Predicted of a heat shock cognate 70 kDa protein by PHYRE\textsuperscript{2} and SWISS-MODEL software from \textit{Elaeis guineensis} and \textit{E. oleifera}

R. Hayati\textsuperscript{1}, M Basyuni\textsuperscript{2,*}, Lisnawita\textsuperscript{3} and E Munir\textsuperscript{4}

\textsuperscript{1} Doctoral Program of Agricultural Sciences, Faculty of Agriculture, Universitas Sumatera Utara, Jl. Prof. A. Sofyan No. 3, Padang Bulan 20155, Indonesia  
\textsuperscript{2} Department of Forestry, Faculty of Forestry, Universitas Sumatera Utara, Jl. Tri Dharma Ujung No. 1 Medan, North Sumatera, 20155, Indonesia  
\textsuperscript{3} Program Study of Agrotechnology, Faculty of Agriculture, Universitas Sumatera Utara, Jl. Prof. A. Sofyan No. 3, Padang Bulan 20155, Indonesia  
\textsuperscript{4} Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, Jl. Bioteknology No. 1, Padang Bulan 20155, Indonesia  

\textsuperscript{*}E-mail: m.basyuni@usu.ac.id

\textbf{Abstract.} Protein modeling was obtained from \textit{Elaeis guineensis} and \textit{E. oleifera}. The identification of heat shock cognate 70 kDa from both very similar to coverage 90% analyzed by SWISS-MODEL. In addition, the data from similarity, identity, range, and Q-mean were distinct. According to the PHYRE\textsuperscript{2} analysis, we observed the longest residue presenting in \textit{E. guineensis} (15-348). Wherein the 3D-ligand site prediction confirms that many amino acids were found (GLY, THR, TYR, LYS, GLU, ARG, and SER). Interestingly, the heat shock cognate 70 kDa protein in \textit{E. oleifera} has a slight residue (15-18), and also have been three amino acids (GLY, THR, TYR). Appropriately, the heat shock cognate-like protein involved in different species, likewise playing a distinct response to kind the defense. The different homology images and templates were identified as different binding site predictions. These models can be used for supporting the \textit{E. guineensis} and \textit{E. oleifera} proteins modeling, mainly in the molecular study.

\section{1. Introduction}
There are two species of oil palm, \textit{Elaeis guineensis} Jacq. from the tropic of Africa and \textit{E. oleifera} from Northern and Central South America [1]. A total of 4,116 and 4,081 proteins have been reported from \textit{Elaeis guineensis} (D × P) and \textit{E. oleifera}, respectively [2]. The heat shock cognate-protein (LMW Class) have been reported to characterize for developing of somatic embryogenesis to \textit{E. guineensis} Jacq. seeds [3]. The transcript factor of HSP70 (heat shock protein 70 kDa) or \textit{EgRBP42} genes belongs to responsive in encoding RNA in \textit{E. guineensis} when abiotic stresses [4]. A total of 947 proteins including the heat shock 70 kDa have been reported as the allergenic potent for immunology studying in macadamia of nut used the shotgun proteomic approach to the allergenic potent [5].
Furthermore, the heat shock proteins were found to stabilize other membranes and protein in *E. oleifera* and *E. guineensis* under plant conditions of stress [6]. The heat shock protein 90 kDa (HSP90) also used in cloning the transcripts of upregulated in heat stress to *Pinellia ternata* leaves, and obtained the full-length cDNAs of PtsHSP to transcripts of cloning upregulation [7]. The full-assembled constrained silencing of heat shock protein bindings for transcript target along with the complementary of sequence miRNA from argonaute proteins to the plant cytoplasm [8]. The heat shock protein protected the cells with the supporting in an ATP manner for the process folding of the protein in pathological, physiological, and environmental [9]. The heat shock response 70 kDa (HSP70) and 25 kDa (HSP25) have a playing role to defend proteostasis through the cell stress with a sensor as the regulator from protein [10].

Besides, the plant type transgenics also can produce heat shock proteins (LimHSP16.45 genes) in responses when the advance temperature (abiotic stresses) in plants [11]. Several of the heat shock protein 60 kDa using bioinformatic analysis considered as a key the chaperone protein to maintaining the mitochondrial function, and regulating of homeostasis for mitochondrial [12]. Another case, the small heat shock proteins (sHsps) integrated activity of prevented aggregation in the folding protein activity to the in vitro system for mechanism and action from plants [13]. Given the function of heat shock cognate-like protein found in several plants, becomes interesting to understand the protein modeling. By reason of, the distinct proteins from *E. guineensis* and *E. oleifera* have not been studying. This paper describes the heat shock cognate 70 kDa protein modeling of different online software occurrences in *E. guineensis* and *E. oleifera* species.

### 2. Materials and method

#### 2.1. Protein collection

Heat shock cognate 70 kDa proteins from National Center for Biotechnology Information (NCBI) database cDNA and amino acid (https://www.ncbi.nlm.nih.gov/) from *Elaeis guineensis* and *Elaeis oleifera* were collected. The Accession numbers of protein sequences followed the GenBank accession: XM\_010924055; XM\_010939344 (*E. guineensis*), and ES414651; ES370602 (*E. oleifera*).

#### 2.2. The template sources proteins and modeling

The template selection and builds a model of the 3-dimensional structure of the protein was performed by online software the PHYRE\(^2\) (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) [14], compared the SWISS-MODEL (https://swissmodel.expasy.org/) [15]. The highest quality of the template was selected to model protein structure.

#### 2.3. The 3D ligand site prediction

The 3D ligand-binding site prediction of the protein build structure analyzed using online software http://www.sbg.bio.ic.ac.uk/3dligandsite/ [16].

### 3. Results and Discussions

#### 3.1. Heat shock cognate 70 kDa proteins modeling

The heat shock cognate 70 kDa protein modeling analyzed by PHYRE\(^2\) from *Elaeis guineensis* (c3d2fC) and *Elaeis oleifera* (c3c7nB) summary was shown in Table 1. Both protein genes have not similar results except in confidence value (100.00). The *E. guineensis* has an aligned sequence of showed 95.00, but *E. oleifera* showed 93.00. On the other hand, the highest coverage and residue value were found in *E. guineensis* (0.95 and 613).
Table 1. The heat shock cognate 70 kDa protein modelling by PHYRE$^2$

| ID protein | Species       | Sequence aligned | Coverage | Residue | Confidence |
|------------|---------------|------------------|----------|---------|------------|
| c3d2fC     | E. guineensis | 95.00            | 0.95     | 613     | 100.00     |
| c3c7nB     | E. oleifera   | 93.00            | 0.94     | 149     | 100.00     |

However, the SWISS-MODEL was shown in Table 2. *E. oleifera* showed the highest value in similarity sequence (0.54), identity (79.58), and Q-mean (0.82), but compared the *E. guineensis* (similarity sequence: 0.44), (identity: 52.39), and (Q-mean: 0.75). The range of *E. guineensis* sequence was obtained from 84-599, but *E. oleifera* from 85-168.

Table 2. The heat shock cognate 70 kDa protein modelling by SWISS-MODEL

| ID protein | Species       | Similarity | Identity | Coverage | Range   | Q-mean |
|------------|---------------|------------|----------|----------|---------|--------|
| 2kho.1.A   | *E. guineensis* | 0.44       | 52.39    | 0.90     | 84-599  | 0.75   |
| 3i33.1.A   | *E. oleifera*  | 0.54       | 79.58    | 0.90     | 85-168  | 0.82   |

3.2. The image of protein modeling

A big homology model in which 613 residues (100% confidence) resulted from 95% coverage a single the highest template score in *E. guineensis* analyzed by PHYRE$^2$ was shown in Figure 1A. Whereas, the homology model *E. oleifera* was shown in Figure 1B. The lowest coverage value by SWISS-MODEL predicted that 90% were shown in Figures 2A and 2B, respectively. Recently, the distinct triterpene proteins have also been reported using PHYRE$^2$ and SWISS-MODEL for isoprenoids synthesis in mangrove species [17].

![A](image1.png) ![B](image2.png)

Figure 1. The PHYRE$^2$ build model from *E. guineensis*, model dimensions (Å): $X$:65.035 $Y$:72.494 $Z$:105.632 (A) and *E. oleifera*, model dimensions (Å): $X$:51.400 $Y$:39.242 $Z$:41.679(B). Image colored by rainbow N → C terminus.
Figure 2. *E. guineensis* (A) and *E. oleifera* (B) proteins by SWISS-MODEL

3.3. The 3D ligand-binding site prediction

The heat shock cognate 70 kDa ligand-binding site prediction from *E. guineensis* and *E. oleifera* was shown in Table 3. The seventeen residues from 15-348 confirm adding several amino acids to *E. guineensis* (GLY, THR, TYR, LYS, GLU, ARG, SER). A variety of contact form 13-23. Moreover, the longest average distance was found in SER (0.50). The cycloartenol synthase genes from KcCAS have been reported in the highest average distance 3D-Ligand binding site 0.69 in *Kandelia obovata* [18]. In another case of *E. oleifera*, the residues from 15-18 were found the amino acids of GLY, THR, and TYR. Thus, the short average distance from 0.01-0.02 with a variety of contact value from 21-23.

| Species  | Amino acid | Residue | Average distance | Contact |
|----------|------------|---------|------------------|---------|
| *E. guineensis* | GLY        | 15      | 0.04             | 23      |
|           | THR        | 16      | 0.00             | 22      |
|           | THR        | 17      | 0.00             | 22      |
|           | TYR        | 18      | 0.00             | 22      |
|           | LYS        | 74      | 0.33             | 20      |
|           | GLY        | 207     | 0.01             | 22      |
|           | GLY        | 208     | 0.00             | 22      |
|           | GLY        | 209     | 0.06             | 22      |
|           | GLY        | 236     | 0.09             | 22      |
|           | GLU        | 237     | 0.09             | 19      |
|           | GLU        | 274     | 0.00             | 22      |
|           | LYS        | 277     | 0.00             | 22      |
|           | ARG        | 278     | 0.47             | 19      |
|           | SER        | 281     | 0.50             | 13      |
|           | GLY        | 345     | 0.00             | 22      |
|           | ARG        | 348     | 0.00             | 22      |
| *E. oleifera*  | GLY        | 15      | 0.01             | 21      |
|           | THR        | 16      | 0.01             | 21      |
|           | THR        | 17      | 0.02             | 22      |
|           | TYR        | 18      | 0.02             | 23      |
4. Conclusions
This study demonstrated the protein modeling presents of differentially proteins between Elaeis guineensis and E. oleifera species. According to the PHYRE² analysis, we can observe the longest residue presenting in E. guineensis. Interestingly, the heat shock cognate 70 kDa protein in E. oleifera has a low residue by SWISS-MODEL analysis. Appropriately, the heat shock cognate-like protein involved playing a distinct response to kind the defense.

Acknowledgment
The PMDSU project supported this study: 11/E1/ KP.PTNB/2019, from Directorate of Research and Higher Education, Republic of Indonesia.

References
[1] Soh AC., Wong CK., Ho YW. and Chong CW. 2009. Oil palm, in Handbook of plant breeding. ed. by Vollman J and Racjcan I. Springer, London, 4 333-367
[2] Singh R., Ong-Abdullah M., Low ET., Manaf MA., Rosli R., Nookiah. R. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature 500 335–339.
[3] Silva CR., Carmo LST., Luis ZG., Silva LP., Scherwinski-Pereira JE., and Mehta A. 2014. Proteomic identification of differentially expressed proteins during the acquisition of somatic embryogenesis in oil palm (Elaeis guineensis Jacq.). J. Proteomics 104 112-127.
[4] Yeap WC., Ooi TEK., Namasivayam P., Kulaveerasingam H., and Ho CL. 2012. EgRBP42 encoding an hnRNP-like RNA-binding protein from Elaeis guineensis Jacq. is responsive to abiotic stresses. Plant Cell Rep. 31 1829-1843.
[5] Rost J., Muralidharan S., and Lee NA. 2020. A label-free shotgun proteomics analysis of macadamia nut. Food Res. Int. 129 108838.
[6] Almeida RF., Santos IR., Meira FS., Grynberg P., Lopes R., da-Cunha RNV. and Mehta A. 2019. Differential protein profiles in interspecific hybrids between Elaeis oleifera and E. guineensis with contrasting responses to somatic embryogenesis competence acquisition. Plant Cell. Tiss. Org. 137 11-21.
[7] Lu H., Xue T., Zhang A., Sheng W., Zhu Y., Chang L. and Xue J. 2013. Construction of an SSH library of Pinellia ternata under heat stress, and expression analysis of four transcripts. Plant Mol. Biol. Rep. 31 185-194.
[8] Budak H. and Akpinar BA. 2015. Plant miRNAs: biogenesis, organization and origins. Funct. Integr. Genomic. 15 523–531.
[9] Elmallah ML., Cordonnier M., Vautrot V., Chanteloup G., Garrido C. and Gobbo J. 2020. Membrane-anchored heat-shock protein 70 (Hsp70) in cancer. Cancer Lett 469 134.
[10] Crul T., Csobo B., Gombos I., Marton A., Peter M., Balogh G. and Vigh L. 2020. Modulation of plasma membrane composition and microdomain organization impairs heat shock protein expression in B16-F10 mouse melanoma cells. Cells 9 951.
[11] Yang R., Yu G., Li H., Li X. and Mu C. 2020. Overexpression of small heat shock protein LimHSP16. 45 in Arabidopsis hsp17. 6II mutant enhances tolerance to abiotic stresses. Russ. J. Plant. Physi. 67 231-241.
[12] Wu X., Guo J., Chen Y., Liu X., Yang G., Wu Y. and Deng H. 2020. The 60-kDa heat shock protein regulates energy rearrangement and protein synthesis to promote proliferation of multiple myeloma cells. Br. J. Haematol. 10 16569.
[13] Avelange-Macherel MH., Rolland A., Hinault MP., Tolleter D., and Macherel D. 2020. The mitochondrial small heat shock protein HSP22 from pea is a thermosoluble chaperone prone to co-precipitate with unfolding client proteins. Int. J. Mol. Sci. 21 97.
[14] Kelley LA., Mezulis S.,Yates CM., Wass MN., Sternberg MJE. 2015. The Phyre² web portal for protein modeling, prediction and analysis. Nat. Protocols 10 845–858.
[15] Arnold K., Bordoli L., Kopp J. and Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22** 195–201.

[16] Wass MN., Kelley LA. and Sternberg MJ. 2010 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acid Res.* **38** 469–473.

[17] Basyuni M., Wati R., Sulistiyono N., Hayati R., Oku H., Baba S. and Sagami H. 2018. Protein modelling of triterpene synthase genes from mangrove plants using Phyre² and Swiss-model. *J. Phys. Conf. Ser.* **978** 012095

[18] Basyuni M., Sulistiyono N., Wati R., Oku H., Baba S. and Sagami H. 2018. Predicted cycloartenol synthase protein from *Kandelia obovata* and *Rhizophora stylosa* using online software of Phyre² and Swiss-model. *J. Phys. Conf. Ser.* **978** 012077