Indonesian Journal of Biotechnology

VOLUME 25(2), 2020, 84-94 | RESEARCH ARTICLE

Cloning and in silico analysis revealed a genetic variation in osmotin-encoding genes in an Indonesian local cacao cultivar

Imam Bagus Nugroho[1,*], **Fahrurrozi**[2]

[1]Indonesian Research Institute for Biotechnology and Bioindustry (IRIBB), Jalan Taman Kencana No 1, Bogor, Jawa Barat 16128, Indonesia
[2]Indonesian Institute of Sciences (LIPI), Jalan Raya Bogor Km. 46, Cibinong, Jawa Barat 16119, Indonesia
*Corresponding author: ibagusnugroho@iribb.org

SUBMITTED 31 January 2020 **REVISED** 10 May 2020 **ACCEPTED** 24 July 2020

ABSTRACT *Theobroma cacao* L. is an important Indonesian estate crop, which suffers from biotic and abiotic stresses. *TcOSM*, which encodes osmotin as a response to pathogens and environmental stresses, is therefore a focus of interest in this research, aiming to characterize *TcOSM* in an Indonesian local cacao cultivar. Bioinformatics queries for putative *TcOSM* were performed against the reference genome of a Criollo-type cacao cultivar. Based on nucleotide sequence determination, our results revealed two genes, *TcOSM1* and *TcOSM2*, which have the highest similarity (≥ 90%) to the cacao reference genes. Heterozygosity was detected in the *TcOSM1*-encoding gene, which contained two overlapping peaks in Sanger-sequencing chromatograms. One of the alleles resulted from a single nucleotide change (G to A), leading to a same-sense mutation that did not substitute corresponding alanine residue. Homology modeling using Phyre2 and structural alignment (superimposition) was conducted to examine the influence of genetic variations in *TcOSM* sequences upon the global protein structures. The result showed no significant changes (RMSD ≤ 0.206 Å, TM-score > 0.5) in tertiary protein structures. Altogether, this research succeeded in characterizing *TcOSM* while providing a fundamental study for future cacao biotechnology endeavors.

KEYWORDS Bioinformatic; heterozygosity; homology modeling; local cacao cultivar; pathogenesis-related proteins

1. Introduction

Cacao (*Theobroma cacao* L.) is a native plant of the Amazon basin (Souza et al. 2018), which cultivation is dated back to ancient Mayan and Aztec civilizations. It spreads throughout the world through domestication (Zhang and Motilal 2016) and becomes one of the most valuable estate crop commodities of Indonesia (Mithöfer et al. 2017). Cacao product derivatives are used as foods, liquors, and additives for food flavoring and coloring. Unfortunately, the productivity of cacao in Indonesia suffers from pathogen attacks and drought. It has been reported that cacao plantation suffers greatly from pathogenic fungal infections, i.e. (i) *Moniliophthora perniciosa* (syn. *Crinipellis perniciosa*) causes the Witches' Broom disease (Farquharson 2014); (ii) the Black Rot Pods caused by *Phytophthora* spp. (Ali et al. 2017); (iii) *Moniliophthora roreri*, which causes the Frosty Pod Rots disease (Bailey et al. 2018); and (iv) Vascular-Streak Dieback disease caused by *Ceratobasidium theobromae* (syn. *Oncobasidium theobromae*) (Ali et al. 2019).

Conventional breeding has successfully produced cacao cultivars, which have better fungal resistance, productivity, and performance amidst environmental stressors (Wickramasuriya and Dunwell 2018), albeit at a relatively long time. The plant molecular breeding studies can address the low turn-over of conventional breeding by implementing known plant molecular trait markers. However, the detailed information of the trait markers, such as Quantitative Trait Loci (QTL) or Simple Sequence Repeats (SSR), is required for the basis in engineering plant phenotypes. Moreover, the full sequence of known resistance gene sequence, its biochemical properties, genetic expression, and working mechanism are also necessary to be obtained in the effort to create a better resistance cacao cultivar against a pathogen.

Cacao, as well as other plant species, can defend against fungal invasion using sentinel plant systems. One of the defense mechanisms among higher plants is to produce the pathogenesis-related (PR) family proteins. The expression of PR-5 protein, such as osmotin, is activated by abiotic environmental stresses (van Loon et al. 2006). Researchers have succeeded in expressing the osmotin-encoding gene from Tobacco (*Nicotiana tabaccum* L.) (Tzou et al. 2011) and Nightshade (*Solanum nigrum* L. var. *americanum*) (Campos et al. 2002), employing the heterologous expression system of the full or

truncated osmotin-encoding gene (Campos et al. 2008). The previous study reported partial gene encoding osmotin from cacao (*Theobroma cacao* L.) and engineered peptides derivation, which showed a significant inhibitory effect on pathogenic fungal strains (Falcao et al. 2016). However, the study of osmotins from Indonesian local cacao cultivar, which derived from Criollo, Trinitario, or Forastero cultivars, remains unexplored.

This research aimed to characterize the osmotin proteins from Indonesian local cacao cultivar using combined bioinformatics and molecular biology approaches. The bioinformatic study included gene determination using BLAST, multiple sequence alignment, and protein homology modeling. Furthermore, the targeted gene was cloned and confirmed by Sanger DNA sequencing analysis. This research has succeeded in determining and characterizing two osmotin proteins from local Indonesian cacao cultivar. Understanding the profile of osmotins will provide a new insight for selection in breeding strategies based on the existing genetic variation pools. This insight is valuable to overcome the issues in cacao plantation and other endeavors.

2. Materials and Methods

2.1. Sample collection and preparation

Ripe cacao fruits, indicated by purple to brownish pod color and harvested from healthy mature trees (three-to-five years of age after planting), were bought from a local trader (PD Petani Kakao Lampung), Pringsewu Region, Lampung Province, Sumatera, Indonesia. Samples were wrapped using plastics, transported to the lab, and stored in -20 °C freezer until use.

2.2. Bioinformatic queries of TcOSM in the public database

Bioinformatic queries were conducted using the osmotin (AP24) amino acid sequence of *N. tabaccum* (AAB23375.1). The retrieved N. tabaccum osmotin amino acid sequence was further used for protein queries using BLASTP against the available sequence of *T. cacao* L. cultivar Criollo (GCF_000208745.1). The results were filtered using percent identity value ≥ 60% and low E-value (≤ 1.0E-90). Domain queries were also performed to confirm the BLAST hits against the Conserved Domain Database (CCD v3.17). Putative osmotins from Indonesian local cacao cultivar were validated as they contained Thaumatin-Like Protein (TLP) motif.

2.3. DNA genomic isolation from cacao beans

The DNA genomic isolation was conducted using Plant Genomic DNA Mini Kit (Geneaid, Taiwan) as per the manufacturer instructions using 100 mg of a fresh plant tissue sample. Cacao beans were extracted from thawed pods. Cotyledons were separated from the testa and endosperms. Cacao cotyledons were weighed up to 100 mg and lysed using mechanical force in a sterile mortar and pestle containing lysis buffer. Further purification of genomic DNA was done using provided column-based protocols to remove the mucilage from cacao beans' testa, which ensures downstream treatment such as PCR amplification. The genomic DNA was eluted in sterile 30 µl ddH2O following the manufacturer's protocol. The intact genomic DNA was visualized using 0.8% agarose electrophoresis (Chiong et al. 2017).

2.4. TcOSM amplification, ligation, cloning and sequencing

The primers (Table 1) were designed based on the predicted cacao osmotins mRNA sequences (XM_018118608.1 and XM_007040100.2) retrieved after bioinformatic queries mentioned before. The primer design followed the protocol described in the study of Nugroho and Handayani (2016). The amplifications were conducted using MyTaq™ HS Red Mix (Bioline, USA) with initial denaturation at 98 °C for five min in 35-programmed cycles as follows: denaturation at 98 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for one min, and a final extension at 72 °C for five min. Direct Sanger DNA sequencing was done to determine the sequence of *TcOSM* genes from the PCR amplicons.

Cloning procedures were conducted by inserting *TcOSM1* and *TcOSM2* into pTA2 cloning vector (Toyobo, Japan) followed by ligation using T4 DNA Ligase (Toyobo, Japan) in 14 °C overnight. The ligation mix was transformed into *E. coli* DH5α competent cells (Takara, Japan). Luria-Bertani medium supplemented with ampicillin was used for screening of positive clones. The resulting plasmid from the positive clone was selected and re-amplified using PCR procedures mentioned above and subjected to sequencing at 1st Base, Selangor, Malaysia. DNA sequencing was repeated three times from three independently sampled positive clones to validate the findings.

TABLE 1 Oligonucleotides used in this study.

Name	Sequences	Predicted Amplicon Size
TcOSM1_F	5'-AAAATTCCAACACTTCTACCGTCCT-3'	822 bp
TcOSM1_R	5'-CAACTTCACTCTAGCACGTCATTCC-3'	
TcOSM2_F	5'-TCTCCCCAACTCCATATCCA-3'	886 bp
TcOSM2_R	5'-TGGGCAATTTGGTTGAGATT-3'	

2.5. Sequence analyses and unrooted Neighbour-Joining (NJ) tree reconstruction

Nucleotide sequences were analyzed using Geneious Prime v2019.1.1 (Biomatters Ltd., New Zealand). Forward and reverse reads from Sanger sequencing were assembled to build Contig. Consensus sequences from several contigs were aligned using MUSCLE (Edgar 2004) against mRNA reference sequences (XM_018118608.1 and XM_007040100.2), osmotins from other taxa, and thaumatins. Multiple sequence alignment was done to assist the sequence analyses for identifying genetic variation and pinpointing residues conservancy among studied proteins.

Open reading frames for *TcOSM1* and *TcOSM2* were deducted to build correct translations, then submitted to BLASTP. The twenty highest BLASTP hits, including *TcOSM1* and *TcOSM2*, were downloaded and subjected to multiple sequence alignment using MUSCLE in MEGAX (Kumar et al. 2018). A Neighbour-Joining unrooted tree was reconstructed using the Jones-Taylor-Thornton matrix to compute evolutionary distance based on 1,000 bootstrap resamplings.

2.6. TcOSM protein structures homology modeling

In silico modeling was conducted to simulate the structural changes caused by variations in amino acids in *TcOSM1* and *TcOSM2*. The amino acid sequence of NtOSM (AAB23375.1), "Wild-type" *TcOSM1* (XP_017974097.1), *TcOSM1*, and *TcOSM2* without signal peptide sequences were submitted to Phyre2 server (Kelley et al. 2015). The protein structures were retrieved as PDB file and analyzed using PyMOL v2.3.2 software (Schrödinger Lnc., USA) to align the models with 120° rotation on the y-axis. The images were captured using [set ray_opaque_background, 0] followed by [png Filename, dpi=600] commands. Structural identities were measured using indicators such as RMSD (Root-Mean-Score Deviation) values and TM-scores (Xu and Zhang 2010).

3. Results and Discussion

3.1. Cacao genome contains two TcOSM paralogs with highly similar amino acid sequences

The implementation of NCBI+blastp (protein BLAST search) has succeeded in identifying two paralogs of osmotin in T. cacao genome cultivar Criollo (Table 2). Based on alignment with *N. tabaccum* osmotin amino acid sequence (NtOSM - AAB23375.1), two putative osmotin sequences of T. cacao, i.e. XP_017974097.1 and XP_007040162.2 were identified with their respective mRNA sequence counterparts as follows: XM_018118608.1 and XM_007040100.2. The mRNA sequences of osmotin paralogs from *T. cacao* showed single coding exons without the detected intron.

Intronless osmotin genes have successfully been elucidated not only in cacao but also in other plant species such as strawberry (Wu et al. 2001). Single exon mRNA within the cacao osmotins brought confidence and ease in the primer design for their amplifications from genomic DNA samples. Our strategy is in agreement with the experiment conducted by Chowdhury et al. (2015), which successfully cloned the osmotin-like gene from Solanum nigrum L using DNA genomic template and similar primer design rationale. The primers used in our study were designed to flank the osmotin ORF; therefore, the start codon (ATG) was not included in the primer as it was located downstream of the priming area (Figure 1). The absence of intron in osmotin made the amplification from the genomic DNA template straightforward in getting full ORF avoiding any introns.

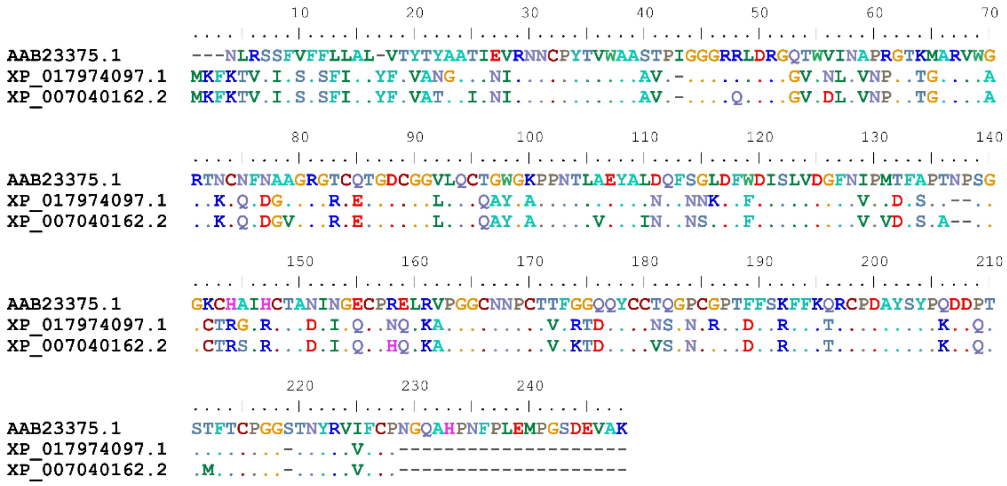
The identified osmotin paralogs were highly similar to another with a ≥ 90% identity (Figure 2a). Additionally, the two paralogs have relatively high identity to NtOSM (≥ 60% and were also confirmed to have the TLP domain according to the CDD Hits. The CDD has been utilized successfully to assign an unknown protein into a known subfamily. The CDD functions based on the unique hallmark or amino acid sequence characterizing a specific domain or motif. A similarity-based CDD search can be used to infer a protein function and its evolutionary aspects (Fong and Marchler-Bauer 2008). Additionally, CDD was also utilized to screen and validate a plant gene called COBRA using the designated domain or motif signature (Putranto et al. 2017). Therefore, the identified paralogs were confidently justified as osmotin encoding genes after it contained conserved TLP domain or motif.

The paralogs also showed the highest similarity to the Thaumatin-Like Protein (TLP) of Durio zibethinus (XP_022773064.1) as they converged to the same branch (Figure 2b) of the reconstructed unrooted NJ tree. The node was well-supported as it displayed a ≥ 70% bootstrap value (Soltis and Soltis 2003). The branch was also complexed with other TLPs of Gossypium spp. This branch represents the common ancestry of osmotin orthologs within the Malvaceae since *T. cacao*, *D. zibethinus*, and *Gossypium* spp. all belong to the same family.

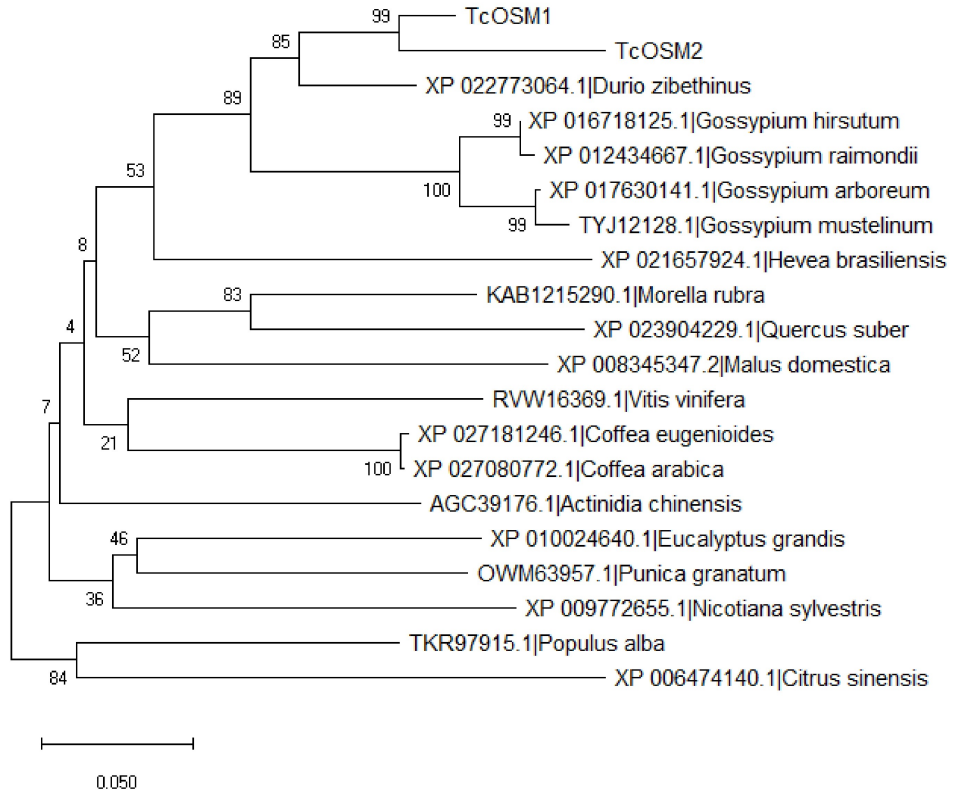
TABLE 2 Summary of bioinformatic queries results.

NCBI+blastp Hits			CDD Hits			
Accession ID	Expectation Value	Percent Identity	Domain Name	Accession ID	Sequence Interval	Expectation Value
XP_017974097.1	8,00E-92	64.86 %	GH64-TLP-SF	cl02511	026-224	3.50E-97
XP_007040162.2	3,00E-99	64.41 %	GH64-TLP-SF	cl02511	026-224	6.36E-101

FIGURE 1 The primers annealing (priming) areas for *TcOSMs* amplification.



(a)



(b)

FIGURE 2 Multiple sequence alignment and the NJ tree of osmotins from different taxa. a. MSA of *Nicotiana tabaccum* osmotin (AAB23375.1) and predicted osmotins from Criollo cacao (XP_017974097.1 and XP_007040162.2). b. Reconstructed Neighbour-Joining (NJ) tree (unrooted) of TcOSM1, TcOSM2, and other osmotins from different taxa.

3.2. *TcOSM1 of Indonesian local cacao cultivar shows a genetic variation*

The genomic DNA isolation from cotyledons of cacao was challenging due to the mucilaginous nature of cacao beans surface, which can hinder DNA precipitation. The column's utilization to remove the mucilage from homogenate has proven successful in retrieving intact genomic DNA from cacao beans (Figure 3a, Lane 1), which later ensured the probability of success in the target gene amplification.

The genomic DNA was used as the template for *TcOSM1* and *TcOSM2* amplifications. Successful amplification was visualized using 0.8% agarose gel electrophoresis, showing 800 bp of PCR amplicons (Figure 3a, Lane 2 and 3). The successful amplification of osmotin genes from the genomic DNA template by other findings (Chowdhury et al. 2015) demonstrates the straightforwardness of the amplification of the intronless eukaryotic gene. The multiple sequence analyses demonstrated the conservancy of essential amino acid residues of TcOSM1

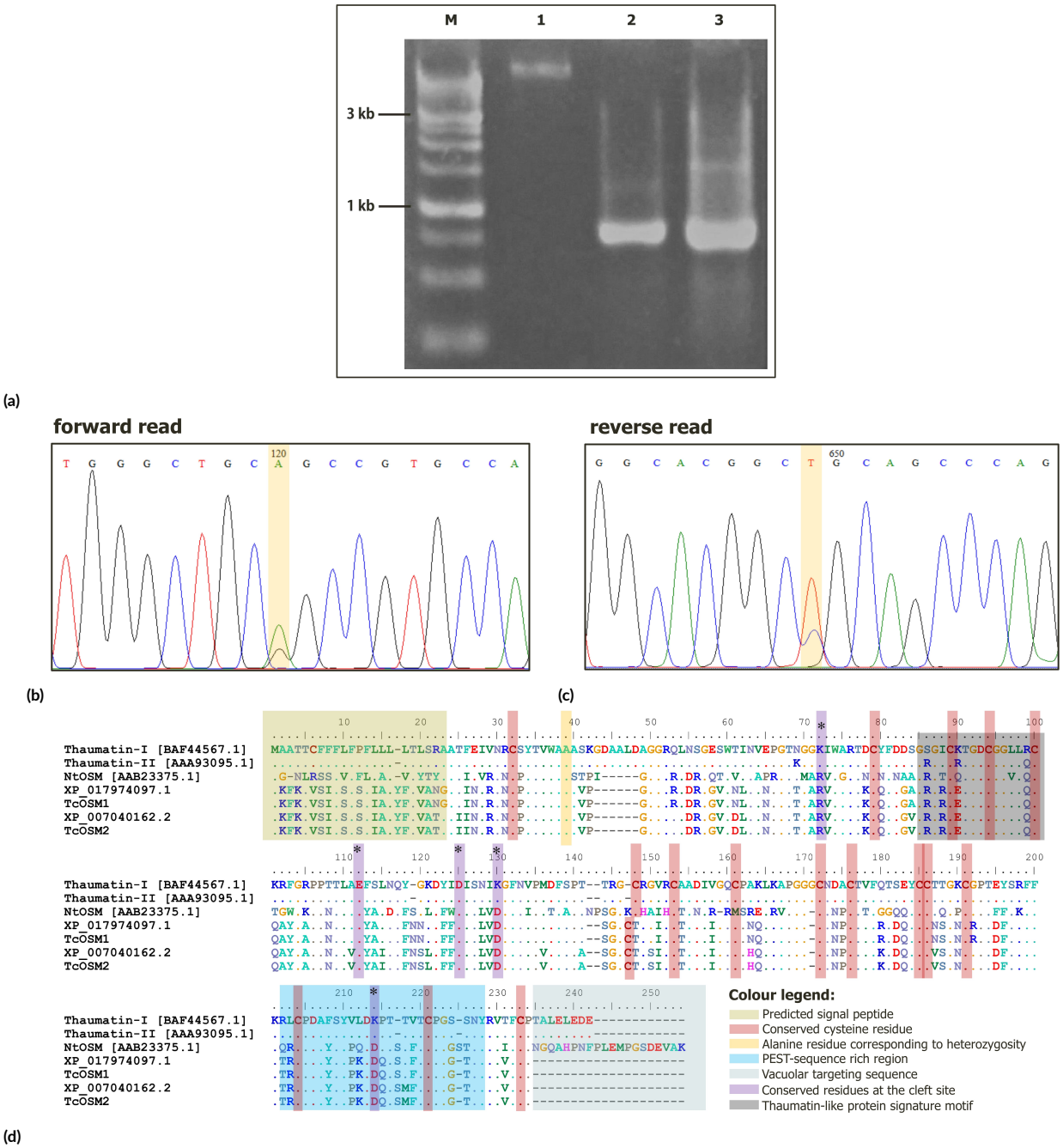


FIGURE 3 Sequence analyses of osmotin genes and proteins. a. The visualization of PCR amplicons in 0.8% agarose gel (M: 1 kb+ DNA marker, 1: Cacao genomic DNA isolated from cotyledons, 2: Amplified *TcOSM1* gene, and 3: Amplified *TcOSM2* gene); b. The forward read; and c. the reverse read after Sanger sequencing of *TcOSM1* showing heterozygosity; and d. Multiple sequence analysis of *TcOSM1* and *TcOSM2* against thaumatins and osmotins from *N. tabaccum* and *T. cacao* reference sequences.

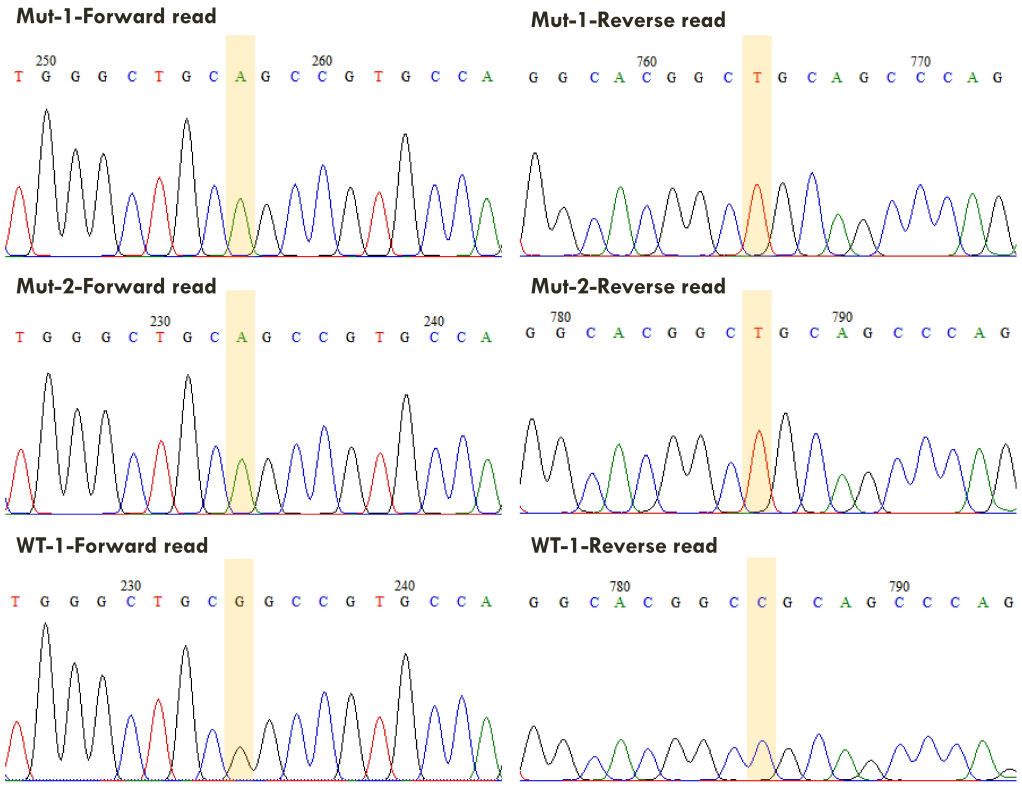


FIGURE 4 Sequence evaluation of the cloned osmotin genes. The chromatograms retrieved after re-sequencing of cloned osmotin genes to validate heterozygosity. Alignment with the reference (Wild-type/WT) sequences was carried out to deduce the mutant (Mut) sequences.

and TcOSM2 comparing to reference sequences. *TcOSM1* sequence depicted a genetic variance as heterozygosity, compared to the predicted gene sequence from the genome of the Criollo-type cacao used as reference (Figure 3b and c). Meanwhile, the *TcOSM2* sequence was exactly identical to the reference sequence (XP_007040162.2).

The finding in *TcOSM1* was confirmed using paired-end (forward and reverse read) Sanger sequencing and confirmed after two-peaks detection, each corresponding to different nucleotides. In general, Sanger sequencing will produce regular spaced and similar height peaks to generate an efficient base calling (Tenney et al. 2007). However, two peaks or double traces can be detected in base substitutions or indels, causing heterozygosity (Hill et al. 2014). Here, the heterozygosity was supported by clear reads of sequencing chromatograms at both left and right flanking bases suggesting high confidence in the base calling results (Figure 3b and c).

The cloning procedures were conducted to circumvent the downside of direct DNA Sanger sequencing from PCR amplicons when double peaks were found. The cloning has succeeded in discerning a nucleotide difference between two strands of DNA, which are otherwise detected as double peaks in the heterozygotic alleles. The independent PCR amplification of each strand enables us to make a distinction between the nucleotide variations. Each amplified strands were independently cloned into vectors and determined the nucleotide sequences. A similar procedure has been implemented to screen heterozygotic and homozygotic mutations caused by indels introduced during gene editing by CRISPR/Cas9 technology (Lawrenson et al. 2015; Ma et al. 2015). Therefore, the obtained heterozygosity in this study is well confirmed (Figure 4).

One of the allelic versions in *TcOSM1* heterozygosity was considered a mutant allele as it differed from the reference sequence (XP_017974097.1). The mutant allele contained a nucleotide change from G to A. This mutation occurred in the codon GCG (wild-type), which mutated the codon to GCA (mutant). However, the mutation did not change the corresponding alanine residue (Figure 3d). Nevertheless, further analysis using homology modeling was conducted to predict the influence of genetic variation in TcOSM1 and TcOSM2 upon the global protein structures; thus also conjecturing their functionalities.

3.3. TcOSM1 and TcOSM2 have conserved essentials amino acid residues belonging to TLP family

The sequence analyses of TcOSM1 and TcOSM2 revealed other interesting aspects (Figure 3d). TcOSM1 and TcOSM2 retained the N-terminal domain, which drives the peptide to the secretory pathway (Campos et al. 2002). Several PR family proteins, as well as TcOSM1 and TcOSM2, lack carboxyl-end signal peptide, which assists the osmotins transport to vacuole. This analysis has predicted that TcOSM1 and TcOSM2 are functionally expressed extracellularly or in the apoplastic regions.

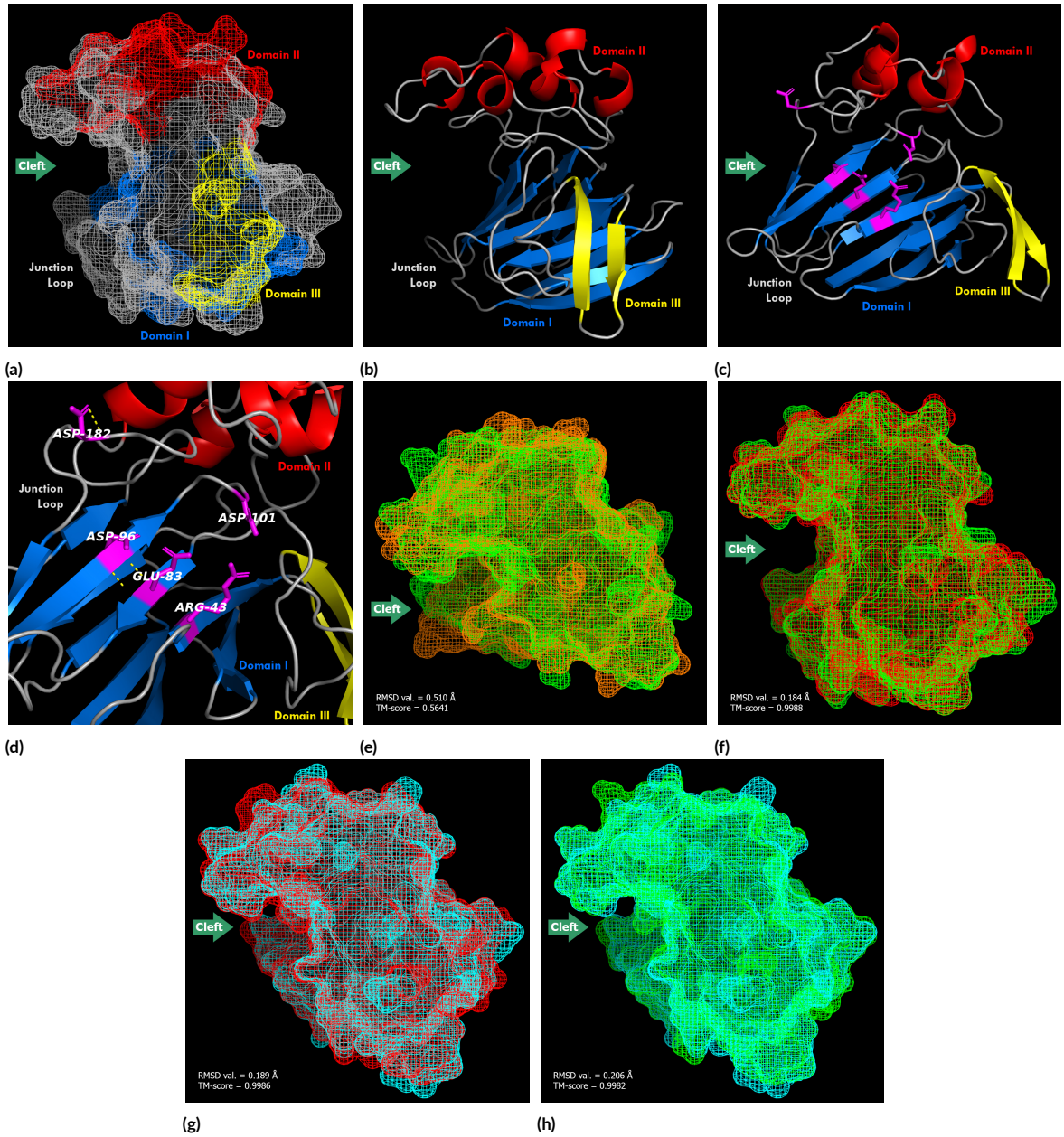


FIGURE 5 The three-dimensional structures of osmotins. The structure of TcOSM1 in mesh (a.) and cartoon ribbon (b.) represents all studied osmotins here, showing a profile consisting of domain I-III. The five highly conserved residues (c.) and (d.) in the osmotin cleft. The structural alignments (superimpositions): e. NtOSM (AAB23375.1) (orange) v.s. "Wild-type" sequence (XP_017974097.1) (green); f. "Wild-type" sequence (XP_017974097.1) (green) v.s. TcOSM1 (red); g. TcOSM1 (red) v.s. TcOSM2 (cyan); and h. "Wild-type" sequence (XP_017974097.1) (green) v.s. TcOSM2 (cyan).

Furthermore, the possibility of eight disulfide bridges in TcOSM1 and TcOSM2 were retained since all sixteen cysteine residues were intact. Moreover, the sequence analysis detected Proline (P) – Glutamic Acid (E) – Serine (S) – Threonine (T)-rich sequence (PEST motif) within TcOSM1 and TcOSM2 (Figure 3d). PEST motif is the tag for protein degradation, such as through ubiquitin-proteasome complex (Zhai et al. 2017). Comparing to other PR family proteins, TcOSM1 and TcOSM2 were predicted to have no potential glycosylated conserved residue. Therefore, they are not post-translationally glycosylated before secretion.

A further comparative study to Thaumatin-I and Thaumatin-II resulted in the identification of conserved TLP motif among osmotins from *N. tabacccum* and *T. cacao*. The TLP motif was firstly identified in the sweet-tasted protein of *Thaumatococcus daniellii* (Benn.) Benth. ex B.D. Jacks, a plant originated from Africa. It has a unique motif of G-x-G-x-C-x-T-G-D-C-G-G-L/V-L-x-C with x correspond to any amino acids (Liu et al. 2010). This result was in agreement with the domain query hits using CCD (Table 2), stating that NtOSM, TcOSM1, and TcOSM2 possessed conserved TLP motif (Accession ID: cl02511).

3.4. In silico modeling revealed implications of the TcOSMs amino acid variations to the protein structures

The global RMSD (Root-Means-Squared Deviation) values retrieved after structural alignments (superimpositions) of paired TcOSMs (Figure 5e, f, g, and h). The values showed insignificant differences between aligned (superimposed) models (RMSD ≤ 0.206 Å). The values depicted the global deviation length of the whole atom-to-atom aligned between the two models. The RMSD value ≤ 0.206 Å reflected highly identical structures even if the figures showed several unaligned protrusions among superimposed models.

It is not surprising that the calculated RMSD values were low, as the protein structures themselves had high identity (≥ 90%). However, The interpretation based on RMSD values can be slightly overestimated due to significant errors (Kufareva and Abagyan 2011). Therefore, it was accompanied by analyses based on TM-scores calculation, which included statistical consideration (Xu and Zhang 2010).

The TM-score > 0.5 between two superimposed protein structures indicates that they have the same fold (Zhang and Skolnick 2004). Xu and Zhang (2010) mentioned that a TM-score of 0.5 equals the probability of one uniquely matched fold among those of 1.8 million random superimposed protein pairs. TM-score > 0.5 signifies that to achieve a significant similarity of structural topology between two superimposed structures, one shall consider more than 1.8 million random pairs. Indeed, it is true, according to the same calculation, when TM-score = 0.6 means more than 90% of superimposed structures fall into the same fold (Xu and Zhang 2010). Therefore, it is safe to assume that all structures studied here are more than 90% to belong correctly to the same fold. Altogether, we can predict that the variations in amino acid sequences were not significantly meaningful in influencing the global structures of TcOSMs models.

The function of a protein can be inferred rationally from the one structure with known function since the distinction among structures reflects the uniqueness of their functions. The structure of NtOSM (AAB23375.1), which has been supported by X-ray crystallography study and other assays (Min et al. 2003), was used as a reference for structure and function. The TcOSMs models were not significantly different in structures toward the NtOSM as the RMSD value is 0.516 Å, while the TM-score value is still above 0.5.

The functionality of the TcOSMs could also be predicted since the clefts (green arrows), which presumably functioned in the binding to the receptors or ligands from fungal counterparts, were intact among the structures. The cleft is a signature of osmotins formed during the protein folding located between Domain I (blue) and Domain II (red) (Figure 5a and b.) (Ghosh and Chakrabarti 2008). Amino acid residues characterize the cleft of osmotin with acidic R-group. The residues within the cleft/pocket binding, i.e. arginine, glutamic acid, and three aspartic acids (Liu et al. 2010), were highly conserved in NtOSM and TcOSMs but varied in thaumatins (see Figure 3d and Figure 5c and d, residues were highlighted in purple and with asterisks). These differences between thaumatins and osmotins reflect their functionality and presumably evolutionary. Moreover, two residues, the aspartic acids (Asp 101 and Asp 182), are located within the junction loops, possibly offering dynamic interactions in osmotin binding. Meanwhile, the other residues are located within the beta-sheets secondary structures in Domain I. The residues also are predicted to support the global three-dimensional structure of osmotin via two distinct polar contacts: Asp 182 to protein backbone and two parallel residues, Glu 83 to Asp 96 (see Figure 5d, yellow dashed-lines).

3.5. Discussion

Osmotin is a member of the Pathogenesis-related protein family 5 (PR5) (van Loon et al. 2006), which shows the potential as a basis for molecular plant engineering to address the problems in cacao plantation and production. Plants generally express osmotin to respond to the environmental stresses, e.g. cold and drought, and exhibit an antifungal activity toward Saccharomyces cerevisiae, *Candida albicans*, *Pichia methanolica*, *Cryptococcus neoformans* (Tzou et al. 2011), *Phytophthora nicotianae* var. *parasitica*, *Fusarium solani* f. sp. Glycines, *C. gossypii* var. *cephalosporioides*, *Macrophomina phaseolina*, *Colletotrichum gloeosporioides* (Campos et al. 2008), *Pichia pastoris*, and *M. perniciosa* (Falcao et al. 2016).

The osmotin is a protein exhibiting a wide-spectrum of antifungal affectivity, suggesting a specific target in the surface of the plasma membrane of the target cell. The binding mode of osmotin to target through the cleft's mediation with five highly conserved amino acid residues remains elusive. The pathway of osmotin exerted cell cytotoxicity involves apoptosis in yeast model through activation of Mitogen-Activated Protein Kinase (MAPK), particularly the mating integration modules: Ste4, Ste5, Ste7, Fus3, Ste11, Ste12, St18, Ste20, and Kss1 (Yun et al. 1998). Interestingly, the fungal Ste7 was rapidly phosphorylated after osmotin exposure (Yun et al. 1998), usually at its threonine and tyrosine residues, which later phosphorylated the partner proteins at the same sites (Bardwell 2006). However, the active receptor of osmotin remains to be determined since osmotin can modulate the regulatory fungal mating elements without activating the pheromone receptor nor its associated G proteins.

Osmotin action has been reported by the identification of ORE20/PHO36 as the receptor of osmotin, which promotes a signal cascade pathway (Narasimhan et al. 2005). The ORE20/PHO36 is a surface protein with seven transmembrane domains that regulate lipid and phosphate metabolism of yeast (Yamauchi et al. 2003; Karpichev et al. 2002). The binding of osmotin to ORE20/PHO36 activates the RAS2-cAMP/PKA (Miele et al. 2011), which later, in turn, suppresses the stress response gene expressions (Narasimhan et al. 2001). The suppression of stress

response genes, such as *Msn2* and *Msn4*, leads to an increased susceptibility to oxidative stress caused by radical-oxygen species (ROS), which under metabolic burden will induce apoptosis/necrosis (Nakazawa et al. 2018). Although many of the concerted elements remain to be determined, as well as the more precise mechanism of osmotin binding to ORE20/PHO36 via the five highly conserved residues resided within the cleft, osmotin can affect the yeast's cellular signaling, which in turn kill the yeast itself.

The bioinformatics approaches and molecular biology protocols have uncovered the characteristics of osmotins. The evidence produced by this study sufficiently supported the functionality prediction of the TcOSMs as antifungal protein. Besides, the mutation in one of the *TcOSM1* alleles was considered to be harmless since it did not change the corresponding amino acid residue and located outside the essential regions of osmotin.

Careful interpretations have also been made after the in silico analyses in predicting the intact activity of TcOSMs based on the structural changes. The TM-score (Xu and Zhang 2010), in conjunction with RMSD analysis, has increased the confidence of interpretations made in this study. In silico modeling produced idealized models using assumptions, which can differ compared to the actual conditions *in vivo*. Therefore, better models or structures deduced using biophysical analyses and another method in analyzing structures such as the contact-based method (Ding et al. 2018) should be applied in the future. Further studies are also implied using biochemical assays and heterologous expression in other model organisms.

This research will open numerous other research in producing elite cacao cultivar. Moreover, it will also promote other research and applications in human health since osmotin also affects human fungal pathogens (Viktorova et al. 2017). Furthermore, osmotin resembles adiponectin, which exhibits anti-tumor activity, inhibits endothelial proliferation and migration, and reduces atherosclerosis (Anil Kumar et al. 2015).

4. Conclusions

The research has succeeded in determining the two osmotin proteins, TcOSM1 and TcOSM2, from local cacao cultivars. Moreover, this research provided preliminary study opening opportunities to establish elite cacao cultivars using plant molecular engineering in Indonesia and other applications in human health.

Acknowledgments

IBN thanked Miranti N. Sari and Aryo P. Purwanto for the assistance. IBN also thanked Dr. Riza A. Putranto for the valuable discussions and Dini A. Sari, who provided the Plant Genomic DNA Mini Kit. IBN and F also thanked Tomy Adrianto, who supplied ripe cocoa fruits, and Dr. Yopi Sunarya for the E. coli competent cells.

Authors' contributions

IBN and F designed the research. F supervised the research and contributed to manuscript revision. IBN conducted database gene mining, cloning, sequence analyses, homology modeling, and wrote the manuscript. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interest.

References

Ali SS, Asman A, Shao J, Firmansyah AP, Susilo AW, Rosmana A, McMahon P, Junaid M, Guest D, Kheng TY, Meinhardt LW, Bailey BA. 2019. Draft genome sequence of fastidious pathogen *Ceratobasidium theobromae*, which causes vascular-streak dieback in *Theobroma cacao*. Fungal Biol Biotechnol. 6(1):14. doi:10.1186/s40694-019-0077-6.

Ali SS, Shao J, Lary DJ, Strem MD, Meinhardt LW, Bailey BA. 2017. *Phytophthora megakarya* and *P. palmivora*, causal agents of black pod rot, induce similar plant defense responses late during infection of susceptible cacao pods. Front Plant Sci. 8. doi:10.3389/fpls.2017.00169.

Anil Kumar S, Hima Kumari P, Shravan Kumar G, Mohanalatha C, Kavi Kishor PB. 2015. Osmotin: a plant sentinel and a possible agonist of mammalian adiponectin. Front Plant Sci. 6. doi:10.3389/fpls.2015.00163.

Bailey BA, Evans HC, Phillips-Mora W, Ali SS, Meinhardt LW. 2018. *Moniliophthora roreri*, causal agent of cacao frosty pod rot: frosty pod rot of cacao. Mol Plant Pathol. 19(7):1580–1594. doi:10.1111/mpp.12648.

Bardwell L. 2006. Mechanisms of MAPK signalling specificity. Biochem Soc Trans. 34(5):837–841. doi:10.1042/BST0340837.

Campos M, Ribeiro S, Rigden D, Monte D, Grossi de sa M. 2002. Putative pathogenesis-related genes within *Solanum nigrum* L. var. *americanum* genome: isolation of two genes coding for PR5-like proteins, phylogenetic and sequence analysis. Physiol Mol Plant Pathol. 61(4):205–216. doi:10.1006/pmpp.2002.0430.

Campos MA, Silva MS, Magalhães CP, Ribeiro SG, Sarto RP, Vieira EA, Grossi de Sá MF. 2008. Expression in *Escherichia coli*, purification, refolding and antifungal activity of an osmotin from *Solanum nigrum*. Microb Cell Fact. 7(1):7. doi:10.1186/1475-2859-7-7.

Chiong KT, Damaj MB, Padilla CS, Avila CA, Pant SR, Mandadi KK, Ramos NR, Carvalho DV, Mirkov TE. 2017. Reproducible genomic DNA preparation from

diverse crop species for molecular genetic applications. Plant Methods. 13(1):106. doi:10.1186/s13007-017-0255-6.

Chowdhury S, Basu A, Kundu S. 2015. Cloning, characterization, and bacterial over-expression of an osmotin-like protein gene from *Solanum nigrum* L. with antifungal activity against three necrotrophic fungi. Mol Biotechnol. 57(4):371–381. doi:10.1007/s12033-014-9831-4.

Ding W, Mao W, Shao D, Zhang W, Gong H. 2018. DeepConPred2: an improved method for the prediction of protein residue contacts. Comput Struct Biotechnol. 16:503–510. doi:10.1016/j.csbj.2018.10.009.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797. doi:10.1093/nar/gkh340.

Falcao LL, Silva-Werneck JO, Ramos AdR, Martins NF, Bresso E, Rodrigues MA, Bemquerer MP, Marcellino LH. 2016. Antimicrobial properties of two novel peptides derived from *Theobroma cacao* osmotin. Peptides. 79:75–82. doi:10.1016/j.peptides.2016.03.006.

Farquharson KL. 2014. The fungus, the witches' broom, and the chocolate tree: deciphering the molecular interplay between *Moniliophthora perniciosa* and *Theobroma cacao*. Plant Cell. 26(11):4231–4231. doi:10.1105/tpc.114.133462.

Fong JH, Marchler-Bauer A. 2008. Protein subfamily assignment using the Conserved Domain Database. BMC Res Notes. 1(1):114. doi:10.1186/1756-0500-1-114.

Ghosh R, Chakrabarti C. 2008. Crystal structure analysis of NP24-I: a thaumatin-like protein. Planta. 228(5):883–890. doi:10.1007/s00425-008-0790-5.

Hill JT, Demarest BL, Bisgrove BW, Su YC, Smith M, Yost HJ. 2014. Poly peak parser: method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products: POLY PEAK PARSER. Dev Dyn. 243(12):1632–1636. doi:10.1002/dvdy.24183.

Karpichev IV, Cornivelli L, Small GM. 2002. Multiple regulatory roles of a novel *Saccharomyces cerevisiae* protein, encoded by *YOL002c* , in lipid and phosphate metabolism. J Biol Chem. 277(22):19609–19617. doi:10.1074/jbc.M202045200.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 10(6):845–858. doi:10.1038/nprot.2015.053.

Kufareva I, Abagyan R. 2011. Methods of protein structure comparison. In: AJW Orry, R Abagyan, editors, Homology Modeling, volume 857. Totowa, NJ: Humana Press. p. 231–257. doi:10.1007/978-1-61779-588-6_10.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 35(6):1547–1549. doi:10.1093/molbev/msy096.

Lawrenson T, Shorinola O, Stacey N, Li C, Østergaard L, Patron N, Uauy C, Harwood W. 2015. Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. Genome Biol. 16(1):258. doi:10.1186/s13059-015-0826-7.

Liu JJ, Sturrock R, Ekramoddoullah AKM. 2010. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Rep. 29(5):419–436. doi:10.1007/s00299-010-0826-8.

Ma X, Chen L, Zhu Q, Chen Y, Liu YG. 2015. Rapid decoding of sequence-specific nuclease-induced heterozygous and biallelic mutations by direct sequencing of PCR products. Mol Plant. 8(8):1285–1287. doi:10.1016/j.molp.2015.02.012.

Miele M, Costantini S, Colonna G. 2011. Structural and functional similarities between osmotin from *Nicotiana tabacum* seeds and human adiponectin. PLoS One. 6(2):e16690. doi:10.1371/journal.pone.0016690.

Min K, Ha SC, Hasegawa PM, Bressan RA, Yun DJ, Kim KK. 2003. Crystal structure of osmotin, a plant antifungal protein. Proteins: Struct Funct Bioinf. 54(1):170–173. doi:10.1002/prot.10571.

Mithöfer D, Roshetko JM, Donovan JA, Nathalie E, Robiglio V, Wau D, Sonwa DJ, Blare T. 2017. Unpacking 'sustainable' cocoa: do sustainability standards, development projects and policies address producer concerns in Indonesia, Cameroon and Peru? Int J Biodivers Sci Ecosyst Serv Manag. 13(1):444–469. doi:10.1080/21513732.2018.1432691.

Nakazawa N, Yanata H, Ito N, Kaneta E, Takahashi K. 2018. Oxidative stress tolerance of a spore clone isolated from Shirakami kodama yeast depends on altered regulation of *Msn2* leading to enhanced expression of ROS-degrading enzymes. J Gen Appl Microbiol. 64(4):149–157. doi:10.2323/jgam.2017.11.002.

Narasimhan ML, Coca MA, Jin J, Yamauchi T, Ito Y, Kadowaki T, Kim KK, Pardo JM, Damsz B, Hasegawa PM, Yun DJ, Bressan RA. 2005. Osmotin is a homolog of mammalian adiponectin and controls apoptosis in yeast through a homolog of mammalian adiponectin receptor. Mol Cell. 17(2):171–180. doi:10.1016/j.molcel.2004.11.050.

Narasimhan ML, Damsz B, Coca MA, Ibeas JI, Yun DJ, Pardo JM, Hasegawa PM, Bressan RA. 2001. A plant defense response effector induces microbial apoptosis. Mol Cell. 8(4):921–930. doi:10.1016/S1097-2765(01)00365-3.

Nugroho IB, Handayani NSN. 2016. Primer design and *in silico* analysis using CLUSTALW and MUSCLE for L-arabinose isomerase (*araA*) gene detection in thermophilic bacteria. In: AIP Conference Proceedings. p. 140007. doi:10.1063/1.4958568.

Putranto RA, Martiansyah I, Saptari RT. 2017. *In silico* identification and comparative analysis of *Hevea brasiliensis* COBRA gene family. In: Proceeding International Conference on Science and Engineering,

volume 1. p. 39–47. URL http://sunankalijaga.org/prosiding/index.php/icse/article/view/267.

Soltis DE, Soltis PS. 2003. Applying the bootstrap in phylogeny reconstruction. Stat Sci. 18(2):256–267. doi:10.1214/ss/1063994980.

Souza PAd, Moreira LF, Sarmento DH, da Costa FB. 2018. Cacao — *Theobroma cacao*. In: Exotic Fruits. Elsevier. p. 69–76. doi:10.1016/B978-0-12-803138-4.00010-1.

Tenney AE, Wu JQ, Langton L, Klueh P, Quatrano R, Brent MR. 2007. A tale of two templates: automatically resolving double traces has many applications, including efficient PCR-based elucidation of alternative splices. Genome Res. 17(2):212–218. doi:10.1101/gr.5661407.

Tzou YM, Huang TS, Huggins KW, Chin BA, Simonne AH, Singh NK. 2011. Expression of truncated tobacco osmotin in *Escherichia coli*: purification and antifungal activity. Biotechnol Lett. 33(3):539–543. doi:10.1007/s10529-010-0453-z.

van Loon L, Rep M, Pieterse C. 2006. Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol. 44(1):135–162. doi:10.1146/annurev.phyto.44.070505.143425.

Viktorova J, Rehorova K, Musilova L, Suman J, Lovecka P, Macek T. 2017. New findings in potential applications of tobacco osmotin. Protein Expr Purif. 129:84–93. doi:10.1016/j.pep.2016.09.008.

Wickramasuriya AM, Dunwell JM. 2018. Cacao biotechnology: current status and future prospects. Plant Biotechnol J. 16(1):4–17. doi:10.1111/pbi.12848.

Wu J, Khan AA, Shih CYT, Shih DS. 2001. Cloning and sequence determination of a gene encoding an osmotin-like protein from strawberry (Fmgaria x ananassa Duch.). DNA Seq. 12(5-6):447–453. doi:10.3109/10425170109084473.

Xu J, Zhang Y. 2010. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics. 26(7):889–895. doi:10.1093/bioinformatics/btq066.

Yamauchi T, Kamon J, Ito Y, Tsuchida A, Yokomizo T, Kita S, Sugiyama T, Miyagishi M, Hara K, Tsunoda M, Murakami K, Ohteki T, Uchida S, Takekawa S, Waki H, Tsuno NH, Shibata Y, Terauchi Y, Froguel P, Tobe K, Koyasu S, Taira K, Kitamura T, Shimizu T, Nagai R, Kadowaki T. 2003. Cloning of adiponectin receptors that mediate antidiabetic metabolic effects. Nature. 423(6941):762–769. doi:10.1038/nature01705.

Yun DJ, Ibeas JI, Lee H, Coca MA, Narasimhan ML, Uesono Y, Hasegawa PM, Pardo JM, Bressan RA. 1998. Osmotin, a plant antifungal protein, subverts signal transduction to enhance fungal cell susceptibility. Mol Cell. 1(6):807–817. doi:10.1016/S1097-2765(00)80080-5.

Zhai Z, Liu H, Shanklin J. 2017. Phosphorylation of WRINKLED1 by KIN10 results in its proteasomal degradation, providing a link between energy homeostasis and lipid biosynthesis. Plant Cell. 29(4):871–889. doi:10.1105/tpc.17.00019.

Zhang D, Motilal L. 2016. Origin, dispersal, and current global distribution of cacao genetic diversity. In: B Bailey, L Meinhardt, editors, Cacao Diseases: A History of Old Enemies and New Encounters. Springer. p. 3–31. doi:10.1007/978-3-319-24789-2_1.

Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. Proteins: Struct Funct Bioinf. 57(4):702–710. doi:10.1002/prot.20264.