia* Roxb. and *Curcuma aeruginosa* Roxb. Int J Green Pharm 8: 65–71. doi: 10.4103/0973-8258.126828

Joshi RK, Kuanar A, Mohanty S, Subudhi E, Nayak S (2010) Mining and characterization of EST derived microsatellites in *Curcuma longa* L. Bioinformation 5: 128–131. doi: 10.6026/97320630005128

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: An overview of the recent progress in plants. Euphytica 177: 309–334. doi: 10.1007/s10681-010-0286-9

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30. doi: 10.1093/nar/28.1.27

Kaur R, Vaidya E (2015) Analysis of frequency and functional annotation of SSRs in ESTs of Cole Crops. J Appl Life Sci Int 2: 189–196. doi: 10.9734/jalsi/2015/11873

Khumaida N, Syukur M, Bintang M, Nurcholis W (2019) Phenolic and flavonoid content in ethanol extract and agromorphological diversity of *Curcuma aeruginosa* accessions growing in West Java, Indonesia. Biodiversitas 20: 656-663. Doi: 10.13057/biodiv/d200306

Lan TTP, Huy ND, Luong NN, Nghi NV, Tan TH, Quan LV, Loc NH (2018) Identification and characterization of genes in the curcuminoid pathway of *Curcuma zedoaria* Roscoe. Curr Pharm Biotechnol 19: 839-846. doi: 10.2174/1389201019666181008112244

Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S (2006) EGassembler: Online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res 34: W459–462. doi: 10.1093/nar/gkl066

Moektiwardoyo M, Tjitraresmi A, Susilawati Y, Iskandar Y, Halimah E, Zahryanti D

(2014) The potential of dewa leaves *(Gynura Pseudochina* (L) D.C*)* and temu ireng rhizomes *(Curcuma aeruginosa* Roxb.*)* as medicinal herbs for dengue fever treatment. Procedia Chem 13: 134–141. doi: 10.1016/j.proche.2014.12.017

Palmer NA, Saathoff AJ, Kim J, Benson A, Tobias CM, Twigg P, Vogel KP, Madhavan S, Sarath G (2012) Next-generation sequencing of crown and rhizome transcriptome from an upland, tetraploid switchgrass. Bioenerg Res 5: 649–661. doi: 10.1007/s12155-011-9171-1

Pan L, Huang T, Yang Z, Tang L, Cheng Y, Wang J, Ma X, Zhang X (2018) EST-SSR marker characterization based on RNA-sequencing of *Lolium multiflorum* and cross transferability to related species. Mol Breeding 38: 80 doi: 10.1007/s11032-018-0775-4

Rachmanita N (2015) The effect of ethanolic extract of temu ireng (*Curcuma aeruginosa* Roxb) on breast tumor growth and tumor histophatology on white rats induced by 7,12-dimetilbenz(α)anthracene [tesis]. Depok (ID): Indonesia University

Rai PS, Chaudhary S, Sharma PC (2018) Expressed sequence tags (ESTs)–based computational identification of novel and conserved microRNAs in turmeric *(Curcuma longa* L.*)*. J Appl Biotechnol Bioeng 5: 108–115. doi: 10.15406/jabb.2018.05.00125

Resmi MS, Soniya EV (2012) Molecular cloning and differential expressions of two cDNA encoding type III polyketide synthase in different tissues of *Curcuma longa* L. Gene 491: 278–283. doi: 10.1016/j.gene.2011.09.025

Sahoo A, Jena S, Kar B, Sahoo S, Ray A, Singh S, Joshi RK, Acharya L, Nayak S (2017) EST-SSR marker revealed effective over biochemical and morphological scepticism towards identification of specific turmeric (*Curcuma longa* L.) cultivars. 3 Biotech 7: 84. doi: 10.1007/s13205-017-0701-1

Sakthipriya M, Sabu KK (2018) Development and cross-genera transferability of ginger EST-SSR markers for cardamom. Curr Bioinform 13: 95-99. doi: 10.2174/1574893611666161128123827

Selvakumari E, Jenifer J, Priyadharshini S, Vinodhini R (2017) Application of DNA fingerprinting for plant identification. J Acad Ind Res 5: 149–151. Corpus ID: 41401879

Sen S, Dehury B, Sahu J, Rathi S, Yadav RNS (2017) *In silico* mining and characterization of simple sequence repeats (SSRs) from *Euphorbia esula* expressed sequence tags (ESTs): A potential crop for biofuel. Plant Omics J 10: 53–63. doi: 10.21475/poj.10.02.17.pne370

Septaningsih DA, Darusman LK, Afendi FM, Heryanto R (2018) Liquid chromatography mass spectrometry (LC-MS) fingerprint combined with chemometrics for identification of metabolites content and biological activities of *Curcuma aeruginosa*. Indones J Chem 18: 43-52. doi: 10.22146/ijc.25456

Setiyono A, Bermawie N (2013) Potensi tanaman obat untuk penanggulangan flu burung: Uji in vitro pada sel vero. J Sains Veteriner 31: 27–34.

Siju S, Dhanya K, Syamkumar S, Sasikumar B, Sheeja TE, Bhat AI, Parthasarathy VA (2010) Development, characterization and cross species amplification of polymorphic microsatellite markers from expressed sequence tags of turmeric (*Curcuma longa* L.). Mol Biotechnol 44: 140–147. doi: 10.1007/s12033-009-9222-4

Silvia D, Katharina K, Hartono SA, Anastasia V, Susanto Y (2016) Pengumpulan data base sumber antioksidan alami alternatif berbasis pangan lokal di Indonesia. Surya Octagon Interdisciplinary J Technol 1: 181–198

Simoh S, Zainal A (2015) Chemical profiling of *Curcuma aeruginosa* Roxb. rhizome using different techniques of solvent extraction. Asian Pac J Trop Biomed 5: 412-417. doi: 10.1016/S2221-1691(15)30378-6

Sint D, Raso L, Traugott M. (2012) Advances in multiplex PCR: Balancing primer efficiencies and improving detection success. Methods Ecol Evol 3: 898–905. doi: 10.1111/j.2041-210X.2012.00215.x

Sofyan FF, Supriyatna, Moektiwardoyo M (2013) Peningkatan sikap positif

masyarakat dalam pemanfaatan tanaman obat pekarangan rumah di Desa Sukamaju Dan Girijaya Kabupaten Garut. Dharmakarya 2: 107–117. doi: 10.24198/dharmakarya.v2i2.8223

Suphrom N, Pumthong G, Khorana N, Waranuch N, Limpeanchob N, Ingkaninan K (2012) Anti-androgenic effect of sesquiterpenes isolated from the rhizomes of *Curcuma aeruginosa* Roxb. Fitoterapia 83: 864-871. doi: 10.1016/j.fitote.2012.03.017

Taheri S, Abdullah TL, Rafii MY, Harikrishna JA, Werbrouck SPO, Teo CH, Sahebi M, Azizi P (2019) *De novo* assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. Sci Rep 9**:** 3047. doi: 10.1038/s41598-019-39

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11: 1441–1452. doi: 10.1101/gr.184001

Terryana RT, Nugroho K, Rijzaani H, Lestari P (2018) Karakterisasi keragaman genetik 27 genotipe cabai berdasarkan marka SSR (Simple Sequence Repeat). Berita Biol 17: 183–194. doi: 10.14203/beritabiologi.v17i2.3313

Victoria FC, da Maia LC, de Oliveira AC (2011) *In silico* comparative analysis of SSR markers in plants. BMC Plant Biol 11**:** 15. doi: 10.1186/1471-2229-11-15

Wang YW, Samuels TD, Wu YQ (2011) Development of 1,030 genomic SSR markers in switchgrass. Theor Appl Genet 122: 677–686. doi: 10.1007/s00122-010-1477-4

Wasternack C, Feussner I (2018) The oxylipin pathways: Biochemistry and function. Annu Rev Plant Biol 69: 363-386. doi: 10.1146/annurev-arplant-042817-040440

Wee CC, Roslan HA (2012) Expressed sequence tags (ESTs) from young leaves of *Metroxylon sagu*. 3 Biotech 2: 211–218. doi: 10.1007/s13205-012-0048-6

Wennman A, Magnuson A, Hamberg M, Oliw EH (2015) Manganese lipoxygenase of *F. oxysporum* and the structural basis for biosynthesis of distinct 11-hydroperoxy stereoisomers. J Lipid Res 56: 1606–1615. doi: 10.1194/jlr.M060178

Yadav R, Jain GC (2011) Effect of contragestative dose of aqueous extract of *Curcuma longa* rhizome on uterine biochemical milieu of female rats. Indian J Fundam Appl Life Sci 1: 183–187. Corpus ID: 2029434

Yan Z, Wu F, Luo K, Zhao Y, Yan Q, Zhang Y, Wang Y, Zhang J (2017) Cross-species transferability of EST-SSR markers developed from the transcriptome of *Melilotus* and their application to population genetics research. Sci Rep 7: 17959. doi: 10.1038/s41598-017-18049-8

Yang Y, Jiang XT, Zhang T (2014) Evaluation of a hybrid approach using UBLAST and BLASTX for metagenomic sequences annotation of specific functional genes. PLoS One 9: e110947. doi: 10.1371/journal.pone.0110947

You Y, Liu D, Liu H, Zheng X, Diao Y, Huang X, Hu Z (2015) Development and characterisation of EST-SSR markers by transcriptome sequencing in taro (*Colocasia esculenta* (L.) Schoot). Mol Breeding 35: 134. doi: 10.1007/s11032-015-0307-4

Yustinadewi PD, Yustiantara PS, Narayani I (2018) MDR-1 gene 1199 variant primer design techniques in pediatric patient buffy coat samples with LLA.J Metamorfosa 5: 105–111. doi: 10.24843/metamorfosa.2018.v05.i01.p16

Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99: 193–208. doi: 10.3732/ajb.1100394

Zhou H, Beevers CS, Huang S (2011) The targets of curcumin. Curr Drug Targets 12: 332–347. doi: 10.2174/138945011794815356