MOLECULAR CLONING AND RE-SEQUENCING OF TRYPSIN INHIBITOR FROM A TROPICAL TIMBER TREE RED KELAMPAYAN
(\textit{Neolamarckia macrophylla})

Ho Pei Yin

Bachelor of Science with Honours
(Resource Biotechnology)
2014
Molecular Cloning and Re-sequencing of Trypsin Inhibitor from A Tropical Timber Tree *Neolamarckia macrophylla* (Red Kelampayan)

Ho Pei Yin  
(30385)

Bachelor of Science with Honours  
Resource Biotechnology  
2014
ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my supervisor, Dr. Ho Wei Seng and my co-supervisor, Dr. Pang Shek Ling for their motivation, patient guidance, and helpful advices throughout the process of completing my research and writing of dissertation. Other than that, I would like to send well-deserved thank you to them for sparing their time to help me in solving my doubts promptly. The process of completing my Final Year Project in the Forest Genomics and Informatics Laboratory (fGiL) was fruitful in the aspects of improving my knowledge and practical skills. I am thankful to have supportive parents as they always give me unlimited care and support whenever I need them, thank you very much. The help and encouragement from Ms. Natalie Anak Gali (Msc. Student) and Ms. Maslini Binti Japar Ali (PhD. Student) are gratefully acknowledged as well. I thank the Universiti Malaysia Sarawak (UNIMAS) and the Faculty of Resource Science and Technology for supporting me and giving genuine impact in many ways on my life. Furthermore, a special thank you goes to my batch mates for supporting and helping me constantly. Lastly, I would like to warmly acknowledge other people who have helped me in completing my Final Year Project. Each of them had helped me and shown their goodwill in every aspect.
DECLARATION

I hereby declare that this thesis is based on original work except for the quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UNIMAS or other institutions.

HO PEI YIN, 30385

Date: 25 June 2014

Resource Biotechnology

Department of Molecular Biology

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak (UNIMAS)
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section/Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>DECLARATION</td>
<td></td>
<td>II</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td></td>
<td>III</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td></td>
<td>VII</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>IX</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>ABSTRACT / ABSTRAK</td>
<td></td>
<td>XIV</td>
</tr>
</tbody>
</table>

### SECTION I

**INTRODUCTION**

1

### SECTION II

**LITERATURE REVIEW**

2.1 Family Rubiaceae

5

2.2 *Neolamarckia macrophylla*

5

2.3 Proteinase Inhibitor (PIs)

6

2.4 Trypsin Inhibitors

8

2.4.1 Kunitz-type Trypsin Inhibitor

9

2.4.2 Woody Plant Species with Trypsin Inhibitor

10

2.4.3 Positive Effects of Application of Trypsin Inhibitor

11

2.4.4 Negative Effects of Application of Trypsin Inhibitor

12

2.4.5 Insect Pests Overcome Effects of Trypsin Inhibitor

13

2.5 Miraculin-like Protein

14

2.5.1 Potential Application of Miraculin-like Protein

15
2.6 Miraculin protein

2.7 Single Nucleotide Polymorphisms (SNPs)

2.7.1 Applications of Single Nucleotide Polymorphisms (SNPs)

SECTION III MATERIALS AND METHODS

3.1 Materials and Reagents

3.2 Plant Material Selection

3.3 Total RNA Extraction

3.4 RNA Quantification

3.5 Band Visualization

3.6 Primer Design

3.7 PCR Optimization

3.8 Reverse Transcription – Polymerase Chain Reaction (RT – PCR)

3.8.1 Reverse Transcription

3.8.2 Polymerase Chain Reaction (PCR)

3.9 PCR Product Purification

3.10 Single Nucleotide Polymorphisms (SNPs) for Trypsin Inhibitor

3.10.1 DNA extraction

3.10.2 Polymerase Chain Reaction (PCR)

3.11 PCR Product Purification

3.12 DNA Sequencing and Nucleotide BLAST
SECTION IV RESULTS AND DISCUSSION

4.1 Plant Materials 27
4.2 Total RNA Extraction and Quantification 27
4.3 Genomic DNA Extraction and Quantification 29
4.4 Band Visualization 29
4.5 Primer Design 30
4.6 Reverse Transcription – Polymerase Chain Reaction (RT–PCR) 31
4.7 PCR Products Purification 32
4.8 DNA Sequencing and Nucleotide BLAST 33
4.9 Comparison between Miraculin-Like Protein and Trypsin Inhibitor 34
4.10 Gene Structure Comparison 35
4.11 Neighbour-Joining (NJ) Phylogenetic Tree 38
4.12 Domain of Miraculin-Like Protein 40
4.13 Comparative Analysis of Miraculin-Like Protein from Neolamarckia macrophylla and Neolamarckia cadamba 41
4.14 Sequence Variations in Miraculin-Like Protein of Neolamarckia macrophylla 43
   4.14.1 Optimization of Polymerase Chain Reaction (PCR) 43
   4.14.2 Single Nucleotide Polymorphisms (SNPs) for Miraculin-Like Protein 46
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Ampere</td>
</tr>
<tr>
<td>AGE</td>
<td>Agarose Gel Electrophoresis</td>
</tr>
<tr>
<td>BBI</td>
<td>Bowman-Birk Inhibitor</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>CDD</td>
<td>Conserved Protein Database</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CIA</td>
<td>Chloroform-isoamyl alcohol</td>
</tr>
<tr>
<td>cm</td>
<td>Centimeter</td>
</tr>
<tr>
<td>CTAB</td>
<td>Cetyl trimethyl ammonium bromide</td>
</tr>
<tr>
<td>dbh</td>
<td>Diameter at breast height</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>Double-distilled water</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DBM</td>
<td>Diamondback moth</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium bromide</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>InDel</td>
<td>Insertion/Deletion</td>
</tr>
<tr>
<td>MAS</td>
<td>Marker Assisted Selection</td>
</tr>
<tr>
<td>mM</td>
<td>Millimolar</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>Magnesium chloride</td>
</tr>
<tr>
<td>MTI2</td>
<td>Mustard Trypsin Inhibitor 2</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>NJ</td>
<td>Neighbour-Joining</td>
</tr>
<tr>
<td>PCI</td>
<td>Phenol: chloroform: isooamylalcohol</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PIs</td>
<td>Plant proteinase inhibitors</td>
</tr>
<tr>
<td>ppm</td>
<td>Parts per million</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative Trait Locus</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse Transcription Polymerase Chain Reaction</td>
</tr>
<tr>
<td>ST</td>
<td>Stepwise Clustering</td>
</tr>
<tr>
<td>STI</td>
<td>Serine trypsin inhibitor</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris base, acetic acid and EDTA</td>
</tr>
<tr>
<td>μl</td>
<td>Microliter</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet light</td>
</tr>
<tr>
<td>V</td>
<td>Voltage</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>PCR thermal cycling profile</td>
<td>20</td>
</tr>
<tr>
<td>3.2</td>
<td>Components needed for synthesizing first strand cDNA</td>
<td>21</td>
</tr>
<tr>
<td>3.3</td>
<td>Components for PCR</td>
<td>21</td>
</tr>
<tr>
<td>3.4</td>
<td>PCR reaction conditions</td>
<td>22</td>
</tr>
<tr>
<td>3.5</td>
<td>Components for PCR</td>
<td>25</td>
</tr>
<tr>
<td>3.6</td>
<td>PCR reaction conditions</td>
<td>25</td>
</tr>
<tr>
<td>4.1</td>
<td>BLASTn output for amplified partial trypsin inhibitor cDNA</td>
<td>34</td>
</tr>
<tr>
<td>4.2</td>
<td>CDD search output for amplified partial trypsin inhibitor cDNA</td>
<td>40</td>
</tr>
<tr>
<td>4.3</td>
<td>Optimized thermal cycling profile</td>
<td>44</td>
</tr>
<tr>
<td>4.4</td>
<td>BLASTn output for amplified partial trypsin inhibitor DNA</td>
<td>48</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Condition of forest cover in Sarawak, Sabah, and Brunei in 2009.</td>
<td>2</td>
</tr>
<tr>
<td>2.1</td>
<td><em>N. macrophylla.</em> (a) 2.5-year-old and (b) rough texture and dark colour of trunk.</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Ribbon structure of Kunitz trypsin inhibitor. The β strands of the proteins are represented as the green arrows. The pink ropes around the structure represent the loops while the red ribbon is representing the $3_{10}$ helix. The two disulphide bonds are shown in the figure as blue labellings of C39 – C86 and C136 – C145 respectively whereas the R63(P1) is Arginine 63.</td>
<td>9</td>
</tr>
<tr>
<td>4.1</td>
<td>Agarose gel electrophoresis (AGE) of PCR products synthesized from first strand cDNA. Lane M: 100 bp DNA marker. Lane 1 – 4: PCR product synthesized from first strand cDNA.</td>
<td>31</td>
</tr>
<tr>
<td>4.2</td>
<td>AGE for purified PCR product synthesized from first strand cDNA. Lane M₁: 100 bp DNA marker. Lane M₂: <em>λ HindIII</em> Lane 1: Purified PCR product synthesized from cDNA</td>
<td>32</td>
</tr>
<tr>
<td>4.3</td>
<td>AGE for purified PCR product synthesized from genomic DNA. Lane M₁: 100 bp DNA marker. Lane M₂: <em>λ HindIII</em> Lane 1: Purified PCR product synthesized from DNA.</td>
<td>33</td>
</tr>
<tr>
<td>4.4</td>
<td>Comparison between amplified partial trypsin inhibitor cDNA and DNA. RKTI 1 – 7: amplified partial miraculin-like genomic DNA. RRKTI: amplified partial miraculin-like cDNA.</td>
<td>36</td>
</tr>
<tr>
<td>4.5</td>
<td>The chart above shows the gene structure comparison of partially isolated cDNA encoded for miraculin-like gene from <em>N. cadamba</em>.</td>
<td>37</td>
</tr>
</tbody>
</table>
(RWKTI) and *N. macrophylla* (RRKTI) and complete cds mRNA (*Coffea arabica* Genbank accession no.: DQ993351.1) respectively. The dotted lines above yellow portions represent small gaps of nucleotides. 1 unit axis: 141 bp.

**4.6 Unrooted Neighbour-Joining (NJ) Phylogenetic Tree**

constructed through ClustalX 2.0 and MEGA 6. The sequence name marked with darken dot represents miraculin-like gene isolated from *Neolamarckia macrophylla*. The tree was constructed based on the gene from miraculin-like protein and trypsin inhibitor.

Abbreviations: Nm: *Neolamarckia macrophylla*; Nc: *Neolamarckia cadamba*; Nb: *Nicotiana benthamiana*; Cj: *Citrus japonica*; Cs: *Citrus* cv. *Shiranuhi*; Cu: *Citrus* unshiu; Ca: *Coffea Arabica*; Cp: *Citrus* x *paradise*; Cja: *Citrus jambhiri*; Gm: *Glycine max*; Mk: *Murraya koenigii*; Os: *Oryza sativa* subsp. *Japonica*; Zm: *Zea mays*; At: *Arabidopsis thaliana*; Lc: *Luffa cylindrical*; Ace: *Allium cepa*; Mj: *Marsupenaeus japonicas*; Bm: *Brugia malayi*; Bov: *Bombina variegata*; Ms: *Medicago scutellata*; Bav: *Bauhinia variegata*; Ps: *Pisum sativum*; Mt: *Medicago truncatula*; Ah: *Arachis hypogaea*; Pt: *Psophocarpus tetragonolobus*; Vu: *Vigna unguiculata*; Dre: *Delonix regia*; Syt: *Symphytum tuberosum*; Cm: *Citrus maxima*; Cl: *Curcuma longa*; Ci: *Ciona intestinalis*; Ac: *Acacia confusa*; Pni: *Populus nigra*; Ti: *Tamarindus indica*; Nt: *Nicotiana tabacum*;
PbxPd: *Populus balsamifera* subsp. *Trichocarpa* x *Populus deltoids*;
Bma: *Bombina maxima*; Ptre: *Populus tremuloides*;
Ptr: *Populus trichocarpa*; Tg: *Toxoplasma gondii*;
St; *Solanum tuberosum*; Za: *Zonotrichia albicollis*;
Dar: *Danio rerio*; Ft: *Fagopyrum tataricum*; Fl: *Solanum lycopersicum*;
Dro: *Droypetes roxburghii*; lb: *Ipomoea batatas*;
Ahy: *Arachis hypogaea*; Ds: *Descurainia sophia*;
Am: *Astyanax mexicanus*; Pn: *Pundamilia nyererei*;
Pc: *Phaseolus coccineus*; Ch: *Cocculus hirsutus*;
Hv: *Hordeum vulgare*; Si: *Setaria italic*;
Bd: *Brachypodium distachyon*.

The accession number of all the sequences are shown in Appendix II.

**4.7** Nucleotide sequence alignment output of red (query) and
white (subject) kelampayan.

**4.8** Amino acid sequence alignment output of red (query) and
white (subject) kelampayan.

**4.9** Gel electrophoresis of 1.5% agarose gel for optimization of annealing
temperature for PCR. Lane M: 100 bp DNA marker. Lane 1: 61.0°C
Lane 2: 61.5°C. Lane 3: 62.5°C. Lane 4: 63.5°C. Lane 5: 64.5°C.
Lane 6: 65.5°C. Lane 7: 66.5°C.

**4.10** Gel electrophoresis on 1.5% agarose gel for optimization of
concentration for DNA sample. Lane M: 100 bp DNA marker.
Lane 1: 5 ng/µl. Lane 2: 10 ng/µl. Lane 3: 15 ng/µl. Lane 4: 20 ng/µl.
Lane 5: 25 ng/µl. Lane 6: 30 ng/µl. Lane 7: 35 ng/µl.
Lane 8: 40 ng/µl. Lane 9: 45 ng/µl. Lane 10: 50 ng/µl.
Lane 11: 55 ng/µl. Lane 12: 60 ng/µl.
4.11 Gel electrophoresis on 1.5% agarose gel for optimization of concentration for Magnesium chloride (MgCl\textsubscript{2}).

Lane M: 100 bp DNA marker. Lane 1: 0.5 mM. Lane 2: 1.0 mM.
Lane 3: 1.5 mM. Lane 4: 2.0 mM. Lane 5: 2.5 mM. Lane 6: 3.0 mM.
Lane 7: 3.5 mM. Lane 8: 4.0 mM.

4.12 Multiple alignments of 7 sequences of miraculin-like protein gene from \textit{N. macrophylla}. There were 8 SNPs detected.

RKTI 1 – 7: amplified partial miraculin-like genomic DNA.
Molecular Cloning and Re-sequencing of Trypsin Inhibitor from A Tropical Timber Tree *Neolamarckia macrophylla* (Red Kelampayan)

**HO PEI YIN**

Resource Biotechnology  
Faculty of Resource Science and Technology  
Universiti Malaysia Sarawak

**ABSTRACT**

*Neolamarckia macrophylla* (red kelampayan) is a relatively unexplored fast-growing woody plant species with superior traits which can be used in reforestation programmes. Thus, the identification of the ability of it in resisting insect pests is needed. Insect pest resistant plants containing trypsin inhibitor, member of proteinase inhibitor can be useful in enhancing the economic traits of the plants. The presence of it helps in minimizing or excluding the usage of pesticides by utilizing the natural defence system of the plants. Therefore, the partial cDNA encoding for trypsin inhibitor from *N. macrophylla* was isolated and characterized by using in silico tools. Based on the nucleotide sequence analysis, the cDNA isolated was miraculin-like gene which has the size of ~ 323 bp. The miraculin-like protein belongs to the Kunitz serine trypsin inhibitor family. It has the same function as trypsin inhibitor in protecting the plants from insect pests naturally. At the DNA level, the rapid and mass discovery of single nucleotide polymorphisms (SNPs) has becoming important in the studies of plant genetics. The results of re-sequencing indicate that there were 8 in silico SNPs identified from 7 samples of *N. macrophylla*.

**Keywords:** *Neolamarckia macrophylla*, cDNA, trypsin inhibitor, miraculin-like protein, single nucleotide polymorphisms (SNPs)

**ABSTRAK**

*Neolamarckia macrophylla* (red kelampayan) merupakan spesis tumbuhan berkayu yang jarang diterokai dan tumbuh dengan cepat serta mempunyai sifat-sifat unggul yang sesuai untuk perladangan hutan. Oleh itu, keupayaan *N. macrophylla* untuk menahan serangan serangga perosak perlu dikenalpastikan. Tumbuhan yang mengandungi trypsin inhibitor, ahli proteinase inhibitor dapat meningkatkan nilai komersialnya. Kehadiran trypsin inhibitor dapat mengurangkan atau mengecualikan penggunaan racun perosak dengan adanya system pertahanan semula jadi. Dengan itu, cDNA separa daripada *N. macrophylla* yang menyandui trypsin inhibitor telah didapati dan seterusnya dicirikan dengan menggunakan alat in silico. Berdasarkan urutan nukleotida analisis, cDNA separa yang didapati merupakan gen miraculin-like dengan saiz yang dijangka sepanjang ~ 323 bp. Protein miraculin-like tersebut dikategorikan dalam keluarga Kunitz serine trypsin inhibitor. Protein tersebut mempunyai fungsi yang sama dengan trypsin inhibitor dalam melindungi tumbuh-tumbuhan daripada serangga perosak. Pada peringkat DNA, penemuan single nucleotide polymorphisms (SNPs) telah menjadi semakin penting dalam pengajian genetik tumbuh-tumbuhan. Hasil re-sequencing telah menunjukkan bahawa terdapat 8 in silico SNPs telah dikenalpasti daripada 7 sampel *N. macrophylla*.

**Kata kunci:** *Neolamarckia macrophylla*, cDNA, trypsin inhibitor, protein miraculin-like, single nucleotide polymorphisms (SNPs)
SECTION I
INTRODUCTION

The world is growing with ever-increasing demand of more people towards wood to fulfil their needs. The decrease in the number of trees is correlated to the increment of rate of forest exploitation for commercial purposes such as pulp and paper (Goulao et al., 2011). According to Achard et al. (2002), the estimated deforestation rate of tropical forest is 5.8 ± 1.4 million hectares each year from 1990 to 1997. In addition, another 2.3 ± 0.7 million hectares of degraded forest will be visible through satellite imaginary. A recent study by Bryan et al. (2013) revealed that starting from 2000 to 2005, the coverage of tropical forest lost was about 27 million hectares for plantation. Furthermore, roughly 398 million hectares of land were cleared for logging purposes. The percentage of intact forest for Sarawak and Brunei is 3% and 54%, respectively (as shown in Figure 1.1) which shows that the protection of forest is relatively ineffective in Sarawak (Bryan et al., 2013). The occurrence of this issue is due to timber and oil palm industries which makes Sarawak to be one of the forest-losing hotspots (Bryan et al., 2013). In addition, the peatland forests in Sarawak had experienced drastic level of deforestation due to plantation development starting from 2005 to 2010 (Miettinen et al., 2011). The continuous activity of deforestation has caused serious consequences to our environment. Therefore, the replant of forest with fast-growing tree species is becoming more acute and necessary in meeting human needs and protecting natural environment (Goulao et al., 2011).

The usage of pesticides in forest plantations is intensive (Kruger and Seville, 2013). It is a drawback for plantations utilizing pesticides which brings health issues to the workers, causes death to non-targeted plants, and soil and water pollution (Amoguis, 2010; Rajvaidya and Markandey, 2008). The subsequent effects of the pollutions caused by pesticides are severe and can be transferred in successive levels of food chain (Rajvaidya...
and Markandey, 2008). Thus, it is necessary to decrease or avoid the usage of pesticides. In order to do so, the pest-resistant trait of the plant must be identified followed by application of genetics studies to improve the quality of the plants. For example, selective breeding is useful in producing plants with good quality traits.

Bradshaw et al. (2000) stated that tree biology is studied due to its economic and ecological importance. *Neolamarckia macrophylla* is expected to be much superior in various aspects compared to other fast growing species such as *Anthocephalus cadamba*, and *Paraserianthes falcataria* (Halawane et al., 2011). The quality of the wood produced is relatively high despite of the fast growing rate. It is also not susceptible to the attack of fungi *Uromycladium tepperianum* which destroyed over 15 ha of *Albizia chinensis* within two months in Indonesia. Apart from that, it can be harvested after 5 to 6 years which is 5
times faster than other natural wood-producing plants (Halawane et al., 2011). Therefore, it is suitable to be used as timber stock in forest replantation.

Trypsin inhibitor is a type of plant proteinase inhibitors (PIs) which helps in fighting against herbivorous insects (Fan and Wu, 2005). Moreover, it is classified as the secondary plant compound which is synthesized in response to initiate plant defense system (Ryan, 1990). Besides that, the endogenous proteinases in the plants can be regulated by trypsin inhibitor as well (Altpeter et al., 1999). Although the woody plant species are susceptible to the attack of insect pests throughout the stages of its growth (Rao et al., 2000), but the presence of trypsin inhibitor could be able to protect the plant from insect pests. There is presence of trypsin inhibitors in many woody plant species. Thus, trypsin inhibitor is suspected to be present in Neolamarckia macrophylla, a woody plant species, in fighting against insect pests.

Trypsin inhibitor is a natural pesticide that does not bring irreversible adverse effects to the environment. Thus, the plants having special mechanisms such as the expression of trypsin inhibitor genes are able to overcome the attack of insect pests by hindering their growth and causing death. This is especially effective towards insect pests from orders Lepidoptera, Coleoptera, and Homoptera (Gatehouse and Gatehouse, 1998).

Besides trypsin inhibitor, there are many other types of Kunitz proteinase inhibitors found in different plants, for instance miraculin-like protein which is in the same family of trypsin inhibitor is one of them. The miraculin-like protein has the similar function as trypsin inhibitor in plant defense by inhibiting proteases in insect pests and it is found in Rubiaceae plant family (Gahloth et al., 2010; Gahloth et al., 2011; Mondego et al., 2011; Selvakumar et al., 2012). Additionally, miraculin-like protein also has antifungal property (Wang and Ng, 2002). The miraculin-like protein was found to be present in Neolamarckia
*cadamba* which is from the same family (family Rubiaceae) as *Neolamarckia macrophylla* (Perera, 2011). Compared to *N. cadamba*, *N. macrophylla* has insufficient information available to the public because it is not properly explored. Thus, more research needs to be done on it to identify its useful traits for various applications.

Other than genetic engineering, plants with desired traits can be produced through selective breeding. In plant breeding, the usage of phenotypic markers is limited because the phenotypes are affected by the environmental factor (Jehan and Lakhanpaul, 2006). This limitation is overcame with the evolvement of sequencing technology that allows the discovery of single nucleotide polymorphisms (SNPs) and insertions or deletions (InDels) which shows the differences between the alternative forms of the gene (Rafalski, 2002). These alternative forms are useful in many applications especially as DNA markers for plant molecular genetics. This powerful approach ensures the correct and optimal selection of seed lines of *N. macrophylla* that are not susceptible to insect pests. Subsequently, the economic traits of *N. macrophylla* can be enhanced.

The objectives of this research were:

1. To isolate a partial cDNA encoding for trypsin inhibitor of *Neolamarckia macrophylla*.
2. To *in-silico* characterize the cDNA encoding for trypsin inhibitor of *Neolamarckia macrophylla*, and
3. To identify single nucleotide polymorphisms (SNPs) for trypsin inhibitor of *Neolamarckia macrophylla*.
SECTION II

LITERATURE REVIEW

2.1 Family Rubiaceae

There is about 13,000 plant species from family Rubiaceae found all around the world (Kainulainen et al., 2013; Karou et al., 2011). Besides that, this family has 630 genera found in tropical regions. In Africa, the plants from this family have medicinal values other than ornamental values. The researchers are expecting to find out quinine similar compounds which have anti-malarial property from plants in this family. Other than that, the biological properties of Rubiaceae species are studied and documented as anti-plamodial, anti-bacterial, anti-inflammatory, anti-diabetic, anti-hyperthermic, anticonvulsive, and anti-pyretic. However, there are cases where toxic of medicinal plants in family Rubiaceae causes poisonings (Karou et al., 2011). According to Karou et al. (2011), the anti-microbial property of plants from family Rubiaceae might be useful in fighting against AIDS opportunistic.

2.2 Neolamarckia macrophylla

The Neolamarckia macrophylla (red kelampayan) has other names such as Anthocephalus macrophyllus, Bancalus macrophyllus, and Nauclea macrophylla (Sosef, 1993). It is from family Rubiaceae. It has straight trunk due to its special characteristic where the shedding of leaves allows the tree to grow into tall tree without ascending branches (approximately 30 m from ground) as shown in Figure 2.1 (Halawane et al., 2011). Apart of that, its bark is in dark colour and rough texture as shown in Figure 2.1.

It is a type of hardwood-producing tree which has leaves size ranging from 20 to 60 cm (Sosef, 1993). The leaves surface of N. macrophylla is hairy with slight reddish in colour.
The rate of growth may reach up to 20 to 30 cm dbh within 4 to 5 years. The height may reach up to 45 m whereas the diameter of the bole may reach up to 100-160 cm. The size of flowers of this plant can reach up to 7 cm (Sosef, 1993). On top of that, the wood can be used to produce pulp and paper, wood craft, laminate boards, and plywood (Halawane et al., 2011). Moreover, it can be used for light construction as well.

The current knowledge on the mechanism of defense towards insect pests used by *N. macrophylla* remains unknown and insufficient.

**Figure 2.1** *N. macrophylla*. (a) 2.5-year-old and (b) rough texture and dark colour of trunk [Adapted from Halawane et al., 2011].

### 2.3 Proteinase Inhibitors (PIs)

In plants, there are many types of antimicrobial proteins being produced in response to fight against insect pests for survival. The examples of antimicrobial proteins are antifungal proteins, ribosome-inactivating proteins, and proteinase inhibitors (Kim et al.,...
2009). Proteinase inhibitors (PIs) are secondary metabolites produced by the plants as first line chemical defense to fight against insect pests by using anti-nutritional interaction (Kim et al., 2009; Mithöfer and Boland, 2012; Odeny et al., 2010). Besides that, proteolysis in plant development is regulated by PIs as well (Mondego et al. 2011; Odeny et al., 2010). The catalytic activity of the proteolytic enzymes is inhibited by the PIs (de Oliveira et al., 2012). There are four classes of PIs namely serine, cysteine, and the recently emerging aspartic and metalloprotease inhibitors (Kulkarni and Rao, 2009; Sharma and Suresh, 2011). During the occurrence of insect pests’ attack, the level of PIs can be up-regulated to a higher level (Khandelwal, 2011). The resistance of plants toward insect pests is correlated to the high level of PIs (Kim et al., 2009). Its application for pest control by targeting the physiological processes in the pests has been widely studied together with the merits and demerits (Howe and Jander, 2008; Khandelwal, 2011).

The gene activation is needed prior to the production of PIs (Khandelwal, 2011). When the plant is wounded, the saliva of insects is able to trigger the secretion of hormones followed by series of activation processes before PIs can be produced. The components triggering hormones release are known as elicitor molecules (Khandelwal, 2011). However, the knowledge regarding the activation processes are limited and not properly understood (Howe and Jander, 2008). Following the release of PIs, the anti-proteolytic activity of the PIs such as trypsin or chymotrypsin inhibitors is maintained by the formation of intermolecular disulphide bond (Kim et al., 2009). The intermolecular disulphide bond formed leads to the formation of dimer which is essential in inhibiting the action of proteases such as trypsin, chymotrypsin, and papain (Kim et al., 2009).

According to Kim et al. (2009), the rise of resistance of pathogens toward antibiotics has led to the study of antimicrobial proteins. Subsequently, novel antimicrobial agents can be developed by using protease inhibitors as the lead compounds. The antimicrobial agents
have the potential to inhibit the growth of various types of bacteria and fungi with its antibiotic activity. In order to produce antimicrobial agents, the protease inhibitors must be extracted. Most of the time, methods such as salt extraction, ultrafiltration, and \( C_{18} \) reverse phase chromatography are used to purify the protease inhibitors for further analysis (Kim et al., 2009).

In medical field, the protease inhibitors isolated from plants have the ability to inhibit the growth of tumor cells (Kim et al., 2009). Other than that, it can be used to control food intake in humans as well. This is due to the functional and non-toxic properties of the proteinase inhibitors in humans. Therefore, the utilization of protease inhibitors as the functional antimicrobial agents as well as an approach to protect and improve plants against pests is proposed by Kim et al. (2009).

In order to increase new sources of information regarding the structures and properties of Pls, many taxa needs to be studies. This is because it might be a useful source for drug design (de Oliveira et al., 2012).

### 2.4 Trypsin Inhibitor

#### 2.4.1 Kunitz-type Trypsin Inhibitor

Kunitz-type trypsin inhibitor is a type of widely studied serine protease inhibitors with approximately 20kDa of molecular mass containing two disulphide bridges and one reactive site as shown in Figure 2.2 (de Oliveira et al., 2012; Selvakumar et al., 2012; Sharma and Suresh, 2011). In addition, the content of cysteine in the Kunitz inhibitors is low (de Oliveira et al., 2012). The Kunitz-type trypsin inhibitors have the ability to coagulate blood and aggregate platelet. Being anti-carcinogenesis is another biological function of Kunitz-type trypsin inhibitors (Selvakumar et al., 2012). The variations in selective pressure toward Kunitz family point out that rapid evolutionary changes is