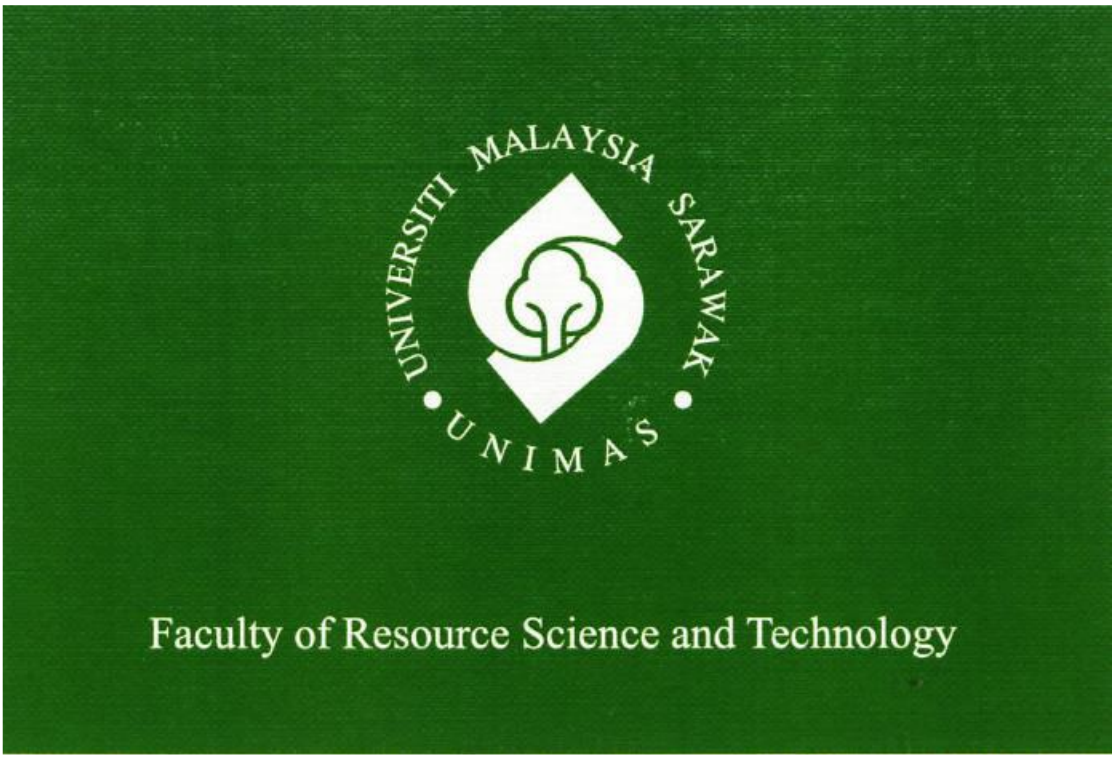

In Silico Characterization and Primer Design of *Thaumatin-like Protein* (TLP) Gene in Kelampayan

Muhammad Dzikri A'zim bin Ramli

Bachelor of Science with Honours

(Resource Biotechnology)

2022

DECLARATION

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)

Final Year Project Report	√
Masters	
PhD	

DECLARATION OF ORIGINAL WORK

This declaration is made on the 15th day of July 2022.

Student's Declaration:

I Muhammad Dzikri A'zim bin Ramli (72405) (Faculty of Resource Science and Technology)
(PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, *In silico* Characterization and Primer Design of Thaumatin-like Protein (TLP) Gene in Kelampayan is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

15th July 2022
Date submitted

Muhammad Dzikri A'zim bin Ramli (72405)
Name of the student (Matric no.)

Supervisor's Declaration:

I, Associate Professor Dr. Ho Wei Seng (SUPERVISOR'S NAME), hereby certify that the work entitled, *In silico* Characterization and Primer Design of Thaumatin-like Protein (TLP) Gene in Kelampayan (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of Bachelor of Science with Honours (Resource Biotechnology) (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: Assoc. Prof. Dr. Ho Wei Seng (Name of the supervisor) Date: 15th July 2022

I declare this Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
☐ **RESTRICTED** (Contains restricted information as specified by the organization where research was done)*
☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitise the content to for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student himself / herself neither third party on this Project/Thesis once it becomes sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student's signature  Supervisor's signature: 

(15th July 2022) (15th July 2022)

Current Address:
No 11, Lorong Aman 8, Taman Aman, Kg. Bahagia, 36000 Teluk Intan, Perak

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument was duly prepared by The Centre for Academic Information Services]

ACKNOWLEDGEMENTS

First and foremost, I want to thank God the Almighty, for his showers of blessings and never forsaken me in times of need throughout the completion of my project successfully. Without His grace, this project would not have been possible.

I would like to also express my deep and sincere gratitude to my final year project research supervisor, Associate Professor Dr. Ho Wei Seng for giving me the opportunity to do this research and spending a lot of time in providing invaluable guidance, support, ideas, and patience throughout the whole process. I would also like to extend my gratitude to my examiner, Professor Dr. Edmund Ui Sim Hang for his insightful comments and suggestions, which helped me to improve my thesis report writing. I am very thankful for the expertise and continuous encouragement that had been extended and given to me. A special thanks to the committee members of Forest Genomics and Informatics Laboratory (fGiLab), especially my mentor, Hong Zixin for their help and support directly or indirectly.

Moreover, I would like to take this opportunity to express my gratitude and appreciation to my friends, Mohd Jais bin Sarifuddin, Nor Shazlina binti Mohamed Mizan and Yap Zi Jian who have directly or indirectly shared ideas, doing the discussion together and lent their hand in completing this project. I thank all of you for your support, friendship, and kindness for giving me attention and helping me out in times of need.

Finally, I am greatly thankful to my family members, especially my parents for their endless care, moral support, encouragement, and advice throughout the process of completing this project.

In Silico Characterization and Primer Design of Thaumatin-Like Protein (TLP) Gene in Kelampayan

Muhammad Dzikri A'zim bin Ramli

Resource Biotechnology Programme
Facuty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Neolamarckia cadamba which is generally known as Kelampayan is a large and sturdy tree which has a rapid growth rate. Kelampayan provides many benefits and uses, ranging from the timber industry, carpentry, and medicinal properties. The problem statement for this study is the lack of genetic information regarding the function of thaumatin-like protein (TLP) in Kelampayan. There are two objectives of this study, firstly is the characterization of TLP gene in Kelampayan by using *in silico* method for further understanding of the gene. The second objective is to design a primer of TLP gene for many uses such as PCR. The methodology of this study was done in two sections with the first section being the *in silico* characterization of TP gene. The characterization consists of the phylogenetic tree analysis using MEGA 11 Software, the Domain search using CD Search tool, and Motifs search using MEME Online Tool. The second section, primer design was designed using Primer Blast tool. From 50 TLP nucleotide sequences, the phylogenetic tree shows three distinct groups of TLP, differentiated by their function in their respective species. The primer design also provides the most suitable primer for every nucleotide sequence of TLP in Kelampayan. From this study, the genetic information of TLP was understood greatly from the characterization process and the primer was designed successfully.

Keyword: Kelampayan, *In silico* characterization, TLP gene, primer, analysis.

Pencirian In Silico dan Reka Bentuk Primer Gen Thaumatin-Like Protein (TLP) dalam Kelampayan

ABSTRAK

Neolamarckia cadamba yang umumnya dikenali sebagai Kelampayan merupakan pokok yang besar dan tegap serta mempunyai kadar pertumbuhan yang pesat. Kelampayan memberi banyak faedah dan kegunaan, mulai dari industri perkayuan, pertukangan, dan berkhasiat perubatan. Pernyataan masalah kajian ini ialah kekurangan maklumat genetik berhubung fungsi protein seperti thaumatin (TLP) di Kelampayan. Terdapat dua objektif kajian ini, pertama ialah pencirian gen TLP di Kelampayan dengan menggunakan kaedah in silico untuk pemahaman lanjut tentang gen tersebut. Objektif kedua adalah untuk mereka bentuk primer gen TLP untuk banyak kegunaan seperti PCR. Metodologi kajian ini dilakukan dalam dua bahagian dengan bahagian pertama adalah pencirian dalam siliko gen TLP. Pencirian terdiri daripada analisis pokok filogenetik menggunakan Perisian MEGA 11, carian Domain menggunakan alat Carian CD, dan carian Motif menggunakan Alat Dalam Talian MEME. Bahagian kedua, reka bentuk primer telah direka menggunakan alat Primer Blast. Daripada 50 jujukan nukleotida TLP, pokok filogenetik menunjukkan tiga kumpulan TLP yang berbeza, dibezakan oleh fungsi mereka dalam spesies masing-masing. Reka bentuk primer juga menyediakan primer yang paling sesuai untuk setiap jujukan nukleotida TLP di Kelampayan. Daripada kajian ini, maklumat genetik TLP sangat difahami daripada proses pencirian dan primer telah direka dengan jayanya.

Kata kunci: Kelampayan, pencirian in silico, gen TLP, primer, analisis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABA	Abscisic Acid
BLASTn	Basic Local Alignment Search Tool (nucleotide)
CDD	Conserved Domain Database
cDNA	Complementary DNA
Cys	Cysteine
DNA	Deoxyribonucleic acid
GC	Guanine-Cytosine
GTR	General Time Reversible
HR	Hypersensitive Response
HKY	Hasegawa-Kishino-Yano
JC	Jukes-Cantor./div
K2	Kimura 2-parameter
MEA	Motif Enrichment Analysis
MEGA	Molecular Evolutionary Genetic Analysis
MEME	Multiple Em for Motif Elicitation
NCBI	National Center for Biotechnology Information
NJ	Neighbor-joining
PCR	Polymerase Chain Reaction
pI	Isoelectric point
PR	Pathogenesis-related
RNA	Ribonucleic acid
RR	Regulatory Region
SMART	Simple Modular Architecture Research Tool

TFBSs	Transcription Factor Binding Sites
TLP	Thaumatin-like protein
TN93	Tamura-Nei
T92	Tamura 3-parameter
UNAM	Mexico National Autonomous University

CHAPTER 1

INTRODUCTION

Neolamarckia cadamba (Roxb.) Bosser, which is generally known as Kelampayan is a plant from the *Rubiaceae* family, one of the fastest growing plants for the project of forest development in Sarawak. Kelampayan is proved to have various benefits including its medicinal effects, production of papers and one of the best wood materials for carpentry industry because of its characteristics (Ho et al., 2015).

The problem statement for this research is the limited amount of genetic information of TLP gene in Kelampayan which makes the mechanism of the TLP gene in Kelampayan not understood properly. Tobacco which is one of the most studied plants regarding the TLPs was identified to have antifungal activity from its leaves and cells which inhibits fungi such as *Cercospora berticola, Candida albicans, Neurospora crassa, Trichoderma reesei* and *Phythopthora infestans.* Aside from that, the TLPs in Tobacco also shown to regulate the TLP expression in response to microbial infection, wounding, osmotic stress, abscisic acid (ABA) and ethylene and lastly with salicylate, methyl jasmonate and elicitors which were found after extensive research for years (Velazahan et al., 1999).

The objectives of the research are to analyze the sequence of TLP genes by using *in silico* characterization technique. The second objective is to design a primer of TLP gene which are functional and can be used for Polymerase Chain Reaction (PCR). The hypothesis of this research is the TLP gene in Kelampayan is expected to have a slightly different mechanism and regulation from the TLP genes in other species.

CHAPTER 2

LITERATURE REVIEW

2.1 *In silico* characterization

The term "*In silico*" was derived from a component in computer, namely silicium which gives the meaning of *in silico* as a method or prediction by using computational approaches (Amberg, 2013). The are several advantages of using *in silico* method which is firstly, this method can make a quick prediction for a large set of compound in a high-throughput mode. Moreover, *in silico* method is capable of making early prediction based on the structure of a compound even before the compound has been synthesized. Therefore, *in silico* method is very suitable to be used on the early stage of various research process, including characterization of a gene and primer design.

The history of the term "*in silico*" started back during the year 1987. An American computer scientist, Christopher Langton used the word to characterize artificial life in the announcement for a symposium on the topic at the Los Alamos National Laboratory's Center for Nonlinear Studies. Two years later, a Mexican mathematician from Mexico's National Autonomous University, Pedro Miramontes used "*in silico*" term to describe biological experiments conducted solely by using a computer during the workshop "Cellular Automata: Theory and Applications" in Los Alamos, New Mexico (UNAM). He uses the term while he was doing a presentation of his report titled "DNA and RNA Physicochemical Constraints, Cellular Automata and Molecular Evolution" which this work later becomes his dissertation.

The term "*in silico*" becomes frequently used after these two events. The European Community Commission has used "*in silico*" in reports to promote the development

of bacterial genome projects. In 1991, a French team published the first publication in which "*in silico*" is mentioned (Danchin et al., 1991). Hans B. Sieburg wrote the first book chapter mentioning "*in silico"* in 1990 and presented it at the Santa Fe Institute's Summer School on Complex Systems.

2.2 Bioinformatics Tools

A biological database is regarded as a sizable and well-organized body of information that is typically linked to computer software. It was created to update, query, and retrieve data elements kept in a particular system. Numerous database and software tools have been made available to provide information, particularly for bioinformatics literature, medical, and research biology. Consequently, it can be challenging to stay current with bioinformatics tools, yet doing so is essential to modern data analysis, particularly in the fields of biology and medicine (Duck *et al.*, 2016). More accurate and reliable information is provided by current databases and software. These days, databases like SwissProt Protein Sequence, EMBL Nucleotide Sequence Data, and NCBI GenBank are often used in bioinformatics research. Even though some of the published databases, like the Database of Databases (DoD) and the BioMed Central databases, were no longer accessible (Duck *et al.*, 2016). One of the most comprehensive databases to date is GenBank, which is publicly accessible with more than 300,000 organisms.

Generally, submissions on GenBank can be made through the NCBI submission portal, BankIt, and the *tbl2asn* tool that are obtained from either an individual laboratory or batch submissions from extensive sequencing projects, including environmental sampling projects and the whole genome shotgun (WGS). In the International Sequence Database Collaboration (INSDC), GenBank, DNA

Data Bank of Japan (DDBJ) and EMBL-EBI European Nucleotide Archive (ENA) were partners so, the regular data exchange within them provides worldwide coverage of a comprehensive and uniform sequence collection. Consequently, the NCBI enables free and open access to GenBank data so that anybody can research information relating to genomes, taxonomy, the biomedical literature found in PubMed, and protein sequences and structures (Sayers *et al*., 2019).

On the internet, there are many web-based tool that can be utilized for finding structural and functional domains in protein sequences, one of them is the Conserved Domain Search service (CD-Search). Unlike other web-based tools, CD-Search leverages BLAST heuristics in order to offer the best performance for its users – a quick, interactive service and able to search a large database of domain models. Pairwise alignments between the query and domain-model consensus sequences and domain architecture cartoons are provided with search results (Marchler-Bauer & Bryant, 2004). There are many sources of CD-Search's alignment, mainly imported collections such as Pfam and SMART, as well as automatic alignments provided using Clusters of Orthologous Groups (COGs) classifications. CD-Search give many different results, ranging from the summaries, individual pairwise alignments between the user query and search model consensus, and also a tabular list of hits (Marchler-Bauer & Bryant, 2004).

Molecular Evolutionary Genetics Analysis (MEGA11) is a software that contains large collection of methods and tools for computational molecular evolution. It is a comprehensive bioinformatics tool that is capable of building timetrees of species, pathogens, and gene families by utilizing the rapid relaxed-clock methods (Tamura et al., 2021). The developer also implemented the methods for estimating divergence times and confidence intervals to use the probability densities

for calibration constrains which is used for various process, including node-dating and sequence sampling dates for tip-dating analyses. In addition, MEGA11 also have the Bayesian method which is usually used to estimate the neutral evolutionary probabilities of alleles in a species using multi species sequence alignments and also a machine learning method to test for the autocorrelation of evolutionary rates in phylogenies (Tamura et al., 2021).

With each successive release, MEGA has developed to take use of new technical advancements and the processing capabilities of personal computers. The MS-DOS character-based MEGA interface was replaced with a powerful graphical user interface (GUI) for the Microsoft Windows operating system. Then it was revamped to become activity-driven, and web technologies were included to guarantee a uniform look and feel across Microsoft Windows, Linux, and macOS. MEGA GUI now runs natively on Windows, Linux, and macOS and is entirely cross-platform (Tamura et al., 2021).

2.3 Primer design

Primer is a short sequence of DNA nucleotides which is usually 18 to 24 base pairs long and it can be employed in a variety of different experimental procedures. Scientists uses primers in PCR which functions to target a locus for the amplification of gene sequence that can be used for additional examination of said sequences. Primer can also be used to sequence a sequencing procedure to target a very precise location and then analyze the DNA molecule's extension. A primer needs to be designed before conducting PCR for a specific gene (Wright, 2021).

The main purpose of primer design is to determine a set of primers that can amplify the sequence of target while avoiding other sequences (non-target) for an optimal PCR. If the main purpose is not possible, the aim becomes predicting

probable cross-amplification with non-target sequences with a high accuracy. This knowledge allows for an informed evaluation of the many primer choices that may be employed to reduce non-target interference. The chosen primers might then be produced and tested using PCR or quantitative PCR (qPCR) in the lab (Wright, 2021).

2.4 Kelampayan (*Neolamarckia cadamba* (Roxb.) Bosser)

Neolamarckia cadamba (Roxb.) Bosser, which is commonly known as Burflower Tree, Kaddam, Leichhardt Pine and locally referred to as Kelampayan is a plant from the *Rubiaceae* family. Kelampayan is one of the fastest growing plants in Sarawak due to the time taken for it to fully matured is within the range of only five to ten years of age. Under normal environment, the height of Kelampayan tree can reach up to 17 meters and the diameter of the tree can grow until about 25 centimeters at breast height (dbh). For a mature Kelampayan tree, the height of the tree can exceed 20 to 30 meters with the highest ever recorded height of Kelampayan was about 45 meters, while the diameter of the trunk can be as large as one meter with a maximum width of 160 cm ever recorded. Its leaves are about 13 to 32 cm long. Kelampayan is an angiosperm plant, which means that the plant is capable to produce flower and the flowering process usually begins when the tree is about 4 to 5 years old

The distinct characteristic of this tree is that it is a type of lightweight tree with medium texture, less shiny, odourless, clean and has a fine surface. Kelampayan provides many benefits, mainly used as raw material of plywood and furniture industry. Because of the fine and c lean characteristics, Kelampayan is easy to process, either by using machine tools or by only using hands. It also cuts evenly, has an excellent surface and subtle to nail (Ho et al., 2015).

Figure 2.1: *Neolamarckia cadamba* (Kelampayan) and its fruit (Dalin, 2016)

Aside from the benefits provided by the trunks of the tree, Kelampayan's leaves and bark offers medicinal properties which sometimes used by the locals for traditional remedies. The dried bark may be used as a tonic and to reduce fever, while the leaf extract can be used as a mouthwash. The species is preferred in tree plantation programs because of its multifunctional function and utility (Ho et al., 2015).

2.5 Thaumatin-like protein (TLP)

Throughout their lives, plants are frequently subjected to a variety of biotic and abiotic stress situations. Plants have evolved an incredibly intricate and sophisticated defensive systems to combat these environmental constraints. When attacked by phytopathogenic fungi, the defense responses include reactive oxygen species (ROS) induced localized cell death, also known as the hypersensitive response (HR), accumulation of antimicrobial phytochemicals (phytoalexins), and expression of a group of proteins known as pathogenesis-related (PR) proteins at the infection site (Misra et al., 2016).

Based on their amino acid content, structure, and metabolic activity, PRs are divided into 17 families from PR1 to PR17. Among all the PR families, PR5 members have a significant sequence resemblance to thaumatin, a protein that have an intense sweet taste which can be found in the Miracle Berry *(Thaumatococcus*

danielli), a plant from West Africa. Because of the similarity, the PR5 members are referred to as thaumatin-like proteins (TLPs). TLPs are a low-molecular-weight (20–26 kDa) proteins having sixteen cysteine (Cys) residues that are conserved. Cys residues mediate eight intramolecular disulfide bonds, which are thought to stabilize the protein under high pH and temperature conditions (Misra et al., 2016).

TLPs are not generally seen in young and turgid plants' leaves, but they rapidly accumulate to a large number in response to biotic or abiotic stress. Older potato leaves, on the other hand, were proven to collect TLPs in the absence of purposeful stress during the experiment that was conducted. However, it was not ruled out that the potato leaves may have been imposed to stress when growing in a greenhouse or growth chamber (Velazahan et al., 1999). Furthermore, several TLPs have been proven to have antifungal action and have been highly expressed in plants to see if they might give fungal pathogen resistance. However, the chemical mechanism behind the antifungal action is still unknown, as is the biological function of TLPs in plants (Misra et al., 2016).

TLPs are proteins that has a high solubility which accumulate to high quantities (one to twelve percent) in certain tissues or subcellular compartments. It can also release into the medium or extracellular space under certain circumstances. They are soluble even in acidic environments and have a high resistance to proteolysis. In previous investigations, these qualities made it simple to see tobacco TLPs in acidic extracts of infected leaves after electrophoresis in polyacrylamide gels and staining with Coomassie Brilliant Blue (Velazahan *et* al., 1999).

CHAPTER 3

MATERIALS AND METHODS

3.1 EST Data Mining

Four thaumatin-like protein (TLP) gene (*cn25, cn26, Ncdx043C10*, and *Ncdx036H08*) of Kelampayan were predicted using a contig mapping approach based on the ESTs obtained from the Kelampayan transcriptome database (NcdbEST) (Ho *et al.,* 2014; Pang *et al.*, 2015).

3.2 Identification of TLP in Kelampayan

The TLP nucleotide sequences of Kelampayan (*cn25, cn26, Ncdx043C10,* and *Ncdx036H08*) were used as query sequences and BLASTn was carried out in the NCBI Genbank database. The retrieved TLP sequences were then analysed in a phylogenetic tree to determine the correlation of the sequences.

3.3 Phylogenetic Analysis of TLP Evolution

A total of 50 TLP genes, which includes the four TLP genes (*cn25, cn26, Ncdx043C10,* and *Ncdx036H08*) obtained from NcdbEST. Another 46 TLP genes were retrieved from NCBI (National Center for Biotechnology Information) which consists of 7 *Malus x domestica clone,* 1 *Malus domestica,* 4 *Pyrus pyrifolia clone,* 2 *Pyrus pyrifolia,* 1 *Pyrus pyrifolia cultivar Huobali,* 1 *Pyrus pyrifolia mRNA,* 2 *Prunus persica,* 2 *Prunus persica clone,* 1 *Prunus avium,* 1 *Populus szechuanica,* 4 *Picea likiangensis.* 7 *Cryptomeria japonica,* 1 *Pseudotsuga menziesii,* 5 *Actinidia deliciosa isolate,* 3 *Vitis vinifera,* 1 *Vitis vinifera cultivar Riesling,* 1 *Vitis vinifera cultivar Regent,* 1 *Litchi chinensis,* and 1 *Sambucus nigra clone.* The full length of all sequences was integrated in MEGA 11 (Kumar *et al.,* 2018) software

for the construction of the phylogenetic tree by using the neighbour-joining (NJ) method and the parameters set are using *p-distance* and *pairwise deletion* at bootstrap value of 1000.

3.4 Conserved Domain and Motifs Identification

The Conserved Domain Database in NCBI was used to analyse the presence of thaumatin-like protein domain (pfam00134) among the TLP in Kelampayan. The conserved motifs in 50 TLP sequences were identified by using the MEME suite server online tool (https://meme-suite.org/meme/tools/meme), with parameters set to a maximum of five number of motifs, zero or one occurrence per sequence, and an optimum width character from 10 to 70. Moreover, the logo diagram of every motif was annotated in MEME.

3.5 Primer Design of Thaumatin-like Protein (TLP)

The primer design of TLP gene were done to the four genes retrieved from (NcdbEST) in the Primer Blast tool in NCBI Database (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The four TLP gene sequences (*cn25, cn26, Ncdx043C10*, and *Ncdx036H08*), which is in FASTA format were run in the primer designing tool to generate the primer. The reason only the four nucleotide sequences retrieved from the NcdbEST is used in primer design analysis is because these are the only TLP genes sequences provided for this project. The parameter used for the designation of primer were the melting temperature, which was set between 55^{o}C to 65^{o}C, the length of primer between 18 until 24 bp, GC content of 45%-55% and GC clamp of 3'.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 TLP Gene Family Identification in Kelampayan

To obtain homology sequences of TLP from the NCBI Genbank database, four TLP Gene sequence in Kelampayan (cn25, cn26, Ncdx043C10, Ncdx036H08) (Appendix A) were respectively subjected to search homology via BLASTn in NCBI Genbank (www.ncbi.nlm.nih.gov). The result obtained from BLASTn analysis (Table 4.1, 4.2, 4.3 and 4.4) of each TLP gene in Kelampayan shows high degree of similarity to several other plants, especially in *Coffea arabica*, and *Coffea euginioides*. Each output of BLASTn homology search was standardized in five sequences except for the fourth gene, Ncdx036H08 which has only three genes with a high degree of similarity from the BLASTn search. Fassler and Cooper (2011) define the E-value or expectation value as the expected number of different alignments that occur by chance in the database. A lower e-value indicates that the gene has a better alignment quality, and the score is considered significant.

Table 4.1: BLASTn sequence homology analysis of *cn25* sequence (926 bp) with NCBI nucleotide database

Organism	Gene	GenBank accession no.	Similarity (%)	E-value
Coffea arabica	LOC113742355	XM_027270186.1	89.42	0.0
Coffea eugenioides	LOC113767439	XM_027311541.1	89.18	0.0
Olea europaea var. sylvestris	LOC111372323	XM_022994608.1	84.04	0.0
Olea europaea var. sylvestris	LOC111365272	XM_022985714.1	83.16	2e-172
Citrullus lanatus	tlp28	MF445021.1	83.04	8e-171

Table 4.2: BLASTn sequence homology analysis of *cn26* sequence (564 bp) with NCBI nucleotide database

Organism	Gene	GenBank accession no.	Similarity (%)	E-value
Coffea eugenioides	LOC113767439	XM_027311451.1	90.93	9e-129
Coffea arabica	LOC113742355	XM_027270186.1	90.93	9e-129
Curcubita pepo subsp. pepo	LOC111801507	XM_023685521.1	85.20	1e-87
Curcubita maxima	LOC111490349	XM_023138931.1	84.59	3e-84
Curcubita moschata	LOC111457163	XM_023099327.1	84.29	1e-82

Table 4.3: BLASTn sequence homology analysis of *Ncdx043C10* sequence (706 bp) with NCBI nucleotide database

Organism	Gene	GenBank accession no.	Similarity (%)	E-value
Coffea arabica	LOC113725444	XM_027248596.1	87.70	5e-157
Coffea arabica	LOC113725444	XM_027248595.1	87.70	5e-157
Coffea arabica	LOC113730461	XM_027255155.1	86.86	9e-150
Coffea eugenioides	LOC113763463	XM_027307288.1	86.65	4e-148
Coffea arabica	LOC113730461	XM_027255154.1	86.65	4e-148

Table 4.4: BLASTn sequence homology analysis of *Ncdx036H08* sequence (731 bp) with NCBI nucleotide sequence

Organism	Gene	GenBank accession no.	Similarity (%)	E-value
Coffea arabica	LOC113712377	XM_027235775.1	80.39	3e-110
Solanum tuberosum	LOC102601558	XM_006342001.2	77.52	5e-38
Cynara cardunculus var. scolymus	LOC112525278	XM_025135337.1	74.02	1e-18

4.2 Phylogenetic Analysis of Kelampayan TLP Genes

A phylogenetic tree of TLPs in Kelampayan was generated with the retrieved homology nucleotide sequences from BLASTn (Appendix B) based on the alignment using the neighbour-joining method in MEGA 11 Software and using 1000 bootstrap replicates (Figure 4.1). In Figure 4.1, the number next to the nodes indicates the node's level of support. Form this information, a high-value number indicates that the sequences to the right of the node cluster together to the exclusion of all other sequences. Moreover, a fewer bootstrap

value indicates that a distance is so small that the probability of the substitution of a nucleotide is very little while a higher bootstrap value shows a higher confidence level of the clade in the phylogenetic tree.

From the data, we can see that the TLP genes in Kelampayan are more closely related to the TLP genes of *Coffea* species (Ncdx036H08) and *Curcubita* (cn25) species compared to *Olea* and *Cynara* species, respectively. Two of the four TLP sequences in Kelampayan, *cn26* and *Ncdx043C10* can be seen to be closely related to each other. On the other hand, *cn25* are related but a bit further from both sequences (*cn26* and *Ncdx043C10*) and the last of the four TLP sequences, *Ncdx036H08* shows a distant relation compared to the other three TLP genes, yet they all shared a common ancestor long ago.