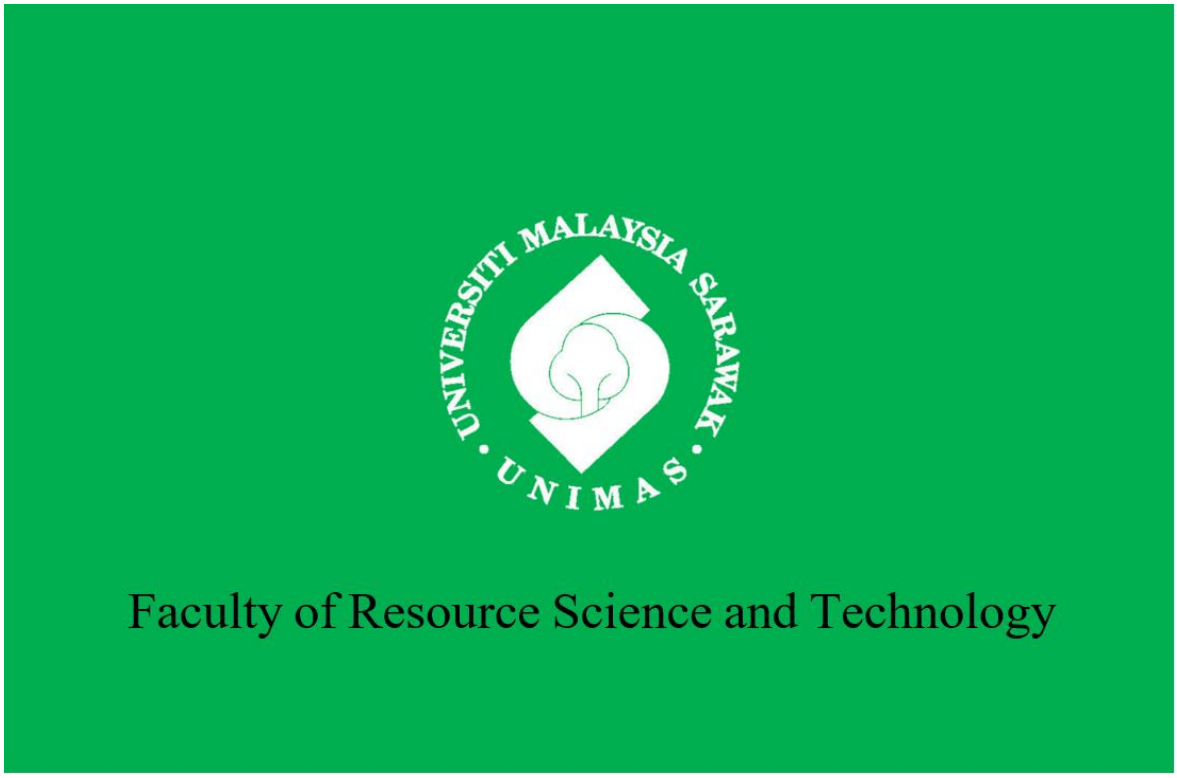


# *In Silico* Characterization and Primer Design of *Fasciclin-Like Arabinogalactan proteins* (FLAs) gene in Kelampayan

Nor Shazlina binti Mohamed Mizan (72477)

Bachelor of Science with Honours
(Resource Biotechnology)
2022

# *In Silico* Characterization and Primer Design of *Fasciclin-Like Arabinogalactan proteins* (FLAs) gene in Kelampayan

Nor Shazlina binti Mohamed Mizan

A thesis submitted in partial fulfilment of the Requirement of The Degree Bachelor of Science with Honours

(Resource Biotechnology)

**SUPERVISOR: ASSOC. PROF. DR. HO WEI SENG**

**Programme of Resource Biotechnology**

**Faculty of Resource Science and Technology**

UNIVERSITI MALAYSIA SARAWAK

**2022**

**UNIVERSITI MALAYSIA SARAWAK**

Grade: _____

**Please tick (√)**

| | |
|---|---|
| Final Year Project Report | √ |
| Masters | |
| PhD | |

**DECLARATION**

This declaration is made on the ……18…. day of……July…… 2022.

**Student's Declaration:**

I ---**Nor Shazlina binti Mohamed Mizan (72477) 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 Fasciclin-Like Arabinogalactan proteins* (FLAs) 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.

**18 July 2022**
Date submitted

**Nor Shazlina binti Mohamed Mizan (72477)**
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 *Fasciclin-Like Arabinogalactan proteins* (FLAs) gene in Kelampayan**---(TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * ~~partial~~/full fulfilment for the conferment of ---**Degree 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: **AP Dr. Ho Wei Seng** Date: **18 July 2022**
(Name of the supervisor)

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 organisation 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: 

**18 July 2022** (Date 

Current Address:
**159, Jalan Rambutan, Kampung Endah, 42700 Banting Selangor.**

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]

## ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful.

All thanks to Allah for blessing me and guiding me along my journey to complete my project successfully. Without His grace, this project would not have been possible. I would also want to convey my heartfelt appreciation to Associate Professor Dr. Ho Wei Seng, my final year project supervisor, for allowing me to do this study and offering valuable advice, support, and patience during the process. I would like to thank Professor Dr. Edmund Sim Ui Hang, my examiner, for his insightful remarks and recommendations, which helped me improve my thesis report writing. I am grateful for the knowledge and constant support that has been offered to me.

A particular thank you to my mentor for this project, Ms Hong Zixin, for her unwavering support and advice, both directly and indirectly. In addition, I want to use this opportunity to convey my thanks and appreciation to my friends, Mohd Jais bin Sarifuddin, Muhammad Dzikri A'zim bin Ramli, and Yap Zi Jian, who have exchanged ideas and assisted in the completion of this project. I appreciate everyone's support and politeness in paying attention to me and assisting me when I needed it. Not to forget my forever comfort BTS boys who always motivated me to finish my project every time I felt discouraged and demotivated.

Lastly, I owe the greatest gratitude to my parents and family for the absolute love and support that I have ever received. The road would be exceedingly tough without their unending prayers and encouragements. They made several sacrifices to allow me to focus on my education, research, and writing. For all these people and many others, I am profoundly grateful Alhamdulillah.

# *In Silico* Characterization and Primer Design of *Fasciclin-Like Arabinogalactan proteins* (FLAs) gene in Kelampayan

Nor Shazlina binti Mohamed Mizan

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

## ABSTRACT

The *fasciclin-like arabinogalactan proteins (FLAs)* have crucial functions in the growth and environmental adaption of the Kelampayan tree. The Kelampayan tree is a big, rapidly expanding tree that is high in demand due to its many advantages, including in medicine and the timber industry. The problem statement for this research is the lack of genetic information related to the function of the FLAs gene specificallythe in Kelampayan tree compared to other species. By using bioinformatics methods, the research focused on the *in silico* characterization and primer design of the *FLAs* gene in Kelampayan. To have a better understanding of the evolution of the *FLAs* gene family of the Kelampayan ESTs, the Conserved Domain Database and the MEME suite serve were used to characterise the motif and conserved domain analysis of 49 *FLA-related* genes from several plant species. Primer-BLAST was used to compare the criteria for each *FLAs* gene in Keltoorder to choose the optimal primer. The efficiency of the primer was confirmed using OligoAnalyzer and BLAST. The classification of these *FLAs* into two classes is mainly based on phylogenetic analysis of plant *FLAs*. This research discovered two distinct clades based on the similarity of genes across species. After this study, the *FLAs* gene characterization and primer design in Kelampayan will enable future research to determine how *FLAs* genes will contribute to Kelampayan tree growth. The results provide insight into the potential future applications of the data, including full-length *FLA* gene isolation, SNP assay, etc.

***Keywords**: Neolamarckia cadamba, in silico, FLAs, conserved domain, primer design*

# *Pencirian In Silico dan Reka Bentuk Primer Protein Arabinogalaktan Seperti Fasciclin (FLA) di dalam Kelampayan*

## ABSTRAK

*Protein arabinogalaktan seperti fasciclin (FLA) mempunyai fungsi penting dalam pertumbuhan dan penyesuaian persekitaran pokok Kelampayan. Pokok kelampayan merupakan pokok besar yang tumbuh dengan cepat dan mendapat permintaan tinggi kerana banyak kelebihannya termasuk dalam bidang perubatan dan industri perkayuan. Pernyataan masalah bagi kajian ini ialah kekurangan maklumat genetik berkaitan fungsi gen FLA khususnya di pokok Kelampayan berbanding spesies lain. Dengan menggunakan kaedah bioinformatik, penyelidikan tertumpu kepada pencirian in silico dan reka bentuk primer gen FLA di Kelampayan. Untuk memahami dengan lebih baik tentang evolusi keluarga gen FLA, 49 gen seperti FLA daripada beberapa spesies tumbuhan telah dikenalpasti. Menggunakan "Conserved Domain Database" dan pelayan suite MEME, 49 gen seperti FLA telah dimasukkna bagi analisis motif dan domain terpelihara. Primer-BLAST digunakan untuk membandingkan kriteria bagi setiap gen FLA di dalam Kelampayan untuk memilih primer yang optimum. Kecekapan primer telah disahkan menggunakan OligoAnalyzer dan BLAST. Pengelasan FLA ini kepada dua kelas kebanyakannya berdasarkan analisis filogenetik FLA tumbuhan. Dua kategori berbeza ditemui dalam penyelidikan berdasarkan persamaan gen dari pelbagai spesies. Selepas kajian ini, pencirian gen FLA dan reka bentuk primer di dalam Kelampayan akan membolehkan penyelidikan pada masa hadapan untuk menentukan dengan tepat bagaimana gen FLA menyumbang kepada pertumbuhan pokok Kelampayan. Hasil daripada kajian ini dapat memberikan gambaran tentang potensi aplikasi data ini, termasuk pengasingan gen penuh, ujian SNP, dsb.*

***Kata kunci**: Neolamarckia cadamba, pencirian in silico, FLA, domain terpelihara, Reka Bentuk Primer*

## TABLE OF CONTENT

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AGP | Arabinogalactan protein |
| *At* | *Arabidopsis thaliana* |
| BLAST | Basic Local Alignment Search Tool |
| *Cc* | *Cuscuta campestris* |
| CDS | Coding sequence |
| cm | Centimeters |
| DDBJ | DNA Data Bank of Japan |
| DNA | Deoxyribonucleic acid |
| DoD | Database of Databases |
| DPPH | 2,2-diphenyl-1-picrylhydrazyl |
| Dw | Duckweed |
| EMBL | European Molecular Biology Laboratory |
| ENA | European Nucleotide Archive |
| EXTs | Extensins |
| FASTA | FAST-All |
| *FLA* | *Fasciclin-like arabinogalactan* |
| FRIM | Forest Research Institute of Malaysia |
| GC | Guanine-Cytosine Content |
| *Gh* | *Gossypium hirsutum* |
| GPI | Glycosylphosphatidylinositol |
| HRGPs | Hydroxyproline-rich glycoproteins |
| ID | Identity |
| ISSDC | International Sequence Database Collaboration |

| | |
|---|---|
| MEME | Multiple Expectation Minimization for Motif Elicitation |
| NCBI | National Centre for Biotechnology Information |
| NJ | Neighbor-joining |
| ORF | Open Reading Frame |
| *Palxt* | *Populus alba x tremula* |
| PCR | Polymerase Chain Reaction |
| PRPs | Proline-rich proteins |
| qPCR | Quantitative Polymerase Chain Reaction |
| *Ta* | *Triticum aestivum* |
| $T_m$ | Melting temperature |
| UI | User Interface |
| XYLP | xylogen-like AGPs |
| *Wa* | *Wolffia australiana* |

# CHAPTER 1: INTRODUCTION

*Neolamarckia cadamba* also called the Kelampayan tree, is a massive and fast-growing tree. It takes around nine years to reach a height of approximately 17 m and a diameter of 25 cm at breast height (dbh) (Mojiol et al., 2018). Kelampayan was chosen for large-scale forest planting in Sarawak due to its multifunctional function and value such as in traditional medicine, plywood industry, pulp, and paper industry (Joker, 2000). Moreover, the genomes of plants can be sequenced quickly and cheaply. *In silico* approaches allow efficient gene characterisation and analysis development by utilising this knowledge.

The problem statement for this research is the lack of genetic information related to the function of the *FLAs* gene specifically in the Kelampayan tree compared to other species. Several studies have shown that *FLAs* genes are engaged in tissue-specific processes and generic responses to abiotic or biotic environmental stresses in a variety of plant growth and development processes (Jun & Xiaoming, 2012; Wang et al., 2015; Zhang et al., 2015; Showalter et al., 2016; Takahashi et al., 2016; Xue et al., 2017). Moreover, *FLAs* are plant-specific arabinogalactan proteins having one or two AGP domains and one or two fasciclin (FAS) domains. FAS domains were found in *Drosophila melanogaster*, microbes, plants, and mammals (Johnson et al., 2003). The hypothesis behind this study is that the *FLAs* gene in Kelampayan could regulate and perform in a slightly different mechanism than *FLAs* genes in other species. The study focused on the *in-silico* characterisation and primer design of the *fasciclin-like arabinogalactan proteins (FLAs)* gene in Kelampayan to better comprehend the development of the *FLAs* gene family.

The objective of this study is to characterise the *FLAs* genes in Kelampayan against the *FLAs* genes retrieved from NCBI GenBank using *in silico* tools and to design the primer that can be utilized for future use, such as full-length *FLA* gene isolation, SNP assay, and etc.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 *In silico* characterization

The *in silico* approach is used in biology and other experimental disciplines to accomplish calculations on a computer or via simulation software. The term *in silico* is derived from the pseudo-Latin phrase "*in silicon*," which refers to the silicon in computer chips. To honour Latin idioms like "*in vitro*," *"in vivo*," and "*in situ*," all of which are often used in biological research, this term was invented back in 1987. The study of biology, medicine, and bioinformatics has used a wide range of database and software resources. As a result, maintaining current bioinformatics resources is challenging, but it is a necessary component of recent big data, particularly in biology and medical fields (Duck et al., 2016).

It is common in bioinformatics studies to use various databases these days, including the NCBI GenBank and the SwissProt Protein Sequence Databases from EMBL (Feltes et al., 2020). More than 300,000 species are included in GenBank, which is one of the most comprehensive databases. International Sequence Database Collaboration (ISSDC) collaborators included the European Nucleotide Archive (ENA), the DNA Data Bank of Japan (DDBJ), and GenBank offer global coverage of a complete and consistent sequence collection due to the constant data sharing. The NCBI provides open and free access to GenBank data, such as taxonomy, genomes, biomedical research publications, and protein sequences and structures, to anybody who intends to study them (Benson et al., 2012).

Experiments conducted within and outside of living beings and their genesis in nature were the focus of these earlier expressions. A variety of industries has benefited from using *in silico* computer-based modelling tools, including bioprocess design and validation, the identification of a potential therapy for disorders like Covid-19, and whole-cell research among both eukaryotic and prokaryotic organisms. In addition, an analysis of Dermacentor andersoni p36 immunosuppressive protein by Oyugi and colleagues (2018), was performed by *in silico*, which

was a first step in understanding the processes of immune recognition, which was considered a challenge in vaccine development. Another study by Tchin et al. (2018), researched to identify and characterize the entire sequence of Kelampayan's C4H gene.

### 2.2 *In silico* Primer Design

In the polymerase chain reaction (PCR), the efficiency and specificity of the primers used to amplify a particular gene/DNA fragment are primarily dependent on their effectiveness. Nucleic acid analysis may benefit significantly from the selective amplification of small amounts of nucleic acid molecules. Most importantly, primer design is critical to amplification success. DNA polymerase attaches to primer, which is the initial nucleic acid in the chain. Amplifying the correct primer sequence and the correct amount of primer is crucial for PCR to be as specific and efficient as possible. A poorly designed primer may provide no result due to nonspecific amplification and amplification that becomes viable enough to inhibit the process. The selection of oligonucleotide primers is a critical step in performing DNA sequencing, PCR, and oligo-hybridization (Basu, 2015).

Primers can be designed using the *in silico* process. A variety of PCR primer design programs are available. However, a few are commercial programs (Singh & Kumar, 2000; Abd-Elsalam, 2019). A program may be a stand-alone program or a complex integrated networked version of the commercial software. These web-based tools, such as "Primer3" and "Web Primer," could be used for constructing primers. All sequence databases, such as the RefSeq database, are suitable for creating a primer based on DNA templates. *In silico* validation could be performed using the BLAST and Gene Runner software to determine the efficiency and specificity of the tools. As a result, it is feasible to evaluate the primers that have been designed *in silico*. The concerned gene or DNA fragment may subsequently be amplified using these verified primers.

The freely available Primer-BLAST software comprises two modules: creating applicant primer pairs and validating the produced primer pairs' target specificity. Primer3 is used in order to determine the viable primer pairings that match to a given template strand. Using a combination of BLAST and the Needleman-Wunsch (NW) global alignment approach, the selectivity testing module searches for matches between targets and primers (Needleman & Wunsch, 1970). Primer-BLAST targets primers to one or more specific areas, including bases, using the Primer3 programming interface and the NCBI C++ toolkit. Primer-BLAST provides a variety of abilities not found in other software programmes, such as the ability to custom 3' end region with a specified number of mismatches and to define the number of mismatches a primer pair must have with unwanted targets. Primer-BLAST is the only programme that will design primers on different exons to avoid genomic DNA amplification, and it is also the only software that enables adjustment of the number of nucleotides matching on either side of an exon or exon junction (Ye et al., 2012).

The OligoAnalyzer is the main calculator for *in silico* primer validations, as it predicts various types of data about an oligonucleotide sequence. The user could provide a nucleotide sequence and associated conditions. When the oligonucleotide hybridizes to the corresponding sequence, a melting temperature is predicted. When the “ANALYZE” button is pressed, the oligonucleotide's physical parameters include oligonucleotide length, extinction coefficient at 260 nm, GC base composition, complementary sequence, molecular weight, and melting temperature shown. The results of alterations are represented in the oligonucleotide extinction coefficient using values derived from the literature or by Integrated DNA Technologies (Owczarzy et al., 2008). After selecting the “HAIRPIN” button, the user will be presented with an input form for estimating secondary oligonucleotide structures. The NCBI BLAST button transmits the sequence to the NCBI website, which uses the brief nearly exact matches

approach to scan different databases (Altschul et al., 1990). This approach could estimate the oligonucleotide's annealing locations within a genome or other cluster of candidate sequences.

The Open Reading Frame Finder (ORF Finder) (www.ncbi.nlm.nih.gov) is a graphical analytic tool that identifies all available ORF of a user-specified minimum size inside a user's sequence or an existing sequence in the database. This tool identifies all ORF using either the common or a different genetic coding system. Using the basic local alignment search tool (BLAST) service, the amino acid sequence may be stored in various formats and searched against the sequence database. The ORF Finder should facilitate the submission of accurate and comprehensive sequences to the primer design procedure. The Sequin sequence submission software is included as well (sequence analyzer).

### 2.3 Kelampayan (*Neolamarckia cadamba)*

*Neolamarckia cadamba* or locally known as the Kelampayan tree is a multifunctional tree that self-prunes and thrives in overgrazed regions such as logged-over woods and plantations in Malaysia, especially in Sarawak. In 1984, Jean Marie Bosser named *N. cadamba* to honour the Asian genus Lamarck, which matched Achille Richard's description of the species as *Anthocephalus indicus* (Razafimandimbison & Sylvain, 2002*).* The tropical evergreen Kelampayan tree, also known as Kadam or cadamba, is a South and Southeast Asian indigenous species. There are 44 chromosomes in *N. cadamba*, which is also known as *A. cadamba Roxbis*. It belongs to the *Rubiaceae* family and is naturally tetraploid (Eng et al., 2021). It is also known as Laran, burflower tree, and Leichhardt pine in English (*Neolamarckia Cadamba*, 2020). It is a relatively light hardwood with a density of 290–560 kg/m$^3$ with a moisture content of 15% (Joker, 2000).

Kelampayan trees are deciduous with huge spreading branches that may grow fast over a six to eight year period, reaching a height of 45 m with branch lengths exceeding 25 m (Figure

2.1). Hand and machine tools could be used to get a beautiful finish on the wood, which has a fine or medium texture (Lal et al., 2010). The circumference ranges from 100 to 160 cm but is usually smaller and buttresses may be present. The crown is formed like an umbrella, and the branches are organized in layers. Simple opposite leaves, generally with domatia, are petiolate or sessile. The buds are conical but without raphides. The leaves are ovate, elliptic, or oblong-elliptic in form, with an acute to the acuminate apex and a length of 13-32 cm shown in Figure 2.2 (Joker, 2000). The flowers are bisexual, monomorphic, and actinomorphic and are subsessile on a glabrous receptacle (Figure 2.3). Furthermore, the flowers are tiny and yellow or orange (Mojiol et al., 2018). Flowering typically occurs when the tree reaches the age of five years. Wind, rain, floods, and rivers all contribute to seed dispersal. The fruits are tiny capsules that create a plump, golden, or orange-coloured infructescence with around 8,000 seeds (Figure 2.4). The small capsules break into four halves at maturity, releasing the seed. Approximately 20,000 seeds are included in each gram (Joker, 2000).

***Figure 2.1*** Kelampayan tree.

***Figure 2.2*** Leaves of Kelampayan tree.

***Figure 2.3*** Flower of Kelampayan tree.

***Figure 2.4*** Fruits of Kelampayan tree.

*Note.* The tree structure of Kelampayan tree or known as *Neolamarckia cadamba*. Adapted from 8 Health benefits of Kadamba Tree (Burflower) by Slyvie, 2017, Health Benefits.

*N. cadamba*, is a hardwood subtropical plant species native to South Asia, Southeast Asia, and Australia. The natural range extends from India, Nepal, and eastward via the Malaysian Archipelago to Papua New Guinea and India through Thailand and Indochina. It has been introduced successfully into Central America and Africa (Mojiol et al., 2018). It is a typical pioneer that grows in forest areas. In addition, it can thrive in dry areas with as little as 200 mm of annual precipitation and a range of soil types. It is found below 1000 m altitude and often where there is more than 1500 mm of precipitation. Furthermore, light exposure is essential for Kelampayan's ideal growth development, with the lowest temperature ranging from three to 15.5°C and the highest temperature ranging from 32 to 42°C. Likewise, it requires a lot of light and is sensitive to frost (Joker, 2000). It can endure flooding and thrive in a variety of soil types. It survives near riverbanks and in the zone between wetlands, frequently flooded areas, and dry loams. (Haruni Krisnawati et al., 2011).

The fruits of Kelampayan can be sliced into tinier pieces and dehydrated in the sun to make dried fruits using a fine sieve in Indonesia and Laos. Besides, its fresh fruits also were taken and dissolved in water until the seeds are visible and then dried up under the sun in the Philippines (Joker, 2000). Kelampayan has lightweight hardwood with a low strength evaluation, making it unsuitable for outdoor use. Generally, the wood grain is delicate with a smooth texture. These properties contribute to the effective and practical use of wood as a raw

material in the wood industry. The pulp is used chiefly to manufacture low- and medium-grade paper sheets. It is ideal for light construction projects such as boxes, rafters, packing cases, tea chests, wooden shoes, ceiling boards, toys, pencils, and chopsticks (Mojiol et al., 2018). Moreover, suppose the hardwood is adequately polished. In that case, it is appropriate for use as a material for upholstery and suitable for light building work. Still, it should be used only inside due to its perishability when in touch with dirt.

Kelampayan trees may also provide shade for dipterocarp line plantings, in agroforestry systems for replanting in watersheds, windbreaks, and eroded regions. For medicinal purposes, the leaves and bark are often used since it is a plant that has been used for centuries to cure a variety of ailments (Pandey & Negi, 2016). In the past years, Ahmed et al. (2011) published a report based on the study of the pharmacological properties, phytochemical contents, and the biological process of extracts from different sections of *N. cadamba*. Phytochemical research conducted previously on this plant resulted in the identification and isolation of new indole alkaloids. Two previously described unique natural products, aminocadambines A and B, may be further changed by the amino acids cadamine, isocadambine, cadambine, neolamarckines A and B, dehydraisodihydrocadambine, nitrocadambines A and B, spirocadambine (including glucoalkaloids), and 3b-isodihydrocadambine (Liu et al., 2008; Dubey et al., 2011; Qureshi et al., 2011; Pang et al., 2015). In the phytochemical screening, Zayed et al. (2014) found 26 compounds, including n-hexadecanoic acid (44.8%), hexadecenoic acid ethyl ester (17.9%), and octadecanoic acid ethyl ester (11.71%).

The phenolic content of Kelampayan leaves and fruit extracts is high, and they show substantial DPPH free radical and $Fe^{2+}$ chelating properties, comparable to conventional antioxidants. Additionally, the antioxidant capabilities of the extracts were shown to be linked to their phenolic concentration (Ganjewala et al., 2013). The methanolic extract of

Kelampayan leaf lowered hyperglycemic mice's blood glucose levels, with the maximum anti-hyperglycemic action at 400 mg per kg of body weight, equivalent to glibenclamide (10 mg/kg) (Ahmed et al., 2011). Lastly, the leaves extract proved antifungal against *Aspergillus fumigatus Fresen* and *Candida albicans* (C. P. Robin) Berkhout (Pandian et al., 2013). (Pandian et al., 2013).

### 2.4 *Fasciclin-Like Arabinogalactan proteins (FLAs)*

The *fasciclin-like arabinogalactan proteins (FLAs)* are required for the growth and adaptability of plants to their environment. This gene family's evolutionary history in plants is currently unknown (He et al., 2019). *FLAs* have conserved fasciclin (FAS) domains and AGP-like glycosylated regions. Both fasciclin domains and arabinogalactan protein (AGP) sections have been identified in several plant species. HRGPs are identified by a protein backbone high in hydroxyproline (Hyp). Based on the varying degrees of O-glycosylation, the HRGPs superfamily may be classified into three major subfamilies; proline-rich proteins (PRPs) arabinogalactan proteins (AGPs), and extensins (EXTs) (Nothnagel et al., 2000; Showalter, 2001, Jamet et al., 2008). These amino acids are organised as Ala–Pro, Ser–Pro, and Thr–Pro, which were inserted as arabinogalactan (AG) glycomodules. (Schpak et al., 2001; Ellis et al., 2010; Ma et al., 2017). AGPs have carbohydrate side chains that are richer in arabinose and galactose and are linked to Hyp (Showalter et al., 2018). Based on their distinct protein backbones, AGPs may be classified as chimeric AGPs, AGP-EXT hybrids classical or AGPs. Based on their conserved domains, the chimeric AGPs may be further separated into three subclasses: phytocyanin-like AGPs (PAGs), fasciclin-like AGPs (FLAs), and xylogen-like AGPs (XYLPs) (Kobayashi et al., 2011; Ma et al., 2011; MacMillan et al., 2015). *FLAs* are a subclass of chimeric AGPs that have both fasciclin and AGP domains. Most plant species consist of at least one fasciclin domain, which is comprised of two highly conserved region

sections (H1 and H2) containing around 10 amino acids each and one [Phe/Tyr]-His motif (Johnson et al., 2011). It has been shown that this domain functions as extracellular matrix adhesion motifs (Kim et al., 2000). The first protein containing fasciclin domains was discovered in grasshoppers, while the first adhesion factor was found in fruit flies (Bastiani et al., 1987; Elkins et al., 1990). Ever since, fasciclin domains have been discovered in a large variety of proteins from plants, bacteria, yeast, and mammals (Seifert, 2018).

*FLAs* comprise the majority of fasciclin-like proteins found in plants. Among their roles are oedema and interpolymer interconnection, secondary cell wall development and composition, salt stress sensing in the root system, organ formation, and male gametophyte growth (Seifert, 2018). *FLAs* discovered in a variety of plants so far, including in thale cress (*Arabidopsis thaliana*), rice (*Oryza sativa*), zinnia (*Zinnia elegans*), wheat (*Triticum aestivum*), cotton (*Gossypium raimondii*), poplar (*Populus trichocarpa*), sea island cotton (*Gossypium barbadense*), Chinese cabbage (*Cannabis sativa*). In evolutionary history, it has been hypothesised that there was a divergence event, indicated by the decreased number of fasciclin domains in terrestrial plants compared to algae. In addition, introns in *FLA* genes are deleted during plant evolution, particularly during the transition from green algae to land plants. Furthermore, gene duplication events are revealed to be crucial for the growth of *FLA* gene families. The duplicated gene pairs in the *FLA* gene family develop mostly by purifying selection (He et al., 2019).

# CHAPTER 3: MATERIALS AND METHODS

## 3.1 Data Mining

The *FLAs-related* gene sequences were obtained from the transcriptome database (NcdbEST) (Ho et al., 2014; Pang et al., 2015). The *FLAs* nucleotide sequences of Kelampayan were used as query sequences and BLASTn was carried out against the NCBI (National Center for Biotechnology Information) GenBank database.

## 3.2 Phylogenetic Analysis of *FLAs* Evolution

A total of 49 *FLAs-related* genes were obtained from NCBI database including six *FLAs* genes of *Arabidopsis thaliana* (*AtFLAs),* 12 *FLAs* genes of *Cuscuta campestris (CcFLAs),* ten *FLAs* genes of *Populus alba x Populus tremula (PaxPtFLAs),* seven *FLAs* genes of *Triticum aestivum* (*TaFLAs*), five *FLAs* genes of *Wolffia Australiana* (*WaFLAs*), one *FLAs* gene of Duckweed (*DwFLAs*), one *FLAs* gene of *Gossypium hirsutum* (*GhFLAs*), one outgroup fungi (*Lachnellula hyaline),* two BLASTn output (*Coffee arabica* and *Coffee eugenioide)* and *FLAs* in Kelampayan (*cn1335, cn1339, cn1391,* and *cn1517*). Full-length *FLAs* nucleotide sequences of *Arabidopsis thaliana* (At), *Cuscuta campestris (Cc), Populus alba x Populus tremula (PalxPt), Triticum aestivum (Ta), Wolffia australiana (Wa), Duckweed (Dw), Gossypium hirsutum (Gh),* and *FLAs* in Kelampayan were aligned with MUSCLE tools using MEGA11 software (Kumar et al., 2018) and the phylogenetic tree was then constructed by neighbour-joining (NJ) method (Nei and Kumar, 2000) with the parameters p-distance and pairwise deletion at the bootstrap value of 1000.

### 3.3 Conserved Domain and Motifs Identification

The Conserved Domain Database in NBCI was applied to analyze the presence of the *FLA-related* domain among the *FLAs* in Kelampayan. The conserved motifs in 49 *FLA*-related genes were identified using the MEME suite servers (https://memesuite.org/meme/tools/meme) online tool with preset parameters of a maximum of five number motifs, zero or one occurrence per sequence, and an optimum width character from 10 to 70.

### 3.4 Primer design

Primers were designed using online bioinformatics software, Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast). The forward and reverse primers for non-redundant *FLAs* genes were designed using Primer-BLAST based on the following core criteria: (1) annealing temperature ($T_a$) between 52˚C and 65˚C; (2) GC content between 40% and 60%; (3) expected products size ranging from 100bp to 500 bp; (4) primer length ranging from 18 bp to 24 bp (Yuliani & Jannah, 2017).

### 3.5 Primer Design Validation

Characteristics of primer analysis applied by online OligoAnalyzer Tool (https://sg.idtdna.com/calc/analyzer). The application of $T_m$ (melting temperature), hairpin, secondary structure, and GC ratio values was investigated. Furthermore, the primers' specificity was confirmed using BLAST against the NCBI's (nr) database. The plant was used as the organism in the BLAST search, which was run against the nonredundant (nr) database.