ISOLATION AND EXPRESSION ANALYSIS OF OIL PALM *SUPPRESSOR OF PPI1 LOCUS1* (*SP1*)

MOHD AZRAI LOH BIN MOHD AZLAN LOH

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

ISOLATION AND EXPRESSION ANALYSIS OF OIL PALM *SUPPRESSOR OF PPI1 LOCUS1 (SP1)*

MOHD AZRAI LOH BIN MOHD AZLAN LOH

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **MOHD AZRAI LOH**

Matric No: **17202240/1 (SOC190001)**

Name of Degree: **MASTER OF SCIENCE (BIOTECHNOLOGY)**

Title of Dissertation:

ISOLATION AND EXPRESSION ANALYSIS OF OIL PALM *SUPPRESSOR OF PPI1 LOCUS 1 (SP1)*

Field of Study:

BIOTECHNOLOGY

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work:
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be the owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without a written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work, I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 23/09/2022

Subscribed and solemnly declared before,

Witness's Signature Date: 23/09/2022

Name:

Designation:

ISOLATION AND EXPRESSION ANALYSIS OF OIL PALM *SUPPRESSOR OF PPI1 LOCUS 1* (*SP1*)

ABSTRACT

SUPPRESSOR OF PPI1 LOCUS1 (SP1) is an important regulator of the translocon at the outer envelope membrane of chloroplasts (TOC) machinery, which is responsible for the import of precursor protein into plastid which effects the plastid biogenesis and functions. Previous studies on *Arabidopsis thaliana* had reported the upregulation of *SP1* would accelerate plastid differentiation during de-etiolation and senescence. SP1 potential effect in improving the plastid differentiation efficiency, makes it an interesting enzyme to study for crop improvement at large. In this study, data mining using NCBI and PalmXplore was able to identify a putative oil palm *SP1* homolog namely *Elaeis guineensis SP1* (*EgSP1)* (LOC105060424) that encodes two protein isoforms: X1 and X2 variants. This SP1 homolog contains the C_3HC_4 (RING-finger) domain that is conserved in SP1 proteins and showed a high sequence identity (73.76% and 73.47% for X1 and X2 variants, respectively) with the Arabidopsis SP1. Moreover, a phylogenetic tree analysis of SP1 proteins from several selected plants showed that EgSP1 clustered together with other monocots species. Next, the coding sequence of *EgSP1* was successfully isolated from cDNA derived from leaves RNA. Using specific primers, amplicons with the size of ~1000 bp were obtained, isolated, and cloned into pGEM-T Easy Vectors. After the transformation of pGEM-T Easy Vector harbouring *EgSP1* was performed into *E. coli* cells, seven PCR positive colonies were selected and subjected for plasmid extraction. From sequencing analysis of the plasmids, it was found that only the X1 transcript variants were able to be cloned and none of the plasmids harbour the X2 variants. In addition, quantitative expression analysis (qPCR) of *EgSP1* in several vegetative tissues revealed that *SP1* transcripts (variant X1) were highly expressed in the leaf tissues,

followed by the roots, with the lowest expression in the stem tissues. The findings uncovered in this study may provide the basis for future agricultural applications of the *SP1* gene in the development of oil palm genetic improvement.

Keywords: SP1 protein, plastid protein imports, TOC machinery, plastid biogenesis and differentiation, oil palm.

ISOLASI DAN ANALISIS EKSPRESI *PENINDAS PPI1 LOKUS 1* (*SP1*) DARI KELAPA SAWIT

ABSTRAK

SUPRESSOR PPI1 LOCUS1 (SP1) ialah pengawal selia penting bagi translocon pada membran selaput luar jentera kloroplas (TOC), dan bertanggungjawab dalam pengimportan protein asas ke dalam plastid yang mempengaruhi penjanaan dan fungsi plastid. Kajian terdahulu terhadap *Arabidopsis thaliana*, melaporkan bahawa peningkatan kadar ekspresi *SP1* juga meningkatkan kadar perubahan plastid semasa proses 'de-etiolation' dan penuaan kloroplast. SP1 boleh dianggap sebagai enzim yang menarik untuk dikaji kerana potensinya bagi pembaikan kecekapan kadar perubahan plastid pada tanaman pertanian secara am. Pada kajian yang dijalankan ini, proses penemuan data mengunakan NCBI dan PalmXplore dapat mengenalpasti homolog andaian *SP1* kelapa sawit iaitu *EgSP1* (LOC105060424) yang mengekodkan dua 'isoform' protein: varian X1 dan X2. Homolog SP1 ini mengandungi domain C_3HC_4 yang dipelihara dalam protein SP1 dan menunjukkan identiti jujukan yang tinggi (masing-masing 73.76% dan 73.47%) dengan protein SP1 dari Arabidopsis. Selain itu, analisis pokok filogenetik protein SP1 daripada beberapa tumbuhan terpilih menunjukkan EgSP1 berkelompok bersama tumbuhan monokot yang lain. Seterusnya, jujukan nukleotida untuk *EgSP1* telah berjaya diisolasi daripada cDNA yang terhasil daripada RNA daun. Mengunakan 'primer' khusus, 'amplicon' dengan saiz ~1000 bp telah diperoleh, kemudian diisolasi dan diklonkan ke dalam vektor 'pGEM-T Easy'. Selepas transformasi vektor 'pGEM-T Easy' yang mengandungi *EgSP1* ke dalam sel *E. coli*, tujuh koloni yang positif PCR telah dipilih dan pengekstrakan 'plasmid' turut dilakukan pada sampel tersebut. Daripada analisis penjujukan 'plasmid', didapati hanya varian transkrip X1 telah dapat diklonkan dan tiada 'plasmid' yang mengandungi varian X2. Di samping itu, analisis ekspresi kuantitatif

(qPCR) *EgSP1* pada beberapa tisu vegetatif mendedahkan bahawa transkrip *SP1* (varian X1) diekspresikan pada tahap yang paling tinggi dalam tisu daun, diikuti oleh akar, dengan ekspresi terendah dalam tisu batang. Penemuan daripada kajian ini dapat memberikan dasar untuk aplikasi pertanian masa hadapan daripada gen *SP1* dalam pengembangan penambahbaikan genetik kelapa sawit.

Kata kunci: protein SP1, pengimportan protein plastid, jentera TOC, pembezaan dan biogenesis plastid, kelapa sawit.

ACKNOWLEDGEMENTS

First, I thank Allah the Almighty for all His providence for carrying out this work successfully.

I would like to express my deepest gratitude to my supervisor, Dr. Nur Ardiyana Binti Rejab and also to Dr. Najiah Binti Mohd Sadali for guidance, support, patience, and encouragement through the course of this work.

I am grateful and touched for the attention and support from my father, Mohd Azlan Loh Bin Abdullah and my sisters. Their presence made my life made it slightly more meaningful than it is. I am graciously grateful and indebted to late mother who recently passed, Julainatul Shita Binti Ahmad, though your passing was sudden, your constant love and encouragement will always be remembered. I place on record to dedicate this thesis to caring parents and siblings.

Also, my warm thanks to my fellow lab mates for the help and suggestion, they had provided during my short stay in CEBAR lab that made my progress that much smoother. Finally, my sense of gratitude to one and all, who directly or indirectly have lent their hand in this venture.

TABLE OF CONTENTS

ORIGINAL LITERARY WORK DECLARATION…….……..... ii

ABSTRACT……..... iii

ABSTRAK…….…….…..... v

ACKNOWLEDGEMENTS…….…….…..... vii

TABLE OF CONTENTS……...……..... viii

LIST OF FIGURES…….……..... xii

LIST OF TABLES…….…..... xiv

LIST OF SYMBOLS AND ABBREVIATIONS…….……..... xv

LIST OF APPENDICES……...……..... xvii

CHAPTER 1: INTRODUCTION…….…..... 1

1.1 Problem Statement..…….... 2

1.2 Objectives…….. 3

CHAPTER 2: LITERATURE REVIEW……..... 4

2.1 African Oil Palm, *Elaeis guineensis*……...……..….... 4

2.1.1 History…….. 4

2.1.2 Oil Palm ……..……. 5

2.1.3 Production of Palm Oil……. 6

2.2 Plastids…….…….…….... 8

2.2.1 Plastid Biogenesis……...……..… 9

2.2.2 The Role of SP1 in Model Plants…….... 12

2.3 SP1 Influence on Plastid Differentiation……..….... 14

2.4 Biotechnological Improvements in Oil palm……..……...……..……... 16

2.4.1 SP1 Utilization Prospect in Oil Palm…….. 18

CHAPTER 3: METHODOLOGY……..……. 19

3.1 Identification of SP1 Homolog from Oil Palm ……..…….……. 19

3.1.1 Identification of Putative Oil Palm SP1 Homolog.…...……. 19

3.1.2 Phylogenetic Analysis……..……...……... 20

3.2 Isolation of Oil Palm *SP1* (*EgSP1*) Coding Sequence (CDS)…….. 20

3.2.1 *EgSP1* Primer Design……..…….……... 20

3.2.2 Extraction of Total RNA……..…….……... 21

3.2.3 Quantity and Quality Assessment of Total RNA…….…...……..... 23

3.2.4 cDNA Synthesis and Confirmation by PCR…….……... 23

3.2.5 Isolation of *EgSP1* Coding Sequence…… 25

3.2.6 PCR Product Purification…..……..…..…….…….……... 26

3.2.7 3’ Poly-A Overhang Addition……...……...……... 26

3.3 Gene Cloning…..……...……..…….……... 27

3.3.1 Preparation of DH5α Competent Cell and Glycerol Stock…….…... 27

3.3.2 Ligation and Transformation into *E. coli*……...…... 28

3.3.2.1 Colony PCR…….…..... 29

3.3.3 Plasmid Isolation from Positive Colonies…….……..... 30

3.3.3.1 Sequencing and Analysis of Isolated Plasmid……...……. 31

3.4 Expression Pattern Analysis of *SP1* Transcript in Different Tissues of Oil Palm…… 32

3.4.1 RNA Extraction and cDNA Synthesis……..……..…... 32

3.4.2 Primer Design for Quantitative PCR (qPCR)...…...……...…... 32

3.4.3 Primer Efficiency Test…… 33

3.4.4 Expression Pattern Analysis of *EgSP1*_X1.…….…….....…... 34

CHAPTER 4: RESULTS…….. 36

4.1 Identification of SP1 Homolog from Oil Palm……..... 36

4.1.1 Identification of Putative Oil Palm SP1 Homolog……...……...…... 36

4.1.2 Phylogenetic Analysis……..…..……...……..……..... 40

4.2 Isolation of Oil palm SP1 (*EgSP1*) Coding Sequence…....……..... 41

4.2.1 *EgSP1* Primer Design……...……..…..……...……...……... 41

4.2.2 Extraction of Total RNA……....…..……...……...……..... 43

4.2.3 Successful cDNA Synthesis Confirmation by PCR..…….....…….. 44

4.2.4 Isolation of *EgSP1* Coding Sequence.....…..…..……..…….... 45

4.3 Cloning of *EgSP1* Coding Sequence……..……...…….... 47

4.3.1 *E. coli* Transformation and Colony PCR…..……...…...……..... 47

4.3.2 Plasmid Isolation and Sequencing…..……...…….…..... 49

4.4 Expression Pattern Analysis of *EgSP1* Transcript in Different

Oil Palm Tissues……..... 53

4.4.1 Primer Design for Quantitative PCR (qPCR)……..…… 53

4.4.2 Primer Efficiency Test…… 56

4.4.3 Expression Pattern Analysis of *EgSP1*_X1…… 60

CHAPTER 5: DISCUSSION……..... 62

5.1 Identification of SP1 Homolog from Oil Palm…...……... 63

5.2 Isolation and Sequencing of *EgSP1* Coding Sequence……..……... 64

5.3 Expression Pattern Analysis of *EgSP1* Transcript in Different

Oil Palm Tissues……..……..... 66

5.4 Challenges and Prospect of Oil Palm Biological Improvements.……..……... 68

CHAPTER 6: CONCLUSION AND FUTURE WORK……...……..……. 71

6.1 Conclusion……..…...……..... 71

6.2 Future Work……..…...……..... 71

REFERENCES…….……..... 73

APPENDIX……..... 79

LIST OF FIGURES

Figure 2.1	:	African oil palm (*Elaeis guineensis* Jacq.) and scientific classification. Adapted from (Mohammed et al., 2011) …..	6
Figure 2.2	:	An overview of a plastid biogenesis and differentiation pathway. Adapted from (Jarvis & López-Juez, 2013) …..…	10
Figure 2.3	:	An overview of the TOC Complex GTPase receptor. Adapted from (Agne & Kessler, 2009)……..	11
Figure 2.4	:	An overview of the TOC machinery reorganization during de-etiolation. Adapted from (Kessler, 2012) …….	13
Figure 4.1	:	Blastp search result from NCBI database in *Elaeis guineensis* using AtSP1 (At1g63900) amino acid sequence as query. ……...	37
Figure 4.2	:	Blastp search result from PalmXplore database search using AtSP1 (At1g63900) amino acid sequence as query ……..	37
Figure 4.3	:	Schematic representation of the Intron-Exon Boundary in *EgSP1* gene and the *EgSP1* X1 and *EgSp1* X2 transcript variants…….....	38
Figure 4.4	:	Amino acid sequence alignments of predicted EgSP1 protein with SP1 from other selected plant species and outgroup (MULAN)…….....……...	39
Figure 4.5	:	Phylogenetic tree of predicted SP1 amino acid sequence, constructed with Neighbour-Joining method…….....	40
Figure 4.6	:	PCR primer stats for the designed forward and reverse *SP1* primer pairs…….……..	42
Figure 4.7	:	Gel electrophoresis of total RNA extracted from various tissue types of oil palm……..	44
Figure 4.8	:	Gel electrophoresis of PCR product with housekeeping, *Cyp2* primers for the newly synthesized cDNA from various oil palm tissues……..	45
Figure 4.9	:	Gel electrophoresis of Q5 High-Fidelity Taq Polymerase PCR product for isolated *SP1* CDS……..…..	46
Figure 4.10	:	Gel electrophoresis for the first and second PCR purification steps…….....	47
Figure 4.11	:	Blue-white screening of transformed E. coli (*DH5α*) and colony master plate of positive colonies……	48

Figure 4.12 : Detection of the presence of *EgSP1* in the plasmid of transformed *E. coli* by colony PCR. ……..... 49

Figure 4.13 : Nucleotide sequence alignment of isolated plasmid insert sequences against *EgSP1* transcript variant sequence X1 and X2…….... 52

Figure 4.14 : Amino acid sequence alignment of translated plasmid sample insert sequences against oil palm SP1 isoform X1 and X2……... 53

Figure 4.15 : PCR primer stats for the designed Real-Time PCR, forward and reverse *EgqSP1* primer pairs…….. 55

Figure 4.16 : Standard curve C_t versus ln (cDNA) of primer pairs *Cyp2* and *EgqSP1*……... 57

Figure 4.17 : Real-Time PCR melt curve analysis of *Cyp2* primer pair…….…… 58

Figure 4.18 : Real-Time PCR melt curve analysis for *EgqSP1* primer pair……..…….. 59

Figure 4.19 : Real-Time PCR of *EgSP1* transcripts (variant X1) in leaf, stem, and root tissue of oil palm…… 60

Figure 4.20 : Melt curve analysis of primer pairs, *Cyp2* and *EgqSP1* for housekeeping gene and *SP1* transcript variant X1 respectively…… 61

LIST OF TABLES

Table 4.1	:	Details on the identified oil palm SP1 isoforms……..	38
Table 4.2	:	Product details of designed primer set…….	41
Table 4.3	:	Assessment of total RNA purity and concentration using Nanodrop…….	43
Table 4.4	:	Nanodrop analysis of first and second PCR purification steps…….	47
Table 4.5	:	Nanodrop analysis of isolated plasmid……...	51
Table 4.6	:	Product details of designed primer pair……..	54

LIST OF SYMBOLS AND ABBREVIATIONS

μl	:	Microliters
ng	:	Nanograms
^oC	:	Degree Celsius
A	:	Ampere
V	:	Voltage
t	:	Metric tonne
g	:	Gram
ha	:	Hectare
ml	:	Millilitres
mM	:	Millimolar
s	:	Seconds
U	:	Enzyme unit
Aa	:	Amino acids
nt	:	Nucleotides
bp	:	Base pairs
T_m	:	Melting temperature
AGE	:	Agarose gel electrophoresis
BLAST	:	Basic Local Alignment Search Tool
CDS	:	Coding Sequence
cDNA	:	Complementary DNA
CRIPSR/Cas9	:	Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR-associated protein 9
C_t value	:	Cycle Threshold value
Cyp2	:	Cyclophilin 2
DEPC	:	Diethyl pyrocarbonate

DNA	:	Deoxyribonucleic acid
DGAT		Diacylglycerol Acyltransferase
dNTP	:	Deoxynucleotide triphosphate
GDP	:	Gross domestic product
GMO	:	Genetically modified organism
LB	:	Lysogeny broth
MEGA-X	:	Molecular Evolutionary Genetic Analysis (Version X)
M-MuLV	:	Moloney Murine Leukemia Virus
MPOB	:	Malaysia Palm Oil Board
MULAN	:	Human Mitochondrial Ubiquitin Ligase Activator of NFKB 1
NCBI	:	National Center for Biotechnology Information
NTC	:	No template control
ORF	:	Open reading frame
PCR	:	Polymerase chain reaction
ppi1	:	Plastid protein import 1 mutant
qPCR	:	Quantitative polymerase chain reaction
RCF	:	Relative centrifugal force
RING	:	Really Interesting New Gene
RNA	:	Ribonucleic acid
RPM	:	Revolutions per minute
SP1	:	Suppressor of PPI1 Locus 1
SNPs	:	Single nucleotide polymorphisms
TBE	:	Tris-Borate-EDTA
TIC	:	Translocon at the inner chloroplast membrane
TOC	:	Translocon at the outer chloroplast membrane
UPS	:	Ubiquitin-proteasome system

LIST OF APPENDICES

Appendix A	:	Sequence alignment of EgSP1 isoform X1 from NCBI and PalmXplore…….....	79
Appendix B	:	Oil palm *SP1* transcript variant X1 sequence……	80
Appendix C	:	Table list of identified organism containing putative SP1 from NCBI database……..	81
Appendix D	:	Map of pGem-T Easy Vector……...…….	82
Appendix E	:	Electropherogram of sequenced plasmid……..	83
Appendix F	:	Amplification plot of primer efficiency test (*Cyp2*)……..	125
Appendix G	:	Amplification plot of primer efficiency test (*EgSP1*).....	126
Appendix H	:	Amplification plot of expression analysis (*SP1* transcript variant X1)……	127

CHAPTER 1: INTRODUCTION

1.0 Introduction

The African oil palm (*Elaeis guineensis* Jacq.) is one of the world's most productive oil-bearing crop and is accounted for more than 34% of the world's total vegetable oil production (Kushairi et al., 2018). Palm oil that are extracted from oil palm fruits are known to have a broad range of applications for high demand products in both in the food and chemical industries. Unlike most notable oil-bearing crops grown around the world, oil palm plant does not require an annual sowing due to its nature as a perennial plant. This translates to the crop's all-year-round productivity which is reflected on the plants annual global edible oil production rate of around 72.3 million tonnes per year of oil palm grown, significantly higher to that of other oil-bearing crops such as rapeseed, soybean, and sunflower (Murphy et al., 2021; Tang & Al Qahtani, 2019). An estimated of over eighty-five percent of the world palm oil productions are concentrated in the tropical regions of Malaysia and Indonesia as large-scale plantations (Ghulam et al., 2020; Murphy et al., 2021).

In Malaysia, the oil palm industry is of high commercial value as it contributes to 43.1% of the Gross Domestic Product (GDP) generated from the agricultural sector (Tang & Al Qahtani, 2019). In 2019, it is estimated that 5.9 million hectares of land was allocated for the oil palm plantations alone, making it the most commercially important crop to the country's agro-industry (Ghulam et al., 2020). With the expected rise in the world's demand for palm oil, as well as the increasing scrutiny on the sustainability of oil palm land expansions. Many methods are being explored for the improvements of oil palm yield which includes the use of new crop growing technique as well as the possible application of genetic modification (Ghulam et al., 2020; Tang & Al Qahtani, 2019).

SUPPRESSOR OF PPI1 LOCUS 1 (SP1) encodes for the E3 ligase enzyme which mediates the ubiquitination of the translocon at the outer envelope membrane of chloroplasts (TOC) complex machinery for degradation, which regulates the uptake of plastid precursor proteins that are needed for plastid differentiation and biogenesis (Ling et al., 2012). The regulatory functions of this enzyme are crucial for response to photosynthetic developments as well as stress conditions. The loss of SP1 function in plants delayed plastid differentiation during both de-etiolation and senescence, as well as reduced tolerance to salinity, osmotic and oxidative stresses. In contrast to that, SP1 overexpressed plants promoted de-etiolation, senescence, and abiotic stress tolerance. These characteristics perceive SP1 among the possible candidates for crop improvement strategies (Ling & Jarvis 2015a; Ling & Jarvis, 2015b).

This study aims to isolate and perform molecular characterization of *SP1* gene to understand plastid biogenesis in oil palms. This was achieved by the isolation of the coding sequence (CDS) of *SP1* derived from RNA of oil palm issue and cloning them using a bacterial system. Next, the validation steps were performed by DNA sequencing and BLAST analysis. Findings of this study may uncover the possible agriculture application of *SP1* gene and could aid in the formulation of oil palm genetic improvements.

1.1 Problem statement

SP1 is known to be an important enzyme for the regulation TOC complex machinery, that regulates the uptake of plastid precursor proteins that are needed for the differentiation and biogenesis of plastids (Ling et al., 2012). The regulatory functions of this enzyme are crucial for response to photosynthetic developments as well as stress conditions. This will ultimately influence crop yield and biomass gain. Although we

already have a better understanding of SP1 role in the model plant *Arabidopsis thaliana*, *SP1* gene has yet to be identified and characterized in oil palm, which limits its accessibility for further study including its functional analysis and its possible application in the agricultural context. Identification and characterization of the gene in oil palm is an important first step towards understanding the role of this gene in the regulation of the plant's plastid biogenesis and differentiation. Findings from this research may facilitate the future modification of this crop that can enhance productivity.

1.2 Objectives

In general, this study aims to isolate and perform molecular characterization of *SP1* coding sequence (CDS) to understand plastid biogenesis in oil palms.

The specific objectives of the research were:

1. to identify the *SP1* homolog gene(s) from oil palm sequence database (NCBI and PalmXplore);
2. to isolate and clone the coding sequence of *SP1* homolog from oil palm;
3. to evaluate the expression pattern of *SP1* transcript in different vegetative tissues (leaf, stem, and root) of oil palm;

CHAPTER 2: LITERATURE REVIEW

2.1 African oil palm, *Elaeis guineensis*

2.1.1 History

The oil palm *Elaeis guineensis* is also known as the African oil palm is one of two accepted species of the *Elaeis* genus and was brought into Malaysia initially as an ornamental plant in the early 1870s. The plant's origin was noted to be from the West African region and was suggested to be cultivated and used by human for thousands of years prior. The plant dispersion was largely contributed by human activities as the plant are seen around human settlements and walkways as semi-domesticated palm. In the current age, the oil palm is largely grown for its flexible application in the food and chemical industries, and currently is one of the largest sources of industrially used vegetable oil (Corley & Tinker, 2015; Alam et al., 2015).

The first commercial oil palm plantation establishment in Malaysia was founded at the Tennamaram estate of the Selangor state of Malaya around the year 1917 by a Frenchman, Henri Fauconnier. Since then, this crop has been successfully grown commercially in Malaysia, which lead to a source of income for many landowners and industries. Oil palms grown in Malaysia are mainly of the *tenera* fruit type, which was commercially adopted in the 1960s, which replaced the *dura* fruit type due to the prospect of increase oil yield of up to 30%. The term fruit type is to describe the internal appearance of the palm fruit. The *dura* has the hard-shelled type, while *tenera* has the thinnest shelled, with respect to the difference in the mesocarp percentage between the two fruit types (Corley & Tinker, 2015).

2.1.2 Oil palm

The African oil palm was first formally introduced into the botanical classification in 1763 by Nicholas Joseph Jacquin. *Elaeis guineensis* is described to be from the genus *Elaeis*, from the palm family *Areacaceae*, and clade under the monocotyledonous angiosperm (Figure 2.1) (Barcelos et al., 2015). These perennial plants are known to have an expected productive lifetime of around 21-23 year when grown on plantations and are known to reach the height of over 15 meters (Woittiez et al., 2017; Barcelos et al., 2015). Oil palms are monoecies which are known to produce both male and female inflorescences but are unisexual at alternating cycle, depending on both genetic and environmental factors. The pollination of the said inflorescences is aided by vector insects specifically weevils (*Elaeidobius kamerunicus*), which feeds on the male flower and inflorescence tissue (Corley & Thinker, 2015). Post-pollination, the fruit would mature into sessile droops within 5-6 months where they can be harvested for palm oil and kernel oil from the mesocarp and kernel, respectively (Corley & Thinker, 2015; Barcelos et al, 2015).

Scientific Classification

Kingdom: *Plantae*

Family: *Arecaceae*

Subfamily: *Arecoideae*

Tribe: *Cocoeae*

Genus: *Elaeis*

Figure 2.1: African oil palm (*Elaeis guineensis* Jacq.) and scientific classification. A**dapted from (Mohammed et al., 2011).**

2.1.3 Production of palm oil

Globally, palm oil contributes to around one-third of the global edible oil supply, with the span of over 19.4 million ha, this crop is grown mostly in Indonesia and Malaysia as well as other countries close to the equator such as Colombia and Nigeria (Kushairi et al., 2018: Murphy et al., 2021). These native African plants are grown as tropical cash crops due its high profitability, easy establishment, and not to mention the plants tolerance for growing on a range of soil type and even in less fertile soils (Corley & Thinker, 2015; Dislich et al., 2016). Over the years, Malaysia has been one of two leading suppliers of the world palm oil (Ghulam et al., 2020), where a large base of experience has been established in the country within the overall production of palm oil, with upstream and downstream industries such as plant breeding, tissue culture, oil fractionation and

oleochemicals that are depended on this industry (Kushairi et al., 2017). As of 2019, Malaysia still stands as the world second largest supplier of palm oil, supplying an estimate of over two-fifth of the world palm oil demand (Ghulam et al., 2020; Shevade & Loboda, 2019).

The oil palm is known to be a highly productive crop that is attributed to its nature as an equatorial plant with a 12-month growing season and its efficient oil-accumulating trait. Compared to other major crops such as soybean, rapeseed, and sunflower, the oil palm shows a significantly higher annual average oil yield of around 3.72 t/ha/year as compared to 0.4 t/ha/year for soybean, 0.72 t/ha/year for rapeseed, and 0.55 t/ha/year for sunflower respectively (Tang & Al Qahtani, 2019). This, in turn, mitigates the cost of land and operation cost per unit of production.

With the continual increase of the world population, it is estimated that the demand for palm oil for edible uses would increase to at least 93 million tonne as by the year 2050 (Murphy et al., 2021). This growth when coupled with the increasing industrial demand secures palm oil as an important source of vegetable oil for as long as the world population continue to rise. The palm oil industry is also a significant contributor to the Malaysian economy and global trade. In 2019, the industry was estimated to contribute around RM 64.84 billion to the Malaysian economy with an export volume of 16.88 million tonnes, due to the strong demand from importing countries such as India, China, and the European Union (Ghulam et al., 2020).

2.2 Plastids

Plastids are doubled-membrane organelles containing non-nuclear DNA and provide plant cells their defining features, such as the ability to perform photosynthesis, roots gravitropism and fruit ripening, based on the differentiation pathways the plastid is directed towards. Much of the plastid's ancestral genomic DNA are entrusted to the genome of the host cell, and plastids engage in retrograde signalling with the host nucleus for the coordination of the expression of the nuclear and plastome genes (Watson et al., 2018; Hirosawa et al., 2017).

There are many types of plastids, where these organelles often characterized by their colour, morphology, and ultrastructure. Where proplastids are known for their role as precursors to other more differentiated plastids such as the chloroplast and are observed under the microscope as small, colourless and undifferentiated plastids (Sadali et al., 2019; Jarvis & López-Juez, 2013). Chloroplasts are one of the most well studied plastids and are essential for photosynthetic organisms. These plastids are well recognized for their accumulation of chlorophyll and large grana ultrastructure which are important components for photosynthesis reactions. The second type of plastid is etioplast that serves as an intermediate between proplastids and chloroplast differentiation pathways. This plastid is often found in seeds grown in darkness and are known to rapidly differentiate into chloroplast upon exposure to light (Plöscher et al., 2011; Smolikova & Medvedev, 2016). Unlike elaioplast which differentiate into chloroplast, gerontoplast are chloroplast derived plastids and are often found in senescence plant leaves and are said to be important for the recycling of plastids. They are characterized for the breakdown chlorophyll as well as the thylakoid membrane that normally found in the chloroplast (Jarvis & López-Juez, 2013). Chromoplast is another type of plastids noted for providing colour to fruits and flowers, where these plastids accumulate a large amount of carotenoids which are stored in their distinctively large plastoglobuli (Sadali et al., 2019).

Another group of plastids are termed as leucoplast, which is derived from the Greek word for white, as appropriately describes how leucoplast appears as white plastids or colourless under the microscope (Solymosi et al., 2018). Generally, leucoplast morphologies are described to have a minimal amount of thylakoid, a lesser developed inner membrane system to that of chloroplast and the presence of a large storage compartment (Markus & Turner, 2013). The term leucoplast can be associated for plastid such as elaioplast, amyloplast, and proteinoplast which are characterized for their synthesis and storage of lipids, starch, and protein respectively (Zhu et al 2018; Vigil et al., 1985; Sadali et al., 2019).

2.2.1 Plastid biogenesis

Plastids have the ability to interconvert between distinct plastid types (amyloplast, chloroplast, chromoplast, elaioplast, etioplast, gerontoplast, and proplastid) according to the plant's developmental cues. As shown in Figure 2.2, proplastid which are undifferentiated plastids are able to differentiate into all other plastid types and served as the progenitor to most plastid types. Differentiation pattern such as etioplast to chloroplast, chloroplast to gerontoplast and chloroplast to chromoplast are often seen during seedling de-etiolation, leaf senescence, fruit ripening (Sadali et al., 2018; Jarvis & López-Juez, 2013; Hayashi & Nishimura, 2009). Other lesser seen differentiation pathways are observed in some root crops such as carrots where amyloplast are known to differentiate into chromoplast (Oleszkiewicz et al., 2018).

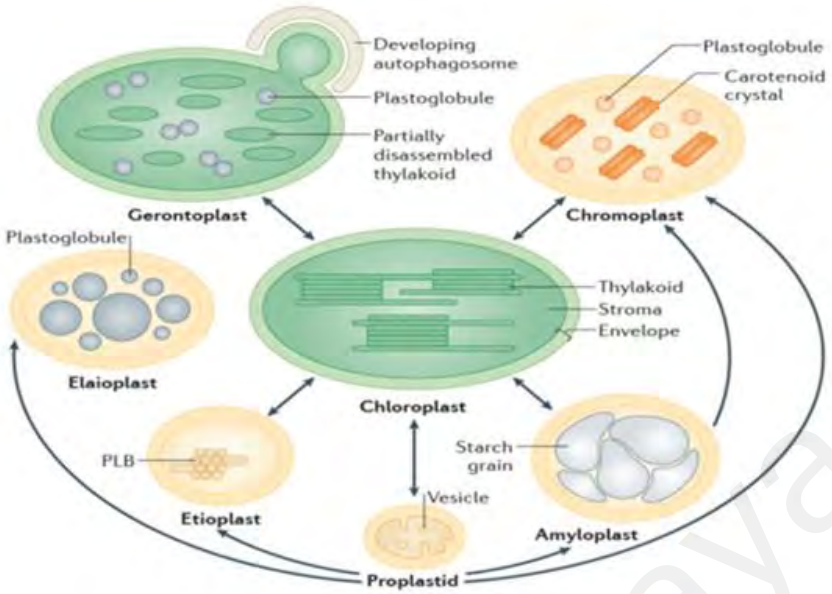


Figure 2.2: An overview of a plastid biogenesis and differentiation pathway. Adapted from (Jarvis & López-Juez, 2013).

Proplastid are undifferentiated plastids that serve as a progenitor to all plastid types. Proplastid may further differentiate into amyloplast, chloroplast, chromoplast, etioplast, or elaioplast. Some differentiated plastids may also undergo interconvertible differentiation between other plastid types such as from etioplast to chloroplast and amyloplast to chromoplast. Plastid differentiation is regulated by both environmental cues as well as retrograde signalling between the organelle and the host nucleus.

Plastid differentiation and regulation are highly dependent on the regulation of the nucleus-encoded proteins (precursors protein) brought into the plastid, where the uptake of these precursor protein may cause changes in the plastid proteome composition which effects the plastids functions and development (Richardson & Schnell, 2020). The translocon at the outer envelope membrane of chloroplasts (TOC) complex machinery found on the outer membrane of the chloroplast envelope mediates the early protein import processes and is one of the checkpoints for the regulation of these precursor proteins (Sadali et al., 2019; Jarvis & López-Juez, 2013).

The major pathway of precursor protein import into the plastids is mediated by TOC and TIC (translocon at the inner envelope membrane of chloroplasts) multiprotein complex. The TOC complex machinery is of more interest, as it functions prior the TIC

complex and represents the first step in the control of nucleus-encoded precursor protein access to the plastid's inner envelope membrane. The TOC complex is comprised of three main proteins units, which are Toc75, a β-barrel membrane channel, Toc159 and Toc33 which are the TOC membrane-bound GTPase primary receptors for the nucleus-encoded precursor proteins. All precursor proteins are initially bound to the Toc159 and Toc33 receptors before undergoing translocation through the Toc75 β-barrel membrane channel into the inner membrane (Li & Chiu, 2010). The TOC GTPases are known to be specific in the recognition of precursor protein as well as being discriminative toward different types of the precursor protein. The assembly of different Toc33 and Toc159 isoforms allows for the generation of TOC complexes for distinct selectivity towards the precursor proteins as described in Figure 2.3. This is useful for allowing entry of precursor proteins of specific classes (photosynthetic proteins or housekeeping proteins) thus providing additional control to the plastid's differentiation and biogenesis pathways (Kessler & Schnell, 2009; Chu & Li, 2018).

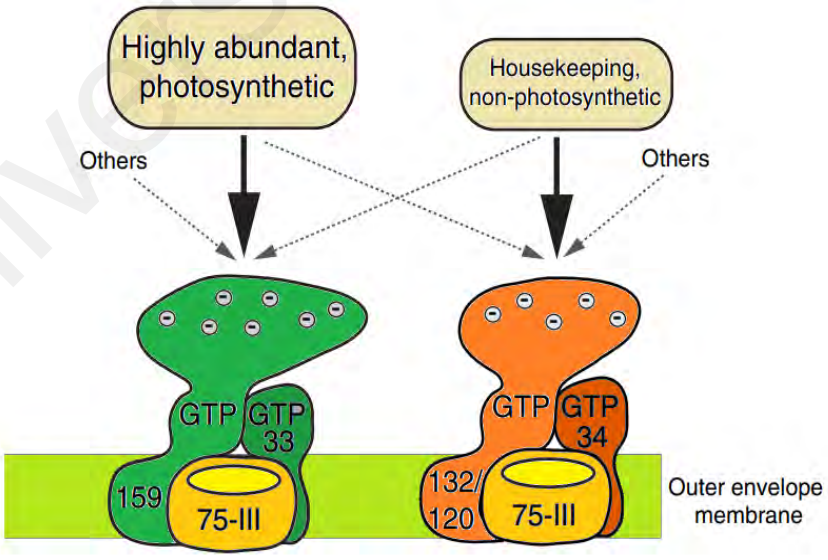


Figure 2.3: An overview of the TOC complex GTPase receptor. Adapted from (Agne & Kessler, 2009)

Depending on the developmental cues directing the plastid's differentiation pathways, the reassembly of the TOC machinery allows for the preferential imports of precursor protein based on the TOC GTPase isoform. The assembly of the TOC complex with GTPase isoform atToc159 and atToc33 allows for the preferential imports of photosynthetic preproteins, whereas the assembly of atToc132/120 and atToc34 show preference toward the imports of non-photosynthetic preproteins.

2.2.2 The role of SP1 in model plant

Development of plastids relies on the proteins imported from cytosol which is regulated by TOC and TIC machineries as previously described. One of the proteins identified to mediate the activity of TOC is SP1 protein (Ling et al., 2012). SP1 was identified by Ling et al., (2012) during a screening for a second site suppressor of the *ppi1* mutant (atToc33 mutants) in Arabidopsis to identify novel regulators of plastid protein import. The *ppi1* mutant had reduced ability to import photosynthetic proteins hence displayed a pale-yellow phenotype due to the reduced efficiency in chloroplast development. SP1 was noted to be a suppressor to the *ppi1* mutant, as the *sp1 ppi1* double mutant resulted in the recovery of chloroplast development as well as an increase in protein imports in the mutant plant (Ling et al., 2012; Broad et al., 2016).

SP1 gene encodes for a RING-finger (RNF) E3 ligase that is located on the outer envelope membrane of the chloroplast and is anchored by 2 transmembrane domains while presenting the C-terminal RING-finger domain to the cytosol. The C-terminal RING-finger domain enables the enzyme access to the cytosolic Ubiquitin–Proteasome System (UPS) component and interaction with their substrate as illustrated in Figure 2.4 (Ling & Jarvis, 2015a: Kessler, 2009). UPS is a major proteolytic system utilized by the eukaryotic organism for the removal of precursor proteins in the cytosol (Lee et al., 2013; Lee et al., 2009). Plastids organelles present within the plant cell also utilize the UPS for the regulation of their surface protein translocation machinery (Ling & Jarvis, 2015a). The proteolysis of a targeted protein on the plastid surface is generally initiated by the attachment of ubiquitin on the target protein, where the attachment process is sequentially mediated by 3 UPS enzymes: E1 activase, E2 conjugate and E3 ligase. E1 activase is an ubiquitin activating enzyme that binds to and activates ubiquitin which is then attached to an E2 conjugating enzyme. E3 ubiquitin ligase acts to confer substrate specificity and to catalyse the transfer of ubiquitin from the E2 enzyme to a targeted protein. The

successive addition of the ubiquitin monomers to the target generates a polyubiquitin which is then targeted for degradation by 26s proteasome (26SP) (Vierstra, 2009; Broad et al., 2016).

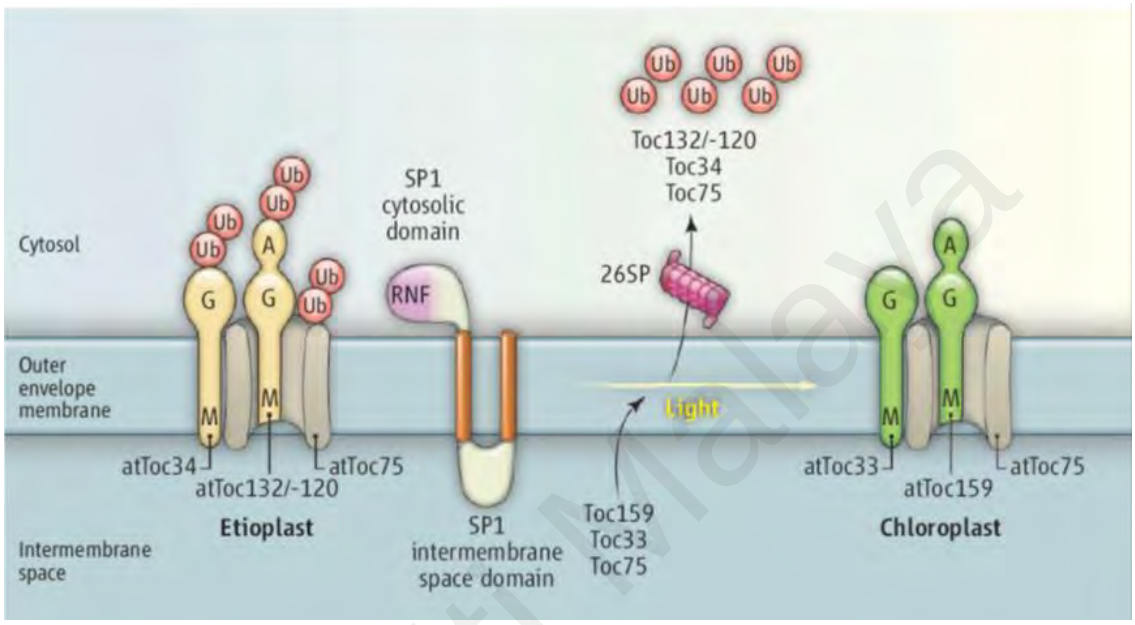


Figure 2.4: An overview of the TOC machinery reorganization during de-etiolation. Adapted from (Kessler, 2012).

Prior to de-etiolation, the etioplast TOC machinery on the outer plastid membrane consists of the Toc75 β-barrel membrane channel as well as GTPase-receptors Toc132/120 and Toc34 which exhibit selectivity toward imports of housekeeping precursor proteins onto the plastid. Environmental cues such as light triggers the retrograde signalling between the plastid and nucleus for the differentiation of etioplast into chloroplast. SP1 E3 ligase is recruited for the remodelling of the TOC machinery and does this by ubiquitinating the TOC machinery components Toc75, Toc34 and Toc 132/120, which marks the proteins for degradation by the 26S Proteasome. This in turn then allows the TOC machinery to reorganize with TOC GTPase-receptors Toc33 and Toc159, which enable the imports of photosynthetic precursor proteins needed for the plastid transition.

SP1 is one of the known E3 ligases which are responsible for target protein identification and the regulation of plastid translocation machinery. The substrate of the SP1 E3 ligase is the TOC proteins, where SP1 catalyse the ubiquitylation of the TOC proteins and is subsequently degrade by 26SP (Sadali et al., 2019; Jarvis & López-Juez, 2013). As a RING-finger E3 ligase, SP1 is also known to perform auto-ubiquitination which acts as a way to self-regulate its own abundance (Guerra & Callis, 2012; Broad et al., 2016). The targeting of the TOC machinery by SP1 E3 ligase enables the

reorganization of the TOC machinery, thereby influencing precursor protein access into the organelle, and ultimately the developmental and differentiation of the plastid.

2.3 SP1 influence on plastid differentiation

Etioplast is a progenitor to chloroplast that develops in tissues that are grown in darkness or are light deprived. Etiolated leaves are known to be yellow due to the presence of carotenoid coupled with the lack of chlorophyll pigments. Chloroplast biosynthesis under normal condition is generally from the differentiation from proplastid without the etioplast stage (Solymosi & Aronsson, 2013). In *A. thaliana*, atToc132 and atToc120, TOC protein that are suggested to be specific toward housekeeping proteins, are predominantly present in tissue that are less exposed to light such as roots (Demarsy et al., 2014). On the other hand, atToc159 and atToc33 are known to be highly expressed in the juvenile development stages where high amount of photosynthetic protein imports is required (Bauer et al., 2000; Sjuts et al., 2017).

De-etiolation is a process in which etioplast undergoes differentiation from etioplast to chloroplast due to the exposure to light which results in the bulk import of photosynthesis associated proteins. This leads to a major proteome change with the increase in chlorophyll synthesis and development of the thylakoids within the plastid (Plöscher et al., 2011; Broad et al., 2016). SP1 is shown to be an important moderator in promoting plastid differentiation by selectively ubiquitinating the TOC receptor isoforms (Toc33 and Toc159; Toc34 and Toc132/120) on the chloroplast outer membrane to allow complex machinery rearrangement to better accommodate precursor proteins that are needed during plastid transitions (Ling et al., 2012). During de-etiolation, the ratio changes between the imports of housekeeping and photosynthetic precursor proteins lead to the rapid transition of the plastid into chloroplasts. An illustrative overview on the

process is described in Figure 2.4. Likewise, during leaf senescence process, rearrangement of the TOC complex mediated by SP1 caused a reduction in the imports of photosynthetic precursor protein followed by the transition of chloroplasts to gerontoplast (Ling et al., 2015a).

Chromoplasts are plastids that are specialized in the synthesis as well accumulation of carotenoid pigments. These pigments often function in aiding with plants in their dispersal of seeds and pollen by attracting vectors for their transmission (Broad et al., 2016). Chromoplast are also one of the few extensively researched plastids due to their importance in fruit ripening and quality. During fruit ripening, chloroplast is known to differentiate into chromoplast with the most apparent changes seen in chloroplast are the disruption of the thylakoids, degradation of chlorophyll and the upregulation of carotenoid biosynthesis which involves the reorganization of protein import machinery to accommodate the changes that required for the plastid differentiation (Sadali et al., 2019). Moreover, knockdown of the plastid ubiquitin E3 ligase SP1, delays tomato fruit ripening, as judged by green-to-red colour changes, whereas overexpression of SP1 accelerates ripening (Sadali, 2018). These data highlighted the critical role of SP1 in chromoplasts differentiation during fruit ripening and provide a theoretical basis for engineering crop improvements.

Elaioplast is a type of plastid involving in the biosynthesis of fatty acids in oil seeds and terpenes in citrus fruits as well as their role in the storage of the said compounds. In a study on citrus fruit elaioplast proteome by Zhu et al. (2018), members of the TOC machinery including Toc33, Toc75 and Toc159 were identified in the proteomic analysis. The presence of the TOC machinery in the elaioplast proteome composition indicates the elaioplast developmental process requires imports of precursor proteins through the TOC components into the plastid. Considering the close association of SP1 with TOC

components, it is highly possible that SP1 is involved in the import machinery during elaioplast development of an important oil-producing crop such as oil palm.

2.4 Biotechnological improvements of oil palm

Biotechnology application has long been present in the palm oil industry since the first application of tissue culture for clonal propagation of oil palm. Since then, the technology has expanded in the industry with the identification of molecular markers, understanding of gene expression and utilization of the oil palm genome (Low et al., 2017). One of the significant applications of molecular markers in the palm oil industry begun with the development of a diagnostic assay for the *SHELL* gene (Singh et al., 2013: Low et al., 2017). The identification of the *SHELL* gene enables the development of an assay for the screening of non-*tenera* plants before they are planted in the commercial fields. Oil palm with the *tenera* fruit forms is much more desirable with an increased yield of ~30% and ~100% compared to the *dura* and *pisifera* fruit forms, respectively. The diagnostic tool, in turn, improves yield per hectare by reducing the contamination rate of the undesired non-*tenera* plant from the commercial fields (Low et al., 2017).

Genetic engineering of oil palm has long been pursued by MPOB as one of the many efforts to produce palm oil that are higher in percentage of unsaturated fatty acids. This is due to the prospect of penetrating into the growing modern health-conscious food market such as salad oil and oil supplements (Parveez et al., 2015). Several of the proposed objective for genetic manipulation in oil palm includes increasing the productivity and oil yield of each individual plant. This is followed by modification on the oil fatty acid and carotenoid contents to meet the differing demands of the industry. Finally, several efforts were done for the generation of oil palm crops that are more pest

and disease tolerant for maintenance of the crop productivity as well as reducing losses (Murphy, 2014; Parveez et al., 2015).

Genetic manipulation in oil palm has previously been done with the focus on finding ways to produce palm oil with a higher content of oleic and stearic acids. Fatty acid profile of plants often differs from each other despite having the same fatty acid biosynthesis pathways due to the different regulation mechanism involved in their biosynthesis pathway. The understanding of key enzymes in the biosynthesis pathways is the key to manipulating these regulatory mechanisms to produce palm oils with custom traits. Palmitoyl-ACP thioesterase, β-Ketoacyl ACP synthase II (KASII) and stearoyl-ACP desaturase are the 3 target enzymes suggested that could be manipulated to produce palm oil with a higher content of oleic or stearic acids (Parveez et al., 2015).

The latest technology in genetic engineering is the CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein 9) genetic editing tool. It has been used for characterization of gene and traits though the modification of genomic material to produce gene knock out (or knock in) by means of precise and irreversible modification of target base without the introduction of double stand breaks (Yarra et al., 2020; Yin et al., 2017). CRISPR technology has recently gain traction for its application for the improvement of crop trait, because the technology showcase several properties which are advantageous to that of conventional genetic manipulation method. In a recent study performed by Budiani et al. (2018) on oil palm, CRISPR/Cas9 was used for the modification for 2 genes involving the isoflavone reductase and metallothionein for the improvement of oil palm tolerance against *Ganoderma* infections. The study reported the successful transformation of both constructs into the oil palm callus. The use of CRISPR technologies may offer several advantages in terms of increasing public acceptance. This technology is often highlighted for its ability to produce genetically modified crops that are considered non-transgenic, as the cassette DNAs used for gene editing are technically

derived from the same plant (Budiani et al., 2018). As GMO plants are often met with caution by the public, being non-transgenic might improve public acceptance towards the modified crop (Amin et al., 2013).

2.4.1 SP1 utilization prospect in oil palm

Plastid organelles are very diverse in both metabolism and biosynthesis which enable each plastid type to contribute to the synthesis of broad range of metabolites such as fatty acids, starch, pigments, amino acids as well as assimilation of nitrogen and sulphur (Bräutigam and Weber, 2009; Broad et al., 2016). In oil palm, elaioplast and chromoplast are a major concern, as these plastids are involved in the synthesis of the key nutritional component of the palm oil which are the lipid and carotenoid contents in the fruit as well as their role in fruit ripening (Sundram et al., 2003; Sadali et al., 2018).

Previous studies have shown the importance of SP1 in chloroplast biogenesis in Arabidopsis plants (Ling et al., 2012). In recent study in tomato (Ling et al., 2021), manipulation of SP1 has been shown to have effects on fruit ripening and leaf senescence. Moreover, based on data on elaioplast proteome in citrus, the presence of TOC machinery components indicates that, just as chloroplast and chromoplast, other plastid type including elaioplast, might also be dependent on the TOC mediated import of precursor proteins (Zhu et al., 2018). Functional studies of SP1 homolog in oil palm will allow researchers to further assess the extent to which its functions are conserved in different species, and to elucidate possible practical applications in crops.

CHAPTER 3: METHODOLOGY

3.1 Identification of SP1 homolog from oil palm

3.1.1 Identification of putative oil palm SP1 homolog

The homolog of *SP1* gene from oil palm was identified using BLASTp searches on two gene databases including the National Centre for Biotechnology Information (NCBI) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and PalmXplore (http://palmxplore.mpob.gov.my/palmXplore/) (Sanusi et al., 2018). In this data mining procedure, amino acid sequence of *Arabidopsis thaliana* SP1 (AtSP1) (Locus no: At1g63900) was used as a query to identify the possible translational products from the databases. Data set that meets the criteria of having sequence similarity of over 70% with query sequence were shortlisted and the corresponding amino acid sequences were retrieved for further analysis. From the collected data set, protein sequences were further screened by manually eliminating sequences that lack the conserved domain of C_3HC_4-type (RING-finger). Oil palm protein sequence obtained from PalmXplore was compared to the associated sequence obtained from NCBI for cross-checking purposes (Appendix A). Subsequently, gene and transcript (cDNA) sequences of the oil palm *SP1* homolog was retrieved from the NCBI database.

The predicted oil palm SP1 homologs protein sequences (X1 and X2 isoforms) were aligned with SP1 sequences from nine selected plant species (including dicots plants such as Arabidopsis, tomato, soybean, apple, rapeseed, and monocot plants such as banana, sorghum, maize, and wheat) and an outgroup, MULAN (a human mitochondrial ubiquitin ligase activator of NFKB 1) using ClustalW in the MEGA-X (Molecular Evolutionary Genetic Analysis) software (Kumar et al., 2018). Reference sequence used in this analysis are listed in (Appendix C). The alignment data was highlighted for amino acid similarity and C_3HC_4-type (RING-finger) domain were indicated by asterisk.

3.1.2 Phylogenetic analysis

Next, a phylogenetic tree was constructed using the Neighbour-Joining method using the MEGA-X software. In this analysis, amino acid sequence of oil palm SP1 (referred as EgSP1) as well as SP1 putative homologues from other plant species which have homology with oil palm SP1 were retrieved from the NCBI database. *In silico* searches for SP1 homologues from other plant species (26 angiosperms and one lycophyte) were carried out using BLASTp in which AtSP1 was used as query. The sequences were then manually curated for alignment based on the C_3HC_4 type (RING-finger) domain. Twenty-nine amino acid sequences (plant SP1 and outgroup) were aligned using ClustalW (Hall, 2013) with default settings. A phylogenetic tree was constructed with Neighbour-Joining method with 1000 bootstrap replications (Khan, 2017; Horiike, 2016). The reference sequences obtained from NCBI are listed in Appendix C. In this analysis, MULAN was used as an outgroup.

3.2 Isolation of oil palm *SP1* (*EgSP1*) coding sequence (CDS)

3.2.1 *EgSP1* primer design

Due to high similarity (99.4%) between the two oil palm SP1 isoforms (X1 and X2), subsequent primer design and gene cloning were carried out based on the transcript sequence of X1 isoform that encodes longer amino acid sequence. However, it is important to note that the primers designed in this section will not be able to separately amplify X1 and X2 coding sequence.

First, the *SP1* mRNA sequence corresponding to X1 isoform (XM_010944106.3) was subjected to an analysis using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/software) to identify all possible open reading frames present on the mRNA sequence. Based on prior knowledge of the SP1 protein sequence length, the open reading frame covering the appropriate nucleotide length was

identified from the listings. Forward and reverse primers were designed to amplify sequence from the start (ATG) until stop codon (TGA) to flank the coding sequence of the *SP1* mRNA sequence. Designed primers (named as F-*EgSP1* and R-*EgSP1* for forward and reverse, respectively) were evaluated individually via *in silico* analysis using PCR Primer Stat tool in Primer Stat Sequence Manipulation Suite online software (https://www.bioinformatics.org/sms2/pcr_primer_stats.html) (Stothard, 2000). The designed primer sequence with the size of 20-24 bp (5' to 3') were then sent to IDT (Integrated DNA Technologies) to be synthesized.

3.2.2 Extraction of total RNA

For the purpose of gene cloning and expression analysis (described in Section 3.3 and 3.4), total RNA was extracted from *in vitro* oil palm plantlets (~1 month old) obtained from the Malaysian Palm Oil Board (MPOB). Only RNA from leaf tissues was used as material in gene cloning section while RNA extracted from leaves, stem, and roots were used in expression analysis. First, samples were harvested, flash frozen in liquid nitrogen before being stored in -80 ^{o}C to minimize sample degradation. Upon extraction, pre-weighted sample of oil palm tissue that was previously kept at -80 ^{o}C was first submerged in liquid nitrogen and was subsequently transferred into a liquid nitrogen pre-chilled mortar. The frozen sample was then crushed with a pestle into a powdered form with constant addition of liquid nitrogen to prevent the sample from thawing. Subsequently, grounded samples were extracted using a commercial plant RNA extraction kit, Spectrum™ Plant Total RNA Kit (Sigma, United States) based on the manufacturer's instruction with some modifications. The grounded samples in mortar were then added with 500 µl of lysis solution with 5 µl of 2-mercaptoethanol (2-ME) and the mixture was then stirred vigorously at room temperature. When the mixture had finally mixed well, the mixture was then pipetted into a 2 ml microcentrifuge tube. The mixture was then

incubated on a heat block at 56 ^{o}C for 5 minutes to enable the lysis step. Next, the lysate mixture was then centrifuged at 22, 500 RCF for 3 minutes to pellet out the cellular debris formed during the lysis step. While avoiding the pellets, the supernatant was then pipetted into a filtration column attached to a 2 ml microcentrifuge tube provided by the extraction kit. The filtration column was then centrifuged for 1 minute at 22, 500 RCF to filter out the supernatant from any remaining cellular debris present. The filtration column was discarded from the microcentrifuge tube and the clarified lysate was then mixed with 500 µl of binding solution by pipetting the solution for 5 times. Next, 700 µl of clarified lysate/ binding solution mixture was then pipetted into a binding column attached to a fresh 2 ml microcentrifuge tube provided by the kit.

The binding column was then centrifuged for 1 minute at 22, 500 RCF to allow binding for RNA in the mixture to bind to the column. The flow-through was decanted, and the microcentrifuge tube was reattached to the binding column. This step is repeated for the remaining clarified lysate/ binding solution mixture remaining. To begin the on-column DNase digestion step, 300 µl of Wash Solution 1 was pipetted into the binding column. The column was centrifuge for 1 minute, the flow-through decanted, and the microcentrifuge tube returned to the column. Then, 80 µl of DNase I/ DNase digestion buffer mixture was directly pipetted into the centre of the binding column and the binding column was incubated at room temperature for 15 minutes to allow the on-column DNase digestion step. The first column washing step was initiated by pipetting 500 µl of Wash Solution 1 into the binding column. The binding column was then centrifuged for 1 minute at 22, 500 RCF, flow-through decanted and the microcentrifuge tube returned to the column. The second washing step, was performed by pipetting 500 µl of diluted Wash Solution 2 into the binding column, centrifuging the column for 30 s, decanting the flow-through, and returning the microcentrifuge tube to the binding column. This step was repeated for a second time to allow the third washing step of the binding column.

After the final washing step, drying step was performed by centrifuging the binding column for 1 minute at 22,500 RCF followed by transferring the dried binding column into a clean 1.5 ml microcentrifuge tube. Total RNA bounded to the binding column was then eluted out by adding 30 µl of elution buffer into the centre of the dried binding column. The column was then incubated with the elution buffer for 1 minute and was then centrifuged for 1 minute to elute out of the total RNA from the binding column. The extracted RNA was stored -80 ^{o}C until further use.

3.2.3 Quantity and quality assessment of total RNA

Nanodrop was used to measure the concentration and assessment of quality of the extracted RNA samples. Elution buffer was used as blank for the measurement of the absorbance ratio of A_{260}/A_{280} and A_{260}/A_{230}. In addition, quality of the extracted RNA was analysed using agarose gel electrophoresis (AGE). A 1% (w/v) agarose gel was prepared by dissolving 0.2 g of agarose with TBE (Tris-borate EDTA) buffer solution (pH 8) with GelRed® (USA) as the nucleic acid gel stain. Prior to well loading, the concentration of the RNA samples was adjusted to 40 ng/µl by adding DEPC treated water and then mixed with 1 µl of 2X loading dye (Biolabs). A total volume of 4 µl of the diluted RNA sample were then run on the 1% (w/v) agarose at 100 V, 400 A for 30 minutes. Samples were run with Sigma transcript RNA marker (0.2k-10k nt) (Sigma, USA). The bands were then visualized using a gel documentation system, from AlphaImager HP imaging system (Proteinsimple).

3.2.4 cDNA synthesis and confirmation by PCR

Reverse transcription of total RNA to cDNA strands was performed on all total RNA obtained from Section 3.2.2 along with a negative control (water) via Viva cDNA synthesis kit (Vivantis, Malaysia) according to the manufacturer protocol. Subsequent

mixing steps were performed on ice. First, 1 µg of oil palm total RNA (volume varied and adjusted with RNase free water) was pipetted into a PCR tube to act as the template for the cDNA conversion. Then, 1 µl of 40 µm oligo dT_{18} and 1 µl 10 mM dNTP mix were added into the PCR tube containing the template total RNA. The final volume was then brought up to 10 µl with the addition of nuclease-free water. The mixture was mixed well by pipetting and was incubated at 65 ^{o}C for 5 minutes on a heat block and 2 minutes on ice. During the incubation step, a mixture of cDNA synthesis reagent was prepared by pipetting 2 µl of 10X buffer M-MuLV, 7.5 µl of nuclease-free water, and 0.5 µl of 200 U/µl M-MuLV reverse transcriptase into a PCR tube. The mixture was mixed well via pipetting, briefly centrifuged, and left on ice for 2 minutes. After the template mixture has chilled, 10 µl of the cDNA synthesis reagent mixture was then added to the template mixture and was mixed well via pipetting. With the final volume of 20 µl, the mixture was then incubated in a thermocycler at 42 ^{o}C for 60 minutes and followed by 85 ^{o}C for 5 minutes to terminate the reverse transcription process. After the incubation step, the PCR tube containing the synthesised cDNA was then chilled on ice for 2 minutes, briefly centrifuged and finally stored at -20 ^{o}C for future uses.

In order to assess the success of the cDNA conversion, an oil palm housekeeping gene *Cyclophilin 2* (*Cyp2*) was used in PCR that target fragment amplification of *Cyp2* with an amplicon size of 163 bp (Yeap et al., 2014). PCR reactions were conducted in a 20 µl reaction volume containing 10 µl of 2X power Taq PCR Master Mix (Bioteke, China), 1 µl of the template cDNA, 1 µl of 10 mM forward (5'-CTCGTCTGATGTCGTCTA-3') and reverse (5'-CTGCTGGTACTCTGGTAA-3') *Cyp2* primers (Yeap et al., 2014) and Milli-Q water to bring the volume up to 20 µl. The PCR profiles were as follow; initial denaturation at 95 ^{o}C for 5 minutes, denaturation, annealing, and extension steps for 30 cycles at 95 ^{o}C for 30 s, 56 ^{o}C for 30 s, and 72 ^{o}C for 1 minute, respectively and followed by final extension for 10 minutes at 72 ^{o}C. The amplified PCR products were then run on

a 1.5% (w/v) agarose gel at 70 V, 400 A for 90 minutes and stained using Novel Juice, DNA stain (GeneDireX, USA). To estimate the PCR product size, 100 bp ladder (Bioteke, China) was used as DNA marker.

3.2.5 Isolation of *EgSP1* coding sequence

To isolate the *SP1* coding sequence from oil palm cDNA, a high-fidelity Taq polymerase (Q5 Taq Polymerase, Biolabs (USA)) was used in the Polymerase Chain Reaction (PCR). Forward and reverse primers from Section 3.2.1 were used to amplify *EgSP1* cDNA fragment (referred onwards as CDS) with a target size of 1032 bp. PCR reactions were carried out in a 50 µl reaction volume containing 10 µl of Q5 reaction buffer, 1 µl of 10 mM dNTP, 2.5 µl of 10 mM forward and reverse *SP1* primers, 2.5 µl of DNA template, 0.5 µl of Q5 High fidelity DNA polymerase and nuclease free water to bring the volume up to 50 µl. Two replicates were prepared. The PCR profiles are as follow; initial denaturation at 98 ^{o}C for 30 s, denaturation, annealing and extension steps for 35 cycles at 98 ^{o}C for 10 s, 61 ^{o}C for 30 s and 72 ^{o}C for 40 s, respectively and followed by final extension for 2 minutes at 72 ^{o}C. The annealing temperature of the *SP1* primer pairs with the high-fidelity polymerase was calculated using the NEB T_m calculator software (http://tmcalculator.neb.com/#!/main) as suggested by the manufacturer. The PCR reaction was performed in two replicates of 50 µl volume along with a negative control (Milli-Q water). The amplified PCR products were then run on a 1.5% (w/v) agarose gel at 100 V, 400 A for 50 minutes and stained using Novel Juice, DNA stain (GeneDireX, USA). To estimate the PCR product size, 100 bp ladder (Bioteke, China) was used as DNA marker.

3.2.6 PCR product purification

PCR purification was performed using FavorPrep™ Gel/PCR purification kit (Favorgen, Taiwan). At room temperature, ~100 µl PCR products from Section 3.2.5 were added with 5 volumes of FADF buffer and well mixed by vortexing. The mixture was then transferred into a FADF column attached to a collection tube and was centrifuge at 20 ^{o}C and 10, 500 RCF for 30 s. The flowthrough was discarded and the 750 µl of wash buffer was added to the FADF column. The column was again centrifuge at 20 ^{o}C and 10, 500 RCF for 30 s and the flow-through was discarded. The FADF column was dried by centrifugation at 20 ^{o}C and 17, 500 RCF for an additional 3 minutes to dry the column matrix. The FADF column was then placed into a 1.5 ml microcentrifuge tube and 20 µl of Milli-Q water was added to the centre of the FADF column membrane. The FADF column was left to stand for 1 minute before being centrifuge at 20 ^{o}C and 17, 500 RCF for 1 minute to elute. The quality of the purified PCR product was determined under a Nanodrop and then separated on a 1.5% (w/v) agarose gel at 110 V, 400 A for 40 minutes and stained using Novel Juice, DNA stain (GeneDireX, USA). To estimate the PCR product size, 100 bp ladder (Bioteke, China) was used as DNA marker.

3.2.7 3’ Poly-A overhang addition

As the use of the Q5 high-fidelity polymerase results in the generation of PCR products with blunt ends, an additional 3’ Poly-A overhang addition step was performed on the purified product obtained from Section 3.2.6. The addition of 3’ Poly-A overhang was carried out using Taq DNA Polymerase (EURx®, Poland) in a 20 µl reaction volume containing 15 µl of (~25.1 ng/µl) purified PCR products, 2 µl of 2 mM dATP, 1 µl of Taq polymerase, and 2 µl of Taq buffer solution. First, the reaction mixture was well mixed via pipetting, briefly centrifuged and followed by incubation on a heat-block at 72 ^{o}C for 30 minutes. The reaction product was then removed from the heat-block and purified

using the PCR purification kit as describe in Section 3.2.6 and was eluted into a 10 µl volume. The quality of the purified PCR product was determined using a Nanodrop and then run on a 1.5% (w/v) agarose gel at 110 V, 400 A for 40 minutes and stained using Novel Juice, DNA stain (GeneDireX, USA). To estimate the PCR product size, 100 bp ladder (Bioteke, China) was used as DNA marker.

3.3 Gene cloning

3.3.1 Preparation of DH5α competent cells and glycerol stocks

To perform bacterial cloning, *E. coli* cells must first be made competent to enable uptake of plasmid containing gene of interest via $CaCl_2$ treatment technique. An aliquot of the 100 µl of DH5α *E. coli* glycerol stock was added to a 10 ml LB (lysogeny broth) broth in a 25 ml centrifuge tube. The mixture was then incubated overnight in an incubator at 37 ^{o}C with 200 rpm agitation for 20-24 hours. To prepare DH5α competent cell for bacterial cloning, 1 ml of the overnight culture was added with 10 ml of LB in a 25 ml centrifuge tube. The culture was then incubated in an incubator at 37 ^{o}C with 200 rpm agitation for an approximate of 30 minutes to 1 hour until the OD_{600} value reaches 0.25-0.3. From the DH5α culture, an aliquot of 1 ml was pipetted into a 1.5 ml microcentrifuge tube and left on ice for 15 minutes. The sample was then centrifuged at 4 ^{o}C and 1, 100 RCF for 10 minutes, supernatant discarded, pellets resuspended in 500 µl of ice-cold 0.1 M $CaCl_2$ solution, and then left on ice for 30 minutes. After 30 minutes, the sample was again centrifuged at 4 ^{o}C and 1, 100 RCF for 10 minutes. The supernatant was discarded, and the pellets were resuspended in 100 µl of ice cold 15% (v/v) glycerol 0.1 M $CaCl_2$ solution. The competent cells were then flash-frozen in liquid nitrogen and stored at -80 ^{o}C for future use. In addition, a glycerol stock was prepared from the overnight culture by aliquoting 850 µl of DH5α overnight culture into a 1.5 ml microcentrifuge tube containing 150 µl of 100% glycerol. The mixture was well mixed by pipetting until a

single homogenous solution was obtained, the glycerol stock was stored at -80 ^{o}C for future use.

3.3.2 Ligation and transformation into *E. coli*

To perform bacterial transformation, the purified PCR products from Section 3.2.7 were ligated into a cloning vector before transformed into *E. coli* (DH5α) via heat-shock treatment. The cloning vector chosen for this transformation step is pGEM-T Easy Vector (Promega, USA) that allows for blue-white screening of transformed bacteria. The ligation reaction was prepared on ice by adding 10 µl reaction volume containing 5 µl 2X ligation buffer, 1 µl of pGEM-T Easy Vector, 3 µl of (25.1 ng/µl) purified PCR product, and 1 µl of T4 DNA ligase. The mixture was mixed by pipetting, briefly centrifuged, and was incubated overnight at 4 ^{o}C.

On the next day, while minding the aseptic techniques, transformation step was performed by mixing 2 µl of the ligation mixture into the microcentrifuge tube containing 100 µl of DH5α competent cells. A positive and negative controls were prepared using existing plasmids in laboratory (pGEM-T Easy-*MaSOC1*) and water, respectively. The tubes were flicked slowly to mix, then left on ice for 30 minutes prior to the heat-shock treatment. After 30 minutes the tubes were heat-shocked by half-submerging the microcentrifuges in a 42 ^{o}C water bath for 45 s and was immediately transferred on ice for 2 minutes. The microcentrifuge tubes were then added with 950 µl of LB broth and incubated on a rotating shaker (220 rpm) at 37 ^{o}C for 3 hours. After the incubation step, 50 µl of the culture was spread on a LB media plate containing 0.05 mg/ml of ampicillin, 0.12 mg/ml IPTG and 0.08 mg/ml X-gal. The remaining culture as well as the positive and negative control was centrifuged at 1, 600 RCF for 2 minutes, the supernatant was removed until approximately 100 µl was left. The cell pellets were then resuspended, plated on the prepared LB media plate, and was appropriately labelled. The plates were

incubated in an incubator at 37 ºC overnight. Successful resultant clones were selected by blue-white screening method.

3.3.2.1 Colony PCR

Colony PCR was performed on selected positive white colonies using M13 universal forward and reverse primer pairs. Prior to PCR, 15 white colonies were picked from the agar plate with a sterile toothpick and inoculated onto a master plate with appropriate labelling and were subsequently transferred into PCR tubes containing 20 µl of autoclaved Milli-Q water. The mixtures were mixed by pipetting and briefly vortexed. Next, colony PCR was conducted in a 20 µl reaction volume containing 10 µl of 2X power Taq PCR Master Mix (Bioteke, China), 3.2 µl of the template DNA (from the 20 µl Milli-Q water inoculated with selected colonies), 1 µl of 10 mM M13 forward (5'-GTTTTCCCAGTCACGAC-3') and M13 reverse (5'-GGAAACAGCTATGACCATG -3') and Milli-Q water to bring the volume up to 20 µl. The PCR profiles was as follow; initial denaturation at 95 ºC for 5 minutes, denaturation, annealing and extension steps for 30 cycles at 95 ºC for 30 s, 56 ºC for 30 s and 72 ºC for 1 minute, respectively and followed by a final extension for 10 minutes at 72 ºC. The amplified PCR products was then run on a 1.5% (w/v) agarose gel at 100 V, 400 A for 45 minutes and stained using Novel Juice, DNA stain (GeneDireX, USA). To estimate the PCR product size, 100 bp ladder (Bioteke, China) was used as DNA marker.

3.3.3 Plasmid isolation from PCR positive colonies

Ten PCR positive colonies identified from the previous section were then subjected to plasmid extraction via Favorprep™ plasmid extraction mini kit (Favorgen, Taiwan) based on adjusted manufacturer protocol. Positive colonies were inoculated into a 50 ml centrifuge tube containing 10 ml of LB broth containing 0.05 mg/ml of ampicillin. The inoculated cultures were incubated at 37 ^{o}C and 220 rpm for 16-18 hours. From the overnight culture, glycerol stocks were prepared for future use by mixing 0.5 mL of overnight culture with 0.5 mL of 50% (v/v) glycerol to obtain a 25% (v/v) glycerol stock. Glycerol stocks prepared were flash frozen in liquid nitrogen and stored at -80 ^{o}C. Plasmid extraction step begins by centrifuging 5 ml of overnight culture at 4, 900 RCF for 10 minutes at 20 ^{o}C. The supernatant was then discarded, and the pellet obtained was then resuspended with 200 µl of FAPD 1 buffer via pipetting. The resuspended pellet was then transferred into a 1.5 ml microcentrifuge tube and was added with 200 µl of FAPD 2 buffer. This was followed by immediately inverting the microcentrifuge tube for 10 times before it was allowed to incubate for 3 minutes at room temperature. After the incubation step, 300 µl of FAPD 3 buffer was immediately added to the microcentrifuge tube containing the sample and the tube was again inverted for 10 times to the neutralized the FAPD 2 enzymatic activities from the previous reaction. The microcentrifuge tube containing the lysate was then centrifuged at 17, 500 RCF for 5 minutes at 20 ^{o}C to clarify the lysate. After centrifugation, the supernatant was then transferred into the prepared FAPD column in a collection tube, where the column and collection tube were then centrifuge at 17, 500 RCF for 1 minute at 20 ^{o}C. The column flowthrough was discarded, and the collection tube was then returned to the column. 400 µl of W1 solution was pipetted into the column, where it was allowed to incubate for 2 minutes at 20 ^{o}C. The column and collection tube were again centrifuge at 17,500 RCF for 1 minute at 20 ^{o}C, the supernatant discarded, and the collection tube was returned to the column. 700 µl of

wash buffer was then added to the column, and the column and collection tube were centrifuged again at 17, 500 RCF and for 1 minute at 20 ^{o}C. The flowthrough was discarded, and the collection tube was returned to the column. To dry the column, the column and collection tube were again centrifuged at 17, 500 RCF for 3 minutes at 20 ^{o}C. The column was then transferred to a new 1.5 ml microcentrifuge tube and 50 µl of autoclaved Milli-Q water was added to the centre of the column. The column was allowed to rest for 1 minute before being centrifuged at 17, 500 RCF for 1 minute at 20 ^{o}C to elute plasmid from the column into the microcentrifuge tube. The isolated plasmids were then analysed via Nanodrop to assess their quantity and quality before being sent for sequencing. The isolated plasmids were store at -20 ^{o}C.

3.3.3.1 Sequencing and analysis of isolated plasmids

The isolated plasmids harbouring the gene of interest (pGEM-T Easy-*EgSP1*) were then submitted for sequencing (1st Base DNA Sequencing Services). Isolated plasmid samples with a minimum concentration of 100 ng/µl were sent for sequencing. Forward and reverse single pass sequencing of the plasmid insert were performed with the supplied M13 forward primer (5'-GTTTTCCCAGTCACGAC-3') and reverse primer (5'-CAGGAAACAGCTATGAC-3'), respectively. The forward and reverse single pass sequencing were used to allow for an overlap sequence read from both ends of the target sequence.

The electropherogram data of each reaction were then analysed with 2 bioinformatics software, Unipro UGENE and MEGA-X. Each pair of electropherogram were carefully analysed and converted to sequence, where the pairs would be aligned and analysed to obtain a single proofread sequence. Sequences obtained from each sample sets were then aligned against the reference sequence of oil palm *SP1* transcript variants X1 and X2 to validate sequence of the isolated *SP1* coding sequence cloned into the plasmid.

3.4 Expression pattern analysis of *SP1* transcript in different tissues of oil palm

3.4.1 RNA extraction and cDNA synthesis

In order to analyse the expression pattern of *SP1* transcript in oil palm, total RNAs were extracted from different tissues including leaves, stem, and roots from *in vitro* plantlets as described in Section 3.2.2. Three biological replicates were used for each tissue type. Subsequent quality and quantity assessments were carried out as described in Section 3.2.3. Then, cDNAs were synthesized as described earlier using 1 μg total RNA as starting material and validated by PCR using housekeeping gene *Cyp2* (Yeap et al., 2014) as described in Section 3.2.4.

3.4.2 Primer design for quantitative PCR (qPCR)

For the purpose of gene expression analysis via qPCR, a primer pair with an amplicon size of between 150-200 bp was designed to target the transcript of *EgSP1* variant X1 (*EgSp1_X1*). The primer set was named as q*SP1* that contains forward (F-*EgqSP1*) and reverse (R-*EgqSP1*) primers. These primers were designed with the forward primer to overlap on the region which differs between the *EgSP1 X1* and *X2* transcript variants which is located ~1168–1174 bp of the *X1* variant transcript sequence. Primer pair were designed with NCBI primer-blast software and were evaluated individually via *in silico* analysis using PCR Primer Stat tool in Primer Stat Sequence Manipulation Suite online software (https://www.bioinformatics.org/sms2/pcr_primer_stats.html). The designed primer sequence with the size of 20-24 bp (5’ to 3’) were then sent to IDT (Integrated DNA Technologies) to be synthesized.

3.4.3 Primer efficiency test

Prior to expression pattern analysis, primer efficiency test was performed on both primers targeting the housekeeping gene *Cyp2* (Yeap et al., 2014) as well as the designed primer targeting partial sequence of the *EgSP1_X1* transcripts (F-*EgqSP1* and R-*EgqSP1*). All Real-Time PCR reactions were performed using the Applied Biosystems 7500 Real-Time PCR System hardware and 7500 (V2.3) associated software. Both PCR efficiency test and expression pattern analysis follow almost similar protocol in both master mix preparation and qPCR thermocycling conditions, with differences in sample setup and output data *in silico* analysis. Real-Time PCR was performed using dye-based method using qPCRBIO SyGreen Blue Mix separate-ROX (PCR Biosystems, United Kingdom). Prior for use, 2 µl of 50 µM ROX Additive was added to the 1 ml 2X qPCRBIO SyGreen Blue Mix as instructed in the manual provided.

PCR efficiency test was performed by preparing master mixes for each housekeeping gene (*Cyp2*) primer pairs and target gene (*EgSP1_X1*). Master mix for a 20 µl reaction was prepared by adding 10 µl of 2x qPCRBIO SyGreen mix with 1 µl of (10 µM) forward primer, 1 µl of (10 µM) reverse primer, and 7 µl of Milli-Q water. A serial dilution was prepared from a stock of 20 µl (1000 ng) cDNA template (derived from root RNAs, as RNA from roots was available excessively compared to other tissues), with serial dilution ranging from 2x, 4x, 8x and 16x was prepared for each primer set being tested. 1 µl of undiluted root cDNA template was added into 1 well of a qPCR strips (Novas Bio, USA) followed by addition of a 19 µl of master mix (containing *Cyp2* primer set) into the well containing the cDNA template. This was repeated for each serial diluted (2x, 4x, 8x, and 16x) root cDNA template and a no template control (NTC). The whole set up was repeated with another set of root cDNA templates (undiluted, 2x, 4x, 8x, and 16x) and an NTC but instead with a master mix containing *EgqSP1* primer set. For both reactions set up containing the *Cyp2* and *EgqSP1* primer sets, qPCR thermocycling condition was as

followed; initial denaturation at 95 ^{o}C for 2 minutes, followed by 40 cycles of denaturation at 95 ^{o}C for 5 s and annealing / extension 60 ^{o}C for 30 s. Default software settings was used for the melt curve analysis. Data collected from amplification curve and melt curve are compiled and further analysed through *in silico* analysis. PCR efficiency was determine analysing the standard curve generated by plotting the C_t value of each dilution against the natural log of the cDNA template. Using the data derived from the standard curve the PCR primer set was determined with the Equation 3.1, where the value of 2 is ideal and shows the doubling of DNA at each cycle (Pfaffl, 2001).

$$\text{Efficiency} = 10^{(-1/\text{slope of standard curve})} \quad (3.1)$$

3.4.4 Expression pattern analysis of *EgSP1*_X1

The expression pattern analysis of oil palm *SP1* was performed with similar protocol set up as in Section 3.4.3 with the preparation of 2 master mix containing housekeeping gene (*Cyp2*) primer set as a normalizing reference gene and (*EgqSP1*) primer set as the target gene being studied. Master mix for a 20 µl reaction was prepared by adding 10 µl of 2x qPCRBIO SyGreen mix with 1 µl of (10 µM) forward primer, 1 µl of (10 µM) reverse primer and 7 µl of Milli-Q water. Three biological replicates of oil palm tissues (leaf, stem, and root) cDNA templates were respectively pooled in equal amounts into a single tube to produce a homogenous mixture containing an average concentration between the 3 biological replicates. A technical triplicate of 1 µl of diluted (diluted at 1:1 ratio with Milli-Q water) pooled leaf cDNA template was added into 3 wells of a qPCR strips (Novas Bio) (USA) followed by addition of a 19 µl of master mix (containing *Cyp2* primer set) into each well containing the cDNA template. This was repeated for the pooled stem cDNA template, pooled root cDNA template and a no template control (NTC). The whole set up was repeated with another triplicate set of pooled (leaf, stem, and root)

cDNA templates and an NTC but instead with a master mix containing Egq*SP1* primer set. For both reactions set up containing the *Cyp2* and Egq*SP1* primer sets, qPCR thermocycling condition was as followed; initial denaturation at 95 ^{o}C for 2 minutes, followed by 40 cycles of denaturation at 95 ^{o}C for 5 s and annealing / extension 60 ^{o}C for 30 s. Default software settings was used for the melt curve analysis. Data collected from amplification curve and melt curve were compiled and further analysed. Expression analysis was performed on the collected data via *in silico* analysis using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001). *Cyp2* housekeeping gene was used for the normalization of target gene expression, whereas the fold change ratio was calibrated with the *EgSP1_X1* transcripts expression in the leaf.

CHAPTER 4: RESULTS

4.1 Identification of SP1 homolog from oil palm

4.1.1 Identification of putative oil palm SP1 homolog

From the data mining process performed on the two online databases; NCBI and PalmXplore using the full amino acid sequence of AtSP1 (At1g63900) as the query, a single gene was identified (LOC105060424) that encodes two isoforms designated as SP1 isoform X1 (XP_010942408.1) and SP1 isoform X2 (XP_010942409.1) (Table 4.1). These two isoforms were identified by BLASTp search from NCBI database (Figure 4.1), while BLAST search from PalmXplore resulted in the identification of only one relevant entry designated as (p5_sc00067_p0005_1) (Figure 4.2) which identical to the SP1 isoform X1 identified from NCBI (as confirmed by multiple sequence alignment showed in Appendix A). SP1 isoform X1 and X2 have 73.76% and 73.47% sequence identity to the query sequence Arabidopsis SP1 (At1g63900), respectively (Figure 4.1). The homolog encoding these two isoforms is referred as EgSP1 in this study.

Further analysis of *EgSP1* sequence indicated this gene to be 12,146 bp in length and contain 11 exons and 10 introns. Schematic representation *EgSP1* genomic region and the predicted transcript is shown in Figure 4.3. The transcript variants X1 and X2 CDS length was predicted to be 1032 nt and 1026 nt in length encoding 343 and 341 amino acids, respectively (Table 4.1). Multiple sequence alignment of amino acid sequence of EgSP1 X1 and X2 together with other putative SP1 homologues from nine plant species including oil Arabidopsis, tomato, soybean, apple, rapeseed, banana, sorghum, maize, and wheat showed the conservation of C_3HC_4-type RING-finger domain (an important domain for SP1) in all species (Figure 4.4).

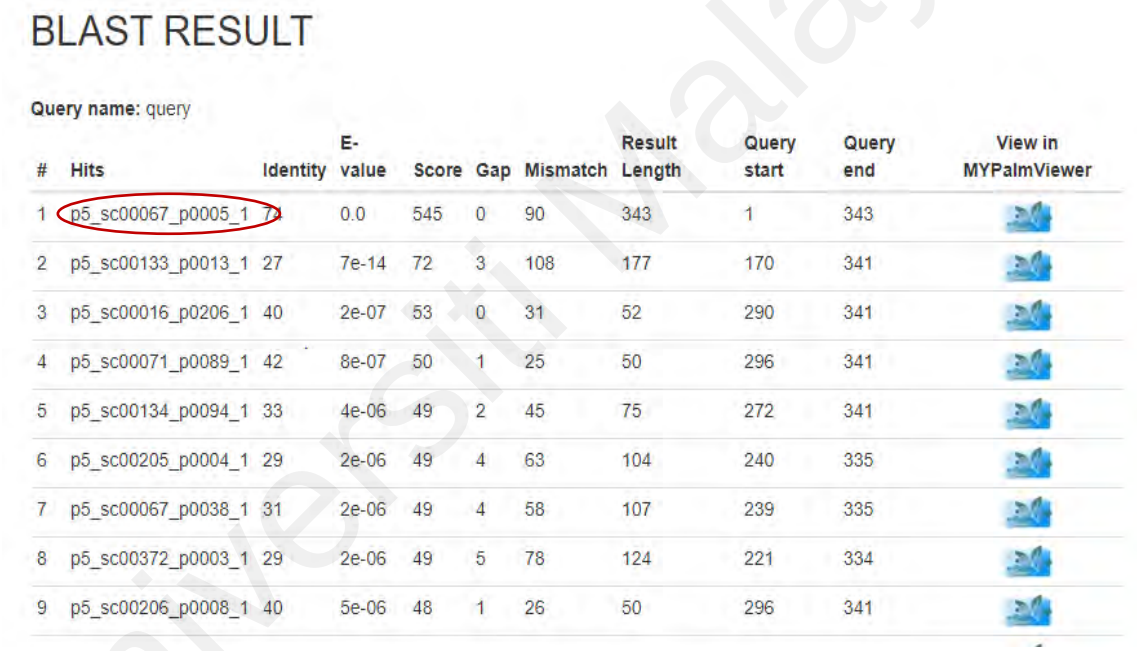
Sequences producing significant alignments | Download | Manage Columns | Show 100

select all 2 sequences selected | GenPept | Graphics | Distance tree of results | Multiple alignment

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☑	E3 ubiquitin-protein ligase SP1 isoform X1 [Elaeis guineensis]	545	545	100%	0.0	73.76%	XP_010942408.1
☑	E3 ubiquitin-protein ligase SP1 isoform X2 [Elaeis guineensis]	538	538	100%	0.0	73.47%	XP_010942409.1

Figure 4.1: Blastp search result from NCBI database in *Elaeis guineensis* using AtSP1 (At1g63900) amino acid sequence as query.

Two subjects identified as SP1 isoform X1 and SP1 isoform X2 with 73.76% and 73.47% sequence identity to AtSP1 (At1g63900), respectively.



BLAST RESULT

Query name: query

#	Hits	Identity	E-value	Score	Gap	Mismatch	Result Length	Query start	Query end	View in MYPalmViewer
1	p5_sc00067_p0005_1	74	0.0	545	0	90	343	1	343	
2	p5_sc00133_p0013_1	27	7e-14	72	3	108	177	170	341	
3	p5_sc00016_p0206_1	40	2e-07	53	0	31	52	290	341	
4	p5_sc00071_p0089_1	42	8e-07	50	1	25	50	296	341	
5	p5_sc00134_p0094_1	33	4e-06	49	2	45	75	272	341	
6	p5_sc00205_p0004_1	29	2e-06	49	4	63	104	240	335	
7	p5_sc00067_p0038_1	31	2e-06	49	4	58	107	239	335	
8	p5_sc00372_p0003_1	29	2e-06	49	5	78	124	221	334	
9	p5_sc00206_p0008_1	40	5e-06	48	1	26	50	296	341	

Figure 4.2: Blastp search result from PalmXplore database search using AtSP1 (At1g63900) amino acid sequence as query.

One subject identified (circled) as SP1 isoform X1 with 74% sequence identity to AtSP1 (At1g63900).

Table 4.1: Details on the identified oil palm SP1 isoforms

Database	Gene ID	Isoform	Accession number	Transcript ID	Coding sequence length	Amino acid length	Amino acid sequence similarity	Similarity of C3HC4-type (Conserved region)
NCBI/ PalmXplore	LOC105060424	X1	XP_010942408.1/p5_sc00067_p0005_1	XM_010944106.3	1032	343	99.42%	100%
NCBI	LOC105060424	X2	XP_010942409.1	XM_010944107.3	1026	341	99.42%	100%

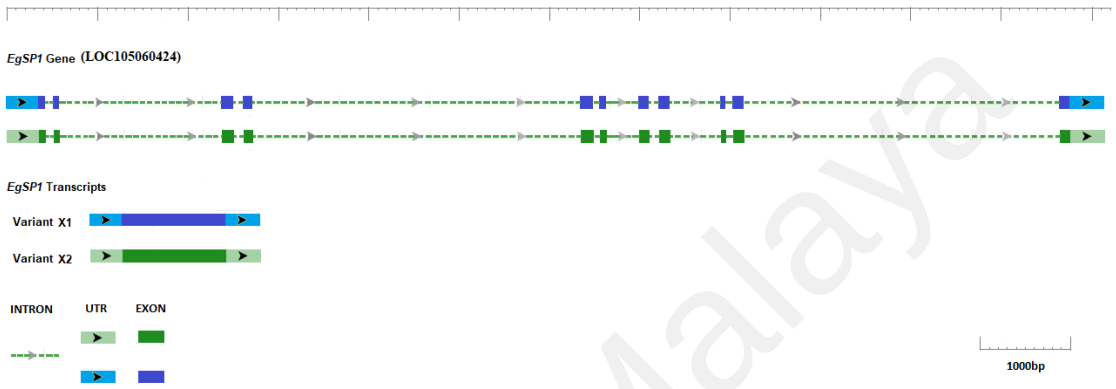


Figure 4.3: Schematic representation of the Intron-Exon Boundary in *EgSP1* gene and the *EgSP1* X1 and *EgSp1* X2 transcript variants.

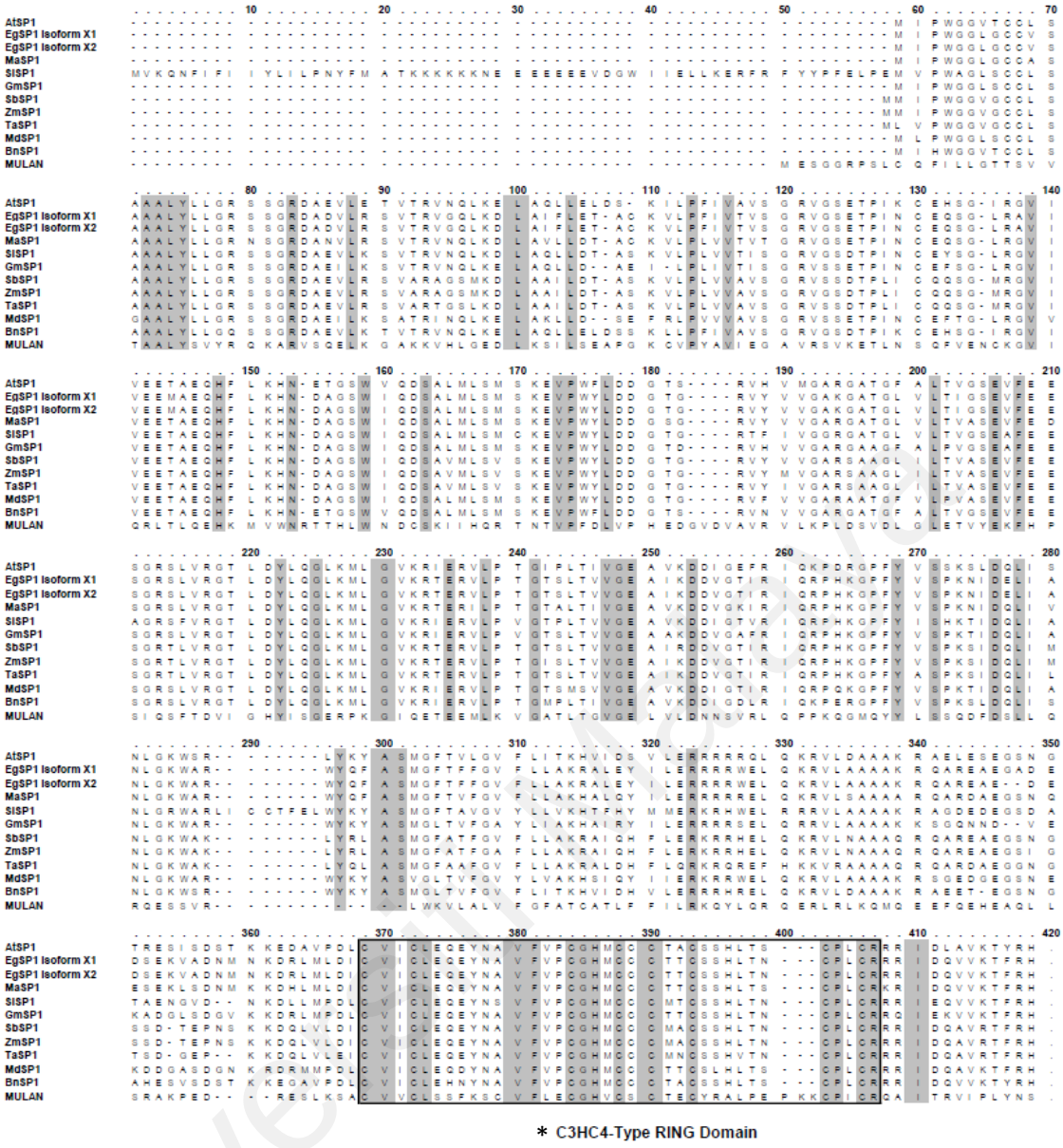


Figure 4.4: Amino acid sequence alignments of predicted EgSP1 protein with SP1 from other selected plant species and outgroup (MULAN).

C_3HC_4-type RING-finger conserve domains located between 369-407 mark are indicated by a box. Accession numbers are as follows: AtSP1(*Arabidopsis thaliana*, NP_176574.2) EgSP1 isoform X1 (*Elaeis guineensis*, XP_010942408.1), EgSP1 isoform X2 (*Elaeis guineensis*, XP_010942409.1), MaSP1 (*Musa acuminata subsp. Malaccensis*, XP_009410049.1), SlSP1 isoform X1 (*Solanum lycopersicum*, XP_010323032.2), GmSP1 (*Glycine max*, XP_003536411.1), SbSP1 (*Sorghum bicolor*, XP_002463293.1), ZmSP1 (*Zea mays*, XP_008670617.1), TaSP1(*Triticum aestivum*, QIQ08122.1), MdSP1(*Malus Domestica*, XP_008386573.1), BnSP1 (*Brassica napus*, XP_013747701.1), and MULAN (*Homo sapiens*, NP 078820.2).

4.1.2 Phylogenetic analysis

Phylogenetic analysis of SP1 proteins showed four distinct clades for angiosperms (eudicot and monocots), non-angiosperm plant (lycophyte) and outgroup MULAN. With reference to Figure 4.5, out of the 26 angiosperms in the phylogenetic tree 12 orders were present and was properly branched between the eudicots and monocots grouping. In this analysis, EgSP1 (X1 and X2 variants) fall in the Arecales order within the monocot clade and clustered together with SP1 proteins of the monocots including Asparageles, Zingiberales, Poales, Orchidaceae and Arecales. Within the monocot clade, the closest related species to the oil palm was found to be *M. acuminata* (dwarf banana) with a bootstrap value of 89%. *S. moellendorffii* being a lycophyte was as expectedly found to had branched away from the rest of the angiosperms. With reference to the amino sequence alignment in (Figure 4.4) that shows MULAN (*H. sapiens*) having the C_3HC_4-type RING-finger conserve region similarly to SP1, it was found that the sequence was still distinct enough as an outgroup.

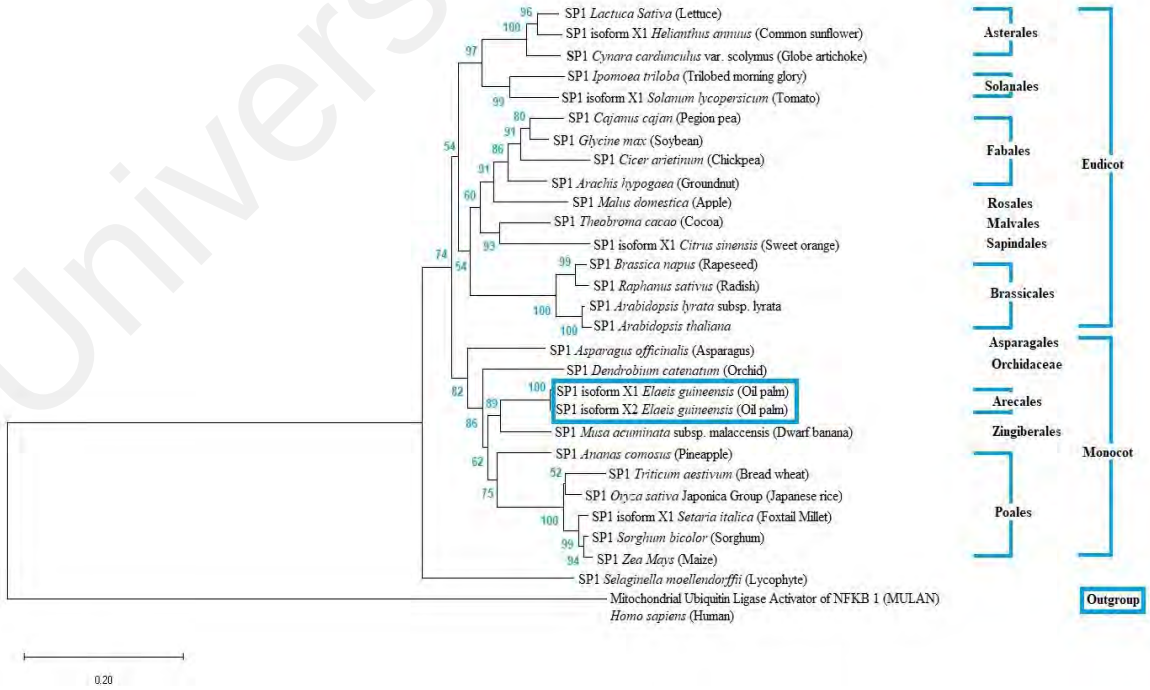


Figure 4.5: Phylogenetic tree of predicted SP1 amino acid sequence, constructed with Neighbour-Joining method.

Phylogenetic tree was constructed using MEGA-X software with 1000 bootstrap replications. Bootstrap values are indicated at nodes based on 1000 bootstrap replicates.

4.2 Isolation of oil palm *SP1* (*EgSP1*) coding sequence

4.2.1 *EgSP1* primer design

A specific primer pair was designed to amplify coding sequence of oil palm *SP1* transcript variants X1 (XM_010944106.3) with the expected size of 1032 bp (from start and stop codon). *In silico* evaluation of the designed primer using PCR Primer Stat Sequence Manipulation Suite (Stothard, 2000) showed that the designed primers passed almost all requirements for PCR suitability except that both primers have a high a T_m (> 58 ^oC) and the forward primer has runs of base 'G' which may cause mispriming (Figure 4.6). This was further reflected in Section 3.2.5 where the NEB T_m calculator software predicted the melting optimal annealing temperature to be at 61 ^oC. Moreover, the PCR suitability test shows the primer designed has no noticeable concerns for self-annealing or hairpin formations that may interfere with primer hybridization with the template.

Table 4.2: Product details of designed primer set.

Primer	Sequence	Length (mer)	Melting Temperature T_m, ℃	Molecular Weight, g/mol	Targeted product size (from cDNA)
Forward Primer (F-*EgSP1*)	5' – ATGATTCCATGGGG GGGATT – 3'	20	56.3	6228.1	1032 bp
Reverse Primer (R-*EgSp1*)	5' – TCAATGGCGAAATG TCTTTAC – 3'	21	51.2	6420.2	

```
General properties:
-------------------
                    Primer name: F-EgSP1
                Primer sequence: ATGATTCCATGGGGGGGATT
                Sequence length: 20
                    Base counts: G=8; A=4; T=6; C=2; Other=0;
                 GC content (%): 50.00
     Molecular weight (Daltons): 6228.12
                      nmol/A260: 5.04
                micrograms/A260: 31.41
           Basic Tm (degrees C): 52
   Salt adjusted Tm (degrees C): 47
Nearest neighbor Tm (degrees C): 64.07

PCR suitability tests (Pass / Warning):
---------------------------------
               Single base runs: Warning: Contains run of G's;
         Dinucleotide base runs: Pass
                        Length: Pass
                    Percent GC: Pass
          Tm (Nearest neighbor): Warning: Tm is greater than 58;
                      GC clamp: Pass
                 Self-annealing: Pass
               Hairpin formation: Pass
```

(a)

```
General properties:
-------------------
                    Primer name: R-EgSP1
                Primer sequence: TCAATGGCGAAATGTCTTTAC
                Sequence length: 21
                    Base counts: G=4; A=6; T=7; C=4; Other=0;
                 GC content (%): 38.10
     Molecular weight (Daltons): 6420.26
                      nmol/A260: 4.94
                micrograms/A260: 31.72
           Basic Tm (degrees C): 49
   Salt adjusted Tm (degrees C): 43
Nearest neighbor Tm (degrees C): 59.68

PCR suitability tests (Pass / Warning):
-----------------------------------
               Single base runs: Pass
         Dinucleotide base runs: Pass
                        Length: Pass
                    Percent GC: Warning: %GC is less than 40;
          Tm (Nearest neighbor): Warning: Tm is greater than 58;
                      GC clamp: Pass
                 Self-annealing: Pass
               Hairpin formation: Pass
```

(b)

Figure 4.6: PCR primer stats for the designed forward and reverse *SP1* primer pairs.

Primer stats data showing the properties of designed (a) forward primer, F-*EgSP1* and (b) reverse R-*EgSP1*.

4.2.2 Extraction of total RNA

Total RNA was successfully isolated from leaf, root, and stem tissues of ~1 month oil palm plantlets. Table 4.3 shows nanodrop readings of the 9 RNA samples that were extracted using Spectrum™ Plant Total RNA Kit (Sigma) showing concentration ranged between 135-487 ng/µl. All samples showed an acceptable A_{260}/A_{280} ratio range of 1.97 - 2.07 that indicates the RNA were of high quality and free from protein contaminants. As for the A_{260}/A_{230}, all samples were within the general acceptable range except for leaf 1 that has slightly higher ratio that might indicates low polysaccharide and phenol contaminations.

Table 4.3: Assessment of total RNA purity and concentration using Nanodrop.

Sample	A_{260}/A_{280} Ratio	A_{260}/A_{230} Ratio	Concentration (ng/µl)
Leaf 1	2.09	2.81	135
Leaf 2	2.04	2.27	487
Leaf 3	1.97	1.59	152
Stem 1	2.02	2.04	387
Stem 2	1.97	1.61	167
Stem 3	2.07	2.07	373
Root 1	1.97	1.85	373
Root 2	1.97	1.81	181
Root 3	1.97	2.03	210

Note: Total RNA extraction was done with Spectrum™ Plant Total RNA Kit (Sigma) with addition of DNase. Sample was obtained from MPOB and was pre-weighted at ~100 mg.

In addition to nanodrop reading, quality of RNA was evaluated by agarose gel electrophoresis. Gel electrophoresis results indicated a good quality of the extracted RNA based on the presence of 28S and 18S RNA bands in all the 9 total RNA samples (Figure 4.7). The 28S RNA band was noted to be near the 3000 bp mark where the 18S RNA band was observed to be near the 2000bp and 1500 bp mark (indicated by arrow in Figure 4.7). From this analysis, residual genomic DNA were noted in several wells, although the absence of prominent bands above the 28S RNA band indicates the samples were adequately fit for use.

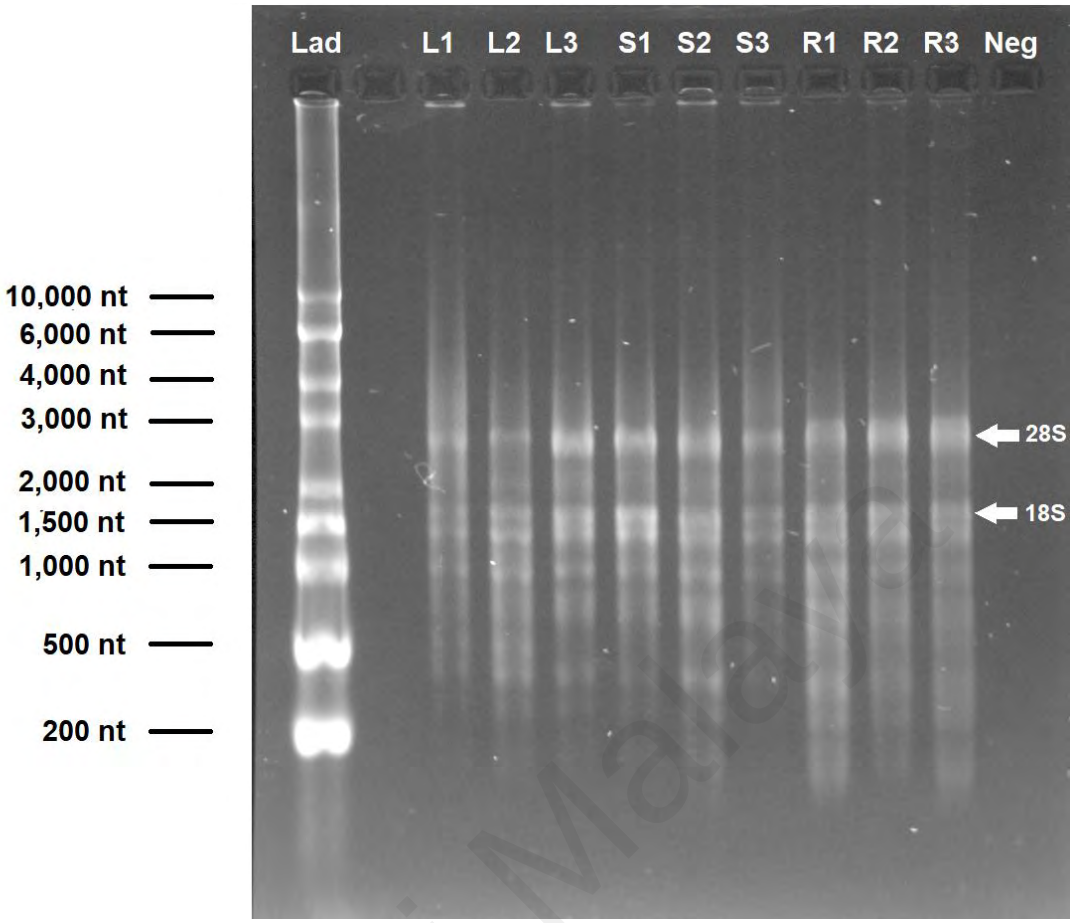


Figure 4.7: Gel electrophoresis of total RNA extracted from various tissue types of oil palm.

Samples were analyzed on 1% (w/v) agarose with TBE buffer and loading concentration of ~40 ng/µl of sample per well. Sigma transcript RNA marker (0.2-10 kbp) was used as ladder (LAD). Leaves, stems, and roots sample designated as "L1, L2, L3", "S1, S2, S2" and "R1, R2, R3", respectively.

4.2.3 Successful cDNA synthesis confirmation by PCR

Successful conversion of RNA (in Section 3.2.4) to cDNA was confirmed using PCR using housekeeping gene, *Cyp2*. Amplification of *Cyp2* fragments with a targeted size of 163 bp was observed in all oil palm cDNA samples (Figure 4.8). In addition, negative control showed no observable bands, which indicate no notable contamination had happened during the cDNA synthesis process.

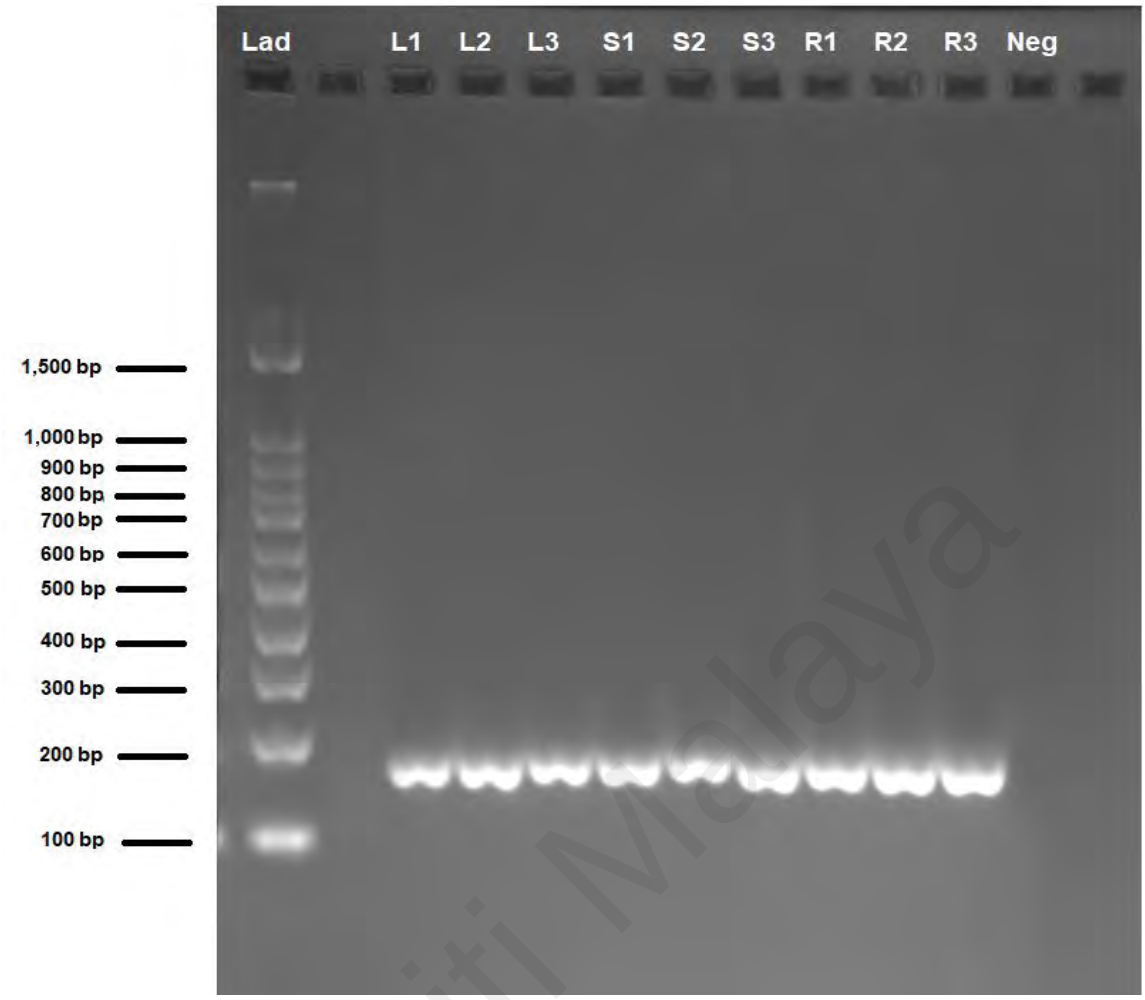


Figure 4.8: Gel electrophoresis of PCR product with housekeeping, *Cyp2* primers for the newly synthesized cDNA from various oil palm tissues.
Samples were analyzed on 1.5% (w/v) agarose with TBE buffer and 3 µl PCR product loading per well. 100 bp ladder (Bioteke, China) was used for the gel electrophoresis and designated as ladder (LAD). Leaves (L), stems (S), and roots (R) sample designated as "L1, L2, L3", "S1, S2, S2" and "R1, R2, R3", respectively.

4.2.4 Isolation of *EgSP1* coding sequence

PCR reactions using the designed primer pair (Table 4.2) were successful at amplifying the targeted amplicons (*EgSP1 X1* variant). Agarose gel visualisation of PCR products appeared as a clear single band observed slightly above the 1000 bp mark which was appropriate for the targeted 1032 bp length of oil palm *SP1* coding sequence amplicon size (Figure 4.9). No other observable bands were noted for the negative control indicating that there is no noticeable contaminates present.

Next the PCR products containing *EgSP1* were successfully purified using column (referred as first purification) before subjected to 3' poly-A overhangs addition step (Figure 4.10a). Then, a second purification was carried out and AGE confirmed the presence of the bands at ~1000 bp as shown in Figure 4.10b. These purification steps were performed to purify PCR reaction materials that may interfere during the 3' poly-A overhang addition step and the following vector ligation steps. Both Nanodrop results for the first and second purification step performed are shown in Table 4.4, in which the concentration of the PCR products were 54 ng/µl and 21.5 ng/µl, after first and second purification, respectively. No bands were observed in the negative control indicating that there is no noticeable contamination present.

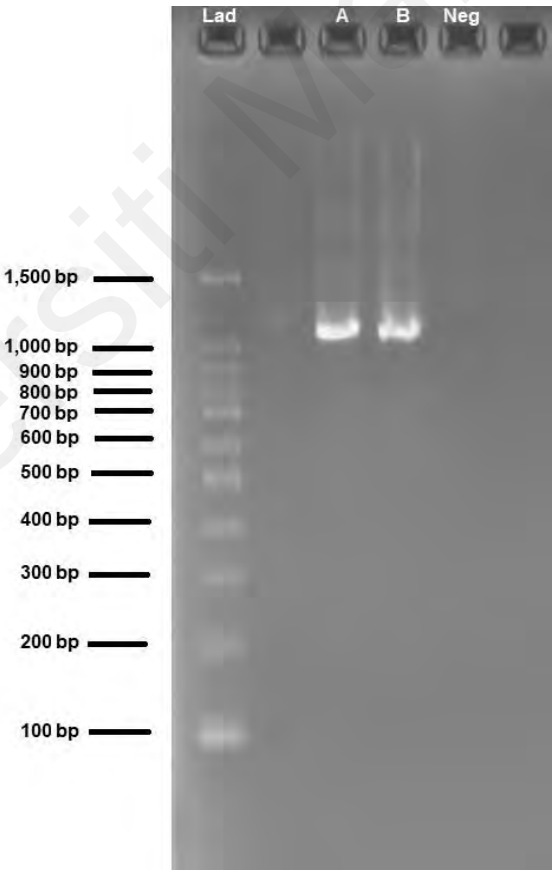


Figure 4.9: Gel electrophoresis of Q5 High-Fidelity Taq Polymerase PCR product for isolated *SP1* CDS.

Samples were analyzed on 1.5% (w/v) agarose with TBE buffer and 2.5 µl PCR product loading per well. 100 bp ladder (Bioteke, China) was used for the gel electrophoresis and designated as ladder (Lad). PCR product was designated as A and B representing two replicates along with a negative control designated as (Neg).

Table 4.4: Nanodrop analysis of first and second PCR purification steps.

Sample	A_{260}/A_{280} Ratio	A_{260}/A_{230} Ratio	Concentration (ng/µl)
First purification	1.76	1.00	54.0
Second purification	2.00	0.40	21.5

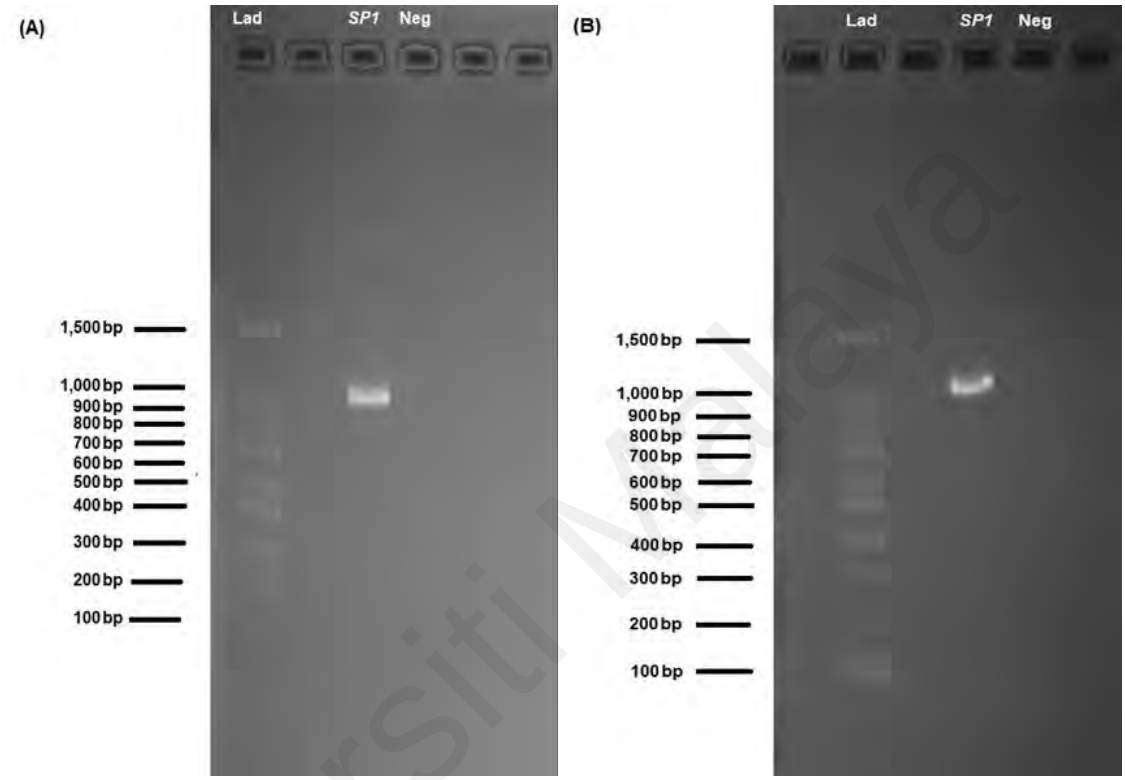


Figure 4.10: Gel electrophoresis for the first and second PCR purification steps.
The purified products prior and after the 3' poly-A overhangs addition steps were analyzed on two separate 1.5% (w/v) agarose with TBE buffer. (A) Shows the gel visualization of the purified *SP1* PCR products prior to 3' poly-A overhangs addition step. (B) Shows the gel visualization of the purified *SP1* PCR products after the 3' poly-A overhangs addition step. In each of the agarose gel run, 2.5 µl and 2 µl of purified PCR products were loaded in each gel for the first and second purification step, respectively. For both gel, 100 bp ladder (Bioteke, China) was used for the gel electrophoresis and designated as ladder (Lad). Purified PCR product is designated as '*SP1*' with negative control is designated as 'Neg'.

4.3 Cloning of *EgSP1* coding sequence

4.3.1 *E. coli* transformation and colony PCR

The isolated CDS of *EgSP1* was then ligated into the pGEM-T Easy Vector and subsequently transformed into the *E. coli* DH5α host from Section 3.3.2. Figure 4.11a shows high number of colonies grown on a LB growth media supplemented with

ampicillin with addition of IPTG and X-gal for blue-white screening. As the reference sequence for the *EgSP1* transcript is known prior from the data mining process, only fifteen white colonies were subjected to testing for presence of insert DNA in the plasmid construct with colony PCR. These colonies were then subsequently isolated on master plate and glycerol stock (Figure 4.11b).

Prior to plasmid extraction and sequencing step, transformed colonies were selected by colony PCR. Visualization of colony PCR of fifteen isolated colonies using M13 primers shows clear banding at the targeted size of ~1300 bp mark (Figure 4.12).

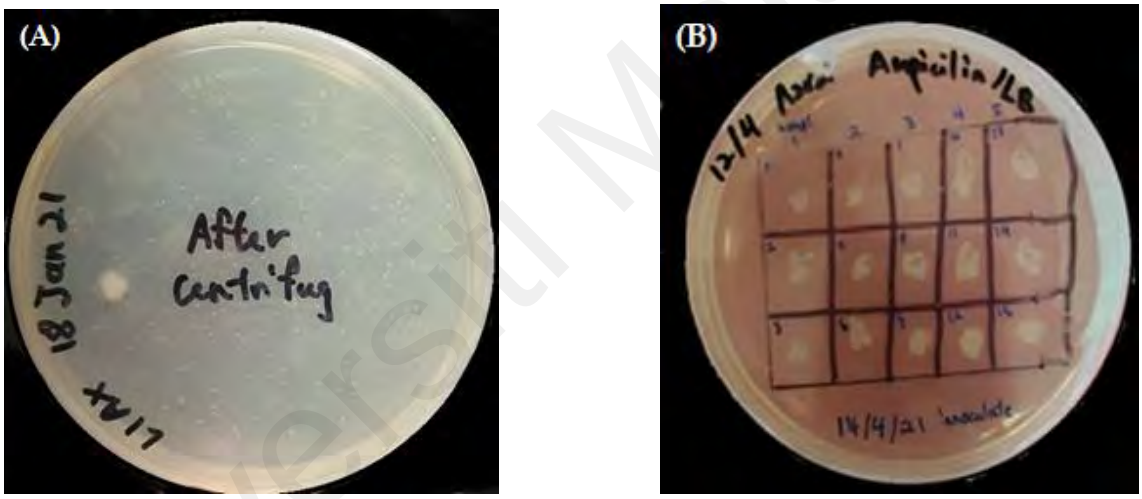


Figure 4.11: Blue-white screening of transformed *E. coli* (*DH5α*) and colony master plate of positive colonies.
Colonies of *E. coli* DH5α transformed with pGEM-T Easy Vector grown on LB growth media containing ampicillin with addition of IPTG and X-gal. (B) Master plate containing 15 colonies.

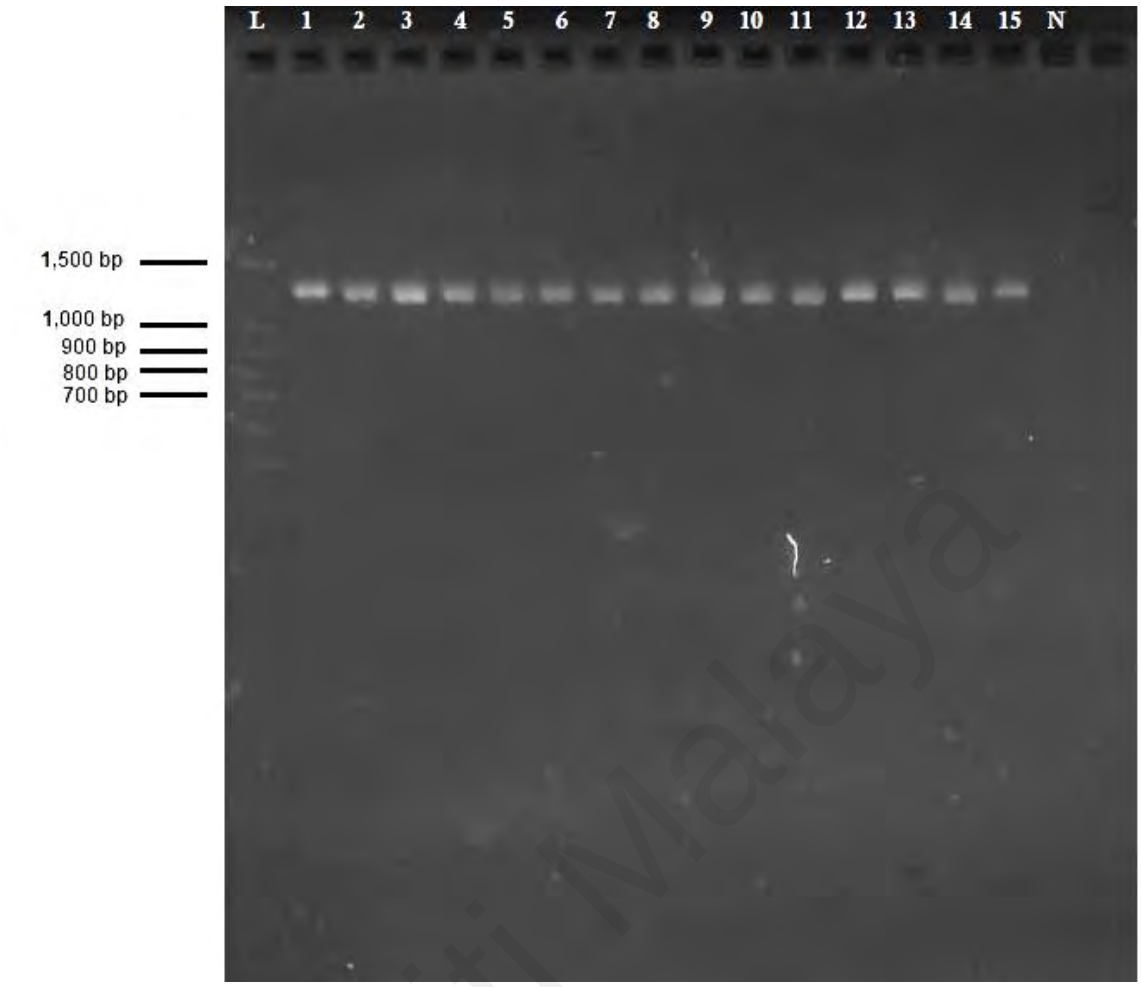


Figure 4.12: Detection of the presence of *EgSP1* in the plasmid of transformed *E. coli* by colony PCR.
Gel visualization of colony PCR of fifteen colonies using the M13 primer set with clear banding at ~1300bp. Colony PCR product were run on a 1.5% (w/v) agarose with TBE buffer. For the gel visualization, a 100 bp ladder (Bioteke, China) was used for the gel electrophoresis and designated as ladder (L), colony PCR sample designated with respective numbering with negative control designated as 'N'.

4.3.2 Plasmid isolation and sequencing

Fifteen PCR positive colonies, ten colonies were selected to be grown and subjected to plasmid extraction. Table 4.5 shows the nanodrop reading of plasmids extracted with Favorprep™ plasmid extraction mini kit (Favorgen, Taiwan) showing concentration ranges between 67.8 – 488 ng/µl. All sample show a consistent reading of A_{260}/A_{280} ratio (~1.8) and an acceptable reading of A_{260}/A_{230} ratio (1.6-2.2). These ratios indicated that the samples were of fine quality for sequencing step. From 10 colonies, only 7 colonies (colonies 1, 2, 5, 8, 10, 13 and 14) with high concentration and good plasmid quality were sent for the subsequent sequencing analysis.

From the sequencing process of 7 plasmid sample with 2 sequencing runs with forward and reverse M13 primers for each sample, resulted in 14 sequencing data (Appendix E). The electropherograms showed a clear sequence reading with no truncated sequence read or indel. *In silico* analysis and sequence alignment of both forward and reverse sequencing run of each sample results in a single proofread sequence for each of the 7 respective colony sequence. The 7 proofread sequences were then aligned with oil palm *SP1* transcript variant X1 and X2 as a comparative reference (Figure 4.13). In Figure 4.13, it can be observed that all of the 7 samples had shown closer sequencing patterns to that of the *SP1* transcript variant X1 than that of X2. All of the 7 sequences had the same length as X1 variant including the 6 bp region that is missing in X2 sequence. From the same alignment result, it was also observed that 4 of the 7 colonies had shown similar mutations at 3 locations (indicated by '*' in Figure 4.13) compared to the reference sequence of X1 transcript variant which maybe a type of single nucleotide polymorphisms (SNPs). To assess whether the presences of the said SNPs would show any significant changes to the amino acid sequence between the 7 samples, the nucleotide sequences of the 7 samples were first translated and then realigned against the amino acid sequences of the oil palm SP1 isoform X1 and X2 (Figure 4.14). From the amino acid alignment results obtained, it was found that sample sequences with or without the SNP, showed no significant differences in the amino acid sequence with 100% sequence identity between each of the 7 samples.

Table 4.5: Nanodrop analysis of isolated plasmid.

Sample	A_{260}/A_{280} Ratio	A_{260}/A_{230} Ratio	Concentration (ng/µl)
1	1.849	2.197	381
2	1.845	2.227	488
4	1.857	1.926	76.1
5	1.839	1.984	123
7	1.829	1.738	67.8
8	1.833	2.062	113
10	1.806	1.710	173
11	1.833	1.646	102
13	1.833	1.977	209
14	1.839	1.968	179

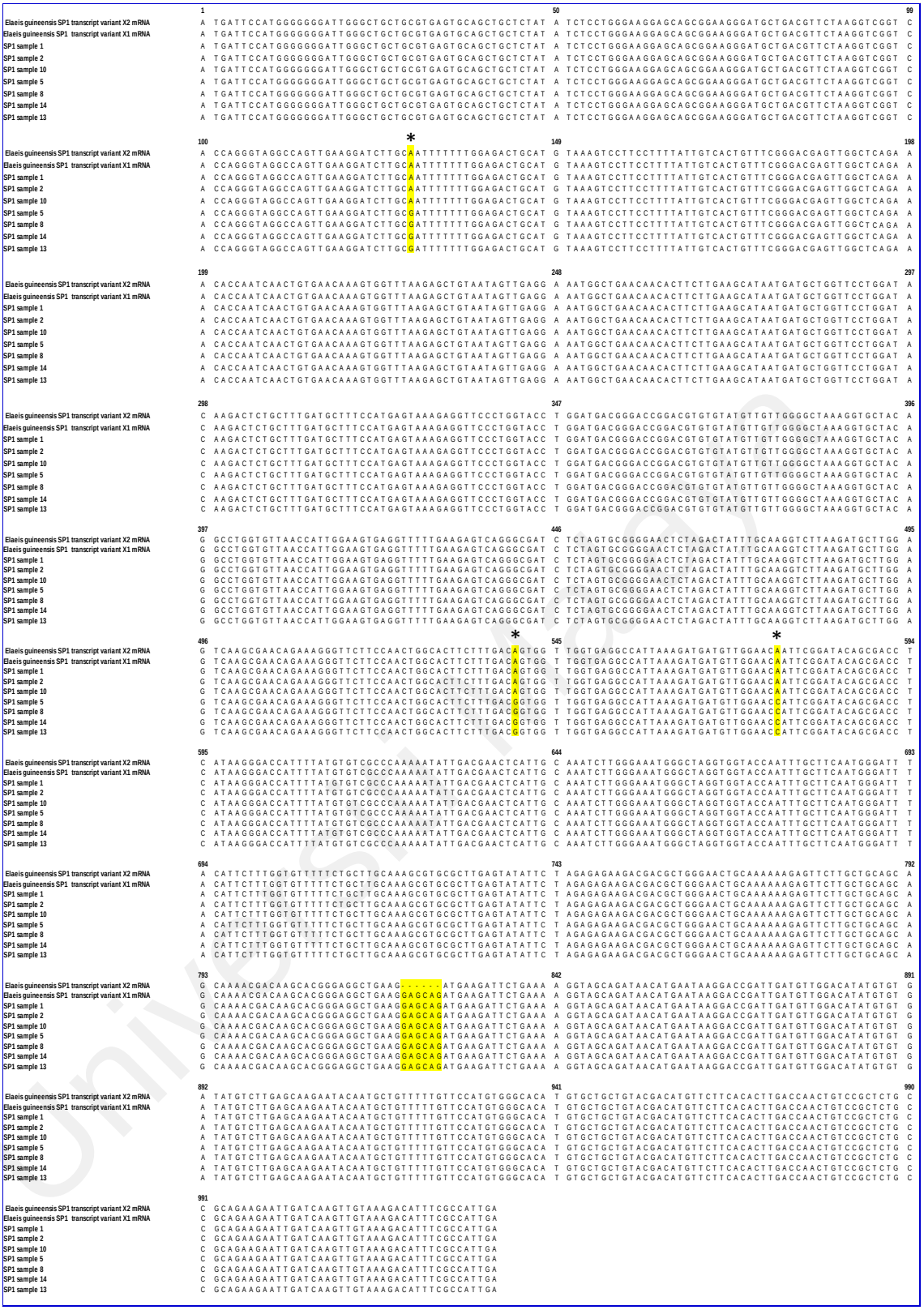


Figure 4.13: Nucleotide sequence alignment of isolated plasmid insert sequences against *EgSP1* transcript variant sequence X1 and X2.

Sequence variation between *SP1 X1* and *X2* transcript variants as well as the 7 sequenced plasmids are indicated as being highlighted in yellow. '*' indicates the locations of mutation or SNPs. The sequence alignment spans from the *SP1* start codon (ATG) and stop codon (TGA). Whole sequence length of alignment is 1032 and 1026 bp for X1 and X2 transcript variants respectively.

Figure 4.14: Amino acid sequence alignment of translated plasmid sample insert sequences against oil palm SP1 isoform X1 and X2.

Amino acid sequence variation between SP1 X1 and X2 isoforms as well as 7 translated plasmid sequences are indicated as being highlighted in yellow. The amino acid sequences span a range of 343 and 341 amino acid for X1 and X2 isoforms, respectively.

4.4 Expression pattern analysis of *EgSP1* transcript in different oil palm tissues

4.4.1 Primer design for quantitative PCR (qPCR)

For the purpose of determining the expression pattern of oil palm *SP1* X1 transcript variant in the oil palm vegetative tissues (leaf, stem, and root), a specific primer pair was designed to amplify a partial sequence of the *EgSP1* transcript variant X1 that from Section 4.3.2. The primer pair were designed with the expected size of 179 bp with one of its primer pair to overlap on the region which differs between the oil palm *SP1* X1 and X2 transcript variants which is located ~1168–1174 bp region. This primer design approach was done to ensure that only the X1 variant transcripts will being amplified during the expression analysis. *In silico* evaluation of the designed primer using PCR Primer Stat Sequence Manipulation Suite (Stothard, 2000) showed that the designed primers passed almost all requirements for Real-Time PCR suitability (Figure 4.15). The PCR suitability test shows the primer designed has no noticeable concerns for self-annealing or hairpin formations that may interfere with primer hybridization with the

template. The primer pair is designated as *EgqSP1* to differentiate it from the primer pair designed to target the *SP1* coding sequence from Section 4.2.1.

Table 4.6: Product details of designed primer pair.

Primer	Sequence	Length (mer)	Melting Temperature T_m, ℃	Molecular Weight, g/mol	Targeted product size (from cDNA)
Forward Primer (F-*EgqSP1*)	5' – GAGGCTGAAGGAGCAG ATGAAG – 3'	22	60.74	6922.58	179 bp of *EgSP1* variant X1
Reverse Primer (R-*EgqSp1*)	5' – CAGAGCGGACAGTTGG TCA – 3'	19	59.63	5877.88	

```
General properties:
-------------------
                    Primer name: F-EgqSP1
                Primer sequence: GAGGCTGAAGGAGCAGATGAAG
                Sequence length: 22
                    Base counts: G=10; A=8; T=2; C=2; Other=0;
                 GC content (%): 54.55
     Molecular weight (Daltons): 6922.58
                      nmol/A260: 4.27
                micrograms/A260: 29.58
           Basic Tm (degrees C): 57
   Salt adjusted Tm (degrees C): 52
Nearest neighbor Tm (degrees C): 65.24

PCR suitability tests (Pass / Warning):
-----------------------------------
               Single base runs: Pass
         Dinucleotide base runs: Pass
                        Length: Pass
                    Percent GC: Pass
          Tm (Nearest neighbor): Warning: Tm is greater than 58;
                      GC clamp: Pass
                 Self-annealing: Pass
              Hairpin formation: Pass
```

(a)

```
General properties:
-------------------
                    Primer name: R-EgqSP1
                Primer sequence: CAGAGCGGACAGTTGGTCA
                Sequence length: 19
                    Base counts: G=7; A=5; T=3; C=4; Other=0;
                 GC content (%): 57.89
     Molecular weight (Daltons): 5877.88
                      nmol/A260: 5.25
                micrograms/A260: 30.84
           Basic Tm (degrees C): 53
   Salt adjusted Tm (degrees C): 48
Nearest neighbor Tm (degrees C): 64.73

PCR suitability tests (Pass / Warning):
-----------------------------------
               Single base runs: Pass
         Dinucleotide base runs: Pass
                        Length: Pass
                    Percent GC: Pass
          Tm (Nearest neighbor): Warning: Tm is greater than 58;
                      GC clamp: Pass
                 Self-annealing: Pass
              Hairpin formation: Pass
```

(b)

Figure 4.15: PCR primer stats for the designed Real-Time PCR, forward and reverse *EgqSP1* primer pairs.

Primer stats data showing the properties of designed (a) forward primer, F-*EgqSP1* and (b) reverse *EgqSP1*.

4.4.2 Primer efficiency test

Prior to expression analysis, primer efficiency test was performed on both the primers targeting the housekeeping gene (*Cyp2*) as well as the designed primer pair *EgqSP1*, targeting the *SP1* transcript variant X1 in separate single runs. PCR efficiency was determined from the standard curve C_t value of each dilution against the natural log of the cDNA template (Figure 4.16), where 2.0 is the ideal value for the expected doubling of DNA at each cycle. Primer efficiency for primer pair *Cyp2* and *EgqSP1* were calculated to be 1.69 and 1.90 respectively (Pfaffl, 2001). R^2 value of the cycle threshold value versus ln (cDNA) as shown in Figure 4.16 show a good inverse corelation between template concentration and cycle threshold value for both the *Cyp2* and *EgqSP1* primer pairs. Melt curve analysis of both *Cyp2* and *EgqSP1* primer pairs shown in Figure 4.17 and Figure 4.18 both showed a single peak at approximately 80 ^oC and 82 ^oC respectively, indicating the presence of only a single significant amplification product for each primer pairs.

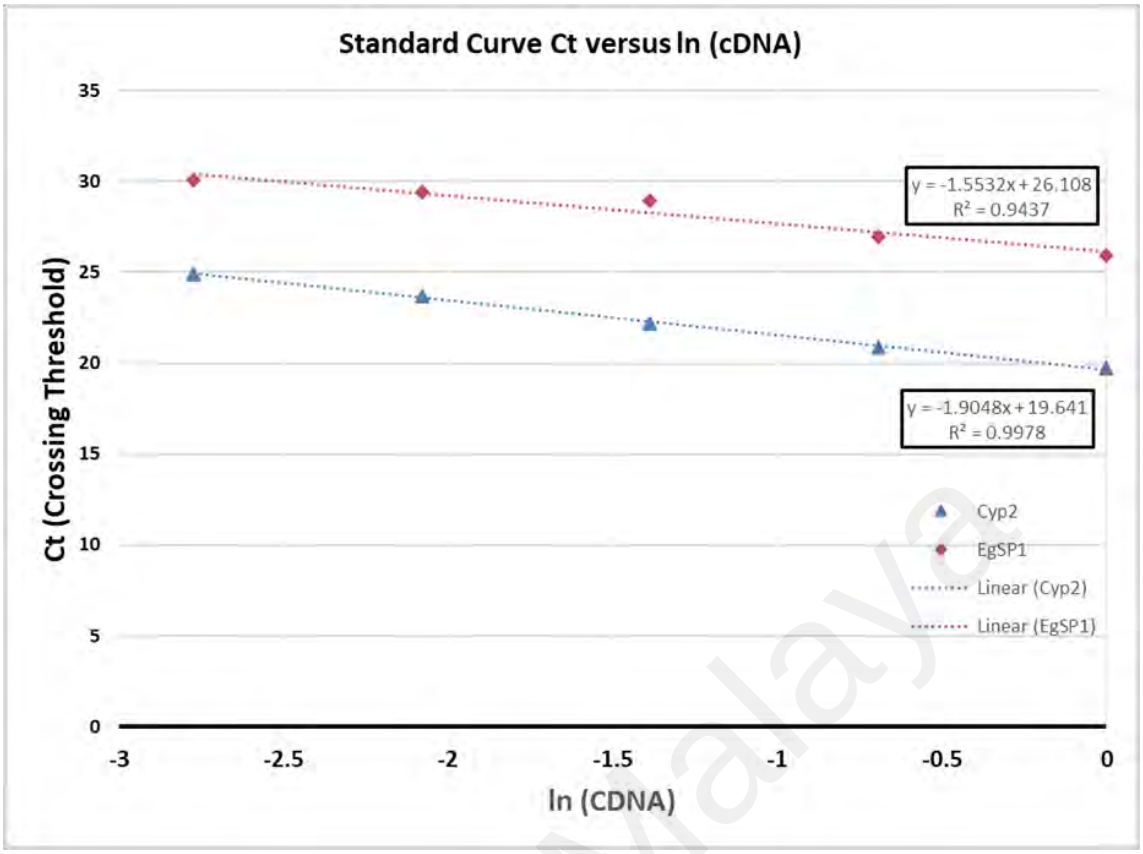


Figure 4.16: Standard curve C_t versus ln (cDNA) of primer pairs *Cyp2* and *EgqSP1*.

PCR efficiencies were determined in single run for each primer pair in 2 independent experiments. Each PCR run were performed with 5 samples representing dilution factors of 2X, 4X, 8X, 16x and undiluted template cDNA. Both standard curves for *Cyp2* and *EgqSP1* primers demonstrated an inverse relationship between template concentration and C_t value, with a R^2 value of 0.9978 and 0.9437, respectively.

Figure 4.17: Real-Time PCR melt curve analysis of *Cyp2* primer pair.

Melt curve analysis of *Cyp2* primer pair post Real-Time PCR reaction of 40 cycle and at annealing temperature of 60 ºC, for 5 samples representing dilution factors of 2X, 4X, 8X, 16x and undiluted template cDNA.

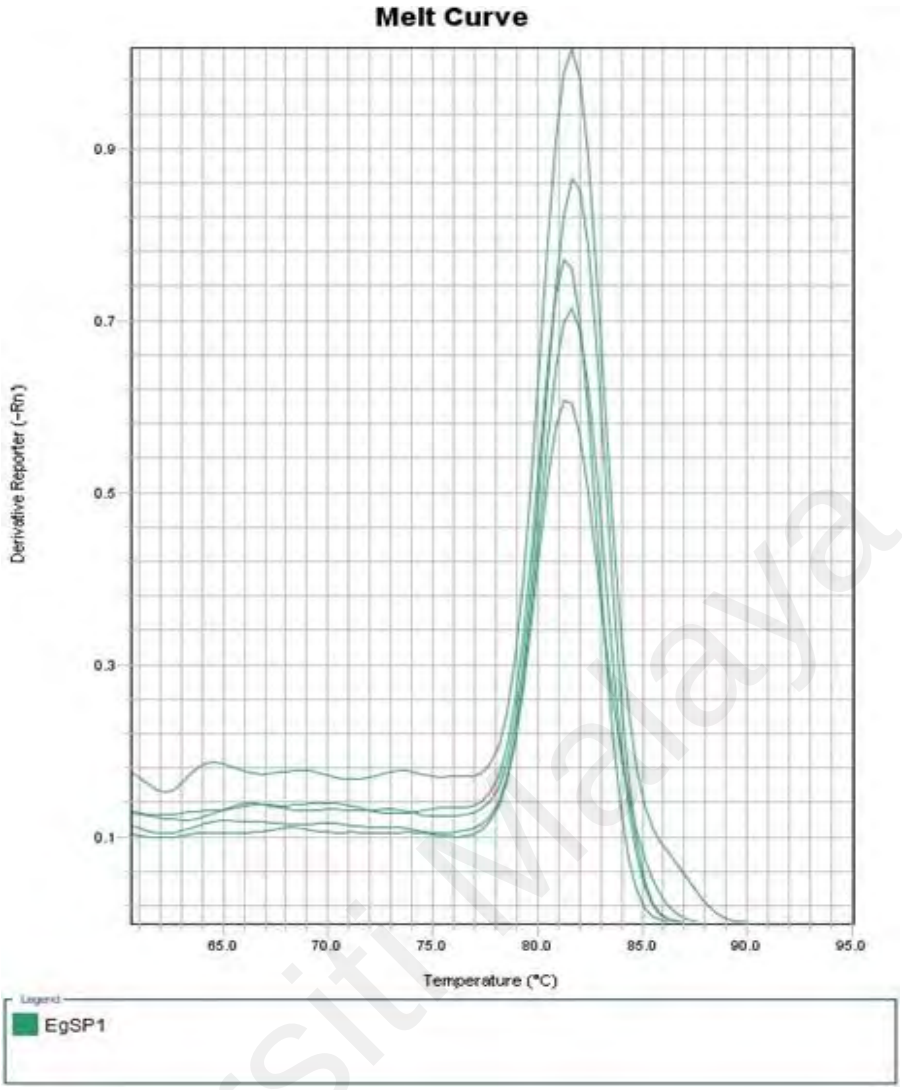


Figure 4.18: Real-Time PCR melt curve analysis for *EgqSP1* primer pair.

Melt curve analysis of *EgqSP1* primer pair post Real-Time PCR reaction of 40 cycle and at annealing temperature of 60 ºC, for 5 samples representing dilution factors of 2X, 4X, 8X, 16x and undiluted template cDNA.

4.4.3 Expression pattern analysis of *EgSP1*_X1

Expression pattern analysis was performed on pooled cDNA samples of vegetative tissues such as the leaves, stems, and roots of ~1 month old oil palm plantlets. While normalizing the expression level *SP1* transcript to the housekeeping gene *Cyp2*, *EgSP1* was found to be highly expressed in leaf tissues, followed by the roots and finally the stem tissues (Figure 4.19). Melt curve analysis of both *Cyp2* and *EgqSP1* primer pair shown in Figure 4.20 shows a single peak at approximately 80 ^{o}C and 82 ^{o}C respectively, similar to the results obtained from primer efficiency test in Section 4.4.2, which indicates the presence of a single significant amplification product for each respective primer pair.

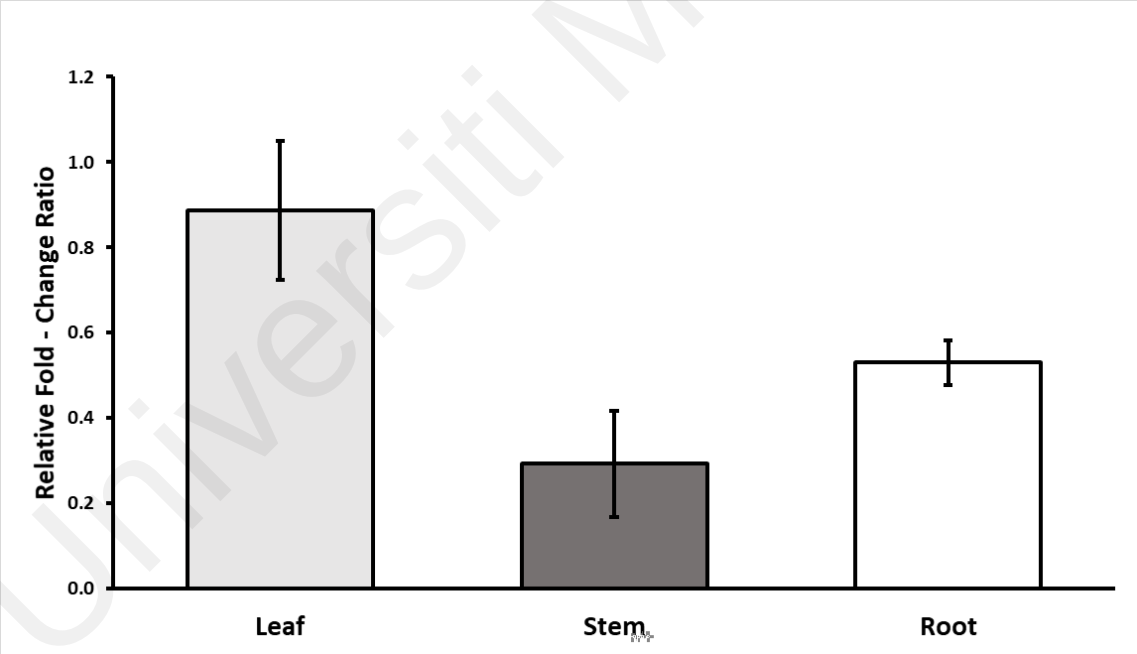


Figure 4.19: Real-Time PCR of *EgSP1* transcripts (variant X1) in leaf, stem, and root tissues of oil palm.

Note: Target gene expression level was normalized with *Cyp2* reference gene. Fold change value was calculated by calibrating the normalised expression of all transcripts against leaf tissue. Error bars indicate the standard deviation for 3 technical replicates values of each respective pooled tissue samples.

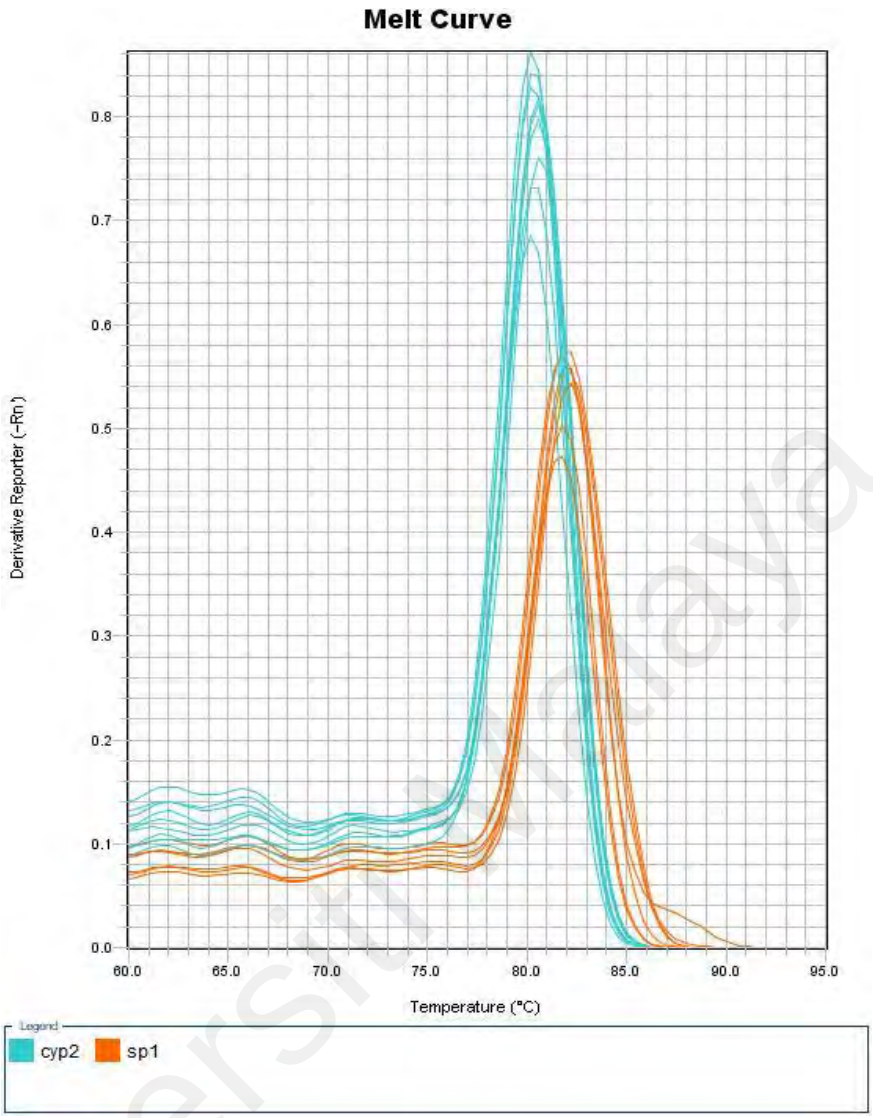


Figure 4.20: Melt curve analysis of primer pairs, *Cyp2* and *EgqSP1* for housekeeping gene and *SP1* transcript variant X1 respectively.

Melt curve analysis of EgqSP1 primer pair post Real-Time PCR reaction of 40 cycle and at annealing temperature of 60 ºC, for 5 samples representing dilution factors of 2X, 4X, 8X, 16x and undiluted template cDNA

CHAPTER 5: DISCUSSION

Oil palm is one of the most important crops in Malaysia. Many efforts were taken to improve this crop in terms of yield and disease tolerance. In an attempt to contribute to the improvement efforts, this study aimed to isolate and characterize homologues of SP1 protein that was first identified in Arabidopsis (Ling et al., 2012). Recent studies of SP1 protein further revealed and suggested ways to improve yield-related parameters in a major crop, tomato (Ling et al., 2021). Interestingly, homologues of SP1 are widely distributed in plants suggesting that its important functions in crop improvement may be well conserved and are likely to have relevance in other important crop species, such as oil palm.

SP1 encodes an E3 ubiquitin ligase located on the outer membrane of plastids. As fatty acids are synthesized in plastids, understanding plastid regulation in plants particularly in the crop itself, may help future improvement. Furthermore, based on the work in tomato and Arabidopsis, it was proposed that SP1 may be useful for control of fruit ripening as well as the improvement of plant's tolerance to abiotic stress (Ling et al 2021: Ling & Jarvis, 2015b). Before future functional analysis of this gene and subsequent efforts on the improvement of the crop can be carried out, it is important to identify and isolate the SP1 homolog in oil palm. The current study aimed at isolating the coding sequence of oil palm *SP1* (*EgSP1)* and investigate the expression pattern of its transcripts in selected vegetative tissues to help understand the role of this gene in plastid biogenesis in oil palms.

5.1 Identification of SP1 homolog from oil palm

In this study, one putative oil palm SP1 homolog (named as *EgSP1*) was successfully identified from the online sequence database NCBI and PalmXplore. *EgSP1* encodes two SP1 isoforms, X1 and X2, and these two isoforms only differ in 2 amino acids at the position ~274–275 of the SP1 isoform X1 amino acid sequence. The identified homolog contains the conserved domain of C_3HC_4-type (RING-finger) similar to the SP1 isolated from tomato (Ling et al., 2021), Arabidopsis (Ling et al., 2012) and wheat (Sadali, 2018). The identified EgSP1 also share ~73% sequence identity with Arabidopsis SP1. This is similar to what have been reported for SP1 homolog found in tomato (*Solanum lycopersicum*) where the tomato SP1 (SlSP1) shared 73.3% similarity with Arabidopsis SP1 (Ling et al., 2021). In wheat, the full-length protein sequence, TaSP1 homologues shares 69.1-69.4% identity with AtSP1 (Sadali, 2018). Furthermore, multiple sequence alignment of EgSP1 with other putative SP1 homologs from other plants revealed that the C_3HC_4-type (RING-finger) domain was conserved in all of the identified SP1, where amino sequence alignment showed no variation within the RING-finger motif (Figure 4.4) similar to what have been reported by Joazeiro & Weissman, (2000). The C_3HC_4-type conserved domain is known to be found in many key regulatory proteins including SP1 in plants and the outgroup MULAN in humans (Yuan et al., 2013). Both SP1 and MULAN are known to be found on the outer membrane of chloroplasts and mitochondria, respectively, where the C_3HC_4-type domain often exposed to the cytosol and is possibly kept conserved by nature due to its particular role in substrate binding (Berg and Shi, 1996; Yuan et al., 2013; Kessler, 2012; Ling et al., 2012).

To obtain information on the importance of SP1 in plants, *in silico* searches of putative SP1 homologs was done. It was found that SP1 occurs and can be identified in 26 plants from 12 plant orders suggesting that SP1 is conserved through the plant kingdom. The current set of data was limited to mostly crops plant due to the context of

the study being related to the possible use of SP1 for crop improvements. The actual data set found for identified putative SP1 containing organisms from the NCBI library was much broader with inclusion of hardwood plant and even non-angiosperm such as *Lycophytes* which was include into the phylogenetic tree and an outgroup for the angiosperms. One interesting topic that could be explored, is whether single-celled photosynthetic organism such as algae also express SP1 for the regulation of its plastids, and what are the effect SP1 manipulation to these organisms.

Phylogenetic analysis of EgSP1 with other putative SP1 from the *in silico* searches (as well as tomato SP1 identified by Ling et al., (2021) showed that EgSP1 is grouped together with SP1 proteins from the monocot species and was closest to banana SP1. The initially in a previous constructed phylogenetic tree, the closest relative of EgSP1 was identified to be the date palm from the Arecales order similar to that of oil palm but was subsequently removed following an update in NCBI database. The current closest relative to EgSP1 is from banana (*Musa acuminata*), where both of these plant species fall under the Commelinids clade along with other monocot plants in the Poales order such as wheat, sorghum, and maize (D'Hont et al., 2012).

5.2 Isolation and sequencing of *EgSP1* coding sequence

Following the identification of the oil palm SP1 isoforms, PCR primers were successfully designed to target the *SP1* coding sequence. Variants X1 and X2 have very high sequence identity (99.4%) and are identical throughout the sequence except for the 6 bp located between ~1168–1174 bp of the X1 variant transcript sequence. Primers designed to isolate the full length CDS were not able to differentially amplify the two variants. The isolation of *SP1* coding sequence was performed via PCR reaction using a high fidelity Taq polymerase (Q5-High Fidelity Taq polymerase) on cDNA derived from

leaf tissue sample. The PCR reactions resulted in the isolation of single amplicon around the size of ~1000bp. The product was then successfully cloned into a pGEM-T Easy Vector and transformed into DH5α *E. coli* host cells. From seven *E. coli* colonies sequenced, only a single variant sequence type (X1) was able to be isolated from oil palm (Figure 4.14). All of the sequence contained the extra bases that are lacking in the X2 variant (5'-GAGCAG-3'). It is possible that the X1 variant is more predominantly expressed in oil palm, specifically in the leaf tissues. With the absence of the X2 variant within the current tested samples, the question remains if there were any implications to the absence of the said variant. The X1 and X2 variants are known to be largely similar with close to 99% similarity, which the 2 variants only differ in around 6 bp in length which translates to around 2 amino acids. The sequence position of the 6 bp missing in the X2 variant was investigated, and it was discovered that the sequence did not fall on any functional domain, making the difference between the 2 variants presumably innocuous. It is speculated that, the minor differences found between the 2 transcript variant, may had come about by alternative splicing of the *EgSP1* pre-mRNA, where as much of the sequence isolated were that of the X1 variant, it is possible X2 variant maybe had been expressed in a much lower level, in which a larger sample size may be needed to validate the presence of this variant.

In the sequence analysis performed on the plasmids from 7 colonies, it was observed that there might be 2 types of X1 variants that differ in the nucleic acid sequence at three positions (Figure 4.13). Three out of seven colonies demonstrated 100% match with the reference sequence from database (X1). The remaining colonies showed the same single nucleotide polymorphisms at 3 positions (denoted as X1- α variants). Despite this, further amino acids sequence alignment of these translated sequences showed no significant differences between the amino acid sequences read of the 7 colonies which defines the SNPs as a synonymous substitution. The SNPs observed are likely genuine variations and

not mutation generated during SP1 coding sequence isolation step as precautionary steps were taken to minimize the introduction of erroneous amplification and sequencing of the target sequence, which includes the use of Q5 high fidelity polymerase to reduce the chance of copy error during *SP1* coding sequence isolation as well as performing the sequencing step from both ends of the plasmid insert to obtain a single proofread sequence. One of the possible reasons that X1 displayed the α variants is that the X1-α could possibly be an allele that was inherited with the X1 variant. As oil palms are known to be diploids, which means they would carry at least 2 pair of each allele, the identification of alleles is not uncommon and was reported in some gene in oil palm such as the SHELL gene with multiple identified allele (Castilho, 2000; Corley and Thinker, 2015; Singh et al., 2020). Where an amino acid sequence alignment of the translated X1 and X1-α variants revealed no significant difference between the 2 variants. Likely the X1-α variant would been inherited without much harm to the plant due to its redundancy with the X1 variant. It should be noted that the isolation of *SP1* coding sequence was only performed once due to the assumption that *in vitro* samples would carry the closely similar genomic material for each biological samples. Nevertheless, the presence of *SP1* X2 transcripts is still possible in oil palm tissues, and further repeats using more biological materials such as different tissues other than the leaves would be recommended.

5.3 Expression pattern analysis of *EgSP1* transcript in different oil palm tissues

SP1 has been studied in two plant species including *Arabidopsis thaliana* (Ling et al., 2012) and tomato (Ling et al., 2021). In these studies, the expression of SP1 was suggested to be tied to developmental transitions in plants such as during fruit ripening, seed germination and senescence, where plant tissues undergoing developmental changes would require an increased expression of *SP1* to aid in the differentiation process through the rearrangement of the TOC machinery (Ling et al., 2012; Sadali, 2018; Ling et al.,

2021). The current study findings on the expression pattern of *EgSP1* transcript (variant X1) for vegetative tissues (leaf, stem, and root) of oil palm (Figure 4.19) showed that SP1 was highly expressed in the leaf tissue, followed by the roots and the lowest expression was found in the stem. The higher expression of *EgSP1* the leaf tissues as compared to the roots and stem were in line with the expected results, as the leaf samples being used were from those of 1 month old *in vitro* plantlets which are still active in their development. During this stage, the leaves are undergoing developmental changes which entails the increase imports of photosynthetic precursor proteins for the development of chloroplast. With the increased need for TOC machinery turnover for chloroplast development, expression of *SP1* was reasonably expected to be higher in the leaf tissues. In comparison with other plants such as in tomato, the expression of *SP1* in the root and stem tissues oil palm were found to be slightly different, where the expression of *SlSP1* (tomato homolog of *SP1*) in root and stem tissues demonstrated similar level of expression (Sadali, 2018; Ling et al., 2021). A possible hypothesis being proposed is that sample conditioning could be a factor for the higher expression of *EgSP1* observed in the root. In this study, plant material used were from 1-month-old *in vitro* plantlets which were grown in a laboratory condition. *In vitro* plantlets are generally grown on media in glass vessel which the roots are periodically exposed to light (tissue culture light conditioning). Plant roots are known to be sensitive to light stimuli and are observed to display negative phototropism (increased growth root and aversion to light source) as well as light induced metabolism of plants secondary metabolism (Yokawa et al., 2011; Hemm et al., 2004). In a study on roots response to light stimuli, Arabidopsis roots under short exposure (10s) to blue light (e.g fluorescent light) elicit an immediate and abundant generation of reactive oxygen species (ROS) which also activates enzymes relating to adaptive stress response towards light (Yokawa et al., 2014; Yokawa et al., 2011; Chan et al., 2016). That being said, SP1 is known to be involved in plant stress response through

the degradation of TOC machinery components to reduce the import of photosynthetic precursor protein imports under stressful condition (e.g osmotic and chloroplast ROS) (Ling & Jarvis, 2016; Ling & Jarvis, 2015b; Woodson, 2016). It is plausible that the increased expression of *SP1* transcripts in the root compared to stems may had been a by-product of a non-specific systemic stress response, although further studies relating to *SP1* transcriptional regulation in needed to validate this hypothesis. The current study on *SP1* transcriptional expression is mostly limited to those of the vegetative tissues and from *in vitro* samples, where further studies should look towards investigating the expression of *SP1* in both oil palm fruits and flowers which may be useful for future oil palm improvement.

5.4 Challenges and prospect of oil palm biological improvements

The current study's goal is to isolate and analyse the expression of *SP1,* which would serve as a foundation for future research into the use of *SP1* in the biotechnological improvements of oil palm. In the previous literature section, we mentioned how biotechnology could be used to improve the oil properties of oil palm. In this section, we will discuss the obstacles encountered in the pursuit of oil palm enhancements through non-conventional techniques, as well as investigate what other successful techniques have been used to increase oil yield observed in other oil-bearing crops. With the predicted global rise in demand of oil palm, the palm oil industry must increase its production to remain competitive in the market. The improvement in oil yield production is an important endeavour for maintaining competitiveness as wells as to promote the sustainable used of agricultural land, as the opening of new land for agriculture becomes less sustainable over time (Murphy et al., 2021; Vega-Sánchez & Ronald, 2010; Shevade & Loboda, 2019). The first implementation of genetic transformation in oil palm had occurred more than 25 years ago (Masani et al., 2018). Where over time, the potential of

conventional breeding methods was slowly nearing its limits, genetic transformation research is gradually being push forward as the primary method for manipulating oil quality and yield. Genetic transformation is a potent and effective method for production of new transgenic plant with novel traits (Yarra et al., 2019). Transgenic plant development is essentially dependent on the steady incorporation of foreign DNA into the host genome using techniques which include particles bombardment, *Agrobacterium*-mediated transformation, and microinjection of DNA into protoplast (Yarra et al., 2019).

In most studies involving the improvement of oil yield in oil-bearing plants, smaller plants like soybean and canola are often chosen since the transformation of larger plants such as oil palms is fraught with obstacles that impede their development. First, oil palm plants are perennial plants with a long regeneration period, where the introduction of a new trait frequently faces a bottleneck of a long waiting period before a stable generation of transformed plants can be established (Masani et al., 2018; Yarra et al., 2019). In general compared to other oil-bearing plants, oil palm has a relatively low transformation efficiency of between 0.7% and 1.5% (Masani et al., 2018). Due to the low response of oil palm explants to culture media, selection and regeneration are relatively more difficult, as oil palm callogenesis and embryogenesis are extremely sluggish processes. Where callus formation could take up to 12 months, followed by 18 to 24 months of regeneration for a plantlet of sufficient size (~15 cm) could be obtained (Masani et al., 2010). Despite all of this, scientists must also consider the possibility that the plant will silence the transgene. This process is associated primarily with the defence against potential pathogens, particularly viruses, by preventing the expression of foreign genes via destruction their transcripts (Shih et al., 2016).

Currently, there are several known successful studies involving the enhancement of oil yield in oil-bearing crops, although they primarily focused on oil bearing-seed crops. These studies frequently concentrate on the manipulation of the triacylglycerol synthesis

pathway, in which the enzymes in the pathway are overexpressed, which causes the observed increase in oil production (Song, 2017; Miguel, 2010). The best known of these genes is the diacylglycerol acyltransferase (DGAT), which catalyses the final acyl-CoA dependent acylation. Field studies performed on canola and soybean resulted in an increase in the seed oil yield of these crops. Where in transgenic canola the oil content increased from 2.5% to 7% on a dry matter basis from the introduction of DGAT1 from both *A. thaliana* and *B. napus* (Taylor et al., 2009). Meanwhile, transgenic soybean, with the DGAT1 variant increases seed oil content by 3% (Roesler et al., 2016). Both of this study highlights the potential of oil yield improvements through the manipulation of the triacylglycerol synthesis pathways. In plants, fatty acids which servers as substrates to the triacylglycerol pathways occurs within the bounds of the plant plastid (Song, 2017). We hypothesized that the overexpression of SP1 would improve the differentiation and biogenesis efficiency of the plastids such as chloroplast and elaioplast (currently investigated in different study), which would likely show improvements in both the oil yield as well as the storage of lipids. Though more functional study on SP1 is needed to provide a proof of concept of the current proposed hypothesis.

CHAPTER 6: CONCLUSIONS AND FUTURE WORK

6.1 Conclusion

The current study was aimed at the isolation and characterization of *SP1* coding sequence in oil palm to help act as a steppingstone for other studies to build upon with the larger goal of SP1 utilization in the improvement oil palm overall productivity. In a brief take of this study, data mining for putative SP1 homolog resulted in identification of a single homolog that encodes two protein isoforms designated as EgSP1 isoform X1 and X2. Following this, gene isolation and cloning was successfully done using gene specific primers for X1 variant that has 100% sequence identity (amino acid level) with the reference sequence of oil palm from the database. In addition, expression analysis was performed on vegetative tissue of the oil palm where the *SP1* X1 transcripts was shown to be highly expressed in the leaves followed by roots and least expressed in the stems.

6.2 Future work

In order to confirm the function of *EgSP1* in regulating plastid biosynthesis in oil palm, functional analysis must be performed. One of the ways to do this is through complementation analysis. In this experiment, mutant Arabidopsis with knock downed expression of *SP1* gene can be transformed using *Agrobacterium tumefaciens* with a binary vector carrying a promoter and the previously isolated *EgSP1* coding sequence. The identity and functionality can be confirmed when the isolated oil palm SP1 (*EgSP1)* is able to reverse the mutant phenotype and compensate the loss of endogenous *SP1*.

Another study that could be looked into is the generation of seed-specific overexpressor of *EgSP1* in *Arabidopsis thaliana* for the study of elaioplast as well as lipid

synthesis. SP1 is known to be involved in plastid biogenesis and differentiation in plants, where overexpression of SP1 in Arabidopsis was found to positively affect plastid differentiation rate during de-etiolation and senescence (Ling et al 2012) and as well as fruit ripening in tomato (Ling et al., 2021). Where studies on etioplast are generally sparse, this proposed study would look into whether the overexpression of SP1 using the *EgSP1* coding sequence would affect elaioplast efficiency in accumulating and synthesising lipids in the seeds of Arabidopsis as compared to the wild type. Any positive findings on improvement in lipid quality or accumulation would hopefully be able to connect back to oil palm where it could be used for the further improvement of oil palm and other crops as a whole.

REFERENCES

Agne, B., & Kessler, F. (2009). Protein transport in organelles: The TOC complex way of preprotein import. *The FEBS Journal, 276*(5), 1156-1165.

Alam, A. F., Er, A. C., & Begum, H. (2015). Malaysian oil palm industry: Prospect and problem. *Journal of Food, Agriculture & Environment, 13*(2), 143-148.

Amin, L., Md Jahi, J., & Md Nor, A. R. (2013). Stakeholders' attitude to genetically modified foods and medicine. *The Scientific World Journal, 2013*, 1-14.

Bauer, J., Chen, K., Hiltbunner, A., Wehrli, E., Eugster, M., Schnell, D., & Kessler, F. (2000). The major protein import receptor of plastids is essential for chloroplast biogenesis. *Nature, 403*(6766), 203-207.

Barcelos, E., Rios, S. D. A., Cunha, R. N., Lopes, R., Motoike, S. Y., Babiychuk, E., ... & Kushnir, S. (2015). Oil palm natural diversity and the potential for yield improvement. *Frontiers in Plant Science, 6*, Article#190.

Berg, J. M., & Shi, Y. (1996). The galvanization of biology: A growing appreciation for the roles of zinc. *Science, 271*(5252), 1081-1085.

Bräutigam, A., & Weber, A. P. (2009). Proteomic analysis of the proplastid envelope membrane provides novel insights into small molecule and protein transport across proplastid membranes. *Molecular Plant, 2*(6), 1247-1261.

Broad, W., Ling, Q., & Jarvis, P. (2016). New insights into roles of ubiquitin modification in regulating plastids and other endosymbiotic organelles. *International Review of Cell and Molecular Biology, 325*, 1-33.

Budiani, A., Putranto, R. A., Riyadi, I., Minarsih, H., & Faizah, R. (2018). Transformation of oil palm calli using CRISPR/Cas9 System: Toward genome editing of oil palm. *Earth and Environmental Science, 183*, 1-7.

Castilho, A., Vershinin, A., & Heslop-Harrison, J. S. (2000). Repetitive DNA and the chromosomes in the genome of oil palm (*Elaeis guineensis*). *Annals of Botany, 85*(6), 837-844.

Chan, K. X., Phua, S. Y., Crisp, P., McQuinn, R., & Pogson, B. J. (2016). Learning the languages of the chloroplast: Retrograde signalling and beyond. *Annual Review of Plant Biology, 67*, 25-53.

Chu, C. C., & Li, H. M. (2018). Developmental regulation of protein import into plastids. *Photosynthesis Research, 138*(3), 327-334.

Corley, R., & Tinker, P. (2015). *The oil palm* (5th ed.) Malden, MA: Wiley-Blackwell.

Demarsy, E., Lakshmanan, A. M., & Kessler, F. (2014). Border control: Selectivity of chloroplast protein import and regulation at the TOC-complex. *Frontiers in Plant Science, 5*, Article#483.

D'hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., ... & Wincker, P. (2012). The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. *Nature, 488*(7410), 213-217.

Dislich, C., Keyel, A., Salecker, J., Kisel, Y., Meyer, K., & Auliya, M. et al. (2016). A review of the ecosystem functions in oil palm plantations, using forests as a reference system. *Biological Reviews, 92*(3), 1539-1569.

Ghulam, K. A., Hishamuddin, E., Soh Kheang, L., Ong-Abdullah, M., Mohamed Salleh, K., & Izuddin Zanal Bidin, M. et al. (2020). Oil palm economic performance in Malaysia and R&D progress in 2019. *Journal of Oil Palm Research, 32*(2), 159-190.

Guerra, D. D., & Callis, J. (2012). Ubiquitin on the move: The ubiquitin modification system plays diverse roles in the regulation of endoplasmic reticulum-and plasma membrane-localized proteins. *Plant Physiology, 160*(1), 56-64.

Hall, B. G. (2013). Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution, 30*(5), 1229-1235.

Hayashi, M., & Nishimura, M. (2009). Frontiers of research on organelle differentiation. *Plant and cell physiology, 50*(12), 1995-1999.

Hemm, M. R., Rider, S. D., Ogas, J., Murry, D. J., & Chapple, C. (2004). Light induces phenylpropanoid metabolism in Arabidopsis roots. *The Plant Journal, 38*(5), 765-778.

Hirosawa, Y., Ito-Inaba, Y., & Inaba, T. (2017). Ubiquitin–proteasome-dependent regulation of bidirectional communication between plastids and the nucleus. *Frontiers in Plant Science, 8*, Article#310.

Horiike, T. (2016). An introduction to molecular phylogenetic analysis. *Reviews in Agricultural Science, 4*, 36-45.

Jarvis, P., & Lopez-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nature Reviews Molecular Cell Biology, 14*(12), 787-802.

Joazeiro, C. A., & Weissman, A. M. (2000). RING finger proteins: Mediators of ubiquitin ligase activity. *Cell, 102*(5), 549-552.

Kessler, F., & Schnell, D. (2009). Chloroplast biogenesis: diversity and regulation of the protein import apparatus. *Current Opinion in Cell Biology, 21*(4), 494-500.

Kessler, F. (2012). Chloroplast delivery by UPS. *Science, 338*(6107), 622-623.

Kushairi, A., Singh, R., & Ong-Abdullah, M. (2017). The oil palm industry in Malaysia: Thriving with transformative technologies. *Journal of Oil Palm Research, 29*(4), 431-439.

Kushairi, A., Soh Kheang, L., Azman, I., Hishamuddin, E., Ong-Abdullah, M., & Noor Izuddin, Z. et al. (2018). Oil palm economic performance in Malaysia and R&D progress in 2017 - Review article. *Journal Of Oil Palm Research, 30*(2),163-195.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), Article#1547.

Lee, S., Lee, D. W., Lee, Y., Mayer, U., Stierhof, Y. D., Lee, S., ... & Hwang, I. (2009). Heat shock protein cognate 70-4 and an E3 ubiquitin ligase, CHIP, mediate plastid-destined precursor degradation through the ubiquitin-26S proteasome system in Arabidopsis. *The Plant Cell, 21*(12), 3984-4001.

Lee, D. W., Jung, C., & Hwang, I. (2013). Cytosolic events involved in chloroplast protein targeting. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1833*(2), 245-252.

Li, H. M., & Chiu, C. C. (2010). Protein transport into chloroplasts. *Annual Review of Plant Biology, 61*, 157-180.

Ling, Q., Huang, W., Baldwin, A., & Jarvis, P. (2012). Chloroplast biogenesis is regulated by direct action of the ubiquitin-proteasome system. *Science, 338*(6107), 655-659.

Ling, Q., & Jarvis, P. (2015a). Functions of plastid protein import and the ubiquitin–proteasome system in plastid development. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1847*(9), 939-948.

Ling, Q., & Jarvis, P. (2015b). Regulation of chloroplast protein import by the ubiquitin E3 ligase SP1 is important for stress tolerance in plants. *Current Biology, 25*(19), 2527-2534.

Ling, Q., & Jarvis, P. (2016). Plant signalling: Ubiquitin pulls the trigger on chloroplast degradation. *Current Biology, 26*(1), R38-R40.

Ling, Q., Sadali, N. M., Soufi, Z., Zhou, Y., Huang, B., Zeng, Y., ... & Jarvis, R. P. (2021). The chloroplast-associated protein degradation pathway controls chromoplast development and fruit ripening in tomato. *Nature Plants, 7*(5), 655-666.

Khan, N. T. (2017). MEGA-core of phylogenetic analysis in molecular evolutionary genetics. *Journal Phylogen Evolution Biology, 5*(2), Article#1000183.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods, 25*(4), 402-408.

Low, E. T. L., Jayanthi, N., Chan, K. L., Sanusi, N. S. N. M., Halim, M. A. A., Rosli, R., & Kushairi, A. (2017). The oil palm genome revolution. *Journal Oil Palm Research, 29*(4), 456-468.

Markus, L. B., & Turner, G. W. (2013). Terpenoid biosynthesis in trichomes-current status and future opportunities. *Plant Biotechnology Journal, 11*(1), 2-22.

Masani, M., Izawati, A., Rasid, O., & Parveez, G. (2018). Biotechnology of oil palm: Current status of oil palm genetic transformation. *Biocatalysis and Agricultural Biotechnology, 15*, 335-347.

Mohammed, M. A. A., Salmiaton, A., Azlina, W. W., Amran, M. M., Fakhru'l-Razi, A., & Taufiq-Yap, Y. H. (2011). Hydrogen rich gas from oil palm biomass as a potential source of renewable energy in Malaysia. *Renewable and Sustainable Energy Reviews, 15*(2), 1258-1270.

Murphy, D. J. (2014). The future of oil palm as a major global crop: Opportunities and challenges. *Journal of Oil Palm Research, 26*(1), 1-24.

Murphy, D., Goggin, K., & Paterson, R. (2021). Oil palm in the 2020s and beyond: Challenges and solutions. *CABI Agriculture and Bioscience, 2*(1). 1-22.

Oleszkiewicz, T., Klimek-Chodacka, M., Milewska-Hendel, A., Zubko, M., Stróż, D., Kurczyńska, E., ... & Baranski, R. (2018). Unique chromoplast organisation and carotenoid gene expression in carotenoid-rich carrot callus. *Planta, 248*(6), 1455-1471.

Parveez, G. K. A., Rasid, O. A., Masani, M. Y. A., & Sambanthamurthi, R. (2015). Biotechnology of oil palm: Strategies towards manipulation of lipid content and composition. *Plant Cell Reports, 34*(4), 533-543.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research, 29*(9), 2002-2007.

Plöscher, M., Reisinger, V., & Eichacker, L. A. (2011). Proteomic comparison of etioplast and chloroplast protein complexes. *Journal of Proteomics, 74*(8), 1256-1265.

Richardson, L. G., & Schnell, D. J. (2020). Origins, function, and regulation of the TOC–TIC general protein import machinery of plastids. *Journal of Experimental Botany, 71*(4), 1226-1238.

Roesler, K., Shen, B., Bermudez, E., Li, C., Hunt, J., & Damude, H. et al. (2016). An improved variant of soybean type 1 diacylglycerol acyltransferase increases the oil content and decreases the soluble carbohydrate content of soybeans. *Plant Physiology, 171*, 878-893.

Sadali, N. M. (2018). *Analysis of the roles of the plastidic E3 ligases SP1 and SPL2 in tomato and wheat* (Unpublished Doctoral dissertation, University of Oxford, United Kingdom).

Sadali, N. M., Sowden, R. G., Ling, Q., & Jarvis, R. P. (2019). Differentiation of chromoplasts and other plastids in plants. *Plant Cell Reports, 38*(7), 803-818.

Sanusi, N. S. N. M., Rosli, R., Halim, M. A. A., Chan, K. L., Nagappan, J., Azizi, N., ... & Low, E. T. L. (2018). PalmXplore: Oil palm gene database. *Database, 2018*, 1-9.

Shevade, V., & Loboda, T. (2019). Oil palm plantations in Peninsular Malaysia: Determinants and constraints on expansion. *PLOS ONE, 14*(2), 1-22.

Shih, P., Liang, Y., & Loqué, D. (2016). Biotechnology and synthetic biology approaches for metabolic engineering of bioenergy crops. *The Plant Journal, 87*(1), 103-117.

Singh, R., Low, E. T. L., Ooi, L. C. L., Ong-Abdullah, M., Ting, N. C., Nagappan, J., ... & Martienssen, R. A. (2013). The oil palm *SHELL* gene controls oil yield and encodes a homologue of *SEEDSTICK*. *Nature, 500*(7462), 340-344.

Singh, R., Low, E. T. L., Ooi, L. C. L., Ong-Abdullah, M., Ting, N. C., Nookiah, R., ... & Sambanthamurthi, R. (2020). Variation for heterodimerization and nuclear localization among known and novel oil palm *SHELL* alleles. *New Phytologist, 226*(2), 426-440.

Sjuts, I., Soll, J., & Bölter, B. (2017). Import of soluble proteins into chloroplasts and potential regulatory mechanisms. *Frontiers in Plant Science, 8*, Article#168.

Smolikova, G. N., & Medvedev, S. S. (2016). Photosynthesis in the seeds of chloroembryophytes. *Russian Journal of Plant Physiology, 63*(1), 1-12.

Song, Y., Wang, X., & Rose, R. (2017). Oil body biogenesis and biotechnology in legume seeds. *Plant Cell Reports, 36*(10), 1519-1532.

Solymosi, K., Lethin, J., & Aronsson, H. (2018). Diversity and plasticity of plastids in land plants. *Plastids, 1829*, 55-72.

Solymosi, K., & Aronsson, H. (2013). Etioplast and their significance in chloroplast biogenesis. In Plastid development in leaves during growth and senescence. *Advances in Photosynthesis and Respiration, 36*, 39-71.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques, 28*(6), 1102-1104.

Sundram, K., Sambanthamurthi, R., & Tan, Y. A. (2003). Palm fruit chemistry and nutrition. *Asia Pacific Journal of Clinical Nutrition, 12*(3), 355-362.

Tang, K., & Al Qahtani, H. (2019). Sustainability of oil palm plantations in Malaysia. *Environment, Development and Sustainability, 22*(6), 4999-5023.

Taylor, D., Zhang, Y., Kumar, A., Francis, T., Giblin, E., & Barton, D. et al. (2009). Molecular modification of triacylglycerol accumulation by over-expression of DGAT1 to produce canola with increased seed oil content under field conditions. *Plant Biotechnology Institute. Botany, 87*(6), 533-543.

Vega-Sánchez, M., & Ronald, P. (2010). Genetic and biotechnological approaches for biofuel crop improvement. *Current Opinion in Biotechnology, 21*(2), 218-224.

Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nature Reviews Molecular Cell Biology, 10*(6), 385-397.

Vigil, E. L., & Ruddat, M. (1985). Development and enzyme activity of protein bodies in proteinoplast of tobacco root cells. *Histochemistry, 83*(1), 17-27.

Watson, S. J., Sowden, R. G., & Jarvis, P. (2018). Abiotic stress-induced chloroplast proteome remodelling: a mechanistic overview. *Journal of Experimental Botany, 69(*11), 2773-2781.

Woodson, J. D. (2016). Chloroplast quality control-balancing energy production and stress. *New Phytologist, 212*(1), 36-41.

Woittiez, L. S., van Wijk, M. T., Slingerland, M., van Noordwijk, M., & Giller, K. E. (2017). Yield gaps in oil palm: A quantitative review of contributing factors. *European Journal of Agronomy, 83*, 57-77.

Yarra, R., Jin, L., Zhao, Z., & Cao, H. (2019). Progress in tissue culture and genetic transformation of oil palm: An overview. *International Journal of Molecular Sciences, 20*(21), Article#5353.

Yarra, R., Cao, H., Jin, L., Mengdi, Y., & Zhou, L. (2020). CRISPR/Cas mediated base editing: A practical approach for genome editing in oil palm. *Biotech, 10*(7), 1-7.

Yokawa, K., Kagenishi, T., Kawano, T., Mancuso, S., & Baluška, F. (2011). Illumination of Arabidopsis roots induces immediate burst of ROS production. *Plant Signalling & Behavior, 6*(10), 1460-1464.

Yokawa, K., Fasano, R., Kagenishi, T., & Baluška, F. (2014). Light as stress factor to plant roots–case of root halotropism. *Frontiers in Plant Science, 5*, Article#718.

Yeap, W. C., Loo, J. M., Wong, Y. C., & Kulaveerasingam, H. (2014). Evaluation of suitable reference genes for qRT-PCR gene expression normalization in reproductive, vegetative tissues and during fruit development in oil palm. *Plant Cell, Tissue, and Organ Culture, 116*(1), 55-66.

Yin, K., Gao, C., & Qiu, J. L. (2017). Progress and prospects in plant genome editing. *Nature Plants, 3*(8), 1-6.

Yuan, X., Zhang, S., Liu, S., Yu, M., Su, H., Shu, H., & Li, X. (2013). Global analysis of ankyrin repeat domain C_3HC_4-type RING finger gene family in plants. *PloS One, 8*(3), 1-11.

Zhu, M., Lin, J., Ye, J., Wang, R., Yang, C., Gong, J., ... & Zeng, Y. (2018). A comprehensive proteomic analysis of elaioplast from citrus fruits reveals insights into elaioplast biogenesis and function. *Horticulture Research, 5*. 1-11