PROTEOMICS ANALYSIS OF SOMATIC EMBRYOGENESIS IN OIL PALM (*ELAEIS GUINEENSIS* JACQ)

TAN HOOI SIN

DOCTOR OF PHYLOSOPHY OF SCIENCES

THE UNIVERSITY OF NOTTINGHAM MALAYSIA CAMPUS

2015

ACKNOWLEDGEMENT

First of all, I would like to express my gratitude to my main supervisor, Associate Professor Dr. Chin Chiew Foan for her guidance, comments and suggestions during my PhD study. I appreciate her valuable insight and lending me a pair of helping hand whenever I needed, especially in my laboratory work. With herhelp, I able to get financial support to Bruker Proteomic Training in Australia as well as attending Plant Proteome Conference in Tsukuba, Japan that have enhanced my knowledge.

Besides that, my sincere appreciation and gratefulness also goes to my co-supervisor, Dr. Susan Liddell from The University of Nottingham United Kingdom for her precious guidance on proteomics study. Her guidance has transformed me from a fresh graduate without knowledge in proteomic field to a protein expert. This valuable experience would allow me to expel and able to handle any challenges confidently in the future. I am very grateful to have another helpful co-supervisor from Malaysia Palm Oil Board (MPOB), Dr Meilina Ong Abdullah who is an expert plant molecular biologist in oil palm. I really appreciate her generosity by allowing me to use consumables and facilities in the MPOB in these three years.

Furthermore, I am in debt to members of Biotechnology Research Centre (BRC) from The University of Nottingham Malaysia Campus as well as Advanced Breeding Biotechnology Centre from MPOB, particularly to Miss Kok Sau Yee who always therewhenever I am in difficultly. Nevertheless, cooperation, encouragement, and support from my wonderful course and project mates are highly appreciated.I also want to extend my sincere thanks to anyone who has contributed directly or indirectly in this project as well as MPOB GSAS sponsorship for the university fees and allowance.Last but not least, I am grateful to my family for their encouragement, support and patience.

ABSTRACT

Oil palm is an important commercial crop in Malaysia where Malaysia is the second largest producer and exporter of palm oilin the world. In order to meet the increasing demand for palm oil, elite oil palm planting materials with higher palm oil yield are the desirable planting materials. Hence, the oil palm plantation companies have incorporated *in vitro* micropropagation technique through somatic embryogenesis in producing elite oil palm. However, low embryogenesis rate has hampered large production of elite oil palm ramets. In this study, proteomic technology was deployed to compare protein expression and identify differential expressed protein between high and low proliferated embryogenic lines of oil palm tissue culture. From the study, total protein of oil palm young and old leaves was extracted using an optimized trichloroacetic acid/acetone precipitation protocol followed by polyethylene glycol (PEG) fractionation to isolate low abundance proteins. Then, the extracted proteins were separated on two-dimensional (2D) gel electrophoresis and protein profiles between the high and low proliferated embryogenic lines were compared. Total of 40 differentially expressed protein spots were isolated from the 2D gel for mass spectrophotometry (MS/MS) identification. However, only 26 out of 40 protein spots were identified and just 8 of the identified protein spots were isolated from young leaves. Quantitative real-time PCR were conducted on 17 proteins candidates to study on the relationship between the protein and mRNA expression level. There was 29% of the 17 proteins' expression showed linear correlation with their mRNA expression. These proteins candidates were highlighted for further validation in the future.

.

TABLE OF CONTENTS

LIST OF TABLES

Pages

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

%	= Percentage
°C	= Degree Celcius
µl	= Microlitre
µM	= Micromolar
µg	= Microgram
g	= Gram
L	= Litre
m	= Metre
M	= Molar
mg	= Milligram
Min	= minute
ml	= milliliter
mm	= millimeter
cm	= Centimeter
n	= Number of sample
v/v	= Volume per volume
w/v	= Weight per volume
Gb	= Gigabase
kDa	= Kilo Dalton

CHAPTER 1
INTRODUCTION

The oil palm, *Elaeis guineensis*, native to West Africa and was first introduced to Malaysia as an ornamental plant in 1870. Since then, the oil palm plantation in the country increased rapidly. Tennamaran Estate in Selangor was the first commercial oil palm estate established in Malaysia in 1917. Malaysia was the first country to invest in large-scale planting of oil palm and processing of its products. Today, Malaysia is one of the world's largest producers of palm oil after Indonesia (Pakiam, 2013). Cultivation of oil palm in Malaysia has expanded rapidly as palm oil becomes a major source of sustainable and renewable raw material for the world's food and biofuel industries (Mosarof *et al.,* 2015).

Over the past 50 years, research and development (R&D) activities and advances in technology have played important roles in increasing palm oil production (Basiron *et al.,* 2007). Many studies have been carried out to improve the yield of oil palm especially through cross breeding between species, tissue culture, and genetic engineering in order to improve the yield and quality of palm oil. Genetic improvement of oil palm is time consuming and costly due to the 10 years long breeding cycle. This brought about the interest in vegetative propagation of oil palm. Hence, commercial propagation of oil palm through tissue culture has played a vital role to increase the production of quality ramets through rapid multiplication of uniform planting materials with desired characteristics.

Currently, the market has demand more than 100 million tissue culture plantlets in Malaysia and other countries (Zamzuri *et al.,* 1999). Tissue culture methods are linked to an effective oil palm breeding with desired explants. Based on

the field performance data, clonal materials has increased oil yield between 20-30% compared to seedling planting materials (Soh *et al.,* 2001).

However, some challenges still exist in this technology. For examples, the low efficiency of callusing and embryogenesis exist in large scale micropropagation of oil palm. Even though tissue culture materials have been subjected to optimized conditions, the conversion rate of oil palm explant to callus is reported to be only 19% while the rate of embryogenic competent callus to embryos is as low as 6% (Kushairi *et al.,* 2010). Such low conversion rate causes major problems in tissue culture laboratories especially due to increases in the expenditure in terms of culture media, growth regulators and also electricity consumption.

Despite much understanding is achieved in the practical aspects of somatic embryogenesis to increase production, the mechanisms involved in somatic embryogenesis of plants have much more to explore and investigate. Hence, the key questions focus on the mechanism that causes cells to change their destiny and become embryogenic. Many studies have focused on gene expression during somatic embryogenesis (Zuo *et al.,* 2002; Boutilier *et al.,* 2002), transcription factors that participate in somatic embryogenesis (Stone *et al.,* 2001, 2008) and potential molecular markers for embryogenesis competent cells (Schmidt *et al.,* 1996; McCabe *et al.,* 1997; Braybrook and Harada, 2008).

Vital cellular functions in somatic embryogenesis require coordinated actions of a large number of proteins that interact together in protein-protein interaction networks and it needs to be studied especially with the dynamic behavior of cells at a particular time and place. More recently, proteomic techniques have evolved as a powerful application in the study of proteins expressed within a cell, tissue or organism. The use of high resolution two-dimensional electrophoresis for protein separation combined with identification of proteins by advanced instrument such as mass spectrometry is a powerful and reliable approach to study a large diversity of protein molecules. In this project, proteomic approach was employed to study the

differential protein expression of somatic embryogenesis in tissue culture of oil palm with the aim of identifying those proteins that are differentially expressed in embryogenic oil palm explants.

In Chapter 3, the optimal protein extraction method for oil palm leaf samples was identified. In this study, two main materials, namely the mature and young leaf samples were studied. With the best extraction method selected, the proteins were extracted and separated using two-dimensional gel electrophoresis (2-DE). There were a few parameters for 2-DE gel running optimized. The 2-DE protein gels for two categories (high proliferation rate and low proliferation rate samples) were compared using the PDQuest software and protein species were identified using mass spectrometry and discussed in Chapter 4.

After comparison, most of the differential proteins identified were involved in photosynthetic roles and present in higher abundance in the leaf samples. It is not surprised as total protein extraction will extract most of the abundance proteins. Yet, in Chapter 5, with the aims to improve protein species between the two categories (high proliferation rate and low proliferation rate samples), polyethylene glycol (PEG) based fractionation was carried out to remove the high abundance proteins to allow low abundance protein to be shown in the 2-DE gels. PEG fractionation is able to provide a supplement data to the total protein extraction method and allows more differential protein species to be identified.

Lastly, there were a total of 12 protein species selected to undergo the real time reverse transcription-polymerase chain reaction (RT-qPCR) to study on their mRNA expression in Chapter 6. The protein candidates presenting a concordance between their protein abundance and mRNA expression were highly recommended as a potential biomarker to differentiate the high and low proliferation rate samples in oil palm ortet prior to tissue culture.

Objectives of the study:

The aim of this study is to deploy a proteomic approach to determine the expression profiles of oil palm callus tissue that displays low and high regeneration rates in order to find the major proteins that take part in the induction of plant regeneration in oil palm. The output from this study can further be used to identify and isolate markers associated with embryogenesis of oil palm. In order to achieve the overall aim of the study, the following specific objectives were defined:

a) To optimize the protein extraction protocol and two-dimensional gel electrophoresis for oil palm leaves
b) To compare the proteome maps between oil palm with high and low proliferation rate samples
c) To study the low abundance proteins that exhibit differential expression levels in high and low proliferation rate samples
d) To investigate the corresponding mRNA expression of the differential proteins identified in oil palm leaves

CHAPTER 2
LITERATURE REVIEW

2.1 History of oil palm

The family of palms, the Arecaceae, are placed in the order Arecales (Cronquist, 1981) which comprises three accepted species. The first two species are *Elaeis oleifera* and *Elaeis guineensis* and the third species was known previously as *Barcellaodora*, but was renamed *Elaeis odora* by Wessels-Boer (1965), *Barcellaodora* is less known because this species was less cultivated. The genus *Elaeis* was initiated on palms and received its botanical name from Jacquin (Bailey, 1933). *Elaeis* is derived from the Greek word *elaion*, meaning oil, while the specific name *guineensis* indicates its origin of the Guinea coast in West Africa.

The American oil palm (*Elaeis oleifera*) is native to central and South America while the African oil palm (*Elaeis guineensis* Jacq.) is a perennial oleaginous monocotyledonous plant, which is originated from West Africa and has been predominantly cultivated across Latin America and Southeast Asia (Konan *et al.,* 2006). In the past, African oil palm (*Elaeis guineensis* Jacq.) has been considered as food crop and consumed locally in West Africa. The fruit quality is judged by the thickness of the mesocarp layer, however, there is little effort put in to extract the oil.

The foundation of the oil palm industry is generally accredited to M. Adrien Hallet, is a Belgian with vast knowledge of oil palm plantings in Africa. He pioneered the planting of oil palms with Deli origin, Africa in 1911 in the first large commercial plantation in Sumatra, Indonesia. Hallet reported that the oil palms grown in Deli are more productive and had superior fruit composition compared to the ordinary *dura* palms of the west coast. An oil content of 30 percent in the Deli fruits was recorded (Leplae, 1939).

Initially, there were 6,500 acres of oil palm planted. However, progress was slow due to insufficient information on the oil palm extraction method and probable profit obtained. The First World War has also affected the economic progress in the oil palm industries. While, M. H. Fauconnier is the person who first brought oil palm of Deli origin into Malaysia and started to plant them as ornamentals in Rantau Panjang in the Tennamaram state of Selangor with seedling obtained from 1911 and 1912 importation (Sambanthamurthi *et al.,* 2009).

The Second World War put the whole of the Far Eastern industry out of the export market and caused restrictions in shipping. After the war, there was a big fluctuation in prices for both palm oil and kernel which can be four and five times of their prewar level. The palm oil and kernel therefore increases to a record high production. In Malaysia, the acreage and planting of oil palm proceeded at a fast rate. In 1980, oil palm planted area in Malaysia achieved close to one million hectares. The total planted area for oil palm was 5.392 million hectare in year 2014 (Figure 2.1).

OIL PALM PLANTED AREA BY STATE AS AT DECEMBER 2014 (HECTARES)

State	Mature	%	Immature	%	Total	%
Johore	651,242	88.8	82,225	11.2	733,467	13.6
Kedah	80,767	93.7	5,415	6.3	86,182	1.6
Kelantan	99,783	68.9	44,979	31.1	144,762	2.7
Malacca	49,501	93.7	3,348	6.3	52,849	1.0
Negeri Sembilan	142,503	84.1	26,865	15.9	169,368	3.1
Pahang	623,269	86.6	96,344	13.4	719,613	13.3
Perak	348,794	89.6	40,370	10.4	389,164	7.2
Perlis	189	64.1	106	35.9	295	0.0
Penang	13,309	93.7	895	6.3	14,204	0.3
Selangor	126,805	91.6	11,677	8.4	138,482	2.6
Terengganu	139,410	82.5	29,538	17.5	168,948	3.1
Peninsular Malaysia	2,275,572	91.6	341,762	8.4	2,617,334	48.5
Sabah	1,355,541	89.7	155,969	10.3	1,511,510	28.0
Sarawak	1,058,208	83.8	205,183	16.2	1,263,391	23.5
Sabah & Sarawak	2,413,749	87.0	361,152	13.0	2,774,901	51.5
MALAYSIA	**4,689,321**	**87.0**	**702,914**	**13.0**	**5,392,235**	**100.0**

Figure 2.1: Total planted area of palm oil in Malaysia in 2014 (Million hectares)
(Sourced from: http://www.mpob.org)

To date, oil palm has become a valuable commercial crop in Southeast Asia, especially in Malaysia and Indonesia. Both countries are now the principal exporters of palm oil. Palm oil has been labeled as the largest international traded vegetable oil globally in the main market in China, European Union, Pakistan, India, Japan and Bangladesh (Soh *et al.,* 2009).

Malaysia has an established oil palm industry. The infrastructure of the oil palm industry was greatly developed to improve the market value of palm products. There are local refineries to process, fractionate, and extract the oil from the oil palm fruits. In order to minimize the pollutants caused by the oil palm mills, reliable and proper methods of effluent disposal according to the recommendation of environmental impact assessment of mills are always in practice.

2.2 Botany of Oil Palm

2.2.1 The plant

Elaeis guineensis (E. guineensis) is a large feather-palm with short internodes at solitary columnar stem. Short spines are formed on the leaf base and within the fruits bunch. Irregular set of leaflets on the leaf becomes the distinct characteristic of the *E. guineensis*. It grows well in the humid tropics in the coastal belt between 10 degrees north latitude and 10 to 20 degrees south latitude with small rainfall (Hartley, 1988). Oil palm trees are single- stemmed and can reach 18-24 m in height in the wild, but rarely more than 10 m in cultivation (referred to Figure 2.2(a)).

The leaves are pinnate and up to 7 m in length with 200-300 leaflets per leaf. Leaflets take over 2/3 of the leaf and the remaining lower part of leaves is filled with spines which increase in length acropetally. Thirty leaves per annum will grow on a young palm tree and the production of leaves decreases by 30% when the palm trees are over 10 years old. Like other palms, early growth of the oil palm focuses on the formation of the stem base rather than internode elongation. The rate of extension of the stem depends on both the environment and hereditary factors. Stem growth will be slower under shaded and low temperature conditions. Under standard plantation conditions, palm height increases by 0.3 to 0.6 m per annum.

Figure 2.2: (a) *Elaeis guineensis*, oil palm tree (b) Oil palm fruit
(c) Oil palm male and female flowers

(Sources: http://toptropicals.com/cgibin/garden_catalog/cat.and
http://www.gourmetindia.com/topic/526-konkan-fruit-festival-in-goa)

2.2.2 The fruits

Oil palm fruits (referred to Figure 2.2 (b)) known as drupes, vary in shape from nearly spherical to ovoid or elongated at the top and develop into compact and dense bunches. The seed is surrounded by a thick layer of mesocarp which is rich in nutrient content including oil. This enables the oil palm seed to support a growing seedling for many weeks after germination. The fruits develop gradually in size. They have a small embryo palm seed and cotyledon. The cotyledon apex will subsequently enlarge to absorb food as storage for endosperm (Tomlinson, 1961).

In general, palm fruits take 5 to 6 months from pollination to reach maturity. Before the Nigerian population palm fruit ripen, they show deep violet to black in the apex and pale greenish yellow at the base. The palm fruits will grow in large bunches with a reddish appearance after maturation. The exocarp of the fruit on the outside bunch is more pigmented than the internal fruit. The mesocarp of all fruits consists of fibres which run longitudinally through the oil-bearing tissue.

The oil palm seed is the nut that remains after removal of the soft oily mesocarp from the fruits. It consists of an endocarp and one, two or three kernels. The size of the nut greatly depends on the thickness of the shell and size of the kernel and it might differ across oil palm species. Typical African *dura* nuts are 2-3 cm in length and an average of 4 g in weight. Deli and some African *dura* nuts are larger with an average weight of 13 g. African *tenera* nuts are usually less than 2 cm in length and have an average weight of 2 g (Hartley, 1988).

2.2.3 The flowers

Oil palm exhibits monoecious with either male or female, but sometimes hermaphrodite inflorescences developing in the axils of the leaves. The flowers are in dense clusters. There are three sepals and petals each for an individual single flower. Female flowers are arranged spirally around the rachis of the spikelet and subtended by a bract; sharp spines will be formed from these bracts. There is an average of about ten flowers per spikelet in 3 year old palms, increasing to over 15 flowers after 10 years (Corley and Gray, 1976).

The male inflorescence is borne on a longer pedicle with long, finger-like, cylindrical spikelets (referred to figure 2.1(c)). Each male inflorescence contains from 25 to 100 g of fresh pollen. The pollen remains viable for at least 4 to 6 days after their release (Hardon and Turner, 1967). Thirty to sixty percent of flowers will successfully develop into fruits, depending on the pollination efficiency. Sets of bunches can carry from 500 to 4000 fruits and the fruit bunch weight increases as the palm grows older.

Some of the palm tends to produce mixed inflorescences. The mixed spikelets have the male flower at the apex and followed by the female flower at the base. Besides that, there may be pairs of male flowers present close to each other without any female flowers between them (Beirnaert, 1935). William and Thomas (1970) found that of these mixed inflorescences, hermaphrodites only occur when a male transition into the female phase during the flowering cycle, while andromorphous inflorescences occur only during the reverse change.

2.3 Industries of palm oil

Palm oil is a versatile oil with 80% of its volume contributing to food production as cooking oil, margarine, vegetable ghee and shortening. The remaining 20% is used as non-food derivatives, such as oleochemicals which function as mineral oil in the detergent, cosmetics and plastics industries. The high demand of vegetable oil has led to the expansion in hectare for planting oil palm, as it can produce high quality oil from its fruits and kernel. Crude palm oil can be obtained from the mesocarp of the fruit, while the palm kernel oil is extracted from the kernel of the nut (Vijaya *et al.,* 2010).

The cost of oil production from oil palms is the lowest among all the oil producing crops. High productivity yield of oil palm can reach up to 5-7 tonnes oil per hectare per annum under optimum agro-ecological conditions. Palm oil became prominent in the world oils and fat trade due to its minimal cost of production, large supply for exportation, competitive pricing and technically superior in terms of the wide range of uses (Basiron *et al.,* 2004). Additionally, by-products from the palm oil industry have generated many new industries. These include in the pharmaceutical field to produce antioxidant health supplements. Palm oil also serves some of the higher-value market products, such as high vitamin A and E oils. It is also involved in. oleichemicals such as bioplastics and ricinoleate oil (Kushairi and Sambanthamurthi, 2006). Even the kernel oil that is mostly used in non-edible applications such as detergents is increasingly incorporated into edible applications as high-energy sport drinks and infant food formulations (Murphy, 2007). On the other hand, the waste kernel of the oil palm can be processed as animal feeds and organic fertilizers.

Another factor for the increase demand in palm oil is due to the growth of population and wealth (Lee, 2012). Food security is the imminent issue that concerns people. Food demand is estimated to be increased by approximately 50% in year 2030 (Dyson, 1999; Pandya-Lorch *et al.,* 2001). Mielke (2001) predicted that the demand for palm oil will double by 2020 as the demand for fats and oils would increase proportionally with the demand for food.

Furthermore, with the rising trend in petroleum prices, an alternative way to substitute renewable energy instead of petroleum has been sought. Therefore the demand for palm oil, a potential source as a biofuel, is expected to increase in the future (Soh *et al.,* 2009). Production of biodiesel from superior *tenera* oil palm is not an impossible mission but has been initiated in some countries like Thailand. The "biodiesel effect" is now distorting the oil palm market, not only in Malaysia. Since the beginning of 2006, investment of $ 515 million has been approved by the Malaysian Government on 20 biodiesel projects. In summary, this is a good prospect for expanding oil palm production in the next decade.

2.4 The importance of conducting research studies in the oil palm industry

Although oil palm is native to Africa, Malaysia was the first country to initiate large-scale oil palm planting with technology together with the incorporation of innovative policies. A key element of this policy was research and development (R&D) through establishment of research institute in the country. As in most industries, R&D plays an important role in generating information, discovering new methods to increase production, and investigating the obstacles that are faced.

Since Malaysia is the leading country in the oil palm industry, other countries has no readily available custom-made technologies or relevant R&D findings. This makes the development process of oil palm industry is slow. In order to overcome the obstacles, government intervention provided quality control legislation, which was followed by the implementation of public research programmes led by the Palm Oil Research Institute of Malaysia (PORIM) in 1979, now known as Malaysia Palm Oil Board (MPOB).

Research and development has contributed to yield improvement, estate management of oil palm and cost effectiveness (Chew, 2001). Other than that, crop improvements of oil palm have been practiced through breeding, genetic engineering, and tissue culture. In addition, intensive research on palm oil contributes significantly to the oils and fats industry globally (Wahid *et al.,* 2004)

2.5 Selection and breeding of oil palm in Malaysia

Due to land limitation for oil palm plantation, there is a need to produce high genetic potential planting material for oil palm. These come across to the breeding activities in oil palm plantation. The main objective of oil palm breeding is to increase the oil and kernel production and thus contribute to plantation economics. There will be some criteria that breeders choose to select for breeding purposes. The breeder will select parent palms for high oil and kernel content, high production of fruit and make crosses between the best individuals.

Since long time ago, oil palm breeders made crosses between different cultivars with the aim to get the highest oil yielding palms. Races of *E. guineensis* were

differentiated by their fruit pigmentation and their characteristics. In Malaysia, the most common cultivars are *Dura* (D), *Tenera* (T) and *Pisifera* (P) (referred to figure 2.3). They are classified according to their shell thickness and mesocarp content. *Pisifera* palms have no shell with 95% high mesocarp content, *tenera* has 0.5-3 mm thick endocarp and high mesocarp content of 60-95% while *Dura* palms have the thickest shell (2-8 mm) and low mesocarp content which is only 35-55% of the fruit weight (Latiff, 2000).

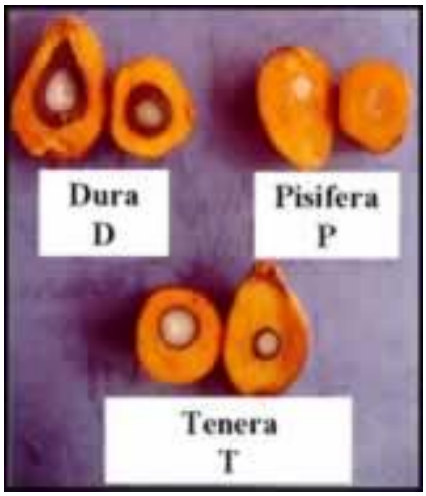


Figure 2.3: Thick shelled dura (D), shell-less pisifera (P), and thin shelled hybrid tenera (T) fruits of oil palm.
(Source from: http://aarsb.com.my/AgroMgmt/OilPalm/PlantBreeding/Intro.html)

Breeders can now rely on marker-assisted selection to obtain the few plants that are likely to express the desired characteristics from amongst tens of thousands of progeny without having to wait for the palm to grow to maturity in order to examine their phenotype. Thousands of young plantlets can be screened for some useful trait like disease resistance, improved nutritional quality or higher yield with specific marker without having to undertake any physiological or biochemical assays. This DNA marker-assisted selection can decrease the timescale of crop breeding programs for years and at the same time is cost effective for crops like oil palm with lengthy life cycles before reaching maturity and fruiting stages. Recently, a 1.8 gigabase (Gb) genome sequence of the *E. guineensis* had published, this is very

useful especially in the path to discovery of useful genes for important traits (Rajinder *et al.,* 2013).

2.6 *In vitro* propagation of oil palm through somatic embryogenesis

Cultivation of oil palm has expanded rapidly due to the increase demand for palm oil in Malaysia and globally. There is an estimation of ready market for oil palm plantlet with a demand of more than 100 million tissue culture plantlets annually (Zamzuri *et al.,* 1999). However, current annual production capacity is about 2.5 million ramets from 12 commercial oil palm tissue culture laboratories in Malaysia. There are about 3.5 million ramets produced annually worldwide (Table 2.1) (Kushairi *et al.,* 2010).

Table 2.1: Estimated World Production of Oil Palm Tissue culture plantlet

Country	Number of plantlets per year (million)
Malaysia	2.5
Costa Rica	0.5
Indonesia	0.5
Total	3.5

Source: Kushairi *et al*. (2010).

Oil palm starts to bear fruits at about 7 years old. After that, it has an average lifespan of 25 years and can still produce fruit till it reached the age 50 years. Due to the longer growth life span in conventional practices, mass propagation through tissue culture can serve as a faster and cheaper alternative to supply the elite oil palm plantlets (Thuzar *et al.,* 2011). First clonal palms were planted in Malaysia in the year 1977 and replications of the clones have been planted in other fields in subsequent years. Favourable results were obtained (Corley *et al.,* 1979). This led to a rapid expansion of adoption of clonal tissue culture plantlets in the field. More than ten tissue culture laboratories in Malaysia were established within a period of 10 years (Wooi, 1990). A number of studies on successful in regeneration of oil palm

using tissue culture techniques have been reported (Paranjothy and Othman., 1982; Duval *et al.,* 1988; Rival *et al.,* 1998; Tarmizi *et al.,* 2004).

At the early stage of tissue culture practice, the root, inflorescences and leaf explants were used as the starting material for tissue culture. Generally, the current tissue culture process that practice in MPOB was using the leaf explants as the starting material. Young leaves explants has less contamination risk but it has to be cut from the unopened part of the bases of old leaves which can cause a huge damage to the growth of palm. Roots are not commonly used as they are more likely to get contaminated with soil fungus or bacteria. The young inflorescences will not cause damage to the palm trees but the initiation of the callus and embryogenesis occur slower than the leaf explants. Currently, young leaf spears are the most preferred choice as a starting material for mass propagation of oil palm in tissue culture due to high clonability rate (Rajanaidu *et al.,* 1997). Based on the reported field data, clonal material D x P from tissue culture has contributed to an oil yield increment of 20% to 30% over planting material derived from seeds (Soh *et al.,* 2001).

Plant growth regulators have been extensively used previously to increase the cloning efficiency. However, incidences of plant abnormality caused by plant growth regulator have been reported (Kushairi *et al.,* 2010). The abnormality of plantlets cannot be fully eliminated but with good cultural practice, it is able to reduce the occurrence of abnormality down to a level below 5% (Maheran *et al.,* 2005). In addition, liquid culture technology has been established for oil palm clonal production and it offer favourable reproducibility as well as the amenability for scalling up the production of oil palm clones (referred to Figure 2.4). Besides, liquid system provided shorter duration of micropropagation cycle compared to gel system. In the meantime, Malaysian Palm Oil Board (MPOB) has used the bioreactor technology with semi or full automation process for mass oil palm clonal production (Tarmizi *et al.,* 2003).

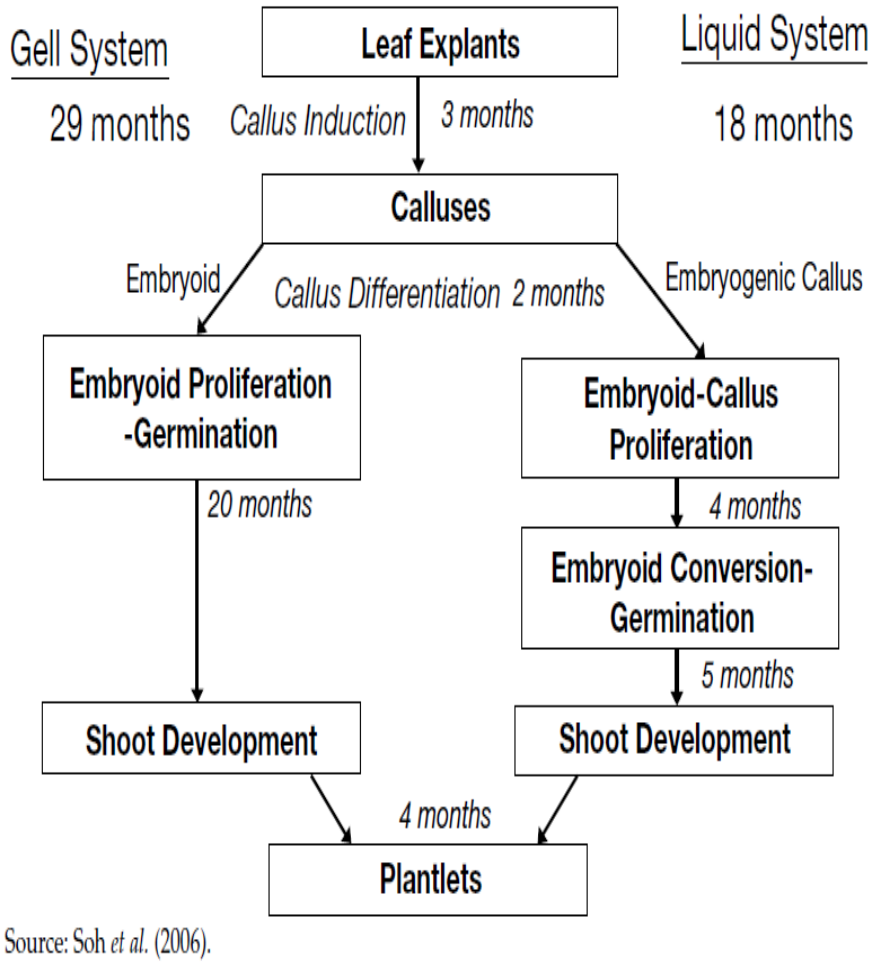


Figure 2.4 Culture stage and duration in gel vs liquid systems to produce 5000 shoots.

Somatic embryogenesis is a process by which somatic cells can be regulated to differentiate into embryos with similar morphological appearance as parental plant by the use of plant growth regulators. There are three main developmental stages of somatic embryogenesis, namely, induction of embryogenesis from undifferentiated mass of cells known as callus, maturation of somatic embryos, and plantlets regeneration (referred Figure 2.5).

The somatic embryos can produce either directly from a cell or a group of cells or indirectly through the production of an intervening callus. There are some publications for plants that cultured through direct somatic embryogenesis, such as *Arachis hypogea* (Hazra *et al.,* 1989), *Brassica juncea* (Eapen *et al.,* 1989), and *Coffea canephora* (Quiroz-Figueroa *et al.,* 2006).

Indirect somatic embryogenesis is more common compared to direct somatic embryogenesis methods. This method has applied to many plant species including *Coelogyne Cristata* orchid (Naing *et al.,* 2011) and *Dendrobium* Chiengmai Pink (Chung *et al.,* 2005). Another special type of indirect somatic embryogenesis is the secondary somatic embryogenesis. This method uses somatic embryos produced as the initial material to proceed through the tissue culture process and frequently show low conversion rates to plants (Vicient and Martínez, 1998).

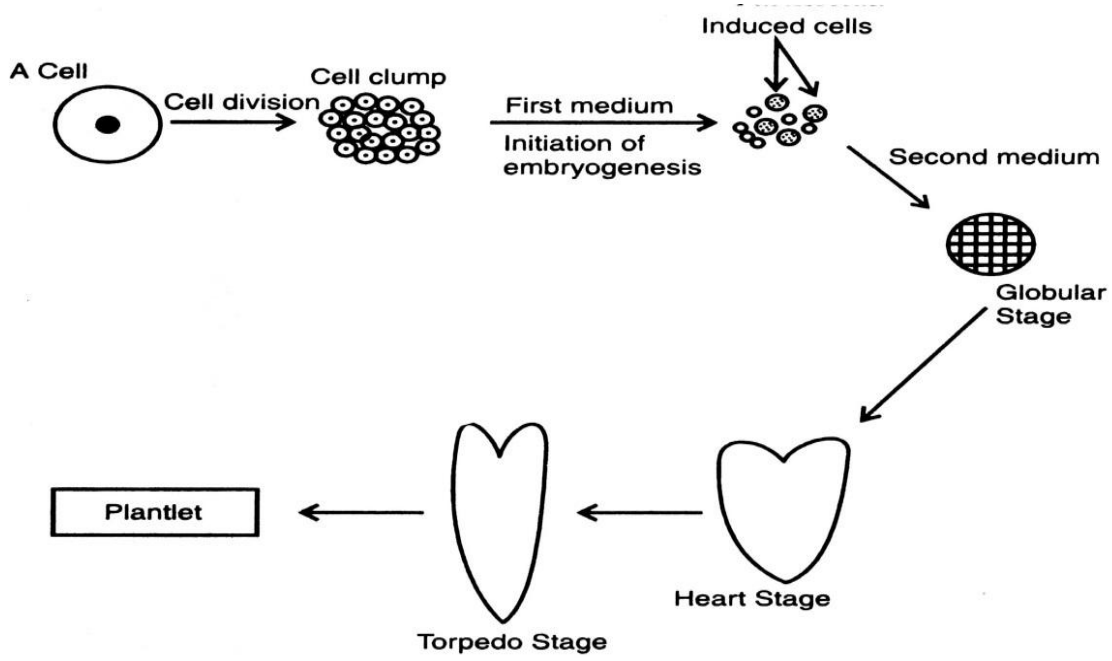


Figure 2.5 Event of somatic embryogenesis occur in tissue culture

Applications of somatic embryogenesis include clonal propagation of genetically uniform plants, large production of quality plant materials, plant virus elimination, metabolite production, and might contribute to the *in vitro* mychorrhizal initiation (Vicient and Martínez, 1998). Somatic embryogenesis provides an alternative method to the plant breeders to produce genetically uniform plants with selected superior genotypes.

Somatic embryogenesis has been widely applied in many plant species such as mass propagation of timber-yielding leguminous tree, Dalbergia sissoo Roxb (Singh and Chand, 2003), improve salt and disease resistance in Citrus (Litz *et al.,* 1985), virus resistance in sugarcane (Oropeza and de Gracia, 1996), metabolite production such as toxoids compound from Taxus species (Lee and Son, 1995; Wann and Goldner, 1994), germplasm preservation on hormone-free medium in some species including asparagus (Dalbreil *et al.,* 1994), *Hevea brasilensis* (Cailloux *et al.,* 1996) in and medicinal *Ephedra foliate* (Dhiman *et al.,* 2010).

2.6.1 Problems of somatic embryogenesis

In vitro propagation of oil palm has been initiated in Malaysia since 1980s. The regeneration efficiency of oil palm through *in vitro* propagation in tissue culture is depending on the genotypes used. At present, there are several bottlenecks encountered in the large scale production of oil palm tissue culture materials. Early results showed low yields of fruits and emergence of fruiting abnormalities. Another prevalent issues occurred in the tissue culture process are the low rate of embryogenesis and clonal abnormality.

Besides, plantlets derived from the *in vitro* culture have the tendency to exhibit abnormal phenotypes, which is known as somaclonal variation. A number of critical issues needed to be considered besides somaclonal variation. For examples, cloning efficiency, ortet selection efficiency, feasibility of recloning, condition or acclimatization of plant produced *in vitro* and field testing. To investigate the problems, a study on the potential effect of somaclonal variation on genome size has been done for tissue culture and seed derived plant (Fabienne *et al.,* 2006) using cytometric approach since it is a rapid and powerful method to estimate nuclear DNA content (Galbraith *et al.,* 1983). However, the results show that the ploidy level of both is same which reinforced the hypothesis that an epigenetic origin is involved for the somaclonal variation in oil palm.

Table 2.2: The success rate (%) of callogenesis, embryogenesis and shoot regeneration in palms and explants (Source: Kushairi *et al.,* 2010).

Culture Stage	**Success rate (%)**			
	Based on Palms		Based on Explants	
	Ortets [a(n=216)]	Reclones [b(n=110)]	Ortets [a(n=400,000)]	Reclones [b(n=200,000)]
Callogenesis	100	100	19 (1 to 67)	14 (1 to 41)
Embryogenesis	72	88	3 (1 to 6)	7 (6 to 20)
Shoot Regeneration	56	85	96	95

a = Cloning of palms derived from seeds
b = Recloning of palm derived from tissue cultured ramets
c = Based on embryogenic lines
( ) = range

According to the data collected for oil palm tissue culture, only some of the oil palm species are able to produce callus with high proliferation rate, while many others produce low proliferation callus (Kushairi *et al.,* 2010). This account for a very low callogenesis and embryogenesis rate of oil palm explants i.e. at 19% and 6% respectively from the proliferating callus culture (Table 2.2). This has been one of the major problems hampering the progress of oil palm tissue culture which

subsequently will affect the production for oil palm seedlings in order to meet the demands.

2.6.2 Studies of somatic embryogenesis on genome level

Since somatic embryogenesis is a complex mechanism, there are areas needed to be explored and understood especially at the cellular level. Since 1980s, scientists were started to investigate on the somatic embryogenesis until now, it is still very little information gained. Previous studies have been conducted on embryogenic and callus specific protein in carrot (Sung *et al.,* 1981), rice (Chen *et al.,* 1987) and grass (Hahne *et al.,* 1988).

Most of the studies use model plants to investigate the process of somatic embryogenesis. Through the studies, several potential molecular markers on embryogenic competent cells have been identified such as somatic embryo receptor-like kinase (SERK) gene (Schmidt *et al.,* 1996), LEAFY COTYLEDON (LEC1), FUSCA3 (FUS3), ABA INSENSITIVE3 (ABI3) and Late embryogenesis abundant (LEA) gene (Braybrook and Harada, 2008).

Other studies indicated that overexpression of WUSCHEL (WUS) genes (Zuo *et al.,* 2002) and BABY BOOM (Boutilier *et al.,* 2002) proved to have a role in inducing somatic embryogenesis by promoting stem cell identity and enhances apical meristem development. WUS act as meristem organizers or embryos organizers and first localized to the shoot meristem in the heart stage embryo, and it makes plant continue produce organs by regulating the stem cell pool. At the same time, somatic embryogenesis-related genes have been successfully isolated from orchardgrass (*Dactylis glomerata* L.), they are *Dactylis glomerata* embryogenesis 1 (DGE1) and *Dactylis glomerata* embryogenesis 2 (DGE2). DGE1 showed 81% homolog with the WRKY DNA-binding protein 21 in *Arabidopsis thaliana* (Alexandrova and Conger, 2002).

Malaysia Palm Oil Board (MPOB) has been actively working on gene expression studies in both embryogenesis and clonal abnormality in oil palm. Some possible gene markers have been identified but functionality of these proteins associated with embryogenesis is still unknown (Sambanthamurthi *et al.,* 2009). Initially, specific gene approaches have been applied to identify potential genes related to somatic embryogenesis in oil palm such as auxin-inducible genes and stress response genes (Meilina and Ooi, 2006). Following that, Ooi (2008) found that a typical embryogenesis marker, serine kinase has been found to be involved in the signal transduction pathway in oil palm somatic embryogenesis.

Improvement in the technology made it possible to identify several interesting expressed sequence tags (ESTs) such as granule-bound starch synthase (GBSS) and putative transcription factor (Myb1). These ESTs can be used to screen calli for the embryogenic potential. More oil palm tissue culture ESTs have been analysed, marked and brought together to identify genes associated with callogenesis and embryogenesis (Low *et al.,* 2008). Nonetheless, the actual mechanisms controlling plant gene expression in somatic embryogenesis remain unclear. Thus future trends involve characterization of development-specific gene in somatic embryogenesis are required (Khurana and Chugh, 2002).

So far, most of the studies on somatic embryogenesis were focused on the genomic level. The advance development of proteomics techniques provides a new alternative way to understand the mechanism involved in somatic embryogenesis at the protein level. Hundreds of proteins can be translated by a particular gene. Thus, identification of the differential gene expression in somatic embryogenesis may provide a lead to fully understand the mechanism involved. Since the proteins play a vital role in regulating the biological mechanism and determining the phenotypic traits, there is a need to understand the somatic embryogenesis process by looking at the protein species that are highly expressed or vice versa.

2.7 Proteomics Approaches and their Application

2.7.1 Introduction to Proteomics

The word “proteome” is derived from PROTEins expressed by a genOME. Proteomics is the characterization of the entire protein complement expressed by a genome of a given organism (Wilkins *et al.,* 1996). The proteome is the time- and cell-specific protein complement of genome in a cell at any given time. It is the large scale study of protein properties such as expression, modification and interaction of proteins to gain an overview of cellular processes at protein level.

The first complete genome of unicellular eukaryote, *Saccharomytes cerevisiae* (Goffeau *et al.,* 1996), followed by *Escherichia coli* (Blattner *et al.,* 1997) and even human (Lander *et al.,* 2001) have now been fully sequenced. With the complete genome sequence, it provides good starting materials for research scientists to study precisely on all the gene functions. The study of proteomic is therefore relevant. However, there are additional challenges in proteomic analysis as compared to the genomic. Firstly, the genome is static while each proteome is dynamic and changes in response to metabolic state and intracellular and extracellular signal molecules (Michael *et al.,* 2008). The second concern is the relative amounts of the components within the genome and proteome. Some proteins are often expressed in low quantity and with high turnover rates, especially those proteins that are involved in signal transduction mechanisms. This makes a very high variability in protein profiling as proteins can alter over very short periods of time.

Plant biologists in all disciplines are now incorporating high-throughput “-omics” technologies principally transcriptomics, proteomics and metabolomics to facilitate the discovery pathways for functionality of genes in a systematic manner. Proteomics aids in giving an understanding towards cellular function at the level of

any cells, organs, tissues and even organisms. It has been described as one of the strong approaches which are able to complement and relate to transcriptomics and metabolomics (Agrawal *et al.,* 2011). Proteomics approaches becoming one of the main directions for researchers to get the fundamental molecular-level knowledge that can be used in characterizing plant subspecies and used wisely to identify molecular markers especially in breeding program (Subhra *et al.,* 2015).

Currently, the most intensive studies on proteomics have been performed on model plants *Arabidopsis thaliana* and rice which give the highest number of publications in the proteome references. Publication of the proteome research on plant is steeply increasing especially after the publication of draft genome sequence of Arabidopsis (The *Arabidopsis* Genome Initiative, 2000) and rice (Goff *et al.,* 2002 and Yu *et al.,* 2002) in 2000 and 2002 respectively. More recently, many researchers have employed proteomics approaches such as maize (Majeran *et al.,* 2010), wheat (Peng *et al.,* 2009), barley (Møller *et al.,* 2011 and Rasoulnia *et al.,* 2011), soy bean (Mohammadi *et al.,* 2012), chickpea (Subba *et al.,* 2013) and date palm (Marondedze *et al.,* 2014).

Identification of proteins will be expanded with the existence of significantly increasing amounts of genomic DNA and EST sequences being deposited in the public databases. Proteomics would not be feasible without the previous accomplishment of genomics (Tyers and Mann, 2003). Lacking database information especially in novel plant species will bring huge challenges in the protein identification (Jorrín *et al.,* 2007). There are a lot of ongoing efforts to create a comprehensive plant database but until now there are more that can represent a complete collection of proteins in any plants (Subha *et al.,* 2015).

Even though proteomics is beginning to achieve at an advanced level, there are limited fraction of cells in a few biological system that have been fully characterized and there are still a lot more proteins remains to be investigated. In general, proteomics can be subdivided into different areas including descriptive proteomics, differential expression proteomics, posttranslational modifications, and interactomics (Jorrín-Novo *et al.,* 2009).

2.7.2 Gel-based proteomics approaches

The steps for proteomics involve protein extraction, depletion, purification, separation, mass spectrometry analysis and finally application of bioinformatics tools for data analysis and database searching. Generally, in order to create a protein profile or a particular sample, proteomics two-dimensional gel electrophoresis (2-DE) technique coupling with mass spectrometry can be used. The gel-free method allows extracted protein to be immediately channalised through liquid chromatography mass spectrometry/mass spectrometry (LC MS/MS). While, gel-based proteomics requires running the extracted proteins in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels. Interestingly, gel based proteomics with IPG strips remain the only technique that can consistently be applied for expression profiling of complex protein mixtures (Wang *et al.,* 2008). Moreover, 2D gels have been extensively used in differential proteomics expression studies since the commercialization of the DIfference Gel Electrophoresis (DIGE) technology in early 2000 (Freidman and Lilley, 2009; Lilley and Friedman, 2004; Unlu *et al.,* 1997).

Two-dimensional gel electrophoresis has long been widely used to analyze protein expression in organs and tissues. A normal plant cell comprises more than 20,000 different individual polypeptides but only a limited number of spots count i.e. in the thousands are able to produce with the 2-DE PAGE gels. The proteomic approaches currently used are mostly restricted to organ, tissue and cell specific proteome before building into a complete proteome map of a species (Newton *et al.,*

2004). Most studies come to the aim to discover possible markers of different genotypes or phenotypes. Conversely, most of the identified proteins could not come into a conclusion due to technically difficult in the mass spectrometry analysis. Apart from that, a lot of proteins remain unknown and unidentified when blast through the existence databases. These cause huge challenges for researchers to identify their proteins and unable to process the unknown proteins.

2.7.2.1 Protein extraction for plant tissues

There are three critical phases in plant protein extraction. Firstly, the tissue disruption phase. This involves crushing of the plant cell wall that is made up of mainly polysaccharides in liquid nitrogen to reduce protein degradation during grinding. Higher protein yields can be obtained with finer ground tissue (Wang *et al.,* 2003). Once fine powder is obtained, the second phase to consider is the efficiency of the extraction buffer. Vascular plants are abundant secondary metabolites which normally accumulate in the vacuoles of plant cells. There are approximately 8000 kinds of different compounds grouped into plant phenolics comprising lignin, stilbenes, tannins, flavonoids and phenols (Stalikas, 2007). In order to effectively remove the phenolic contaminant, several methods of cleanup have been suggested such as TCA/acetone with 10% TCA precipitation, water soluble polyvinylpyrrolidonen (PVP) (King, 1971) or polyvinyl polypyrrolidone (PVPP) (Wang *et al.,* 2009). The final phase is protein solubilization. The best solubilization methods allow most of the non-covalent bound protein complexes to separate into single polypeptides, thus performing well in 2D gels. Application of reducing agents such as Dithiothreitol (DTT), in the solubilization buffer can effectively reduce disulfide bonds of proteins. The combination of different degree and concentration of the solubilizing buffers will be very useful in solubilize a good number of proteins in this process.

2.7.2.2 Two-Dimensional Gel Electrophoresis (2-DE)

Two-dimensional gel electrophoresis separates protein in the first dimension by their intrinsic isoelectric point (pI) using isoelectric focusing (IEF), followed by the second dimension by their molecular weight (Mr). 2-DE is one of the most powerful tools for protein profiling especially to visualize protein isoforms that result from charged posttranslational modification (Huang *et al.,* 2008).

Isoelectric focusing (IEF) is an electrophoretic separation method which separates amphoteric molecules such as protein and peptide based on their charge (Westermeier *et al.,* 2009). The isoelectric point defined as the pH at which there is zero net electric charge of a protein or protein become immobilize in an electric field. In this IEF process, the extracted protein will need to rehydrate overnight with a rehydration buffer using a strips known as Immobilised pH gradient strips (IPG). The IPG strips technology has the pH gradient which made by acidic and alkaline buffering groups which copolymerized with the polyacrylamide matrix during gel preparation (Westermeier, 2005). The IPG strips are dried for long term storage. The rehydration buffers containing high chaotrope concentration, a zwitterionic detergent, a reducing thiol, and carrier ampholytes to prevent the formation of aggregates between proteins. Carrier ampholytes play a role to provide an amphoteric buffer. Besides, it also helps to improve conductivity and protein solubility.

After first dimensional separation, IEF, it is then move to second dimensional separation based on the molecular weight which is SDS-PAGE. SDS-PAGE has long been a method for resolving intact proteins according to their electrophoretic mobility (Laemmli, 1970) (a function of the length of a polypeptide chain and its charge) and widely used in biochemistry, forensic, and molecular biology. SDS is a best solubilizing detergent that would allow hydrophobic proteins, basic proteins and all proteins to migrate in the same direction, and separate according to their molecular

weight. The proteins will form anionic micelles with a constant negative net charge per mass unit when excess SDS is added to the samples. The secondary and tertiary structures of the protein will be disrupted and the polypeptides become unfolded. Reducing agent such as DTT and 2-mercaptoethanol can be added to break the disulfide bonds between cysteines (Freidman *et al.,* 2009).

2.7.2.3 Protein Identification by Mass Spectrometry (MS)

Currently, mass spectrometry has been established as a primary method for protein identification from complex proteins. Proteins are proteolyic digested to produce the peptide sequence and the mass/charge (m/z) of the peptide sequence was captured by the mass spectrometry analyzer which facilitate the protein identification. The most commonly used enzyme is trypsin, which targets the C-terminal side of lysine and arginine.

The development of mass spectrometry has been rather limited due to the lack of reliable methods to conduct soft ionization and effectively transfer ionized molecules from condensed phase to the gaseous phase. Recent advances in instrumentation and software analyze have facilitated mass spectrometry (MS) identification and imaging of biological molecules. A mass spectrometry basically consists of an ion source, a mass analyzer that measure mass-to-charge and a detector which measure the quantity and abundance of particular ionized ions.

The inventions of matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) enabled the polypeptides to volatize and ionized for mass spectrometry analysis. A proton source that promotes sample ionization, and a matrix will co-crystallize with the protein samples before the MS analysis. MALDI will sublimate the samples out of dry, crystalline matrix via laser pulse. Mass analyzer must have the ability to generate rich information of ion mass spectra from peptide fragment by measuring the travel time of each peptide. The spectrum of masses of all peptides is then compared with databases of known proteins (Mann *et al.,* 2001).

2.7.3 Proteomics in somatic embryogenesis

The current progress in the proteomics field provides a platform to study somatic embryogenesis in plant at molecular level. Two dimensional gel electrophoresis have been applied in somatic embryogenesis study in carrot (Choi and Sung, 1984), *cichorium* (Helleboid *et al.,* 2000), Vitis (Gianazza *et al.,* 1992), *Cupressussempervirens* (Sallandrouze *et al.,* 1999) and *Cyclamen persicum* Mill (Fuhua *et al.,* 2010).

There are three major somatic embryogenesis related proteins that have been identified in *Cichorium* as pathogenesis-related (PR) proteins. The study showed that direct somatic embryogenesis in Cichorium '474' is associated with the increase in the level of β-1,3-glucanases, 32 kDa chitinase and 25 kDa osmotin-like proteins. This study indicated that specific proteins involved in somatic embryogenesis might be involved in other pathways as well (Helleboid *et al.,* 2000).

In a study on *Picea glauca*, a group of scientists from Canada found that early somatic embryogenesis was associated with a total of 79 proteins that show differential expression when subjected to MS/MS analysis. Unfortunately, there are only 48 proteins successfully identified. In this study, they found that a storage protein, vicilin has shown to increase in abundance throughout the maturation of somatic embryos. This study also found enolase protein expressed during embryo development (Lippert *et al.,* 2005).

In another study, proteomic analysis has been carried out on somatic embryos induced in *Medicago truncatula* culture that grown under 6-Benzylaminopurine and 1-Naphthaleneacetic acid treatment. The study found that 54 proteins are differentially expressed and only 16 proteins are able to be identified with high confidence. The results showed a decrease trend of Rubisco small chain proteins, expression of Thioredoxin H protein in the early development of somatic embryogenesis and indicated that 1-Cys Peroxiredoxin might play a role in embryogenesis (Imin *et al.,* 2005).

Two years later, *Medicago truncatula* protoplasts were used on proliferation study. A total of 886 protein spots showed significant differential changes but only 89 proteins were able to be identified using MALDI-TOF MS. The identified proteins were grouped based on their function. The results showed that more than 50% of the identified proteins were involved in energy metabolism. Several pathogenic related (PR) proteins were shown to be highly abundance, i.e. up to five fold, which might be due to the effect of wound or stress response from protoplast isolation (Jong *et al.,* 2007).

More recently, in a study by Fuhua *et al* (2010) on *Cyclmen persicum* Mill, it shown that 460 total proteins resolved in 2-DE gels, 35 protein spots were found to be significant in expression. And there were 10 out of 35 protein spots were able to be identified by MALDI-TOF-MS and MALDI-TOF-TOF MS. There are few interesting proteins found such as proteasome subunit which is closely related to cell proliferation processes (Amsterdam *et al.,* 1993). This protein has also been reported in previously on early somatic embryo development in *Picea galuca*. Besides, there are some proteins found to be involved in metabolism mechanism such as triosephosphate isomeras which is one of the the key regulatory enzyme that took part in glycolysis and tricarboxylic acid cycle (Ito *et al.,* 2003).

2.7.4 Real-time Reverse transcription Polymerase Chain Reaction (RT-qPCR) in proteomics study

Quantitative reverse transcription Polymerase Chain Reaction (RT-qPCR) commonly used to quantify the relative expression levels of gene of interest in an organism. By using the qPCR coupled with a RT step approaches, messenger RNA (mRNA) expression for a particular gene is quantified. Quantification of total RNA is necessary prior to the real time RT-PCR step to ensure standardised RNA amount is used during verification analysis amongst samples. Despite the quantities of the RNA, quality of the RNA is very important for analysis. If the RNA sample is shown to be partially degraded, it is advisable not to use it because the integrity of the samples will affect the overall expression level by underestimating lower transcript level. This affects the assay sensitivity leading to the incorrect target ratio (Stephen *et al.,* 2009).

Real time RT-qPCR which also known as messenger RNA approach is a powerful tool as it allowing massive screening of a group of genes at one time. It provides the quantification of expression level for a particular transcript within a tissue at specific time as what proteome does. However, it is important to understand that there is dissimilitude between the messenger (transcript) and protein (final effector). Phenotypic behavior of an organism is highly affected by the expression on final effector like proteins or metabolites rather than mRNA.

Since there is no direct information given by the mRNA expression level on the protein abundance in particular samples, many researches will start to question on the necessity of carrying out the real time RT-qPCR. There are a lot of studies showed that the poor correlation between the mRNA and protein abundance (Tien *et al.,* 2004., Carpentier *et al.,* 2008), but in the other hand, some of the proteomic

studies did show the correlation between mRNA and metabolites in their samples (Goossens *et al.*, 2003 and Hirai *et al.,* 2004). This is a challenge to researchers as the information that is provided through transcript profiling cannot be independent, and it needs a bridge to connect the information between transcripts and proteins. There are a lot of information needs to be compiled such as the mRNA level, protein expression level and even possible posttranslational modifications in order to give a full understanding on the mechanism that occurred in an organism.

CHAPTER 3

OPTIMISATION OF PROTEIN EXTRACTION AND TWO-DIMENSIONAL GEL ELECTROPHORESIS FOR OIL PALM LEAF

3.1 Introduction

In proteomic studies, the quality of samples is a vital factor. The ideal protein extraction method should be able to present a comprehensive set of protein profiles for the species to be studied. Unfortunately, there is no single protein extraction protocol that can capture the full proteome (González-Fernández *et al.,* 2010). Sample preparation is a crucial step prior to electrophoresis. Protein extraction from the plant samples are challenging mainly due to low cellular protein content and presence of high level of contaminant like proteases and interfering component such as phenolics, pigments, lipids, nucleic acid, and other secondary metabolites (Darmeval *et al.,* 1988; Shaw and Riederer, 2003; Gorg *et al.,* 2004). Oil palm leaf are rich in polysaccharides and secondary metabolites. The presence of secondary metabolites highly affects the performance of protein extraction as well as separation in 2-DE gels. According to Vâlcu and Schlink (2006), phenolics are likely to form irreversible complexes with proteins and oxidation of phenolics contributes to streaking and artefactual spots in protein gels. This study is focused on the optimisation of extraction methods for oil palm leaf in order to produce good quality proteins. Two-dimensional gel electrophoresis (2-DE) was employed to separate proteins present in total cell extracts. Several modifications, such as nuclease treatment has been assessed for oil palm leaf sample in order to get high quality 2-DE gel profiles.

3.2 Materials and methods

3.2.1 Plant Materials

Young (cabbage) and old leaf of oil palm (*Elaeis guinensis* Jacq.) were collected from Advanced Agriecological Research Sdn Bhd (AAR) and Malaysian Palm Oil Board (MPOB), Malaysia. Young cabbage leaf sample was collected from the unopened leaf of the palm. While, the old leaf samples with the frond number 17 in the oil palm tree were used. All plant samples were stored at –80°C until use. Cabbage stage unopened leaf was defined as young leaf used in this study while the old leaf were referred to Frond number 17 (F17). Both old and young leaf samples were provided with high and low proliferation rate group (Table 3.1). Total proteins were extracted from the samples using the four methods as described in the next section. Leaf were transferred to a pre-chilled mortar, and ground into a 0.2 g/ml fine powder using liquid nitrogen. This was used as the first step for all the protein extraction methods.

Table 3.1: Old leaf and cabbage samples with high and low proliferation rate categories were collected from the AAR and MPOB.

AAR samples (Old and cabbage samples)		MPOB samples (Old and cabbage samples)	
High proliferation	Low proliferation	High proliferation	Low proliferation
AN 25	AN 27	226	73
AN 28	AN 29	290	285
		294	295

*High proliferation = more than 30 embryogenic lines

Low proliferation = less than 10 embryogenic lines

3.2.2 Optimisation of the protein extraction methods

Total proteins were extracted using four extraction methods, namely TCA/acetone precipitation method, phenol extraction method, lysis buffer and combination of both TCA/acetone and phenol method. All the collected samples were transferred to a pre-chilled mortar, frozen in liquid nitrogen and ground into a fine powder. This finely ground powder were used for all protein extraction methods.

3.2.2.1 TCA/acetone precipitation method

Proteins were extracted according to a protocol modified from Gómez-Vidal *et al.* (2008). Five volumes of ice-cold 10% v/v trichloroacetic acid (TCA) in acetone containing 20 mM dithiothreitol (DTT) were added to 0.2 g/ml finely powdered leaf tissues. Proteins were precipitated for two hours at –20ºC and centrifuged at 13,000 x *g* for 15 minutes at 4ºC. The pellets were washed with acetone containing 20 mM DTT and incubated overnight. Samples were centrifuged at 13,000 x *g* for 15 minutes at 4°C. The supernatants were discarded and the pellets were washed twice with ice-cold acetone containing 20 mM DTT with each incubation time of 30 minutes. The pellets were air dried and dissolved in 2-DE rehydration solution (8 M urea, 4% 3-[3-Cholamidopropyl) dimenthylammonio]-1-propanesulfonate (CHAPS), 0.2% carrier ampholytes, and 18 mM DTT). Subsequently, sonication with 230V/50Hz (Fisherbrand Ultrasonic bath) for 15 minutes was performed to enhance the solubility of the proteins in the rehydration buffer. After sonication, the samples were rocked at room temperature for 30 minutes and centrifuged at 20,000 x *g* for 30 minutes at 18ºC. The supernatants were collected and treated with nuclease mix (GE healthcare life sciences, USA) for 30 minutes on ice before storing at -80ºC.

3.2.2.2 Lysis buffer extraction method

A total of 0.2g powdered tissue were mixed with 1 ml of lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 100 mM DTT, 40 mM Tris, protease inhibitor cocktail (Sigma Aldrich, USA) and Nuclease Mix (GE healthcare life sciences, USA). The mixtures were incubated on ice for 15 minutes. The sample was centrifuged at 20,000 x *g* at 15°C for 20 minutes. The supernatant of the samples were collected and kept at -80°C until further use.

3.2.2.3 Phenol method

A phenol extraction procedure was carried out according to Wang *et al.* (2003). Powdered tissues were suspended in 1ml of cold acetone twice and centrifuged at 12,000 x *g* for 10 minutes. Pellets were resuspended in 1 ml of 1:1 phenol: Tris-SDS buffer (sucrose, 2% SDS, 0.1 M Tris-HCl, pH 8.0, 5% 2-mercaptoethanol) for extraction. After thorough mixing by vortexing, and phase separation, the upper phenolic phase was collected. The procedure of extraction with an equal volume of 1:1 phenol/dense SDS buffer was repeated twice. Eventually, only half of the original volume remained. The samples were precipitated with 0.1 M ammonium acetate in cold methanol, incubated at -20°C for 30 minutes and centrifuged at 20,000 x *g* for 10 min. It was then followed by pellet washing with cold 0.1 M ammonium acetate in methanol and 80% acetone twice. The final pellet was dissolved in 2-D Electrophoresis rehydration buffer solution as described in section 3.2.2.1.

3.2.2.4 Combination of TCA/acetone and phenol method (TCA/Ph)

Proteins were extracted according to the protocol of Wang *et al.* (2006). The plant tissue powder was washed with 10% TCA/acetone, followed by washing in 0.1 M ammonium acetate in 80% methanol, and a further wash with 80% acetone. After washing, the powder was air dried at room temperature. Phenol/SDS solution (30% sucrose, 2% SDS, 0.1 M Tris-HCl, pH 8.0, 5% 2-mercaptoethanol) in a ratio of 1:1 was added to the mixture and incubated for 5 minutes. After centrifugation at 20,000 x *g* for 10 minutes at 4°C, the upper phenol phase (0.2-0.4 ml) was collected and transferred into a fresh tube. Methanol containing 0.1 M ammonium acetate was added and the samples were incubated overnight at -20°C. After centrifugation at 12,000 x *g* for 10 minutes at 4°C, the pellets were washed with 100% methanol, following by 80% acetone wash. The pellet was dissolved in 2-D Electrophoresis rehydration buffer solution as described in Method (A).

3.2.3 Quantification of protein content

Protein concentrations were measured using Bradford Protein method (Bradford, 1974). In this method, a stock solution of 1 mg/ml of Bovine Serum Albumin (BSA) was prepared and make into a serial of dilution to create a protein standard curve. Ten µl of the sample solution was added to 250 µl of Bradford reagent and the mixtures were vortexes and incubated at room temperature for 10 minutes. The blank standard was prepared whereby 10 µl of the buffer was used instead of the sample solution. The protein absorbance was measured at absorbance 595 nm using spectrophotometer. The protein concentration was determined by reference to a standard curve.

3.2.4 One Dimensional gel electrophoresis – Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

One dimensional SDS-PAGE analysis was performed as described by Laemmli *et al.* (1970). Approximately 20 mg solubilized protein was loaded in the wells of 1.0 mm thick SDS-PAGE gel. The 7% stacking and 12% resolving polyacrylamide home-cast gels (Figure 3.1) was used to perform the electrophoresis for one dimensional gel electrophoresis (1-DE) with the duration of 1 hour 10 minutes with 120 V running voltage in a 7cm Mini-Protean 3 cell (BioRad) in order to confirm there are presence of proteins from different protein extraction protocol prior to 2-DE.

3.2.4.1 Preparation of Sodium Dodecyl Sulphate (SDS) - Polyacrylamide Gel

The SDS-polyacrylamide gel needed for one dimensional SDS-PAGE and two dimensional SDS-PAGES are slightly difference. One dimensional SDS-PAGE gel consists of two layers where stacking gel is the upper part and lower part will be the separating gel. For the two dimensional SDS-polyacrylamide gel, only separating gel was used. The SDS-polyacrylamide gel was prepared one day before the running time to make sure the gel polymerize completely. During preparation, the monomer solution (distilled water, 2 M Tris-HCL, pH 8.8, 30% Acrylamide, SDS) was degassed for 20 minutes using sonicator instrument with the purpose of prevent the formation of air bubbles during polymerization. The fresh prepared 10% Ammonium persulfate (APS) and TEMED as the source of free radicals and a stabilizer was added last to initiate polymerization.

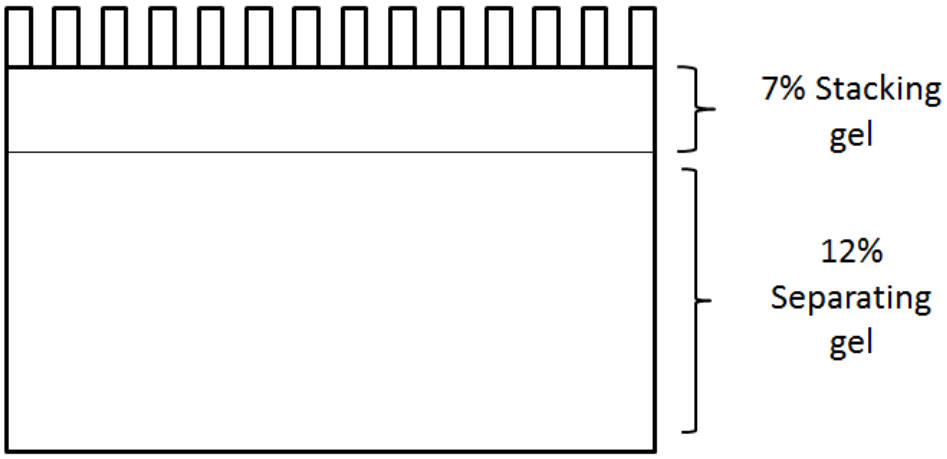


Figure 3.1: The schematic SDS-PAGE gel which consists of 7% stacking gel and 12% separating gel.

3.2.5 Two-Dimensional Gel Electrophoresis

3.2.5.1 First Dimensional Gel Electrophoresis

The protein concentration of all samples was measured using the Bradford method (Bradford, 1976). For 2-DE, the Isoelectric focusing (IEF) stage was conducted on 7 cm Immobilised pH gradient (IPG) strips (GE healthcare, USA) with a linear pH gradient of 3-10 on the PROTEAN®IEF System (BioRad). The IPG strips were rehydrated at a constant 50 µA per strip at 20°C for 16 hours. The focusing program used a linear increase from 0 to 250 Volt (V) over 20 minutes, 250 to 4000 V for 2 hours and then a rapid gradient to 4000 V until 10000 Volt hours (Vh) had accumulated.

3.2.5.2 Second Dimensional Gel Electrophoresis

After IEF, the IPG strips were incubated in an equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M Urea, 30% w/v glycerol, 2% w/v SDS) containing 2% w/v DTT for 15 min, followed by incubation for 15 min in the same buffer containing 135 mM iodoacetamide instead of DTT. The strips were transferred to 12% SDS-PAGE gels for the second dimension separation using SDS electrophoresis buffer (250 mM Tris, pH 8.3, 1.92 M Glycine, 1% SDS) with 150 V applied for approximately 1 hour 5 minutes. The experimental molecular mass (M_r) values were calculated by compared the mobility with protein standard markers (SDS-PAGE Standards Broad Range, BioRad), while the isoelectric focusing point (p*I*) were determined using 3-10 and 4-7 linear scales over the total dimension of the IPG strip (Figure 3.2).

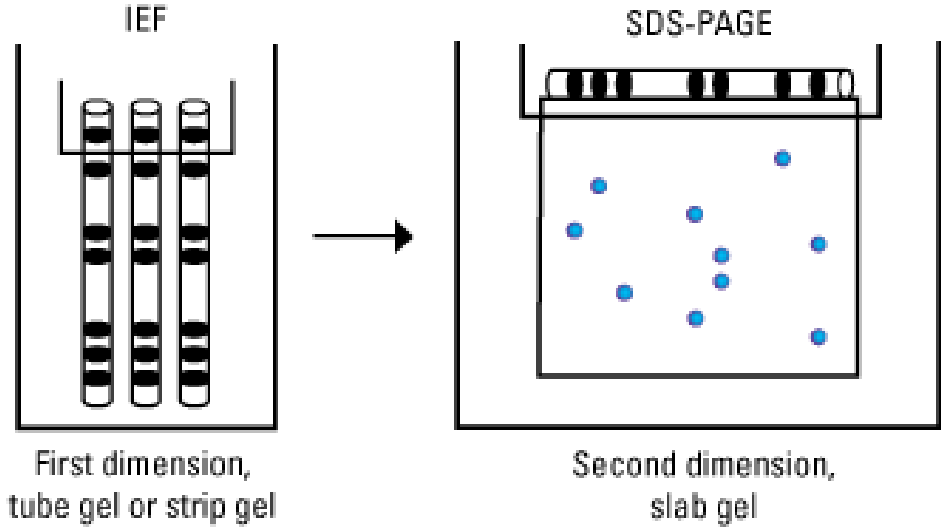


Figure 3.2: Diagram showing two dimensional gel electrophoresis workflow.

3.2.5.3 Gel Staining Protocols

Coomassie staining

After electrophoresis, the 2D gels were removed gently from the glass plate with distilled water. The gels were placed on a clean container and washed with 200 mL of ultrapure distilled water, 18 Mega ohm (Ω) grade water and rocked gently for 5 minutes using the orbital shaker. The washing step was repeated for three times with ultrapure distilled water. After 15 minutes washing step, the gels were stained with 100 mL of the Coomassie stain solution (BioRad, USA) at room temperature for at least 1 hour or overnight. The staining solution were then removed and replaced with the ultrapure distilled water. The gels were subsequently rinsed with ultrapure distilled water for several times until clear spots appeared with transparent background.

Silver staining

The silver staining procedures were performed with some modifications according to the protocol published by Yan *et al.* (2000). This silver staining protocol is compatible with the subsequent matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization-mass spectrometry (ESI-MS). All the solution used in the silver staining protocol has to be prepared fresh and all the chemicals must be fully dissolved. Purity of the water would affect the staining results, thus, ultrapure water 18 Mega ohm (Ω) grade water was used for the whole staining process. After gently removing the 2D gels from the glass plate, gels were rinsed with ultrapure water for 5 minutes. The washing step was repeated for three times, a total of 15 minutes washing step. It is then followed by fixing the gels with fixative solution for 30 minutes and gently shaked on an orbital shaker.

After fixation, decant fixative solution was replaced with sensitizer solution for the subsequent 30 minutes. Next, the gels were washed with ultrapure water for 5 minutes and repeat for three times. After that, the gels were incubated with silver solution for 20 minutes. After silver impregnate, the gels were washed with ultrapure water for twice. The chilled developing solution was added to the silver-stained gels for 1-3 minutes. In order to stop the reaction, developing solution will be decanted and rinsed with the stop solution for 10 minutes. The gels were then kept in ultrapure distilled water.

3.2.5.4 Imaging and data analysis

The silver stained proteins were visualized as digital images captured on a GS-800^{TM} calibrated densitometer (Figure 3.3) (BioRad, USA). All the imaging process analysis was carried out according to the PDQuest analysis Quick Guide. The protein spots on the 2-DE gels were automatically matched and edited accordingly. Analysis set manager such as Student's t-test was then used to analyse the protein spots with the help of PDQuest version 8.0.1 Analysis Software (Figure 3.4) (BioRad, USA). The comparison analysis was carried out for the two different categories (high and low proliferation rate group) in term of protein relative density. The molecular weight (M_r) and pH (p*I*) of each protein was determined by referencing to protein marker appeared on separate gel. There are two biological replicates for high proliferation and low proliferation rate samples. And, all the 2-DE gels were run in three technical replicates.

Figure 3.3 An image of Densitometer GS800

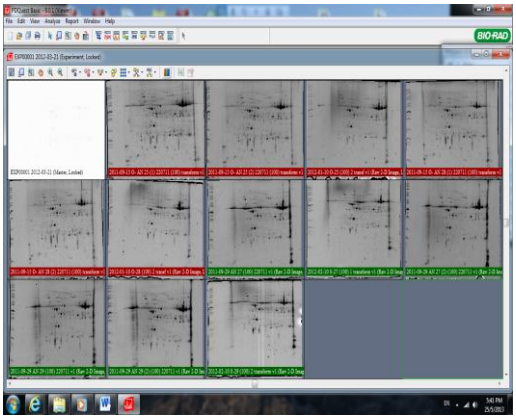


Figure 3.4 A snap photo of on going analysis using PDQuest software

3.2.6 Optimisation on Two-dimensional Gel electrophoresis

3.2.6.1 Four different protein extraction protocols

The protein samples extracted using the four protocols from the young and old oil palm leaf were separated in 2D gels. From the 2D gels, further comparisons can be made in term of number of protein spots present in these four extraction methods together with the patterns obtained in the 2D gels profiles.

3.2.6.2 The optimal range for the IPG strips

There are many different pH ranges for the Immobilized pH Gradient (IPG) strips available in the commercial market. In this experiment, only two different pH ranges of the IPG strips were selected for the 2D gels, namely pH 3-10 and pH 4-7.

3.2.6.3 Clean-up treatment with the ReadyPrep 2D clean-up kit

The clean-up for the samples were performed using ReadyPrep 2D clean-up kit. One hundred microgram of the protein was transferred to the 1.5ml centrifuge tube and clean-up was done according to the manufacturer instructions. The cleaned proteins were then subjected to isoelectric focusing (IEF) and two-dimensional gel electrophoresis.

3.2.6.4 Focusing time of the IPG strips

During the IEF run, four different focusing times were performed for the oil palm leaf samples which were 8,000 Vh, 10,000 Vh, 12,000 Vh, and 14,000 Vh.

3.2.6.5 Horizontal streaking treatment for the acidic regions

In order to reduce the horizontal streaking at the acidic end of the 2D gels, the extracted proteins were treated with nuclease mix after final extraction step and incubate for 45 minutes on ice. A few optimised conditions for the nuclease mix were done such as the additional acetone precipitation. After treatment, the 2D gels were performed to see the results.

3.3 Results

3.3.1 Protein Yield obtained with Different Protein Extraction Protocols

To the best of our knowledge, limited reports on protein extraction of oil palm for 2-DE studies are currently available. Therefore were applied several procedures to extract and solubilise the proteins present in oil palm leaf. There were TCA/Acetone precipitation method, Phenol extraction method, Lysis buffer extraction method and combination of both TCA/Phenol extraction methods. Each extraction protocol was repeated for six replicates for young leaf (Cabbage) and old leaf as well. The four protein extraction methods in this study produced samples that varied in their protein contents. These are presented in Table 3.2. Both young and old leaf samples exhibit the same pattern among these four protein extraction methods.

Table 3.2: Protein yield from four different protein extraction methods

Method	Protein yield (mg/g)	
	Young Leaf	Old Leaf
TCA/Acetone	$18.15^{a} \pm 3.02$	$31.01^{a} \pm 2.47$
Phenol	$5.18^{b} \pm 1.24$	$6.87^{b} \pm 1.35$
Lysis Buffer	$14.28^{a} \pm 1.88$	$29.00^{a} \pm 5.19$
TCA/Phenol	$8.32^{c} \pm 1.37$	$11.76^{c} \pm 3.64$

The results in Table 3.2 show that TCA/acetone extraction method produced a significantly higher protein yield (18.15 mg/g fresh weight) as compared to the other methods. The second highest yield was from the Lysis buffer extraction method (14.28 mg/g fresh weight), followed by the combination of TCA/Phenol extraction method and lastly is the Phenol extraction method.

3.3.2 One Dimensional Gel Electrophoresis – SDS-PAGE

All the proteins obtained from the four different extraction protocols were used for the one dimensional gel electrophoresis. Generally, one dimensional gel electrophoresis was performed prior to the two dimensional gel electrophoresis for the following reason. Firstly, SDS-PAGE gel were run to make sure there are proteins present in the samples before proceeding to 2-DE. It is an efficient method in term of time and energy to confirm the present of proteins in samples. Secondly, the pattern of the protein profiles can be estimated where all the proteins are separated by their own molecular weight as shown in the Figure 3.5.

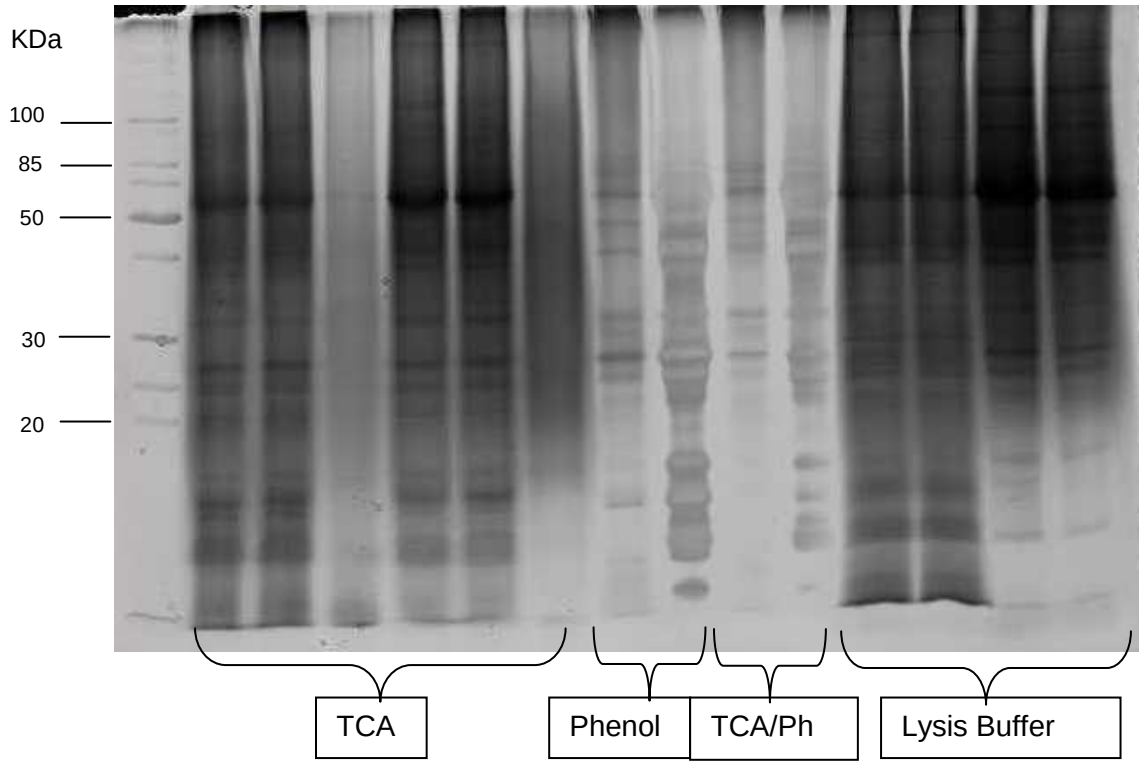


Figure 3.5: One Dimensional gel electrophoresis, SDS-PAGE for the four different extraction methods namely TCA/acetone, Phenol, TCA/Phenol and Lysis Buffer.

3.3.3 Optimised for the Two Dimensional Gel Electrophoresis (2-DE)

To date, only a few studies have described oil palm protein extraction for 2-DE analysis. Therefore, several procedures were used to extract and solubilize the proteins present in oil palm for both young and old leaf samples in an attempt to optimise the best extraction protocol. Four protein extraction methods were used and the proteins obtained were run on 2-DE analysis. The gel images obtained are shown in Figure 3.6a and the gels were subjected to PDQuest analysis to identify the number of protein spots present in each of the methods.

In previous work, Tan *et al.* (2011) reported that TCA/acetone extraction method produced the highest protein yields compared to the other three protein extraction methods assessed in oil palm. But, there are no assessment was made of the 2-DE gels profiles for four extraction methods tested in previous study. In this study, 2-DE gels analyses were carried out and the number of protein spots focused on 2-DE for each of the protein extraction methods were compared. After 2-DE separation, the 2D gels were analysed using PDQuest software version 8.0.1 (BioRad, USA).

The data in Table 3.3, shows that for samples extracted from the old leaf the highest number of protein spots resolved on 2-DE was obtained using the TCA/acetone precipitation method, followed by the lysis buffer method, phenol extraction method and lastly is a combination of TCA/phenol extraction methods which exhibited the lowest numbers of protein spots. There was no significant difference for the number of protein spots produced from the TCA/acetone methods compared to the other methods in old leaf. On the other hand, young leaf samples have significant difference with their highest number of protein spots presence as compared to the other three methods. Even though TCA/acetone method has no significant difference in the statistics for the old leaf, but those gels exhibited the highest level of reproducibility with the amounts of protein spots present in a 2D gels and the pattern for the 2D gels were close to each

other with the smallest standard error for the number of protein spots shown in each gel.

Table 3.3: Spot number in 2-DE protein gel for four different protein extraction methods

Spot numbers	TCA/acetone	Lysis buffer	Phenol	TCA/Ph
Old leaf	514.0 ± 15.3[a]	269.0 ± 140.9[a]	255.0 ± 62.5[a]	237.0 ± 119.8[a]
Young leaf	269.0 ± 14.0[a]	144.0 ± 29.0[b]	162.0 ± 1.0[b]	24.0 ± 7.5[c]

Different results were observed for samples from young leaf. The highest number of protein spots was obtained through the TCA/acetone precipitation method, followed by the phenol extraction and the lysis buffer extraction method, and the lowest number of protein spots was shown in the TCA/Ph methods in young leaf samples.

Despite this observation, the greatest protein concentration of samples did not contribute to the highest resolution profile on 2-DE gels. This was especially notable in the lysis buffer extraction (Figure 3.6) gel. This method produced samples with the second highest protein content, nevertheless the 2-DE resolution profiles was less clear than that produced by the phenol extraction method which had lower protein yields. In this study, the TCA/acetone precipitation method has been shown to be a highly effective protein extraction protocol for oil palm leaf samples, showing the highest protein yield and greatest number of protein spots on 2-DE gels.

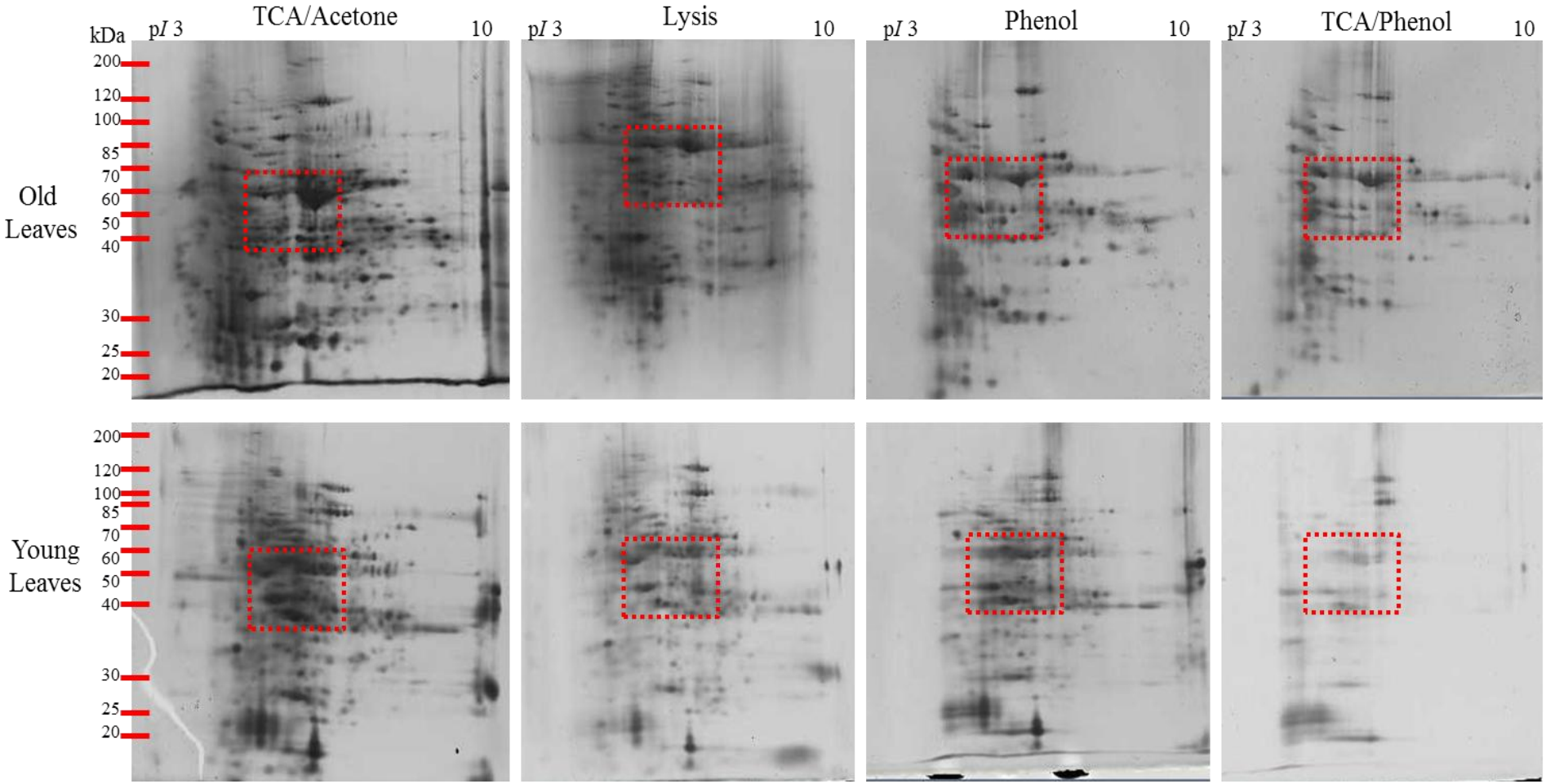


Figure 3.6: 2-DE protein profiles for oil palm leaf (old and young) for four different protein extraction methods using linear immobilized pH gradient strips of 3-10

(The enlargement image of the red dot boxes were shown in Figure 3.7)

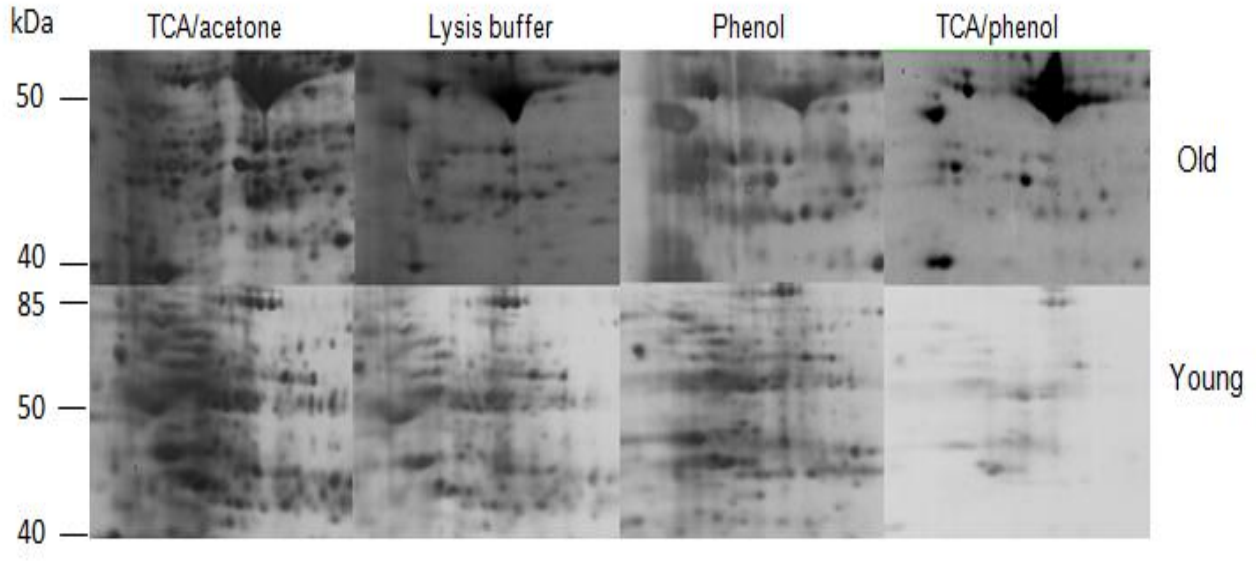


Figure 3.7: Close up 2-DE protein gels image for different extraction methods for old and young leaf samples.

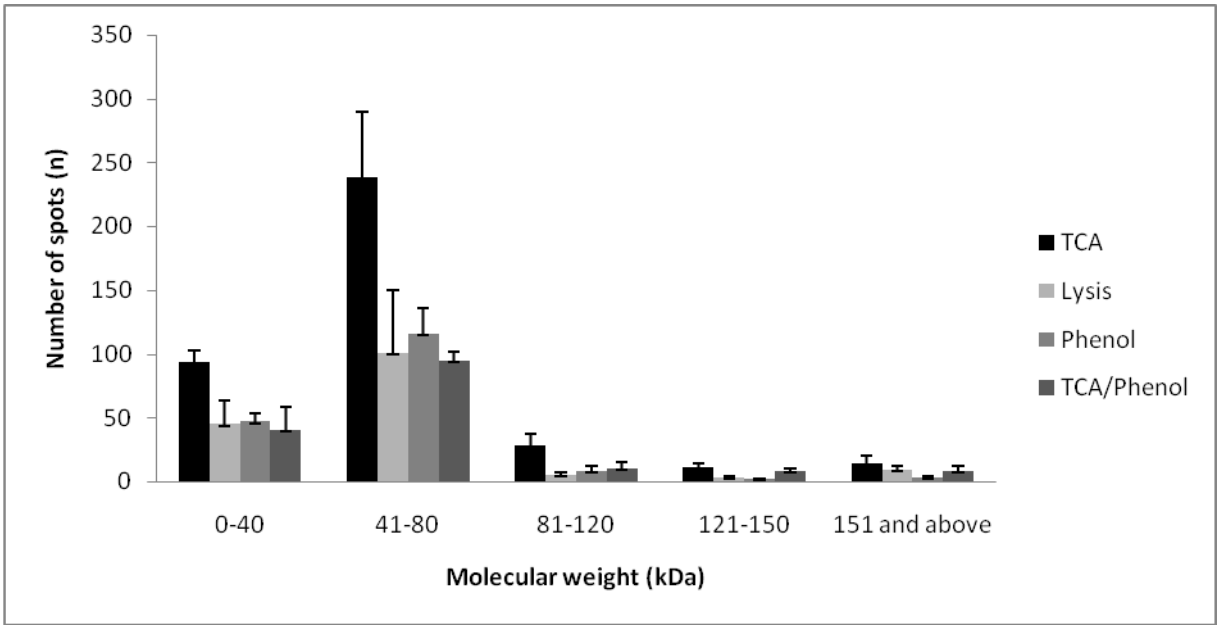


Figure 3.8: Analysis of the number of 2-DE protein spots detected in different molecular weight ranges with samples obtained from TCA/acetone, lysis buffer, phenol and TCA/Ph extraction methods

The ideal protein extraction protocol for any type of plant tissue will provide the highest amount of protein concentration and number of individually separable protein spots in a 2-DE gel. For all the protein extraction procedures assessed, the majority of proteins focused in the molecular weight range from 10 to 80 kDa (Figure 3.8).

Surprisingly, the highest number of protein spots exhibit the molecular weight from 41 to 80 kDa, followed by the smallest molecular weight 0 to 40 kDa. The proteins with molecular weight higher than 81 and above are very low in quantity.

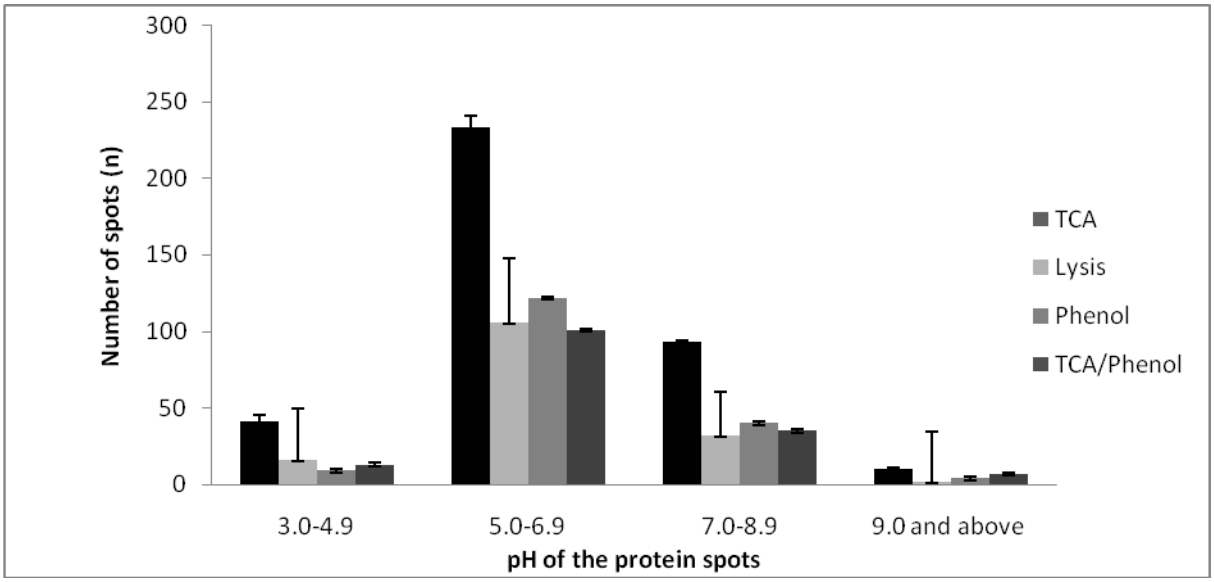


Figure 3.9: Analysis of protein spots in different pH range with samples obtained from TCA/acetone, lysis buffer, phenol and TCA/Ph extraction method.

The distribution of protein spots according to their pH are shown in Figure 3.9. Majority of the proteins are located between pH 5.0 to 8.9. However, the highest number of protein spots present in the pH ranges from 5.0 to 6.9, secondly is the more basic region, i.e. pH 7.0 to 8.9, followed by pH 3.0 to 4.9. These results were observed in all the protein extraction methods. The most acidic and the most basic regions have the limited number of protein presence.

3.3.3.2 Two dimensional gel clean up procedure for old and young leaf

In using plant tissues, clean up for the samples is often necessary in order to get a well resolved two-dimensional gel. Hence, in this project, clean up had performed using 2D clean up kit (BioRad, USA) for both the young and old plant materials with the aim to improve the gel resolution.

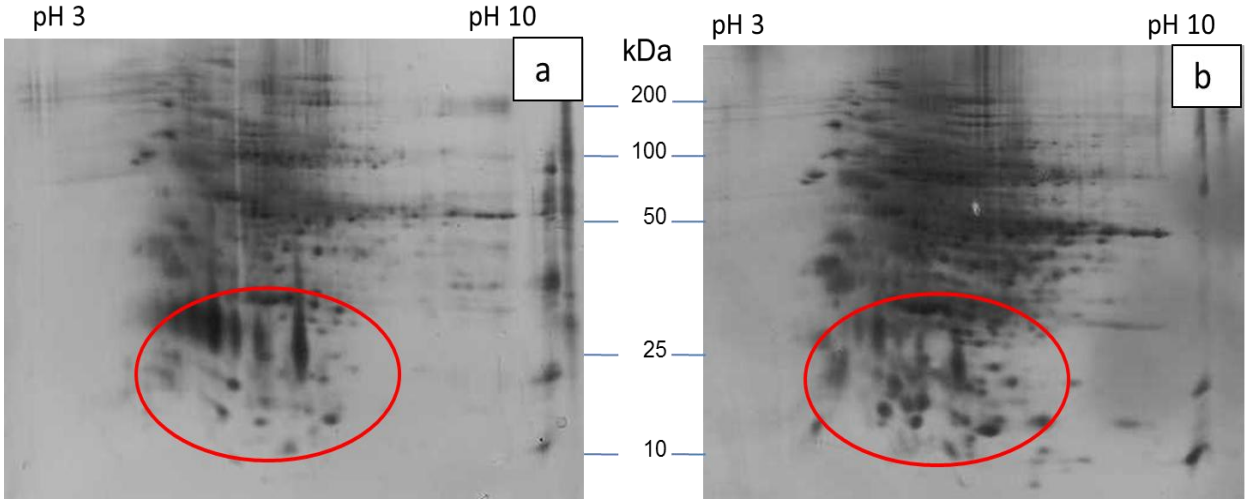


Figure 3.10: 2-DE protein profiles of young leaf samples (a) before clean up (b) after clean up

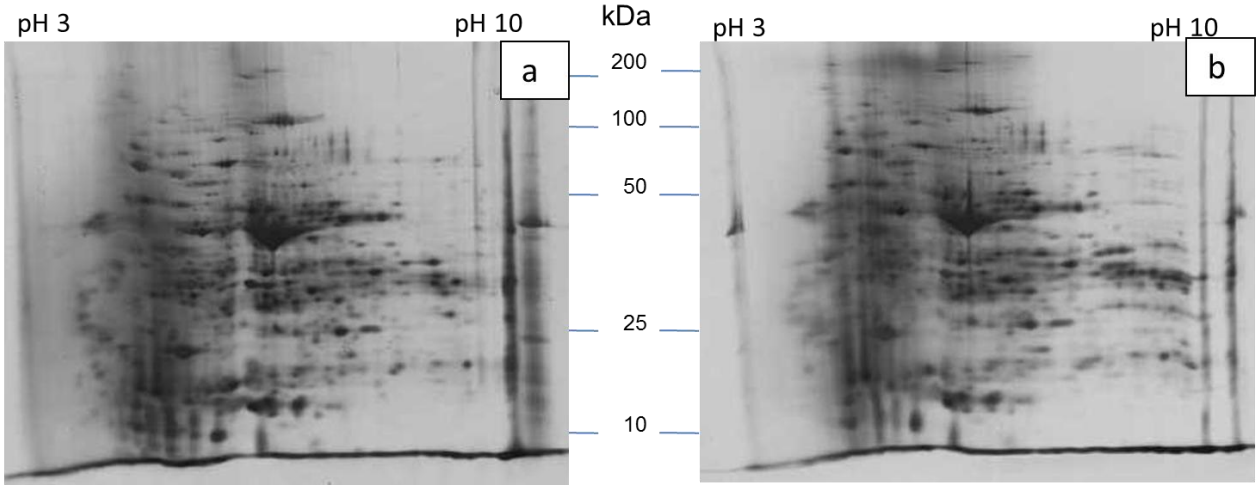


Figure 3.11: 2-DE protein profiles of old leaf samples (a) before clean up (b) after clean up

As shown in the Figure 3.10, there was more low abundance proteins appear in the young leaf samples (cabbage) after 2D clean up. However, the horizontal streaking problem remained and the background of the gels remained noisy. While the old leaf

samples (Figure 3.11), there was not much improvement shown after clean up. The condition of the overall vertical streaking problem remained in the gel.

3.3.3.3 Different range of the pH strips

Although the protein profiles were somewhat resolved the quality of the overall separation was unsatisfactory. The protein spots are clumped together (Figure 3.10) and this causes difficulties when further work such as comparing, quantifing or identify the protein spots need to be done. Therefore, a further modification of the 2-DE gels using IPG strips of a different pH range i.e.pH 4-7 was carried out. The outcome of the results (Figure 3.12) had given a better resolution of the protein spots resolved in 2D gels.

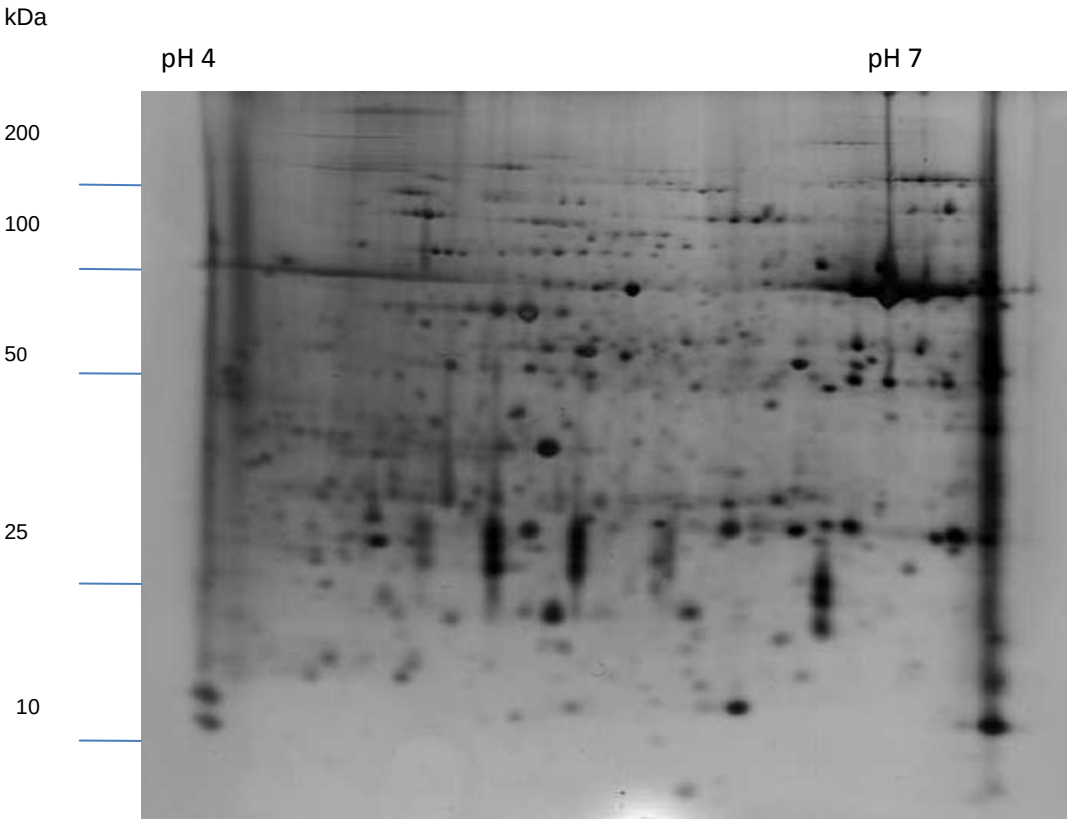


Figure 3.12: 2-DE protein profiles for old leaf sample on pH 4-7 IPG strip

3.3.3.4 Different focusing time

In addition to the extra clean-up step performed, different focusing times had been involved to further optimise the protocol in the hope of reduce the horizontal streaking on the 2D gels. There were a total of four different focusing times used to run the 2D gels, which are 8 kVh, 10 kVh, 14 kVh and 16 kVh. The results were shown in Figure 3.13.

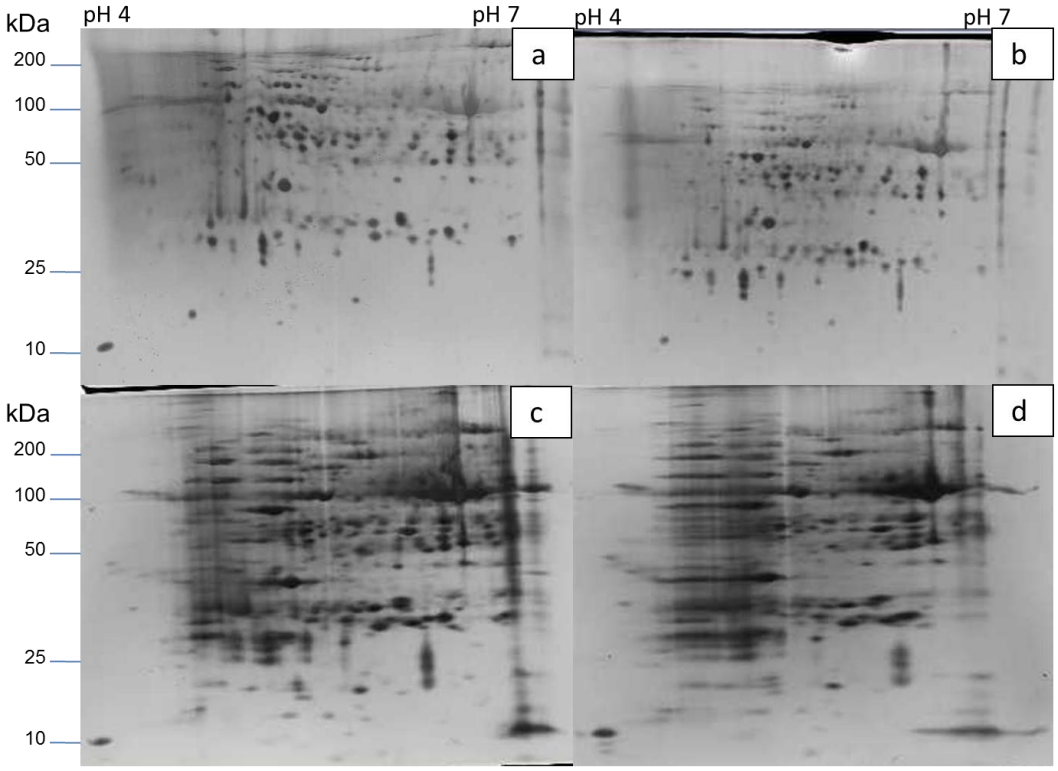


Figure 3.13: 2-DE protein profiles of oil palm leaf samples (cabbage) with different focusing times

(a) 8000 Vh (b) 10 000 Vh (c) 14 000 Vh (d) 16 000 Vh

Optimisation of focusing time gave a good resolution and focusing protein spots at 10 000 Vhr (refer to Figure 3.13 (b)). The 8000 Vhr gave second preferable focusing results and clear resolution as well, but there a slightly unfocused part at the positive end (Figure 3.13 (a)). As shown in the Figure 3.13 (c) and (d), the proteins in the gels start to be over-focusing with horizontal streaking as excess focusing time was applied.

3.3.3.5 Nucleic acid Elimination

Even though the optimal focusing time were applied, there was still a presence of horizontal streaking on the positive end (acidic) especially in the young leaf samples (cabbage). According to Görg *et al.* (2007), this might be due to the presence of impurities such as nucleic acid. In order to eliminate the nucleic acid from the samples, the effect of adding a mixture of nuclease was tested.

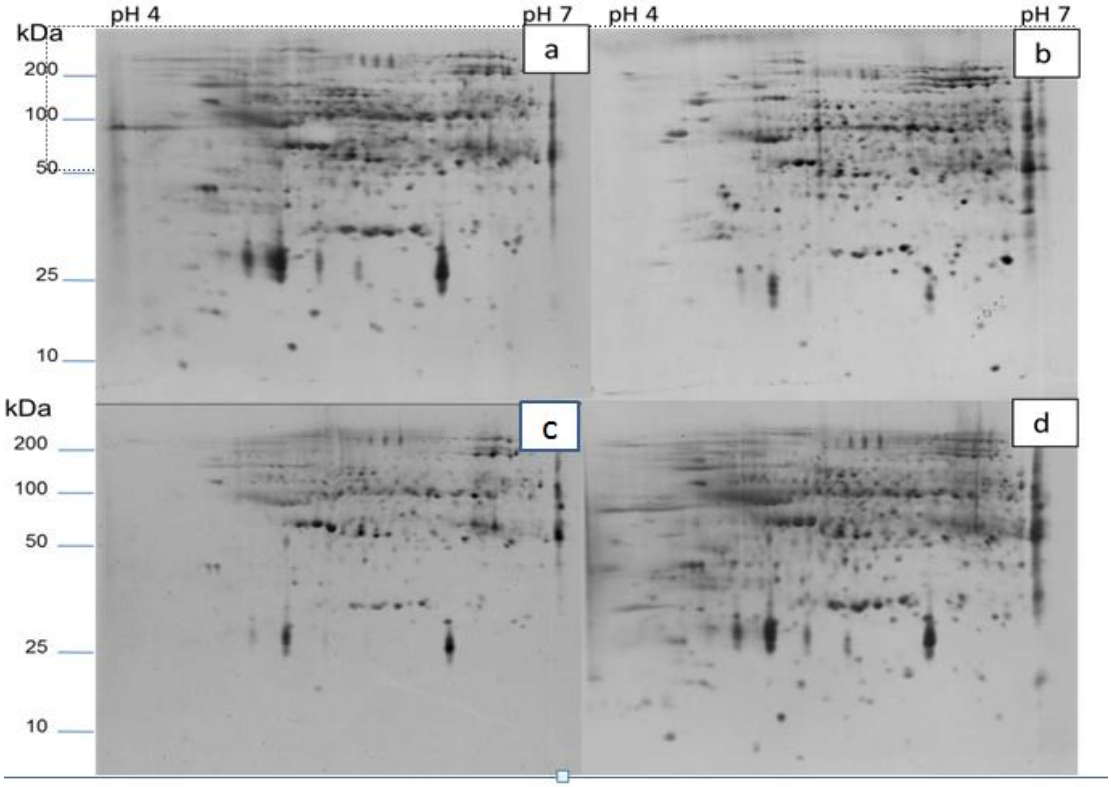


Figure 3.14: 2-DE protein profiles for samples (a) Untreated (b) nuclease treated (c) nuclease treated and acetone precipitated (d) acetone precipitation only

The positive end streaking problem was improved after nuclease treatment. Figure 3.14(c) shows that the addition of an acetone precipitation step after nuclease treatment did improve the horizontal streaking on the positive site but protein loss occurred as compared with nuclease treatment only (Figure 3.14(b)). The acetone precipitation alone did not improve the streaking problem at the positive end as shown in Figure 3.14(d).

3.4 Discussion

3.4.1 Improvement on protein yield obtained from different protein extraction protocols

The TCA/acetone extraction method produced a significantly higher protein yield. The same result is obtained from the date palm, *Phoenix dactylifera* L. leaf, a recalcitrant material which also found that TCA/Acetone precipitation method provides more efficient resolubilisation of date palm leaf proteins (Gómez-Vidal *et al.,* 2008). The second most efficient yield was from the Lysis buffer extraction method which has a shorter extraction period as compared to the TCA/acetone extraction method. This shorter extraction period may cause the incomplete extraction of full proteins from the samples, at the same time the elimination of impurities is not effective enough. The combination of the TCA/acetone and phenol method and phenol extraction method alone produced lower protein yields, this might be due to the time consuming procedures and longer steps involved. There are plenty washing step in the protocol, which are indirectly contributed to the loss of proteins prior to washing. The combination of both TCA/acetone and phenol methods has successfully extracted great amount of proteins in most of the plant species (Wang *et al.,* 2006). However, in this project, for the oil palm leaf, TCA/acetone precipitation methods show greater protein yield which are more than two fold increase in amount of protein extracted as compare to the other two methods. Several findings are in contrasting results, which suggested that the phenol extraction method is the technique of choice for recalcitrant plant tissues (Wang *et al.,* 2003; Saravanan and Rose, 2004; Carpentier *et al.,* 2005, Jellouli *et al.,* 2010). Other studies in agreement with our current findings show that the optimised TCA/acetone precipitation gave a relatively high protein yield, clear and high resolution profiles and highest number of protein spots (Xiang *et al.,* 2010; Wu *et al.,* 2011). TCA/acetone precipitation has been reported to be useful for minimising protein degradation and removing interfering compound (Gómez-Vidal *et al.,* 2008).

3.4.2 One Dimensional Gel Electrophoresis – SDS PAGE

From the SDS-PAGE gels, there are difference band patterns shown for the four extraction methods. Different protein extraction protocol will contribute to the dissimilarity of the total proteins extracted. Each extraction method will have their favour criteria that tend to extract difference ranges of the proteins which vary in their solubility and acidity in nature.

There are a range of extremely high metabolites present in the oil palm. Thus, it is not practical to cut the protein bands directly from the gels and send for identification using mass spectrometry. This might contribute to complexity in protein identification when multiple bands exist at the same position with identical molecular weight. Hence, two-dimensional gel electrophoresis need to perform in order to separate the proteins further in two dimensions so that a single protein spots with their particular pH and molecular weight can be observed.

3.4.3 Optimisations for the Two-Dimensional Gel Electrophoresis (2-DE)

A total of four protein extraction protocols were performed for the two dimensional gel electrophoresis were performed for the four protein extraction protocols. From the 2-DE gels, TCA/acetone shows the greatest number of protein spots on the 2-DE gels as compared to the other three methods, which is lysis buffer, phenol extraction method and a combination of the TCA/Phenol method. It seems logical where TCA/acetone should exhibit highest number of protein spots as it has the highest protein yield. However, even though a combination of TCA/Phenol has the higher protein yield than Phenol method. Phenol method has higher number of protein spots than combination TCA/Phenol method. These results show that the quantity of proteins is not correlate with the number of protein spots.

In addition to this, even TCA/acetone and Lysis extraction method presented a good protein yield, the quality of the 2-DE profiles were different. Two-dimensional gels from the TCA/acetone method show less horizontal streaking as well as the vertical streaking compared to the Lysis extraction method. Horizontal streaking mainly cause by the first dimension electrophoresis, i.e. the isoelectric focusing (IEF). This means that protein extracted using TCA/acetone methods performed well during the IEF compare to the Lysis extraction method. This results from the few replicates show that 2D gels from the TCA/acetone method was more consistent compared to the Lysis buffer extraction method. Hence, TCA/acetone was chosen as the most suitable protein extraction method in the study. This extraction method was subsequently standardised for both young and old leaf in further studies. In conclusion, the optimal protein extraction protocol for a good 2-DE gel should not be only based on the protein yield, it needs to be further confirmed by 2-DE gels before making a final decision.

Purity of the sample is very important during the 2D gel electrophoresis. There are several factors that might interfere with the 2D gels. One of the main factors is the presence of high impurities such as salt, phenolic compound; nucleic acid in the sample that may cause a serious streaking on the 2D gel. High impurities in the sample cause poor first dimensional separation through isoelectric focusing (IEF). The samples are unable to reach their optimum voltage and current during the IEF run with the high salt content. These causes the protein samples were unable to focus well at their own pI position. Even though there is a pre-focusing step during the IEF run, this steps would not be sufficient to eliminate all the impurities exist in a highly contaminant sample. Thus, the dirty samples can go through an extra clean up step. In order to clean up the impurities such as salt, further 2D clean up procedure is performed to eliminate any impure substances that might interrupt the isoelectric focusing run by using the 2D cleanup kit (Bio-Rad).

As shown in Figure 3.10, there is more low abundance proteins appear in the young leaf samples (cabbage) after 2D clean-up. This is one of the benefit of 2D clean up as it will help to remove the non-protein contaminant and increase the percentage of low abundance proteins present in a sample. This indicates that some of the samples required further clean up to have a better solubility and resolution of the proteins especially for the young leaf samples. However, the horizontal streaking problem remained and the background of the gels remains noisy. The horizontal streaking can cause by the incomplete focusing during first dimensional separation, IEF. However, for the old leaf sample, there is not much improvement shown after clean up (Figure 3.9b). The condition of the overall vertical streaking problem remained in the gel. This could indicate that the 2D clean-up kit might not work effectively for old leaf samples. Moreover, after 2D clean up procedure, it increases the chance of protein loss due to the excessive clean up steps involved.

3.4.3.3 Different range of pH strips

All the 2-DE gels in this study focused proteins across the pH ranges from 3 to 10. Further examination shows that protein spots were concentrated in the middle region of these 2-DE gels. This can cause difficulties and confusion when doing 2-DE gels comparison. Therefore, the narrower range of pH 4 to 7 was used instead of the normal range pH 3 to 10. The gel in Figure 3.8 shows that the pH ranges 4-7 provided improved separation for the protein spots. All the protein spots are presented as a single spots which can be clearly identified during spots picking. The improvement in the resolution of the 2-DE gel profiles make it possible to make any comparison between treatment species. A further advantage is that on narrow pH range gels, greater separation can be obtained between adjacent spots, which facilitate the excision of individually resolved separate proteins. It can be concluded that the IPG strips with pH 4-7 can provide good separation where by the entire gel can give a full image of the protein spots population in a 2D gel.

3.4.3.4 Different focusing time

In order to solve the horizontal streaking on the 2D gels, further optimisation on the focusing time is required for the leaf samples. Proteins exhibited an amphoteric characteristic with either positive, negative, or zero net charge. The isoelectric point (pI) is the pH where protein has a zero net charge at that pH. A protein with positive net charge will migrate toward cathode and slight change to less positively charged as it move through the pH gradient and reach their own p*I*. Similarly for the negatively net charge proteins migrate to the opposite direction toward the anode site.

The proteins in the IPG strips will be well focused and locate themselves at their own pI value with a particular focusing time. However, the proteins might lose their stability in position and results in electroendosmotic water and protein movement with extensive period of focusing, also known as over-focusing. This might cause a vertical streaking at the left end of the protein spots.

It is necessary to study suitable focusing time for a particular protein sample in order to avoid over-focusing or under-focusing condition for the protein species. A solid and distinct protein spots can be captured from the 2D gels with the optimum focusing time. With the well-focused protein, it will make the PDQuest software analysis more accurate and reliable when the spots differences were identified.

3.4.3.5 Nucleic acid Elimination

For the young leaf samples (cabbage), there are horizontal streaking present on the positive end (acidic) (Figure 3.10a). This may be due to nucleic acid contamination (Görg *et al.,* 2007). In the presence of nucleic acid, protein might bind together with the nucleic acid and formed a protein-nucleic acid mixture which can interrupt the IEF run. Figure 3.10 show that addition of the nuclease mix treatment effectively removes quite a large number of horizontal streaking. Hence, nuclease mix treatment will be used to treat the protein samples with streaking on the 2-DE gel in order to reduce the nucleic acid contaminants.

3.5 Conclusion

The TCA/acetone extraction procedure provided the most efficient and reliable method for preparing samples for 2-DE protein separation for both young and old oil palm leaf. The highest protein content and superior protein profile on the 2-DE gels were obtained where nuclease treatment was used. Our results also demonstrated that nucleic acid removal is effective in reducing horizontal streaking in the acidic region. This method that involved precipitation using TCA/acetone allowed instant elimination of proteolytic enzyme. An optimal isoelectric focusing time of 10,000 Vhr for pH 4-7 IPG strips, 7 cm in length was determined that provided good focusing profiles for protein samples extracted from the F17, old oil palm leaf samples. The 2D clean-up procedure shows a little improvement in the cabbage samples but not for the old leaf samples. In this chapter, the most efficient protein extraction for the old and young leaf samples were identified coupling with the optimal parameters for the two-dimensional gel electrophoresis. This optimising stage is very important to make sure all the samples produce high quality 2-DE protein prior to the downstream proteomics analysis.

CHAPTER 4

COMPARISON OF PROTEIN PROFILES BETWEEN OIL PALM LEAF SAMPLES WITH HIGH AND LOW PROLIFERATION RATES

4.1 Introduction

Oil palm (*Elaeis guineensis* Jacq.) is an important commercial crop in Malaysia. It is a diploid monocotyledon with single vegetative apex (Low *et al.,* 2008). Due to its high commercial value and high demand for high yielding planting material, somatic embryogenesis has become one of the alternatives in clonal propagation of oil palm to supply the elite oil palm plantlets (Thuzar *et al.,* 2011). Generally, young leaves of oil palm are chosen as a starting material for tissue culture to induce callus due to their efficiency. However, the callus could develop into soft, granular and translucent tissues which do not have embryogenic potential and incapable of regeneration into a new plantlet. Thus, the callus formation and development of somatic embryos become one of the major bottlenecks in oil palm tissue culture as well as the clonal abnormality problem. A very low rate approximately 19% of callogenesis of oil palm has been reported by Corley and Tinker (2003) and an average rate of embryogenesis in leaf derived callus is only 6% (Wooi, 1995). To date, there are limited information available about the molecular changes associated with callogenesis and embryogenesis in oil palm. Malaysia Palm Oil Board (MPOB) has been actively working on the gene expression studies in both embryogenesis and clonal abnormality and some potential gene markers have been identified but functionality of these proteins associated with embryogenesis is still unknown (Sambanthamurthi *et al.,* 2009). There is still too early for researchers to solve the obstacles in somatic embryogenesis due to limited scientific finding regarding the somatic embryogenesis. Thus, the aim of this experiment is to employ a proteomic approach to determine the protein expression profiles of the low and high proliferation rates in embryogenesis lines of oil palm tissue culture in order to

understand more on the biological processes occur through the protein expression study.

4.2 Materials and methods

4.2.1 Plant Materials

The plant materials used in these experiments were collected from Advanced Agriecological Research Sdn Bhd (AAR) and United Plantation (UP), Malaysia. In this experiment, two types of samples collected which were young and old oil palm leaves. Two categories of the oil palm old leaf samples have been classified as high and low proliferation rates based on previously recorded tissue culture performances. Oil palm samples producing proliferation embryogenic line of more than 30 were categorised as high proliferation rate samples while low proliferation rate samples were shown with proliferation embryogenic line of less than 10. Samples AN 25 and AN 28 were categorized as high proliferation rate, while AN 27 and AN 29 were grouped as low proliferation rate in old leaf samples. In young leaves samples, there were three biological replications. The high proliferation rate samples (933, 948, and 944), and low proliferation rate samples (194, 1086 and 352) were compared and for each of the samples. All the samples were extracted using the optimised protocol, TCA/acetone method as described in Chapter 3.

4.2.2 Two-dimensional gel electrophoresis

The extracted proteins were rehydrated overnight using 7 cm IPG strips for the old leaf samples and 17 cm IPG strips for young leaf samples. The IEF was then carried out on the next day for 6 hours and subjected to second dimension

separation using SDS-PAGE. All the optimal running parameters were used as described in the Chapter 3 (Section 3.3.5.2). For the large 2-DE 17 cm gel, the IEF running was 60,000 Vhr which was about 9 hours running time. There was standard electrophoresis condition for 17 cm (16 mA running per gel for a 30 minutes, followed by 24 mA per gel for 5 hours). It was then stained with Coomassie Blue G-250 or silver stain. The images of the 2-DE gels were digitalised using the densitometer (Bio-rad). The digital images were subjected to the PDQuest analysis version 8.0.1 and Progenesis SameSpot software for spot analysis. The differential protein expression profiles for particular spots between the high and low proliferation rate samples were selected followed by protein identification using mass spectrometry (Figure 4.1)**.**

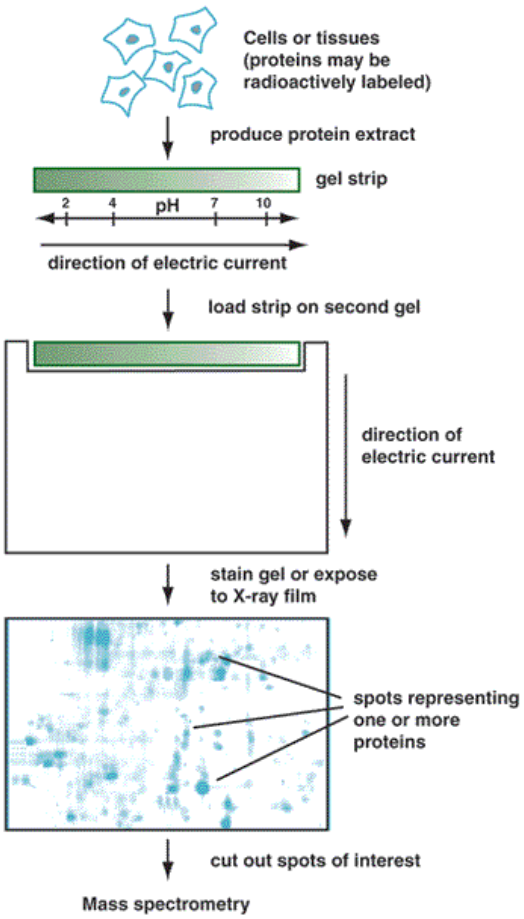


Figure 4.1 Proteomic workflow showing from the two-dimensional gel electrophoresis to mass spectrometry (Susanne et al., 2008)

4.2.3 Excision of protein spots from the polyacrylamide gels

Differences in protein spot expression between high and low proliferation rates in oil palm samples were analysed through the PDQuest software 8.0.1. The protein spots that exhibited significant difference were excised using ProteomeWorks spot cutter (BioRad) and placed in a clean microcentrifuge tube before trypsin digestion (Soskie *et al*, 1992). The protein spots were then pooled together as 4 spots to 8 spots prior to the mass spectrometry analysis. This is to confirm the protein spots can reach the concentration threshold in MALDI TOF/TOF instrument. The individual protein spots were performed in-gel digestion using trypsin enzyme. The peptides were then extracted and spotted on the MALDI plate and send to mass spectrometry to identify the mass spectrums of the peptide. The mass spectra from the peptides were used to blast through the MASCOT search engine with accessible protein database such as Swissprot and NCBInr (Figure 4.2).

4.2.4 *In situ* digestion of proteins

4.2.4.1Destain of gel pieces

a) Coomassie stain gels

Prior to peptide digestion, the stained gel plugs were went through destaining procedures. For the coomassie staining approach, gel plugs were destained with 200 mM ammonium bicarbonate. Two hundred microliter of the ammonium bicarbonate (NH_4CO_3) was added to each tube and incubated at 37 °C for 30 minutes. The solution was then discarded and repeated the steps until the gel plugs were cleared from blue coomassie stain.

b) Silver stained gels

The destaining procedure of silver stained protein in gel pieces were performed as described (Farzin *et al.,* 1999). The gel plugs were destained with chemical reducers to remove the silver ion. Potassium ferricyanide and sodium thiosulfate were used as reactive substances of the chemical reducers. The solution of the chemical reducers which made up of 30 mM potassium ferricyanide and 100 mM sodium thiosulfate were prepared freshly. For each tube, 100 µL destain solution was added and incubated at room temperature under dark until clean gel plugs obtained. The destaining solution was then discarded and gels were washed with 150 µL of 200 mM NH_4CO_3 for 20 minutes.

4.2.4.2 Reduction and alkylation

Reduction of the protein were carried out using 100 µl 10 mM DTT/ 50 mM NH_4CO_3, followed by vortexing and brief centrifugation. It was then incubated at 56 °C for 45 minutes to reduce the protein. The supernatant was discarded and the gels were washed with 100 mM NH_4CO_3 with vortexing and brief centrifugation. In-spot reduction was recommended even the proteins were reduced prior to an electrophoresis run. It was then alkylated by vortexing with 100 µL of 55 mM iodoacetamide in 50 mM NH_4CO_3 for 30 minutes at room temperature under dark condition. The solution was discarded and the gel pieces were washed with 50 mM NH_4CO_3 in 50% acetonitrile (ACN). Additional washing might be required if the residual staining was observed.

4.2.4.3 *In-gel* digestion with trypsin

The reduced and alkylated gel plugs were then added with 150 µL ACN and incubated for 10 minutes. After incubation, discarded solution and allowed to dry under vacuum in Speed Vac (Eppendorf, German) for 5 minutes. The gel plugs were rehydrated with 80 µL digestion buffer (10% ACN in 50 mM NH_4CO_3) consisted of 12.5 ng/µL trypsin which was sufficient to cover gel plug. It was then vortexed gently for 5 minutes. Lastly, all the tubes were incubated overnight for 16 to 18 hours at 37 °C.

4.2.4.4 Peptides Extraction

Peptide extractions were carried out on the second day after overnight incubation. The gel slurry was centrifuged briefly and supernatant was transfered to a clean 1.5 mL Eppendorf tubes. Secondly, 100 µL of freshly made 10% ACN in 50 mM NH_4CO_3 were added to the gel plug and vortexed for 10 minutes before centrifuged it down to collect the supernatant. The supernatant that collected from the first round centrifugation was pooled together with the second extracted supernatant. It was continued with the third extraction with 50 µL 0.5% trifluoroacetic acid (TFA). Lastly, the gel plugs were further extracted with 80% ACN and vortexed vigorously for 2 minutes before centrifugation. All the collected supernatant was pooled together and dried completely under vacuum.

4.2.4.5 Identification of proteins by Mass Spectrometry

The digested peptide samples were re-constituted with 0.1% TFA before subjected to the MALDI TOF/TOF instrument (Bruker, Germany). Matrix preparation was performed as introduced by (Hillenkamp and Karas, 1988) with little modifications. The saturated matrix was prepared freshly and diluted before use. Prior to the acquicisition of spectra, 1 µL of each peptide mixture was mixed with 1 µL of matrix solution and a droplet of the resulting mixture (1 µL) was placed on MALDI target plate. The mixture was then allowed to dry at room temperature.

The MALDI plate was loaded into the mass spectrometry instrument for protein identification. The instrument was equipped with a smart beam laser and acquisition laser power was optimized using the peptide mass calibration mixture before collection of sample data. Mass spectra were acquired in positive ion reflection mode in mass range 480 – 5500 Da with a suppression mass gate set to 450 Da to prevent detector saturation from matrix cluster peaks. Default operating conditions were as follow: ion source 1, 25.0 kV; ion source 2, 21.80 kV; lens voltages, 9.5 kV; reflector voltage, 26.3 kV and reflector 2 voltage, 13.7 kV.

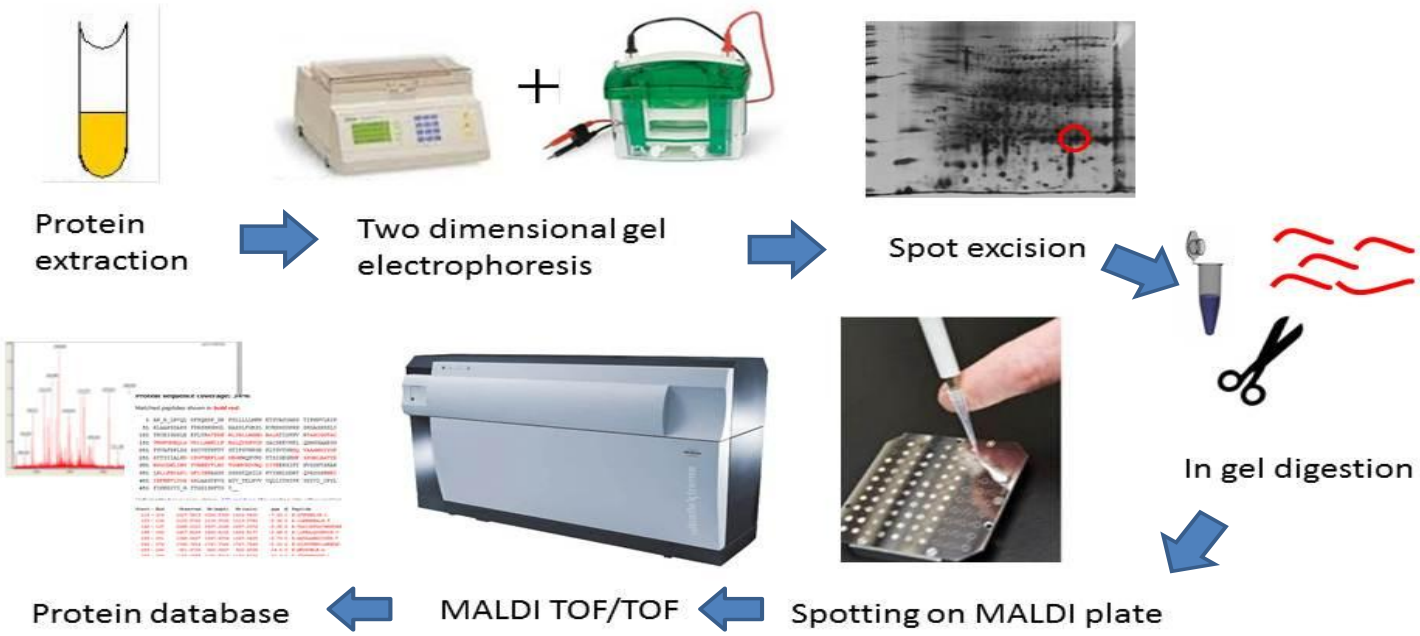


Figure 4.2: Proteomic workflow from spot excision to protein identification and database search.

All acquisitions were generated automatically in the instrument software and based on 800 shots. Data were calibrated externally with the peptide mass calibration mixtures. All spectra were processed and analsysed using Flex Analysis and Biotool softwares. The masses of the peptides come from the trypsinzed protein spots were subjected to MASCOT search engine against in-house (MPOB Transcript), Swiss-Prot (Figure 4.3) and NCBInr databases.

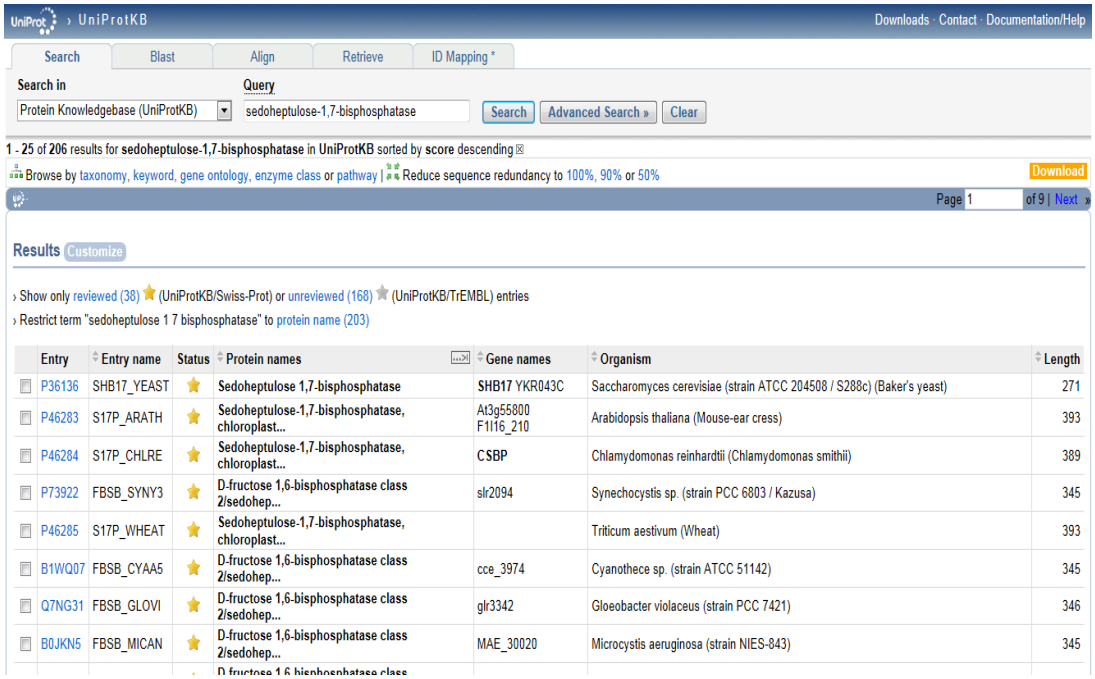


Figure 4.3: UniProtkb, Swissprot database

4.3 Results

4.3.1 Global profiling of the old leaf sample in oil palm

The protein extracted from oil palm samples were subjected to the optimised IEF and SDS-PAGE separation as described in Chapter 3. The proteome profiles obtained are shown in Figure 4.4. There was limited number of protein spots present in the 2-DE gels which only enabled a certain degree of proteins to be accessed at a particular time. Approximately, a total of 100 to 150 individual protein spots were detected in the samples. The protein profiling through 2-DE gels provides general information for the protein populations present in an old leaf sample especially oil palm.

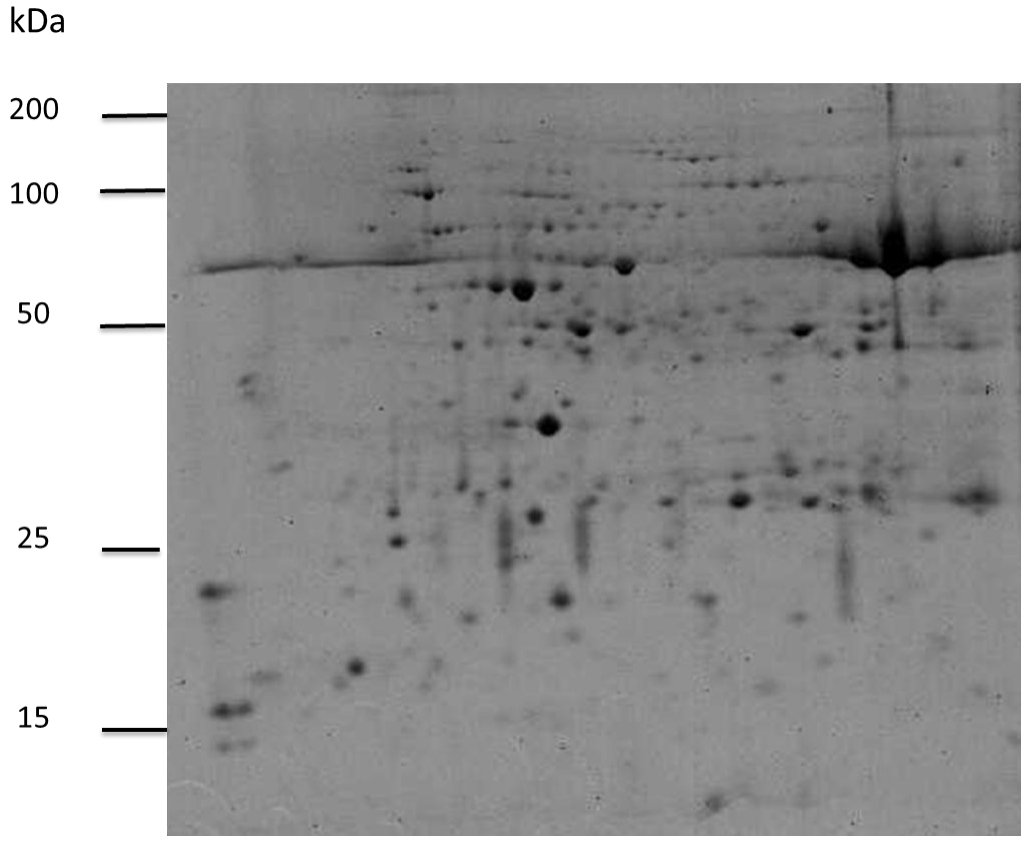


Figure 4.4: Two-dimensional protein gel for the oil palm old leaves sample

The identification rate for protein interests was 50%. The identities of each protein spot are shown in Table 4.1. The results provide p*I* and molecular weight (MW) for each of the identified proteins. In addition, the biological and molecular functions of the proteins as well as their cellular location have been given. This information provides further understanding on the protein populations that were expressed in the old leaf samples of the oil palm species. All of the identified proteins occurred abundantly in the oil palm leaf samples.

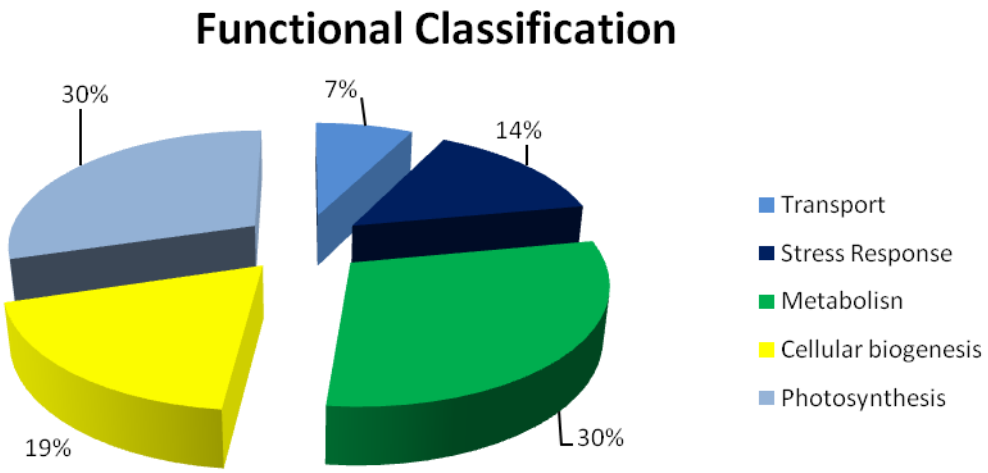


Figure 4.5: Functional classification of total protein expressed in oil palm leaf samples.

The 27 identified protein spots were classified into five groups namely transport, stress response, metabolism, cellular biogenesis, and photosynthesis based on their functions in the plant species (Figure 4.5). Most of the identified proteins were involved in metabolism and photosynthesis (30%), followed by cellular biogenesis (19%), stress response (14%) and lastly transportation (7%). Similar results from Shim *et al.* (2010) who reported high detection of enzymes involved in photosystem reactions in chloroplast for the protein profiling of *Piper sarmentosum* plant.

Table 4.1: Protein profiling from the oil palm leaf samples with 27 spots identified.

Spot No	Protein name	Biological process	Molecular function	Cellular location	Reference organism	Accession No	MW	p*I*	MOWSE
1.SSP 2107	RuBisCo large subunit-binding protein subunit alpha	Stress response	Binds RuBisCo small and large subunits and implicated in the assembly of the enzyme oligomer	Plastid, Chloroplast	Chlamydomonas reinhardtii	Q42694	61999	5.57	65
2. SSP 2202	2-Cys peroxiredoxin BAS1-like	Defence response to bacterium and cold	Antioxidant Peroxidase activity Oxidoreductase	Apoplast, Chloroplast stroma, stromule	Arabidopsis thalinana	Q9C5R8	29932	5.4	104
3. SSP 2305	Chlorophyll a-b binding protein of LHCII type III	Photosynthesis	Light receptor, captures and delivers excitation energy to photosystem	Plastid, Chloroplast membrane, Thylakoid	Hordeumvulgare	P27523	28798	4.99	51
4. SSP 3305	Triosephosphate isomerase	Glycolysis	Catalyzesinterconversion of dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate. Isomerase	Cytoplasm	Gossypium hirsutum	D6N3G7_GOSHI	51648	8.46	105
5. SSP 3306	Oxygen-evolving enhancer protein 1	Photosynthesis Photosystem II stabilization	Calcium ion binding	Chloroplast thylakoid membrane, Oxygen evolving complex	Fritillaria agrestis	O49079	35076	6.3	84
6. SSP 3401 And 7. SSP 3402	Sedoheptulose-1,7-bisphosphatase	Carbohydrate metabolism; Calvin cycle	Light activation through pH changes, Mg2+ levels and light-modulated reduction of essential disulphide groups via ferredoxin-thioredoxin f system.	Plastid, Chloroplast	Triticum aestivum	P46285	42547	6.0	49 And 188

8. SSP 4405	Sedohepulose-1,7-bisphophatase	Calvin cycle, carbohydrate metabolism, Reductive pentose-phosphate cycle	Hydrolase, metal ion binding	Plastid, Chloroplast	Spinacia oleracea	O20252	42568	5.87	106
9. SSP 4202	Oxygen=evolving enhancer protein 2	Photosynthesis	Regulation of Photosystem II	Plastid, Chloroplast thylakoid membrane	Solanum tuberosum	P93566	28158	8.27	76
10. SSP 4505 And 11. SSP 5401	Ribulosebisphosphate carboxylase/ oxygenaseactivase B	Leaf senescence, Response to jasmonic acid stimulus, light stimulus	ADP and ATP binding, Ribulose-1,5-bisphosphate carboxylase/oxygenase activator activit	Cell wall, chloroplast envelope, nucleus, plastoglobule, stromule	Arabidopsis thaliana	At2g39730	47426	8.6	97 And 153
12. SSP 5403	Photosystem II stability/assembly factor HCF136	Photosynthesis	Essential for photosystem II (PS II) biogenesis, required assembly of an early intermediate in PSII assembly and Chlorophyll a binding.	Chloroplast membrane, Plastid, Thylakoid	Arabidopsis thaliana	O82660	44133	6.79	42
13. SSP 5404	Phosphoribulokinase	Reductive pentose-phosphate cycle Defence response to bacterium/cold	ATP binding, Phosphoribulokinase activity	Apoplast, chloroplast envelope, stroma, thylakoid, stromule	Chlamydomonas reinhard	P19824	42151	9.0	103
14. SSP 6201	Cytochrome B6-F complex iron sulphur subunit 2	Electron transport	Iron sulphur protein, Rieske domain involved in electron transfer, metal ion binding, ubiquinol-cytochrome-c reductase activity.	Integral to membrane, thylakoid membrane	Sonneratia ovate	122816	60147	8.60	135

15. SSP 6202	Oxygen-evolving enhancer protein 2	Photosynthesis	Regulation of Photosystem II	Plastid, thylakoid, chloroplast membrane	Fritillaria agrestis	O49080	28265	8.31	75
16. SSP 6304 And 17. SSP 7304	L-ascorbate peroxidase 2	Embryo development ending, Hydrogen peroxidase catabolic process, Response to cadmium ion, heat and salt stress	L-ascorbate peroxidase activity, Heme binding, metal ion binding	Cell wall, chloroplast stroma, plasma membrane	Oryza sativa japonica	Q9FE01	27215	5.1	130 And 149
18. SSP 6501	Elongation factor TuB	GTP catabolic process	GTP binding, GTPase activity, Translation elongation factor activity	Chloroplast	Nicotiana sylvestris	Q43364	53076	6.3	95
19. SSP 7307	Ribulosebisphosphate carboxylase large chain	Photorespiration Reductive pentose-phosphate cycle	Magnesium ion binding, Monooxygenase activity, Ribulose-bisphosphate carboxylase activity	Chloroplast	Spinach Oleracea	P00875	49258	6.6	58
20. SSP 7402	Malate dehydrogenase, Mitochondria	Glyoxylate cycle, Malate metabolic process, Tricarboxylic acid cycle	L-malate dehydrogenase activity, nucleotide binding	Glyoxysome	Citrullus lanatus	P19446	36406	9.6	148
21. SSP 7403	Glutamine synthetase root isozyme 4	Glutamine biosynthetic process	ATP binding, Glutamate-ammonia ligase activity	Cytoplasm	Zea mays	P38561	39241	5.1	218

22. SSP 7406	Glutamine synthetase N-1	Glutamine biosynthetic process, nitrogen fixation	ATP binding, glutamate-ammonia ligase activity	Cytoplasm	Phaseolus vulgaris	P00965	39488	6.1	124
23. SSP 7504	Phosphoglycerate kinase, chloroplastic	Glycolysis, reductive pentose-phosphate cycle	ATP-binding, phosphoglycerate kinase activity	Chloroplast	Spinach oleracea	P29409	45658	5.8	97
24. SSP 7702	ATP synthetase subunit alpha, chloroplastic	ATP hydrolysis coupled protein transport, Plasma membrane ATP synthesis coupled	ATP binding, proton-tranporting ATP synthase activity, proton-transporting ATPase activity, rotational mechanism	Chloroplast thylakoid membrane, proton-transporting ATP synthase complex, catalytic core	Acorus americanus	A9LYH0	55376	5.1	215
25. SSP 7606 And 26. SSP 8605	Ribulosebisphosphate carboxylase large chain	Carbon dioxide fixation, photorespiration, photosynthesis, Calvin cycle	Primary event in carbon dioxide fixation, oxidative fragmentation in photorespiration	Plastid, Chloroplast	Acacia farnesiana	P93998	50860	6.22	220
27. SSP 8405	Fructose- biphosphate aldolase 1	Glycolysis	Fructose-bisphosphatealdolase activity	Thylakoid lumen, apoplast, chloroplast envelope	Arabidopsis thaliana	Q9SJU4	43075	6.18	45

4.3.2 Comparison of protein profiles between high and low proliferation rate samples

The proteins of these two categories of samples were extracted and run in two-dimensional gel electrophoresis to get a proteome profile and compare their differential protein expression. Comparison of the protein profiles of high and low proliferation rates of the oil palm samples were conducted using leaves. Different octets and genotypes of oil palm were reported to have different embryogenic potency rate. Recent publication by Jayanthi *et al.,* (2015) reported that the embryogenesis percentage obtained from *tenera* palms (6.8 - 9.35%) were higher than dura palms (0.33 – 4.98%) under the same media treatment.

4.3.2.1 Old leaf samples

Proteins were extracted from the high and low proliferation rate group samples and separated using 2-DE gels. Images of the 2-DE gel protein profiles were digitised using densitometer prior to the analysis using 2-DE analysis software. Figure 4.6 shows the 2-DE gel images for the two sample categories, high proliferation rate (Figures 4.6a and 4.6b) and low proliferation rate (Figures 4.6c and 4.6d) samples. Each sample was run in triplicate. The protein profiles obtained from the old leaf samples are highly reproducible across the technical replicates.

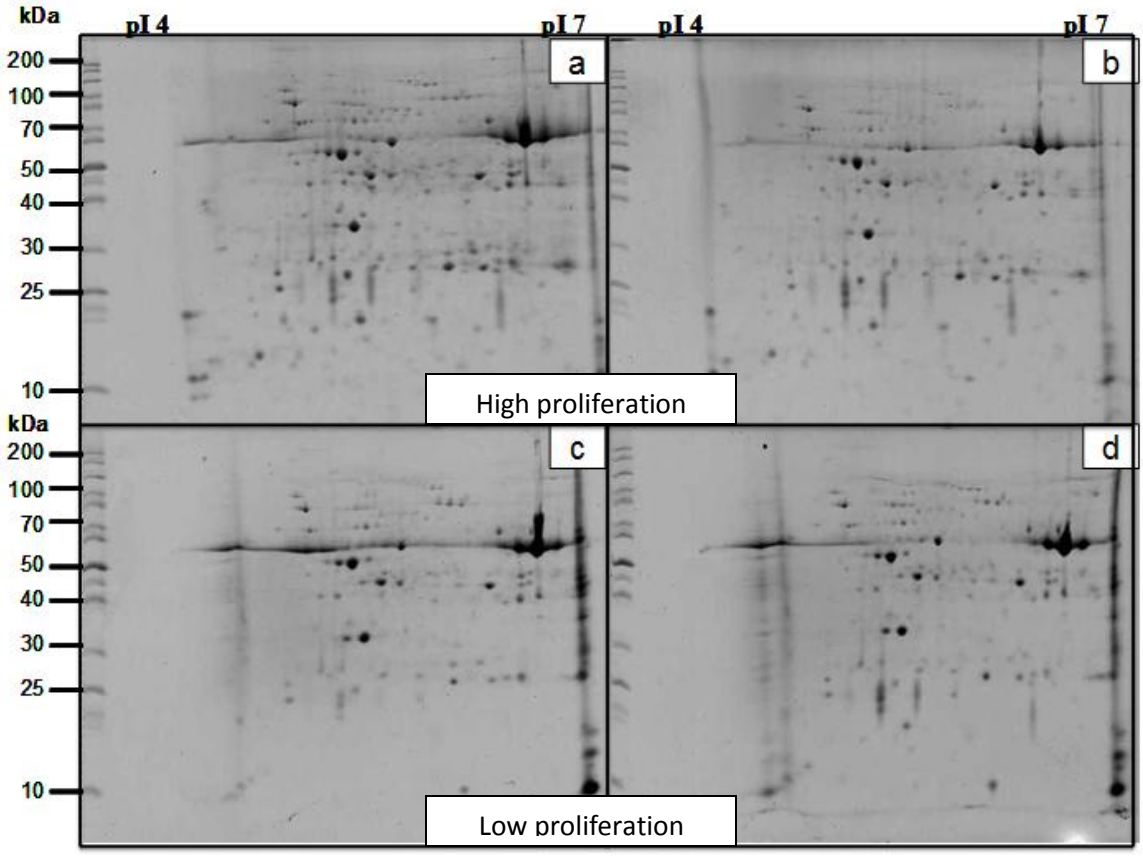


Figure 4.6: 2-DE gel images for AN 25 (a), AN 28 (b) which were high proliferation rate samples, AN 27 (c) and AN 29 (d) which were low proliferation rate samples

The distributions of the protein spots across the four biological samples were quite similar but the differential expressions between protein spots were not detected through naked eyes. All the gel images were subjected to PDQuest software analysis (Figure 4.7) for automated spot detection, matching, normalization and quantification. After matching all the protein spots, the gels were analyzed for statistically significant in fold change differences between samples together with the student t-test analysis.

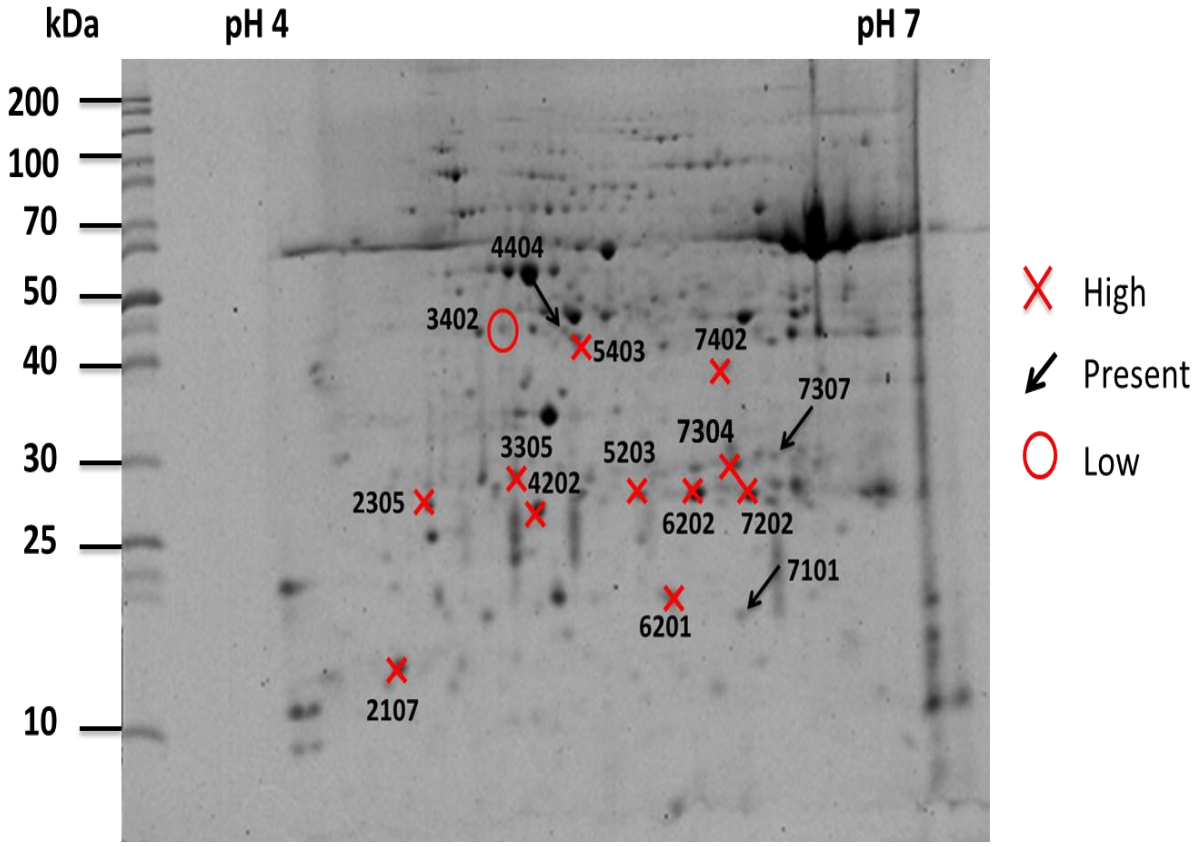


Figure 4.7: Protein spots that exhibit significant differences between high and low proliferation rate categories are shown.

PDQuest software provided the relative density of individual protein spots across all the protein gels in each category. There are a total of 16 protein spots expressed significantly with 10-fold change difference and pass the student t-test analysis at 95% confidence level. Thirteen protein spots were found to be highly expressed in high proliferation rate samples. Two proteins were detected only present in the high proliferation rate samples and there was one protein spot (SSP 3402) having low expression in high proliferation rate samples (Table 4.2). All the proteins were sent for MALDI TOF/TOF analysis and there were 13 protein spots identified out of 16 protein spots. The identification rate was roughly 81% for the old leaf samples (Table 4.4).

Table 4.2 Number of protein spots that exhibit differential expression in the high and low proliferation rate samples.

Protein expression	Total protein spots	Protein number
High expression in High Proliferation rate samples	13	2107, 2305, 3305,4202, 5203, 6201, 6202, 5403, 7304, 7402, 7202, 8405, 8605
Low expression in High Proliferation rate samples	1	3402
Present only in High	2	7101, 7307
Total	**16**	

4.3.2.2 Young leaf samples

The extracted proteins from three biological samples in young leaf samples were run on 17cm long 2-DE gel respectively. Young leaf samples was found to be higher challenges as compared to the old leaves due to its low protein concentration during initial protein extraction stage. All the protein gels in young leaves were stained using conventional silver stained method (Section 4.2.4.1b). All the samples were run in triplicate. Figure 4.8 shows the young leaf protein profiles for three biological replicates for high and low proliferation rate group samples.

Based on the young leaf 2-DE profiles, there were always presence of four low molecular weight subunits across the biological replicate gels. The protein extraction and 2-D gel running were quite consistent.

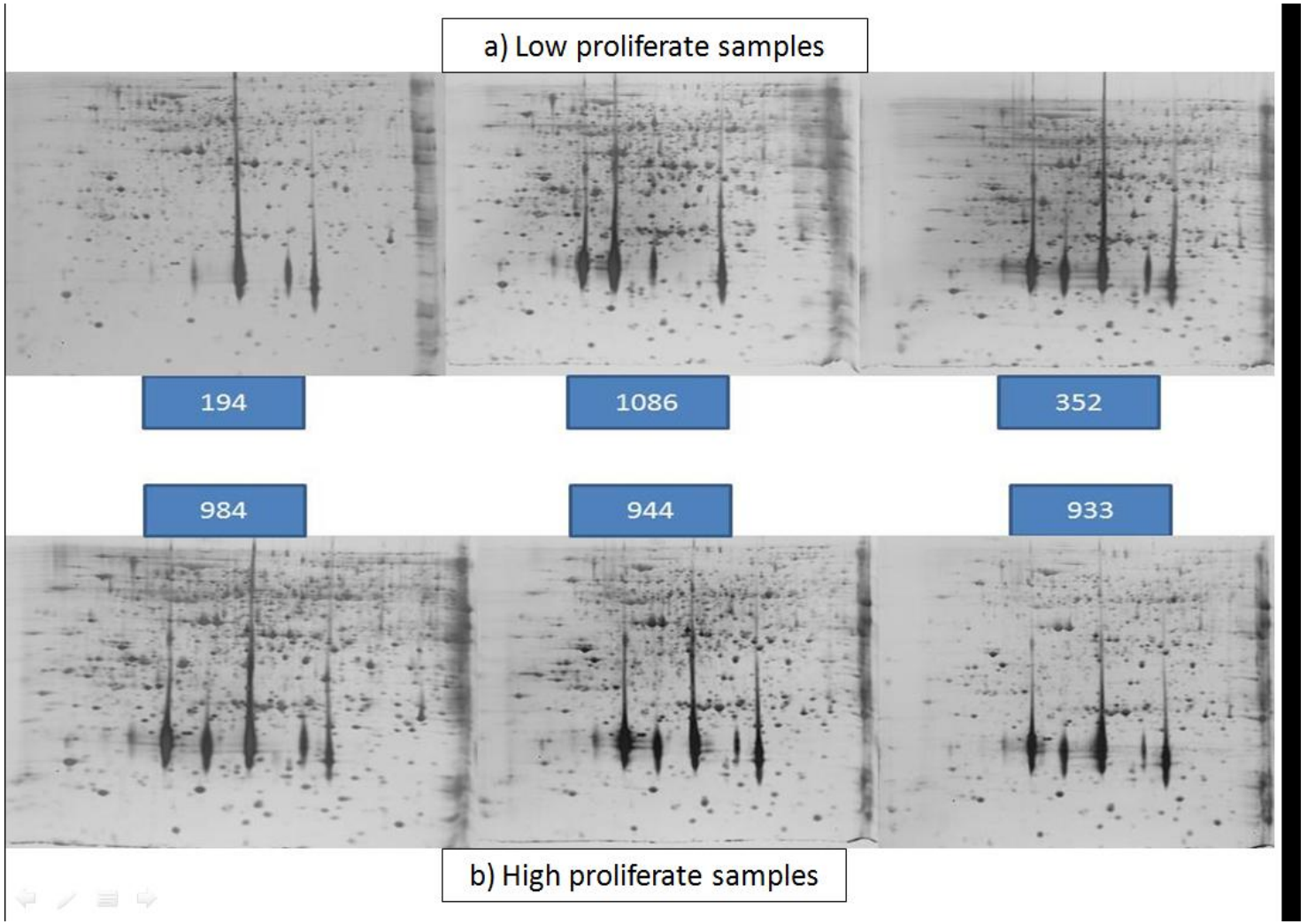


Figure 4.8: Protein gels for two categories of samples. a) Low proliferate samples (194, 1086 and 352) b) high proliferate samples (984, 944 and 933)

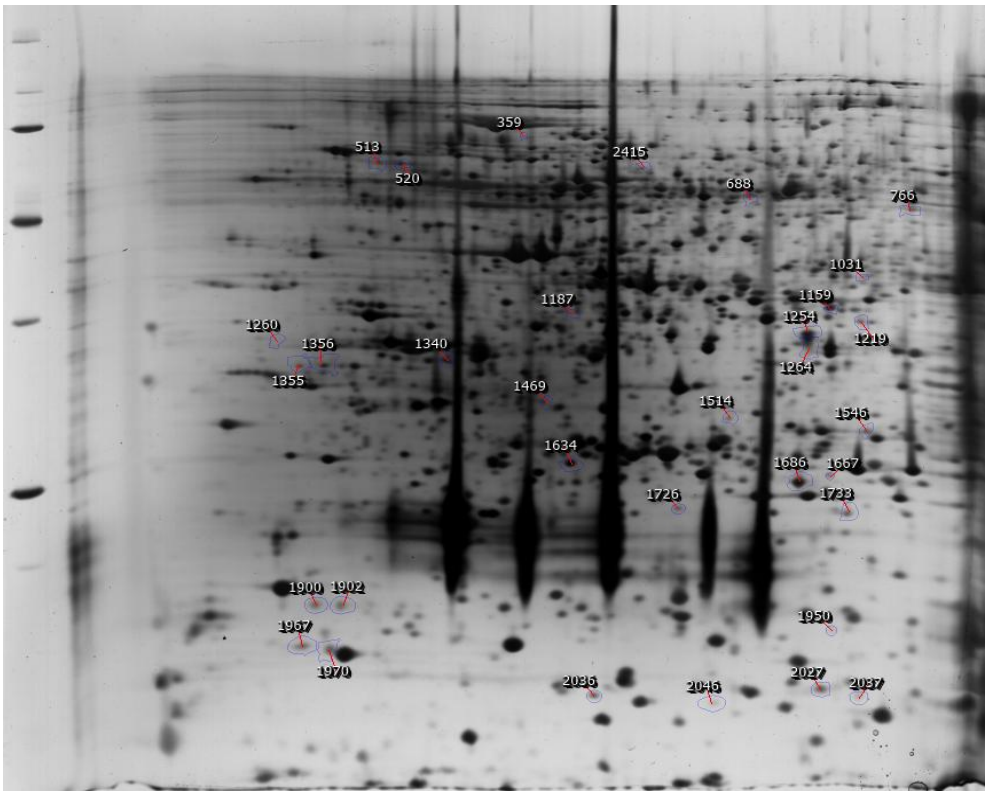


Figure 4.9: Protein spots that exhibit significant differences between high and low proliferation rate categories in young leaves are shown.

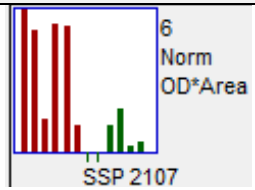
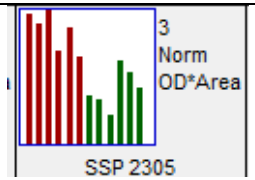
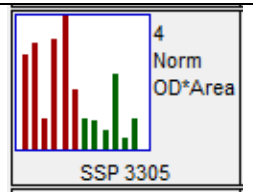
All the 2-DE gels from young leaf samples were analysed using the Progenesis Samespot software. Both Progenesis software and PDQuest software were powerful software in 2-DE gels analysis. They are very sensitive to the parameter with respect to the tendency of finding false positive spots through their spot detection, gel matching and spot quantification (Rosengren *et al.,* 2003). For the young leaf samples, Progenesis software were used due to their high number of protein spots in the 2-DE gels. Progenesis software worked well by its highly advance and automated image alignment for all the 2-DE gels. It also provided automatic analysed and ranked a list of significant spots based on ANOVA p-value.

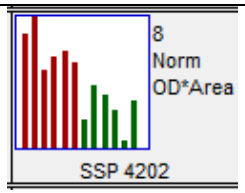
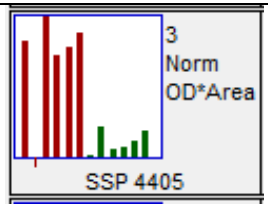
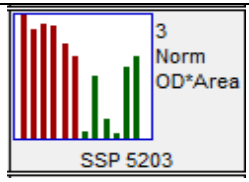
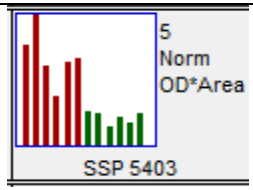
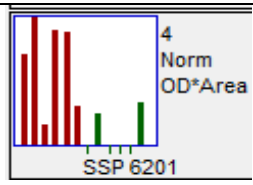
Table 4.3: Protein spots that exhibit significant difference between high and low proliferation rates in young leaf samples.

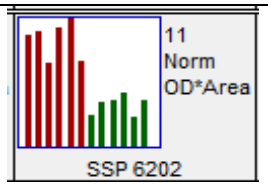
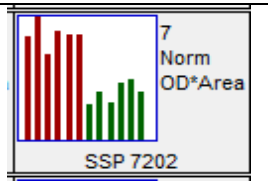
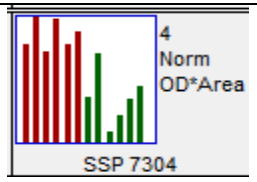
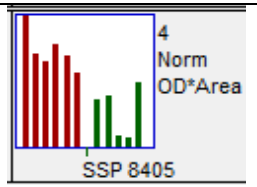
Protein expression	Total spots	Protein number
High expression in High Proliferation rate samples	14	688, 1187, 1219, 1260, 1355, 1356, 1514, 1634, 1733, 1900, 1902, 1967, 1970, 2415
High expression in Low proliferation rate samples	18	513, 520, 766, 1031, 1159, 1254, 1264, 1340, 1469, 1546, 1667, 1686, 1726, 1950, 2027, 2036, 2037, 2046
Total	**32**	

Table 4.3 provided a list of protein spots that exhibit significant difference between high and low proliferation rate samples in young leaves. A total of 32 protein spots were found to be expressed significantly between high and low proliferation rate group at p-value less than 0.01. Out of the 32 protein spots, there are larger amount of the protein spots were highly abundant in low proliferation rate group (56%) as compared to the 14 protein spots which highly abundant in high proliferation rate samples. All the protein spots were subjected to MALDI TOF/TOF instrument for further protein identification and the protein identified were shown in Table 4.4. The protein identification rate for the young leaf samples as 25% which means 8 proteins were successfully identified (Table 4.5). The spot diagrams for each of the identified spots were given in the Table 4.5.

Table 4.4: Sixteen protein spots that exhibit significant differences in the high and low proliferation rate old leaf samples.

Spot No	EST accession	Protein Identities (Biological function)	Homologous Protein SwissProt Accession (Sequence identiy %)	Score	MW (Exp/theory)	pI (Exp/theory) P	Covered sequence % (No of peptide)	Relative Protein abundance (PDQuest software) 8-fold changes with P<0.01 (red: high; green: low)	
High expression in high proliferic group									
2107	EoV2B_isot21105	RuBisCo large subunit-binding protein subunit alpha (Stress responses)	Q42694	65	15.3/61.9	4.55/5.57	15		PMF
2305	DuV2A_isot17667	Chlorophyll a-b binding protein of LHCII type III (photosynthesis)	P27489 (86)	51	27.6/28.7	4.7/4.99	2		PMF
3305	TCV2G_isot18884	Triosephosphate isomerase (metabolism)	P46225(77)	105	29.3/51.6	5.13/8.46	30		MSMS

4202	PiV2x_isot32349	oxygen evolving enhancer protein 2 (photosynthesis)	P29795(80)	112	25.9/51.2	5.2/9.39	35		MSMS
4405	TeV2D_isot11522	Sedoheptulose-1, 7-bisphosphatase (metabolism)	P46285(61	105	46.6/49.5	5.2/8.31	36(15)		PMF
5203	DuV2A_isot06934	20 kDa chaperonin, chloroplastic (Stress responses)	O65282(73)	108	27.0/39.4	5.7/8.54	30(8)		MSMS
5403	PiV2x_isot29459	Photosystem II stability/assembly factor HCF136 (photosynthesis)	O82660(62)	94	45.2/49.2	5.4/8.95	28(11)		MSMS
6201	PiV2x_isot52480	Oxygen-evolving enhancer protein 2 (photosynthesis	O49080(81)	135	18.3/60.1	5.85/8.6	3(1)		MSMS

6202	DuV2A_isot14951	Oxygen-evolving enhancer protein 2 (photosynthesis)	O49080(81)	87	26.6/37.8	6.2/8.03	34(9)		PMF
7202	EoV2B_isot08961	Pentatricopeptide repeat-containing At2g03880, mitochondrial Flags: Precursor (Others)	Q9SI53(82)	79	26.7/26.3	6.2/6.5	24(6)		PMF
7304	DuV2A_isot13776	L-ascorbate peroxidase, cytosolic (Stress responses)	P48534(72)	187	29.7/43.8	6.13/7.28	8(2)		MSMS
7402	No hit								
8405	DuV2A_isot12680	Fructose-bisphosphate aldolase, chloroplastic (metabolism)	Q40677(77)	80	45.6/52.1	6.8/8.86	27(12)		PMF

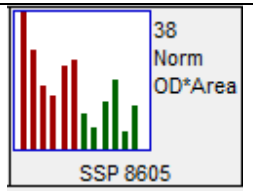
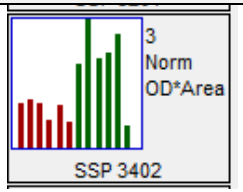
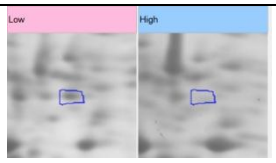
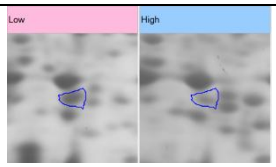
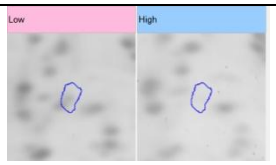
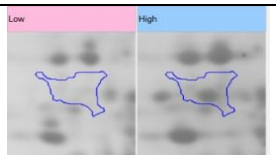
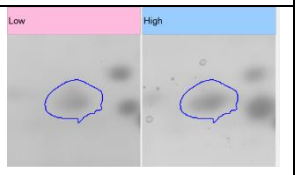
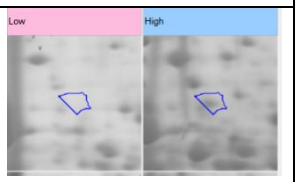
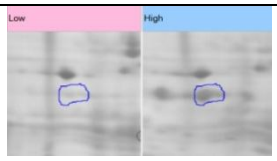
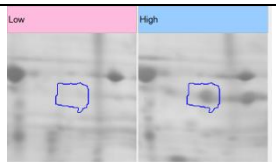
8605	PiV2x_isot39316	Perakine reductase (metabolism)	Q3L181(62)	78	/45.4	/8.03	34(11)		PMF
High expression in low proliferic group									
3402	PiV2x_isot29025	Sedoheptulose-1, 7-bisphosphatase (metabolism)	P46285(61)	106	/50.7	/8.49	34(15)		PMF
4404	No hit								

Table 4.5: Eight protein spots that expressed significantly difference in high and low proliferation rate groups of young leaves.

Expression Pattern	Protein spots	Protein name	Spot diagram
High expression in low proliferate samples	1. Spot 1031	Predicted: Uncharacterized protein LOC105053449	
	2. Spot 1159	Predicted: Uncharacterized protein LOC105053449	
	3. Spot 1469	Probable receptor-like protein kinase at3g55450	
High expression in high proliferate samples	4. Spot 1356	Elongation factor 1-beta-like	
	5. Spot 1967	TCV2G_iost35105	
	6. Spot 2415	Ketol-acid chloroplastic-like	
	7. Spot 520	Fibrous Sheath cabry-binding	
	8. Spot 513	Fibrous Sheath cabry-binding	

4.4 Discussion

4.4.1 Global proteome profiling of the old leaf sample in oil palm

A total of 55 differentially expressed proteins were submitted to MALDI TOF/TOF analysis for protein identification, 27 proteins were successfully identified by MSMS via MASCOT database searching. The database applied was SwissProt, one of the most prominent databases in proteomics. In addition to this, there is an in-house database (MPOB transcript database) used to search for these proteins. The MPOB transcripts database consists of ESTs and transcripts that successfully sequenced from the oil palm tissues during the past research studies on oil palm genome and has recently been published by Rajinder *et al.* (2013). It has helped to enhance protein identification in this study. When the mass spectra of digested peptides hit the sequences in the MPOB transcript database, it will directly link to the details of the gene sequences for that particular protein. Through the gene sequences, second round searches into the Swissprot and NCBI database aided in protein annotation. If there is no hit to any proteins in both available databases, it was categorised as unknown protein that needs to be subjected for further characterization. In the leaf samples, most of the identified proteins were metabolism and photosynthetic related, followed by stress response proteins and transportation proteins. It is well understood that leaf part is the main place for photosynthesis to be taken place in order to produce energy for plant to carry any metabolism process in cellular level of plants.

4.4.2 General functions of the protein population that has been expressed in oil palm leaf samples

Until now, limited study was done on the protein expression profiles of the oil palm leaves. A proteomics study was carried out on the high level of oxidative phosphorylation activity in storage oil production in oil palm samples (Loei *et al.,* 2013). A group of researchers studied the comparative proteomics on oil palm leaves that are infected by *Ganoderma* diseases. Comparative 2-DE from the *Ganoderma* disease enable 51 protein spots identified and majority were involved in photosynthesis, carbohydrate metabolism, followed by the immunity and defense system (Leona *et al.,* 2015). Recently, a thesis study on the proteomics profiling of chloroplast in oil palm fruit samples has enabled a better understanding on regulation of oil palm fatty acid biosysthesis. There are 162 proteins were identified from the chromoplast and only 10% of them were related to the fatty acid synthesis (Benjamin *et al.,* 2015). Here, in this study, the oil palm leaf samples were profiled to provide a better overview of the protein population distribution in oil palm leaf samples. Using the two-dimensional gel electrophoresis approaches run on the total protein extraction of oil palm leaf samples, protein spots that appeared in the 2-DE gel are those that highly abundant in the leaf samples. This is because using the 2-DE approaches, the abundance proteins tend to compete with the low abundance protein to be separated and finally stained out by the Coomassie staining. All the expressed proteins were sorted into few categories based on their major biological functions. Those proteins were involved in metabolism, photosynthesis, cellular biogenesis, stress response and finally transportation.

Metabolism

The results in this study show that most of the identified proteins in oil palm leaves were involved in carbohydrate metabolism. Photosynthetic carbon metabolism plays an important role in plant development and yield production. The carbon fixations mainly take part in the Calvin cycle (Miyagawa *et al.,* 2001). Several identified proteins identified to be involved in carbon metabolism were sedoheptulose-1,7-bisphosphate, phosphoribulokinase, ribulosebisphosphate carboxylate large chain, malate dehydrogenase and glutamine synthetase. For these metabolism group of proteins, phosphoribulokinase functions as a catalyst for the conversion of ɣ-phosphoryl group of ATP to the C-1 hydroxyl group of ribulose 5-phosphate, ribulose 1,5-bisphosphate (Hirasawa *et al.,* 1998). The enzyme also involves in the defence response against bacterium and cold condition. In plants, nitrogen is an essential building block of nucleic acids and proteins, which is necessary for reproduction and plant growth. Nitrogen stored within enzymes involved in carbon fixation. One of such enzymes is glutamine synthetase, which is a key enzyme for nitrogen metabolism. Glutamine synthetase involves in synthesizing of glutamine from ammonium (Zhang *et al.,* 2009). Besides, sedoheptulose-1,7-bisphosphatase that found in the leaf samples, is an enzyme that catalyses the removal of phosphate group to produce an intermediate in pentose phosphate pathway, i.e. sedoheptulose 7- phosphate to generate pentoses. On the other hand, ribulose bisphosphate carboxylate large chain is found to be mainly assisted in the photorespiration, magnesium ion binding and ribulose-bisphosphate carboxylate activities. Another metabolism protein namely malate dehydrogenase (MDH) is an enzyme involving in the oxidation of malate by reduce NAD^+ to NADH molecules.

Protein involved in the Photosynthesis

Photosynthesis is a crucial process for plants because it harnesses solar and convert it into chemical energy which is stored in the form of glucose. Through a series of cellular respiration the energy is converted to form as ATP, which can power biological processes such as active transport, cell division and carbon metabolism (McGinley *et al.,* 2010). The light-harvesting complex associated with photosystem II (LHCII) is the most abundant pigment-protein complex in chlorophyll plant. Its function is light capture and transfers the efficient energy to the reaction centers (Paulsen *et al.,* 1990). There are five proteins in this study that take part in the photosynthesis pathway. First, is the chlorophyll a-b binding proteins, which are involved in the light receptor that capture and deliver excitation energy to photosystems. Oxygen evolving enhancer proteins (OEEs) consist of three subunits, Oxygen evolving enhancer protein 1 (33 kDa), Oxygen evolving enhancer 2 (OEE 2) (23 kDa) and Oxygen evolving enhancer 3 (OEE 3) (23 kDa). Current study has found that there are two subunits i.e. oxygen evolving enhancer protein 1 and 2 in oil palm leaves. In general, the expression of oxygen evolving enhancer protein 1 is necessary for oxygen evolving activity and vital to maintain the stability of photosystem II (Mizobuchi and Yamamoto, 1989). The oxygen evolving protein 2 have been reported to play a role in salt adaptation process (Murota *et al.,* 1994). Another protein is photosystem II stability/assembly factor HCF 136, which is essential for photosystem II biogenesis and aid in the chlorophyll a binding.

Cellular biogenesis

Biogenesis is a process to synthesis new living organelles in cells. Cellular biogenesis refers to the biosynthesis of constituent macromolecules at cellular level. This includes macromolecular modifications and assembly of macromolecules for cellular component. Based on functional characterization of identified protein from this study, there were few proteins that involved in cellular biogenesis progression such as glycolysis, GTP catabolic process and hydrogen peroxidase catabolic process. The identified proteins were triosephosphate isomerase, L-ascorbate peroxidase 2, elongation factor TuB, phosphoglycerate kinase, and fructose-bisphosphate aldolase 1. Triosephosphate isomerase catalysed the interconversion of dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate during glycolytic pathway. It can also kinetically enzyme in isomerisation to facilitate the rate of reaction (Berg, 2007). L-ascorbate peroxidase 2 assists in the metal ion binding and smoothes the responses from heat and salt stress.

Stress response and transportation

RuBisCo large subunit-binding protein subunit alpha is classified as stress response protein in biological function. The protein facilitates the binding between small and large RuBisCo subunits. The protein has also been implicated in the assembly of the enzyme oligomer. When subjected to stress from the bacterium and cold condition, 2-Cys peroxiredoxin BAS1-like will be responsible to overcome the stress environment. At the same time, it has the antioxidant properties for the plant (Chow *et al.,* 2016). On the other hand, Ribulose bisphosphate carboxylase/oxygenase activase B is responsible to the light and jasmonic acid stimulus, as well as the leaf senescense (Shan *et al.,* 2011). In term of transportation, cytochrome B6-F complex iron sulphur subunit 2 was

found to be involved in electron transport in ubiquinol cytochrome-c reductase activity (Kuras and Wollman, 1994). Another is ATP synthetase subunit alpha, which is a factor in plasma membrane acts as proton-transporting ATP synthase complex (Jonckheere *et al.,* 2012).

4.4.3 Proteins expressed differentially in high and low proliferation rate samples

Old leaf samples were used in this study because it is considered as an ideal sample to use as early indicator for tissue culturist to identify high proliferation rate samples in particular palm. A total of 27 spots exhibited a statistically significant difference ($p < 0.01$) with at least a 10-fold change between the high and low proliferation rate samples. Each of the selected spots was checked manually on every single gel to avoid mismatch. Apart from that, a total of 13 proteins were found to be high in abundance, 1 protein was low in abundance in high proliferation rate and 2 protein spots are present only in the high proliferation rate samples (Table 4.2, Figure 4.7). All the protein spots of interest were excised and subjected for the mass spectrometry. The protein spots were undergone a series of destain, trypsin digestion, peptide extraction processes and lastly spotted on the MALDI plate for mass spectrometry analysis.

A total of 13 out of 16 protein spots were identified using MALDI TOF/TOF analysis. Most of the identified proteins in oil palm leaf proteome were located at chloroplast (84%), while only 8% in cytoplasm and 8% at other locations. There are more than two isoforms highly expressed in the high proliferation rate samples which are oxygen-evolving enhancer proteins (Heide *et al.,* 2004) and sedoheptulose-1,7- bisphosphate. Both of the proteins were located in the

chloroplast and essential for photosynthesis and Calvin cycle (Lefebvre *et al.,* 2005).

Young leaf samples (cabbage) were selected in this study because it is the nearest stage prior to the somatic embryogenesis to be occurred in the explants during tissue culture process. In the young leaf samples, a total of 32 spots exhibited a statistically significant difference ($p < 0.01$) between the high and low proliferation rate samples. Each of the selected spots was checked manually on every single gel to avoid mismatch.

Apart from that, there were 14 proteins of interest were found to be high in abundance, 18 protein was low in abundance in high proliferation rate samples (Table 4.3, Figure 4.9). The successfully rate of protein identification for the young leaf protein spot was 25% using MALDI-TOF/TOF instrument. It was far lower compared to the old leaf samples. There are a total of 8 out of 32 protein spots were identified using MALDI TOF/TOF analysis. It is known that silver staining method able to stain out 100 X more sensitive than Coomassie blue staining method. Yet, even more proteins spots have been selected at this young leaf samples but the identity of the proteins were unlikely. The main reason that caused the low identification rate was because the individual protein spots in the young leaf protein gels were low in concentrations and thus the 2-DE gels were stained using silver staining.

A total 11 individual proteins that were highly expressed in the high proliferation rate group compared to the low proliferation rate samples have been identified in this study. The identified proteins were chlorophyll a-b binding, trisephosphate isomerise, oxygen evolving enhancer protein 1 and 2, photosystem II stability/assembly factor HCF136, cytochrome B6-F complex iron sulphur subunit

2, L-ascorbate peroxidise, malate dehydrogenase, fructose-bisphosphate aldolase, and lastly ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo).

RuBisCo exhibited the highest abundance in the high proliferation rate samples. It is a major enzyme which plays important roles to incorporate carbons from the carbon dioxide into organic compound during carbon fixation process. In plant, energy is needed for all the cellular functions. Tricarboxylic acid cycle (TCA) which also known as Kreb's cycle is crucial for the macronutrient metabolism and energy conversion for carbohydrates, protein, and fats. In high proliferation rate leaf samples, malate dehydrogenase that catalyzes the conversion of oxaloacetate and malate by utilising NAD/NADH coenzyme system in tricarboxylic acid cycle (TCA) was found to be higher in expression. The results suggest that high proliferation rate leaf samples have higher energy metabolism occurred as compared to low proliferation rate leaves.

In this study, sedoheptulose-1,7-bisphosphatase (SBPase) was expressed in the leaf samples. A study showed that increase in sedoheptulose-1,7-bisphosphatase activities can regulate photosynthesis and biomass growth for up to 30% in plant development. Mature plants with increased SBPase activity were found to be able to fix 6-12% the carbon sources for the plant. Therefore, the presence of more SBPase activities will enhance the photosynthesis within the plant and aid in optimal plant growth (Stephane *et al.,* 2005).

Other than those above proteins, a few other proteins that directly link to the efficiency of photosynthesis were found to be expressed significantly high in the high proliferation rate category samples. For example, the photosystem II (PSII) stability with assembly factor HCF 136, it is a genetic mosaic consists of nuclear encoded subunit protein complex. PSII is normally found in thylakoid membrane

of chloroplasts and plays a role in water oxidation during oxygenic photosynthesis (Komenda *et al.,* 2008). Follow by another protein, cytochrome B6-F complex iron sulphur subunit 2 that functions as electron transport for photosystem. This protein is also expressed abundantly in high proliferation rate group. Both proteins are related to each other as the cytochrome Bf complex that converts the redox potential energy of plastoquinol into a transmembrane electrochemical charge gradient of protons for ATP synthesis (Kallas, 1994).

Chlorophyll a-b binding protein, which is one of the main components in light harvesting complex (LHC) that aids in delivering excitation energy to photosystem I or II for further reaction to take place. There are a total of three types of oxygen evolving enhancer proteins (OEEs) subunits. Two subunits (oxygen-evolving enhancer protein 1 and 2) out of three were detected in this study which expressed abundantly in high proliferation rate samples. Both Oxygen-evolving enhancer protein 1 (OEE1) and oxygen-evolving enhancer protein 2 (OEE2) are enormously needed for high level of photosynthetic oxygen evolution. Absence of OEE2 can pay compensation by photosystem II core and peripheral proteins, while the absence of OEE1 will cause cells unable to grow photosynthetically. Thus, OEE1 strongly affects the stoichiometry of PSII core proteins (Mayfield, 1991).

Two proteins that involved in glycolysis were shown high expression in high proliferation rate group. There were triosephosphate isomerase and fructose-bisphosphate aldolase (FBA). FBA is a key enzyme for glycolysis and gluconeogenesis and the pentose phosphate cycle in cytoplasm as well as play a role in calvin cycle in the plastid (Lu *et al.,* 2012). FBA catalyzed an aldol cleavage of fructose-1,6-bisphosphate to dihydroxyacetone-phosphate and glyceraldehydes 3-phosphate. Activities of FBA have marked consequences for photosynthesis, carbon positioning and growth (Konishi *et al.,* 2004).

Lastly, the identified protein which involved in plant metabolism is L-ascorbate peroxidase 2. This protein plays an important role in the metabolism of hydrogen peroxidase (H_2O_2) in higher plant. It acts as a scavenging defense system in plant toward all the excess active oxygen species that might cause harm in plant itself. It is an isoenzyme which distributed at four distinct cellular compartments like stromal APX, thylakoid membrane-bound APX, cytosolic APX and microbody membrane-bound APX (Shigeoka *et al.,* 2002). This protein expressed abundantly in high proliferation rate samples and suggests that the presence of ascorbate peroxidase is important as one of the scavenging defense system for plant to have a favorable condition to grow with an optimal proliferative property.

For the young leaves, there is less finding in this project due to the low identification rate for the protein identity attributed by the low concentration of protein spots. Few spots were found to be highly expressed in the high proliferation rate samples which were elongation factor 1-beta like, ketol-acid chloroplastic-like, and fibrous sheath protein. Probable receptor-like protein kinase, and uncharacterized proteins were found to be at low abundance in high proliferation rate samples.

Elongation factor 1-beta like protein is highly conserved and multifunctional. It involves in the protein biosynthesis transportation, chaperone activities in protecting proteins from aggregation due to stress environment and aid in renaturation of proteins during normal conditions. Elongation factor 1-beta gene is up regulated by abiotic stresses in plants and it responses well toward the stress condition. One of the papers shown that elongation factor plays an important role in improving the heat tolerance of plant (Fu *et al.,* 2012). Hence, it was speculated that high proliferation rate samples would have higher tolerance

level towards any abiotic stresses as the presence of high abundance of elongation factor 1-beta like proteins.

Another protein ketol-acid chloroplastic-like was found abundantly in high proliferation rate samples. This protein is mainly involved in the valine and isoleucine biosynthesis. Fibrous sheath protein was also found expressed abundantly in high proliferation rate samples. It mainly found in the cell wall of the plant. On the other hand, the protein that found to be expressed low in high proliferation rate samples was receptor-like kinases (RLPs) protein which is a signaling protein; it receives external signal and transduces them into the plant cell. It is very essential in the regulation of development, recognition event and response toward pathogen attack (Morillo *et al.,* 2006). This is very important protein toward the plant development.

The analysis results were fascinating as the proteins that expressed abundantly in old leaf samples and young leaf samples were difference among each other. In old leaf samples, most of the proteins that exhibited differential expression were involved in photosynthesis while in the young leaf samples, the proteins that revealed highly abundance were involved in plant defenses system. This results might due to the old leaf samples having the high photosynthetic properties. The 2-DE gel approach was able to detect the highly abundance proteins that expressed in the old leaf samples, namely the photosynthetic proteins. For the young leaf samples, also known as “cabbage” which is plant tissue that embedded in the core of the palm and they are not exposed to sunlight directly. This type of plant tissue is classified as not fully mature, the youngest part of the palm and it is a commonly used as starting plant material in tissue culture. The proteins that showed differential expression were mostly involved in plant defenses proteins which protect toward pathogen and abiotic stress in high proliferation rate samples.

4.5 CONCLUSIONS

The global proteome profiling of mature oil palm leaf samples was done. A total of 27 protein spots were identified out of the 50 spots of interest. The identified proteins mostly involve in photosynthesis and metabolism, and followed by the cellular biogenesis, stress response and lastly are the transportation function. Further leaf proteome comparison was carried out between high and low proliferation rate samples. A total of 16 protein spots were found to be statistically different between the two categories. Those identified proteins were involved in the photosynthesis mechanism too (45%), and 27% was taken by the metabolism function and followed by the stress response and lastly involved in transportation for the electrons. In old leaf samples, most of the photosynthetic proteins were found to be highly expressed in high proliferation rate samples, this suggests that the photosynthesis rate for the high proliferation palm is definitely higher than the low proliferation palm. Apart from the photosynthesis proteins, young leaf samples were found to be expressed more in stress responsive proteins especially the high proliferation rate samples. These indicate that the young leaves with high proliferative ability tend to response towards pathogen attack and abiotic stress. As preliminary assumptions, high proliferation rate samples have higher ability toward the defenses system during young stage and tend to have higher expression in photosynthetic proteins at mature stage. With the 2-DE gel approaches, the proteins that are able to be detected were mostly abundantly present in the leaf samples. To further understand on the biological function of the plant in high proliferation rate samples, it is interesting to investigate the relationship between protein and (messenger RNA) mRNA expression for the differential proteins that were identified in old and young leaf samples. Several protein candidates with higher differences in protein abundance were selected to the mRNA expression studies.

CHAPTER 5

ENRICHMENT OF THE LOW ABUNDANCE PROTEINS USING POLYETHYLENE GLYCOL (PEG) BASED FRACTIONATION METHOD

5.1 Introduction

Gene or protein identification in response to any experimental condition such as cold stress is the starting point to investigate the molecular changes that undergo in the plant. Proteomics is the systematic analysis of proteins expressed by a genome. Two-dimensional gel electrophoresis analysis is not only powerful tool to show a complete proteome picture of the tissue specific organelles, it also contributes to study the results of different physiological environment for a particular plant species (Lee *et al.,* 2007). In plants, distribution of high abundance proteins such as Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) interfere and prevent in-depth proteomic study in leaf samples. RuBisCo occupies roughly 60% of the total protein content in green leaf tissue and it is a major protein in the plant.

The presence of the highly abundance of RuBisCo brings huge challenges to proteomics analysis, especially in 2-DE gel electrophoresis. This is because high abundance protein interfere the detection of low abundance proteins. Hence, this phenomenon is also one of the issues hampering the study of oil palm leaves proteome. In previous studies, proteins have been a focus as potential biomarkers for diseases and these are usually present in a minimum concentration at ng/mL to pg/mL levels. Such low concentration causes challenges in detecting these amongst abundant proteins, which also exceed the

dynamic range of most analytical techniques. Researchers ordinarily will employ tools or reagents that can specifically remove the high abundance proteins and allow higher sensitivity to discover the low abundance proteins and in that way to "dig deeper into the proteome" (Huang *et al.,* 2005).

A comparison of leaf samples between high and low proliferation rate using 2-DE gel electrophoresis was reported in Chapter 4. There are 16 individual proteins that showed differential expression between those two categories. However, the proteins identified were all present abundantly in the leaf samples. The aim of the work in this chapter is to understand more on the proteins that would play a role in somatic embryogenesis, further fractionation steps to isolate the low abundance proteins are required as most of the functional proteins always exist in low concentration. Enhancement of the low abundance proteins is highly contributed to the discovery the potential group of proteins that are involved in somatic embryogenesis pathway. Since the high abundance proteins rarely participate in gene regulatory activities, they become interfering compounds in 2-DE analysis. Many low abundance proteins are regulatory factors and receptor molecules are present in only 100 molecules per cell and thus not detectable (Tirumalai *et al.,* 2003).

In this century, various pre-fractionations of the high abundance proteins have been practiced in early research, such as SDS-PAGE based size fractionation (Sun *et al.,* 2003), complementary multidimensional technologies (Issaq *et al.,* 2002), three-phase partitioning (TPP) and subcellular fractionation. Some researchers employed immunoaffinity-based protein subtraction chromatography or physiochemical approaches such as affinity chromatography based on the molecular weight or isoelectric point (p*I*). Affinity chromatography is recently the most common pre-fractionation method of choice in proteomics. It can be

classified into two types which are biological (antibodies, protein, peptide, lectin and nucleotide) and non-biological (synthetic dyes, immobilised metal ion complex) (Huang *et al.,* 2005). Furthermore, Solassol *et al.* (2005) has facilitated the identification of low abundance proteins using pre-fractionation of serum proteins by using strong anion exchange chromatography which can enrich the low abundance protein and increase their detection in combination with using high advance mass spectrometry known as Surface-enhanced laser desorption/ionization (SELDI) mass spectrometry.

In the current study, an alternative approach has been used to eliminate the high abundance protein molecules in plant samples. The method used was polyethylene glycol fractionation (PEG) during protein extraction. This approach is one of the most cost effective methods to remove many of the high abundance proteins, especially RuBisCo. This PEG fractionation method has been reported previously in *Arabidopsis thaliana* Columbia (Xi *et al.,* 2006), *Oryza sativa* (Kim *et al.,* 2001) and *Cynara cardunculus* (Acquadro *et al.,* 2009).

Plant proteomics has huge challenges due to the presence of large abundance protein, RuBisCo protein which occupied roughly 60% of the total protein content in green leaf tissue (Huang *et al.,* 2005)**.** Gel-based proteomics is still remaining as one of the favourite approaches to detect the differential protein even it has the limitation to shown the low abundance protein species in samples. This is one of the main reasons to employ extra fractionation method, F3 in this proteomics comparative study. These studies found that PEG fractionation enables detection of a large number of low-abundance proteins, and allows the “hidden proteome” to be investigated.

5.2 Materials and methods

5.2.1 Plant Materials

Old leaves of oil palm (*Elaeis guinensis* Jacq.) were collected from Advanced Agriecological Research Sdn Bhd (AAR) and Malaysian Palm Oil Board (MPOB), Malaysia. The leaf samples were collected at frond no. 17 (F17) which is the outer frond of the palm. The leaf samples were kept at -80°C until further use.

5.2.2 Protein Extraction Protocols

Prior to the protein extraction, the leaf samples were ground to a fine powder in liquid nitrogen using a mortar and pestle. The Mg/NP-40 based protein extraction method was used as described by Acquadro *et al.* (2009) with slight modifications. One gram of the powder was homogenised in 10 mL of cold Mg/NP-40 buffer (2% v/v NP-40, 20 mM $MgCl_2$, 0.5 M Tris-HCl, pH 8.3, 2% v/v β-mercaptoethanol, 1% PVP, 1% v/v protease inhibitor cocktail (Sigma-Aldrich, Saint Louis). The mixture was vortexed and mixed well before incubation on ice for 5 minutes. It was then centrifuged at 12,000 x *g* for 15 minutes at 4°C. The supernatant was then subjected to TCA/acetone extraction protocol as described in the previous section 3.2.2.1. The protein extract was re-dissolved in IPG buffer (7 M urea, 2 M thiourea, 4% CHAPS) and sonicated for 15 minutes. Further centrifugation was carried out at 12,000 x *g* to remove all the cell debris. The protein concentration was then measured using the Bradford method.

5.2.3 PEG fractionation

PEG fractionation was performed as reported in Kim *et al.* (2001). A schematic illustration of the work flow for the PEG fractionation of *Elaeis guineensis* leaves is depicted in Figure 5.1. The initial protein extraction was based on the Mg/NP-40 method. The slurry was then centrifuged at 12,000 x *g* for 15 minutes at 4°C. The supernatant was then treated with 50% (w/v) PEG-4000 stock solution to give a final concentration of 10% PEG. The PEG suspended solution was then placed on ice for 30 minutes to ensure protein precipitation. After incubation, the 10% PEG mixture was centrifuged at 1,500 x *g* for 10 minutes at 4°C. The resultant pellet was taken as fraction 1 (F1). The supernatant from the previous step was made up to 20% PEG 4000 and incubated for another 30 minutes. The centrifugation of the 20% PEG suspended solution was carried out at 12,000 x *g* for 15 minutes at 4°C. The supernatant and pellet from the 20% PEG suspended solution were collected as fraction 2 (F2) and fraction 3 (F3), respectively. All the fractions (Total, F1, F2 and F3) were precipitated using the TCA/acetone method and dissolved in IPG buffer for 2-DE analysis. For all the PEG fractionation experiments, three replicates were performed.

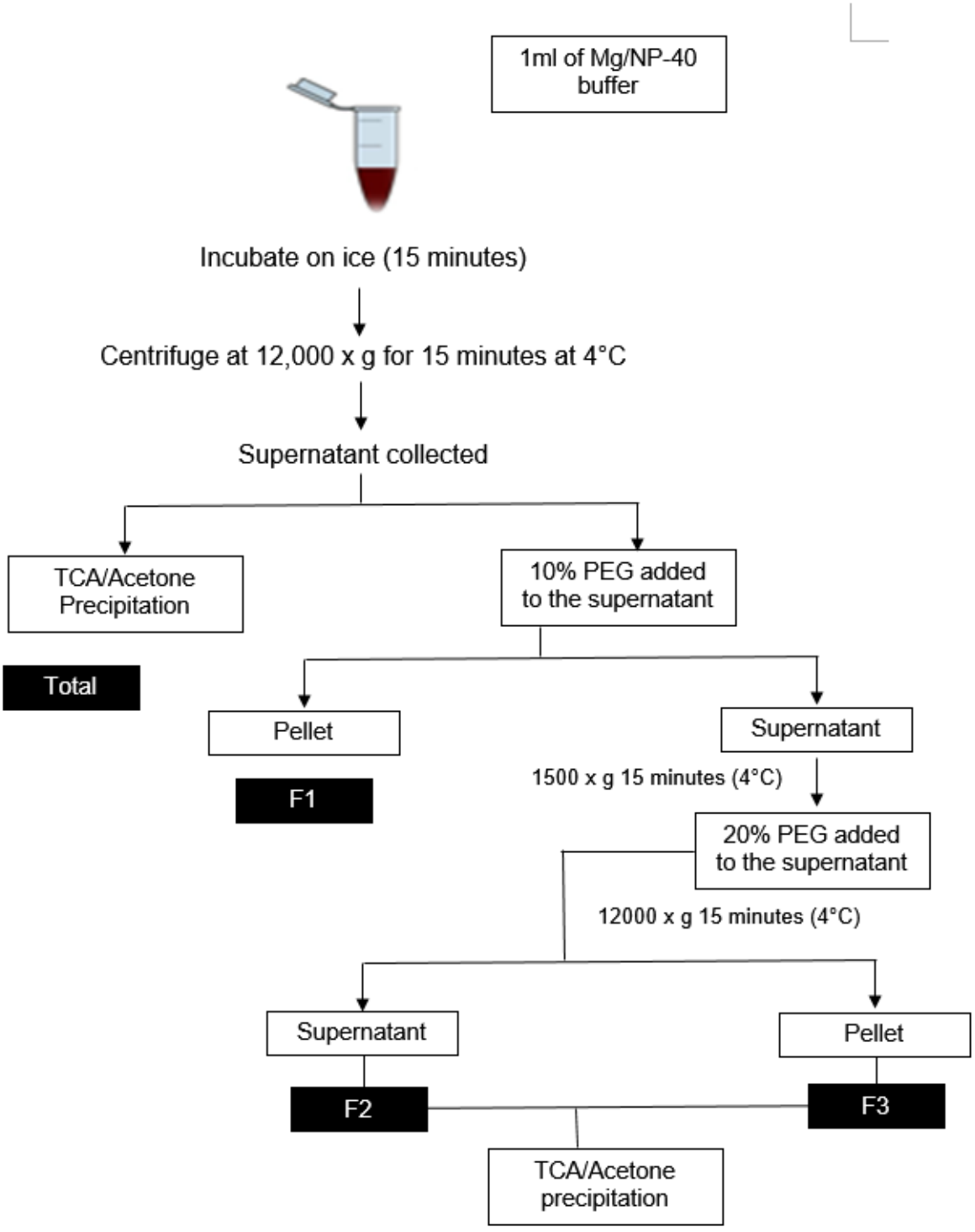


*Figure 5.1: The schematic work-flow for the differential PEG fractionation (*10% and 20% PEG is the final concentration)*

5.2.4 Two-dimensional gel electrophoresis, image and data analysis

Fifty micrograms of the fractionated protein samples was used for 2-DE analysis. The samples were rehydrated overnight on pH 3-10 and pH 4-7 Immobilised pH gradient (IPG) strips. Strips were then run for the first dimensional separation using Isoelectric focusing (IEF) instrument and reaching the 10,000Vhr focusing point. After separation according to the p*I*, the strips were transferred to SDS-PAGE for the second dimensional separation. The running condition for the SDS-PAGE was 120 V for one hour and five minutes. The gels were then stained using silver staining due to its high sensitivity. The 2-DE gel images were captured using a GS-800 densitometer (Biorad, USA) and PDQuest software (Biorad, USA) was used to analyse the gel images. All the biological replicates were carried out twice in this experiment.

5.2.5 Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

After the PDQuest software analysis, the silver stained spots that exhibited differential expression were selected for tandem mass spectrometry analysis to identify the proteins. The LC-MS/MS mass spectrometry was used in for the proteins identification. Before protein identification, in-gel digestion was conducted according to the method described in Chapter 4 (Section 4.2.4.3). Digested peptides were subjected to liquid chromatography coupling with Orbitrap mass spectrometry (University of Florida, USA).

5.3 Results

5.3.1 Protein concentration for the PEG fractionation

PEG fractionation method produced four main fractions which were labelled as Total protein, Fraction 1, Fraction 2 and Fraction 3 (Total, F1, F2 and F3). For each sample, all four fractions were isolated and the protein content was measured as summarised in Table 5.1.

Table 5.1: Protein yield upon PEG fractionation

	Sample	**Protein Yields (µg/µl)**				
		F1	F2	F3	**Sum**	Total
High	AN25	$1.130^{b} \pm 0.13$	$1.944^{a} \pm 0.50$	$0.602^{b} \pm 0.14$	3.675 ± 0.77	2.810 ± 0.48
	AN28	$1.656^{b} \pm 0.14$	$2.653^{a} \pm 0.17$	$0.663^{b} \pm 0.19$	4.972 ± 0.27	3.832 ± 0.10
Low	AN27	$1.139^{a} \pm 0.07$	$1.981^{b} \pm 0.21$	$0.727^{c} \pm 0.23$	3.847 ± 0.41	3.615 ± 0.01
	AN29	$0.971^{b} \pm 0.08$	$1.639^{a} \pm 0.21$	$0.503^{b} \pm 0.06$	3.111 ± 0.11	3.109 ± 0.33

Data represented as the mean values ± standard deviation (SD) of three independent experiments. Samples F1 – F3 indicate the fractions produced from PEG fractionation; Total is the protein samples without going through fractionation.

As shown in the Table 5.1, the sum of the protein yield for all the samples after fractionation was increased as compared to the total protein extract without fractionation. Surprisingly, the F2 has the higher protein yield compared to F1 for all the samples extraction. In F2 fraction, higher percentage of PEG was used which is 20% as compared to F1 fraction (10% PEG). The F3 proteins, the low abundance protein always exhibited the lowest protein yields.

These results indicated that the fractionation procedure could improve the protein yield for leaf samples. The standard deviations for all the fractionation samples and non-fractionation samples were low except for one sample AN 25 which had more than 0.5.

5.3.2 Two-dimensional gel for different fractions

All the protein samples from the Total, Fraction 1 to 3 were collected and run on 2-DE analysis to observe the protein distribution in protein spot profiles. Non-fractionated protein (Total) shows a 2-DE protein profile with all the abundant proteins present. For fractions F1 and F2, the high abundance protein at 50 kDa, which is the RuBisCo large subunit (red dot arrow in Figure 5.2), remained present. The small subunit RuBisCo (indicated by blue arrow in Figure 5.2) not fully eliminate and it is still visible in the fraction 3 samples. Remarkably, the large RuBisCo protein was successfully eliminated from F3. In addition, many low abundance proteins were present in F3 that were unable to be visualised in 2-DE gel of total extracts, as shown in Figure 5.2.

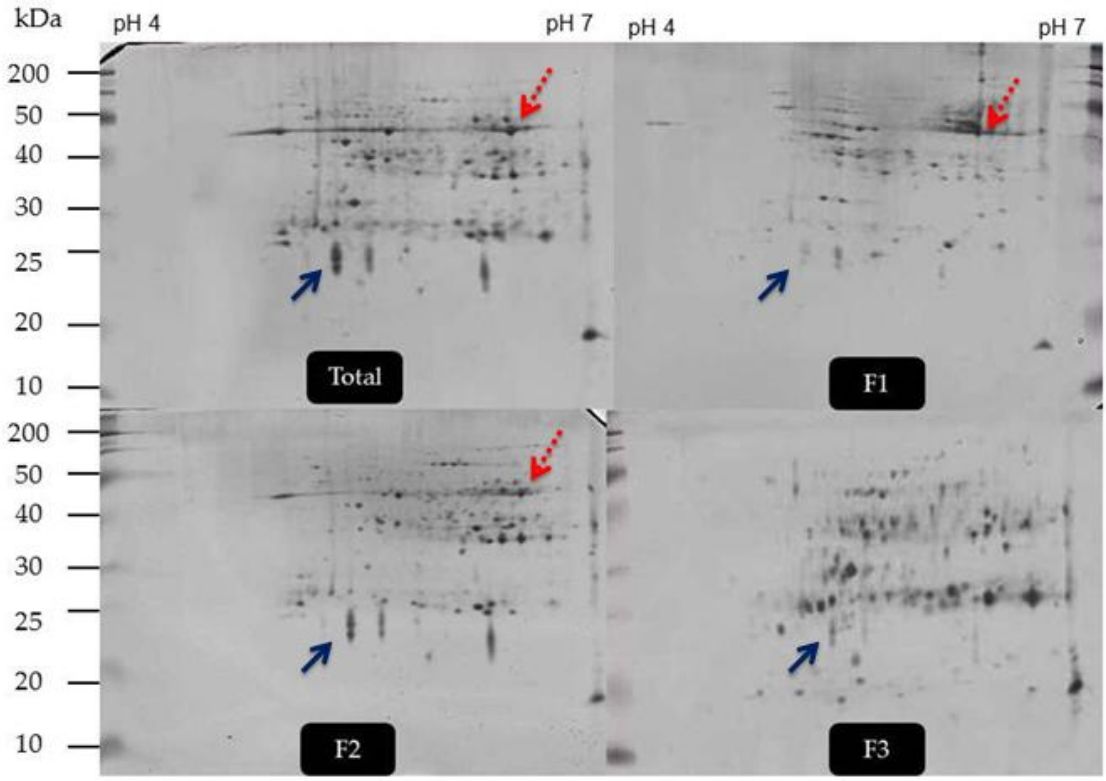


Figure 5.2: 2-DE protein profiles of the Total protein, F1, F2 and F3 proteins (Red dot line arrow indicates the large subunit RuBisCo, Blue arrow: small subunit RuBisCo).

5.3.3 F3 comparison for high and low proliferation rate samples

The purpose of PEG fractionation is to eliminate the high abundance proteins and increase the detection for the low abundance proteins in the samples. Hence, 2-DE gels loaded with the F3 proteins became the main focus in this study. Figure 5.3 showed the 2-DE gels for F3 proteins extracted from the high and low proliferation rate groups.

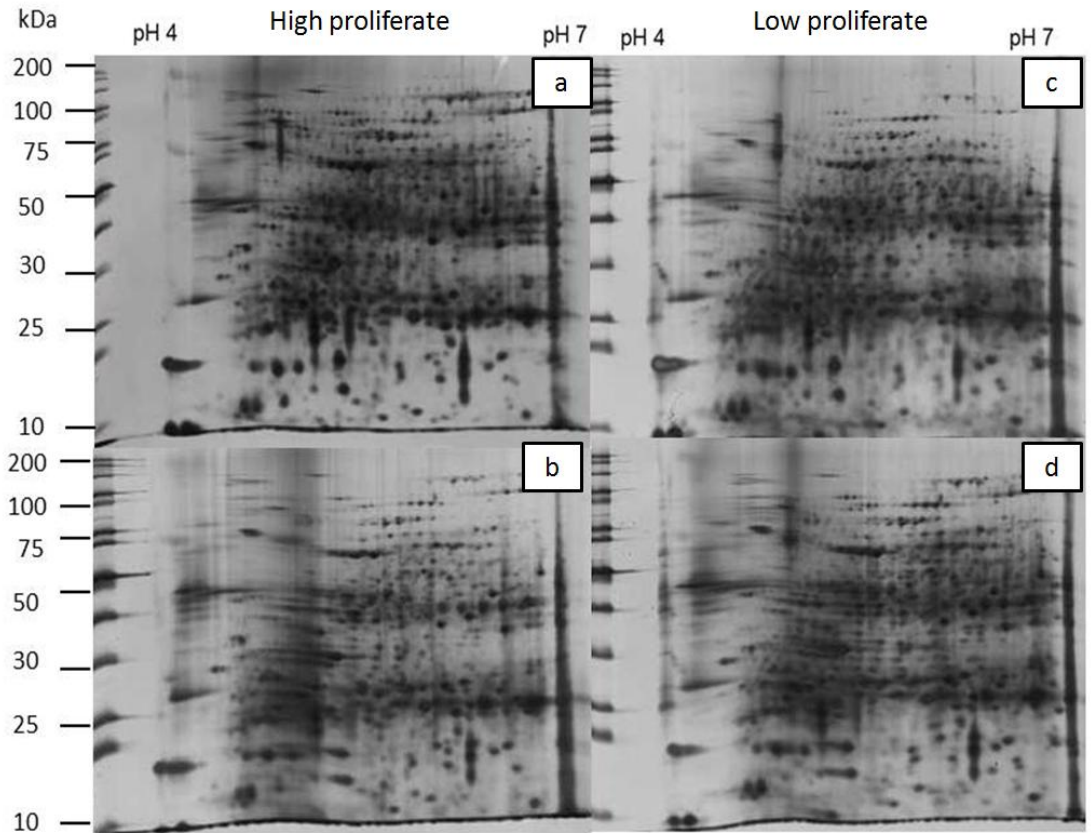


Figure 5.3: The 2-DE protein profiles of the F3 proteins from high proliferation rate samples (a) AN25 (b) AN28, and low proliferation rate samples (c) AN27 (d) AN29

After the 2-DE gel separation, the gels were stained using the silver staining method and digitalized on the densitometer. The gel images captured were then subjected to PDQuest software analysis for comparison of the different protein expression levels between the high and low proliferation rate category samples. All the protein spots were matched across the eight gels. Any smearing protein spots were excluded to ensure a more reliable comparison using the PDQuest Software.

Table 5.2: The 24 protein spots that have significant difference in expression from the F3 were classified into four categories.

Categories	Number of protein spots
More abundant in high proliferation rate sample gel	5 spots
More abundant in low proliferation rate sample gel	6 spots
Present only in high proliferation rate sample gel	10 spots
Present only in low proliferation rate sample gel	3 spots
Total	24 spots

A total of 24 protein spots showed significant difference in the high and low proliferation rate gels for F3 samples. The 24 protein spots were determined using the t-student test with a 95% confidence level and showed a more than two-fold change in level.

5.3.4 Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) identification

Only 13 protein spots out of the 24 proteins of interest were successfully identified by LC-MS/MS. All spectra were summarised using the Scaffold Viewer software (Figure 5.4). All the results from the mass spectrometry were clearly stated the protein identification probability, percentage of protein coverage, and total unique peptide count.

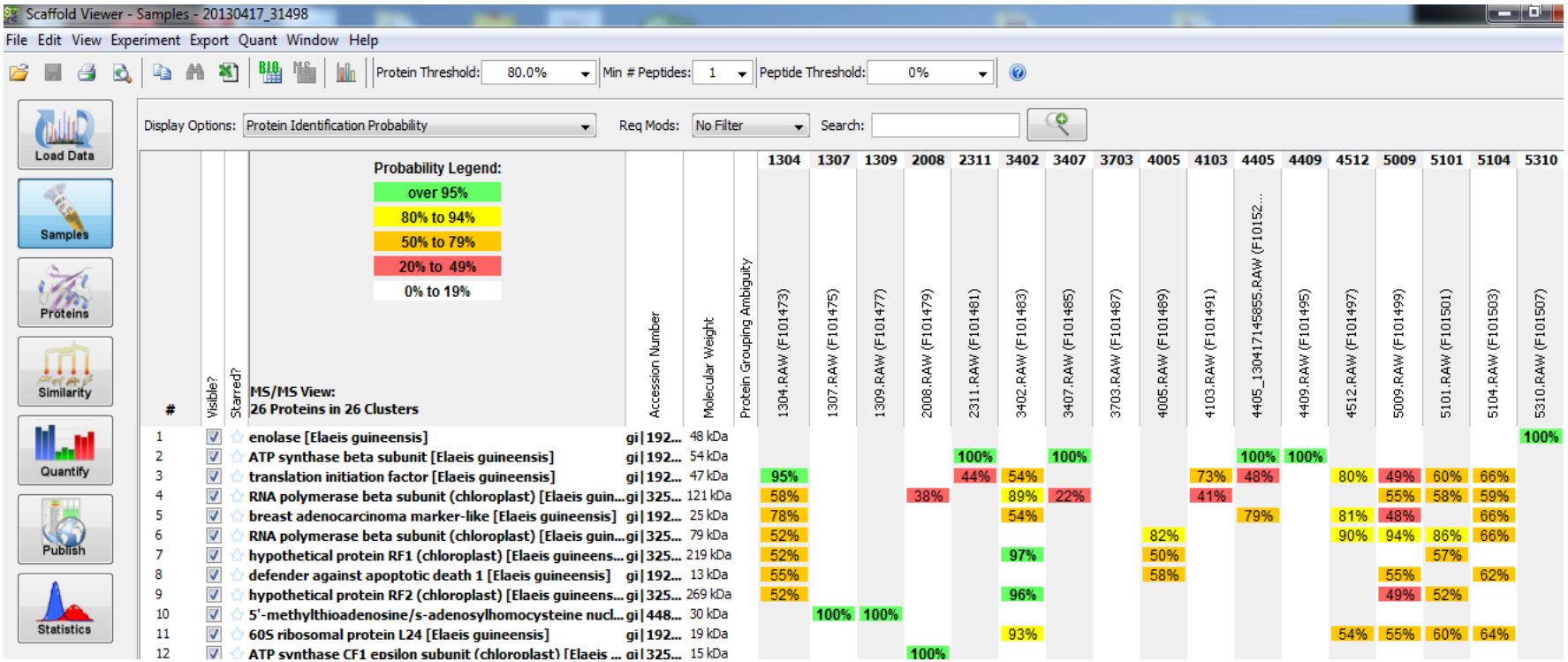


Figure 5.4: Screen capture of the examples of the LC-MS/MS analysis for the selected protein spots in Scaffold Viewer software.

Table 5.3 Proteins that exhibit significant difference in level between the high and low proliferation rate group samples.

Categories	Spot No	Protein Identities
Abundance in High proliferation rate gel	2008	EoV2B_isot01039
	6003	EoV2B_isot01039
	3406	No Hit
	8102	PiV2x_isot00792
	5310	Putative dehydrolipoamide dehydrogenase
Abundance in Low proliferation rate gel	4409	Alcohol dehydrogenase
	4512	Phosphoribulose kinase
	5009	TeV2D_isot25659
	7610	Enolase
	8013	Copper/Zn Superoxide dismutase
	7502	Phosphoglycerate kinase
Present in High proliferation rate gel only	3703	No Hit
	1304	EoV2B_isot01043
	1307	Chaperonin alpha subunit
	1309	Chaperonin alpha subunit
	4005	EoV2B_isot010139
	6109	PiV2x_00805
	4103	No hit
	5101	Phosphoglycerate kinase
	5104	Phosphoglycerate kinase
	5205	No Hit
Present in Low proliferation rate gel only	3408	Sedoheptulose-1,7-bisphosphate
	7007	Superoxide dismutase
	2311	a-a binding protein

For these F3 samples, all the protein spots that sent for LC-MSMS analysis were gone through Swiss-Prot database for protein identification. As referred to the Table 5.3, there was a total of 5 proteins that showed high expression levels in the high proliferation rate samples, 6 proteins had high expression in the low proliferation rate samples, 10 proteins were present only in high proliferation rate samples and only 3 protein spots were present in the low proliferation rate samples that showed significant difference between the high and low proliferation rate samples. The identification rate for the proteins was around 54% for the proteins in F3 through Swiss-Prot database. This relatively low percentage of successful identification was mainly due to the low amount of proteins detected from 2-DE gels, resulting in a low peak list quality produced in the mass spectrometry. This was happening especially for the proteins present in F3 which were those proteins that were expressed in low abundance. Limited availability of potentially matching sequences in the protein database was another factor that contributed to the low identification of proteins. In order to increase the proteins identification, the mascot generic files (.mgf) results for each protein spots were searched against the in-house MPOB database. MPOB database is an in-house database which comprises of the entire transcripts that successfully sequenced in the oil palm tissues. Ninety two percents of the protein spots were found to match with the MPOB transcript database. With these transcripts, the nucleic acid sequence for the proteins can be found but some of the protein identities remained unknown due to the uncharacterised proteins where the function of the coding sequence has not been established.

5.4 Discussion

5.4.1 Protein concentration for the PEG fractionation

The protein yield from the four fractions have been compared and listed in the Table 5.1. The results show that the fractions, F1 to F3 after fractionation had higher protein yield as compared to that of the total protein samples (without fractionation). The same trend of results was obtained in a previous study by Xi *et al.* (2006). This result was satisfactory as the additional fractionation step definitely enhanced the solubility of the sample proteins by reducing the possible abundance protein competitors for allowing the low abundance protein population to be appeared.

By minimising the space used up by large and highly abundant proteins, it increased the solubility of low abundance proteins in the protein extract. The second phenomenon examined was the protein concentration of the F2, it was higher in quantity compared to F1. This might be due to the higher percentage of PEG 20% was used in F2, while the F1 only contained half of the concentration of PEG (10%) in the supernatant that can help to precipitate large abundance proteins.

The percentage of PEG added to the samples strongly affected the solubility of the proteins as well. Higher concentrations of PEG will enhance the protein precipitation for the samples. On the other hand, F3 is the remaining protein pellet after removal of the high abundance proteins. In this fraction most of the low abundance proteins were present at a minimal quantity. The low concentration of the F3 (< 1 µg/µl) was a challenge for the 2-DE gel analysis.

5.4.2 Two-dimensional gel electrophoresis for PEG fractionation

In Fraction 1 (F1) (10% PEG 4000), some of the large abundance proteins were precipitated out from the supernatant. Eventhough the F1 had a relatively high protein yield, the number of protein spots appeared in the 2-DE gels were much less as compared to the F2. This leads to an assumption that during the first precipitation step using PEG, the high molecular weight proteins were effectively eliminated and thus in the 2-DE gel profiles of F1, only the large subunit of the RuBisCo is observable and other low quantity proteins were masked by the RuBisCo molecules. Therefore, the amount of the highly abundant proteins was reduced greatly after F1.

During the precipitation of supernatant in F2 (20% PEG 4000), with reduction of the large subunit RuBisCo in F1, other relatively high abundance proteins were able to precipitate out in a greater quantity (Figure 5.2) and indirectly increase the number of protein spots appear on the 2-DE protein profiles for F2. The protein spots distribution for the F2 exhibit highly similar as compared 2-DE gels from the total protein extract. Some of the abundant proteins spots were still visible but the intensity was greatly diminished after the 10% PEG fractionation. Moreover, the horizontal streaking caused by the highly abundant proteins was significantly reduced in the F2.

Unexpectedly, the F3 shows a different protein distribution compared with the previous fraction F1 and F2 on the 2-DE gels. From the visualisation of 2-DE gels, PEG fractionation effectively removed the highly abundant proteins and enables those undetectable low abundance proteins to be precipitated out in F3. In addition, the protein spots observed in the F3 were mostly the “hidden” spots and these spots contributed to the total detectable protein spot number for the 2-DE analysis.

5.4.3 F3 comparison for high and low proliferation rate samples

In this study, only F3 was selected to further investigate for the downstream proteomic analysis. Due to the low amount of the individual protein in the F3, silver staining was selected as the staining method. This is because the coomassie stain gel was not sensitive enough to detect the presence of low abundance proteins. As shown in figure 5.3, many low abundance proteins were sensitively detected by the silver stain and resulting the number of protein spots in the F3 ranged from 335 to 455 spots.

According to the PDQuest software analysis, the match rate between the high proliferation rate sample gel members were as high as 98% while the low proliferation rate gel members has 93% match rate. This indicates the 2-DE gels used in this experiment were highly repeatable. In addition, the match rate between the high and low proliferation rate group samples was 72% only. This means that there is a difference detectable in the F3 between the high and low proliferation rate groups.

In further matching between the high and low proliferation rate group samples, 24 protein spots showed significant differences with a fold change more than 2.0. The numbers of significant protein spots in F3 were higher as compared to the previous results without fractionation using TCA/acetone extraction methods (Chapter 3, section 3.2.2.1). Hence, PEG fractionation enabled the low abundance proteins to be detectable on the 2-DE gels and the differential expression level of these previously hidden low abundance proteins was able to be examined.

5.4.4 LC-MS/MS analysis

For the F3, the LC-MS/MS instrumentation were selected for the method to identify the low abundance proteins as MALDI TOF/TOF instruments might not be sensitive enough to detect the protein fragment and provide a lower rate of successful protein identifications. LC-MS/MS is combination of the physical separation of peptides using liquid chromatography followed by mass analysis of those peptides by mass spectrometry. It has a very high sensitivity and selectivity. Nevertheless, even the large numbers of mass spectra generated from the LC-MS/MS encountered challenges in providing protein identifications from the existing public databases. This was because of the huge quantities of peptides generated could match more than one protein in the database and hence gave the false positive result. Thus, further selection was needed before the protein identities could be confirmed. In general, for each sample, the highest probability of percentage up to 90% was taken into consideration as well as the protein coverage.

Out of the 24 proteins, only 13 protein spots were successfully identified. Five protein spots were highly abundant in high proliferation rate sample gels (SSP 2008, SSP 6003, SSP 3402, SSP 8102 and SSP 5310), one of the highly abundance proteins in high proliferation rate samples was annotated as putative dehydrolipoamide dehydrogenase. Meanwhile, 6 protein spots were found to be present in low abundance in high proliferation rate samples. Among them, some were identified as alcohol dehydrogenase, phosphoribulose kinase, enolase, phosphoglycerate kinase, and Copper/Zn superoxide dismutase. This result was unexpectedly different from the results obtained from total protein samples which were described in Chapter 4. From the F3, 62% of the proteins identified were

involved in metabolism, followed by 31% stress response proteins and lastly 7% of others. The main difference among the total protein profiles and the F3 profiles was that no photosynthetic proteins were found in the F3 fractionation. This phenomenon indicated that most of the photosynthetic proteins were present at high abundance and that these were excluded from the F3 fraction.

Alcohol dehydrogenase (ADH) genes are expressed with different abundance in various organs and this pattern is not consistent from one plant to another. Generally, the main function of ADH is involved in the NADH metabolism by reduce the acetaldehyde to ethanol under anaerobic condition in plants. The expression of ADH is known to be induced by environment stresses such as wound and cold conditions. In the model plant *Arabidopsis thaliana* the over-expression of ADH gene was studied and the results showed that higher levels of ADH improved the tolerance level of roots toward the low oxygen level (Shiao *et al.,* 2002). Since the ADH was found in lower abundance in the leaves of high proliferation rate oil palm, it could be speculated that high proliferation rate has fewer stresses compared to the low proliferation rate samples, thus the expression of ADH was significant lower in the high proliferate group.

Phosphoribulose kinase (PRK) was found to be expressed high in low proliferation rate samples. In low proliferation rate samples, more PKR was detected. This might due to unfavourable conditions of the low proliferation palm that makes the palm unable to perform optimally and thus more PKR is needed to produce energy for survival. Three main stages of photosynthesis take place, the first is capturing energy from sunlight and this is followed by production of ATP and NADPH. The third stage is the use of Adenosine Tri-Phosphate (ATP) and Nicotinamide adenine dinucleotide phosphate-oxidase (NADPH) compounds to synthesise organic molecules from carbon dioxide in the absence of light. The first two main stages are defined as the light reaction while thethird stage is designated as the Calvin

cycle. PKR is one of the essential enzymes other than RuBisCo that plays a very important role in the Calvin cycle of CO_2 assimilation. PKR catalyzes an irreversible reaction which is the regeneration of Ribulose-biphosphate (RuBP) from ribulose 5-phosphate and serves a role in regulation of the flow of sugar through the carbon dioxide fixation cycle (Hariharan *et al.,* 1998).

Another protein that was expressed abundantly in low proliferation rate samples was enolase. Generally, enolase is known to play a crucial role as an enzyme in the glycolytic pathway (Wang *et al.,* 2015). In addition, it is reported that enolase is up-regulated in the endothelial cell during hypoxia and functions as a cell associated stress protein that protects the plant (Aaronson *et al.,* 1995). Recently, it has been reported that enolase protein level in algae decreases by up to 50% due to its activities in balancing the hyperosmotic salt stress and it is also involved in thermal tolerance (Ruan *et al.,* 2009). Dennis *et al.* (1987) investigated the exposure of plants to anaerobic stress which can cause a shift from oxidative to fermentative mode of carbohydrate metabolism in plant tissues and thus increased the expression of this glycolytic cycle enzymes. The presence of enolase protein at relatively low abundance in the high proliferation rate samples suggests that more glycolytic process is conducted in the high proliferation rate samples to produce energy for metabolism.

In addition, two other proteins were found to be present only in the high proliferation rate samples, chaperonin alpha subunit and phosphoglycerate kinase. Two spots, SSP 1307 and SSP 1309, were identified as chaperonin. It is important to protect the plant especially under abiotic stresses that can cause proteins to function improperly. Chaperonins, also known as heat-shock proteins can assist in protein assembly, degradation and protein refolding under stress conditions to maintain the proteins in their functional conformations (Wang *et al.,* 2004). *In vivo*,

newly synthesized protein chains are unlikely to fold spontaneously in cells, this is where chaperonin takes places to facilitate correct protein folding by encapsulating substrate proteins into native states (Kmiecik and Kolinski, 2011). Hence, the presence of chaperonin is to protect the plant against stress for better survival. In this study, more chaperonins were present in the high proliferation rate samples, this suggests that the protection given against the stress in high proliferation tissue was higher compared to low proliferation tissue, and thus indirectly affect their excellent proliferate performances.

Interestingly, phophoglycerate kinase and superoxide dismutase were each identified in more than one protein spot. Phosphoglycerate kinase proteins (PGK) were found to be identified as same identity for SSP 7502, SSP 5101, and SSP 5104. For the SSP 7502, it showed that PGK was expressed abundantly in low proliferation rate samples but conversely the SSP 5101 and SSP 5104 spots were shown to be present only for the high proliferation rate samples. Overall, the PGK shows higher abundance in high proliferation rate samples with greater protein spots. PGK is one of the major enzymes involved in glycolysis to generate ATP energy and 3-phosphoglycerate from the transfer phosphate group from 1,3-bisphosphateglycerate to adenosine diphosphate (ADP). The presence of PGK proteins was to increase energy production in high proliferation rate samples.

Lastly, sedoheptulose-1,7-bisphosphate (SBPase), amino acid binding protein and superoxide dismutase were found to be present only in low proliferation rate samples. There is an overexpression study on the SBPase which was carried out in tobacco plants and positive results proved that the carbon fixation and electron transport rate were enhanced by increasing the content and activity of the SBPase enzyme (Rosenthal *et al.,* 2011). In this study, the expression of SBPase protein was found presence in higher abundance in the low proliferation rate samples, and this indicated that the low proliferation rate samples had the higher carbon

assimilation as compared to the high proliferation rate samples as described by Rosenthal group (2011).

Two protein spots were identified as superoxide dismutase (SOD) but with different expression pattern in low proliferation rate samples. One of the spots (SPP 7007) was present only in the low proliferation rate samples, while the other spot (SPP 8013) was abundantly expressed. In plant systems, there are natural defence systems against reactive O_2 species (ROS) that produced in both unstressed and stressed cells. The formation and depletion of ROS are balanced during unstressed conditions. Under oxidative stress conditions, SOD acts as the first line of defence against ROS (Alscher *et al.,* 2002). The high expression of SOD in low proliferation rate leaves indicates there might be high stress conditions in the low proliferation rate samples that required more SODs to overcome and balance up the stress conditions in plants.

5.5 Conclusion

PEG fractionation is a time and cost-effective method to eliminate the high abundance proteins. After PEG fractionation, many previously hidden low abundance proteins were detectable on 2-DE gels. A total of 24 protein spots were differentially expressed between high and low proliferation rate samples and were subjected to identification using the LC-MS/MS instrument. Out of the 24 protein proteins spots, only 13 proteins spots were successfully identified. The results were arranged and summarised using the Scaffold viewer software. The functions of the identified proteins were investigated to understand the biological mechanisms that might have occurred between the two different proliferation rate group samples. The results showed that most of the differentially expressed proteins were involved in metabolism (62%), followed by stress response protein (31%), and others (7%). F3 provided an in depth degree of protein profiles than usually could not be observed using general total protein extraction methods. Different groups of proteins were found differentially expressed in the F3 and total protein extracts but both extracts have show the highest percentage of proteins identified were in metabolism categories. There are no photosynthetic proteins were found in the F3 fractionation. This phenomenon indicated that most of the photosynthetic proteins were present at high abundance and that these were excluded from the F3 fraction after PEG treatment. In conclusion, F3 provides complementary results to the total protein extraction results with respects to the proteins that have significant expression level differences between the high and low proliferation rate samples.

CHAPTER 6

QUANTITATIVE EXPRESSION STUDY USING REAL TIME REVERSE TRANSCRIPTION POLYMERASE CHAIN REACTION (RT-qPCR)

6.1 Introduction

In 1993, Kary Mullis, a Nobel Prize winner was the first person to introduce the technology of polymerase chain reaction (PCR). Polymerase chain reaction allows the nucleic acid material, DNA to be amplified to thousands of copies under a series of reaction temperatures with the presence of the enzyme, polymerase and short specific nucleotides called primers. While, quantitative real time PCR (qPCR) is an extended modification to the PCR that allows quantification of amplified gene copies in real time by detection of the fluorescent signal during amplification process (Fraga *et al.,* 2014). In theory, reverse transcription (RT)-PCR differs from PCR only by the additional step of initial conversion of RNA into DNA template by reverse transcriptase. High throughput RT-qPCR has become the benchmark for detection and quantification of RNA targets which is frequently use in novel clinical diagnostic assay (Bustin and Mueller, 2005). .

It is important to study the patterns of expressed genes in order to obtain information on gene regulatory network, which could lead to the identification of novel genes in new biological processes (Vandesompele *et al.,* 2002). Apart from RT-qPCR, there are a few common methods that can be used for quantification of gene transcription, namely northern blotting, *in situ* hybridization, RNAse protection assays, conventional RT-qPCR and microarray. *In situ* hybridization can provide

information on the localization of transcripts in a specific cell and tissues, while the RNAase protection assay is a sensitive method for transcription start-site localization (Carey *et al.,* 2013). On the other hand, northern analysis can provide information on mRNA size, potential alternative splicing and RNA integrity.

The advantages of real time RT-PCR are that they are highly sensitive, specific, and can have broad quantification range in particular for low abundant transcripts in tissues with low RNA concentrations (Mukesh *et al.,* 2006). Microarray allows parallel analysis of thousands of genes at particular time and samples, but it incurred high cost. Real time RT-qPCR analysis has been chosen for this study because it is easy to use, fast, and highly reproductive, with minimal amount of RNA needed and radioactive free (Radonić *et al.,* 2004).

The analysis of gene expression pattern could unravel the functional aspect of gene that is involved in the regulation of biological processes within an organism (Nagavara *et al.,* 2013).There are some criteria that must be gone through prior to the real time RT-qPCR such as the integrity of RNA samples, primer design, housekeeping genes and PCR reaction conditions. Even though real time RT-qPCR is powerful, it has certain pitfalls. The concerns are mostly related to the normalisation of the quantitative RT-qPCR results, which is very essential in order to quantify variations occur in specific gene expression across different samples. Real time RT-qPCR approach is only reliable after the results are normalized. Result normalisation aids in correcting the errors that could be contributed by variations in RNA extraction yield, reverse transcription yield, and efficiency of amplification. The use of reference genes as internal control is the most appropriate strategic for normalising the cellular mRNA data and the gene expression pattern can be experimentally validated for a particular tissueusing a specific experimental design (Bustin *et al.,* 2009).

6.2 Materials and methods

6.2.1 Plant materials

The plant samples used in these experiments were collected from Advanced Agriecological Research Sdn Bhd (AAR) and United Plantation (UP), Malaysia which were the same samples for proteomics analysis as described in Chapter 4 (Section 4.2.1). There were a total of three biological replicates for each high proliferation and low proliferation rate categories. All the samples were ground into powder and kept in -80°C for further used.

6.2.2 DNA extraction

This method was developed by Doyle & Doyle (1990) with some modifications. Approximately 1g of leaf powder was transferred to a 50ml centrifuge tube filled with 5ml of modified CTAB buffer (2% w/v CTAB, 20mM EDTA pH 8, 1.4M NaCl, 100mM Tris-Cl pH8). A few components were added freshly before used, namely 2% w/v polyvinylpyrrolidone-40, 5mM ascorbic acid, 4mM DIECA, 60μl of β-mercaptoethanol and 70μg of RNase A. The mixture was allowed to mixed and incubated at 60°C for 30 minutes. After incubation, the solution was mixed thoroughly with an equal volume of 24:1 chloroform:isoamyl alcohol solution, followed by centrifuging at 13,000 x*g* for 15 minutes. After centrifugation, three phases appeared in the centrifuge tubes. The upper aqueous phase is the DNA content. Next, the upper aqueous phase containing DNA was transferred to a new tube, and 0.6 volume of cold isopropanol was added for DNA precipitation at -20°C for 1 hour. There was no vortex allowed to avoid breaking the DNA strand as DNA started to precipitate at this stage. The precipitate was collected by centrifuging at 13,000 x*g* at 4°C for 15 minutes. Subsequently the pellet was washed with a

solution containing 76% v/v ethanol/10mM ammonium acetate before the pellet was suspended in 4ml of $T_{10}E_1$ buffer (10mM Tris-Cl pH8, 1mM EDTA pH8). It was then followed by 0.5 volume of 7.5M ammonium acetate, and then placed on ice for 20 minutes, before the tube was centrifuged at 14,000x*g* and 4°C for 15minutes. After centrifugation, the solution was transferred to a new tube and 2.5 volume of ethanol was added. The precipitation reaction was incubated for 1 hour at -20°C. A final round of centrifugation resulted in the collection of the DNA pellet at the side of the tube, which was then washed with 5ml of 70% ethanol before being dissolved in 1.5ml of $T_{10}E_1$buffer (10mM Tris-Cl pH 8, 1mM EDTA pH 8). The quality and quantity of extracted DNA were assessed using Nanodrop (Biorad, USA).

6.2.3 RNA extraction

Two different RNA extraction methods were used in this experiment. For the old leaf samples, RNA was extracted using modified CTAB method (Section 6.2.3.1). While, the young leaf samples which is less recalcitrant compared to old leaves was used a simple RNAprep Pure Plant extraction kit to extract RNA (Section 6.2.3.2).

6.2.3.1 RNA extraction using modified CTAB method

This RNA extraction method was carried out with some modifications (Gambino *et al.,* 2008). Approximately 1g of leaf powder was transferred into a 50ml centrifuge tube filled with 10ml of the preheated (65°C) modified CTAB reactants (3% w/v CTAB, 3% PVP (W/V), 20mM EDTA pH 8, 2M NaCl, 100mMTris-Cl pH8, 0.5g/L spermidine, 0.1% DEPC pH8).Freshly prepared 4% β-mercaptoethanol was then added. The mixture was incubated at 65°C for 10 minutes with occasional shaking

to mix the content. After that, the solution was mixed thoroughly with an equal volume of 24:1 chloroform: isoamylalcohol solution and gently mixed for 10 minutes. It was then followed by centrifuging at 12,000 x *g* at 10°C for 10 minutes. The supernatant was transferred to a new tube and extracted again with 10 ml of the chloroform and isoamyl alcohol (24:1) for 10 minutes. After mixed, it was then centrifuged at 12,000 x g at 10°C for 10 minutes. The supernatant was transferred to a new tube containing ¼ volumes of 10M Lithium chloride.

The solution was mixed thoroughly and stored at 4°C for overnight precipitation. After precipitation, it was then centrifuged down at 12,000 x g at 4°C for 30 minutes. The RNA pellet was dissolved very gently with 250 µl 0.5% Sodium Dodecyl Sulphate (SDS), and extracted with chloroform and isoamyl alcohol (24:1). After extraction, it was then centrifuged at 12,000 x g at 4°C for 10 minutes. The supernatant was then transferred to new tube and washed with 2 volumes of ethanol and mix thoroughly for precipitating total RNA at -20°C for at least 2 hours. It was then followed by centrifugation for 30 minutes after the incubation. The supernatant was removed and the pellet was washed twice using 75% ethanol. Lastly, the pellet was dried and re-dissolved in 100 µl RNase-free water.

6.2.3.2 RNAprep Pure Plant kit

This RNA extraction kit was used for the young leaf samples for this experiment. All the extraction procedure was referred to the handbook of the TIANGEN RNAprep Pure Plant kit. The main advantage of using RNAprep Pure Plant kit is due to its minimal amount of samples needed during the RNA extraction. The 1.5 ml Eppendorf tube was loaded with 0.1g leaf samples. The sample was incubated in 500 µl of buffer SL (ensure that β-ME is added to buffer SL before use), and vortexed vigorously. It was then centrifuged for 2 minutes at 13,400 x g. The lysate was then transferred to RNase-Free Filter Column CS which was placed in 2 ml

collection tubes and centrifuged again for 2 minutes at 13,400 x g. The supernatant was carefully transferred to a new microcentrifuge tube without disturbing the cell-debris pellet in the collection tubes.

The cleared lysate was added with 0.4 volume of ethanol and the sample was transferred to RNase-Free Spin Column CR3 placed in a 2ml collection tube and centrifuged for 15 seconds at 13,400 x g. The flow-through was discarded and 350µl Buffer RW1 was added to the spin column CR3. The centrifugation was repeated at 13,400 x g for another 15 seconds. DNase I working solution (10 µl DNase I stock to 70 µl Buffer RDD) was added directly to the centre of spin column CR3 and placed on the bench top (20-30°C) for 15 minutes. It was then washed with 350 µl Buffer RW1 and centrifuged down at 13,400 x g for 15 seconds, followed by washing with 500 µl Buffer RW twice. The flow-through was discarded. Lastly, the spin column allowed centrifugation for 2 minutes at 13,400 x g to dry the spin column membrane. The spin column was then transferred to a new 1.5ml collection tube. RNase –free water (30 – 50 µl) was added directly to the middle of spin column and placed in room temperature for 2 minutes before centrifugation at 13,400 x g for 1 minute in order to get the final RNA. The quantity of RNA was measured using nanodrop ND-1000 (Thermo Scientific, USA).

6.2.4 Nuclear acid Quantification

All the DNA and RNA yield from the samples were read using the Nanodrop ND-1000 (Thermo Scientific Company, USA). The DNA and RNA were also run in 2% agarose gel in order to confirm the presence of DNA and RNA bands. The RNA samples were then quantified using the 2100 Bioanalyzer instruments (Agilent Technologies, USA).

6.2.5 Oligonucleotide design

Following the results from the proteomics in Chapter 4 and Chapter 5, there were a total of 12 candidate genes were selected to be investigated in the current gene expression study. Primers were designed using the Primer 3 software. Prior to oligonucleotide design, the full sequence for each of the candidate genes was searched from the MPOB database (Figure 6.1). In order to make sure all the designed primers were specific to the gene of interest. The un-conserved region of the gene of interest has been selected as priority for primer designed. Primers were designed to have a size of 18-30 base pair (bp), GC content of 40-60% and Tm of 55-65°C. The degree of primer self-annealing had also been taken into consideration. All the oligonucleotides were synthesized by Integrated DNA Technology (IDT) Company. The predicted fragment sizes were ranged from 150 bp to 300 bp.

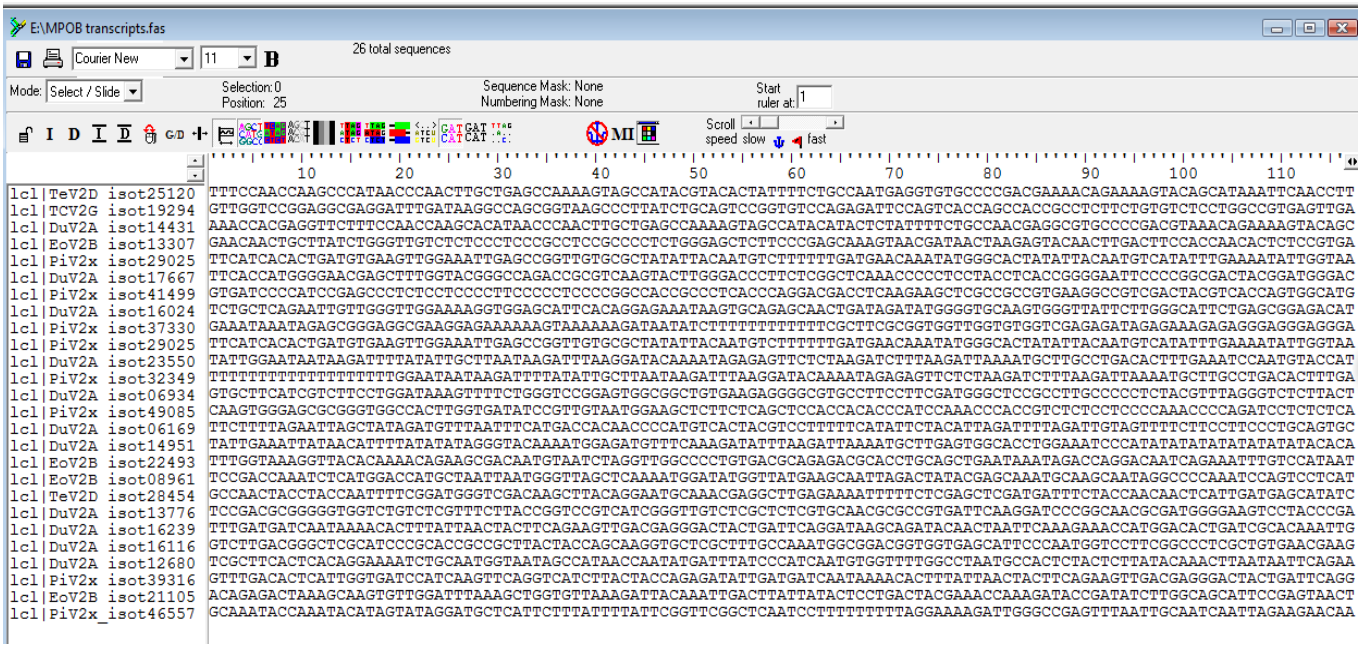


Figure 6.1: Sequences for the selected number of genes for oligonucleotide design

6.2.6 PCR products extraction protocol

The designed primers were used in the PCR reaction with DNA templates extracted from leaf samples. The purpose was to make sure that the designed primers were worked well on the samples prior to do the RT-qPCR analysis. A few primers had been randomly selected for these verification procedures, namely, Primer 3305, 4202, 5403, 7403 and 8605. These primers were used to amplify the gene of interest using conventional PCR method. The PCR products were then fractionated on 1% agarose gel to check the specificity of the primers. If single band was obtained, the PCR products in solution were purified and extracted using MEGAquick-spin™ (Intron, European Biotech Network). The PCR products were added with 5 volumes of BNL buffer and mixed well by vortexing. BNL buffer (100 µL) were added to the 20 µL of PCR products and centrifuged at room temperature to ensure all the content sank to the bottom of the tubes. For sequence less than 200 base pairs, 1.5 volume of isopropanol was added to the samples in order to increase the product yield. The purification was carried using MEGAquick-spin™ column. The PCR products were allowed to bindto the column after centrifugation at 14,500 xg for 1 minute. Next, the column was washed with 700 µl washing buffer followed by centrifugation at 14,500 xg for 1 minute. The centrifugation process was repeated until the membrane was spun dry. The flow-through was discarded and the MEGAquick-spin™ was placed to a new 1.5 ml microcentrifuge tube. It is then followed by the 30- 100 µl of elution buffer directly to the centre of column and incubated at room temperature for 1 minute before centrifugation. Then, the 1.5 ml microcentrifuge tube containing purified PCR products were sent for sequencing (First BASE laboratory Sdn Bhd).

6.2.7 Real time Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR)

6.2.7.1 cDNA synthesis

The conversion of RNA to cDNA was conducted using the QuantiTech Reverse Transcription kit. The cDNA conversion mainly involved 2 main steps which wereelimination of genomic DNA and reverse transcription. Firstly, the purified RNA samples were thawed and incubated with the gDNA Wipeout Buffer reaction (Table 6.1) at 42°C for 2 minutes to remove genomic DNA contaminants. After genomic DNA elimination step, the RNA samples were prepared for reverse transcription step. The master mix prepared from Quantiscript Reverse Transcriptase (1 µl), Quantiscript RT Buffer, 5x (4 µl), and RT Primer Mix (1µl) was added to the RNA template. Quantiscript Reverse Transcriptase enables cDNA synthesis from the range of 10 pg to 1 µg of RNA with high affinity to RNA properties. It was then incubated at 42°C for 30 minutes for the cDNA conversion to carryout followed by the inactivation step at 95°C for 3 minutes. After the conversion, final volume of 20 µl cDNA products was obtained. The cDNA was diluted, to 50 times with RNase-free water. The cDNAwas then ready for the qPCR analysis. All the cDNA products were then kept in -20°C for further use.

Table 6.1 Genomic DNA elimination reaction components

Component	**Volume/reaction**	**Final concentration**
gDNA Wipeout Buffer,7x	2 µl	1x
Template RNA	Variable (up to 1 µg*)	-
RNase- free water	Variable	-
Total volumes	**14 µL**	

* This amount corresponds to the entire amount of RNA present, including any rRNA, mRNA, viral RNA, and carrier RNA present.

6.2.7.2 Real time PCR

Prior to the qPCR reaction, all the cDNA was added with primers, distilled water and master mix solution from KAPA SYBR FAST qPCR kit. KAPA SYBR FAST qPCR master mix (2x) was used in this experiment consisting of KAPA SYBR FAST DNA polymerase, reaction buffer, dNTPs, SYBR Green 1 dye, and 2.5mM $MgCl_2$. KAPA SYBR FAST qPCR kit helps to improve accuracy and reproducibility. It also enables consistent detection of low copy or difficult target. Each of the individual dilutions per primer was carried out in triplicates. There were a total of 17 primers of interest and 3 housekeeping genes carried out for constructing the PCR efficiency curves. A complete real time PCR reaction cycles required 45 to 60 minutes. A three-step experiment run protocol was used: (i) hot-start 95°C for 10 minutes (ii) 40 cycles of (denaturation at 94°C 15 seconds, annealing 56°C 30 seconds, elongation 72°C 30 seconds) (iii) final elongation 72°C 5 minutes.

6.2.7.3 Construction of the PCR efficiency curve for primers

A serial dilution of cDNA was done in order to construct the PCR efficiency curve for each of the primers. The cDNA with (x50) dilution were used as the initial concentration for the PCR efficiency curve. There were a total of 5 series of dilutions carried out for this efficiency curve (Figure 6.2). All the diluted cDNA was aliquot into PCR tubes and ready for the real time PCR analysis using Real time PCR system Mastercycler Ep Realplex (Eppendorf, Germany).

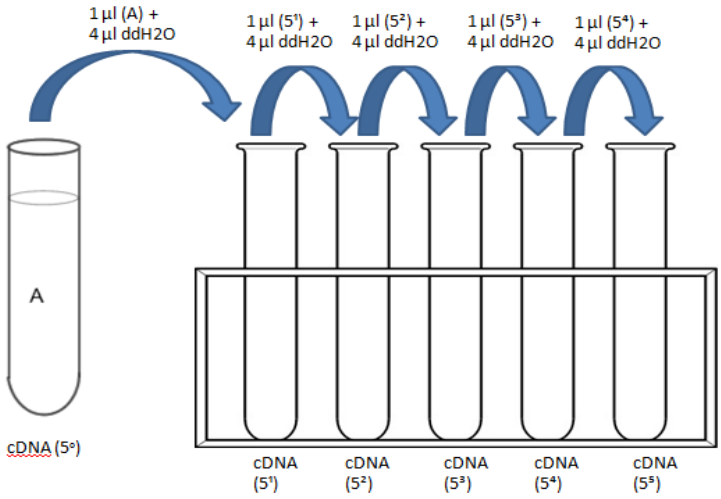


Figure 6.2: Schematic diagram showing a total of 5 serial dilutions (x5) were made from the cDNA products

The negative control (ntc) and non-reverse transcription control (non-rt) were included for each of the qPCR analysis. Both were run in triplicates. The negative control consisting of all the amplification reagents except cDNA template was replaced with distilled water. While non-rt control was the indicator to make sure the primers amplified only the cDNA template in a samples rather than the contaminating genomic DNA that was present in the samples. All samples were run on 96 well plates.

6.2.8 Statistical analysis

All the real time PCR reading was obtained in Ct values which were then normalised with the housekeeping genes. The mRNA expression between samples for different primers was calculated using the formula and the significance level for mRNA expression was determined using the Minitab Statistical Software. One way ANOVA was used in this statistical analysis. The p-value less than 0.1 were used as the significance level for this study. Normality test for each of the data was carried out to ensure the results normally had been distributed.

6.3 Results

6.3.1 RNA extraction

The old leaf samples were extracted using the CTAB modified method (described in section 6.2.3.1), while the young leaf samples were extracted using the RNAprep Pure Plant kit (Section 6.2.3.2). After extraction, the RNA extract were quantified using the Agilent 2100 Bioanalyzer instrument system (Biorad, USA) (Table 6.2).

Table 6.2: RNA concentration for old and young leaf samples

Leaf Samples	Categories	Samples	RNA Yield (ng/µl)	Ratio 28S:18S	RIN* Value
Low Proliferation	Old Leaves	1081 2004 2091	212 771 271	2.3 1.7 1.6	7.5 7.1 7.1
	Young Leaves	194 352 1086	108 184 186	1.6 1.4 1.8	8.1 7.0 8.6
High Proliferation	Old Leaves	933 944 984	413 676 450	1.8 1.6 1.6	7.2 7.1 7.4
	Young Leaves	933 944 984	418 349 944	1.8 1.6 1.3	8.1 8.8 7.0

RIN= RNA integrity number

Table 6.2 shows the RNA yields from oil palm leaf samples.All the RNA yields obtained have a good 28S:18S rRNA ratio of more than 1.3. All the samples reached an acceptance level for downstream analysis with the 28S:18S rRNA ration of more than1.0. Secondly, the RNA integrity number (RIN) value is one of the important information for RNA integrity, it ranged from 1-10. All of the samples

in this studyhad good RIN values above7.1 for old leaf samples while young leaf had the RIN values ranged at 7.0-8.8.

After quantification, the RNA samples were fractionated on 2% agarose gel as shown in Figure 6.3 (a). There were two visible bands shown in the RNA gels i.e. 28S and 18S bands. Visualisation of 28S:18S rRNA ratio on conventional agarose gel is subjective due to the condition of running electrophoresis gel and the amount of RNA loading. The 2100 Agilent Bioanalyzer method was an improved analytical tool for an accurate digital data with the electropherogram showing the ratio of 28S and 18S, concentration of RNA and RIN values. The 2100 Agilent Bioanalyzer had combined the technologies of microfluids, capillary electrophoresis, and fluorescence to calculate the RNA amount and integrity.

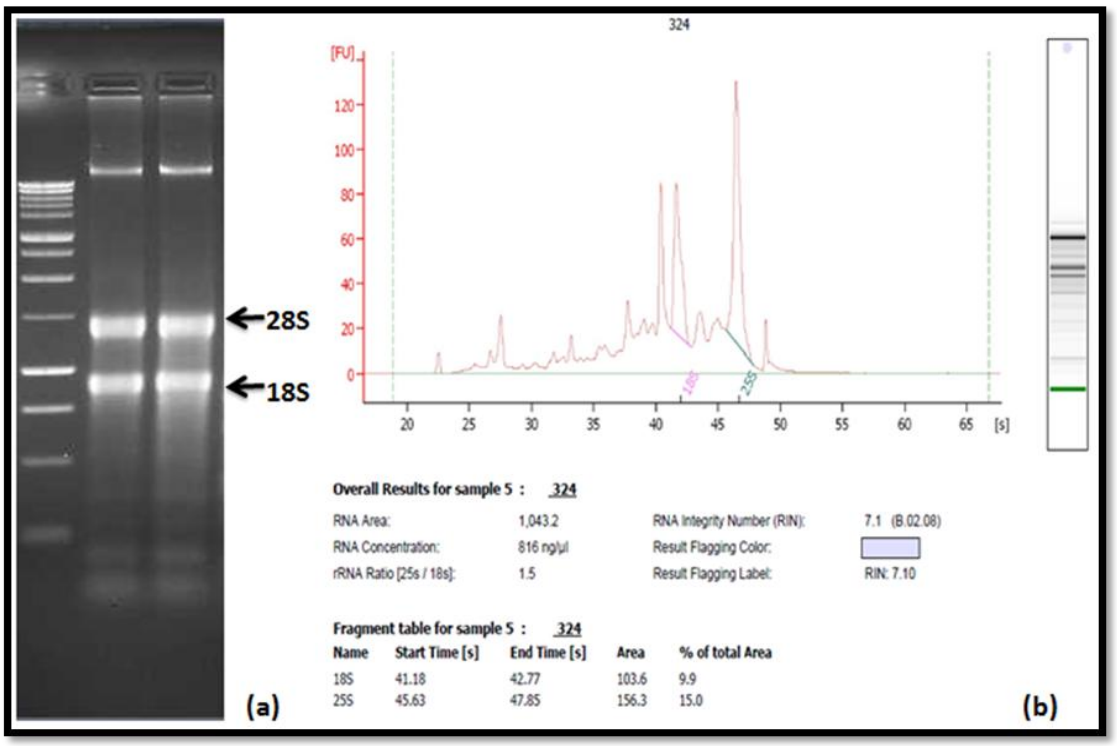


Figure 6.3 RNA fractionation based on (a) Conventional 2% agarose gel method and (b) 2100 Agilent Bioanalyzer method.

6.3.2 Design and validation of primers

Based on the results from Chapter 4 and 5, there were a group of protein candidates that appeared to be significantly different in protein abundance between the high and low proliferation rate groups in both young and old leaf samples. Among these, only the few highest differential abundance values of the proteins were selected for further mRNA analysis using real time RT-PCR. Yet, there were a total of 13 genes selected for the old leaf samples based on the total protein and the PEG fractionated protein. While there were five genes were selected for the young leaf samples.

All the primers were designed using the free access online Primer3 software (online access from: http://simgene.com/Primer3). During primer design, a few considerations were taken in order to increase the specificity of the primers. From the protein candidates, each of their gene sequence was obtained through the MPOB database. After that, the gene sequences of the particular protein were compared with the National Centre for Biotechnology Information (NCBI) database to identify the unconserved regions. The primers were designed based on the unconserved regions within the gene sequences to increase the specificity of the genes amplified. All the design primers were synthesised using services from First Base Laboratories Sdn Bhd (shown as Table 6.3).Using the primers, the extracted DNA was amplified through the Polymerase Chain reaction (PCR) approach and the PCR products were run on the 2% agarose gel to observe the amplification band. All the primers were amplified in single band only (Figure 6.4). There were a few PCR products that have been selected for further sequencing to ensure the perfect match of the amplified sequences with the gene sequences used for primer designation. Prior to the gene sequencing, PCR products were purified and extracted using MEGAquick-spin™ kit. Interestingly, the sequencing results showed that all the primers tested have match the gene sequence ranged from 91% to 99% indicating that the primers were working very well. The high match

rate of the sequencing results showed that the designed primers amplified very well on specific DNA sequences in the samples (Figure 6.5).

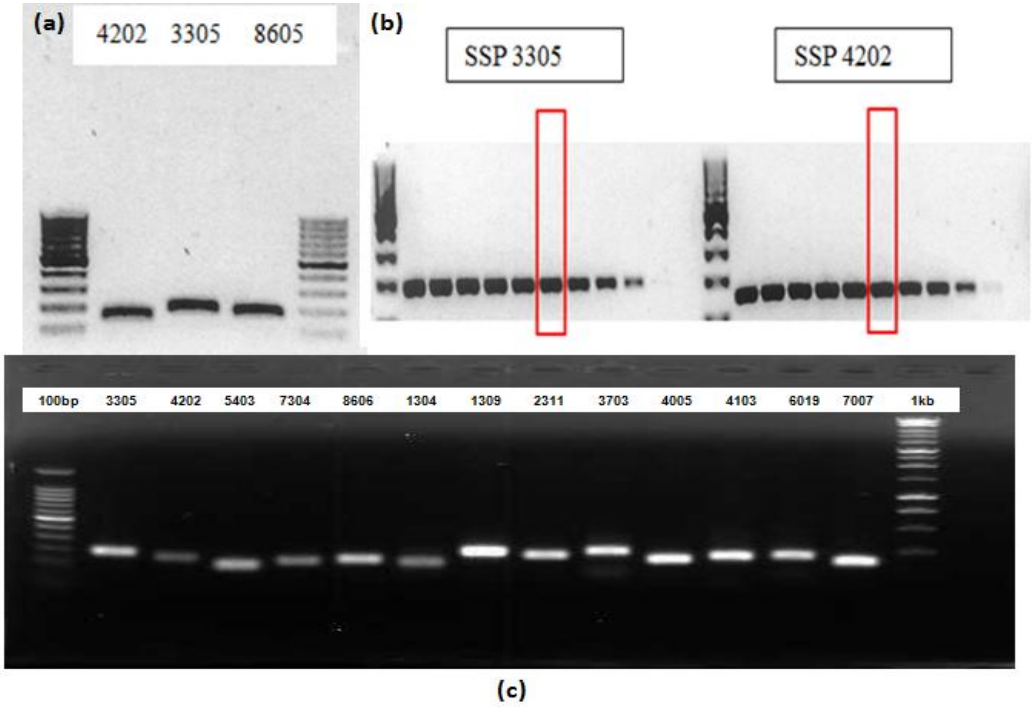


Figure 6.4: (a) Agarose gel with specific DNA band amplified by primers (b) Gradient analysis in temperature during PCR reaction for each of the primers used (c) RNA band amplification by PCR using specific primers.

Sequencing results for SSP 4202

a) Forward sequencing

Sequence ID: lcl|9129 Length: 135 Number of Matches: 1

Range 1: 1 to 135 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
176 bits(95)	7e-49	124/137(91%)	6/137(4%)	Plus/Plus

```
Query  407  CCAtctctctctctccctccccctact-cttctccaatct-ctT-CCTTGGTTGTTGAG  463
            ||||  ||||||||| |||  ||||  |  ||  ||||||||   ||  |||  ||||||||||||||
Sbjct  1    CCATTTCTCTCTCT-CCT-CCCCTTCCTCCTTCTCCATCCTCCTTCCCTTGGTTGTTGAG  58

Query  464  ATTTGGTAAG-TGGCAGTGGGTTATTATGGTGGCTTTGTGATGTGTGGAGGGAAGTGGTT  522
            |||  ||||||  ||||||||||||||||||||||||||||||||||||||||||||||  |||||
Sbjct  59   ATTGGGTAAGTTGGCAGTGGGTTATTATGGTGGCTTTGTGATGTGTGGAGGGAAATGGTT  118

Query  523  TTGCTCCAGATATGGGA  539
            ||||||||||||||||
Sbjct  119  TTGCTCCAGATATGGGA  135
```

Figure 6.5: Sequencing results from the PCR products for SSP4202.

Table 6.3: Sequences of the primers used to amplify transcripts by real time RT-qPCR

Gene	Primer Sequences
3305	Forward: 5'- GGCAGAGTTGAATGCGGATA-3' Reverse: 5'- TTTCTGCCAACGAGGTGT-3'
4202	Forward: 5'- TGGTGGAGGAAGCATGAAGT-3' Reverse: 5'- CCCATATCTGGAGCAAAACC-3'
5403	Forward: 5'- CCTGTGTCTCCAGGTTT-3' Reverse: 5'- GGTCCACAATTGGCTTCACT-3'
7304	Forward: 5'- TTTGCTGGAAGTTGGCTTCT-3' Reverse: 5'- AAGCATCACAAAAACCCAGG-3'
8605	Forward: 5'-AGGCGTGAATGAAACCTTTG-3' Reverse: 5'- CAAAACAAGGATTGGGTTGC-3'
3703	Forward: 5'- AACCCTTTTCCTCGTCGAAT-3' Reverse: 5'- ACTGAGAAAAGCCCCATTGA-3'
1304	Forward: 5'- CGGAAGGAACAATTTTGGAG-3' Reverse: 5'- CACTGCTGAGCATCACAACC-3'
1309	Forward: 5'- CTCCTCTCCCTTTTCTCCGT-3' Reverse: 5'- CGCTCTCCTCCTCAGCAC-3'
4005	Forward: 5'- CTATATCTGGGAGCGGTGGA-3' Reverse: 5'- GAGTCCATGCTGCTCTTTCC-3'
4103	Forward: 5'- ATTGGCACCACATCACAAGA-3' Reverse: 5'- TGTCGTTGTGGTTCCTGAAA-3'
6109	Forward: 5'- AAAAATGGGCAGACATGAGG-3' Reverse: 5'- CAATGTCACCAAACGAGGTG-3'
7007	Forward: 5'- TCCTTTCTGCCCAATGTTTC-3' Reverse: 5'- CCACACCAAGGGCTAAAAGA-3'
2311	Forward: 5'- AGGTGAGAAGCTAGCCACGA-3' Reverse: 5'- TAGGGATCCTCGCAGTTACG-3'
1159/1031	Forward: 5'-GCGTACAAAAGGGCAAACAT-3' Reverse: 5'-CTCTTCGTCTCCTCCACCAG-3'
1469	Forward: 5'-GCACAACTTCGAGGAAGAGG-3' Reverse: 5'-CCCTCTTCAAGAACCAGCAG-3'
520/513	Forward: 5'-TAGGGTTTCTTCGCTGCTGT-3' Reverse: 5'-AGCTCCCGGTTCTTGAAAAT-3'
2415	Forward: 5'-TTTAGCTGGACGACGCTTTT-3' Reverse: 5'-TAAACCCCAGCAGTAAACGG-3'

6.3.3 Housekeeping gene analysis

In this experiment, there were a total of 7 housekeeping genestested across the samples and the optimal two housekeeping genes were used for the subsequent qPCR analysis. Here, the 7 housekeeping genes used, namely Actin (ACT), glyceraldehydes-3-phosphate dehyrdrogenase (GADPH), predicted 40S ribosomal protein S27-2 (PD380), manganese superoxide dismutase (PD569), predicted protein IFH-1 like (EA1332), NADH dehydrogenase subunit 5-like gene (NAD), and polyubiquitin (UBQ) were tested across the samples in order to identify the most suitable reference gene for this experiment. The Ct values for each housekeeping genes were compared and analysed using RefFinder software (shown in Figure 6.6). RefFinder software helps to evaluate reference gene expression by integrating the main four computational programs (GeNorm, Normfinder, BestKeeper, and the comparative Delta CT method). The results were compared in a rank order between four statistical algorithms methods.

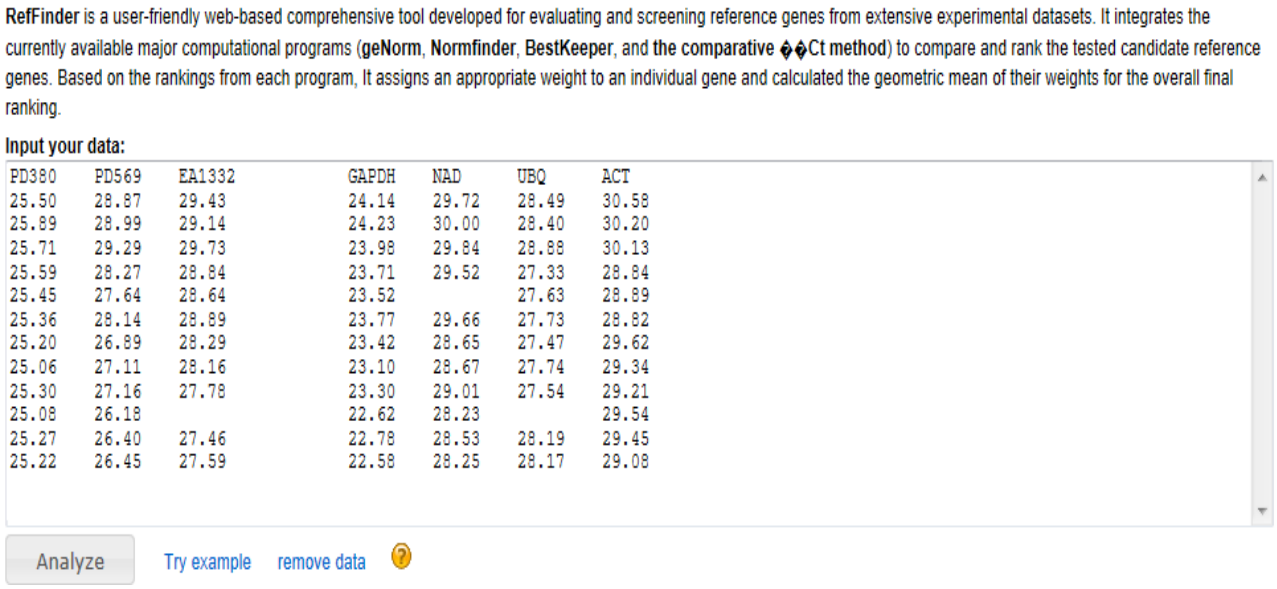
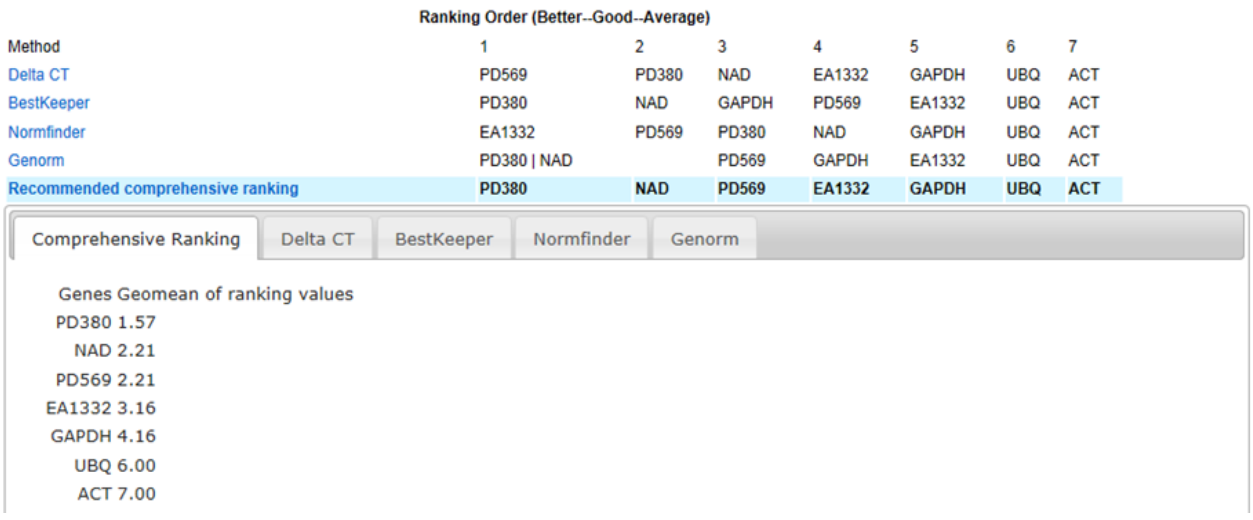


Figure 6.6: RefFinder software analysis for housekeeping genes.

Ranking order from the best to average of housekeeping genes and genes geomean of ranking values for all 7 housekeeping genes were shown in Figure 6.7. Through the RefFinder analysis, the comprehensive ranking order from the four statistical algorithms showed that the best housekeeping genes were PD380 (1.57), followed by NAD and PD 569 (2.21) and thenEA1332 (3.16). It was then followed by the GAPDH (4.16), UBQ (6.00) and lastly ACT (7.00).

GeNorm and Normfinder both were written as a Visual Basic Application (VBA) for Microsoft excel. GeNorm calculates M value for gene expression stability, the most stable gene has lower M value as compared to the less stable genes. The lowest M value belonged to PD380 and NAD in GeNorm analysis. On the other hand, Normfinder used mathematical model-based approaches to estimate the expression variation across samples. It had taken the intra- and intergroup variation to calculate for the expression stability. The results suggested that the most stable genes which represent the lower expression variability which in sequence of EA 1332, followed by PD569 and PD380. While the BestKeeper analysis uses the average Ct values to calculate the coefficient of variance and standard deviation. Gene with lower variation such as PD380 was classified as the most stable one.



Ranking Order (Better--Good--Average)

Method	1	2	3	4	5	6	7
Delta CT	PD569	PD380	NAD	EA1332	GAPDH	UBQ	ACT
BestKeeper	PD380	NAD	GAPDH	PD569	EA1332	UBQ	ACT
Normfinder	EA1332	PD569	PD380	NAD	GAPDH	UBQ	ACT
Genorm	PD380 | NAD		PD569	GAPDH	EA1332	UBQ	ACT
Recommended comprehensive ranking	**PD380**	**NAD**	**PD569**	**EA1332**	**GAPDH**	**UBQ**	**ACT**

Figure 6.7: Result rank orders from the four statistical algorithms methods.

6.3.4 PCR efficiency Curve

Real time RT-qPCR is currently the most accurate method to measure the mRNA expression but this method has generated a series of raw numerical data and the mis-processing of this data may lead to imprecise final results. The PCR standard curve design remains as a reliable method to determine the PCR efficiency in both relative and absolute qPCR (Larionov *et al.,* 2005). All the 17 proteins that selected for the downstream analysis were gone through the RT-qPCR reaction using 5x series dilution of the cDNA to calculate the PCR efficiency for each of the primers. The Ct values obtained from the qPCR reaction were plotted on a graph to get the coefficient value in linear regression (r^2) with a preference of more than 0.98. The real time PCR efficiencies were calculated from the slope using the established equation $E = 10^{[-1/slope]}$ (Bustin *et al.,* 2000). The efficiency of the PCR ranged between 90-100% (-3.6 ≤ slope ≤ -3.3). Figure 6.8 shows the PCR efficiency for primer 3305 which had the slope of -3.567 and R^2 value of 0.996. This primer has good efficiency to proceed for the qPCR analysis.

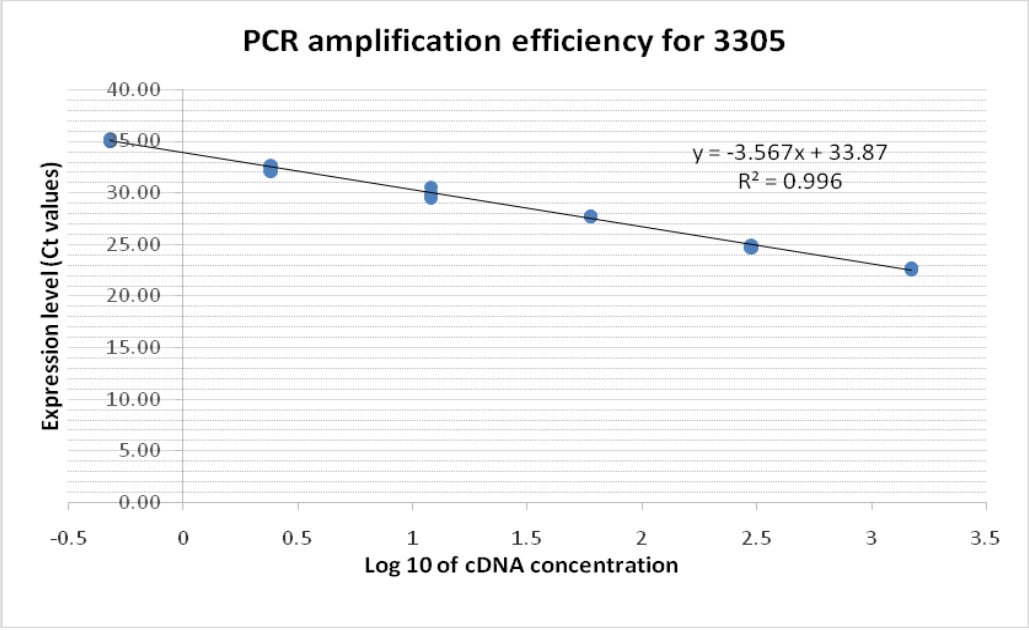


Figure 6.8: PCR amplification efficiency curve for Primer 3305.

6.3.5 Real time PCR analysis

6.3.5.1 Melting curve analysis

Through the RT-qPCR analysis, all the primers were used to amplify the cDNA of samples across the high and low proliferation rate samples. The specificity of the primers across the samples could be identified using the melting curve analysis. After completion of the RT-qPCR reaction, a melting curve was generated by increasing temperature in the reaction which can be monitored from the fluorescent signal at each step. Single peak of the curveindicates the primers were amplifying a single gene at specific degree of temperature. If more than a peak observed in the melting curve, it represents the nonspecific products. All the melting curves for each set of primers were plotted. The results showed that all primers worked very specifically to the samples. Figure 6.9 shows the melting curves of a series of primers set (4202, 5403, 7403, and 8605). All primers showeda single peak at a specific temperature meaning the specificity of the primers set was ensured.

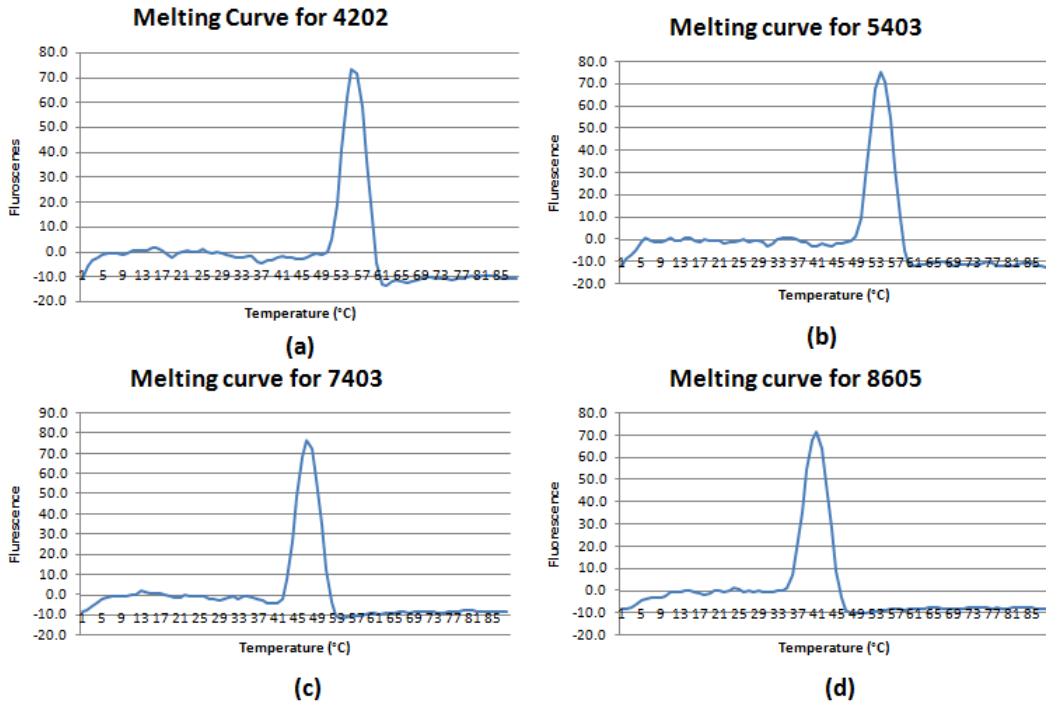


Figure 6.9: Melting curve for examples primers set of (a) 4202 (b) 5403 (c) 7403 (d) 8605.

6.3.5.2 mRNA expression level for high proliferation rate group versus low proliferation rate group

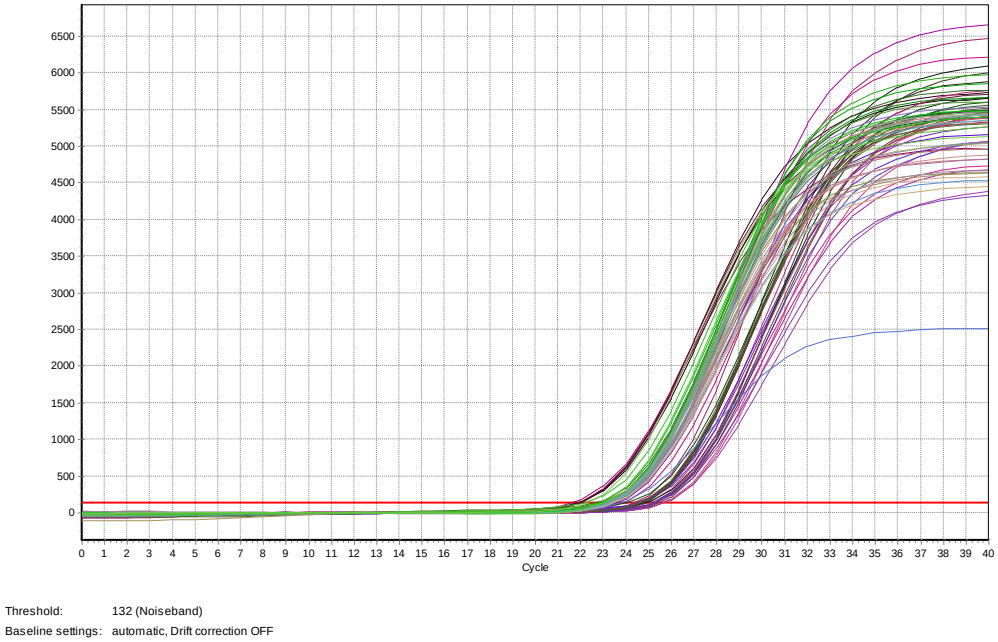


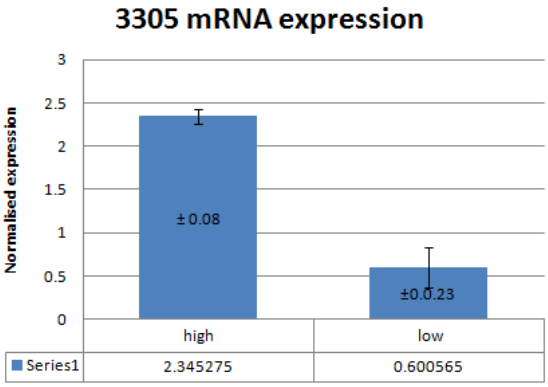
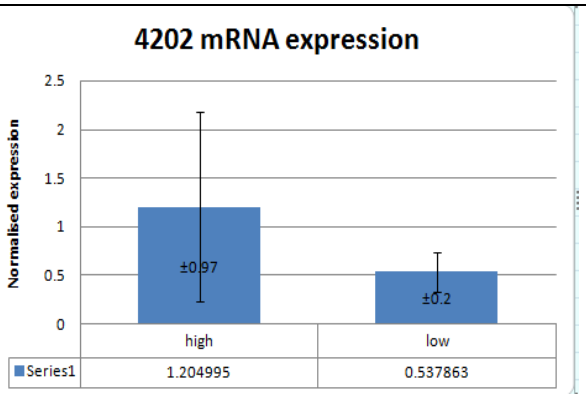
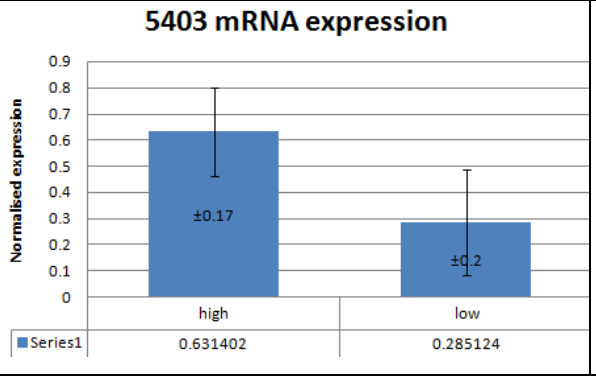
Figure 6.10 Real time RT-qPCR amplification graph versus cycle.

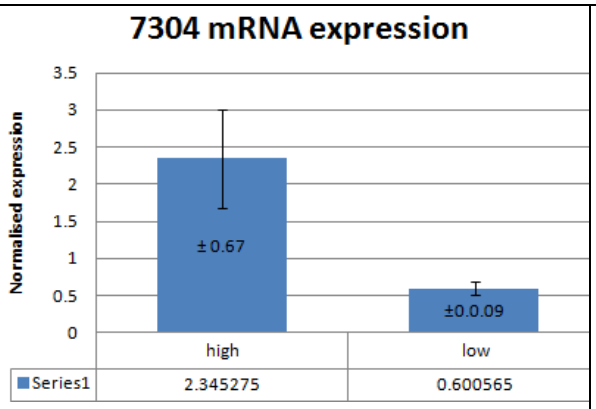
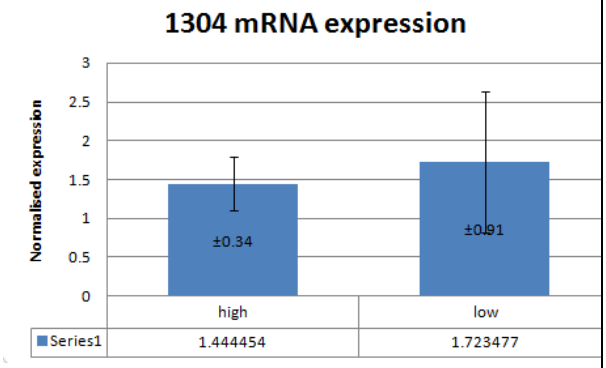
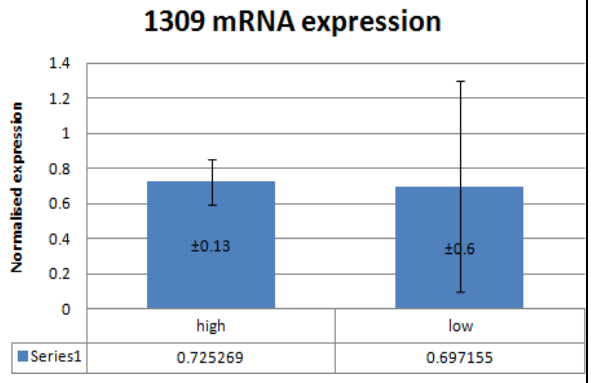
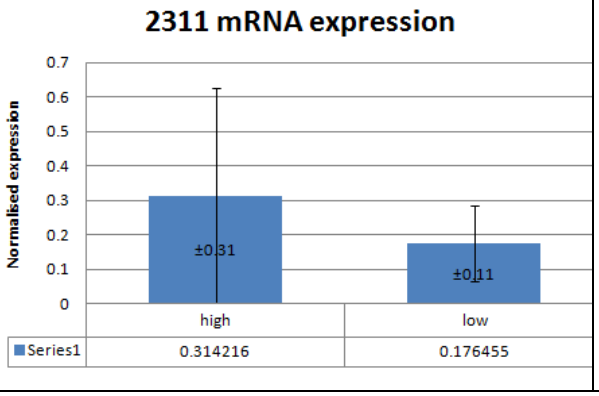
After the primers and RNA integrity had been checked for all samples, the RT-qPCR analysis was carried out accordingly to the experimental design. There were 12 primers tested across the old leaf samples and 7 primers ran on the young leaf samples. Five technical replicates were conducted for each sample to make sure the variations among the replicates measuring less than 0.33. One or two Ct values that contribute variation more than 0.33 were excluded. A minimal of three Ct values encountered for each sample. Each primer was run across a series of high and low proliferation rate samples in the same 96 wells plate to minimize technical errors. After RT-qPCR analysis, all the fluorescent that released during gene amplification per cycle were captured at particular time as shown in Figure 6.10. All mean Ct data and standard deviation among samples were recorded and go for downstream expression calculation using GeNorm analysis. GeNorm analysis is a popular algorithm to calculate the gene expression for each sample based on the geometric mean of reference genes.

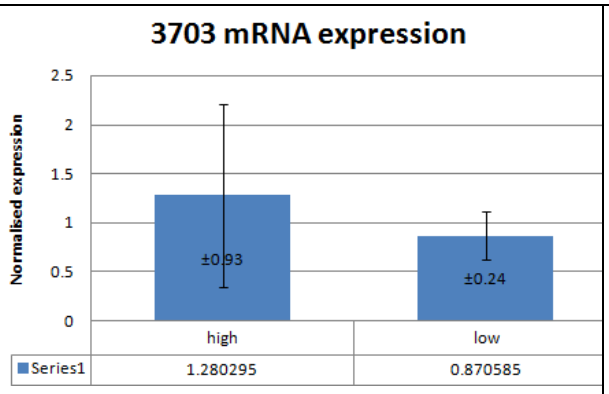
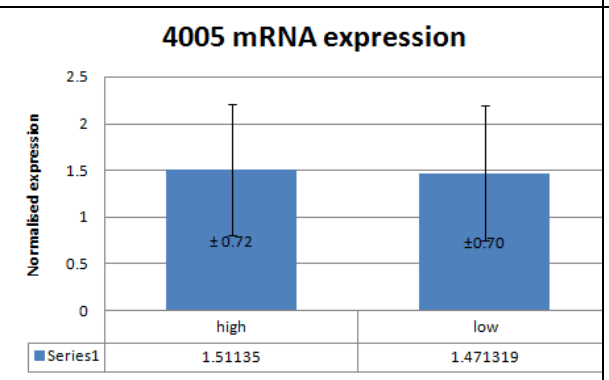
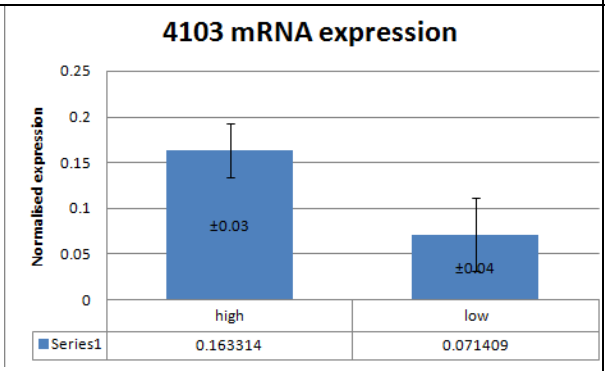
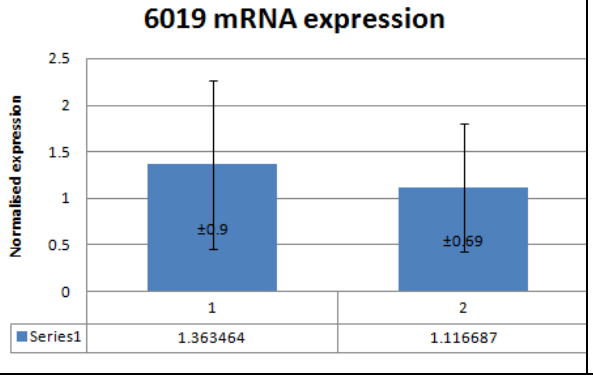
6.3.5.3 Concordance of the mRNA expression level with respective protein abundance

Real time RT-qPCR approach enables the relative quantification of mRNA expression in a gene. There were a total 12 genes from old leaves and 7 genes from young leaves in the gene expression studies. The Ct values that obtained from the experiments were calculated using the GeNorm analysis. The expression value were normalised with the housekeeping genes to correct sample to sample variation. For the old leaf samples as shown in Table 6.4, four out of twelve transcripts were shown to be expressed in concordant with their protein abundance but there are only three transcripts (25%) were shown to be statistically significant. On the other hand, the young leaf samples (Table 6.5) has two out of seven transcripts (28%) were shown to be significant and concordant with their protein abundance. In conclusion, there were a total of six transcriptsthat were concordant in relation with their protein abundance but only four were found to be significant difference using Minitab statistics at p-value < 0.05. Protein candidates that have linear concordance with significant different in statistics with p-value <0.05 were triosephosphate isomerase, L-ascorbate peroxidase, and superoxide dismutase.

Table 6.4: Concordance of the mRNA expression level and protein expression pattern in the high and low proliferation rate group in old leaf samples

Primers	Protein abundance	mRNA expression	Relationship
3305 Triosephosphate isomerase	High abundance in high proliferation rate samples		**Concordant (statistically significant in mRNA)**
4202 Oxygen-evolving enhancer protein 2	High abundance in high proliferation rate samples		**Concordant**
5403 Photosystem II stability	High abundance in high proliferation rate samples		**Concordant**

7304 L-ascorbate peroxidase 2	High abundance in high proliferation rate samples		**Concordant (statistically significant in mRNA)**
1304 EoV2B_isot01043	Present in high proliferation rate gel samples		**Non-concordant**
1309 RuBisCo large subunit binding protein subunit alpha	Present in high proliferation rate gel samples		**Non-concordant**
2311 ACT domain containing protein ACR11	Present in low proliferation rate gel samples		**Non-concordant**

3703 Unidentified	Present in high proliferation rate gel samples		**Concordant**
4005 EoV2B_isot010139	Present in high proliferation rate gel samples		**Non-concordant**
4103 Unidentified	Present in high proliferation rate gel samples		**Concordant**
6019 PiV2x_00805	Present in high proliferation rate gel samples		**Non-concordant**

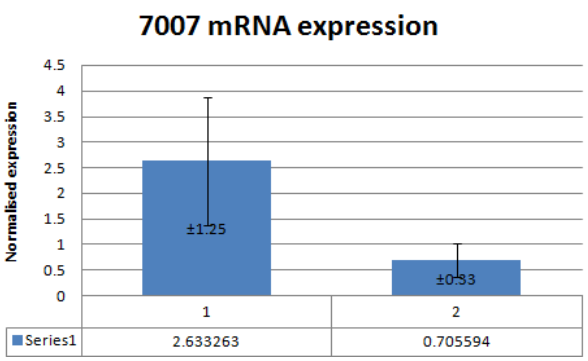
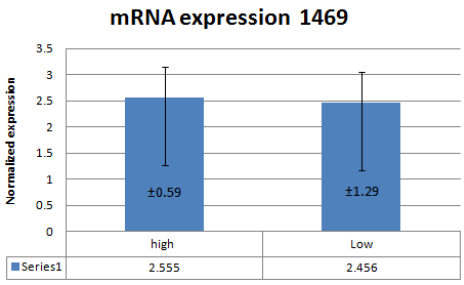
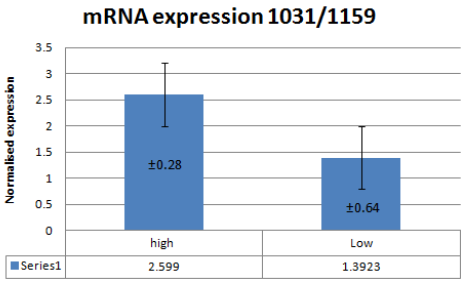
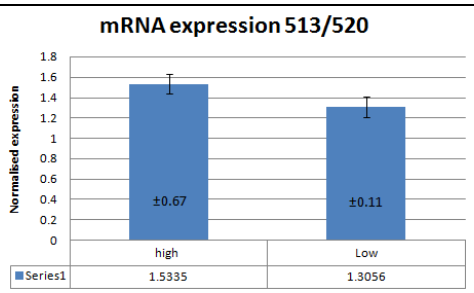
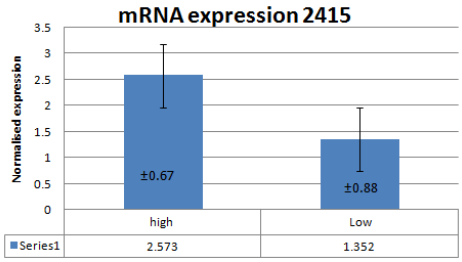
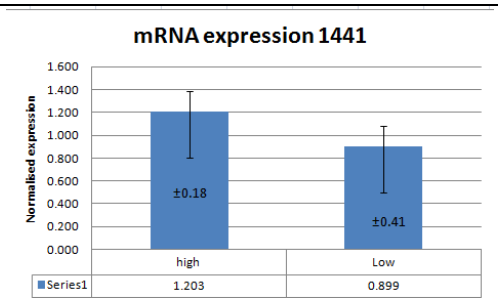
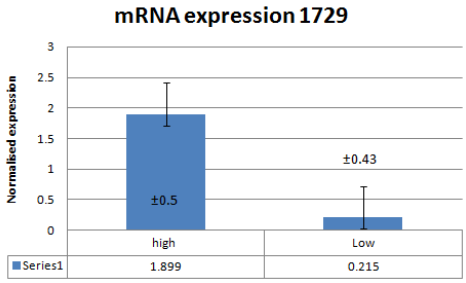
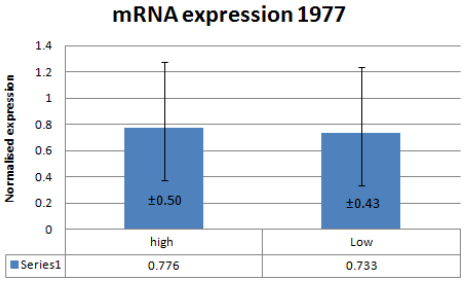
7007 Superoxide dismutase	Present in low proliferation rate gel samples		**Non-concordant (statistically significant in mRNA)**

Table 6.5: Correlation of the mRNA expression level and protein expression pattern in the high and low proliferation rate group in young leaf samples

Primers	Protein expression	mRNA expression	Relationship
1469 Probable receptor-like protein kinase	High abundance in low proliferation rate samples		Non-concordant
1031/1159 Uncharacterised protein	High abundance in low proliferation rate samples		**Concordant***

513/520 Fibrous sheath cabyr-binding	High abundance in high proliferation rate samples		Non-concordant
2415 Ketol-acid Chloroplastic-like	High abundance in high proliferation rate samples		Non-concordant
1441 Stress responsive protein	High abundance in low proliferation rate		Non-concordant
1729 Ferritin chloroplastic-like	High abundance in low proliferation rate		**Concordant***
1977 Germin-like protein	High abundance in low proliferation rate		Non-concordant

6.4 Discussion

6.4.1 RNA extraction

The RNA of the old leaf samples were extracted using CTAB method, while the young leaf samples were extracted using the TIANGEN RNAprep Pure Plant kit. The extraction protocol for young leaf samples was easier as compared to the old leaf samples because young leaves which were less recalcitrant and contained less polysaccharides and contaminants as compared to the old leaf samples. All extracted RNA samples were quantified using the Agilent 2100 Bioanalyzer which provided a clear estimate to the major components that make up total RNA. This is because mRNA constitutes only 1-3% of total RNA samples and it is hardly detectable in experiment. Conventional RNA gel electrophoresis detection method is commonly used for mRNA. This method depends on the assumption of rRNA quality and quantity to reflect the presence of mRNA population (Figure 6.3). There are more than 80% of the total RNA samples are ribosomal RNA with the presence of 28S (5kb) and 18S (2kb) rRNA species, and a ratio of 2:1 for 28S and 18S showing good standard for intact RNA. Total RNA with 28S:18S rRNA ratio of more than 1.0 is well accepted for downstream analysis. Besides the 28S:18S rRNA rations, to test for the integrity of RNA, RNA integrity number (RIN) value are important. The RIN value ranged from 1-10 for the assessment of RNA integrity. The higher the RIN value is the greater of RNA integrity for a sample. RIN values greater than five consider as good total RNA quality and there will be perfect if the RIN value exceeds eight for a downstream application (Simon and Michael, 2006). The quality of the RNA samples is one of the key factors to ensure the reliability of the downstream real time RT-qPCR analysis. All extracted RNA of young and old samples that had good RIN values which are 7 and above that enable to proceed to RT-qPCR analysis.

6.4.2 Housekeeping gene analysis

Ideally, the housekeeping gene should not be affected by the experimental method. But, due to the high dynamic range and sensitivity in real time RT-PCR, many of the well-known housekeeping genes such as GAPDH and β-actin were shown to be influenced by different treatments, different tissue or cell and also biological processes. Hence, it is absolutely important to validate the stability of housekeeping genes with own samples rather than rely on other previous publication (Wong and Medrano, 2005). Normalisation against a single reference gene is not acceptable as stated in the MIQE guidelines for PCR (Bustin *et al.,* 2009). A true housekeeping gene is very important to allow normalisation of differences in (i) amount of cDNA (ii) RNA qualities (iii) enzymatic efficiencies and (iv) variation during RNA and cDNA preparation (Radonić *et al.,* 2004). In this experiment, a total of 7 housekeeping genes were tested across the old and young leaf samples to identify the most stable and reliable control genes for normalisation. After comparing the 7 housekeeping genes, the best three housekeeping genes were PD380, PD569 and EA1332 based on the RefFinder analysis which consists of the comprehensive ranking from the four statistical algorithms. These three housekeeping genes were used for both old and young leaf samples. The similar results were obtained by other group researchers who work on oil palm elite planting materials propagated in tissue culture. From the evaluation of reference genes experiment in oil palm tissue culture by Chan *et al.,*(2014), PD380 and PD569 were selected for accurate and reliable normalization of gene expression data in real time PCR.

6.4.3 Concordance of the mRNA expression level with individual protein abundance

The correlation of mRNA and protein levels was found to be far from perfect in most of the available literatures. There are several main reasons that cause poor correlation between mRNA and protein levels such as post-translational mechanism, post-transcriptional parameters and possibly by noise and experimental error. It is because the mechanism on how genes are transcribed, mRNA processed and translated into functional protein are far from a linear and simple relationship (Maier *et al.,* 2009). Hence, the percentage of correlation between mRNA and protein levels were differ from an organism to another.

A research study on the mRNA-protein expression correlation in yeast and human tissues found that half of the samples showed significant positive correlation for 71 genes (Guo *et al.,* 2008). The correlation for the old leaf samples and young leaf samples in this study has found to be corresponding to 25% and 28%, respectively. This result was similar tothe previous studies that the cellular concentrations of proteins which have roughly 40% correlate with their corresponding mRNA abundance (Vogel and Marcotte, 2012).

For old leaf samples, the mRNAs of three proteins was found to be expressed in linear correlation with their corresponding protein at significant level p-value < 0.05. These were triosephosphate isomerases, L-ascorbate peroxidase, and superoxide dismutase. In this study, triosephoshate isomerase which is an enzyme was abundantly presence in high embryogenic proliferic samples and it is suggested that energy is very important to support proliferation. The same enzyme were reported specifically induced in somatic embryo of Cyclamen persicum Mill together with other three glycolytic enzymes (UDP-glucose pyrophosphorylase, fructose bisphosphate aldolase and glyceraldehydes-3-phosphate dehydrogenase GAPDH) (Winkelmann *et al.,* 2006). L-ascorbate peroxidise play a role in defense

against oxidative stress and photooxidative stress (Vale *et al.,* 2014). High proliferic samples showed higher abundant of ascorbate peroxidise and it is believed that it has better cellular adaptation compare to the low prolific samples. While, Superoxide dismutase (SOD) is an antioxidant and act as first line defense against reactive oxygen species (Alscher *et al.,* 2002). Surprisingly, the study showed that low abundant of SOD were found in high prolific samples. This result supported by Dhir *et al.,* (2014) who found that SOD decreased with the presence of ascorbate peroxidise which act as similar role to protect plant from oxidative stress. Meanwhile, young leaf samples contained ferritin chloroplast-like protein and unidentified protein. Two novel proteins that were unidentified had significant difference in mRNA expression and protein abundance.These proteins could be used as potential candidate markers in both protein and mRNA level for screening proliferic oil palm tissue cultures materials.

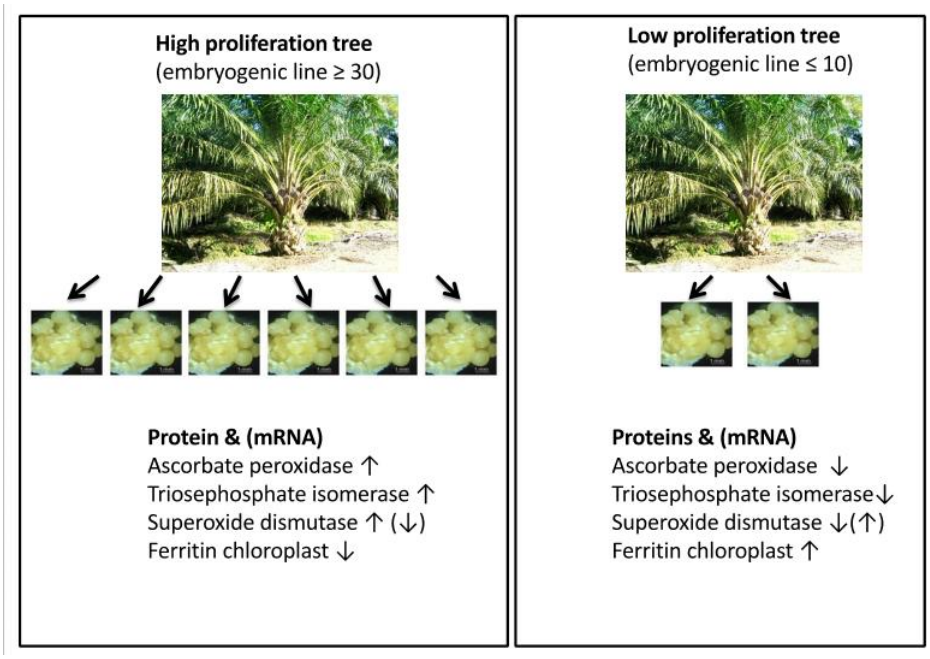


*Figure 6.11: Summary diagram on a group of proteins that exhibit significant in both mRNA expression and protein abundance level in high and low proliferation rate samples (*Upward arrow (↑) indicate high abundance and up regulated; downward arrow(↓) indicated low abundance and down regulated.*

6.5 Conclusion

There were a total of 19 protein candidates that showed significant difference in high and low proliferation rate samples that were undergone RT-qPCR analysis to study their corresponding mRNA expression level. These results were good as there were a total of 5 out of the 17 proteins candidates showed positives concordance between their mRNA expression and protein abundance. These proteins were triosephosphate isomerases, L-ascorbate peroxidase, and superoxide dismutase from old leaf samples. As for young leaf samples, the protein that exhibited concordant correlation between mRNA expression and protein abundance pattern was ferritin chloroplast-like protein. For mRNA expression study, these proteins were significantly difference in expression level between high and low proliferation rate group (p-value < 0.05). These few proteins could be potential candidates as biomarkers to differentiate high and low proliferic starting material for oil palm tissue culture process.

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSIONS

The aim of this study was to use the proteomic approaches to identify a group of proteins that exhibit significantly altered expression in the high and low proliferation rate groups of oil palm samples. This can aid more understanding on the biological processes that occurred in between the two different phenotypic characteristics of the palm. The phenotypic differences observed in these samples were their differences in proliferation performance during somatic embryogenesis in tissue culture process. Prior to applying proteomic approaches, the protein extraction protocols were optimised for the oil palm samples. The TCA/acetone extraction procedure provided the most efficient and reliable method for preparing samples for 2-DE protein separation for both young and old oil palm leaves. All the protein extracts were treated with nuclease mix (DNase and RNase) to remove nucleic acid contaminants. An optimal isoelectric focusing of 10,000 Vhr for pH 4-7 IPG strips that provided good focusing profiles for protein samples extracted from both young and old leaves. The 2D cleanup procedure showed a little improvement in the cabbage samples but not for the old leaf samples.

Through the global profiling for the old leaf samples, 50 spots were assessed for potential alteration in expression level, out of these a total 27 proteins were identified. Most of the identified proteins were involved in photosynthesis and metabolism, followed by proteins involved in cellular biogenesis, stress response and lastly transportation function. Of proteins that have significantly altered expression between the high and low proliferations categories in the old leaves, there were a total of 16 protein spots found to be statistically different. In old leaf samples, mostly photosynthetic proteins were found to be highly expressed in high proliferation rate samples; this suggests that the photosynthesis rate for the high

proliferation palm is higher than the low proliferation palm. Apart from the photosynthesis proteins, in the young leaf samples, more stress responsive proteins were found to be expressed abundantly in high proliferation rate samples. Those proteins responding to pathogen attack and abiotic stress were highly expressed. It indicates that the high proliferation rate samples have higher expression in photosynthetic proteins and a higher ability toward the defense system in plant. From these results, the assumption was made that the high proliferation rate samples have the ability to combat high stress environment with the ability to produce more stress responsive proteins. On the other hand, the high proliferation rate samples appeared to be able to make more food to improve the proliferation rate through their high photosynthetic capabilities.

With the 2-DE gel approaches, most of the identified proteins are present abundantly in the leaf samples. In order to further analyse the low abundance proteins, PEG fractionation is a time and cost-effective method to eliminate the large abundance proteins. With the PEG fractionation, the hidden low abundance proteins were rendered visible on the 2-DE gels. From the PEG fractionation, there was an extra 24 protein spots found to be significantly abundance and 13 of these were successfully identified. This group of low abundance proteins were quite different as compared to the previous proteins that present in old and young leaf samples. Most of the F3 fractionated identified proteins were involved in enzymatic reactions of metabolic processes and some were stress response proteins. Thus, F3 proteins from the PEG fractionation provided another in depth degree of protein profiles which usually could not be observed using general total protein extraction methods. Indeed, analysis of the F3 proteome provided complementary results to the total protein extraction results on the proteins that have significant expression level differences between the high and low proliferation rate samples. Most of the previous studies focused only on the total protein extraction in particular species, different extraction methods will provide a more comprehensive overview of the proteins that are expressed differently in particular categories. From this

experiment, it is suggested that further extraction methods need to be used to study the biological function of a specimen at more in depth level. This will improve the findings and explore more deeply into a dynamic range of the protein population that has been expressed.

With the aim to further understand on the biological function of the plant in high proliferation rate samples, it is interesting to investigate on the relationship between protein and mRNA expression for the differential proteins that were identified in old and young leaf samples. There were 17 protein candidates with significant differences in high and low proliferation rate samples that were chosen to undergo RT-qPCR analysis to study their corresponding mRNA expression level. These results showed that there were 6 out of the 17 protein candidates (35%) showed a positive correlation between their mRNA and protein levels. Hence, in this preliminary finding, the proteins, namely the triosephosphate isomerases, L-ascorbate peroxidase as well as superoxide dismutase from old leaf samples, and from young leaf samples, a ferritin chloroplast-like protein that exhibited a linear correlation in their protein and mRNA expression patterns. These proteins have high potential to further investigate for large population screening and future validation as biomarkers in the high and low proliferation rate group samples.

This preliminary study has found a group of proteins which exhibited significant differences at the proteome level between the high and low proliferation rate samples that were used as explants for tissue culture. In the oil palm industries, there are many individual private sectors with their own practise in the tissue culture laboratory. In this study, the oil palm samples within the high and low proliferation rate groups were obtained from two main companies which are UP and AAR. Prior to concluding the degree of potential utility of the potential biomarkers identified in this work, larger sample populations need to be collected and tested across a group of potential biomarkers in order to stand as an universal biomarker that can be applied in the commercial fields.

REFERENCES

Aaronson, R.M., Graven, K.K., Tucci, M., McDonald, R.J. and Farber, H.W (1995) Non-neuronal enolase is an endothelial hypoxia stress protein. *The Journal of Biological Chemistry* 270(**46**): 27752-27757.

Acquadro, A., Falvo, S., Mila, S., Albo, A.S., Comino, C., Mogila, A. and Lanteri, S (2009) Proteomics in globe artichoke: Protein extraction and sample complexity reduction by PEG fractionation. *Electrophoresis* 30(**9**): 1594-1602.

Agrawal, G.K., Job, D., Zivy, M., Agrawal, V.P., Bradshaw, R.A., Dunn, M.J., Haynes, P.A., Wijk, K.J.V., Kikichi, S., Renaut, J., Weckwerth, W., and Rakwal, R. (2011) Time to articulate a vision for the future of plant proteomics – A global perspective: An initiative for establishing the International Plant Proteomics Organization (INPPO). *Proteomics* 2011. 11(**9**): 1559-1568.

Alexandrova, K. and Conger, B.V. (2002) Isolation of two somatic embryogenesis-related genes from orchardgrass (*Dactylis glomerata*). *Plant Science* 162(**2**): 301-307.

Alscher, R.G., Erturk, N. and Heath, L.S (2002) Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *Journal of Experimental Botany* 53 (**372**): 1331-1341.

Bailey, L.H. (1933) Certain palms of Panama. *Gentes Herb*. 3, Fasc. II.52

Basiron, Y., Balu, N. and Chandramohan. (2004) Palm Oil: The Driving Force of World Oils and Fats Ecomony. *Oil Palm Industry Economic Journal* 4(**1**): 1-10.

Basiron, Y. (2007) Palm oil production through sustainable plantations. European Journal of Lipid Science and Technology 10(**4**): 289-295.

Beirnaert, A. (1935) Introduction à la biologieflorale du palmier à huile (*ElaeisguineensisJacquin*). *Publ. Inst. Nat. Etude agron. Congo Belge, Ser. Sci.,* 5(**3**): 95-110.

Berg, J.M (2007) Structural Biochemistry. Six Edition. New York. 310-323. Retrieved from www.wikibooks.org/wiki/structural_Biochemistry.

Bernard, S. and Habash, D (2009) The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New phytologist* 182(**3**): 608-620

Boutilier, K., Offringa, R., Sharma, V.K. (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14(**8**): 1737-1749.

Braybrook, S.A. and Harada, J.J. (2008) LECs go crazy in embryo development. *Trends in Plant Science* 13(**12**): 624-630.

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T, Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. (1997) The Complete Genome Sequence of Escherichia coli K-12. *Science* 5, 277(**5331**): 1453-1462.

Bustin, S.A (2000) Review: Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25(**2**): 169-193.

Cailloux, F., Julien-Guerrier, J., Linossier, L and Coudret, A. (1996) Long-term somatic embryogenesis and maturation of somatic embryos in *Heveabrasilensis*. *Plant Science* 120: 185-196.

Caref, M.F., Peterson, C.L., and Smale, S.T (2013) The RNase protection assay. *Cold Spring Harbor Protocol* 2013, DOI: 10.1101/pdb.prot071910.

Chan, P.K., Rose, R.J., Abdul Murad, A.M., Zainal, Z., Leslie Low, E.T., Ooi, C.L., Ooi, S.E., Yahya, Suzaini, Y. and Rajinder, S (2014) Evaluation of reference genes for quantitative Real-time PCR in Oil Palm Elite planting materials propagated by tissue culture. PLOS ONE 9(**6**): e99774.

Chen, L.J. and Luthe, D.S. (1987) Analysis of proteins from embryogenic and non-embryogenic rice (*Oryza sativa* L.) calli. Plant Science 48(**3**): 181-188.

Chew, P.S. Soh, A.C., Goh, K.J., and Kee, K.K. (1998) Role of private sector research in oil palm crop production. In: Proceeding 1998 International Oil Palm Conference

Community of the past, today and the future‟ (Ed. By Jatmika *et al*.,) 134-133, Indonesia Oil Palm Res. Inst., Medan, Indonesia.

Chee (2011) The Advancement of Oil Palm Breeding. Agroworld, (translated by Soon, S.H. and reviewed by Wong, C.K.) Issue No. 236: 34-36.

Choi, J.H and Sung, Z.R. (1984) Review and Perspectives: Two-Dimensional Gel Analysis of Carrot Somatic Embryonic Proteins. *Plant Molecular Biology Reporter 2(**3**): 19-25.*

Chow, S.Y., Grondin, M., Averill-Bates, D.A. and Quellet, F. (2016) Plant protein 2-Cys peroxiredoxin TaBAS1 alleviates oxidative and nitrosative stresses incurred during cryopreservation of mammalian cells. *Biotechnology and Bioengineering*, DOI:10.1002/bit.25921.

Chung, H.H., Chen, J.T., and Chang, W.C. (2005) Cytokinins Induce Direct Somatic Embryogenesis of Dendrobium Chiengmai Pink and Subsequent Plant Regeneration. *In Vitro Cellular & Development Biology. Plant* 41(**6**): 765-769.

Corley, R.H.V. and Gray, B.S. (1976) Yield and yield components. In: Oil palm research (Ed. By R.H.V. Corley, J.J. Hardon & B.J. Wood), 7-21, Elsevier, Amsterdam.

Corley, R.H.V., Wooi, K.C., and Wong, K.C. (1979) Progress with vegetative propagation of oil palm*. Plantler*, *Kuala Lumpur* 55: 377-380.

Corley, R.H.V. and Tinker, P.B. (2003) The oil palm. 4th Ed. Oxford, Blackwell Science Ltdp 562. ISBN 0-632-05212-0.

Cronquist, A. (1981) *An integrated system of classification of flowering plants.* Columbia University Press.

Damerval, C., Zivy, M., Granier, F., and de Vienne, D. (1988) Two-dimensional electrophoresis in plant biology, in Advances in electrophoresis (Chrambach, A., Dunn, M., and Radola, B., eds.), VCH, Weinheim, New York, pp. 265-340.

De Jong, F., Mathesius, U., Imin, N. and Rolfe, B.G. (2007) A proteome study of the proliferation of cultured Medicagotruncatula protoplasts. Proteomics 7(**5**): 722-736.

Delbreil, B., Coebel-Tourand, I., Lefrancois, C., and Jullien, M. (1994) Isolation and characterization of long-term embryogenic lines in *Asparagus officinalis L*. *Journal of Plant Physiology* 144: 194-200.

Dennis, E.S., Walker, J.C., Llewellyn, D.J., Ellis, J.G., Singe, K., Tokuhisa, J.G., Wolstenholme, D.R. and Peacock, W.J (1989) Chapter 3: The response to anaerobic stress: transcriptional regulation of genes for anaerobically induced proteins. *Environmental Stress in Plants*, NATO ASI Series, series volume 19: 231-245.

Dhiman, M., Sharma, V., and Moitra, S. (2010) Somatic embryogenesis and Plant regeneration in Ephedra foliate (Boiss.); a non Coniferous Gymnosperm. *Plant tissue culture and Biotechnology*. 20(**2**): 133-143.

Dhir, R., Shekhawat, G.S. and Alam. A (2014) Improved protocol for somatic embryogenesis and calcium alginate encapsulation in *Anethum graveolens* L: a

medicinal herb. *Applied biochemistry and biotechnology* 173(**8**): 2267-2278.

Eapen, S., Abraham, V., Gerdemann, M., and Schieder, O (1989) Direct Somatic Embryogeneis, Plant Regeneration and Evaluation of Plants Obtained from Mesophyll Protoplasts of *Brassica juncea*. *Annals of Botany* 63(**3**): 369-372.

Fabienne, M., Cyrill, G., Hélène, A., Frédérique, R., Rajinder, S., Cheah, S.C., Rival, A., Duval, Y. and James, W.T. (2006) Somaclonal variation in micropropagated oil palm. Characterization of two novel genes with enhanced expression epigenetically abnormal cell lines and in response to auxin. *Tree Physiology* 26(**5**): 585-594.

Filipecki MK, Przybecki Z (1994) Marker protein of somatic embryogenesis in cucumber (*Cucumis sativus* L.). Genetica Polonica 35, 1-9. Department Plant Genetics and Breeding Horticulture, Warsaw Agricultural University, ul. Nowoursynowska 166: 02-766 Warszawa, Poland.

Fraga, D., Meulia, T. and Fenster, S. (2014) Real-Time PCR. Current Protocols Essential Laboratory Techniques. DOI: 10:10.3:10.3.1–10.3.40.

Friedman, D.B. and Lilley, K.S. (2009) Difference gel electrophoresis (DIGE). In "The Protein Protocols Handbook" (Walker John ed.) 3rd edn: 379-408. Human Press, Totowa, NJ.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., Firoozabady, E. (1983) Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220(**4601**): 1049–1051.

Gambino G., Perrone I., and Gribaudo I. (2008) A Rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. *Phytochemical Analysis* 19(**6**): 520-525.

Gantasala, N.P., Papolu, K.P., Thakur, P.K., Kamaraju, D., Sreevathsa, R. and Rao, U (2013) Selection and validation of reference genes for quantitative gene expression studies by real-time PCR in eggplant (*Solanum melongene* L) *BMC Research Notes* 6: 312.

Gianazza, E., De Ponti, P., Scienza, A., Villa, P. and Martinelli, L. (1992) Monitoring by two-dimensional electrophoresis somatic embryogenesis in leaf and petiole explants from Vitis. Electrophoresis 13(**1**): 203-209.

Goffeau, A., Barrell., B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., and Johnston, M. (1996) Life with 6000 genes. Science 274(**546**): 563-567.

Goh, K.J. (2005) Plant Breeding-Germplasm Resources. Retrieved from: http://wordpressdemo.greatbloggingtips.com/category/agro-management/page/12.

Gómez-Vidal,S., Tena, M., Lopez-Llorca, L.V., Salinas, J. (2008) Protein extraction from *Phoenix dactylifera* L. Leaves, a recalcitrant material, for two-dimensional electrophoresis. *Electrophoresis* 29: 448-456.

González-Fernández, R., Prat, E and Jorrín-Novo, J.V (2010) Proteomics of Plant Pathogenic Fungi. *Journal of Biomedical and Biotechnology*, Article ID 932527, Page 1-36.Published online 2010 May 27. DOI: 10.1155/2010/932527, PMC2878683

Görg, A., Weiss, W., and Dunn, M.J. (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4(**12**): 3665-3685.

Görg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J., Madjar, J. (1997) Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. Electrophoresis 18 (**13-14**): 328-337.

Gosden, J. (2007) Rare tropical plants for home and garden – *Elaeis guineensis* [photograph]. Retrieved from: http://toptropicals.com/cgibin/garden_catalog/cat.cgi?uid=*Elaeis_guineensis*.

Guo, Y., Xiao, P., Lei, S., Deng, F., Xiao, G.G., Liu, Y., Chen, X., Li, L., Wu, S., Chen, Y., Jiang, H., Tan, L., Xie, J., Zhu, X., Liang, S and Deng, H (2008) How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochemica et Biophysica Sinica (Shanghai)* 40(**5**): 426-436.

Hahne, G., Mayer, J.E. and Lörz, H. (1988) Embryogenic and callus-specific proteins in somatic embryogenesis of the grass, *Dactylis glomerata* L. Plant Science 55(**3**): 267-279.

Hardon, J.J and Turner, P.D. (1967) Observation on natural pollination in commercial planting of oil palm (*Elaeis guineensis*) in Malaya. *Experimental Agriculture* 3: 105-116.

Hariharan, T., Johnson, P.J., and Cattolico, R.A (1998) Purification and Characterization of Phosphoribulokinase from the Marine Chromophytic Alga *Heterosigma carterae*. *Plant Physiology* 117: 321-329.

Hartley, C.W.S. (1967) The Oil Palm (*Elaeis guineensis* Jacq.) Longman, Green and Co Ltd, London.

Hartley, C.W.S. (1988) The Oil Palm (*Elaeis guineensis* Jacq.) Third Edition. Longman, Green and Co Ltd, London.

Hazra, S., Sathaye, S.S., and Mascarenhas, A.F. (1989) Direct Somatic Embyogenesis in Peanut (*Arachis Hypogea*). *Nature Biotechnology* 7: 949-951

Huang, B.R., Xu, C.P., and Xu Y. (2008) Protein extraction for Two-Dimensional Gel Electrophoresis of Proteomic Profilling in Turfgrass. *Crop Science* 48(**4**): 1608-1614.

Huang, L., Harvie, G., Feitelson, J.S. Gramatikoff, K., Herold, D., Allen, D.L., Amunngama, R., Hager, R.A., Pisano, M.R., Zhang, W.W. and Fang, X.M (2005) Immunoaffinity separation of plasma proteins by IgY microbeads: Meeting the needs of proteomic sample parathionion and analysis. *Proteomics* 5(**13**): 3314-3329.

Heide, H., Kalisz, H.M., and Follmann, H. (2004) The oxygen evolving enhancer protein 1 (OEE) of photosystem II in green algae exhibits thioredoxin activity. *Journal of Plant Physiology* 162(**2**): 139-149.

Helleboid, S., Hendriks, T., Bauw, G., Inzé, D., Vasseur, J. and Hilbert, J.L. (2000) Three major somatic embryogenesis related proteins in Cichorium identified as PR proteins. *Journal of Experimental Botany* 51(**348**): 1189-1200.

Hillenkamp, F. and Karas, M. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* 60 (**20**): 2299-2301.

Hirasawa, M., Brandes, H.K., Hartman, F.C., and Knaff, D.B (1998) Oxidation-Reduction Properties of the Regulatory Site of Spinach Phosphoribulokinase. *Achieves of Biochemistry and Biophysics* 350(**1**): 127-131.

Imin, N., Nizamidin, M., Daniher, D., Nolan, K.E., Rose, R.J., and Rolfe, B.G. (2005) Proteomic analysis of somatic embryogenesis in *Medicagotruncatula*. Explant cultures grown under 6-Benzylaminopurine and 1-Napthaleacetic Acid treatment*. Plant Physiology* 137(**4**): 1250-1260.

Jain, M., Nijhawan, A., Tyagi, A.K., and Khurana, J.P (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications*
345: 646-651.

Jayanthi, M., Susanthi, B., Mohan, N.M. and Mandal, P.K. (2015) In vitro somatic embryogenesis and plantlet regeneration from immature male inflorescence of adult *dura* and *tenera* palm of Elaeisguineensis (Jacq.). *Springerplus* 4: 256-262.

Jonckheere, A.I., Smeitink, J.A.M., and Rodenburg, R.J.T. (2012) Mitochondrial ATP synthase: architecture, function and pathology. *Journal of Inherited Metabolic Disease* 35(2):211-225.

Jong, F., Mathesius, U., Imin, N. and Rolfe, B.G. (2007) A Proteome study of the proliferation of cultured *Medicagotruncatula* protoplasts. *Plant Proteomics* 7(**5**): 722-736.

Jorrín, J.V., Maldonado, A.M., and Castillejo, M.A. (2007) Plant proteome analysis: A 2006 update. *Proteomics* 7(**16**): 2947-2962.

Jorrín-Novo, J.V., Maldonado, A.M., Echevarrí-Zomeňo, S., Valledor, L., Castillejo, M.A., Curto, M., Valero, J., Sghaier, B., Donoso, G., and Redondo, I. (2009) Plant proteomic update (2007-2008): Second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. *Journal of proteomics* 72: 285-314.

Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* 60 (**20**): 2299-2301.

Khaw, C.H. and Ng, C.K. (1997) Performance of commercial scale clonal oil palm (*Elaeisguineensis* Jacq.) plantings in Malaysia. International Symposium on Biotechnology of Tropical and Subtropical Species Part 2. *ISHS ActaHorticulturae* 461.

Khurana, P. and Chugh, A. (2002) Gene expression during somatic embryogenesis – recent advances. *Current Science*, 83 (**6**): 715-730.

Kim, S.G., Wang, Y., Lee, K.H., Park, Z.Y (2013) In-depth insight into in vivo apoplasticsecretome of rice Magnaportheoryzae interaction. *Journal of Proteomics* 78: 58-71.

Kim, S.T., Cho, K.S., Jang, Y.S. and Kang, K.Y (2001) Two-dimensional electrophoresis analysis of rice proteins by polyethylene glycol fractionation for protein arrays. *Electrophoresis* 22(**10**): 2103-2109.

King, E.E (1971) Extraction of cotton leaf enzyme with borate. *Phytochemistry* 10(**10**): 2337-2341.

Kmiecik, S. and Kolinksi, A (2011) Simulation of Chaperonin Effect on Protein Folding: A shift from Nucleation- Condensation to Framework Mechanism. *Journal of the American Chemical Society* 133(**26**): 10283-10289.

Konan, E.E., Gasselin, T.D., Kouadio, J.Y., Flori, A. and Rival, A (2007) A modeling approach of the *in vitro* conversion of oil palm (*Elaeis guineensis*) somatic embryos.
Plant Cell, Tissue and Organ Culture 84 (**1**): 99-112.

Kuras, R. and Wollman, F.A. (1994) The assembly of cytochrome b6/f complexes: an approach using genetic transformation of the green alga Chlamydomonas reinhardtii. *The EMBO Journal* 13(**5**): 1019-1027.

Kushairi, A., and Sambanthamurthi, R. (2006) Breeding & biotechnology: A powerful value additive for the oil palm industry. EuroFedLipid Conference, 1-4 Oktober, Madrid (Spain). http://www.eurofedlipid.org/meetings/madrid/.

Kushairi, A., Tarmizi, A.H., Samsuri, I., Ong-Abdullah, M., Samsul Kamal, R., Ooi, S.E. and Rajanaidu, N. (2010) Production, performance and advances in oil palm tissue culture. Paper presented at the *ISOPB Seminar on Advances in Oil Palm Tissue Culture*. Yogyakarta, Indonesia. 29 May 2010.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.

Lander, E.S with International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Larionov, A., Krause, A., and Miller, W (2005) A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics* 6: 62-78.

Lee, B.S. and Son, S.H. (1995) A method for producing taxol and taxanes from embryo cultures of *Taxus* species. WO PatentN°95/02063.

Lee, D.G., Ahsan, N., Lee, S.H., Kang, K.Y., Lee, J.J. and Lee, B.H. (2007) An approach to identify cold – induced low abundance proteins in rice leaf. *Comptes Rendus Biologies* 330(**3**):215-225.

Lee, L (2012) Economic growth, climate change, confusion, and rent seeking: The case of palm oil. *Journal of Oil Palm & The Environment* 3:1-8.

Lefebvre, S., Lawson, T., Zakhleniuk, O.V., Lloyd, J.C. and Raines, C.A. (2005) Increased Sedoheptulose-1,7-Bisphosphate activity in transgenic tobacco plants stimulate photosynthesis and growth from an early stage in development. *Plant Physiology* 138(**1**):451-460.

Leplae, E. (1939) Le Palmier à huileenAfrique, son exploitation en Congo BelgeetenExtrème-Orient. *Mem. Inst.r.colon.belge Sect. Sci. natméd.,* 7(**3**): 108.

Lilley, K.S., and Freidman, D.B. (2004) All about DIGE: Quantification technology for differential-display 2D-gel proteomics. *Expert Rev. Proteomics* 1, 401-409.
Lippert, D., Zhuang, J., Ralph, S., Ellis, D.E., Gilbert, M., Olafson, R., Ritland, K., Ellis, B., Douglas., C.J. and Bohlmann, J. (2005) Proteome analysis of early somatic embryogenesis in *Piceaglauca*. *Proteomics* 5(**2**): 461-473.

Litz, R.E., Moore, G.A. and Srinivasan, C. (1985) *In vitro* systems for propagation and improvement of tropical fruits and palms. *Horticultural Reviews* 7: 157-200.

Low, E.T., Alias,H., Boon, S>H., Shariff, E.M., Tan, C.Y., Ooi, C.L., Cheah, S.C., Raha, A.R., Wan, K.L. and Rajinder, S. (2008) Oil Palm (Elaeisguineensis Jacq.) tissue culture ESTs: Identifying gene associated with callogenesis and embryogenesis. BMC Plant Biology 8(**62**):1-19.

Maheran, M., Clyde, M.M., and Cheah, S.C. (1999) Application of genomic in situ hybridization (GISH) on *Elaies* hybrids. *Journal of Oil Palm Research*, Special Issue (Oktober 1999), 74-80.

Maier, T., Güell, M and Serrano, L. (2009) Review: Correlation of mRNA and protein in complex biological samples. *Journal of FEBS letters* 583(**24**): 3966-3973.

Mann, M., Hendrickson, R.C. and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annual Review Biochemistry* 70: 437-473.

Mosarof, M.H., Kalam, M.A., Masjuki, H.H. Ashraful, A.M., Rashed, M.M., Imdadul, H.K., and Monirul, I.M. (2015) Review: Implementation of palm biodiesel based on economic aspects, performance, emission, and wear characteristics. *Energy Conversion and Management* 105: 617-629.

McCabe, P.F., Valentine, T.A., Forsberg, L.S., and Pennell, R.I. (1997) Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot. *Plant Cell* 9(**12**): 2225-2241.

McGinley, M., Swarthout, D. and Taub, D.R (2010) “Photosynthesis”. In: Encyclopedia of Earth Editions. Cutler J. Cleveland (Washington, D.C: Environmental Information Coalition, Natianal Council for Science and the Environment). Retrieved from http://www.eoearth.org/article/Photosynthesis

Meilina, O.A., and Ooi, S.E (2006) Biomarkers: finding a niche in oil palm tissue culture. Part 1- Laying the foundation. *Oil Palm Bulletin*. 53: 36-48.

Michael, H., Haike, A., Knut, B. and Jörg, B. (2008) Review: Gel-based proteomics of Gram-positive bacteria: A powerful tool to address physiological questions. *Proteomics* 8(**23-24**): 4958-4975.

Miyagawa, Y., Tamoi, M., and Shigeoka, S (2001) Overexpression of a cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphate in tobacco enhance photosynthesis and growth. *Nature biotechnology* 19: 965-969.

Mizobuchi, A. and Yamamoto, Y. (1989) Assembly of photosystem II polypeptides and expression of oxygen evolution activity in the chloroplasts of

Euglena gracilis Z. during the dark-light transition. Biochemisty Biophysical Acta 977(**1**): 26-32.

Murota, K.I, Ohshita, Y., Watanable, A., Aso, S., Sato, F., and Yamada, Y (1994) Changes related to salt tolerance in thylakoid membrances of phtoautotrophically cultured green tobacco cells. Plant Cell Physiology 35(**11**): 107-113.

Murphy, D.J. (2007) Future prospects for oil palm in the 21st century: Biological and related challenges. *European Journal of Lipid Science and Technology* 109: 296-306.

Naing, A.H., Chung, J.D., and Lim, K.B. (2011) Plant Regeneration through Indirect Somatic Embryogenesis in *Coelogyne Cristata* Orchid. *American Journal of Plant Sciences* 2: 262-267.

Newton, R.P., Brenton, A.G., Smith, C.J. and Dudley, E. (2004) Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry*, 65(**11**): 1449–1485.

Official Palm Oil Information centre (2012) Malaysia Palm Oil Industry. Malaysia Palm Oil Board. Retrieved from: http://www.palmoilworld.org/about_malaysian-industry.html

Ooi, S.E., Harikrishna, K and Meilina, O.A. (2008) Isolation and characterization of a putative serine/threonine kinase expressed during oil palm tissue culture. *Journal of Oil Palm Research.* 1: 14-22.

Oropeza, M and De Gracia, E. (1996) Somaclonal variants resistant to sugarcane mosaic virus and their agronomic characterization. In Vitro Cellular and Development Biology, 32: 26-30.

Pakiam, Ranjeetha (16 September 2013) ."Malaysia Keeps Palm Oil Export Tax Unchanged to Spur Shipments". Bloomberg News. Retrieved22 September 2013.

Paulsen, H., Rumler, U. and Rudiger, W (1990) Reconstitution of pigment-containing complexes from light-harvesting chlorophyll a/b binding protein overexpressed in *Escherichia coli.* Planta 181(**2**):204-211.

Quiroz-Figueroa, F., Monforte-González, M., Galaz-Avalos, R.M., and Loyola-Vargas, V.M. (2006) Direct somatic embryogenesis in Coffeacanephora. *Methods in Molecular Biology* 318: 111-117.

Radonić, A., Thulke, S., Mackay, I.M., Landt, O., Siegert, W and Nitsche, A (2004) Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Reseach Communication* 313(**4**): 856-862.

Rajanaidu, N., Rohani, O., and Jalani, B.S. (1997) Oil Palm clones: current status and prospects for commercial production. *The Planter* 73 (**853**): 163-184.

Rohani,O., Sharifah, S.A.,Rafii, M.Y., Ong, M., Tarmizi, A.H. and Zamzuri,I. (2000) Tissue culture of oil palm. In Advances in Oil Palm Research. Edited by: Basiron,Y., Jalani, B.S., Chan, K.W.,Bangi. Malaysian Palm Oil Board: 238-283.

Rosengren, A.T., Salmi, J.M., Aittokallio, T., Westerholm, J., Lahesmaa, R., Nyman, T.A. and Nevalainen, O.S. (2003) Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. *Proteomics* 3(**10**): 1935-1946.

Rosenthal, D.M., Locke, A.M., Khozaei, M., Raines, C.A., Long, P.L. and Donald, R.O (2011) Over-expressing the C_3 photosynthesis cycle enzyme Sedoheptulose-1,7-bisphosphate improves photosynthetic carbon gain and yield under fully open air CO2 fumigation (FACE). *Biomedical Plant Biology*11: 123-135.

Ruan, K., Duan, J.B., Bai, F.W., Lemaire, M., Ma, X.Z. and Bai, L.H (2009) Function of Dunaliellasalina (Dunaliellaceae) enolase and its expression during stress. *European Journal of Phycology* 44(**2**): 207-214.

Sallandrouze, A., Faurobert, M., El Maataoui, M. and Espagnac, H. (1999) Two-dimensional electrophoresis analysis of proteins associated with somatic embryogenesis development in *Cupressus sempervirens* L. Electrophoresis 20: 1109-1119.

Sambanthamurthi, R., Singh, R., Kadir, A.P.G., Abdullah, M.O., Kushairi, A. (2009) Oppurtunities for the oil palm via breeding and biotechnology. In: Jain, S.M. Priyadarshan, P.M. (eds.) Breeding plantation tree crops: tropical species. *Springer*, New York: 377-421.

Schmidt, E.D.L., Guzzo, F., Toonen, M.A.J., and deVries, S.C. (1996) A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *DevelopmetCambridge England* 124(**10**): 2049-2062.

Shan, X.Y., Wang, J.X., Chua, L.L., Jiang, D., Peng, W. and Xie, D.X. (2011) The role of Arabidopsis Rubisco Activase in Jasmonate-induced leaf senescence. *Plant Physiology* 155(2):751-764.

Shaw, M.M. and Reiderer, B.M. (2003) Sample preparation for two dimensional gel electrophoresis. *Proteomics* 3(**8**): 1408-1417.

Shiao, T., Ellis, M.H., Dolferus, R., Dennis, E.S. and Doran, P.M (2002) Overexpression of alchohol dehydrogenase or pyruvate decarboxylase improves the growth of hairy roots under hypoxia. Biotechnology and Bioengineering 77(**4**):

455-461.

Shim, S.Y., Ali, R.M. and Gam, L.H. (2010) Protein profiling of *Piper sarmentosum*.
Asia-Pacific Journal of Molecular Biology and Biotechnology 18(**3**): 303-313.

Silva, R.C., Carmo, L.S.T., Luis, Z.G., Silva, L.P., Pereira, J.E.S and Mehta, A (2014) Proteomic identification of differentially expressed proteins during the acquisition of somatic embrygenesis in oil palm (Elaeisguineensis Jacq.) *Journal of Proteomics* 73(**1**): 161-177.

Simon, F. and Micheal, Pfaffl, W (2006) Review: RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* 27(**2-3**): 126-139.

Singh, A.K. and Chand, S. (2003) Somatic embryogenesis and plant regeneration from cotyledon explants of a timber-yielding leguminous tree, DalbergiasissooRoxb. *Journal of Plant Physiology* 160: 415-421.

Soh, A.C., Wong, G., Tan, C.C., Chew, P.S., Hor, T.Y., Chong, S.P., and Gopal, K. (2001) Recent advances towards commercial production of elite oil palm clones. Proceedings of PIPOC 2001 International Palm Oil Congress. Agriculture Conference: 33-44.

Soh, A.C., Wong, C.K., Ho, Y.W., and Choong, C.W. (2009) “Oil Palm” Oil Crops, Handbook of Plant Breeding 4. (Eds. Vollmann, J. and Rajcan, I.) *Springer Science*, New York: 333-367.

Stalikas, C.D. (2007) Extraction, separation and detection methods for phenolic acids and flavonoids. *Journal of Separation Science* 30(**18**): 3268-3295.

Stephan, A.B., Vladimier, B., Jeremy, A.G., Jan, H., Jim, H., Mikael, K., Reinhold, M., Tania, N., Michael, W.P, Gregory, L.S., Jo, V., and Carl, T.W (2009)
The MIQE Guidelines: Minimum Information for Publication of Quantittive Real-Time PCR Experiments. *Clinical Chemistry* 55(**4**): 611-622

Stone, S.L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fisher, R.L., Goldberg, R.B. and Harada, J.J (2001) LEAFTY COTYLEDON 2 encodes a B3 domain transcription factor that induces embryo development. *Proceeding of the National Academy of Sciences of the USA* 98(**20**): 11806-11811.

Stone, S.L., Braybrook, S.A., Paula, S.L. (2008) Arabidopsis LEAFY COTYLEDON 2 induces maturation traits and auxin activity: implications for somatic embryogenesis.
Proceedings of the National Academy of Sciences of the USA 105(**8**): 3151-3156.

Subhra, C., Ghasem, H.S., Yang, P.F., Woo, S.H., Chin, C.F., Chris, G., Paul, A.H., Mehdi, M. and Komatsu, S (2015) Proteomics of Important Food crops in the Asia Oceania Region: Current Status and Future Perspectives. *Journal of Proteome Research* DOI: 10.1021.

Sugihara, K., Hanagata, N., Dubinsky, Z., Baba, S. and Karuba, I. (2000) Molecular Characterization of cDNA Encoding Oxygen Evolving Enhancer Protein I Increased by Salt Treatment in the Mangrove *Bruguiera gymnorrhiza*. Plant Cell Physiology 41(**11**): 1279-1285.

Sung, Z.R., and Okimoto, R. (1981) Embryogenic proteins in somatic embryos of carrot. Proceeding of the National Academy of Sciences of United States of America 78(**6**): 3683-3687.

Susanne, H.S., Josip, S., and Dayne, M. (2008) Proteomic Approaches for Studying Alcoholism and Alcohol-Induced Organ Damage. *National Institute on Alcohol Abuse And Alcoholism (NIAAA)* 31(**1**): 36-48.

Tan, H.S., Meilina, O.A., Choo, C.N., Wong, W.C., Susan, L., and Chin, C.F (2011) Optimisation of Protein Extraction Methods for 2-DE Proteomic Analysis Using Oil Palm Leaves*. International Palm Oil Congress (PIPOC) 2011.*

Tarmizi, A.H., Norjihan, M.A., Samsul Kamal, R., Zaitun, R., and Cheah, S.C. (2003) Proceeding of the PIPOC 2003 International Palm Oil Congress. Agriculture Conference 130-144.

Te-chato, S., and Chehmalee, S. (2008) Induction of somatic embryogenesis and plantlet regeneration from cultured zygotic embryo of oil palm. *Journal of Agricultural Technology* 4(**2**): 137-146.

Thuzar, M., Vanavichit, A., Tragoonrung, S. and Jantasuriyarat, C. (2011) Efficient and rapid plant regeneration of oil palm zygotic embryos cv. „Tenera" through somatic embryogenesis. *Acta of Physiology Plant* 33(**1**): 123-128.

Tirumalai, R.S., Chan, K.C., Prieto, D.A., Issaq, H.J., Conrads, T.P., Veenstra, T.D. (2003) Characterization of the low molecular weight human serum proteome. Molecular and Cellular Proteomics 2(**10**): 1096-1103.

Tomlinson, P.B. (1961) *Anatomy of the monocotyledons*. II. *Palmae*. Oxford.

Murphy, D.J. (2007) Future prospects for oil palm in the 21st century: Biological and related challenges. *European Journal of Lipid Science and Technology*. 109: 296-306.

Tyers, M. and Mann, M. (2003) From genomics to proteomics. *Nature* 422: 193-197.

Unlu, M., Morgan, M.E. and Minden, J.S. (1997) Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* 18: 2027-2077.

Vale, M., E. Heringer, A.S., Barroso, T., de Silva Ferreira, A.T., da Costa, M.N., Jonas, E.A.P., Santa-Catarina, C. and Vanildo, S. (2014) Comparative proteomic analysis of somatic embryo maturation in Carica papaya L. *Proteome Science* 12: 37-54.

Vandesompele, J., Preter, K.D., Pattyn, F., Poppe, B., Roy, N.V., Paepe, A.D. and Speleman,F. (2002) Accurate normalization of real-time quantification RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3(**7**): 0034.1-0034.11.

Vâlcu, C.M. and Schlink, K. (2006) Reduction of proteins during sample preparation and two-dimensional gel electrophoresis of woody plant samples. Proteomics 6(**5**): 1599-1605.

Veena (2008) Gourmet India- Konkan Fruit Festival in Goa [photograph]. Retrieved from: http://www.gourmetindia.com/topic/526-konkan-fruit-festival-in-goa.

Vicient, C.M. and Martínez, F.X. (1998) Review: The potential uses of somatic embryogenesis in agroforestry are not limited to synthetic seed technology. *RevistaBrasileira de Fisiologia Vegetal* 10(**1**): 1-12.

Vijaya, S., Choo, Y.M., Halimah, M., Zulkifli, H., Yew, A.T. and Puah, C.W (2010) Life cycle assessment of the Production of Crude Palm Kernel Oil (Part 3a). *Journal of Oil Palm Research* 22: 904-912.

Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomics and transcriptomic analyses. Nature Reviews Genetics 13(**4**): 227-232.

Wahid, M.B., Abdullah, S.N.A., and Henson, I.E (2004) Oil Palm – Achievements and Potential. *"New directions for a diverse planet", Proceeding of the 4th International Crop Science Congress.* Published on CDROM, www.cropscience.org.au.

Wang, W., Scali, M., Vignani, R., Spadafora, A., Sensi, E., Mazzuca, S., and Cresti, M. (2003) Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis* 24(**14**): 2369-2375.

Wang, W., Vignani, R., Scali, M. and Cresti, M. (2006) A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic

analysis. *Electrophoresis* 27(**13**): 2782-2786.

Wang, W., Tai, F., and Chen, S. (2008) Optimizing protein extraction from plant tissues for enhanced proteomics analysis. *Journal of Separation Science* 31: 2032-2039.

Wang, W., Bianchi, L., Scali, M., Liu, L., Bini, L., and Cresti, M. (2009) Proteomic analysis of β-1,3-glucanase in grape berry tissues. *ActaPhysiologiaePlantarum* 31(**3**): 597-604.

Wang, W.X., Li, K.L., Chen, Y., Lai, F.X. and Fu, Q (2015) Identification and function analysis of *enolase* Gene *NIEno1* from Nilaparvatalugens (Stâl) (Hemiptera: Delphacidae) *Journal of Insect Science* 15(**1**): 66.

Wang, W.X., Vinocur, B., Shoseyov, O. and Altman, A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trent in plant science* 9(**5**): 244-252.

Wann, S.R. and Goldner, W.R. (1994) Induction of somatic embryogenesis in *Taxus*, and the production of taxane-ring containing alkaloids thereform. **WO Patent** N° 93/19585.

Wessels-Boer, J.G. (1965) The indigenous palms of Surinam. Thesis, University of Leiden, Leiden.

Westermeier, R. (2005) Electrophoresis in Practice (4^{th}edn) Method 10 – IEF in immobilized pH gradients, pp 257-273. Wiley-VCH.

Westermeier, R., Friedmean, D.B., Hoving, S. (2009) Chapter thirty: Isoelectric Focusing and Two Dimensional gel Electrophoresis. *Methods in Enzymology* 463: 515-540.

Wilkins, M.R., Gasteiger, E. Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., William, K.L. (1996) Progress with proteome projects: why all protein expressed by a genome should be identified and how to do it. *Biotechnology andGenetic Engineering Review* 13: 19-50.

Winkelmann, T., Heintz, D., Van Dorsselaer, A., Serek, M and Braun, H.P. (2006) Proteomic analyses of somatic and zygotic embryos of cyclamen persicum Mill reveal new insight into seed and germination physiology. *Planta* August 224(**3**) : 508-519.

Wooi, K.C. (1990) Oil palm (*Elaeisguineesis* Jacq.): tissue culture and micropropagation. In: *Biotechnology in agriculture and forestry 10: Legumes and oilseed crops I* (Ed. by Y.P.S. Bajaj), pp 569-592, Springer, Berlin.

Wooi, K.C.(1995) Oil palm tissue culture - current practice and constraints. In: Rao V, Henson IE,Rajanaidu N. Eds. Recent developments in oil palm tissue culture and biotechnology. Malaysia, Malaysian Palm Oil Board, pp 21-32.

Wong, M.L and Medrano, J.F. (2005) Review: Real-time PCR for mRNA quantification.
Biotechniques 39(**1**):75-85.

Xi, J.H., Wang, X., Li, S.Y., Zhou, X., Yue, L., Fan, J. and Hao, D.Y (2006) Polyethylene glycol fractionation improved detection of low-abundant proteins by two-dimensional electrophoresis analysis of plant proteome. *Phytochemistry* 67(**21**): 2341-2348.

Zamzuri, I., MohdArif, S., Rajanaidu, N and Rohani, O. (1999) Commercial feasibility of clonal oil palm planting material production. PORIM Occasional paper No 40, pp 52.

Zhang, J.W., Ma, H.Q., Chen, S., Ji, M., Perl, A., Kovacs, L., and Chen, S.W (2009)
Stress response proteins‟ differential expression in embryogenic and non-embryogenic callus of Vitisvinifera L. cv. Cabernet Sauvignon- A proteomic approach. Plant Science 177(**2**):103-113.

Zuo. J., Niu, Q.W., Frugis, G. and Chua, N.H. (2002) The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant Journal* 30(**3**): 349-359.

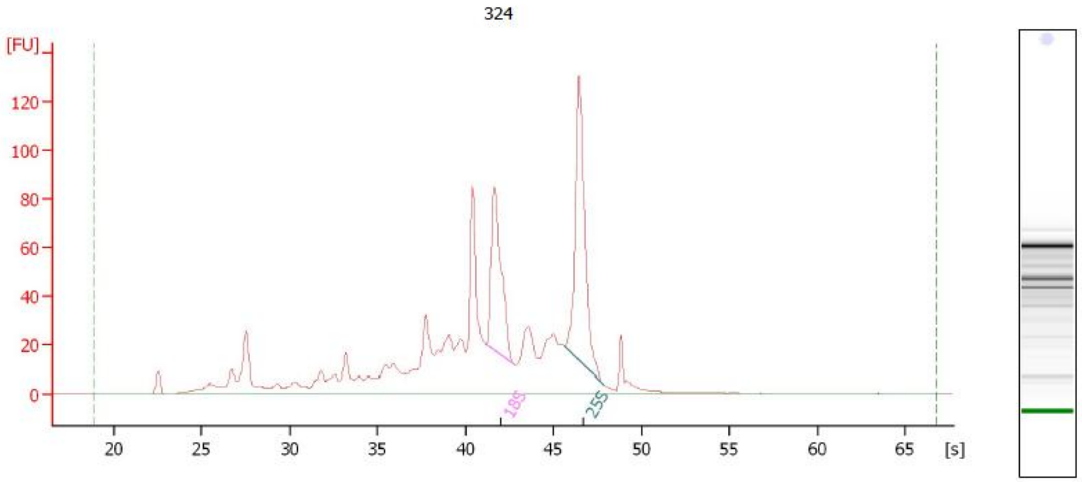
APPENDICES

APPENDIX 1: Preparation for the reagents for silver staining

Staining stage	Final Concentration
Fixative solution	Acetic acid 10% Methanol 40% Top up with distilled water
Sensitizer solution	Sodium thiosulphate (STS) 0.2% Sodium acetate 68 g/L
Silver solution	Silver nitrate 2.5 g/L
Developing solution (4°C)	Sodium carbonate 25 g/L Formaldehyde 0.4% Sodium thiosulphate (STS) 0.27%
Stop solution	Ethylenediamine- tetraacetic acid (EDTA) 14.6g/L

APPENDIX 2(a): Bioanalyzer results for old leaves

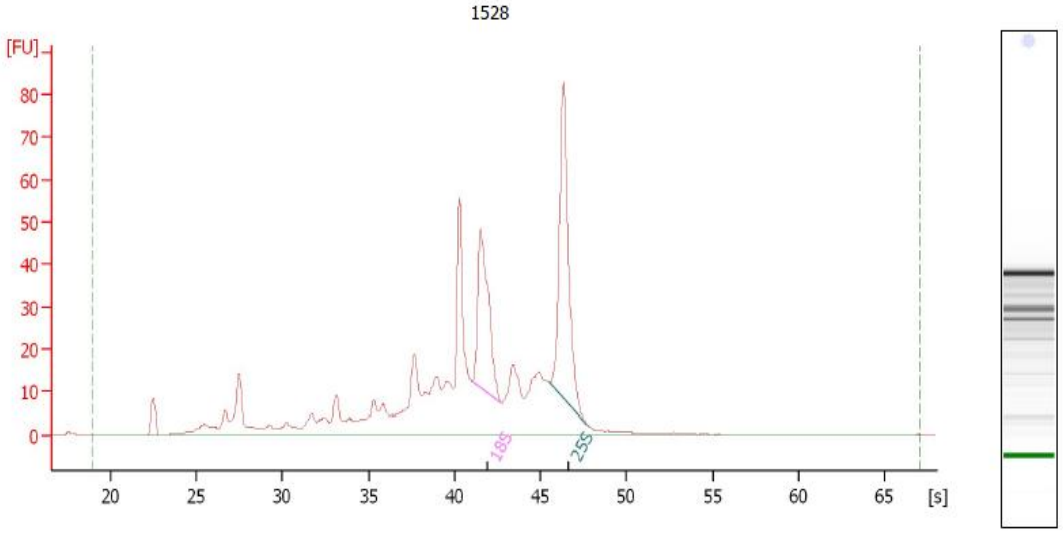
Sample 324



Overall Results for sample 5 : 324

RNA Area:	1,043.2	RNA Integrity Number (RIN):	7.1 (B.02.08)
RNA Concentration:	816 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	1.5	Result Flagging Label:	RIN: 7.10

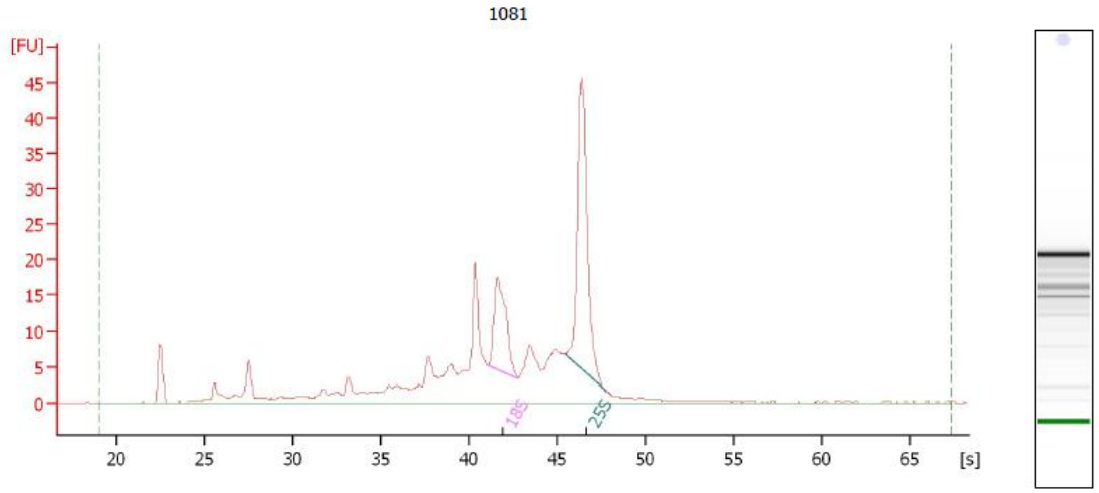
Sample 1528



Overall Results for sample 6 : 1528

RNA Area:	616.1	RNA Integrity Number (RIN):	7.3 (B.02.08)
RNA Concentration:	482 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	1.5	Result Flagging Label:	RIN: 7.30

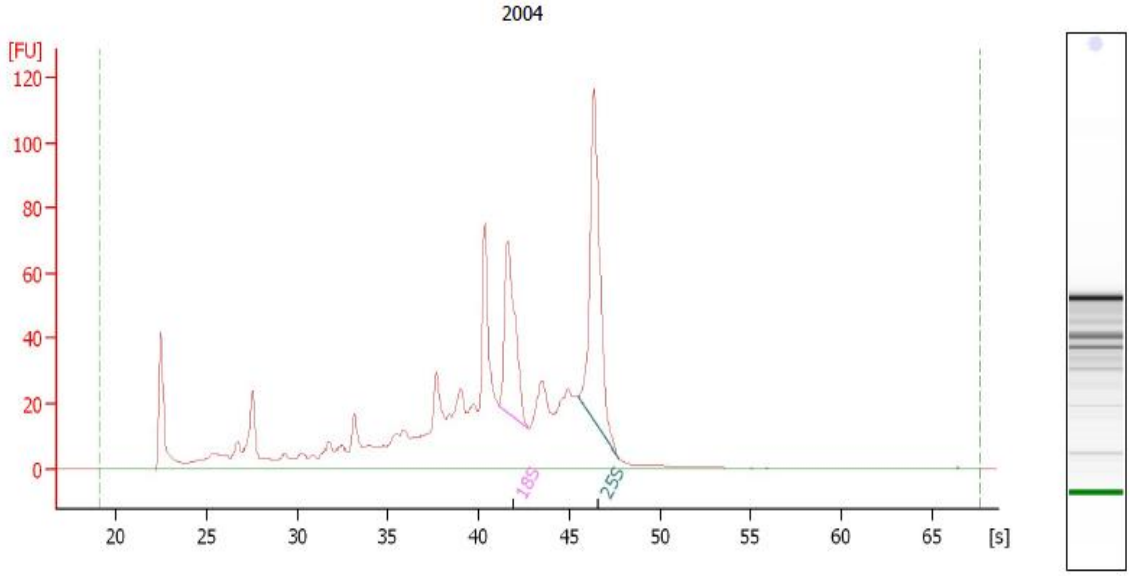
Sample 1081



Overall Results for sample 7 : 1081

RNA Area:	270.3	RNA Integrity Number (RIN):	7.5 (B.02.08)
RNA Concentration:	212 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	2.3	Result Flagging Label:	RIN: 7.50

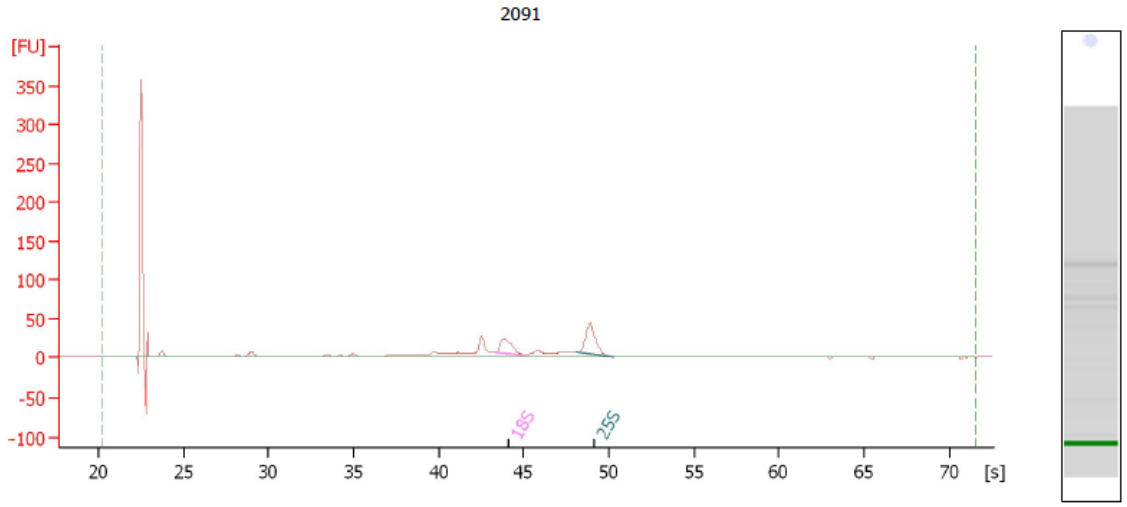
Sample 2004



Overall Results for sample 8 : 2004

RNA Area:	985.0	RNA Integrity Number (RIN):	7.1 (B.02.08)
RNA Concentration:	771 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	1.7	Result Flagging Label:	RIN: 7.10

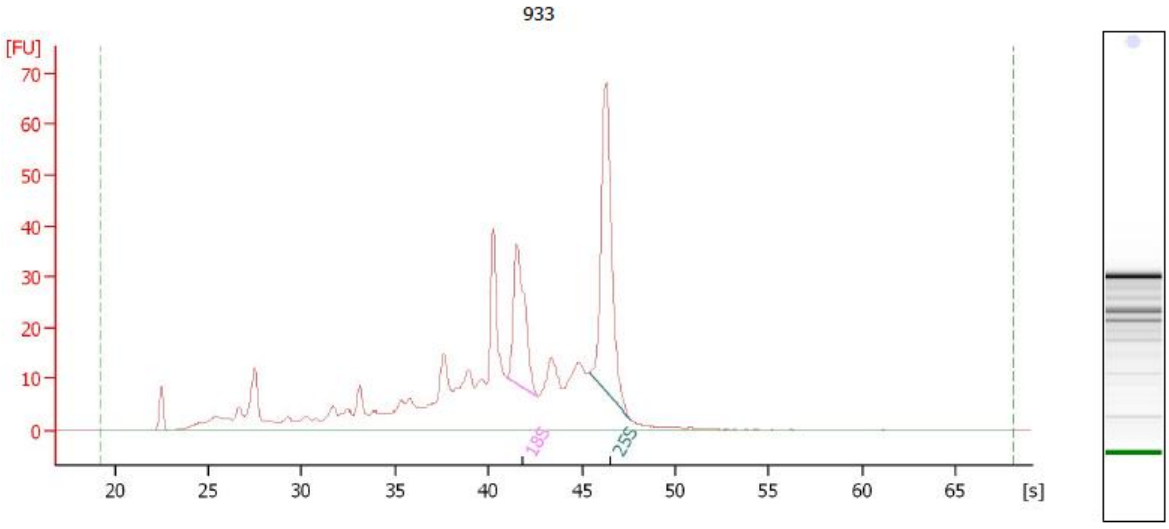
Sample 2091



Overall Results for sample 9 : 2091

RNA Area:	345.8	RNA Integrity Number (RIN):	7.1 (B.02.08)
RNA Concentration:	271 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	1.6	Result Flagging Label:	RIN: 7.10

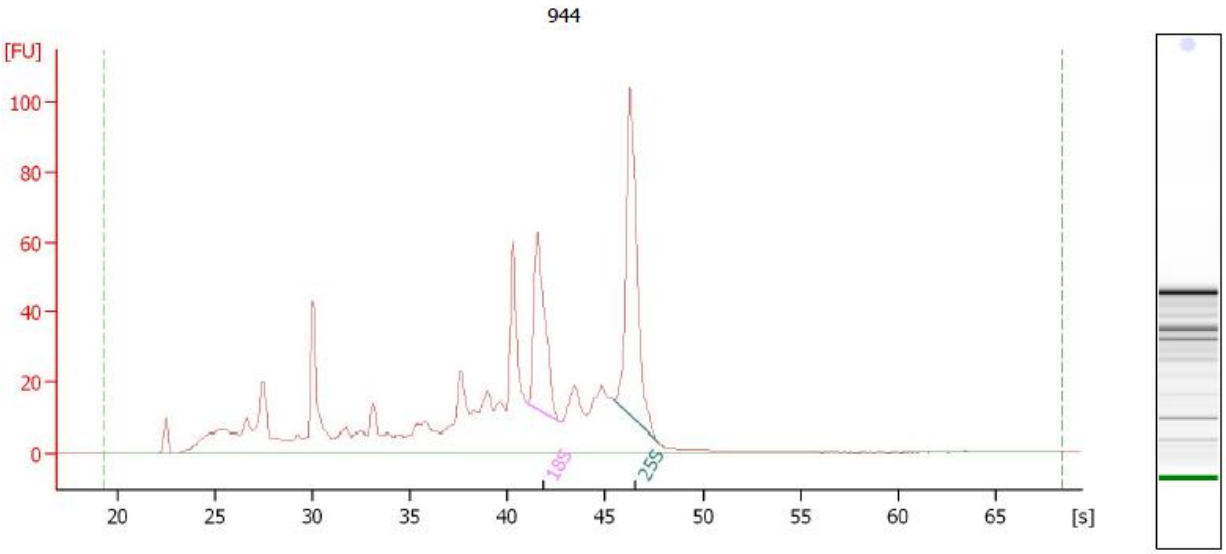
Sample 933



Overall Results for sample 10 : 933

RNA Area:	527.4	RNA Integrity Number (RIN):	7.2 (B.02.08)
RNA Concentration:	413 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	1.8	Result Flagging Label:	RIN: 7.20

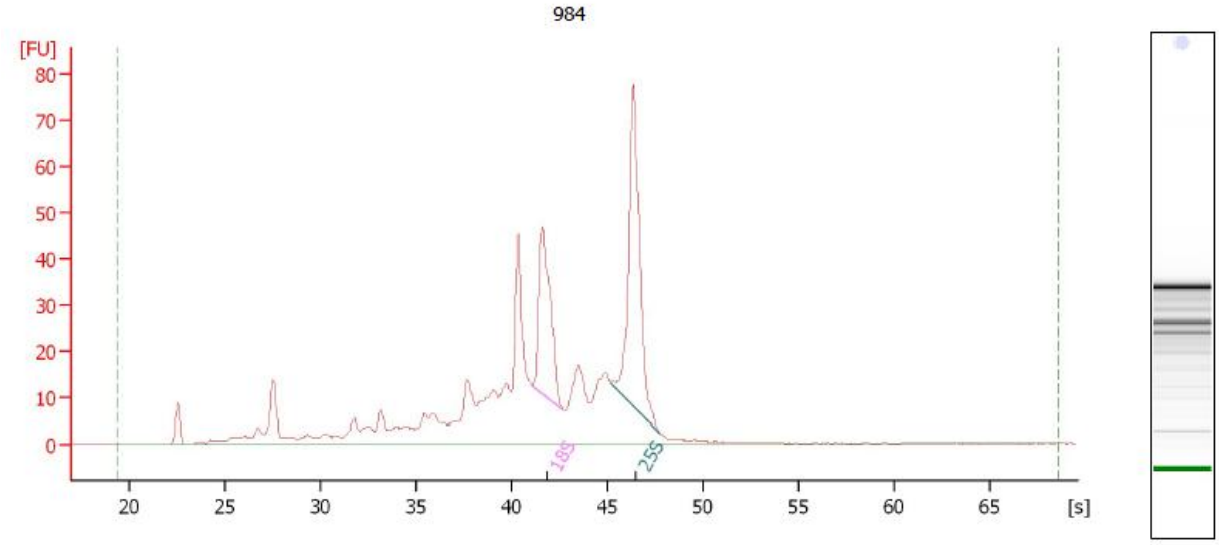
Sample 944



Overall Results for sample 11 : 944

RNA Area:	863.7	RNA Integrity Number (RIN):	7.1 (B.02.08)
RNA Concentration:	676 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	1.6	Result Flagging Label:	RIN: 7.10

Sample 984

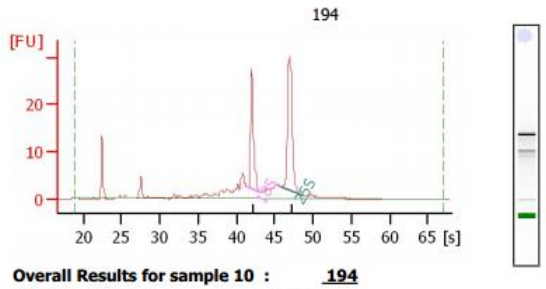


Overall Results for sample 12 : 984

RNA Area:	574.7	RNA Integrity Number (RIN):	7.4 (B.02.08)
RNA Concentration:	450 ng/µl	Result Flagging Color:	
rRNA Ratio [25s / 18s]:	1.6	Result Flagging Label:	RIN: 7.40

APPENDIX 2(b): Bioanalyzer results for young leaves

Sample 194



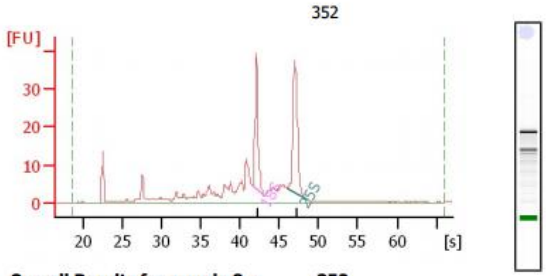
Overall Results for sample 10 : 194

RNA Area:	151.1
RNA Concentration:	108 ng/µl
rRNA Ratio [25s / 18s]:	1.6
RNA Integrity Number (RIN):	8.1 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 8.10

Fragment table for sample 10 : 194

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.55	43.01	27.3	18.1
25S	45.87	48.79	42.7	28.3

Sample 352



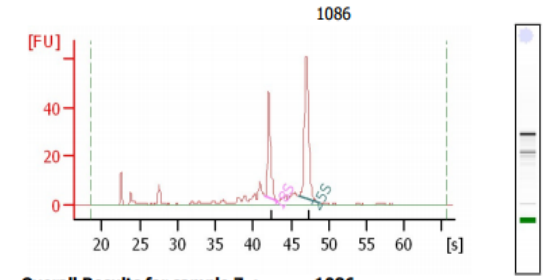
Overall Results for sample 8 : 352

RNA Area:	257.8
RNA Concentration:	184 ng/µl
rRNA Ratio [25s / 18s]:	1.4
RNA Integrity Number (RIN):	7 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN:7

Fragment table for sample 8 : 352

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.55	43.04	37.3	14.5
25S	46.01	48.54	50.4	19.6

Sample 1086



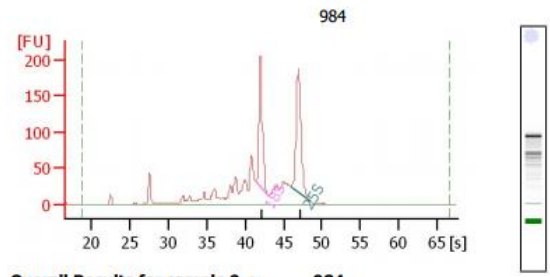
Overall Results for sample 7 : 1086

RNA Area:	259.8
RNA Concentration:	186 ng/µl
rRNA Ratio [25s / 18s]:	1.8
RNA Integrity Number (RIN):	8.6 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 8.60

Fragment table for sample 7 : 1086

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.53	43.24	45.8	17.6
25S	45.95	48.85	82.9	31.9

Sample 984



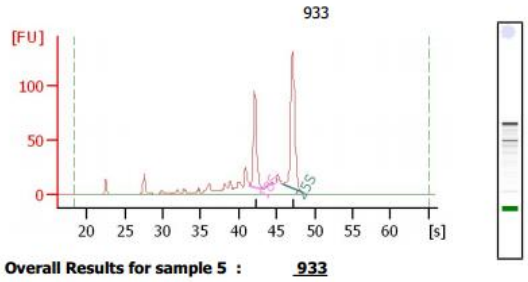
Overall Results for sample 9 : 984

RNA Area:	1,320.0
RNA Concentration:	944 ng/µl
rRNA Ratio [25s / 18s]:	1.3
RNA Integrity Number (RIN):	7 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN:7

Fragment table for sample 9 : 984

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.43	42.92	209.8	15.9
25S	45.92	48.52	267.2	20.2

Sample 933



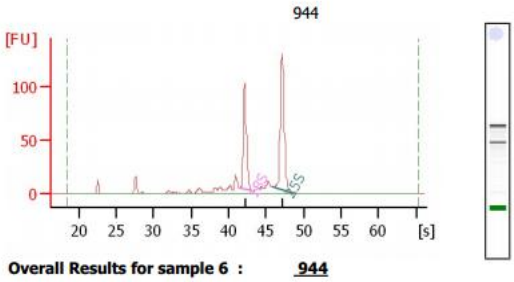
Overall Results for sample 5 : 933

RNA Area:	584.7
RNA Concentration:	418 ng/µl
rRNA Ratio [25s / 18s]:	1.8
RNA Integrity Number (RIN):	8.1 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 8.10

Fragment table for sample 5 : 933

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.46	43.16	97.3	16.6
25S	45.85	48.49	171.6	29.3

Sample 944



Overall Results for sample 6 : 944

RNA Area:	487.7
RNA Concentration:	349 ng/µl
rRNA Ratio [25s / 18s]:	1.6
RNA Integrity Number (RIN):	8.8 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 8.80

Fragment table for sample 6 : 944

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.49	43.15	108.0	22.1
25S	45.85	48.65	169.3	34.7

APPENDIX 3(a): Statistical test for the RT-qPCR in Old leaves

General Linear Model: 3305, 4202, ... versus Proliferation

```
Factor         Type   Levels  Values
Proliferation  fixed       2  H, L
```

3305

Analysis of Variance for 3305, using Adjusted SS for Tests

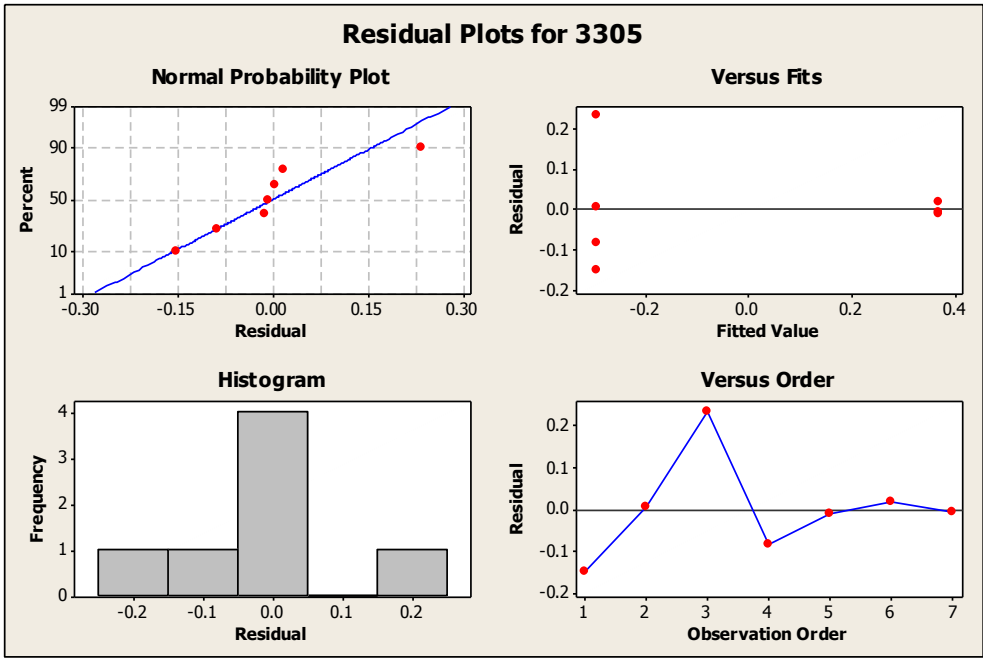
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Proliferation	1	0.75527	0.75527	0.75527	44.01	0.001
Error	5	0.08581	0.08581	0.01716		
Total	6	0.84107				

S = 0.131001 R-Sq = 89.80% R-Sq(adj) = 87.76%

Grouping Information Using Tukey Method and 95.0% Confidence for 3305

Proliferation	N	Mean	Grouping
H	3	0.370	A
L	4	-0.294	B

Means that do not share a letter are significantly different.



7304

Analysis of Variance for 7304, using Adjusted SS for Tests

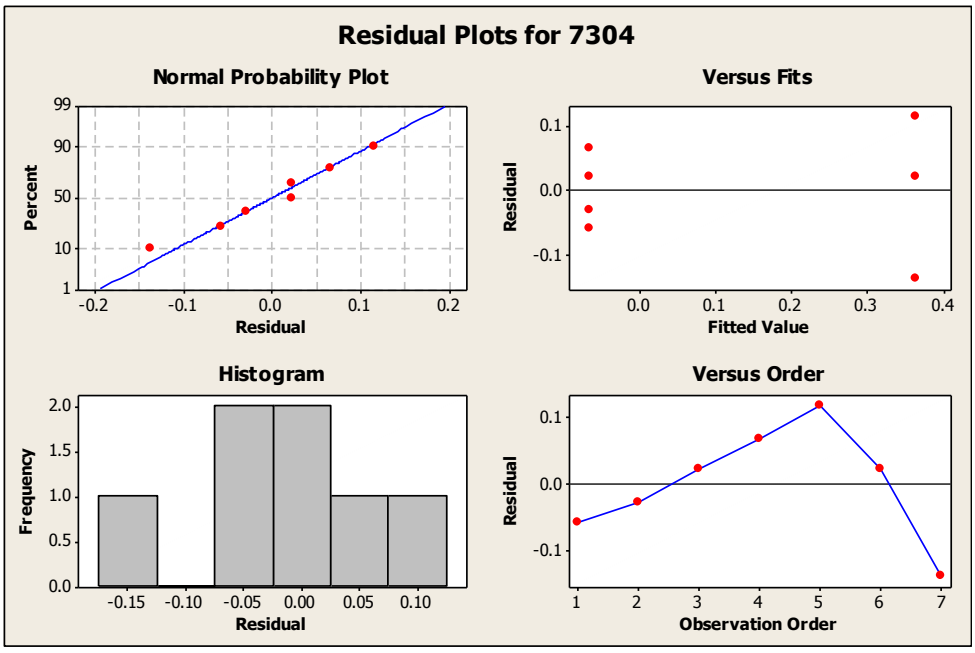
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Proliferation	1	0.31804	0.31804	0.31804	37.67	0.002
Error	5	0.04222	0.04222	0.00844		
Total	6	0.36026				

S = 0.0918860 R-Sq = 88.28% R-Sq(adj) = 85.94%

Grouping Information Using Tukey Method and 95.0% Confidence for 7304

Proliferation	N	Mean	Grouping
H	3	0.364	A
L	4	-0.066	B

Means that do not share a letter are significantly different.



7007

Analysis of Variance for 7007, using Adjusted SS for Tests

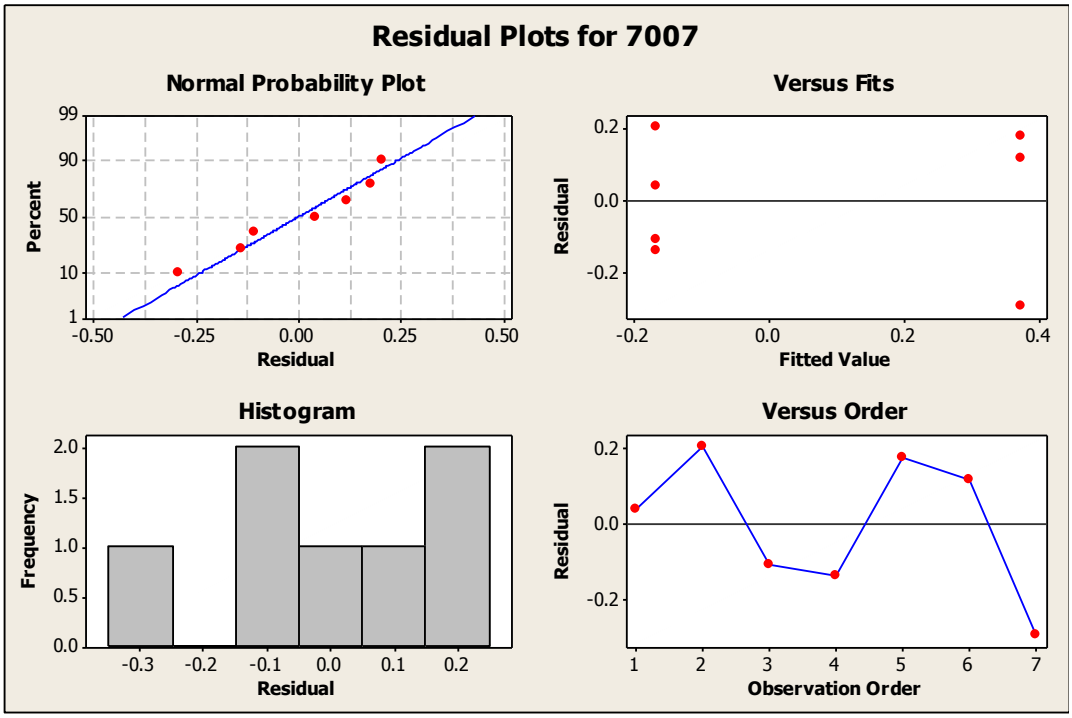
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Proliferation	1	0.50698	0.50698	0.50698	12.28	0.017
Error	5	0.20646	0.20646	0.04129		
Total	6	0.71344				

S = 0.203204 R-Sq = 71.06% R-Sq(adj) = 65.27%

Grouping Information Using Tukey Method and 95.0% Confidence for 7007

Proliferation	N	Mean	Grouping
H	3	0.376	A
L	4	-0.168	B

Means that do not share a letter are significantly different.



APPENDIX 3(b): Statistical test for the RT-qPCR in Young leaves

One-way ANOVA: 12198

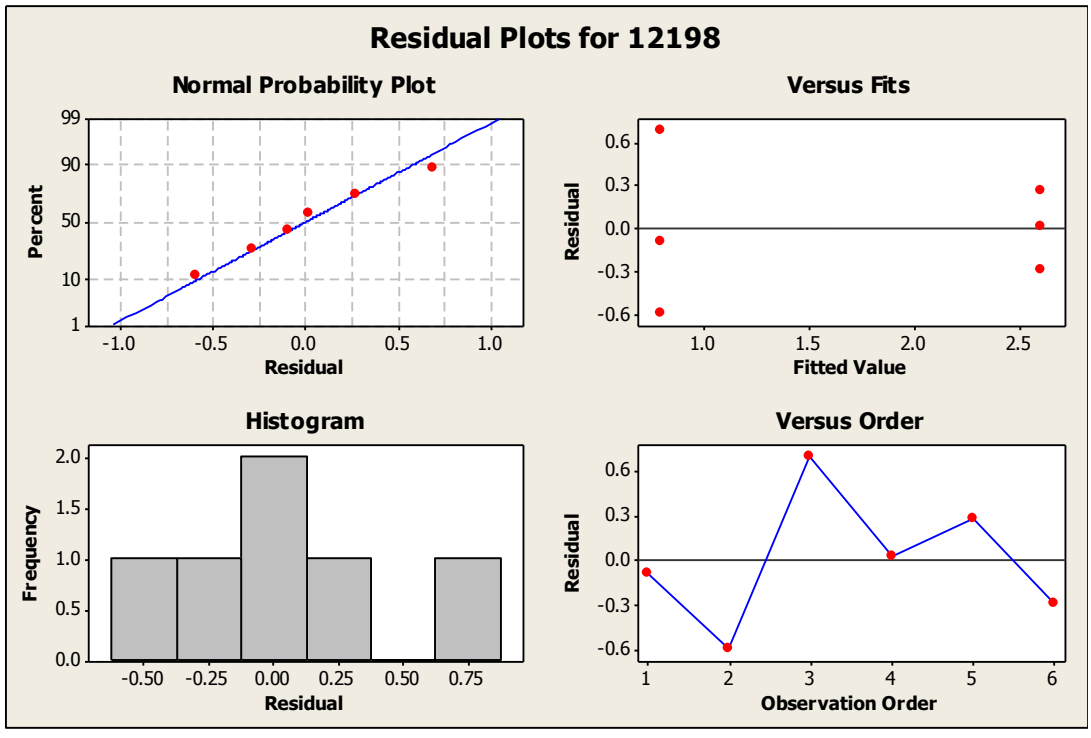
Source	DF	SS	MS	F	P
C1	1	4.876	4.876	19.39	0.012
Error	4	1.006	0.252		
Total	5	5.882			

S = 0.5015 R-Sq = 82.90% R-Sq(adj) = 78.62%

Grouping Information Using Tukey Method

C1	N	Mean	Grouping
L	3	2.5995	A
H	3	0.7965	B

Means that do not share a letter are significantly different.



One-way ANOVA: 51068

```
Source  DF      SS      MS       F       P
C1       1  4.252   4.252   28.23   0.006
Error    4  0.603   0.151
Total    5  4.855

S = 0.3881   R-Sq = 87.59%   R-Sq(adj) = 84.49%

Grouping Information Using Tukey Method

C1  N    Mean  Grouping
L   3  1.8993  A
H   3  0.2155     B

Means that do not share a letter are significantly different.
```