In silico Identification and Comparative Analysis of *Hevea brasiliensis* COBRA Gene Family

Riza Arief Putranto^{1*}, Irfan Martiansyah¹, Rizka Tamania Saptari^{1,2}

¹Indonesian Research Institute for Biotechnology and Bioindustry, Jl. Taman Kencana No.1 Bogor 16128, Indonesia. Tel. +62-251- 8324048, Fax. +62-251- 8328516. ²School of Life Sciences and Technology – ITB, Jl. Ganesa 10, Lebak Siliwangi, Coblong, Bandung, Indonesia 40132. Tel. +62-22- 2511575. *Email: rizaputranto@iribb.org

Abstract. *Putranto R A, Martiansyah I, Saptari R T. 2017. In silico Identification and Comparative Analysis of Hevea brasiliensis COBRA Gene Family. Proc Internat Conf Sci Engin 1: 39-47.* In this paper, the *H. brasiliensis* COBRA gene family, alleged to be involved in laticifer differentiation, was identified from the public rubber tree genome of Reyan 7-33-97 clone. A comparative analysis was carried out against A. thaliana genomic database. This analysis has resulted to the in silico validation of thirteen putative genes encoding glycophosphatidylinositol anchors (GPI) proteins harbored by nine *Hevea* genomic scaffolds. The sequence's similarity of *HbCOBL* against *AtCOBL* genes were ranged from the threshold 50 to 81.58% covering 151 to 458 amino acid residues, respectively. Three partial and ten fullength protein sequences of *HbCOBL* genes were annotated. The partial protein sequences ranged from 89 to 184 amino acid residues as opposed to the full-length proteins ranging from 160 to 471 amino acid residues. Two types of COBRA domains (pfam04833 and cl04787) were found among *HbCOBL* genes. Phylogenetic analysis has clustered two subfamilies. Nine *HbCOBL* genes (*HbCOBL-B, HbCOBL-J, HbCOBL-A, and HbCOBL-A, and HbCOBL-N*) were clustered as COBRA gene subfamily-I. By contrast, four genes (*HbCOBL-O, HbCOBL-P, HbCOBL-E, and HbCOBL-L*) were clustered as COBRA gene subfamily-II. The *HbCOB* subfamily-II was marked by the addition of 203 residues in C-terminal which is different with *Arabidopsis*. The gene *HbCOBL-C* was the putative ortholog to *AtCOB* carrying the unique COBRA domain cl04787 with 74 amino acid residues. Taken together, these results showed that *Hevea* and *Arabidopsis COBRA* genes might share similar functions while differ in gene structure.

Keywords: bioinformatics, COBRA, differentiation, laticifer, rubber tree

Abbreviations: glycophosphatidylinositol anchors (GPI); APETALA2/ETHYLENE RESPONSE FACTORS (AP2/ERF); Calcium-Dependent Protein Kinase (CDPK); Abscisic acid (ABA); Glucose-6-Phospate Dehydrogenase (G6PDH); Small Rubber Particle (SRPP); COBRA-like (COBL); European Nucleotide Archive (ENA); coding sequence (CDS); Conserve Domain Database Search (CDD); John-Taylor-Thornton (JTT); reads per kilobase per million mapped reads (RPKM)

INTRODUCTION

In regard to the importance of para rubber tree (Hevea brasiliensis Müll.Arg.) as the sole commercial source of natural rubber, studies have been carried out to identify the potential factors influencing the high production of latex. Arabidopsis thaliana (L.) Heynh is an important and well-described model plant for molecular biology due to its advantages of short life cycle (6 weeks) and small genome size (~125Mb) with five chromosomes (Initiative, 2000). The in silico comparative analysis between H. brasiliensis and A. thaliana have successfully identified important gene families such as the APETALA2/ETHYLENE RESPONSE FACTORS (AP2/ERF) superfamily (Duan et al., 2010; Piyatrakul et al., 2014; Putranto et al., 2015a; Putranto et al., 2015b; Putranto and Montoro, 2016), Calcium-Dependent Protein Kinase (CDPK) family (Xiao et al., 2017), Abscisic acid (ABA) receptor gene family (Guo et al., 2017), Glucose-6-Phospate Dehydrogenase (G6PDH) family (Long et al., 2016), Sucrose synthase gene family (Xiao et al., 2014), Metacaspase family (Liu et al., 2016), Aquaporin family (Zou et al., 2015), ABC transporter family (Zhiyi et al., 2015), Small Rubber Particle (SRPP) gene family (Lau et al., 2016) and

Protease Inhibitors (PI) family (Martiansyah et al., 2017).

Latex, a milky white containing rubber particle, is synthesized inside specialized latex cells namely laticifers. Latex is harvested by tapping the soft bark of Hevea which truncates the secondary laticifers. The secondary laticifers are differentiated tissues in the secondary phloem of rubber bark. Their abundant numbers were positively correlated with the natural rubber yield (Hao and Wu, 2000). Comprehensive researches have confirmed that the differentiation of secondary laticifers is pivotally regulated by jasmonate (Laosombut et al., 2016; Pirrello et al., 2014; Tian et al., 2015; Wu et al., 2016; Zhang et al., 2015). In the biological knowledge, laticifer differentiation should potentially involve a mechanism of cell expansion and/or wood biosynthesis. Studies on specific gene or gene family involved in such mechanism on rubber tree were limited. However, the transcript of a small gene family namely COBRA was found in the laticifers (Tang et al., 2016). This result suggested that this gene family could be related to the laticifer differentiation.

The COBRA multi-gene family has been intensively studied mostly in *Arabidopsis*, rice and maize (Brady et al., 2007). This family consists of one COBRA gene and eleven homologs namely COBRA-like (COBL) genes (Roudier et al., 2002; Schindelman et al., 2001). The COBRA genes encode glycosylphosphatidylinositol (GPI) anchoring of proteins divided into two subfamilies (Borner et al., 2002; Roudier et al., 2002). Each member of the gene family has an important and redundant role in oriented cell division and cell expansion (Brady et al., 2007). Further analysis has demonstrated their positive expression to induce cellulose synthesis (Ben-Tov et al., 2015; Dai et al., 2011; Liu et al., 2013).

This paper sets out to identify the *H. brasiliensis* COBRA gene family by performing a comparative analysis against *A. thaliana* genomic database. Multiple sequence comparison and BLAST using public platform Galaxy were carried out to identify orthologous sequences between both species. Phylogenetic analysis was constructed to identify two subfamilies of COBRA genes. This work uses the scaffolds of *H. brasiliensis* public genome published by Tang et al. (2016).

MATERIALS AND METHODS

Preparation of Genome Reference Database and Galaxy BLAST

The rubber tree genome of Reyan 7-33-97 with accession number: LVXX01000000 was downloaded from the European Nucleotide Archive (ENA) database (www.ebi.ac.uk/ena). The length of the genome was 1.46 Gb consisting of 7,453 scaffolds; 84,285 contigs and 43,792 genes (Tang et al., 2016 5079). Total genomic scaffolds were stored in the Southern France bioinformatics Galaxy platform (http://galaxy.southgreen.fr/galaxy/). Twelve nucleotide and amino acid sequences of A. thaliana COBRA genes encoding GPI proteins were downloaded from the TAIR database v10 (http://www.arabidopsis.org/). These sequences were used as a comparator for the identification of COBRA gene family in H. brasiliensis. As a model plant in molecular biology, the genome database of A. thaliana has full annotation and demonstrated functions (Piyatrakul et al., 2014; Putranto et al., 2015a). An NCBI MEGA-BLAST + tblastn of A. thaliana COBRA gene family against H. brasiliensis genomic scaffolds was carried out using Galaxy (Cock et al., 2015). The expectation value cutoff was set to 0.001 using scoring matrix of BLOSUM62. The results of tblastn were sorted to ensure no duplicated genes noted. The selection was based on the parameters such as: (1) hits sharing >50% of sequence similarity with minimum 150 bp length, (2) one unique gene for each location in the scaffold, (3) one scaffold hosting more than one gene in the different location, and (4) low evalue. The final selection of scaffolds was separated.

Manual Annotation and Conserve Domain Analysis of *Hevea* Genomic Scaffolds

Manual annotations of *Hevea COBRA* genes were carried out on selected scaffolds encoding GPI proteins using Geneious software version 5.3.6 (Biomatters Ltd,

USA). The annotation included the non-translated region (5' and 3'-UTR), coding region or sequence (CDS), and protein or domain motifs (Kearse et al., 2012; Martiansyah et al., 2017). The putative CDS region for each Hevea COBRA genes was verified using BLAST-n in the NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The analysis of conserve domain in CDS region followed the methodology of (Putranto et al., 2015a). Each scaffold containing COBRA genes was screened using NCBI Conserve Domain Database Search (CDD) (www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Every COBRA gene should carry a protein domain namely "COBRA" with accession number: pfam04833 and cl04787, respectively. The length of COBRA domain was between 70 to 180 amino acid residues (Roudier et al., 2002).

Comparative Phylogenetic Analysis of *Hevea* and *Arabidopsis* COBRA Gene Family

Phylogenetic analysis was carried out to determine the orthology and paralogy of putative Hevea COBRA gene family encoding the GPI proteins. The comparative analysis was performed on 13 sequences of amino acid residues Hevea COBRA genes with 11 sequences of Arabidopsis COBRA genes. The phylogenetic tree was assembled using the Neighbor-Joining algorithm in the MEGA 7.0 program (Pennsylvania, USA) (Kumar et al., 2016). Bootstrap for the consensus tree was made as many as 1000 repetitions (Felsenstein, 1985). The missing data (gap) was treated using pairwise deletion. The evolutionary distances were computed using the John-Taylor-Thornton (JTT) matrix-based method and are in the units of the number of amino acid substitutions per site (Jones et al., 1992). The gene Arabidopsis Auxin Response Factor, AtARF01.2 was used as the out-group. An additional multiple sequence alignment was carried out using MUSCLE algorithm to verify domain or protein motifs (Edgar, 2004).

Data Analysis and Presentation

The workflow analysis was repeated three times to verify the results. Phylogenetic tree analysis was repeated 15 times to ensure minimum software and human error despite the high number of bootstraps (1000 repetitions). Final data was presented in recapitulative tables and figures.

RESULTS AND DISCUSSION

Workflow of *in silico* comparative analysis using *H*. *brasiliensis* and *A. thaliana genome* database

The *H. brasiliensis* genome of the Reyan clone 7-33-97 has been successfully published (Tang et al., 2016). This genome reference was selected because the data provided a high-quality assembly covering 93.8% of the genome (1.37 out of 1.47 Gb) with 7,453 scaffolds (1.28 Mb of scaffold N50) that harbors 43,792 verified genes. The Scheme of *in silico* comparative analysis for the identification of putative COBRA gene family encoding

GPI proteins in *H. brasiliensis* genome was presented in Figure 1.

Nine *Hevea* genomic scaffolds potentially encoding GPI proteins (LVXX01000065, LVXX01000183, LVXX01000195, LVXX01000520, LVXX01000625, LVXX01000656, LVXX01001114, LVXX01001412, and LVXX01003569) were selected after a tblastn analysis between the target and comparator sequences. The length of the scaffolds ranged from 2.714 to 12 Mb

(Table 1). These scaffolds were successfully annotated and resulted in 17 amino acid residues encoding GPI proteins. Thirteen COBRA domains out of these 17 amino acid residues were putatively discovered. A phylogenetic tree analysis was further carried out of 13 amino acid residues of *HbCOBLs* against 11 *AtCOBLs*. One sequence of *AtCOBL-3* (F14F8.10) has not yet clearly described in TAIR.



Figure 1. Scheme of *in silico* comparative analysis for the identification of putative COBRA gene family encoding glycosylphosphatidylinositol protein (GPI) in *H. brasiliensis* genome. EMBL-EBI, NCBI, Galaxy, Geneious, and MEGA 7.0 were used in this analysis.

Identification and Annotation of *H. brasiliensis* COBRA Gene Family

Seventeen putative encoding COBRA genes (HbCOBL-A HbCOBL-O) were identified using in silico to comparative analysis against Arabidopsis COBRA accession numbers (Table 1). These putative HbCOBL genes were harbored by nine H. brasiliensis genomic scaffolds identified in previous analysis. The sequence's similarity of HbCOBL genes against AtCOBL genes were ranged from the threshold 50 to 81.58% covering 151 to 458 amino acid residues, respectively. The gene HbCOBL-N had the shortest amino acid residues while two genes (HbCOBL-O and HbCOBL-P) had the longest amino acid residues. The results of tblastn showed a low E-value ensuring the confidence of the analysis. Three partial and ten full-length protein sequences of HbCOBL genes among the thirteen sequences were confirmed.

The partial protein sequences ranged from 89 to 184 amino acid residues. Meanwhile, the full-length protein sequences ranged from 160 to 471 amino acid residues. Thirteen HbCOBL proteins harbored COBRA domains with two accession types (pfam04833 and cl04787). The length of COBRA domain cl04787 was 74 amino acid residues. By contrast, the dominant number of COBRA domain pfam04833 ranged from 115 to 180 amino acid residues.

	Tang's Scaffold			Galaxy tblastn*			Protein sequence		COBRA domain			
Gene name	Scaffold ID	Length (Mb)	Arabidopsis ID	Similarity (%)	Length (aa)	E-value	Annotated	Length (aa)	CDD Search	Accession	Length (aa)	Reference
HbCOBL-A	LVXX01000065	2,714	AT1G09790.1	53.89	167	7.00E-59	Partial	165	YES	pfam04833	151	(Lalanne et al., 2004)
HbCOBL-B			AT3G29810.1	61.84	228	2.00E-91	Full-length	164	YES	pfam04833	115	(Ben-Tov et al., 2015)
HbCOBL-C			AT5G60920.1	50.33	153	7.00E-36	Partial	89	YES	cl04787	74	(Schindelman et al., 2001)
HbCOBL-D	- LVXX01000183	1,783	AT3G16860.1	53.07	179	3.00E-51	-	-	NO	-	-	-
HbCOBL-E			AT4G16120.1	70.93	454	0	Full-length	471	YES	pfam04833	180	(Borner et al., 2002)
HbCOBL-F	_		AT3G02210.1	81.58	152	5.00E-81	Full-length	160	YES	pfam04833	152	(Roudier et al., 2002)
HbCOBL-G	LVXX01000195	1,700	AT5G60920.1	66.51	209	5.00E-98	-	-	NO	-	-	-
HbCOBL-H	-		AT5G60920.1	77.63	152	1.00E-71	Full-length	159	YES	pfam04833	152	(Roudier et al., 2002)
HbCOBL-I	- LVXX01000520	893	AT3G02210.1	81.58	152	3.00E-80	Full-length	172	YES	pfam04833	152	(Roudier et al., 2002)
HbCOBL-J			AT5G15630.1	76.43	157	2.00E-79	Full-length	171	YES	pfam04833	115	(Taylor-Teeples et al., 2015)
HbCOBL-K	- LVXX01000625	733	AT3G16860.1	54.44	180	1.00E-52	-	-	NO	-	-	-
HbCOBL-L			AT4G16120.1	71.52	453	0	Full-length	471	YES	pfam04833	180	(Roudier et al., 2002)
HbCOBL-M	LVXX01000656	689	AT3G29810.1	50.00	152	1.00E-44	Full-length	181	YES	pfam04833	150	(Roudier et al., 2002)
HbCOBL-N	LVXX01001114	267	AT3G29810.1	52.32	151	2.00E-58	Partial	184	YES	pfam04833	148	(Roudier et al., 2002)
HbCOBL-O	LVXX01001412	126	AT4G16120.1	63.54	458	0	Full-length	451	YES	pfam04833	180	(Roudier et al., 2002)
HbCOBL-P	- LVXX01003569	12	AT4G16120.1	63.54	458	0	Full-length	451	YES	pfam04833	180	(Roudier et al., 2002)
HbCOBL-Q			AT5G49270.1	54.37	160	1.00E-47	-	-	NO	-	-	-

Table 1. Identification of 17 putative encoding COBRA genes using *in silico* comparative analysis of 9 genomic scaffolds of *H. brasiliensis*. The matching of *Hevea* and *Arabidopsis COBRA* genes were carried out using tblastn of Galaxy. The identification of protein domain was carried out using NCBI Conserved Domains Database (CDD) Search.

Thirteen of seventeen putative HbCOBL genes were annotated in nine genomic scaffolds (Table 2). The HbCOBL-A, HbCOBL-B and HbCOBL-C harbored by scaffold LVXX01000065.1 in the position of 82,210-82,704; 748,223-748,714; and 722,236-722,502 were potentially homolog to AtCOBL6, AtCOBL2 and AtCOB, respectively. The HbCOBL-E harbored by scaffold LVXX01000183.1 in the position of 89,232-90,644 was potentially ortholog to AtCOBL7. The scaffold LVXX01000195.1 harbored two confirmed HbCOBL genes (HbCOBL-F and HbCOBL-H) in the position of 995,959-996.438 and 987,960-988,436. These genes were potentially ortholog to AtCOBL1 and AtCOB. The genes HbCOBL-I and HbCOBL-J located in the scaffold LVXX01000520.1 (position of 73,476-73,991 and 77,649-78,161) were putative ortholog to AtCOBL1 and AtCOBL4. Meanwhile, only the gene *HbCOBL-L* closely related with *AtCOBL7* was confirmed in the scaffold LVXX01000625.1 (position of 217,872-219,284). The genes HbCOBL-M and HbCOBL-N putatively ortholog to AtCOBL2 were

harbored by the scaffolds LVXX01000656.1 and LVXX01001114.1 in the position of 533,887-534.429 and 261,457-262,008, respectively. Finally, the genes HbCOBL-O and HbCOBL-P putatively ortholog to AtCOBL-7 were harbored by the scaffolds LVXX01001412.1 and LVXX01003569.1 in the and position of 104,777-106,129 2,362-3,714, respectively.

The putative location of each *HbCOBL* gene in nine *H. brasiliensis* genomic scaffolds were shown in Figure 2. For the scaffold LVXX01000065.1, the gene *HbCOBL-A* was located for more than 666 Mb upstream of *HbCOBL-C* and *HbCOBL-B* while *HbCOBL-C* and *HbCOBL-B* while *HbCOBL-C* and *HbCOBL-B* were separated by approximately 26 Mb distances. For the scaffold LVXX01000195.1, the genes *HbCOBL-H* and *HbCOBL-F* were separated by 8 Mb distances in the downstream part of the scaffold. For the scaffold LVXX01000520.1, the genes *HbCOBL-J* were separated by 4 Mb distances in the upstream part of the scaffold (Table 2, Figure 2).

Table 2. Annotation of 17 putative encoding COBRA genes in 9 genomic scaffolds of H. brasiliensis.

		Tang's Scaffo	Arabidopsis genome			
Gene name	Scaffold ID	Length (Mb)	Position (base to base)	Accession	Gene name	
HbCOBL-A			82,210 82,704	AT1G09790.1	AtCOBL6	
HbCOBL-B	LVXX01000065.1	2.714	748,223 748,714	AT3G29810.1	AtCOBL2	
HbCOBL-C			722,236 722,502	AT5G60920.1	AtCOB	
HbCOBL-D	I VVV01000192 1	1 792	-	AT3G16860.1	AtCOBL8	
HbCOBL-E	LVAA01000185.1	1.785	89,232 90,644	AT4G16120.1	AtCOBL7	
HbCOBL-F			995,959 996.438	AT3G02210.1	AtCOBL1	
HbCOBL-G	LVXX01000195.1	1.700	-	AT5G60920.1	AtCOB	
HbCOBL-H			987,960 988,436	AT5G60920.1	AtCOB	
HbCOBL-I	LVXV01000520 1	202	73,476 73,991	AT3G02210.1	AtCOBL1	
HbCOBL-J	LVAA01000520.1	893	77,649 78,161	AT5G15630.1	AtCOBL4	
HbCOBL-K	LVXV01000625 1	722	-	AT3G16860.1	AtCOBL8	
HbCOBL-L	LVAA01000025.1	155	217,872 219,284	AT4G16120.1	AtCOBL7	
HbCOBL-M	LVXX01000656.1	689	533,887 534.429	AT3G29810.1	AtCOBL2	
HbCOBL-N	LVXX01001114.1	267	261,457 262,008	AT3G29810.1	AtCOBL2	
HbCOBL-O	LVXX01001412.1	126	104,777 106,129	AT4G16120.1	AtCOBL7	
HbCOBL-P	L VVV01002560 1	12	2,362 3,714	AT4G16120.1	AtCOBL7	
HbCOBL-O	LVAA01003569.1	12	-	AT5G49270.1	AtCOBL9	



Figure 2. Putative location of each *HbCOBL* gene in 9 *H. brasiliensis* genomic scaffolds. Yellow triangle locates the CDS of the gene. Red triangle locates the COBRA domain of the gene. For the scaffold LVXX01000065.1, the order of annotation (left to right) was *HbCOBL-A*, *HbCOBL-C* and *HbCOBL-B*. For the scaffold LVXX01000195.1, the order of annotation (left to right) was *HbCOBL-F*. For the scaffold LVXX01000520.1, the order of annotation (left to right) was *HbCOBL-F*. For the scaffold LVXX01000520.1, the order of annotation (left to right) was *HbCOBL-I* and *HbCOBL-J*.

Comparative Phylogenetic Analysis of *Hevea* and *Arabidopsis* COBRA Gene Family

The comparative phylogenetic analysis of 13 *Hevea* and 11 *Arabidopsis* COBRA gene families were exhaustively carried out (Figure 3). Two clusters have been identified and named after the grouping of *AtCOBL* genes: Subfamily-I and –II. Nine *HbCOBL* genes (*HbCOBL-B*, *HbCOBL-J*, *HbCOBL-C*, *HbCOBL-H*, *HbCOBL-F*, *HbCOBL-I*, *HbCOBL-M*, *HBCOBL-A*, and *HbCOBL*-N) were clustered as COBRA gene subfamily-I. By contrast, four genes (*HbCOBL-O*, *HbCOBL-P*, *HbCOBL-E*, and *HbCOBL-L*) were clustered as COBRA gene subfamily-II. The analysis of multiple sequence alignment on HbCOBL amino acid residues confirmed two differentiated groups (Figure 4). The *Hevea* COBRA gene subfamily-II was marked by the addition of 203 residues in C-terminal.



Figure 3. Phylogenetic analysis of 13 and 11 COBRA amino acids from *H. brasiliensis* and *A. thaliana*. The evolutionary history was inferred using the Neighbor-Joining method of MEGA 7.0 The bootstrap consensus tree inferred from 1000 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The bootstrap consensus tree inferred from 1000 replicates. The analysis involved 25 amino acid sequences with *AtARF01.2* as the outrgoup. All ambiguous positions were removed for each sequence pair. There were a total of 779 positions in the final dataset. Two bars with red and blue colors refer to COBRA subfamily-I and –II.

Discussion

The tolerance to recurrent tapping in rubber tree is a unique feature that allows latex to be harvested over the production age (Putranto et al., 2015a). This tolerance is marked by the reproduction of latex flow and also the regeneration of laticifers, especially the secondary layers. Two main hormonal biosynthesis and pathways, ethylene and jasmonate, were comprehensively studied for the past 10 years. However, studies on genes related to structural development of the laticifers were limited. In their study, Tang et al. (2016) has carried out annotation analysis and expression of genes located in *H. brasiliensis* SNP desert. Three scaffolds putatively encoding COBRA genes (LVXX01000065, LVXX01000520, and LVXX01000625) showed high transcript abundance in the laticifer with reads per kilobase per million mapped reads (RPKM) of 0.14, 8.39, and 5.01, respectively. In addition, Rahman et al. (2013) has previously predicted and grouped ten putative *Hevea COBRA* genes as lignocellulose biosynthetic genes. These data raised a question whether COBRA genes could potentially play an important role during laticifer differentiation in rubber tree.



Figure 4. Multiple sequence alignment of *H. brasiliensis* COBRA gene family. The analysis was performed by MUSCLE alignment of Geneious. Two bars with red and blue colors refer to COBRA subfamily-I (SF-I) and –II (SF-II). The orange boxes highlight the additional sequence on C-terminal.

In this study, a comparative analysis using two reference genomes from H. brasiliensis and A. thaliana was systematically carried out for specific gene family. At the beginning, seventeen putative HbCOBL genes were identified. After the CDD search analysis, thirteen genes were in silico validated. These genes carried a COBRA domain in the CDS region. The other four were excluded from the analysis. The phylogenetic analysis has regrouped the HbCOBL genes into two subfamilies as described in the A. thaliana COBRA gene family. The subfamily-II differs from the subfamily-I due to its addition of N-terminal 170 amino acids (Roudier et al., 2002). In this study, the HbCOBL subfamily-II showed additional amino acid residues at C-terminal. This result suggests that Hevea and Arabidopsis COBRA genes might structurally different but share similar functions.

In Arabidopsis, the COBRA gene family was located in the chromosome I, III, IV, and V (Roudier et al., 2002). The chromosome I harbored the genes AtCOBL6 and AtCOBL3 whereas the chromosome IV harbored the genes AtCOBL7 and AtCOBL11. Two chromosomes III and V harbored four genes for both of them. The genes AtCOBL1, AtCOBL8, AtCOBL10, and AtCOBL2 were found in the chromosome III. The genes AtCOBL4, AtCOBL9, AtCOB, and AtCOBL5 were located in the chromosome V. The duplication phenomenon of Arabidopsis COBRA genes has been discussed (Brady et al., 2007; Roudier et al., 2002). In this paper, three genes (HbCOBL-A, HbCOBL-B, and HbCOBL-C) as well as two other groups of genes (*HbCOBL-F* and *HbCOBL-H*; HbCOBL-I and HbCOBL-J) resided the same scaffold. The gene *HbCOBL-C* putatively ortholog to *AtCOB* had the unique COBRA domain cl04787 with 74 amino acid residues. In addition, the gene HbCOBL-H was predicted to be ortholog with AtCOB. In Arabidopsis, the role of gene AtCOB has been demonstrated in cell wall deposition (Liepman et al., 2010; Schindelman et al., 2001). Both genes (HbCOBL-C and HbCOBL-H)

were positioned in different scaffolds. The close relation of both genes was equally illustrated in the phylogenetic tree. The genes *HbCOBL-C*, *HbCOBL-H*, and *AtCOB* shared a small cluster. This suggests a potential duplication of *HbCOBL-C* and *HbCOBL-H* in *H*. *brasiliensis* genome.

CONCLUSIONS

Taken together, comparative genomic analysis, whether partially or totally implemented reveals new cues at the understanding of molecular mechanism inside a given organism. By implementing sequence-to-sequence analysis, the putative *H. brasiliensis* COBRA gene family was identified. The gene family was clustered into two subfamilies with total 13 members. Future perspectives should be focused on demonstrating the implication of *HbCOBL* genes in laticifer differentiation.

ACKNOWLEDGEMENTS

The authors thank Prof. Chaorong Tang and team for providing the rubber genome in the public database. The authors equally thank the Laboratory of Biochemical and Molecular Biology, the Indonesian Research Institute for Biotechnology and Bioindustry in Indonesia and the SouthGreen Bioinformatics Platform in France for the software support.

REFERENCES

Ben-Tov, D., Abraham, Y., Stav, S., Thompson, K., Loraine, A., Elbaum, R., de Souza, A., Pauly, M., Kieber, J.J., Harpaz-Saad, S., 2015. COBRA-LIKE2, a Member of the Glycosylphosphatidylinositol-Anchored COBRA-LIKE Family, Plays a Role in Cellulose Deposition in *Arabidopsis* Seed Coat Mucilage Secretory Cells. Plant Physiology 167, 711-724.

- Borner, G.H.H., Sherrier, D.J., Stevens, T.J., Arkin, I.T., Dupree, P., 2002. Prediction of Glycosylphosphatidylinositol-Anchored Proteins in *Arabidopsis*. A Genomic Analysis. Plant Physiology 129, 486-499.
- Brady, S.M., Song, S., Dhugga, K.S., Rafalski, J.A., Benfey, P.N., 2007. Combining Expression and Comparative Evolutionary Analysis. The COBRA Gene Family. Plant Physiology 143, 172-187.
- Cock, P.J.A., Chilton, J.M., Grüning, B., Johnson, J.E., Soranzo, N., 2015. NCBI BLAST+ integrated into Galaxy. bioRxiv.
- Dai, X., You, C., Chen, G., Li, X., Zhang, Q., Wu, C., 2011. OsBC1L4 encodes a COBRA-like protein that affects cellulose synthesis in rice. Plant Molecular Biology 75, 333-345.
- Duan, C., Rio, M., Leclercq, J., Bonnot, F., Oliver, G., Montoro, P., 2010. Gene expression pattern in response to wounding, methyl jasmonate and ethylene in the bark of *Hevea brasiliensis*. Tree Physiol 30, 1349-1359.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792-1797.
- Felsenstein, J., 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution 39, 783-791.
- Guo, D., Zhou, Y., Li, H.-L., Zhu, J.-H., Wang, Y., Chen, X.-T., Peng, S.-Q., 2017. Identification and characterization of the abscisic acid (ABA) receptor gene family and its expression in response to hormones in the rubber tree. Scientific Reports 7, 45157.
- Hao, B.-Z., Wu, J.-L., 2000. Laticifer Differentiation in *Hevea brasiliensis*: Induction by Exogenous Jasmonic Acid and Linolenic Acid. Annals of Botany 85, 37-43.
- Initiative, T.A.G., 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796-815.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8, 275-282.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution.
- Lalanne, E., Honys, D., Johnson, A., Borner, G.H.H., Lilley, K.S., Dupree, P., Grossniklaus, U., Twell, D., 2004. *SETH1* and *SETH2*, Two Components of the Glycosylphosphatidylinositol Anchor Biosynthetic Pathway, Are Required for Pollen Germination and Tube Growth in *Arabidopsis*. The Plant Cell 16, 229-240.
- Laosombut, T., Arreewichit, P., Nirapathpongporn, K., Traiperm, P., Kongsawadworakul, P., Viboonjun, U., Narangajavana, J., 2016. Differential Expression of Methyl Jasmonate-Responsive Genes Correlates with Laticifer Vessel Proliferation in Phloem Tissue of Rubber Tree (*Hevea* brasiliensis). Journal of Plant Growth Regulation, 1-15.
- Lau, N.-S., Makita, Y., Kawashima, M., Taylor, T.D., Kondo, S., Othman, A.S., Shu-Chien, A.C., Matsui, M., 2016. The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. Scientific Reports 6, 28594.
- Liepman, A.H., Wightman, R., Geshi, N., Turner, S.R., Scheller, H.V., 2010. *Arabidopsis* – a powerful model system for plant cell wall research. The Plant Journal 61, 1107-1121.
- Liu, H., Deng, Z., Chen, J., Wang, S., Hao, L., Li, D., 2016. Genome-wide identification and expression analysis of the metacaspase gene family in *Hevea brasiliensis*. Plant Physiology and Biochemistry 105, 90-101.

- Liu, L., Shang-Guan, K., Zhang, B., Liu, X., Yan, M., Zhang, L., Shi, Y., Zhang, M., Qian, Q., Li, J., Zhou, Y., 2013. Brittle Culm1, a COBRA-Like Protein, Functions in Cellulose Assembly through Binding Cellulose Microfibrils. PLOS Genetics 9, e1003704.
- Long, X., He, B., Fang, Y., Tang, C., 2016. Identification and Characterization of the Glucose-6-Phosphate Dehydrogenase Gene Family in the Para Rubber Tree, *Hevea brasiliensis*. Frontiers in Plant Science 7, 215.
- Martiansyah, I., Putranto, R.A., Khumaida, N., 2017. Identification of putative gene family encoding protease inhibitors by *in silico* comparative analysis in *Hevea brasiliensis* Müll.Arg. genome. E-Journal Menara Perkebunan 85, n/a-n/a.
- Pirrello, J., Leclercq, J., Dessailly, F., Rio, M., Piyatrakul, P., Kuswanhadi, K., Tang, C., Montoro, P., 2014. Transcriptional and post-transcriptional regulation of the jasmonate signalling pathway in response to abiotic and harvesting stress in *Hevea brasiliensis*. BMC Plant Biology 14, 1-17.
- Piyatrakul, P., Yang, M., Putranto, R.-A., Pirrello, J., Dessailly, F., Hu, S., Summo, M., Theeravatanasuk, K., Leclercq, J., Kuswanhadi, Montoro, P., 2014. Sequence and Expression Analyses of Ethylene Response Factors Highly Expressed in Latex Cells from *Hevea brasiliensis*. PLoS ONE 9, e99367.
- Putranto, R.-A., Duan, C., Kuswanhadi, Chaidamsari, T., Rio, M., Piyatrakul, P., Herlinawati, E., Pirrello, J., Dessailly, F., Leclercq, J., Bonnot, F., Tang, C., Hu, S., Montoro, P., 2015a. Ethylene Response Factors Are Controlled by Multiple Harvesting Stresses in *Hevea brasiliensis*. PLoS ONE 10, e0123618.
- Putranto, R.-A., Herlinawati, E., Rio, M., Leclercq, J., Piyatrakul, P., Gohet, E., Sanier, C., Oktavia, F., Pirrello, J., Kuswanhadi, Montoro, P., 2015b. Involvement of Ethylene in the Latex Metabolism and Tapping Panel Dryness of *Hevea brasiliensis*. International Journal of Molecular Sciences 16, 17885.
- Putranto, R.A., Montoro, P., 2016. The *Hevea brasiliensis* AP2/ERF superfamily: From ethylene signalling to latex harvesting and physiological disease response. Menara Perkebunan 84, 49-62.
- Rahman, A.Y.A., Usharraj, A., Misra, B., Thottathil, G., Jayasekaran, K., Feng, Y., Hou, S., Ong, S.Y., Ng, F.L., Lee, L.S., Tan, H.S., Sakaff, M.K.L.M., Teh, B.S., Khoo, B., Badai, S.S., Aziz, N.A., Yuryev, A., Knudsen, B., Dionne-Laporte, A., Mchunu, N., Yu, Q., Langston, B., Freitas, T.A., Young, A., Chen, R., Wang, L., Najimudin, N., Saito, J., Alam, M., 2013. Draft genome sequence of the rubber tree *Hevea brasiliensis*. BMC Genomics 14, 75.
- Roudier, F., Schindelman, G., DeSalle, R., Benfey, P.N., 2002. The COBRA Family of Putative GPI-Anchored Proteins in *Arabidopsis*. A New Fellowship in Expansion. Plant Physiology 130, 538-548.
- Schindelman, G., Morikami, A., Jung, J., Baskin, T.I., Carpita, N.C., Derbyshire, P., McCann, M.C., Benfey, P.N., 2001. COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. Genes & Development 15, 1115-1127.
- Tang, C., Yang, M., Fang, Y., Luo, Y., Gao, S., Xiao, X., An, Z., Zhou, B., Zhang, B., Tan, X., Yeang, H.-Y., Qin, Y., Yang, J., Lin, Q., Mei, H., Montoro, P., Long, X., Qi, J., Hua, Y., He, Z., Sun, M., Li, W., Zeng, X., Cheng, H., Liu, Y., Yang, J., Tian, W., Zhuang, N., Zeng, R., Li, D., He, P., Li, Z., Zou, Z., Li, S., Li, C., Wang, J., Wei, D., Lai, C.-Q., Luo, W., Yu, J., Hu, S., Huang, H., 2016. The rubber tree genome reveals new insights into rubber production and species adaptation. Nature Plants, 16073.
- Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., Handakumbura, P.P., Xiong, G., Wang, C., Corwin, J., Tsoukalas, A., Zhang, L., Ware, D., Pauly, M.,

Kliebenstein, D.J., Dehesh, K., Tagkopoulos, I., Breton, G., Pruneda-Paz, J.L., Ahnert, S.E., Kay, S.A., Hazen, S.P., Brady, S.M., 2015. An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. Nature 517, 571-575.

- Tian, W.-M., Yang, S.-G., Shi, M.-J., Zhang, S.-X., Wu, J.-L., 2015. Mechanical wounding-induced laticifer differentiation in rubber tree: An indicative role of dehydration, hydrogen peroxide, and jasmonates. Journal of Plant Physiology 182, 95-103.
- Wu, S., Zhang, S., Chao, J., Deng, X., Chen, Y., Shi, M., Tian, W.-M., 2016. Transcriptome Analysis of the Signalling Networks in Coronatine-Induced Secondary Laticifer Differentiation from Vascular Cambia in Rubber Trees. Scientific Reports 6, 36384.
- Xiao, X.-H., Yang, M., Sui, J.-L., Qi, J.-Y., Fang, Y.-J., Hu, S.-N., Tang, C.-R., 2017. The calcium-dependent protein kinase (CDPK) and CDPK-related kinase gene families in *Hevea brasiliensis*—comparison with five other plant species in structure, evolution, and expression. FEBS Open Bio 7, 4-24.
- Xiao, X., Tang, C., Fang, Y., Yang, M., Zhou, B., Qi, J., Zhang, Y., 2014. Structure and expression profile of the sucrose

synthase gene family in the rubber tree: indicative of roles in stress response and sucrose utilization in the laticifers. FEBS Journal 281, 291-305.

- Zhang, S.-X., Wu, S.-H., Chen, Y.-Y., Tian, W.-M., 2015. Analysis of Differentially Expressed Genes Associated with Coronatine-Induced Laticifer Differentiation in the Rubber Tree by Subtractive Hybridization Suppression. PLoS ONE 10, e0132070.
- Zhiyi, N., Guijuan, K., Yu, L., Longjun, D., Rizhong, Z., 2015. Whole-Transcriptome Survey of the Putative ATP-Binding Cassette (ABC) Transporter Family Genes in the Latex-Producing Laticifers of *Hevea brasiliensis*. PLoS ONE 10, e0116857.
- Zou, Z., Gong, J., An, F., Xie, G., Wang, J., Mo, Y., Yang, L., 2015. Genome-wide identification of rubber tree (*Hevea brasiliensis* Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber-producing tissue. BMC Genomics 16, 1001.

THIS PAGE INTENTIONALLY LEFT BLANK