Research Article

**In Silico Analysis of Phalaenopsis Orchid Homeobox1 (POH1) Functional Gene for Shoot Development in Phalaenopsis Orchid**

Nuzlan Rasjid¹, Febri Yuda Kurniawan¹, Saifa Usni Putri¹, Aviesta Linggabuwana¹, Ireneus Seno Prasojo¹, Endang Semiarti*¹

¹)Department of Tropical Biology, Faculty of Biology, Universitas Gadjah Mada, Sleman, D.I. Yogyakarta 55281, Indonesia.
²)Study Program of Biotechnology, Graduate School, Universitas Gadjah Mada, Sleman, D.I. Yogyakarta 55281, Indonesia.
* Corresponding author, email: endsemi@ugm.ac.id

**Keywords:**
- POH1
- Conserved Domain
- Transcription Factor
- Protein Modelling
- Phalaenopsis

Submitted: 17 April 2023
Accepted: 26 June 2023
Published: 16 October 2023

**ABSTRACT**

The most favorite ornamental crop in Indonesia is orchid which benefited as floriculture. Therefore, the quality of this crop must be improved. Biotechnology is appropriate to be used to improve the quality and quantity of orchid plants. To conduct this method, researchers must know what genes function in plant development. In *Phalaenopsis* orchids, the gene has been identified as homeobox genes called *Phalaenopsis Orchid Homeobox1* (POH1). This research aims to conduct *in silico* analysis of the gene. The materials were retrieved from mRNA and amino acid databases. Then, the materials are aligned, visualized, motif location analysis, motif function discovery, phylogenetic construction, and protein 3D structural modelling. Based on mRNA and amino acid alignment, there are 4 domain regions that are conserved in POH1 and other homologous genes, such as KNOX1, KNOX2, ELK Domain, and Homeobox KN Domain, which roles as a transcription factor involved in plant development. SWISS-MODEL and ColabFold were used in protein modelling of the protein. By ColabFold modelling, the modelling prediction uses 325 residues, higher than SWISS-MODEL in 59 residues. ColabFold validation by Ramachandra Plot depicts having the most favourite regions is 68.6%, while SWISS-MODEL is 92.3%. Another validation parameter is overall quality factor and QMEAN Score. Protein modelling by ColabFold has overall quality factor 89.252 and QMEAN score 0.41 ± 0.05. However, SWISS-MODEL 3D prediction has overall quality factor 98.039 and QMEAN score of 0.71 ± 0.11.

Copyright: © 2023, J. Tropical Biodiversity Biotechnology (CC BY-SA 4.0)

**INTRODUCTION**

Orchids are the most favourite ornamental crops in Indonesia. Approximately, there are 5,000 orchids out of 30,000 orchids found throughout the world's third-largest area of tropical rainforest (Semiarti et al. 2015). They have been used in the floriculture trade, e.g. cut flowers and potted plants (Chandra De et al. 2014). Furthermore, the quality of the crops should be improved by biotechnology methods. To use this method, we must have a comprehensive understanding of orchid development (Hossain et al. 2013).

All plants during their life cycle will grow into three phases, namely, the embryogenic phase, the vegetative phase, and the generative phase (Howell 1998). Each phase is characterised by phenotypic change as ef-
fect of the expression of functional genes. These genes interact with each other and one of them is the key gene that will drive into the next phase (Howell 1998; Moore 2013). These are called homeobox genes that function in developmental transcription factors and are identified in plants and animals (Holland 2013; Viola & Gonzalez 2016).

In plants, there are 14 classes that are classified as homeobox genes and one of the genes is the class-1 KNOTTED1-like homeobox (KNOX1) gene. It is identified as a transcriptional factor for maintaining Shoot Apical Meristem (SAM) (Scofield et al. 2008). The SAM in plants comprises indeterminate cells that continually regenerate and then form structures above ground. In the next step, it contributes to the formation of plant shoots (Howell 1998). In orchids, there are two growth habits that are known as sympodial and monopodial. *Phalaenopsis* orchids have monopodial growth habit that depicts from one stem and grows pointedly upward (Zahara & Win 2019). A gene that has been identified in *Phalaenopsis* orchids called *Phalaenopsis Orchid Homeobox1* (POH1), which is KNOX1 homologous gene (Semiarti et al. 2008). The gene affects shoot formation of orchid in earlier stages of plant development, mainly in 4–16 weeks and after 48 weeks development, based on its expression analysis (Semiarti et al. 2016). This article intends to perform in silico analysis of the genes.

**MATERIALS AND METHODS**

**Materials**

The research was conducted from December 2022 to March 2023. The materials were retrieved from mRNA and amino acid databases. Then, they would be analysed by bioinformatic tools. The mRNA and amino acid sequence had been obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/) and OrchidBase 6.0 (https://cosbi.ee.ncku.edu.tw/orchidbase6/).

**Methods**

The obtained sequences were aligned and visualized by Multalin on http://multalin.toulouse.inra.fr/multalin/ (Corpet 1988). Motif location analysis use MEME Suite (https://meme-suite.org/meme/tools/meme (Bailey et al. 2015) and motif function analysis use InterPro (https://www.ebi.ac.uk/interpro/about/interproscan/) (Paysan-Lafosse et al. 2023). Then, the phylogenetic tree of mRNA was constructed using the Maximum Likelihood approach with bootstrapping 1000 repeats, Kimura 2-parameter model with 2 Gamma distribution model and amino acid sequence using JTT (Jones Taylor Thornton) with 2 gamma distribution model, by MEGA 11 (Tamura et al. 2021). To carry out protein 3-dimensional structural prediction, ColabFold (Goddard et al. 2018; Mirdita et al. 2022) and SWISS-MODEL web server (https://swissmodel.expasy.org/) were used (Arnold et al. 2006). The prediction structure was validated with Ramachandran Plot on PROCHECK (Laskowski et al. 2013), Overall Quality Factor on ERRAT (Colovos & Yeates 1993), and protein model quality estimation on QMEAN (Benkert et al. 2009). To evaluate the modelled structure, benchmarking to AlphaFold database (https://www.ebi.ac.uk/Tools/sss/fasta/) was conducted.

**RESULTS AND DISCUSSION**

**The Results of POH1 gene Alignment**

According to alignment analysis of POH1 gene shows polymorphism at the nucleotide level. Through the analysis, it was seen that there are mu-
tations and deletions in this alignment. Figure 1 shows coding sequences that would be translated into four domain regions, such as KNOX1, KNOX2, ELK domain and Homeobox KN Domain. The domain is categorized as Three Amino-Acid Loop Extension (TALE) family homeoproteins that are involved in plant growth and development. The gene structure of this gene is highly identical between species (Razzaq et al. 2020). In this case, POH1 gene with other homologous genes is highly identical, it comprises around 900-1400 bp for each species.

The Analysis of Amino Acid Sequence
The POH1 Protein sequence analysis shows a conserved amino acid sequence. Protein motif analysis and its location describe conserved regions on plants. It means that homeobox genes can be found in different species (Holland 2013). Its roles as a transcription factor that regulates gene function to enhance and repress certain genes (Gao et al. 2014; Yuan et al. 2018). Highly conserved amino acids are coloured in red and blue colours, meanwhile non-conserved amino acids are black colours (Figure 2). Motif locations at amino acid sequence indicated that there are 4 important domains to regulate plant development (Figure 3, Table 1). Each domain has specific roles in plant development. KNOX1 function correlates to tissue proliferation and maintains potential meristematic of flowering plant and moss sporophytes, while the function of KNOX2 is to regulate the alternation of generations by suppressing the haploid body plan in the diploid phase, which causes morphological transition to the land plants. Besides, KNOX1 modulation activity is involved in contributing to leaf shape diversity of flowering plants (Sakakibara et al. 2013; Furumizu et al. 2015). Moreover, KNOX2 also acts as negative regulator of secondary cell wall biosynthesis (Wang et al. 2020). ELK Domain acts as a signal for nuclear localization and is included in protein-protein interaction (Nagasaki et al. 2001; Ito et al. 2002; Nookaraju et al. 2022). Homeobox KN Domain acts as transcription factor that is conserved in different plants, which also affects in leaf shape elaboration (Wang & Jiao 2020; Zhang et al. 2022).

Phylogenetic Analysis of POH1 Gene with Other Homologous Genes
The POH1 phylogenetic analysis was performed by mRNA and amino acid sequences. In mRNA sequence phylogenetic construction, there are 2 clades and 1 outgroup. The gene sequence has highly similarities with DOH1 gene from Dendrobium orchid. Moreover, phylogenetic construction by using amino acid sequence reveals that POH1 gene still highly has similarities with DOH1 from Dendrobium orchid (Figure 4a; b, respectively).

Protein Modelling of POH1 Protein and Its Domain Function
The three-dimensional structural prediction was performed by 2 web servers, ColabFold (Figure 5a) and SWISS-MODEL (Figure 5b). Based on the Ramachandran plot, ColabFold structural prediction has residues in most favoured regions accounting for 68.6%, meanwhile, SWISS-MODEL has 92.3%. However, SWISS-MODEL modelling uses 59 residues, compared to ColabFold which uses 325 residues, all the translated amino acids. These results are correlated to algorithms that use in the WEB-SERVERS. ColabFold modelling focus on physical interactions or the evolutionary history (Jumper et al. 2021), meanwhile SWISS-MODEL build a model from templates in their database (Arnold et al. 2006). Moreover, other ways to validate protein modelling are overall
Figure 1. The result of POH1 gene alignment with other homologous genes. Different colours of the box described different domains; □ = KNOX1; □ = KNOX2; □ = ELK Domain; and □ = Homeobox KN Domain.
Figure 2. The result of POH1 protein alignment with other homologous genes. Different colours of the box described different domains;  = KNOX1;  = KNOX2;  = ELK Domain; and  = Homeobox KN Domain.

Figure 3. Domain location in amino acid sequence.
Table 1. The Function of Each Domain.

<table>
<thead>
<tr>
<th>No</th>
<th>Domain</th>
<th>Domain Name</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Homeobox KN domain</td>
<td>Regulation DNA-templated transcription</td>
<td>(Hirayama et al. 2007; Mukherjee et al. 2010)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>ELK domain</td>
<td>Nuclear localization signal; protein-protein interaction domain</td>
<td>(Nagasaki et al. 2001; Ito et al. 2002; Nookaraju et al. 2022)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>KNOX2</td>
<td>DNA-binding</td>
<td>(Nagasaki et al. 2001)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>KNOX1</td>
<td>DNA-binding</td>
<td>(Nagasaki et al. 2001)</td>
</tr>
</tbody>
</table>

Figure 4. Phylogenetic analysis of POH1 gene sequence with other homologous genes. (Construction by using: a. mRNA and b. amino acid sequences).
quality factor and QMEAN score. Overall quality factor of these structures is 89.252 (ColabFold modelling) and 98.039 (SWISS-MODEL modelling) (Figure 6). According to QMEAN scores, SWISS-MODEL modelling has higher score than ColabFold, 0.71 ± 0.11 to 0.41 ± 0.05 (Figure 6 a; b, respectively). Moreover, benchmarking effort to AlphaFold database shows that the modelled protein has 50 proteins that are homolog. In the database, the homologous proteins comprise around 300–400 amino acids. The highest relationship of this modelled protein in structural AlphaFold database found on Homeobox Protein Knotted-1-like 2 in *Dendrobium catenatum* (AlphaFoldDB: A0A2I0XDV6). Based on ColabFold modelling, we investigate the location of domain, which is KNOX1, KNOX2, ELK Domain and Homeobox KN Domain in 3D structural prediction (Figure 7). KNOX1 and KNOX2 domains are formed into MEINOX domain which is separated by a poorly conserved linker sequence (Nookaraju et al. 2022).

MEINOX Domain that is located in N-terminus will interact with POX domain from BEL protein (Zhang et al. 2021). The interaction between POX and MEINOX is hydrophobic (Ezura et al. 2022). The interaction of them affects phytohormone regulation that is related to plant development (Niu & Fu 2022). The heterodimer modulates phytohormone by activating cytokinin and repressing gibberellin (Tadege 2013; Testone et al. 2015; Marsch-Martínez & de Folter 2016), which correlates to shoot development (Wybouw & De Rybel 2019; Arro et al. 2019). Not only for shoot development, but also KNOX protein linked up to the lignification process in dicot and monocot plants (Townsley et al. 2013; Xu et al. 2019), the diversity of leaf shape (Gao et al. 2015; Wang et al. 2022), and a necessary role in starchy storage organs (Dong et al. 2019; Ruscher et al. 2021). Moreover, it involves to environmental response because of high temperature and humidity stress (Shu et al. 2015). To conclude this paragraph, the research on different functions of *POH1* gene, KNOX1 homologous gene in *Phalaenopsis* orchids, has to be broadened to reveal its functions.
CONCLUSIONS

POH1 gene is a gene that contributes to shoot formation in *Phalaenopsis* orchids and has four domains as transcription factors, such as KNOX1, KNOX2, ELK Domain, and Homeobox KN Domain. Three-dimensional structure modelling is performed by SWISS-MODEL and ColabFold. In SWISS-MODEL modelling use 59 residues with the percentage of most favoured regions of 92.3%, overall quality factor of 98.039, and QMEAN score 0.71 ± 0.11, while ColabFold uses 325 residues with a percentage of most favoured regions accounting for 68.6%, overall quality factor 89.252, and QMEAN scores 0.41 ± 0.05.
AUTHOR CONTRIBUTION
ES designed and controlled the research. NR conducted the data retrieval of mRNA and amino acid sequence and data analysis and wrote the manuscript. FYK, SUP, AL, and ISP conducted data and manuscript proof-reading.

ACKNOWLEDGMENTS
This research was supported by Universitas Gadjah Mada RTA (Rekognisi Tugas Akhir) 2022, Contract No: 3550/UN.1.P.III.Dit-Lit/PT.01.05/2022 that was given to ES as principal investigator.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest in this research.

REFERENCES
Furumizu, C. et al., 2015. Antagonistic roles for KNOX1 and KNOX2 genes in patterning the land plant body plan following an ancient gene duplication. PLoS Genetics, 11(2), e1004980. doi: 10.1371/journal.pgen.1004980.


Zhang, Y. et al., 2021. Hormonal regulatory patterns of laknoxs and label1 transcription factors reveal their potential role in stem bulblet formation in LA hybrid lily. *International Journal of Molecular Sciences*, 22(24), 13502. doi: 10.3390/ijms222413502