IMPERIAL COLLEGE LONDON
FACULTY OF NATURAL SCIENCES
DIVISION OF CELL AND MOLECULAR BIOLOGY

SYSTEMIC SIGNALING IN RELATION TO REGULATION OF FLOWERING TIME

ZAIDAH RAHMAT

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
NOVEMBER 2012

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work, except as acknowledged in the text, and that the material presented here has not been submitted, in whole or part, for a degree at that or any other university.

Zaidah Rahmat

Abstract

In *Arabidopsis thaliana*, phloem proteome study is gaining recognition. An optimised EDTA exudation technique from excised leaves was incorporated in this study to formulate a method of obtaining sufficient phloem exudate protein. Total phloem exudate protein and excised gel bands from 1D SDS-PAGE yielded 221 identified proteins via analysis by LC-MS/MS. The majority of the proteins identified have functions in metabolic processes, stress response and defense mechanism. These may have been induced by wounding nevertheless many of these proteins may have been present within the companion cell and sieve element. Similar findings were reported in earlier *Arabidopsis* and other plant phloem proteome studies. A comprehensive list of *Arabidopsis* phloem exudate protein from three studies is presented here. Although the florigen protein FT was not identified in the main study, a different approach using MRM proved the presence of FT in the phloem exudate of *Arabidopsis*. An attempt to quantify FT was performed using MRM to test out if varying level of FT protein in response to photoperiod is produced in accordance to its mRNA expression as previously reported. Level of FT protein detected showed a pattern of expression different from *FT* mRNA level of expression. In an effort to detect long-distance movement of FT protein, transgenic lines carrying FT fused to Dendra2 and mediated by SUC2 and GAS1 promoters were generated. In spite of evidence of transcription of transgene, this experiment was unsuccessful with only weak expression detected in a few lines. Further validation could not be made due to time constraint. In summary, the protein data obtained from this study and the comprehensive list of compiled proteins serve as the basis of an *Arabidopsis* phloem proteome, and showed that quantification of level of FT protein in response to photoperiod is possible.

Acknowledgments

I would like to extend my foremost gratitude to Dr. Colin Turnbull for his guidance, assistance, supervision, advices and patience while I was conducting my research under his lab and throughout my writing period. Special thanks to Mr. Mark Bennett for his help with running all my LC-MS/MS and MRM runs, and the information provided on those topics. A big thank you goes to Coupland's Lab (Germany) for providing me with the initial FT constructs. In addition I would like to thank the late Dr. Judit Nagy who helped us with 2D electrophoresis work and letting us work in her lab.

To lab members; David, thanks for letting me use your protein for standard and all the useful discussions we had, and Sadia, thanks for the times we had inside and outside of the lab. To all friends from IC regardless of your nationalities (particularly Nyuk Ling, Roslinda, Huda, Natalia, Miao, Nan, Nico, Maija, Issariya, Vin, Joe and Naqi), I am very grateful to have all of your companies during my stay and thanks for being such wonderful friends.

My study was sponsored by Ministry of Higher Education Malaysia and Universiti Teknologi Malaysia, and I would like to thank them for giving me this opportunity.

Very special thanks to my family; mak, kak Dewi & kak Nor for all your support, prayers and encouragements. This dissertation is dedicated to my late father who did not have the chance to see his daughter's success; your memories live on and thank you for being the best father a daughter could ever ask for (Al Fatihah). *Alhamdulillah, kejayaan ini untuk kita sekeluarga*. And finally to Adam, thank you for everything; janji telah ditunaikan.

Table of Contents

Declaration ……..2

Abstract ……...3

Acknowledgments ……...4

Table of Contents ……...5

List of Figures ……...11

List of Tables ……14

Abbreviations ……...15

Chapter 1 Introduction.....18

1.1 Flowering ……..18

1.2 Physiological Control of Flowering Time ……18

1.3 Genetics of Flowering Time ……..19

1.4 FLOWERING LOCUS T (FT) – The Florigen Involved in Regulation of Flowering ……..……22

1.5 Phloem and Translocation ……27

1.5.1 Phloem Loading and Unloading ……..28

1.5.2 Sucrose as Transporter in the Phloem ……..……....30

1.2.3 Callose Accumulation and Occlusion ……...31

1.6 Exudation Technique ……32

1.7 Plant Proteomics ……34

1.8 The Phleom Proteome ……..35

1.9 Polyacrylamide Gel Electrophoresis (PAGE) ……...36

1.10 Two Dimensional (2D) Gel Electrophoresis ……36

1.10.1 Sample Preparation and Protein Solubilisation for 2D Gel Electrophoresis ……37

1.10.2 The First-Dimension, Second Dimension, Protein Detection on Gel and Spot Identification ……39

1.11 Mass Spectrometry-Based Proteomics ……41

1.12 Multiple Reaction Monitoring (MRM) ……43

1.13 Photoactivatable Fluorescent Proteins (PAFP) ……44

1.14 Dendra2 ……46
1.15 Aims of Thesis ……..47

Chapter 2 *Arabidopsis* Phloem Proteomics - Method Development.....50

2.1 Introduction 50
2.2 Materials and Methods 51
2.2.1 Mutant Screening 51
2.2.1.1 Plant Growth and Conditions 51
2.2.1.2 Phenotypic Observation 51
2.2.1.3 Genotyping via PCR 52
2.2.1.3.1 PCR 52
2.2.1.3.2 RNA Extraction 53
2.2.1.3.3 Synthesis of First Strand cDNA 54
2.2.1.3.4 rt-PCR Analysis 54
2.2.2 1D Proteomics 55
2.2.2.1 Exudation Technique 55
2.2.2.2 Protein Purification 56
2.2.2.2.1 Acetone Precipitation 56
2.2.2.2.2 Purification using Millipore Filter Centrifuge 56
2.2.2.3 Bradford Assay 56
2.2.2.4 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE).....57
2.2.2.5 Trypsin Digestion 57
2.2.2.5.1 In-Gel Digestion 57
2.2.2.5.1.1 In-Gel Digestion of Coomassie Stained Gel 57
2.2.2.5.1.2 In-Gel Digestion of Silver Stained Gel 58
2.2.2.5.2 In-Solution Digestion 59
2.2.2.6 Mass Spectrometry 60
2.2.3 2D Proteomics 60
2.2.3.1 Protein Extraction 60
2.2.3.2 2D Clean-Up 61
2.2.3.3 Isoelectric Fcusing (IEF) and 2-Dimensional Gel Run 61
2.3 Results 62
2.3.1 Mutant Screening – Phenotypic Analysis 62
2.3.2 Mutant Screening – Genotyping 66

2.3.3 Optimisation and Establishing Exudation Technique 68
2.3.4 Optimisation of 2D Gel Electrophoresis for *Arabidopsis* Phloem Sap Protein 72
2.4 Discussion 73
2.4.1 Mutant Screening 73
2.4.2 Effect of EDTA on Exudation 75
2.4.3 Correlation between Photoperiod and Leaf Morphology and Its Effect on Yield of Exudates 77
2.4.4 Plant Condition and Developmental Stage May Affect Protein Content of Phloem Sap 79
2.4.5 Factors Affecting the Reproducibility of 2D Gel Electrophoresis of *Arabidopsis* Phloem Proteome 80

Chapter 3 *Arabidopsis* Phloem Proteome.....85
3.1 Introduction 85
3.2 Materials and Methods 86
3.2.1 Phloem Sap Collection via Exudation Technique 86
3.2.2 Trypsin Digestion 86
3.2.2.1 In-Gel Digestion 86
3.2.2.1.1 In-Gel Digestion of Coomassie Stained Gel 86
3.2.2.1.2 In-Gel Digestion of Silver Stained Gel 87
3.2.2.2 In-Solution Digestion 88
3.2.3 Mass Spectometry 88
3.2.4 Multiple Reaction Monitoring (MRM) 89
3.2.5 Database Searches 90
3.2.6 2D Proteomics 90
3.3 Results 91
3.3.1 General Comparison of Proteins Identified Across Different Genotypes 91
3.3.2 Comparison between Different Treatments 94
3.3.3 Data Analysis by MASCOT and ProteinPilot 95
3.3.4 Identification of Unique Proteins 114
3.3.5 Functional Categories of Identified Proteins 118
3.3.6 Validity of Proteins 123

3.3.7 Detection of FT via Multiple Reaction Monitoring (MRM) 126
3.3.8 Level of FT in Plant under Photoperiod Treatment 131
3.4 Discussion 133
3.4.1 Data Analysis and Identification of Proteins from Phloem Exudate...133
3.4.2 Number of Proteins Identified Differ According to Source of Protein and Treatments 142
3.4.3 Phloem Exudate Protein Content from Various Genotypes 147
3.4.4 *Arabidopsis* Phloem Exudate Protein and Possible / Predicted Functions of the 16 Homologous Proteins Across Four Phloem Proteomes ...149
3.4.4.1 Proteins Representing Basic Cellular Component in the Phloem 152
3.4.4.2 Proteins Involved in Maintaining the Biochemistry of Phloem 153
3.4.4.3 Stress Responses Proteins 155
3.4.4.4 Defense Proteins 157
3.4.4.5 Protein-Folding and Destruction 159
3.4.4.6 Major Regulator of Cell Growth 160
3.4.4.7 Protein Catabolic Process 160
3.4.5 Proteins from the Phloem Exudate Have Diverse Function in Response to Leaf Wounding and Photoperiod 161
3.4.5.1 Redox Regulation, Anti-Oxidant Response and Amino Acid Metabolism 162
3.4.5.2 Signal Transduction 163
3.4.5.3 Metabolism 164
3.4.5.4 Component of Cell Structure 165
3.4.5.5 Protein Transport 165
3.4.5.6 Nucleic Acid Binding 165
3.4.5.7 Signaling Proteins 166
3.4.6 FT Was Detected Using MRM But Was Not Found Using Standard LC-MS/MS 167
3.4.7 Quantitative Analysis of FT in Relation to Photoperiod is Possible via MRM 168
3.4.8 Reproducibility and Variation in Proteomics Data 170

Chapter 4 Expression of FT Fusion protein Using Dendra2.....173

4.1 Introduction 173
4.2 Materials and Methods 174
4.2.1 Fusion Protien Construction 174
4.2.1.1 Preparation of Promoters and Fluorescent Protein 174
4.2.1.2 Preparation of GV3101 Electrocompetent Cell 176
4.2.1.3 Ligation, Transformation, Selection and Confirmation Via PCR 176
4.2.2 Transformation via Floral Dip 177
4.2.3 Selection of Transformants using Basta 177
4.2.4 Verification of Fusion Protien via PCR 178
4.2.4.1 Genomic DNA Extraction 178
4.2.4.2 RNA Extraction 178
4.2.4.3 Polymerase Chain Reaction (PCR) and Reverse Transcriptase PCR (rt-PCR) 179
4.2.5 Phenotypic Observation 179
4.2.6 Tracking Protein Movement of FT/Dendra2 Fusion Construct 180
4.3 Results 180
4.3.1 Construction of FT:Dendra2 Fusion Proteins 180
4.3.2 Transformation of Fusion Proteins via Floral Dip and Analysis of Fusion Proteins in *Arabidopsis* T_1 Plants 182
4.3.3 Phenotypic Observations of Plants Expressing Fusion Protein Genes 186
4.3.4 Expression Analysis of SUC2::FT:Dendra2 in Wild Type and *ft-10* Plants 189
4.4 Discussion 192
4.4.1 Comparison of Flowering Time and Transcript Level of Expression of SUC2::FT:Dendra2 and GAS1::FT:Dendra2 Lines 192
4.4.2 Could FT and Its Movement Be Tracked Using Dendra2 Fluorescent Protein? 195
4.4.3 Stability of Dendra2 in Fusion Protein to Detect and Track Movement of FT 197

Chapter 5 General Discussion.....200

5.1 General Discussion 200

5.2 Establishing a Phloem Proteome Database for *Arabidopsis thaliana* 204

5.3 Future Work 208

Bibliography.....212

Appendix A (List of Identified Proteins with ESI peptides, scores, frequency found) 241

Appendix B (List of Identified Proteins with assigned biological functions) 271

Appendix C-1 (Overview of GO Term Enrichment) 281

Appendix C-2 (GO term enrichment for Biological Process category) 282

Appendix C-3 (GO term enrichment for Biological Process category spread out over two pages with the first focusing on metabolic process) 283

Appendix C-4 (GO term enrichment for Biological Process category spread out over two pages which covers cellular processes and was linked to stress-related and defense gene (Appendix C-3)) 284

Appendix C-5 (GO term enrichment for Molecular Function category) 285

Appendix C-6 (GO term enrichment for Cellular Component category) 286

Appendix D (Complete list of proteins identified in three *Arabidopsis* studies) 287

Appendix E Result of one way ANOVA performed based on leaf number differences in transgenic lines grown in LD (left) and SD (right) 299

Appendix F (2D Proteomics gels) 300

Appendix G (List of primers used in Chapter 4) 301

List of Figures

Figure 1.1: Control of flowering time in *Arabidopsis* involves 4 pathways that control the expression of floral integrator genes and proteins that allow plant in vegetative stage to produce flower 21
Figure 1.2: The florigen theory 23
Figure 1.3: Photoperiod pathway in *Arabidopsis* plants under LD (long day) to induce flowering 26
Figure 1.4: Primary pathway into and out of phloem 29
Figure 1.5: Schematic diagram of a mixture of proteins before IEF and how they are then separated based on their pI after focusing 40
Figure 1.6: Separation of proteins by SDS-PAGE after separation by IEF 41
Figure 2.1: Mature and healthy leaves of 6 weeks old plants were used for exudation where (a) it is cut at the base of the petiole and (b) up to 7 leaves were located into 1.5 mL tube filled with 10 mM Na_2-EDTA 55
Figure 2.2: Wild type and mutant plants grown in SD and LD 63
Figure 2.3: Mutant rosette leaf count for all mutants and wild type in the initial screening plants grown in LD 64
Figure 2.4: Flowering time of mutants based on total leaves formed by the primary meristem of plants grown under LD and SD 65
Figure 2.5: PCR gel to detect T-DNA insertion in *Arabidopsis* mutant for homozygous lines 67
Figure 2.6: RT-PCR products on 1.2% agarose gel 68
Figure 2.7: Exudation technique of *Arabidopsis* phloem sap protein was optimised under different parameters 70
Figure 2.8: Purification of protein was carried out using acetone precipitation (left) and Amicon filter (right) where it can be seen from the SDS-PAGE that much of the samples were lost in the former method 71
Figure 2.9: Phloem sap protein comparison from the first exudates collected right after wounding, the second exudates collected 30 minutes after wounding and overnight respectively 71
Figure 2.10: 2D gel electrophoresis of *Arabidopsi*s phloem 72
Figure 3.1: Phloem protein obtained from five different genotype 91
Figure 3.2: How gel bands were selected and categorised for LC-MS/MS analysis

.....92
Figure 3.3: Concentration of protein from exudates collected 30 min after wounding, overnight from the same leaves and the pellet of both exudates after ultra-centrifugation was measured using Bradford Assay, and compared against the standard curveperformed on BSA 95
Figure 3.4: Flow diagram of selection of identified proteins following LC-MS/MS analysis 98
Figure 3.5: Comparisons of identified phloem exudate proteins are represented as Venn diagrams 113
Figure 3.6: GO analysis of identified proteins from *Arabidopsis* specifying Biological Process 119
Figure 3.7: GO analysis of identified proteins from *Arabidopsis* specifying Molecular Function 120
Figure 3.8: GO analysis of identified proteins from *Arabidopsis* specifying Cellular Component 121
Figure 3.9: Normed frequencies of functional categories from MapMan classification for the identified proteins 123
Figure 3.10: Comparison of phloem sap proteins from Brassica napus on SDS-PAGE between pure phloem sap obtained using needle wounding method and EDTA exudation technique 124
Figure 3.11: MRM was used to detect FT in the standard, His-FT using three chosen peptides where (a) all peptides were detected in the sample 127
Figure 3.12: MS spectra of FT after MRM analysis using the peptide VEIGGEDLR 129
Figure 3.13: Quantification of FT protein via MRM was conducted under photoperiod Induction based on the highest peak area count in any of the three peptides 133
Figure 3.14: Proteolytic mechanism of trypsin that cleaves the C-terminal to arginine or lysine 143
Figure 4.1: Map of Gateway® Dendra2-At-N used in the study 175
Figure 4.2: Construction of the five fusion proteins for FT accumulation study ...181
Figure 4.3: Basta-treated transformed plants after the last spray 183
Figure 4.4: Presence of fused transgene were confirmed using PCR to amplify out (a) Dendra2 fluorescent protein, (b) SUC2 promoter, (c) GAS1 promoter, (d)

FTcDNA region and (e) pGreen vector from genomic DNA of each lines and controls 185
Figure 4.5: Flowering time of T_2 plants by rosette leaf count for plants both in LD (above) and SD (below) 187
Figure 4.6: Flowering time of T_2 plants were determined by rosette leaf count for plants both in LD and SD for (a) SUC2::FT:Dendra2 line and (b) GAS1::FT:Dendra2 line 188
Figure 4.7: Further verification of T_2 plants carrying the full construct was done on genomic DNA using primer pairs to confirm the presence of (a) Dendra2, (b) SUC2 promoter, (c) GAS1 promoter, (d) FTcDNA and (e) M13 ..189
Figure 4.8: Schematic representation of the location where the primers for the promoters were designed on (a) SUC2::FT:Dendra2 and (b) GAS1::FT:GFP 190
Figure 4.9: rt-PCR carried out on T_2 plants to determine transcription level of (a) Dendra2, (b) FT:Dendra2 in lines mediated by SUC2, (c) FT:Dendra2 in lines mediated by GAS1 and (d) FT in the fusion proteins 191
Figure 4.10: Preliminary analysis of FT:Dendra expression 192

List of Tables

Table 2.1: List of primers used for PCR reaction 53
Table 2.2: List of primers used for rt-PCR reaction 54
Table 2.3: Bonferroni's posttests performed on the four mutants and WT following two way ANOVA analysis 66
Table 3.1: Average number of proteins identified from excised gel bands according to sections and size 93
Table 3.2: Complete list of identified proteins from *Arabidopsis* phloem sap detected in various samples treatments and genotypes 99
Table 3.3: List of identified protein unique to the current study 115
Table 3.4: A list of proteins identified via LC-MS/MS from *Brassica napus* exudates collected using EDTA exudation technique 125
Table 3.5: Result of MRM carried out on the standard and other samples while setting up the procedure to detect FT 128
Table 3.6: Summary of MRM carried out on various treatments for detection of FT ..130
Table 3.7: Summary of quantified FT in the standard (His-FT) and the 12 samples based On the main peptide VVGDVLPFNR 132
Table 3.8: Identified proteins from excised gel band based on gel sections 136
Table 3.9: List of 16 proteins conserved within *Arabidopsis*, brassica, cucurbit and rice 151
Table 4.1: List of adaptor-primer used for construction of fusion protein 175
Table 4.2: Analysis of T_1 plants via PCR to determine incorporation of transgene.....184

Abbreviations

µ	Micro
^{o}C	Degree Celcius
x g	Gravity force
>	More than
<	Less than
1D	One dimensional
2D	Two dimensional
ACN	Acetonitrile
AGI	*Arabidopsis* gene index
ANOVA	Analysis of variance
AP1	APETAALA 1
APS	Ammonium persulfate
ATP	Adenosine triphosphate
bp	Base pair
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
Ca^{2+}	Calcium ion
CAD	collisionally activated dissociation
CC	Companion cell
cDNA	Complementary Deoxyribonucleic acid
CH_3CN	Acetonitrile
CO	CONSTANS
Col-0	Columbia
CXP	Collision cell exit potential
Da	Dalton
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleic triphosphate
DP	Dual-polarity
DTT	Dithiotreitol
EDTA	Ethylenediaminetetraacetic acid
e.g.	Example
EGTA	ethylene-glycol-bis-(b-aminoethyl-ether) N, N9-tetraactetate
EP	Exit potential
ESI	Electrospray ionisation
EST	Expressed sequence tag
etc	Etcetera
EtOH	Ethanol
$FeK_3(CN)_6$	Iron potassium cyanide
fm	FLORAL MERISTEM

FP	Fluorescent protein
FT	FLOWERING LOCUS T
FTL2	FLOWERING LOCUS T-LIKE2
ft-tsf	*Flowering locus T-twin sister of flowering locus T* mutant
g	Gram
GA	Gibberrelic acid
GFP	Green fluorescent protein
GO	Gene ontology
GPM	Global Proteome Machine
h	Hour
HCl	Hydrochloric
IEF	Isoelectric focusing
IDA	Information dependent acquisition
im	Inflorescence meristem
IPG	Immobilised pH gradient strips
iTRAQ	Isobaric tags for relative and absolute quantitation
k	Kilo
K_2-EDTA	Potassium- ethylenediaminetetraacetic acid
L	Litre
LB	Left border
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Tandem liquid chromatography-mass spectrometry
LD	Long day
LP	Left primer
M	Molar
MeOH	Methanol
MIAPE	Minimum information about a proteomics experiment
min	Minute
mg	Milligram
$MgCl_2$	Magnesium chloride
mL	Millilitre
mm	millimetre
mM	Millimolar
mRNA	Messenger ribonucleic acid
MRM	Multiple reaction monitoring
MS	Mass spectrometer
MW	Molecular weight
m/z	Mass to charge ratio
Na_2-EDTA	Sodium-Ethylenediaminetetraacetic acid
$Na_2S_2O_3$	Sodium thiosulfate
NCBI	National centre for biotechnology information

NH_4HCO_3	Ammonium hydrogen carbonate
nm	Nano meter
PAFP	Photoactivatable fluorescent protein
PA-GFP	Photoactivatable green fluorescent protein
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pI	Isoelectric point
PR	Pathogenesis-related
PS-CFP	Photoswitchable cyan fluorescent protein
psi	Pressure unit
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RP	Right primer
rpm	Rotation per minute
RT	Room temperature
rt-PCR	Reverse-transcription PCR
SAIL	Syngenta *Arabidopsis* Insertion Library
SAM	Shoot apical meristem
SD	Short day
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-Polyacrylamide gel electrophoresis
SE	Sieve element
SEL	Size exclusion limit
SEM	Standard error mean
sec	Second
SOC1	SUPPRESSOR OFOVEREXPRESSION OF CONSTANS 1
SRM	Single reaction monitoring
TAIR	The *Arabidopsis* information resource
TBP	Tributyl phosphine
TCTP	Translationally-controlled tumor protein
TEM	Temperature
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
TSF	TWIN SISTER OF FLOWERING LOCUS T
U	Unit
UDP	Uridine diphosphate
UV	ultraviolet
v	Volume
V	Voltage
w	Weight
WT	Wild type

CHAPTER 1
GENERAL INTRODUCTION

1.1 Flowering

Flowering plants are known as angiosperms and gymnosperms. Both are vascular seed-bearing plants. In many plants, flowering often precedes reproduction as pollen and ovules are contained within the flowers. Flowering plants go through two phases; 1) vegetative growth phase to produce more stems and leaves, and, 2) flowering phase for the production of floral organs followed by production of fruit and/or seed (Glover, 2007).

The change from vegetative to reproductive phase is affected by internal and external factors (Erwin, 2006). The conversion of the apical meristem into a floral meristem involves signals within the plants to produce all parts of the flower. Flowering is also influenced by stimulation from internal (biotic factors such as plant's size, age) and external factors (abiotic factor such as day length, temperature, water) (Erwin, 2006). Abiotic factors such as temperature (Franks et al., 2007) and photoperiod (Erwin, 2006; Galloway and Burgess, 2012) may influence flowering time.

The flowering process takes place at the main shoot that serves as a platform to give rise to flowers at its apex. Known as floral evocation, the process is the outcome of floral induction. Flowers are initiated after induction to form flower buds (Erwin, 2006). The flowers can also form on lateral branches or in some cases, both the main shoot and lateral branches. Either way, further development of axillary buds signals the first stage of flower development (Glover, 2007).

1.2 Physiological Control of Flowering Time

The factors that stimulate flowering time can be predictable or unpredictable. Day length is often predictable and vernalization (exposure to cold temperature to induce flowering) can sometimes be predictable or unpredictable. Other factors such as ambient temperature, light integral (day length irradiance) and water availability are less

predictable depending on the environment. The most unpredictable factors are mineral availability and neighbouring plants. Furthermore, plants also perceived endogenous cues such as size, node number and age where biennials are known to flower according to size rather than age (Lacey, 1986). In the control of flowering time, different environmental factors are perceived differently in various organs (Bernier, 1988). Vernalization is not only perceived by the shoot apical meristem of many plants (Searle et al., 2006) but the leaves of some species too (Metzger, 1988; Crosthwaite and Jenkins, 1993). Measurement of day length and light quality are usually by expanded leaves however, the stem is able to perceive day length in the absence of leaves (Bernier and Perilleux, 2005).

Daily rhythms depend on regular alterations in the environment such as diurnal changes in light, temperature and humidity, and are important in flowering process. These are controlled by the circadian clock in approximately 24-hour time. However, study on position of leaf movement suggested circadian rhythms continue in constant environmental conditions undergoing a non-24-hour period that can be reset by light (Harmer, 2009). The clock influences plant sensitivity to light and most genes in the clock pathway play roles in light signalling (Harmer, 2009). For example, nearly one third of expressed genes in *Arabidopsis* are controlled by circadian clock as reported by Covington et al (2008). Further study on flowering control led to the awareness of important roles played by photoperiod (response of an organism in changing light environment) and circadian clock in the regulation of seasonal responses (McClung, 2006).

1.3 Genetics of Flowering Time

Plants go through several developmental transitions in their life cycle before reaching flowering stage. Vegetative phase starts when the seeds germinate and once seedlings are produced it enters juvenile vegetative phase. Although incapable of flowering, this changes as soon as it passes through adult vegetative phase. This is the transition stage where floral inductive signals can trigger response to induce flowering. Once floral transition occurs plants enter the reproductive phase to produce seeds and fruits (Baurle and Dean, 2006).

A plant's ability to respond to its local environmental change is important to regulate flowering time for optimal reproduction. The decision to promote or delay flowering depended upon responses to these changes. The environment plays a big role in plants transition to flowering and this is especially true for plants in seasonal countries where they align their growth and development to suit various conditions all year around.

In the case of *Arabidopsis*, different but linked floral pathways (Fig 1.1) detect major seasonal cues such as day length (also known as photoperiod), temperature and other environmental as well as internal signals. Flowering is regulated by genes downstream of the floral pathways referred to as floral pathway integrator genes that interact with these signals. *FLOWERING LOCUS T (FT), SUPPRESSOR OF OVEREXPRESSION of CONSTANS1 (SOC1)* and *LEAFY (LFY*) are well known floral pathway integrator genes. These integrator genes will then activate floral meristem identity genes such as *APETALA1, APETALA2, FRUITFULL (FUL), CAULIFLOWER (CAL*) and *LFY* to promote flower development via its encoded proteins (Henderson and Dean, 2004; Putterill et al., 1995; Suarez-Lopez et al., 2001; Samach et al., 2000; Blazquez and Weigel, 2000).

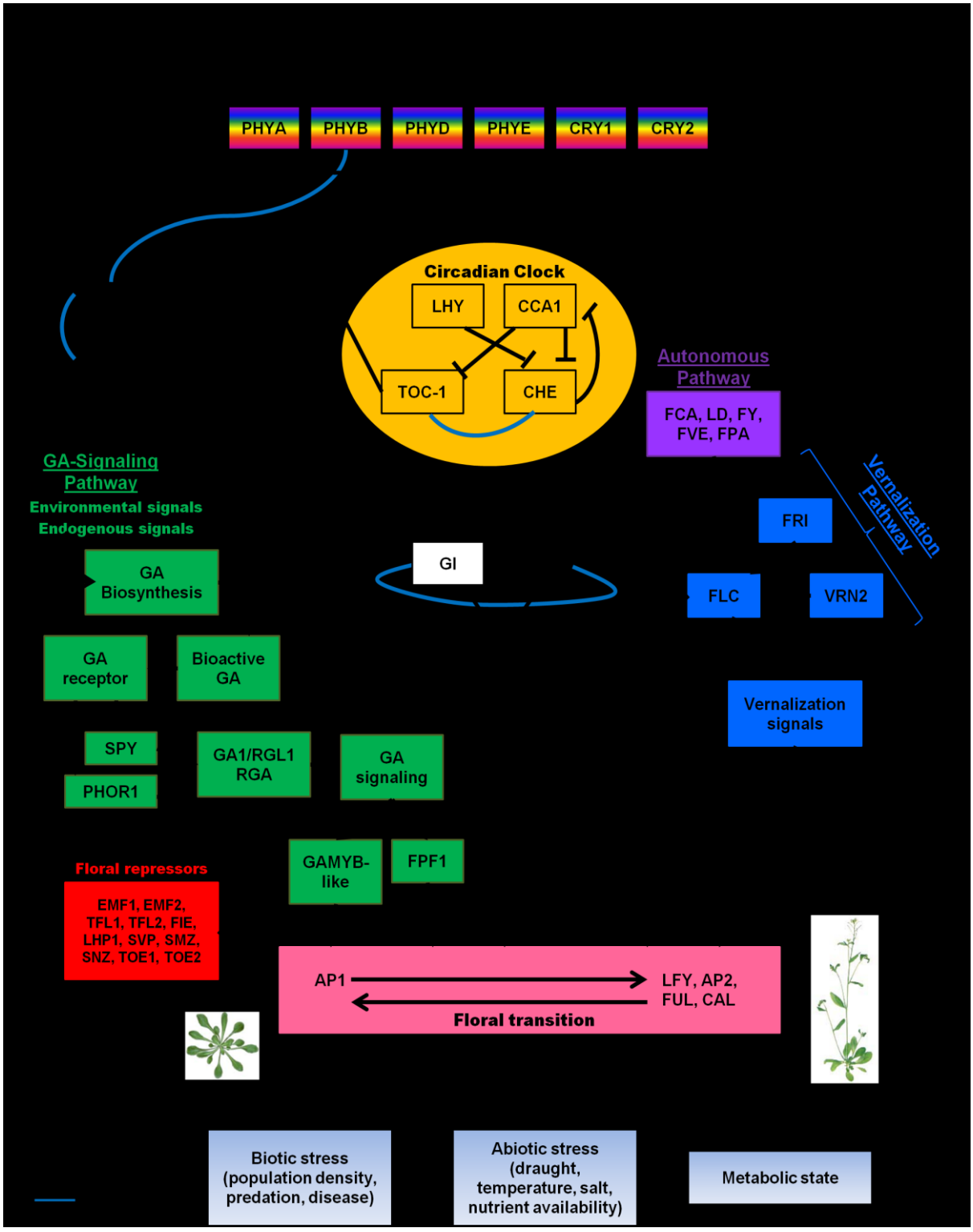


Figure 1.1: Control of flowering time in *Arabidopsis* involves 4 pathways that regulate the expression of floral integrator genes and proteins that allow plant in vegetative stage to produce flower.

Three classes of pathways regulating floral pathway integrators are identified in *Arabidopsis* referred to as promotion, enabling and resetting pathways (Boss et al., 2004). Photoperiod and gibberellins pathways together with ambient temperature and light-quality are known to promote floral transition. Most plants respond to day length in photoperiodism where some processes such as flowering time, onset of bud dormancy and production of storage organs (tubers and bulbs) can be controlled (Putterill et al., 2004, Henderson and Dean, 2004). Being a facultative long day plant, flowering in *Arabidopsis* is promoted by CONSTANS (CO) to upregulate the floral integrator genes (Suarez-Lopez et al., 2001; Samach et al., 2000; Blazquez and Weigel, 2000).

1.4 FLOWERING LOCUS T (FT) – The Florigen Involved in Regulation of Flowering

Regulation of plant growth is influenced by plant hormones produced in certain parts of plants with a specific function to the tissue. Based on that theory, the function of florigen was first proposed to induce flowering through a signal flowing from the leaves (Chailakhyan, 1975).

Developmental and external factors influence plants to produce signals in response to photoperiod. Originating from the leaves, the systemic signal produced is transported via the phloem and triggers the switch to induce flower at the shoot apex (Fig. 1.2). During the time when florigen was first hypothesized, phytohormones were also found to play key roles in plant growth and development as well as mediating plants' responses to the environment. This hypothesis led to an idea that florigen may be a simple compound but with a universal role in flowering. Thus, a florigen should have these properties 1) the signal is initiated in the leaves and travels to the shoot apical meristem, 2) the signal movement rate could be measured, 3) the signal is graft-transmissible and 4) the signal is universal (Willmitzer, 2006).

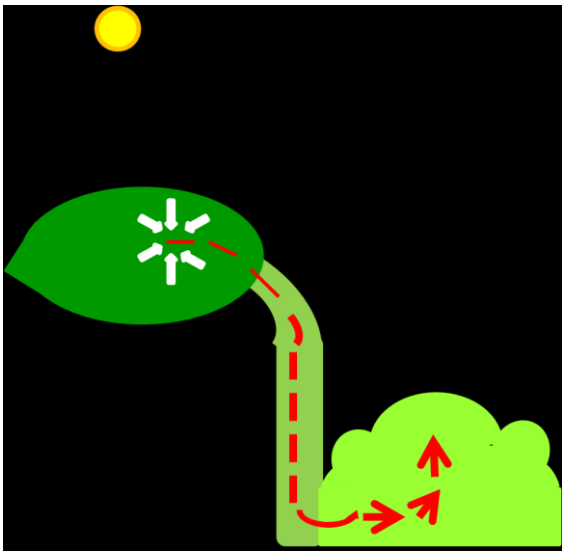


Figure 1.2: The florigen theory. Under environmental cues such as photoperiod and also other developmental processes such as plant size, signal is produced in the leaves. The signal travels along the phloem to the signal target, the shoot apical meristem (SAM) to produce flower (Adapted from Colasanti and Sundaresan, 2000).

Despite being described as universal, unique and specific, earlier studies could not isolate such compound. In the meantime, a multifactorial control hypothesis was proposed suggesting the involvement of several factors, promoters and inhibitors that control shoot apical meristem floral transition (Bernier, 1988; Corbesier and Coupland, 2006). The existence of florigen is supported by studies that found the stimulus is produced in the leaves and traveled to the meristem to promote flowering. Through various studies over the years, FT protein has been proven to be the florigen involved in long-distance signaling in regulation to flowering.

FLOWERING LOCUS T or *FT* is a flowering time gene belonging to a member of CETS protein family (CENTRORADIALIS (CEN), TERMINAL FLOWER 1 (TFL1) and FT) (Pneuli et al., 2001). It shows homology to a highly conserved family of small globular proteins, mammalian phosphatidylethanolamine-binding protein or the Raf kinase inhibitor proteins (RKIPs) (Kardailsky et al., 1999; Kobayashi et al., 1999; Yeung et al., 1999). Other member of CETS includes TWIN SISTER OF FT (TSF) which is closely related to FT (Michaels et al., 2005; Yamaguchi et al., 2005).

The study on ectopic expression of *FT* found an excessive expression of *APETALA 1 (AP1*) and higher levels of *LEAFY (LFY*), suggesting *FT* expression is sufficient to induce flowering (Kardailsky et al., 1999). Late flowering under long day (LD) with 16 hours light was characteristic of mutations in *FT* but was not affected in short day (SD, 10 hours light). Hence, *FT* gene is categorized in the photoperiod pathway where wild type plants are promoted to flower under LD (Kobayashi and Weigel, 2007; Turck et al., 2008). For activation of *FT*, *FD*, a bZIP transcription factor is needed to induce floral meristem identity genes to promote flowering.

Analysis of the *co* mutant confirmed that under photoperiod condition, FT acts downstream of CO. FT expression can only be activated in the presence of CO protein in daylight. Therefore, floral transition occurs during the light period when FT expression is upregulated (Valverde et al., 2004).

In the vascular tissue of the leaf, transcription of *CO*, *FT* and *TSF (TWIN SISTER OF FLOWERING LOCUS T)* takes place (Takada and Goto, 2003; An et al., 2004; Yamaguchi et al., 2005). From the leaves, FT protein moves to the meristem through the phloem to activate floral transition (Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007; Notaguchi et al., 2008). For floral activation, FT binds to FD, forming a heterodimer to induce *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1*) in the meristem (Fig. 1.3). In floral primordia, binding of FT and TSF to FD activates *AP1*, a floral meristem identity gene to confer floral identity (Abe et al., 2005; Wigge et al., 2005). Jang et al (2009) reported the possibility of TSF acting together with FT as a mobile signal as TSF was found to promote flowering when expressed in the phloem of *ft-10 tsf-1* plant. Besides, it also interacts with FD and FDP (FD paralog) to promote flowering. The only caveat is that the function of TSF in promoting flowering is still unclear (Turck et al., 2008). Although both *FT* and *TSF* are transcribed under LD, Yoo et al (2005) suggested CO promotes flowering mostly through *FT* activation.

Other small globular proteins such as green fluorescent protein (GFP) have been known to enter the sieve element from the companion cells and can be transported through the phloem and unloaded into surrounding tissue (Imlau et al., 1999; Stadler et al., 2005). FT is also transcribed in the companion cells (Fig. 1.3) and since GFP is larger in size than

FT (27 and 20 kDa respectively), it was suggested that FT too can move between cells (Kobayashi and Weigel, 2007). Long-distance movement of FT protein fused to GFP was reported as graft-transmissible (Lifschitz et al., 2006; Corbesier et al., 2007; Notaguchi et al., 2008). Corbesier et al (2007) documented movement of SUC2::FT:GFP fusion protein from the phloem companion cells to the meristem in *ft-7* plants. Through confocal microscopy, FT:GFP was detected in the vascular tissue of 6-day-old plants which had not yet undergone the floral transition and in 10-day-old plants which were about to undergo the transition to flowering. In addition, FT:GFP was also detected in the 10-day-old plants at the base of the shoot apical meristem.

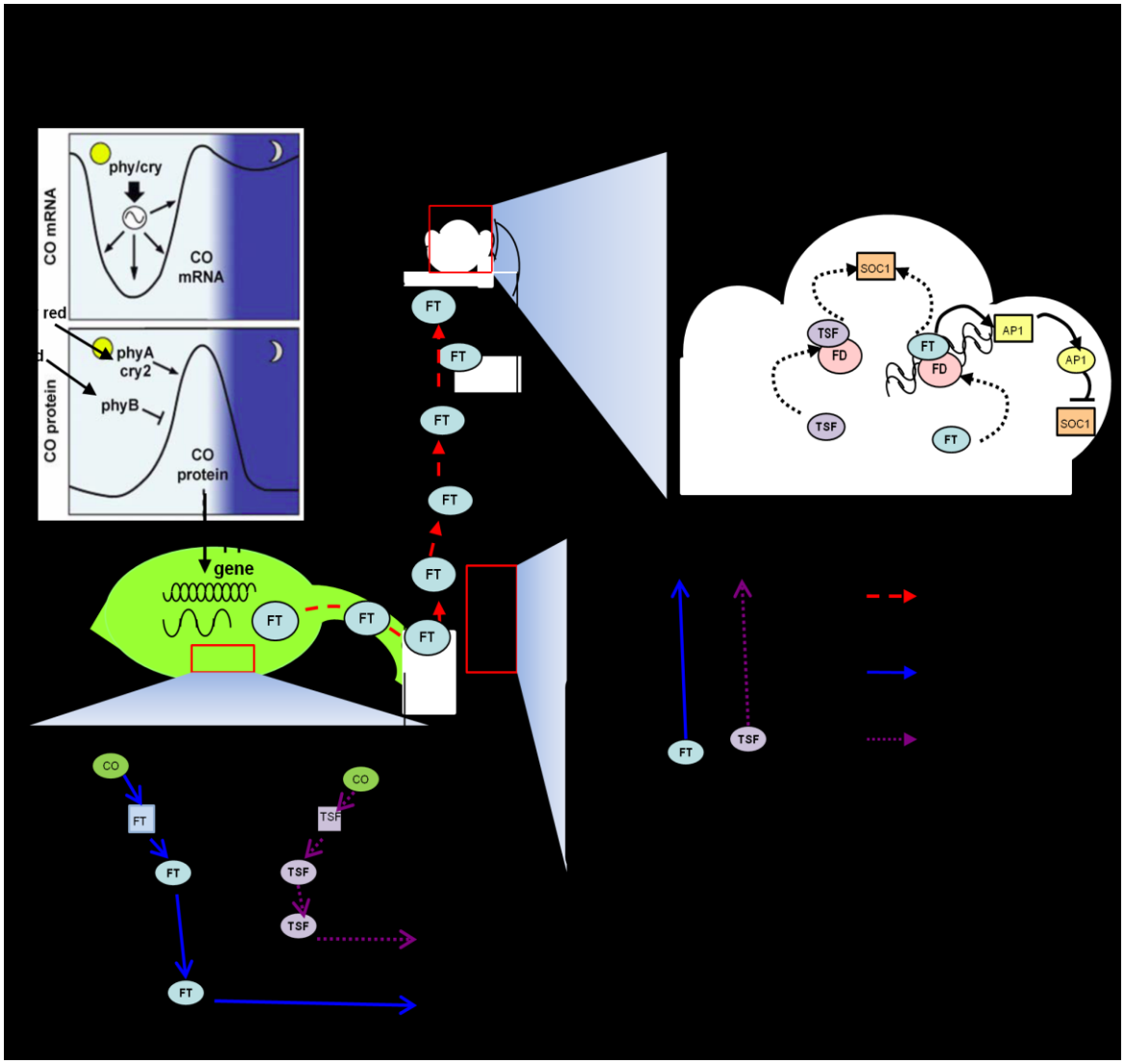


Figure 1.3: Photoperiodic pathway in *Arabidopsis* plants under LD (long day) to induce flowering. In the leaves of LD plants, light source boosts phytochrome and cryptochrome photoreceptors to produce CO mRNA. Far red light promotes expression of *phytochrome A (phy A*) and *cryptochrome 2 (cry 2*) to enhance production of CO protein. However, *phytochrome B (phy B*) which is promoted by the red light will repress CO protein production. Accumulation of CO protein results in the production of transmissible floral stimulus FT protein in the phloem. FT protein is transported via the phloem to the shoot apical meristem (SAM) where it binds to FD, a transcription factor to activate *AP1* and *SOC1* genes to trigger the switch to induce flowering at the shoot apex. AP1-APETALA 1, CC-companion cell, CO-CONSTANS, FT-FLOWERING LOCUS T, FD-bZIP transcription factor, FM-floral meristem, IM-inflorescence meristem, SAM-shoot apical meristem, SE-sieve element, SOC1-SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1, TSF-TWIN SISTER OF FT. (Adapted and modified from Yang et al., 2007 and Turck et al., 2008)

1.5 Phloem and Translocation

Phloem is a name derived from the Greek word *phloos* meaning bark, where in trees, the phloem is found in the innermost layer of the bark. In vascular plants, it is a living tissue that distributes organic nutrients or photosynthate (for example sugars) to all parts of plants in a process known as translocation. A pressure-driven mass flow system enables transportation over long distances in plants. Phloem translocation of secondary plant products in many species may function to defend the plants against pests and predators. Furthermore, the sieve tube is capable of translocating insecticides and herbicides within the plants (van Bel, 2003).

In vascular cambium, the phloem originates from the meristematic cells which will then further grow outwards. It is produced in phases where the primary phloem is established by the apical meristem and later established to secondary phloem by vascular cambium. Its structure comprises of less specialized and nucleate parenchyma cells, sieve tube cells and companion cells. In addition, albuminous cells, fibers and sclereids also make up the structure.

In the phloem, sieve element (SE) and companion cell (CC) forms a functional complex where it is connected to surrounding phloem parenchyma cells and other cell types via plasmodesmata. In most cases, CC is the primary cell connected to surrounding cells. Between SE and CC, the main communication pathway is the plasmodesmata-pore unit, a plasmodesmata modified complex that allows large molecules to pass through between the two cells. The size exclusion limit (SEL) correlates to the maximum size of molecule capable of crossing plasmodesmata where only molecules smaller than 1 kDa move freely among cells of the same domain (Oparka and Roberts, 2001). For molecules larger than 1 kDa to pass through plasmodesmata, it was suggested that they have to have some degree of affinity with proteins embedded in the pore (Concenco and Galon, 2011). There are a large number of plasmodesmata in the phloem but its conductivity depends on the tissue type it connects to and also its developmental stage. With its passive transport, there is no limit for movement of small molecules. As long as the molecule is in the cytosolic compartments limitation of movement between cells is only by a molecule's size, conductivity of the pores and charge (Erwee and Goodwin, 1983; Turgeon and Wolf,

2009). Within a mature SE there is a layer of cisternae covering almost all of the plasma membrane known as sieve element reticulum. It is thought the space between plasma membrane and sieve element reticulum may close off molecules to protect them for the rushing current of sap.

Phloem translocation is a process of transporting water and minerals across the phloem by processes known as phloem loading and unloading. A transport pathway into and out of the phloem starts in the source leaves and moves along the path or conducting phloem before unloading occurs in sink tissues.

1.5.1 Phloem Loading and Unloading

Two types of loading and unloading of metabolites are termed symplasmic or apoplasmic. Symplasmic loading/unloading occurs through plasmodesmata and cell cytosols and depends on the availability of plasmodesmata though concentration gradients could allow diffusion of photosynthates from the mesophyll cells to the sieve tubes. Entrance of photosynthates across the plasma membranes via intervening cell walls allows apoplasmic loading/unloading into the sieve elements – companion cell complex of the minor veins (Cronshaw, 1981; Lalonde et al., 2003).

Loading of solutes into the CC from phloem parenchyma is apoplastically through apoplastic loaders. In apoplastic loading, a reduction in plasmodesmata number and permeability decreases symplastic transport. Transport from CCs to SEs on the other hand is mainly symplastic through high permeability plasmodesmata-pore units. Sometimes a leak of solutes from SEs will be recovered by CCs or other cell types. The contents of the SEs and CCs exchange readily along the path phloem, and efflux is mainly apoplastic as a result of low permeability of CC plasmodesmata. Finally, once solutes reached sink tissues unloading is symplastic most of the time, but depending on the type of sink and its developmental stage, unloading may also be apoplastic (Fig. 1.4) (Turgeon and Wolf, 2009).

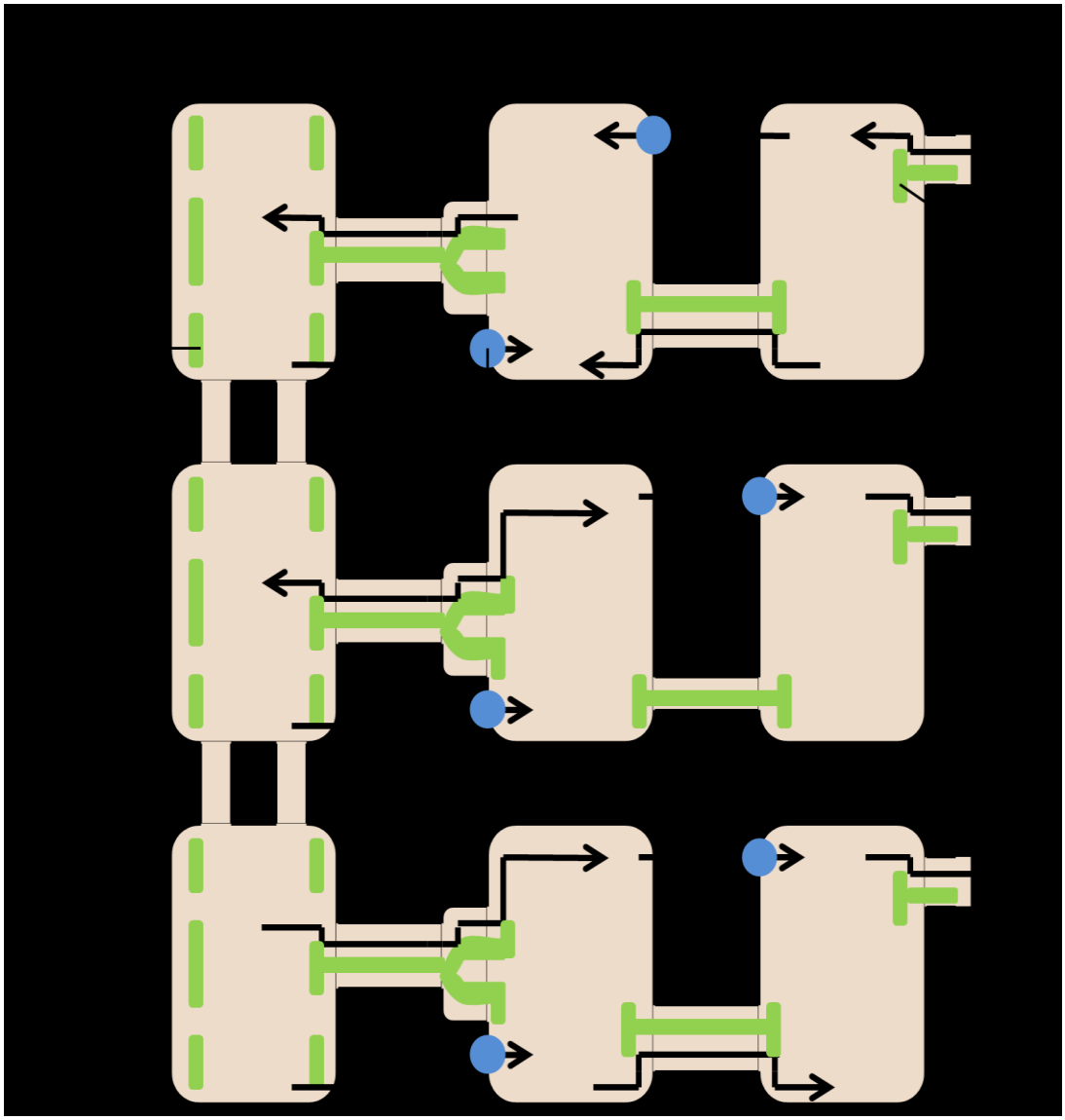


Figure 1.4: Primary pathway into and out of phloem. Starting from the source tissues, solutes from phloem parenchyma are unloaded into CC through plasmodesmata by a specific transporter. By using plasmodesmata-pore unit which has a high permeability, solutes are transported to SE. In the path phloem however, low permeability of the plasmodesmata-pore unit enables the exchange of contents between SEs and CCs. Upon reaching sink tissues, solutes are unloaded and distributed to other tissues in the plant. CC-companion cell, SE-sieve element. (Adapted from Turgeon and Wolf, 2009)

The movement of fluorescent protein such as GFP has been observed to study qualitative connectivity of the symplast (Kim and Zambryski, 2005), a cytoplasmic continuum that arises from opened plasmodesmata (Rutschow et al., 2011). Believed to play vital role in growth and development, symplastic domains occur in meristems and mature tissue (Erwee and Goodwin, 1985; Oparka et al., 1994; Rinne and van der Schoot, 1998). However, gradual declines in symplastic connectivity were observed as cells differentiate and pass from the meristem to mature tissue (Duckett et al., 1994; Kim and Zambryski, 2005). Detection of FT:GFP from the phloem companion cells to the meristem by Corbesier et al (2007) may be provided by sympastic connectivity through plasmodesmata. Notaguchi et al (2008) reported transport of FT protein within a short period of 24 to 48 hours from the scion to the apical region of stock plants suggesting that symplastic loading is involved in FT movement.

1.5.2 Sucrose Transporters in the Phloem

Sucrose is important in the phloem translocation system as a main source of carbohydrate. Transport of synthesized sucrose into the phloem is mediated either by diffusion, pressure force or sucrose transporter. Most phloem-localized sucrose transporters normally operate at acidic pH value for ease of sucrose uptake (Geiger, 2011). Increment in acidity of extracellular space in the cytoplasm along with negative potential (Chandran et al., 2003) is what allows movement of sucrose.

In the phloem, sucrose transporters in general transport sucrose from source to sink cells. During phloem loading, especially apoplastic loading, sucrose transporters transfer sucrose from the cytosol of photosynthetic cells to the apoplast and loading is made into sink cells such as SE and CC (Truernit, 2001; Shiratake, 2007). In fact sucrose transporter has been expressed in various sink organs and tissues in various plant species such as *Ricinus communis*, tobacco, cotton, rice and *Arabidopsis* (Chew et al., 1998; Lemoine et al., 1999; Ruan et al., 2001; Aoki et al., 2003; Baud et al., 2005). Additionally, studies in tobacco and potato, both of which come from the Solanaceae family, demonstrated that sucrose transporter played vital role in carbohydrate partitioning (Burkle et al., 1998; Lemoine et al., 1996).

Sucrose transporters are able to move sucrose against its concentration gradient across plasma membranes as a result of negative potential from proton motive force produced by H^+-ATPases (Gottwald et al., 2000). In apoplastic loading, plasma membrane sucrose-H^+ symporters such as SUTs or SUCs are required (Williams et al., 2000). For example, uptake of sucrose by SUC1 and SUC2 is energy dependent (Sauer and Stolz, 1994). SUC2 is a well-known phloem-specific sucrose transporter in *Arabidopsis* (Truernit and Sauer, 1995; Stadler and Sauer, 1996) where its expression has been observed in the CC.

1.5.3 Callose Accumulation and Occlusion

Earlier on, callose was identified through aniline-blue staining as a polysaccharide in plant cells (Xie and Hong, 2011). Its chemical structure was later confirmed as β-1,3-glucan. During normal plant growth, callose is confined at a basal level on the sieve plates, and is also found in dividing cell plates, pollen tubes and plasmodesmatal channels (Verma and Hong, 2001; Jacobs et al., 2003). Synthesis and accumulation of callose is regulated during cell growth, cell division and differentiation (Xie and Hong, 2011) though accumulation can be induced by chemical or mechanical stress (Hong et al., 2001; Verma and Hong, 2001; Nakashima et al., 2003; Levy et al., 2007).

Callose accumulation or deposition occurred intracellularly within cytoplasm as pollen grains (Cresti and van Went, 1976) while extracellular deposition is around plasmodesmata and sieve pores (Zabotin et al., 2002). During accumulation, callose is not the only compound being deposited but also minor amounts of other polysaccharides, phenolic components, reactive oxygen intermediates and protein (Heath, 2002). Massive accumulation upon exposure to stress occurred rapidly, resulting in plug formation at the sieve pores (Cresti and van Went, 1976). Referred to as a sieve tube occlusion, the process prevents loss of sieve tube sap (Schulz, 1998). The occlusion process is calcium-dependent (Colombani et al., 2004) and occurred along the path of damage-induced electropotential waves (Furch et al., 2007). Additionally, protein plugging also occurred in response to stress though both plugging by callose and proteins could take place in the same species in coordinated time-shift (Furch et al., 2007; Furch et al., 2010).

Exclusive to legumes or Fabaceae family only, sieve pores for this species were reported to be occluded by spindle-like protein bodies called forisomes (Knoblauch and van Bel, 1998). With similar role to prevent loss of photoassimilates during stress, forisomes occlude sieve tube by dispersion. Its dispersion from condensed spindle-shaped form is by conformational change that is reversible and calcium-dependent (Knoblauch et al., 2001; Knoblauch et al., 2003). As forisomes are only found in legumes, adjustment of their size is possible via genetic engineering making it the finest characteristic of forisomes (Tuteja et al., 2010).

1.6 Exudation Technique

Birds and insects were probably the first species to discover the nutrients in the phloem sap by means of sucking it out of the plant. In the phloem, the sieve elements are the principal conduit and the exudates flow under positive hydrostatic pressure in it. Thus, it led to the development of exudation techniques to sample the phloem such as bark incision and aphid stylet technique (Crafts and Crisp, 1971). Since then the technique has been used to investigate phloem sucrose flow (Munier-Jolain and Salon, 2003) and studying phloem sap concentration and composition (Friedman et al., 1986).

A wound inflicted by incision that severs sieve element would release the sap within. However, the amount of sap released is almost undetectable in most plant species due to sealing mechanism causing the wound to close. A sealing or clotting mechanism is the result of P-protein or “slime” (Evert, 1977), a phloem protein drawing itself against the walls (sieve plates) of sieve elements after wounding, causing rapid reduction in hydrostatic pressure. Following that, callose is also formed at the pores of the sieve plates to block it thus sealing the plate (Johnson, 1978; Walsh and Melaragno, 1981).

Phloem sap exudation by incision has been carried out in various plants. However, getting enough sap for analysis is a problem faced by many researchers. Earlier work was mainly done in *Ricinus*, cucurbits and oilseed rape where sufficient amount of phloem sap exudates can be obtained (Kehr, 2006). In *Arabidopsis*, various methods have been applied and so far the use of EDTA chelation technique has been proven useful to obtain enough phloem sap for analysis. Treating cut tissue or organs with the chelating agent prevents

blocking of sieve tubes (Gessler et al., 2004) and inhibits callose formation by binding the divalent ions (Ca^{2+}) (King and Zeevart, 1974; Tully and Hanson, 1979). This method has also been successful in other plants such as sunflower, mung bean, grape vine, orange and soybean (King and Zeevart, 1974; Friedman et al., 1986; Touraine and Astruc, 1990).

However, incision disrupts not only the phloem but xylem as well that might cause the phloem to be diluted and, or contaminated by the apoplasm and the xylem. Najla et al. (2010) reported an enhanced phloem exudation technique by combining HEPES and EDTA. HEPES alone cause a large uptake by xylem in contrast to when used in combination with EDTA making measurement of xylem uptake possible. The combined solution allows an estimation to be made of the balance between xylem uptake and phloem exudates.

On the other hand, aphids are capable of becoming "sinks", the capability of a tissue to acquire or store assimilates (Crafts and Crisp, 1971), and can divert most or all of the sap from plants' own sinks. In addition, their ability to feed on sap at a rapid rate allows the collection of exudates before sealing mechanism takes place (de Wet and Botha, 2007). Increment in the volume flux through stylet and sieve tube was reported (Pritchard, 1996) hence cutting the stylet is an important aspect as it would determine the sink strength of an aphid. Nevertheless, there are limiting factors to the aphid-stylet technique. The aphid has to be compatible with the host plant for it to feed and the method takes into account the growth conditions that would suit both parties. Even then, severing the stylet without disrupting the phloem is not an easy task (Helden et al., 1994). And finally, only a small amount of sample is obtained. To avoid evaporation of phloem droplet, stylets were either immersed in oil or plants have to be placed in saturated water-vapor pressure atmosphere (Fisher and Frame, 1984; Hayashi and Chino, 1986) and this could lead to other problem such as molding on the soil if left too long.

Another technique to study phloem is by using carbon isotope labeling. Isotopically labelled substrates such as $^{14}CO_2$ and $^{11}CO2$ can be used for exposure to the leaves and later detected downstream. Due to its short half-life, ^{11}C is less favored for studies as compared to $^{14}CO_2$ that was used in earlier study by Biddulph and Cory (1957) to identify transported photoassimilates. There is a chance of misidentification of metabolic products

as transport compound due to the fact that labeled compounds can be metabolized in transit but the effect has been reported as minimal over short periods (Webb and Gorham, 1964). Despite determination of a phloem-mobile compound being possible with this technique, no guarantee can be made that the same endogenous compound is mobile as it depends on the availability of a compartment for transport (Turgeon and Wolf, 2009).

1.7 Plant Proteomics

Proteomics is an important area to complement information about genome structure and regulation. Information on post-translational modification and subcellular localization of gene products from proteomics help in understanding many cellular processes and network functions.

In early studies, proteomics was defined as two-dimensional (2D) gels of cell lysates with annotation where the gels are used to visualize differential protein expression. However, with the increasing amount of work and data on genomes, it has been redefined into a wider perspective that includes 1) protein identification in the form of one-dimensional gels (1D), 2D gels, protein chips and protein/protein complexes in solution, and, 2) post-translational modifications such as phosphorylation and glycosylation, all of which utilise the use of mass spectrometry (MS). The function of proteomics has broadened since in various fields (Pandey and Mann, 2000). Proteomics is now defined as the systematic analysis of all proteins in a tissue, cell or subcellular compartment (van Wijk, 2001).

Proteomics analysis complements the analysis of the transcriptome and metabolome. Protein is a product of translated transcript in the central dogma of molecular biology. The positive correlations have been reported associating transcript levels with the abundance of subsets of all proteins (Ideker et al., 2001; Greenbaum et al., 2003; Kleffmann et al., 2004). However, regulation of protein levels can be independent of its transcript level in many pathways. In those cases, the control of mRNA translation, protein stability or antisense transcription may influence the final concentration of a protein (Yamada et al., 2003; Baginsky and Gruissem, 2006).

Protein identification accommodates new insights into the interrelated biological processes of specific proteome. Total protein populations are often the target of plant proteomics studies and as such, require good separation technique. This is partly due to the high dynamic range of protein abundance in eukaryotic cells that could sometimes hinder analysis of low abundance proteins (Hurkman and Tanaka, 2007).

A complete map of a species of plant proteome would provide information on genome activities and gene structures. Expression of a gene exhibited by identified peptides complement the annotation of open reading frames and confirms or corrects gene structure prediction. Furthermore, protein analysis accommodates certain dimensions of information inaccessible from the analysis of other cellular components (Baginsky and Gruissem, 2007).

Proteome alterations in plant including post-translational modifications occurred in response to the plant's perception to signal molecules (Newton et al., 2004). Regardless, plant proteomics has been used for studies such as 1) study of protein variations in different plant organs (Mo et al., 2003), 2) variation in response to physiological events (Gallardo et al., 2001), identifying plant viruses due to their proteome (Cooper et al., 2003) and 4) identification of microtubule binding protein in plants (Chan et al., 2003).

1.8 The Phloem Proteome

As an essential part of plant development, the phloem delivers nutrients to other tissues and organs throughout the plant. Recently, translocation of other macromolecules such as proteins, mRNA, small RNA and viral nucleic acids in the phloem has also been reported (Gomez et al., 2005; Foster et al., 2002; Haywood et al., 2005; Corbesier et al., 2007; Lin et al., 2007; Mathieu et al., 2007; Jaeger and Wigge, 2007; Lin et al., 2009).

FT is an example of a protein translocated through phloem to the shoot apical meristem to induce flowering (Corbesier et al., 2007). Within the phloem translocation stream, there are also other proteins implicated in stress, defense and antioxidant defense system (Walz et al., 2004; Walz et al., 2002). Studies of several plant species have adopted the proteomics approach by analyzing phloem exudates from pumpkin (*Cucurbita*

maxima), castor bean (*Ricinus communis*), rape (*Brassica napus*), cucumber (*Cucumis sativus*), rice (*Oryza sativa*), hybrid poplar (*Populus trichocarpa* x *Populus deltoides*) and lupin and led to the identification of various proteins (Haebel and Kehr, 2001; Barnes et al., 2004; Giavalisco et al., 2006; Walz et al., 2004; Aki et al., 2008; Dafoe et al., 2009; Rodrigues-Medina et al., 2011).

1.9 Polyacrylamide Gel Electrophoresis (PAGE)

In proteomics study, it is common to separate complex protein or peptide samples using denaturing sodium dodecyl sulfate polyacrylamide (SDS-PAGE) followed by liquid chromatography and identification by mass spectrometry (Hunter et al., 2002). SDS-PAGE is widely used to separate proteins according to size and is easy to carry out with good reproducibility.

SDS-PAGE separated proteins based on their size just as normal electrophoresis does. However, one band from SDS-PAGE may represent more than one protein. Phinney and Thelen (2005) reported that a single distinct band produced by SDS-PAGE from a complex protein sample produced multiple proteins when analyzed using mass spectrometry. They noted that volume of any band is in general the collective composition of each unique protein corresponding to the size of the band. Though the proteins may have similar weights they may have different migration rate attributed to their secondary, tertiary or quartenary structure. To solve these complexities the proteins are reduced to their primary structure or linearized form by the action of SDS. As an anionic detergent, in SDS-PAGE SDS is known to coat the proteins with uniform negative charges (Garfin and Heerdt, 2004).

1.10 Two Dimensional (2D) Proteomics

Analysis of complex protein mixtures from cells, tissues and other biological samples has widely utilized two-dimensional (2D) electrophoresis method. 2D electrophoresis separates protein into two independent properties in two distinct steps. Application of isoelectric focusing (IEF) in the first dimension and SDS-PAGE in the second dimension became a basis for protein mapping. When IEF is applied in the first

dimension, proteins are separated according to their isoelectric points (pI) and further separation of proteins in accordance to their molecular weights is achieved using SDS-PAGE in the second dimension. Different from 1D SDS-PAGE that produces protein bands, 2D electrophoresis produces protein spots. In theory each spot from the 2D gel represents a single protein species in the sample. Therefore, separation of huge amounts of different proteins is possible to deliver information such as the protein pI, the apparent molecular weight and the amount of each protein (Gorg, 2004).

Earlier in its emergence, this technique suffered a drawback as despite having a better resolving efficacy it had issues regarding its reproducibility that stemmed from the fragile tube gels used for IEF. Improvement was achieved after the introduction and commercialization of immobilized pH gradient (IPG) strip in place of IEF tube gels. Coupled with programmable focusing unit for IEF, the reproducibility and resolution was strongly enhanced (Thelen and Peck, 2007). 2D electrophoresis resulted in a static proteome reference map with the visual entity being in the form of a spot, making it one preferred approach for proteome profiling and quantitative characterization of complex protein samples.

1.10.1 Sample Preparation and Protein Solubilisation for 2D Gel Electrophoresis

2D gel electrophoresis relies heavily on sample treatment and resolution as the first step is IEF and for that proteins have to retain their own charge. Proteins must reach their pI and stay soluble at the pI (Rabilloud et al., 2007).

Protein solubilisation is probably the most important aspect in 2D gel electrophoresis since proteins have to be well solubilised prior to and throughout the separation procedure (Berkelman et al., 2004). In a biological sample, proteins are generally present initially in their native state, where they are mostly not soluble and this continues even after isolation. Solubilisation functions to prevent protein aggregation by breaking interactions including ionic interactions, hydrophobic interactions, hydrogen bonds, disulfide bonds and van der Waals forces (Rabilloud, 1996).

Another factor to consider is the choice of chemicals that should not change net electric charge in solution across the pH range chosen for IEF to ensure proteins are thoroughly solubilised. The use of chaotropic agents such as urea and thiourea helps solubilisation of proteins by disrupting non-covalent interactions between molecules in the sample (Chevalier, 2010). Hydrogen bonds normally cause aggregation or formation of secondary structures that may affect protein mobility hence the use of chaotropic agents help not only better solubilisation but also to unfold proteins.

The use of detergent in protein solubilisation is to increase solubility of proteins at their pI by interrupting hydrophobic bonds. Wherever possible, any strong ionic detergent must not be present in the buffer. In principle ionic detergent creates a charged coat on protein molecules and is much more efficient as the protein-detergent complexes repel each other via ionic interactions to prevent protein aggregation (Chavelier, 2010). Although sodium dodecyl sulphate (SDS) is a common detergent used for protein work, it is an anionic detergent and not compatible with IEF. As a replacement non-ionic or zwitterionic detergents can be used to allow protein migration according to their own charges. However, the difference between the two is that the former has no charges on the molecule whereas the latter has an equal number of positive and negative charges on the molecule. In IEF, completely zwitterionic detergent is preferable over the pH range of interest (Chavelier, 2010).

Disulfide bonds within proteins and between proteins subunits can also affect solubilisation. Fortunately it can be cleaved using reducing agents. Tributyl phosphine (TBP) or dithiotreitol (DTT) are the two most commonly used reducing agents. At low concentration TBP reduce cysteines stoichiometrically (Herbert et al., 1998) and is active over a wider pH range (Berkelman et al., 2004) thus is much more effective than DTT but it is also chemically much more difficult to handle. On the other hand, reduction of cysteines can only be achieved when DTT is in excessive volume. Even then, not all proteins are completely reduced and its concentration cannot be too high as it can affect the pH gradient (Rabilloud, 1999).

1.10.2 The First-Dimension, Second-Dimension, Protein Detection on Gel and Spot Identification

As amphoteric molecules, proteins carry positive, negative or zero net charge according to their amino acid composition and the pH of the local environment. The sum of all positive and negative charges will give the net charge of a protein. The isoelectric point (pI) of a protein is defined as the specific pH when the net charge is zero. In other words, at its pI, a protein will not migrate in an electric field and at the same time the number and types of charged groups in the protein is determined. At pH values below their pI, proteins are positively charged while at pH value above their pI, are negatively charged (Chevalier, 2010).

The first dimension separation of proteins is achieved via isoelectric focusing (IEF) using immobilized pH gradients (IPGs). In the presence of an electric field, proteins will initially move toward the electrode with the opposite charge. These proteins either gain or lose protons as it migrates through the pH gradient and it will slow down when its net charge and mobility decreased. Upon reaching its destination on the pH gradient that equals to its pI proteins will stop migrating as they are no longer charged (Fig 1.5). If the protein is accidentally diffused to a region lower or higher than its pI, it will be forced to move towards the cathode or anode in the electric field where it is focused into a sharp spot in the pH gradient that correspond to its individual characteristics pI value.

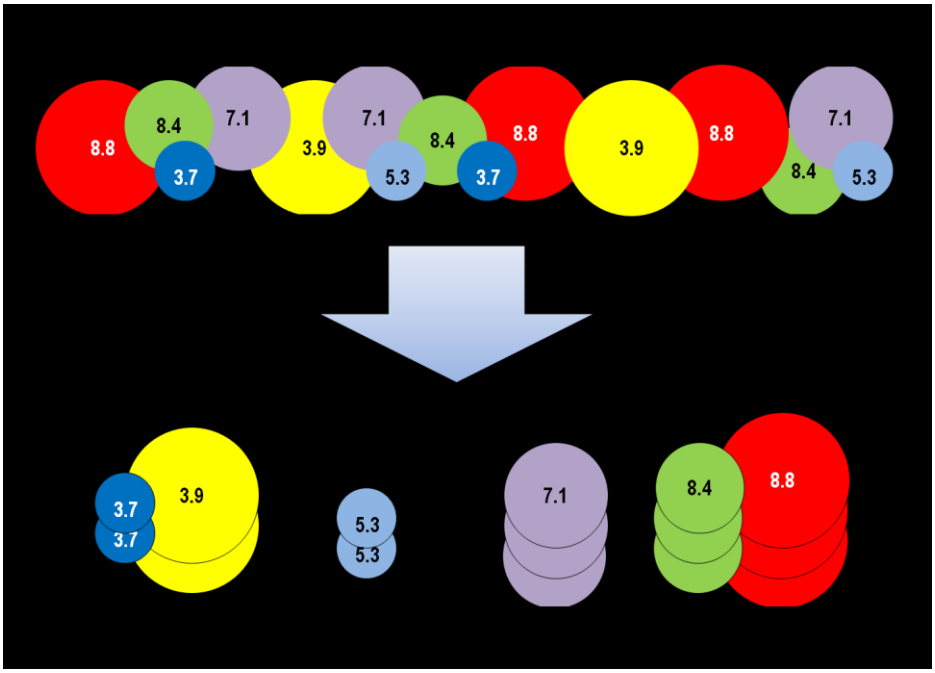


Figure 1.5: Schematic diagram of a mixture of proteins before IEF and how they are then separated based on their pI after focusing.

Prior to second dimension separation IPG strips were equilibrated before embedding it onto second dimension gel. Equilibration is a critical step involving two buffers comprising of saturated SDS that function to denature proteins and form negatively charged protein/SDS complex (Chevalier, 2010). The first SDS buffer contains DTT to maintain a reducing environment while the second contain iodoacetamide, an alkylating agent used to reduce thiol group thus preventing re-oxidation during electrophoresis. Theoretically after equilibration, proteins are completely covered by negative charges and can now be separated based on molecular mass (Fig 1.6) (Chavelier, 2010).

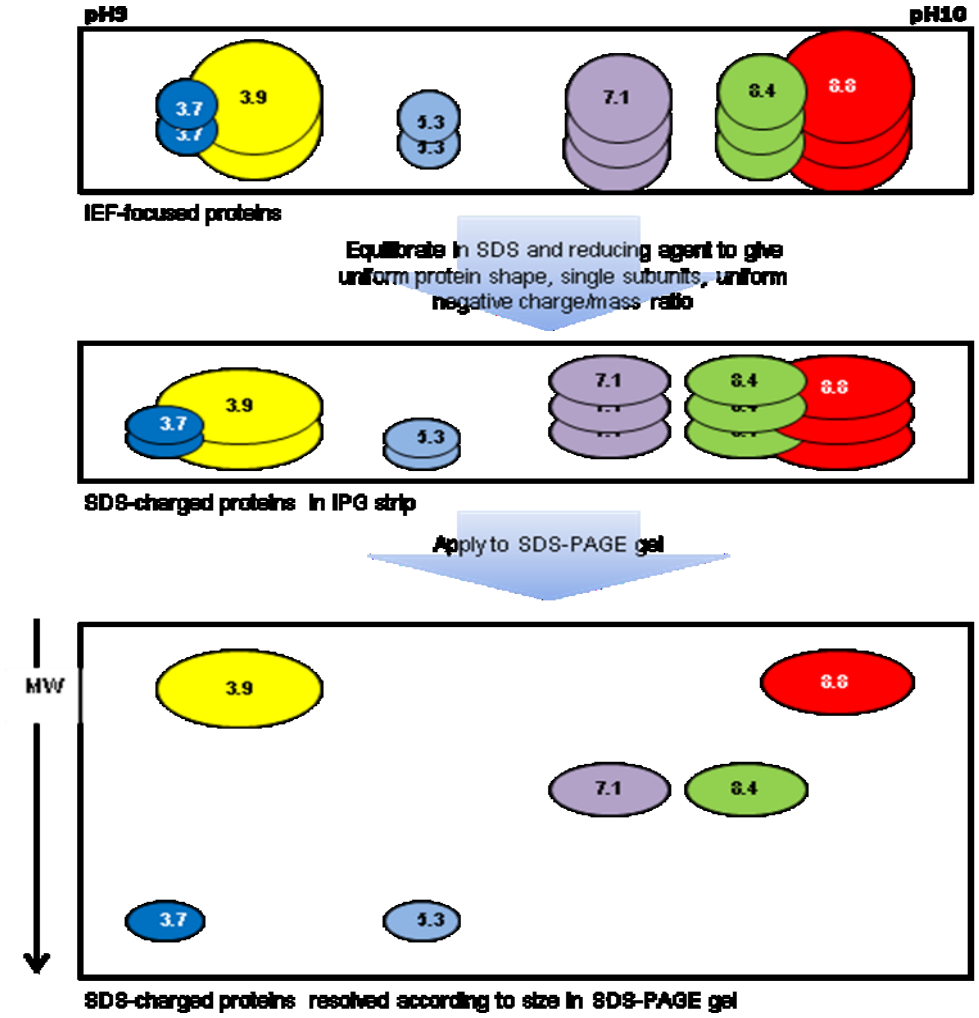


Figure 1.6: Separation of protein by SDS-PAGE after separation by IEF.

1.11 Mass Spectrometry-Based Proteomics

A wide range of applications has employed mass spectrometry (MS) and MS has now played a significant role in the biological sciences. Measurement of mass-to-charge-ratio by the MS has been useful for identification of chemicals based on their molecular mass and predictable fragmentation pattern. In addition, relative abundance of ions can also be measured. The basis of mass spectrometry relies on its key components such as an ion source, an analyzer, an ion detector and a computer for its output (Aebersold and Mann, 2003). Four main functions that carry out the principle of MS are ionization, separation of ions by mass, measurement of mass and measurement of abundance (Newton et al., 2004).

Generally, a sample to be analyzed has to be mixed with a supporting matrix and sprayed into an electric field. The function is to ionize the sample resulting in the molecule of interest to be positively or negatively charged. As it travels into a vacuum tube (the analyzer), an electric field will influence the ions to move based on its mass. Separation of ion species is based on their mass-to-charge (*m/z*) ratio. Velocities of ions differ with mass and arrive at the ion detector at different times. Detection of the ion is then recorded where the data will generate a mass spectrum through a computer. The *m/z* ratio against the percentage relative intensity of the detected ion is represented through each peak on the graph, giving the measurement of ion abundance (Aebersold and Mann, 2003; Newton et al., 2004).

There are a few types of MS available for various analyses. The difference lies in the types of analysers and detectors used. They are;

1) magnetic-sector MS which is the oldest type of MS whereby ions are separated using momentum dispersion and directional focusing,
2) quadrupole MS containing four matching parallel metal rods that separate ions using direct-currect (dc) and radio frequency (rf) potentials creating high-frequency oscillating electric field,
3) time-of-flight MS that separates ions in a long field-free flight tube based on their velocity and measuring the time taken to reach the detector,
4) quadrupole ion-trap MS that manages ions movement in time in an oscillating electric trapping field comprising of three-electrode structure and the ions trapped are stored for separation,
5) linear ion-trap MS which construction is exactly like the quadrupole MS but operates as in quadruople ion-trap MS,
6) fourier-transform ion cyclotron resonance MS which traps ions in a cell known as Penning trap using static magnetic and electric field,
7) orbitrap MS that orbitally traps ions using a pure magnetic field around a spindlelike central electrode and coaxial barrel-like outer electrode, and
8) ion mobility MS which is a hybrid using gas-phase mobility to separate ion under an electric field gradient and cross-flow of a buffer gas (Dass, 2007).

Interpretation of the information encoded in genomes using MS-based proteomics is a useful tool. Consisting of five stages, the proteins to be analyzed are first isolated from the cell or tissue by biochemical fractionation. The proteins are then degraded enzymatically using trypsin, creating peptides with C-terminally protonated amino acids. Separation of the peptides in stage 3 by one or more steps of high-pressure liquid chromatography is carried out in fine capillaries and an electrospray ion source will elute the peptides into small, highly charged droplets. Following evaporation, the protonated peptides enter the MS and a mass spectrum of the peptides is recorded. In the last stage, a tandem MS/MS spectrum is generated by the computer, consisting of isolated peptide ions resulting from fragmentation by high energy collision with gas. For about one second each, the MS and MS/MS spectra are acquired and stored to match against protein sequence databases. The end product is an identified peptide (Aebersold and Mann, 2003).

1.12 Multiple Reaction Monitoring (MRM)

Specific peptides in a complex mixture can be measured by designing an experiment such as multiple reaction monitoring (MRM) to obtain maximum sensitivity for detection of target compounds. Widely used to detect and quantify metabolites especially in the pharmaceutical industry, MRM has been applied to screen for metabolites where metabolite's structure was confirmed by triggering dependent product ion scan. A modified peptide with known mass and structure after post-translational processing can also be studied by utilizing the same principle (Cox et al., 2005; Anderson and Hunter, 2006).

MRM is a process where selective ion monitoring technique is used to detect or screen molecules that have undergone a thorough characterization during discovery phase (Kitteringham et al., 2009). The focus is on analytes of specific masses and excluding all others. Precursor *m/z* and fragment *m/z* (MRM transition) for many common compound molecule can be predicted by selecting a specific tryptic peptide as stoichiometric representative of the cleaved protein and quantitated against a spiked internal standard (Cox et al., 2005; Anderson and Hunter, 2006).

In proteomics analysis, it is recommended the target peptides to not exhibit any enzyme missed cleavage sites and they should not be inclined to post-translational modification. The requirement for the former rule depends heavily on the effectiveness of trpsin and digestion method. However, for quantification of post-translational modification study, the latter rule does not apply. In addition to being in a size that accommodated the mass range of the instrument used, target peptides should uniquely identify the protein of interest. To validate huge number of individual peptide transitions, various *in silico* methods are available. Generally, putative MRM transitions are delivered based on sophisticated algorithms using a combination of theoretical rules and empirical observation for optimal peptide MS/MS (Kitteringham et al., 2009).

1.13 Photoactivatable Fluorescent Proteins (PAFP)

Gaining importance in biological sciences, colorful fluorescent proteins (FPs) are derived from marine invertebrates. Apart from naturally occurring FP, many engineered variants have also been commercialized. Originally, the fluorescent properties of FPs come from an internal tripeptide sequence which is modified to form a cyclized chromophore (Heim et al., 1994). A specific and predictable form of intrinsic protein fluorescence is represented from the fluorescent. The maximum emission of FPs across the visible spectrum ranges from blue to red (440 to 650nm) while under ultraviolet light the irradiation ranges from 280 to 400nm (Mocz, 2007).

New and improved FPs have been produced that are brighter with enhanced photostability, reduced oligomerization, pH insensitivity and cover a broad spectral range (Shaner et al., 2007). Variants modified from wild type GFP emit blue, cyan and yellow regions (Heim et al., 1994; Ormo et al., 1996; Tsien, 1998). Red FP was harder to achiever but was finally produced from a nonbioluminescent reef coral (Matz et al., 1999).

Cloning of many FP genes serves as a platform for construction of chimeric proteins in *in vivo* and *in vitro* experiments (Tsien, 1998). Natural and mutant FPs fused to the target of interest can function as genetically encoded fluorescent markers or fusion tags in various applications. They can trace cell lineage, act as gene expression reporters,

enable visualization of physiological processes as well as monitor subcellular protein localization (Mocz, 2007).

Development of light-controlled fluorescent proteins called photoactivatable fluorescent proteins (PAFP) such as Kaede, mEosFP, Dendra, KFPs, Dronpa and others with spectral properties responding to irradiation are being widely used to study protein mobility. Under certain specific wavelengths of light and intensity irradiation, the PAFP is able to change its fluorescent properties (Lukyanov et al., 2005). Some PAFP are able to self-convert from a low or non-fluorrescent to a bright fluorescent state which is known as photoactivation while photoconversion or photoswitching occurs in some PAFP that change their fluorescent colour. Based on their mechanism of activation PAFP are divided into three classes; 1) irreversible (off to on), 2) photoconversion (wavelength shift) and 3) reversible (photoactivated).

In the first class, fluorescence is initiated by emission from a quiescent or dark state. For example, PA-GFP comprises a photoactivatable variant of avGFP, a green fluorescent protein from jellyfish *Aequorea victoria* (Patterson and Lippincott-Schwartz, 2002). One of its residues is substituted from histidine to threonine at position 203 (Thr203His) to produce the neutral chromophore form (Patterson and Lippincott-Schwartz, 2002) and is distinguished by excitation at 400nm with a maximum emission of 515nm. Excitation at 480-510nm produced almost no fluorescence even though the range coincides with the excitation spectrum of the anionic chromophore form. However, irreversible photoconversion from the neutral to the anionic form can be achieved by intense violet irradiation at 400nm. PA-GFP absorbs at 504nm and fluoresces at 517nm resulting in 100-fold increase in green fluorescent of the anionic form (Lukyanov et al., 2005). Another example is PAmRFP1, a red photoactivatable variant of DsRed from *Discosoma* sp. Compared to mRFP1, it has residue substitutions at positions 148, 165 and 203. In its non-photoactivatable state, weak cyan fluorescence is produced. Irreversible photoactivation can be achieved at 380nm giving a 70-fold increase in red fluorescent within the excitation at 578nm and emission maximum of 605nm (Lukyanov et al., 2005).

Photoconversion occurs from one fluorescent emission bandwidth to another. It allows PAFP in this second class to be tracked before photoconversion for ease of

selection of specific regions to photoactivate. Examples include PS-CFP (photoswitchable cyan fluorescent protein), Kaede, Dendra2, EosFP, mEOS2, KikGR and mKikGR (Chudakov et al., 2004; Lippincott-Schwartz and Patterson, 2009). All of PAFP in this class are derived from stony corals (Ando et al., 2002; Tsutsui et al., 2005; Gurskaya et al., 2006). UV light irradiation causes an irreversible peptide cleavage near a chromophore that initially emits green fluorescence. The electron conjugation system is subsequently extended to give red fluorescence at a new emission peak (Mizuno et al., 2003; Nienhaus et al., 2005). Meanwhile, the last class of PAFP have a switching mechanism whereby photoconversion is reproducible. Its initial green fluorescent form can be switched to a non-fluorescent form with continuous irradiation and applying initial light emission will change the PAFP back its fluorescent state (Lippincott-Schwartz and Patterson, 2009).

PAFPs enable objects to be tracked without constant visualization with three main levels of application: cell, organelle and protein. In all of these cases, movement rate and direction remain the primary parameters of tracking and detection. Tracking a photolabeled PAFP fused to a protein allows determination of rates and preferred directions of its movement inside living cells by timing its motion. Not only that, the rate of protein movement in various organelles can be measured and compared by targeting the PAFP to a particular organelle (Chudakov et al., 2007). Detailed information about protein localization, turnover and mobile and immobile fractions can also be obtained (Lippincott-Schwartz and Patterson, 2003).

1.14 Dendra2

Dendra2 is an improved version of commercially available Dendra (dendGFP), originally cloned from the soft coral *Dendronephthya* sp. (Labas et al., 2002). It is monomeric and able to photoconvert by blue light at 488nm (Gurskaya et al., 2006). A green-to-red convertible protein, Dendra2 has a single substitution at position 224 from valine to alanine (V224A). The result is a more complete chromophore formation and brighter fluorescence before and after photoswitching. The difference between Dendra2 and other PAFP is that it allows the use of blue light for activation hence potentially damaging UV irradiation is not necessary (Chudakov et al., 2007).

In its non-activated state Dendra2 excites and has an emission maxima at 490nm and 507 respectively. Due to this range the 488nm laser line commonly installed on all commercial confocal microscopes can be used for activation of primary visualization. Excitation at 488nm with laser or a blue activating light does not activate photoactivattion when used at short pixel dwelling time and only excites the green fluorescence. To activate photoconversion, excitation at shorter wavelength peak using UV irradiation would result in a very efficient conversion of Dendra2 into red. Another option is to irradiate Dendra2 with intense blue light that corresponds to its absorption peak at 490nm (Chudakov et al., 2007; Zhang et al., 2007). This is because the absorption spectrum peaks of non-activated Dendra2 are between 386 and 490nm which coincides with neutral and anionic GFP-like chromophore state. Selective excitation of red Dendra2 in fluorescent imaging can also be achieved by excitation above 510nm and, both green and red can be excited efficiently between 450 and 500nm (Adam et al., 2009).

Application of Dendra2 as a fusion marker is aided by monomerisation due to three point mutations; N121K, M123T and Y188A, (Adam et al., 2009). This feature allows selective visualization and in *Nicotiana benthamiana*, selective photoconversion of subnuclear regions of interest was reported without disturbing other nuclei in the adjacent cells (Martin et al., 2009). An attempt to use similar fusion protein (RSZp22-Dendra2) as Martin et al was carried out in *Arabidopsis* but without the same result as detection of fusion protein was unsuccessful at 488nm blue light or intense UV light irradiation (Rausin et al., 2010). However, Dendra2 has been used in various other species apart from plants (Mocz, 2007; Zhang et al., 2007; Arrenberg et al., 2009). A new opportunity to study fusion protein movement in plants may be the next step following the success of such work in *Nicotiana benthamiana*.

1.15 Aims of Thesis

This study aims to generate an inventory of phloem exudate proteome to provide an essential data resource for future study of *Arabidopsis* phloem. FT or its homolog had been discovered in several studies in other plant phloem proteomes (Giavalisco et al., 2006; Lin et al., 2009; Aki et al., 2008; Rodriguez-Medina et al., 2011) and an assumption was made that the same protein could also be found in the *Arabidopsis* phloem proteome. Thus,

studying the *Arabidopsis* phloem proteome and protein mobility in addition to FT should facilitate understanding the relationship between systemic signalling and regulation of flowering time. The overall objectives of the study are:

1. To generate an analyzed database of the *Arabidopsis* phloem exudate proteome.

Phloem exudate proteome has been studied in *Brassica*, cucurbit, rice, melon, pumpkin and white lupin (Giavalisco et al., 2006; Walz et al., 2004; Omid et al., 2007; Aki et al., 2008; Lin et al., 2009; Rodriguea-Medina et al., 2011). All previous phloem proteome studies addressed identification of FT (FLOWERING LOCUS T), its homologs or ortholog TSF (TWIN SISTER OF FT). Recently *Arabidopsis* phloem exudate protein studies (Guelette et al., 2012; Batailler et al., 2012) were reported. Therefore this study is to provide additionall data resources and compilation of phloem exudates protein from all reports. The presence of native FT was detected via MRM. Meanwhile, many other proteins with diverse function in the phloem of *Arabidopsis* were detected using LC-MS/MS (liquid chromatography-tandem mass spectrometry) experimentation and spectra were searched using MASCOT and ProteinPilot against *Arabidopsis* database.

2. To provide preliminary quantification of the level of FT protein in the phloem exudate in response to floral-inductive photoperiod treatments.

Arabidopsis growth can be influenced by photoperiod where LD (long day) induces flowering. The highest level of *FT* mRNA was documented by Corbesier et al. (2007) after 3 days transition from SD (short day) to LD (77 hours) as compared to 2 days (54 hours). Nevertheless, the question as to whether the level of FT protein in the phloem sap increases as a result of SD to LD transition is still unclear. In order to detect FT, specific FT-derived tryptic peptides were developed for targeted MS/MS using MRM (multiple reaction monitoring).

3. To define the sites of FT accumulation and provide visualization of routes of movement.

Grafting experiments have shown that FT protein is transmissible across graft junctions from a donor to a recipient plant. Movement of FT-GFP (green fluorescent protein) fusion protein was detected at the apex of the SAM (shoot apical meristem) (Corbesier et al., 2007, Jaeger and Wigge, 2007, Notaguchi et al., 2008) but the movement is only detectable transiently for 24-48 hours after induction (Notaguchi et al., 2008). Although well demonstrated, the accumulation and movement route of FT remain to be investigated. The aim of the study is to determine FT accumulation sites and spatial distribution at the shoot apex. An attempt was made to provide visualization of FT movement routes by tracking its movement using Dendra2, a Photoactivable Fluorescent Protein (PAFP) to detect translocation of protein.

CHAPTER 2
Arabidopsis PHLOEM PROTEOMICS – METHOD DEVELOPMENT

2.1 Introduction

Method development is the establishment of analytical method(s) that will be suitable and applicable for the analysis intended. It plays an important role in the discovery, development and manufacturing processes. The result from these processes can then be used as a guideline for further work. Depending upon the type of method being developed, the steps could commonly involve method development, plan definition, background information gathering, laboratory method development and generation of test procedure(s) (Breaux et al., 2003). Testing of all possible parameters is necessary including scientifically justified and logical step-by-step experimental approaches.

Therefore the aim of this study was to develop suitable methods to study *Arabidopsis* phloem proteome. Combination of established methods in screening new lines of mutant for use in proteomics work and optimising previous protocol that has been used in other plants was conducted to suit this study. Four mutants already in collection for intended use in proteomics work had been determined. However, one of them (*co)* is in a different background (Landsberg erecta) while the rest are in Columbia. Using all mutants from the same background as the WT is necessary to eliminate uncertainty that can arise from variation contributed by different ecotypes. Moreover, Columbia is the most widely used background and is the ecotype sequenced in the *Arabidopsis* Genome Initiative. In addition to replacing one of the existing mutants, it would be advantageous if new line(s) of mutant can be obtained from the screening process whether for use in this study or for future work.

It is important to justify the protocols for proteomics which is one of the main aims. This is because good quality and a large amount of proteins are needed as the method used to obtain protein is by wounding. Sufficient protein without or with less contaminant that could cater for the whole phloem proteome and be used to reproduce would be beneficial to ensure a satisfactory result. Unlike other phloem proteome studies such as in *Brassica* and *Cucurbita* where a large amount of pure phloem protein could be

obtained by wounding plants with a hypodermic needle, small plants such as *Arabidopsis* require an exudation technique in which wounded plants are soaked into a chelating agent to collect the phloem sap. Optimization in terms of chelating agent concentration and time among other factors played important role to ascertain sufficient amount and good quality protein for analysis. Method development for 2D proteomics was also attempted utilising isoelectric focusing (IEF) method from brassica for comparison with that of 1D proteomics.

Developing these methods and optimising the necessary parameters to ensure good quality, ease of use and reproducibility would be advantageous especially for proteomics work. Further work following the outcome of this study could be accelerated using methods developed not only to identify certain protein but also to take *Arabidopsis* phloem proteomics work one step further particularly in relation to systemic signalling and regulation of flowering.

2.2 Materials and Methods

2.2.1 Mutant Screening

2.2.1.1 Plant Growth and Conditions

Six lines of T-DNA insertional mutant (genotype; *CO*[Exon] (N870084), *TSF*[Intron] (N663213), *TSF*[Promoter] (N662738), *FT*[exon] (N876735), *SOC1*[Intron, SAIL lines] (N877250) and *SOC1*[Intron, SALK line] (N657480)) bought from Nottingham *Arabidopsis* Stock Centre were initially grown under LD condition (16h light, 160 µmol) with *Col-0* as control. Seeds were grown on soil to vermiculite ratio of 4:1 and kept at 4^{o}C for 2-3 days. Subsequent growth was done in growth chamber in SD condition at 23^{o}C with 10h light (160µmol) and 60% humidity. Plants that need to be grown in LD condition were exposed to light for 16 h with the same intensity and humidity.

2.2.1.2 Phenotypic Observation

Selection for each line was based on the location of its T-DNA insertion especially those in the exon followed by promoter and intron regions. *co* was chosen to replace the existing *co* already in collection but not in Columbia background, both *tsf* mutants in

addition to *ft* were chosen to in the hope to find a better mutant as compared to the ones in collection that can provide late flowering. On the other hand, *soc1* was chosen as *SOC1* is a gene associated with flowering produced by *FT* hence its protein content would be worth to investigate. Subsequent individual plants selected based on late flowering (with an exception to *tsf* [promoter] where early flowering plants were chosen) from each line bearing seeds after 5 to 10 weeks were then grown back in SD and LD conditions. Flowering time was determined by rosette leaf count of plants after flowering.

2.2.1.3 Genotyping via PCR

Genomic DNA was extracted using Wizard® genomic DNA Purification Kit from Promega as instructed by manufacturer's recommendation. Concentration of DNA was determined using nanodrop spectrophotometer (Nanodrop ND-1000 Spectrophotometer by Nanodrop Technologies) and its quality was assessed by running a 1.2% (w/v) agarose gel electrophoresis.

.

2.2.1.3.1 PCR

PCR amplification was performed with two primer combination mixes for all genotypes. The first mixture had 10 µL GoTaq buffer (Promega), 1 µL 10 mM dNTP, 0.25 µL *Taq* polymerase, 1 µL Left Primer (10 mM), 1 µL Right Primer (10 mM), 1 µL DNA and 34.75 µL dH_2O. For the second mixture, the same components were used except that the Left Primer is not included and replaced with Left Border of the T-DNA insert. A list of the primers used is shown in Table 2.1. The PCR reaction began with initial denaturation at 94^{o}C for 5 min followed by 40 cycles of 94^{o}C for 30 seconds, 55^{o}C for 1 min and 72^{o}C for 1 min 20 seconds followed by a final extension of 72^{o}C for 10 min. All PCR products were run on 1% agarose gel (1% (w/v) agarose (Sigma) in 1X TAE buffer (40mM Tris-Base, 40mM glacial acetic acid, 1mM EDTA) for 45 min at 80V.

Table 2.1: List of primers used for PCR reaction. Primers for *TSF* (Intron) and *FT* were not included as they were omitted out from the screening process after phenotyping. P=promoter.

Gene and Genotypes	Left Primer (LP)	Right Primer (RP)
CO	CCAGTTTCCATGGATGAAATG	CCCCTTCTTTCAGATACCAGC
TSF (P)	TGAGGCTACGAAAGAAAAGAATG	TCCATTGTTTTAAAAGTTTTACTGG
SOC1 (SALK)	TGGACCAAATGTCTGAAACATC	GAAGAAGATATGGTGAGGGGC
SOC1 (SAIL)	TCATCTTGTTTAATCATCTGTCTCTC	TTTCTTCATCATGTTTGCTGC
Left-Border (LB) Primer for SALK lines	TGGTTCACGTAGTGGGCCATCG (LB1) GCGTGGACCGCTTGCTGCAACT (LB2)	
Left-Border (LB) Primer for SAIL lines	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC (LBS1) GCTTCCTATTATATCTTCCCAAATTACCAATACA (LBS2)	

2.2.1.3.2. RNA Extraction

RNA extraction was also performed using TRI Reagent (Sigma). 50-100 mg tissue was frozen and ground to a fine powder. 1 mL of TRI Reagent was added and the sample was let stand for 5 min at room temperature followed by addition of 0.2 mL chloroform and vigorous shaking for 15 s before letting it stand at room temperature for 15 min. The mixture was then centrifuged at 12,000 x *g* for 15 min at 4^{o}C. Three phases separated the mixture after centrifugation comprising of upper aqueous phase (containing RNA), an interphase (containing DNA) and a red organic phase at the bottom (containing protein). The aqueous phase was transferred to a fresh tube and 0.5 mL of isopropanol was added and mixed. The mixture was allowed to stand for 10 min at room temperature before centrifuging it at 4^{o}C for 10 min at 12,000 x *g*. The supernatant was removed and the RNA which was precipitated into a pellet form was washed with 1 mL of 75% (v/v) cold ethanol. The sample was vortexed briefly and centrifuged at 7,500 x *g* for 5 min at 4^{o}C. After discarding the supernatant, the RNA pellet was air-dried at room temperature for 5-10min but not to a complete dry. The RNA pellet was dissolved in an appropriate amount (about 40-50 µL) of TE buffer (10mM Tris-Base pH8.0, 1mM EDTA) by gentle pipetting. The RNA was then stored at -80^{o}C prior to use.

2.2.1.3.3 Synthesis of First Strand cDNA

500 ng of total RNA was used and mixed in an eppendorf tube with 4 µl 2.5mM dNTP mix, 2 µl of Oligo dT and made up to a final volume of 16 µl each with dH_2O. The mixture was mixed thoroughly and spun briefly before heating it up at 70°C for 3 min. The tube was transferred to ice immediately and several reagents were added such as 4 µl 5X Reverse Transcriptase (RT) buffer, 1 µl Placental RNase Inhibitor and 1 µl Superscript II RT. The content of the tube was mixed and spun briefly before incubating it at 40°C for 1 hour. The reaction was stopped by heating it up at 95°C for 10 min. The final product could be stored in –20°C.

2.2.1.3.4 rt-PCR Analysis

Reverse-transcriptase PCR was achieved by preparing a mixture containing 10 µL GoTaq Flexi buffer (Promega), 1 µL 10mM dNTP, 3 µL 25mM $MgCl_2$, 0.5 µL Taq polymerase (5U/µL), 1 µL Forward Primer (10mM), 1 µL Reverse Primer (10mM), 2 µL cDNA and 33.5 µL dH_2O. PCR cycle condition was as followed; 94°C for 4 min, 40 cycles of 94°C for 30s, 55°C for 1 min (annealing temperature varies), 72°C for 1 min 20s and a final extension at 72°C for 7 min. Products were run for 45 min at 80V on 1% agarose gel.

Table 2.2: List of primers used for rt-PCR reaction.

Genotypes	Forward Primer	Reverse Primer
co	TGATAAGGATGCCAAGGAGG	ATCGTGTTGAACCCTTGCTC
tsf (P)	CCTTTCACGAGGTTGGTCTC	AGTAAGAGGCAGCCACAGGA
soc1 (SALK)	AAAGTCTTGTACTTTTTCCCCC	TGGAATCACATCTCTTTGACG
soc1 (SAIL)	TGCTCTTTCGTAGCCAATTTACAC	TTGGTGCTGACTCGATCCTTAG

2.2.2 1D Proteomics

2.2.2.1 Exudation Technique

Prior to exudation, the 6 weeks-old plants which were previously grown in SD condition were exposed to LD condition for 3 days (Corbesier et al., 2007). In the initial method development experiments, after 77 hours exposure to LD, the petiole was cut close to the base (Fig 2.1 (a)). Up to 7 leaves was then arranged and rinsed in 10mM Na_2-EDTA pH 7.0, dabbed onto clean paper towel and soaked into 10mM Na_2-EDTA pH 7.0 (Fig 2.1 (b)) overnight in LD condition. The collected exudates were then frozen in liquid nitrogen, and concentrated with Amicon filter device (Amicon Ultra-15 Filters Ultracel-3 by Millipore) with 3kDa cut-off and stored at -20^{o}C prior to use.

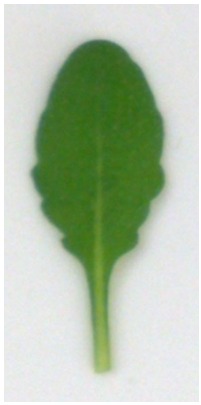
(a)  (b) 

Figure 2.1: Mature and healthy leaves of 6 weeks old plants were used for exudation where (a) it is cut at the base of the petiole and (b) up to 7 leaves were located into 1.5 mL tube filled with 10 mM Na_2-EDTA.

Later experiments were conducted as per initial ones with a slight modification. The leaves were first immersed in 20mM Na_2-EDTA for 30min in LD and later transferred to a fresh solution for overnight exudation. Both proteins obtained from the first and second collections were ultracentrifuged at 100,000 rpm for 1h at 4^{o}C before being purified and concentrated.

For optimization purpose, these parameters were explored; 1) comparison of exudates from SD and LD plants, 2) concentration of EDTA (10 mM and 20 mM), 3) incubation period (3 hours and overnight), 4) LD exposure (2 days and 3 days), 5)

Incubation condition (SD and LD), 6) plants' age (6 weeks and 7 weeks) and 7) purification method (acetone precipitation and using Amicon filter device [Millipore]).

2.2.2.2 Protein Purification

2.2.2.2.1 Acetone Pecipitation

Acetone precipitation was carried out by adding 4 volumes of cold acetone (cooled to -20^{o}C) to one volume of protein solution. The mixture was vortexed briefly and incubated in -20^{o}C for 1 hr or overnight. Following incubation, the mixture was centrifuged at 13,000 rpm for 10 min at 4^{o}C. The supernatant was discarded and access acetone in the tube was allowed to evaporate at room temperature. Care was taken not to over-dry the pellet to avoid problem in dissolving later on. Finally, appropriate amount of 2X SDS-PAGE buffer (0.5M Tris pH 6.8, 20% (v/v) glycerol, 4.4% (w/v) SDS, 2% (v/v) 2-mercaptoethanol and 6% (w/v) bromophenol blue) was added and the pellet was dissolved thoroughly by vortexing.

2.2.2.2.2 Purification using Millipore Filter Centrifuge

The filter centrifuge was used to purify and concentrate protein as per manufacturer's instruction (Millipore). Proteins were transferred into Millipore Filter Centrifuge (Millipore) with a cut-off of 3kDa to minimize the loss of smaller mass proteins. Tubes were centrifuged at 4,000 rpm for 45 to 60 min at 16 to 20^{o}C in a swinging rotor centrifuge (Jouan BR4). Protein solution on the reservoir of the filter was transferred into new 1.5 mL siliconized tube (Sigma) and stored in -20^{o}C prior to use.

2.2.2.3 Bradford Assay

To measure concentration of protein, Bradford Assay was carried out with BSA (bovine serum albumin) as standard. A series of BSA concentration (0 as blank, 1, 2, 4, 6, 8 and 10 μg) was prepared from the stock to a total volume of 0.5 mL. At the same time protein samples were diluted approximately 100x in the same total volume. The same amount of Bradford Reagent (Sigma) was added (0.5 mL) to the standard and protein samples and mixed by shaking the tubes. Incubation took place immediately where the

tubes were left at room temperature for 20 up to 45 min to allow the reaction to occur between protein and Bradford Reagent. Absorbance of the reaction was measured at 595nm using spectrophotometer (Biochrom WPA Lightwave II UV/Visible Spectrophotometer) against the blank. The measurements taken from the standard was plotted against each BSA concentration. Each sample concentration was then determined from the standard graph, taking into account the dilution imposed on each sample.

2.2.2.4 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

To check the quality of protein obtained, 18% PAGE (18% separating and 5% stacking gel: 30%-Bis-acrylamide, 1.5M Tris-HCl pH 8.8 for separating gel and 0.5M Tris-HCl pH 6.8 for stacking gel, 10% SDS, 10% (w/v) APS, TEMED) was done with approximately 5-10μg proteins. An initial 100V for 15min was applied to run the gel followed by 150V for 70min where the gel was later stained overnight in Coomassie stain (Fisher Scientific) and destained for 2X1hr in 30:10 MeOH: acetic acid before a final rinse in deionized H_2O for 1hr.

Silver staining was also employed for comparison with Coomassie staining. Gel was fixed with 40:10 MeOH: acetic acid for 20 min, washed 3X 20min with 30% EtOH, soaked in reductant solution (0.2 mg/L sodium thiosulfate) for 1 min and washed with deionized water for 3X 30 sec. The gel was then soaked for 20 min with silver stain (2 g/L silver nitrate, 0.02% (v/v) formaldehyde) followed by 3X 30 sec wash in deionized water. Next, developer (30 g/L sodium carbonate, 5 mg/L sodium thiosulfate, 0.05% (v/v) formaldehyde) was used and the gel was left in the solution until bands became visible. The last washing step with 3X 30 sec deionized water followed by 1-2 min treatment with stop solution (5% (v/v) acetic acid) completes the process.

2.2.2.5 Trypsin Digestion

2.2.2.5.1 In-Gel Digestion

2.2.2.5.1.1 In-Gel Digestion of Coomassie Stained Gel

Corresponding bands (Fig 3.7) were cut from gel and each of them was diced into smaller pieces of about 1-2mm^2. Approximately 100μL of 25mM NH_4HCO_3/50% ACN

was added to cover the diced gels and vortexed for 10 min. The supernantant was discarded and addition of 25mM NH_4HCO_3/50% ACN followed by vortexing were repeated twice before the gel pieces were dried completely using a speed vacuum for about 20 min. 10mM DTT (in 25 mM NH_4HCO_3) was added to cover the gel, vortexed, spun briefly and incubated at 56^oC for 1 hr. Following incubation, 55 mM iodoacetamide was added to the gel pieces, vortexed, spun briefly and incubated in the dark at room temperature for 45 min. After that, the supernantant was removed and the gels washed with 25mM NH_4HCO_3, vortexed for 10 min and spun briefly before the supernantant was removed. Gels were dehydrated with 25mM NH_4HCO_3/50% ACN, vortexed for 5 min and spun. Dehydration step was repeated once prior to complete drying of the gel with a speed vacuum. Next, trypsin solution (12.5ng/μL trysin in 25mM NH_4HCO_3) was added to cover the gel pieces followed by rehydration of the gels on ice or at 4^oC for 10 min. The samples were spun briefly and 25 mM NH_4HCO_3 was added to completely cover the gel pieces. After a brief spin, the samples were incubated overnight at 37^oC.

The digest solution was transferred into a clean 1.5 mL tube and approximately 30μL 50% ACN/50% formic acid was added to the gel pieces, vortexed 20-30 min, spun and sonicated for 5 min. The solution was transferred into the tube containing the overnight digest solution and the solution was vortexed and spun before the volume was reduced to approximately 10 μL with the speed vacuum. The samples were then cleaned with C18 Zip Tip (Millipore) by rinsing the tips 5X each in ACN, 1% formic acid, samples and 0.1% formic acid followed by elution in 10 μL 0.1% formic acid/50%MeOH and vacuum dried for 15 min. Prior to LC-MS, 10 μL of 3% CH_3CN/0.1% formic acid was added to resuspend the samples.

2.2.2.5.1.2 In-Gel Digestion of Silver Stained Gel

To the diced bands, 30 μL 30mM $FeK_3(CN)_6$: 100 mM $Na_2S_2O_3$ (1:1) was added and the gels soaked in the solution for 10 min with vortexing in between. The solution can be decanted and the step repeated until gel bands are clear. After removal of the destaining solution, the gel pieces were washed three times with 100 μL water followed by vortexing for 15 min each. Once the water was removed, the gel pieces were washed 3X with 100 μL

25 mM NH_4HCO_3/50% ACN and vortexed for 15 min each. Next, the solution was removed and the gel pieces vacuum dried for about 15 min. 10 mM DTT (in 25mM NH_4HCO_3) was added to the dried gel pieces, vortexed, spun briefly and incubated for 1 hr at 56^oC. The supernatant was removed and 55 mM iodoacetamide was added, vortexed, spun and incubated for 45 min at room temperature in the dark. After removal of the supernatant, the gel pieces were washed with 100 µL 25 mM NH_4HCO_3 and vortexed for 10 min. Following removal of supernatant, the gel pieces were dehydrated with 100 µL 25mM NH_4HCO_3/ACN, vortexed for 5 min, spun briefly and discarde. The last step was repeated once and dried completely. Trypsin solution (12.5ng/µL) was added to the samples, rehydrated on ice for 10 min, covered completely with 25 mM NH_4HCO_3 and incubated overnight at 37^oC.

The solution from trypsin digestion was transferred to a new tube and 5% formic acid/50% ACN was added to the gel pieces. The gel pieces were vortexed 15 min, spun and sonicated for 15 min. The solution was then transferred to the same vial containing the trypsin digestion solution, vortexed, spun and the volume was reduced around 10-20 µL by vacuum drying. The C18 Zip Tip step proceeded as in 2.5.1.1.

2.2.2.5.2. In-Solution Digestion

Proteins volume (30 µg) was increased to 100 µL with 50mM NH_4HCO_3 and 5 µL 200 mM DTT (in 100mM NH_4HCO_3) was added to reduce the sample by boiling it for 10 min followed by incubation at room temperature for 45-60 min. Alkylation was done by adding 4 µL 1M iodoacetamide to the samples, vortexed, spun briefly and incubation at room temperature for 45-60 min. Neutralization of iodoacetamide proceeded by the addition of 20 µL 200 mM DTT, vortexed, spun and incubation at room temperature for 45-60 min. Trypsin digestion was carried out in a ratio of 1:50 to 1:20 of trypsin to sample where the samples were vortexed and spun briefly prior to overnight incubation at 37^oC. 2 % formic acid was added to the sample until the pH reached to pH 6.0 using a pH paper indicator. Samples cleanup with C18 Zip Tip followed as in 2.5.1.1.

2.2.2.6. Mass Spectrometry

The second phase was to compare the phloem proteome profiles. Measurement was direct LC-MS analysis. Total proteins (20-50μg) and bands of proteins from PAGE analysis were subjected to in-solution and in-gel digestion respectively prior to LC-MS-MS. Gel bands were run using Agilent 1100 system nanoflow LC where 8μl injection was loaded onto a trap column. From there, the samples moved to C18 nanoflow analytical column which was washed isocratically with 2.5% (v/v) acetonitrile and 0.1% formic acid. The gradient took approximately 59 min at a flow rate of 0.30 μL/min from 5% acetonitrile and 0.1 % (v/v) formic acid to 95% acetonitrile and 0.1% formic acid in water. The analysis was run at 40^{o}C. A linear IT (ion trap) quadrupole mass spectrometer (QTrap, AB Sciex Instruments) with an acquisition program for the four most intense peaks set at *m/z* between 400 to 1500 and charge state +2 to +3 were chosen for collision induced dissociation (MS/MS). Rolling collision energy was applied. For the total proteins, a similar procedure was employed but using capillary flow LC (gradient at approximately 300 min, flow rate 7.00μL/min) and the analysis run at 60^{o}C.

2.2.3 2D Proteomics

2.2.3.1 Protein Extraction

Arabidopsis wild type plants, *Col-0* were grown on ½ MS supplemented with 1% (w/v) sucrose and left in LD for 3 to 4 weeks. Young leaves were harvested for total protein extraction using both Plant Total Protein Extraction Kit (Sigma) according to the manufacturer's instruction and manually.

For manual extraction, 500 mg of leaves were ground to a fine powder in liquid nitrogen using mortar and pestle. The powder was then transferred to a clean frozen 1.5 mL tube and 1 mL of protein extraction buffer (0.1M Tris-Base pH 7.0) was added. The mixture was vortexed for 30s to 1min before centrifugation at 4^{o}C for 15min at 13,000 rpm. The supernatant was carefully transferred to a new clean 1.5 mL tube and 1 mL extraction buffer was added, vortexed and spun as in the initial step. After spinning, the

supernatant was transferred to a new tube and ultracentrifuged at 4°C for 1h at 100,000 rpm. Prior to use, the protein extract was kept at -20°C.

2.2.3.2 2D Clean-Up

500 µg of protein (total leaf and phloem sap protein) in a final volume of 100 µL was used and cleaned up according to manufacturer's instruction (Bio-Rad). Prior to use, 5% (v/v) of IPG buffer pH 3-10 (GE Healthcare) was added to the sample.

2.2.3.3 Isoelectric Focusing (IEF) and 2-Dimentional Gel Run

Up to a week prior to isoelectric focusing, 11cm non-linear IPG strip pH 3-10 was rehydrated. Rehydration solution (GE Healthcare) was mixed with 0.5% (v/v) IPG buffer pH 3-10 (GE Healthcare) and applied onto the strip holder. The transparent cover tape was peeled off from the IPG strip and it was put on top of the rehydration buffer in the sticky-side down condition. After ensuring any bubbles were eliminated, the strip was overlaid with Drystrip Cover Fluid (GE Healthcare) and left at room temperature. Overnight, the sticky side of the strip will form a gel layer following absorption of the rehydration solution.

The aluminium tray of the IGPhor machine (GE Healthcare) was layered with Drystrip Cover Fluid. The strip was then placed gel-side up on the tray. Electrode paper blotter was wet with dH_2O and put on top of the gel at each positive and negative ends of the strip. On top of the blotter, the electrodes were placed and clipped to hold the stacks in place. A sample cup loader was placed on top of the strip at the positive end and was pressed gently down. Cover fluid was poured into the cup to completely cover the gel and the sample was loaded into the cup. Next, the rest of the strip was completely covered with the cover fluid. Focusing was carried out at 100V for 1h, 300V for 1h, 1000V for 23h, 1500V for 30min and 2000V for 10min.

Once focusing was completed, the strip was equilibrated in SDS equilibration buffer (6M urea, 75mM Tris-HCl pH 8.8, 29.3% (v/v) glyverol, 2% (w/v) SDS, 0.002% (w/v) bromophenol blue) with 10 mg/mL DTT with gentle shaking for 15min at room

temperature followed by another 15min gentle shaking at room temperature with SDS equilibration buffer with 25 mg/mL iodoacetamide. Prior to running the gel, the strip was dip briefly in the SDS electrophoresis buffer (25mM Tris-Base, 192mM glycine, 0.1% (w/v) SDS) before placing the strip with its plastic side backing the glass plate. While pushing the strip down towards the gel, air bubbles were avoided and agarose sealing solution (25mM Tris-Base, 192mM glycine, 0.1% (w/v)SDS, 0.5% (w/v) agarose, 0.002% (w/v) bromophenol blue) was used to seal the strip to the gel cast. Once protein marker has been loaded, the gel (4-16%) was run at 100V for 4h. After completion of the electrophoresis the gel was stained in Coomassie stain.

2.3 Results

2.3.1 Mutant Screening – Phenotypic Analysis

Small populations (*n*=12) were used for the initial screening experiment as there were not enough seeds at the beginning of the work. However the problem was overcome by increasing the seed population from initial planting. The wild type plants started to flower at 4 weeks and number of rosette and cauline leaves for each plant was scored. From initial screening of the 6 lines, *tsf* plants supposedly carrying the T-DNA insertion within promoter (1000-UTR5') was found to flower very early, within 3 weeks with an average leaf count of 4 to 6 leaves (Fig 2.2). The other *tsf* line carrying T-DNA insertion at its intron site and *ft* flowered normally as the wild type. The two *soc1* lines with T-DNA insertions in intron gave contrasting results. *soc1* (SALK) gave an average leaf count of 7 to 9 leaves with the latest mutant plant giving 11 leaves. On the other hand, *soc1* (SAIL) produced 3 out of 13 late flowering plants at 18 and 19 leaves respectively and average much later than *Col-0*. One *co* plant which had a T-DNA insertion in the exon was the latest flowering of all lines, at 20 leaves and other individuals were also later than *Col-0*.

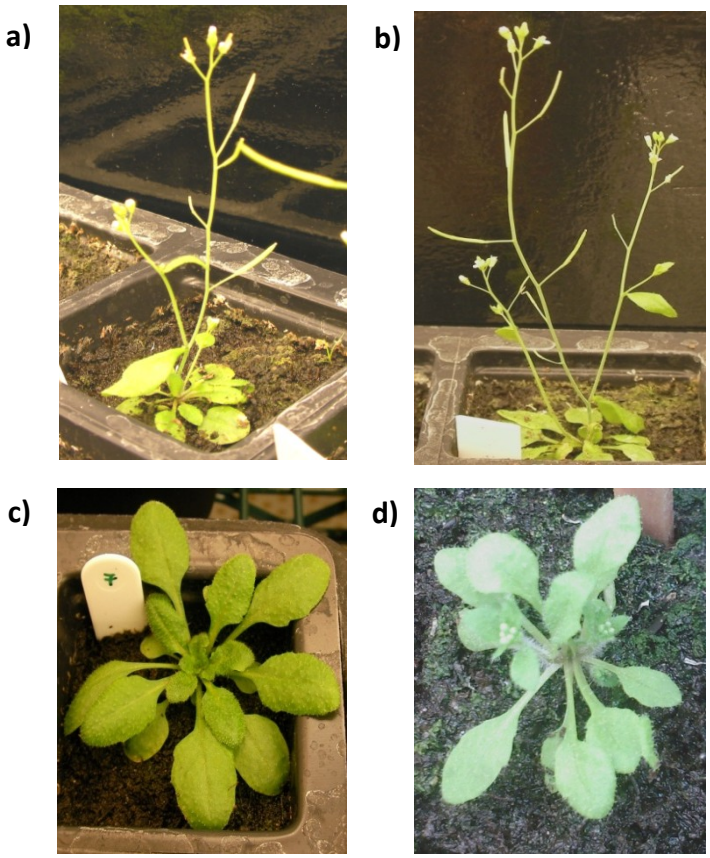


Figure 2.2: Wild type and mutant plants grown in SD and LD. All SD and LD plants were the same age; 5 weeks. A) WT/LD, b) *tsf* (P)/LD, c) WT/SD and d) *tsf* (P)/SD. WT-wild type, SD-short day, LD-long day, P-promoter.

The result of flowering time for each line based on leaf count under LD is represented in Figure 2.3. From the graphs, *tsf* (intron), *ft* and *soc1* (SALK) showed wild type phenotype while *tsf* (promoter) showed early flowering phenotype and, *co* and *soc1* (SAIL) exhibited late flowering phenotype and thus are expected to be homozygous. *co* and *soc1* (SAIL) also showed characteristics of non WT, heterozygous and homozygous in their population.

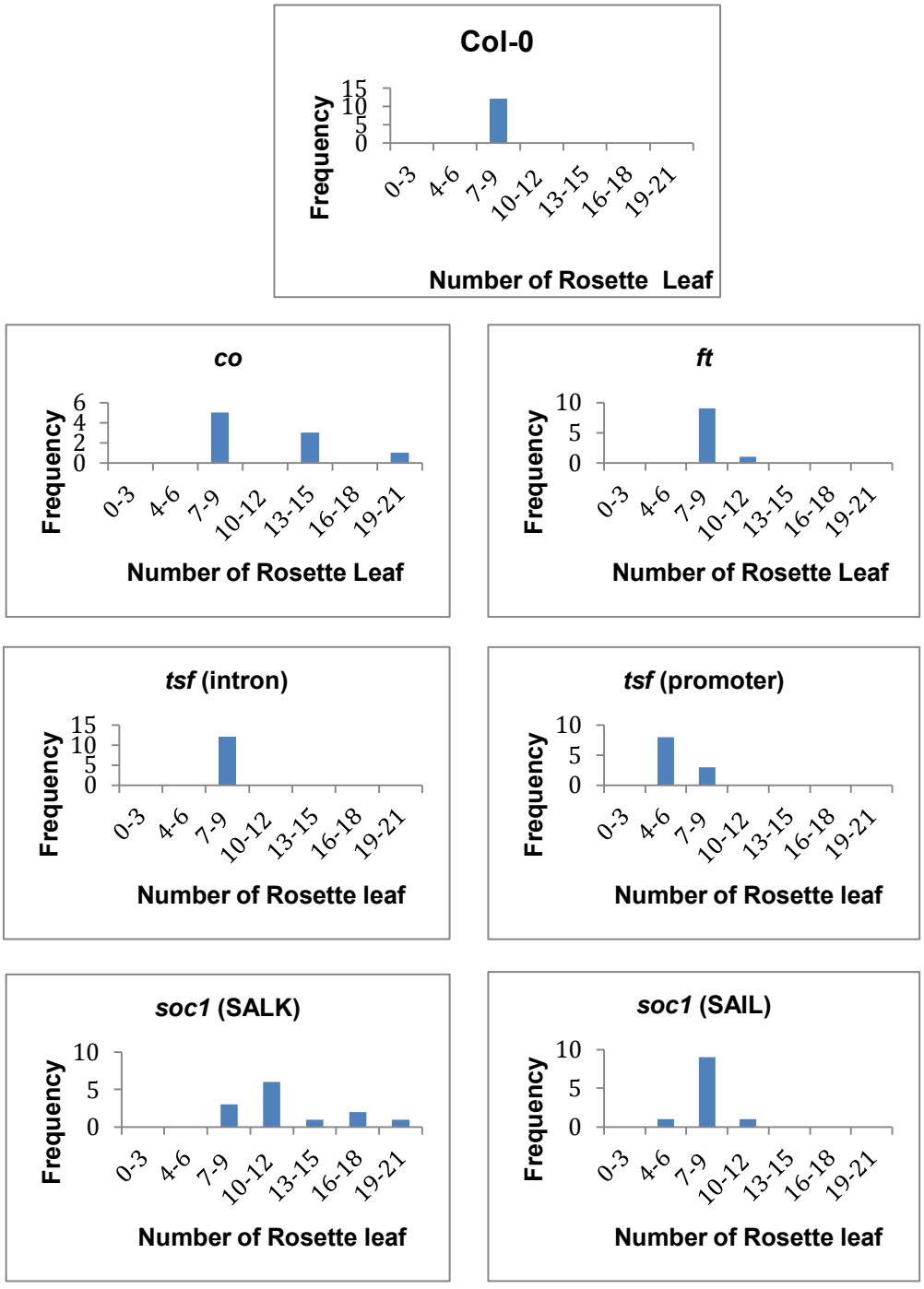


Figure 2.3: Mutant rosette leaf count for all mutants and wild type in the initial screening. All plants were grown in LD. Determination of flowering time by rosette leaf count should be done before the plants start producing flower. This is because as the first bud of flower starts to form, rosette leaves will stop its production. LD-long day.

From the six lines, only *co*, *tsf* (promoter) and both *soc1* were chosen for further work. *ft* and *tsf* (intron) were not chosen as they were showing WT phenotype and not showing any late flowering characteristic especially for *ft*. Seeds from two or more plants that correspond to maximum leaf count (low leaf number for *tsf*) from the graph (Fig. 2.3) were grown under SD and LD condition for comparison.

After third generation, the two *soc1* and *co* mutants showed uniform phenotypes which are late flowering under LD. Interestingly, *tsf* was found to be an early-flowering mutant. The findings were considered to be homozygous populations. Of the three late-flowering mutants, *co* remains the best choice based on flowering time as it produced the most leaves and flowered generally at 7 to 9 weeks in LD. As for *tsf*, it still flowered earlier than wild type (WT) both in LD and SD (Figure 2.4). From two-way ANOVA analysis, LD has a highly significant effect on flowering as compared to SD, $F(4,151) = 2996$, $p<0.001$. Bonferroni post-tests showed that under both LD and SD, three of the mutants; *tsf* (P), *soc1* (SALK) and *co* are significantly different from WT while *soc1* (SAIL) was found to have no significant difference from WT (Table 2.3). Flowers were produced as early as 3 weeks in LD and 4 weeks in SD.

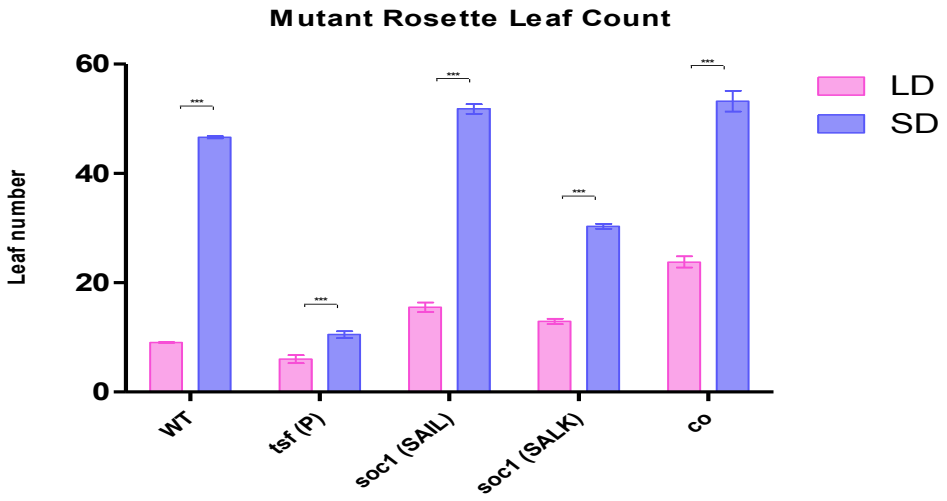


Figure 2.4: Flowering time of mutants based on total leaves formed by the primary meristem of plants grown under LD and SD. Each value represent means ± SEM (N=22). Two way ANOVA with non repeated measurement was performed. *$p<0.01$, **$p<0.05$, ***$p<0.001$. ANOVA=analysis of variance, LD-long day, SD-short day, SEM=standard error median.

Table 2.3: Bonferroni's post-tests performed on the flowering time by rosette leaf count of four mutants and WT following two way ANOVA analysis. *=significant with varying probability (*p*) (*$p<0.01$, **$p<0.05$, ***$p<0.001$), ANOVA=Analysis of variance, LD=long day, ns=not-significant, P=promoter, SD=short day, WT=wild type.

	LD	SD	Mean Difference	p value	Summary
WT	9.050	46.63	37.58	p<0.001	***
tsf (P)	6.000	10.50	4.500	p<0.001	***
soc1 (SAIL)	15.50	51.82	36.32	p<0.001	***
soc1 (SALK)	12.91	30.32	17.41	p<0.001	***
co	23.80	53.22	29.42	p<0.001	***
			Mean Difference	Significant?	Summary
WT vs. tsf (P)			27.08	Yes	***
WT vs. soc1 (SAIL)			1.26	No	ns
WT vs. soc1 (SALK)			20.17	Yes	***
WT vs. co			8.16	Yes	***

2.3.2 Mutant Screening - Genotyping

Genomic DNA of each mutant was subjected to PCR to determine its genotype. A combination of left primer (LP) and right primer (RP) in the PCR reaction is set up for WT and a combination of RP with left border (LB) would determine homozygous genotype (www.signal.salk.edu/tdnaprimers.html). This is because the RP is always on the 3' end of the insertion hence RP is always on the flanking sequence. The LP+RP reaction for WT would give a product about 900-1100 bp and an extra band of about 410+N bp (N is about 300 bp) for heterozygous line. As for the LB+RP reaction, heterozygous line would give the two bands as in the LP+RP reaction and only one band for homozygous (410+N bp).

Results for genomic analysis of all four mutants are shown in Figure 2.5. The expected size of the T-DNA insertion for *co* is estimated between 507 to 807 bp and a band within the range was observed as shown by the red circle. The rest of the mutants also produced a band within the estimated range with the exception of *soc1* from the SALK line which produced a band slightly higher than the expected range (446 to 746 bp). PCR confirmed that all four mutants are homozygous based on a single band detection on WT tested with LP and RP primers and mutants tested with LB and RP primers (Fig. 2.5,

samples designated with WT# and (mutant names)* respectively). In addition, sequencing carried out on all mutants also confirmed each mutant origin (data not shown).

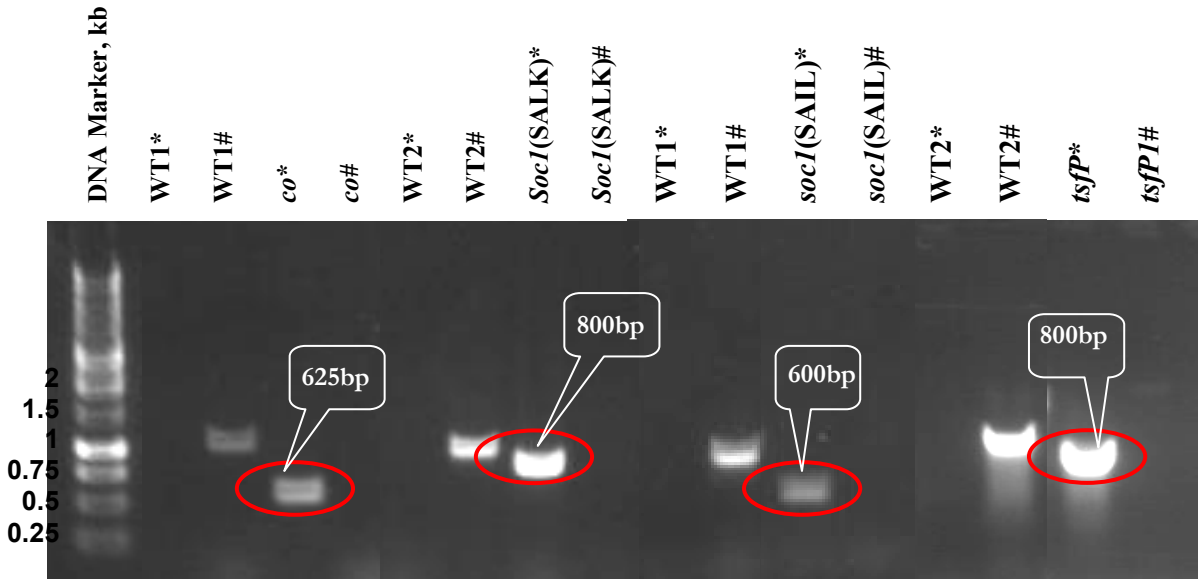


Figure 2.5: PCR gel to detect T-DNA insertion in *Arabidopsis* mutant for homozygous lines. Fragments were amplified from two T_3 generation plants tested with two primer pairs. The designation of * indicates PCR using a right primer specific to each mutant and a T-DNA left border primer while # indicates PCR using left and right primers of each gene. A T-DNA left border primer specific for the SAIL line was used for *co* and soc1 (SAIL) and a T-DNA left border primer specific for the SALK line was used for *soc1* (SALK) and tsf. WT1* and WT1# (wild type 1) had the same mixture as *co** and *co#* respectively while WT2* and WT2# had the same mixture as *soc1** and *soc1#* respectively. The same PCR condition was applied throughout all samples and the PCR products were loaded onto 1.2% agarose gel. Mutants showing homozygous genotype were represented by a single band from the amplification (as shown in red circles) and absence of WT gene product indicates gene disruption. PCR-polymerase chain reaction, T-DNA-transfer-deoxyribonucleic acid, SAIL (Syngenta *Arabidopsis* Insertion Library) insertion lines, WT-wild type, *co-constans, soc1-suppressor of constans 1*, kb-kilo base pair.

Further analysis was carried out by checking the transcript level of each mutant to ascertain its knock-out status. PCR of cDNA synthesized from RNA of each mutant using Actin primer as positive control produced the expected band at 400 bp (Fig. 2.6; yellow rectangles and arrows). Control wild type cDNA was also amplified using each mutant

specific transcript primer and produced expected band size corresponding to specific mutant (Fig. 2.6; blue rectangles and arrows). No band was detected in three mutants; *co* and both *soc1* lines confirming absence of transcript (Fig. 2.6; green arrows). However, three bands were detected when the cDNA of *tsf* was amplified using the mutant's specific primer (Fig. 2.6; green rectangle)

Based on these results, co mutant plant was chosen for further proteomics work as CO is more directly involved in FT regulation and the mutant late flowering characteristic is a good comparison as well as control for further experiments. In addition, as in is already in Columbia background, it would compliment proteomics experiment where the other 3 samples and WT are of the same ecotype.

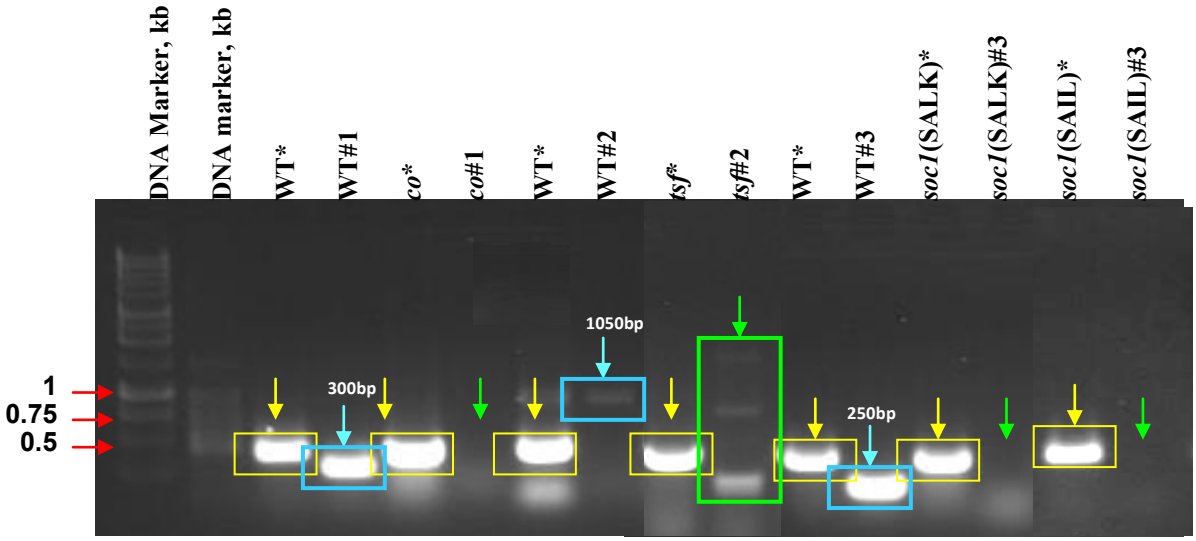


Figure 2.6: RT-PCR products on 1.2% agarose gel. Fragments were amplified from cDNA of mutants using Actin as control (*). The designation of #(1,2,3,4) indicates PCR using specific primer pairs for each mutant. A 400 bp positive control band can be seen on all mutants and WT amplified with Actin (yellow arrows) while WT amplified with specific primer pairs gave the expected band size of each mutant (blue arrows). Lanes without any band (green arrows) indicate null transcript expression which is a characteristic of a mutant.

2.3.3 Optimisation and Establishing Exudation Technique

Parameter optimization of phloem sap protein exudation was carried out on growth conditions, chelating agent concentration (10 and 20 mM), exposure to LD prior to

exudation (2 and 3 days), incubation period during exudation (3 hours and overnight), incubation condition (in SD and LD chamber) and plants age (6 and 7 weeks). Two growth conditions were used; 1) plants were either grown solely in SD until exudation or 2) grown in SD and prior to exudation were exposed to LD for a period of up to 3 days. From the initial experiments, plants exposed to LD up to three days prior to exudation were found to give better quantity of protein compared to plants grown solely in SD condition (Fig. 2.7 (a)). Hence, for subsequent work, plants were exposed to LD two to three days prior to exudation.

No major difference was observed with differing EDTA concentration and incubation period but plants soaked in 20mM EDTA and those incubated overnight produced larger quantity of protein based on Bradford assay (data not shown). As for the duration of exposure to LD condition, plants exposed to LD condition for 3 days generally produced better quantity of protein (Fig. 2.7 (b)). For phloem exudation, the plants incubated for 3 hours or overnight as well as those incubated overnight in SD and LD condition did not differ much in concentration based on Bradford assay done (data not shown) and quality (based on the presence of equal bands and pattern) (Fig. 2.7 (c) and (d) respectively). Finally, plants' age was also studied where 7 weeks-old plants were found to produce a low amount of protein (Fig. 2.7 (e)). In addition, proteins purified by acetone precipitation gave lower yield and poorer quality compared to those purified and concentrated with Amicon filters (Fig. 2.8).

Using the optimized exudation technique, protein were collected in 20mM EDTA for overnight in LD condition from plants grown for 6 weeks in SD and exposed to LD for 3 days prior to exudation. Two extra steps were added before purification using Amicon filters. The first exudates were collected for a period of 30 min and the leaves were dipped in a second EDTA solution of the same concentration. Second exudation was done overnight. Proteins from both first and second exudations were then centrifuged separately at 4472 *x g* (100,000 rpm) for 1 h in an ultracentrifuge producing a pellet containing contaminants such as membranous protein that might leak from the sieve element (SE) and separating it from phloem sap protein in soluble form. The aqueous phase was concentrated with Amicon filter and the pellet was resuspended in 20mM EDTA. In terms of quantity or quality there was not much difference between the first and second exudates

(Fig. 2.9). Meanwhile the pellet was impossible to quantify as it was wax-like after ultracentrifugation and attempts to conduct Bradford assay was not successful. When run on SDS-PAGE all protein bands from other samples apart from the pellet was observed (Fig 2.9).

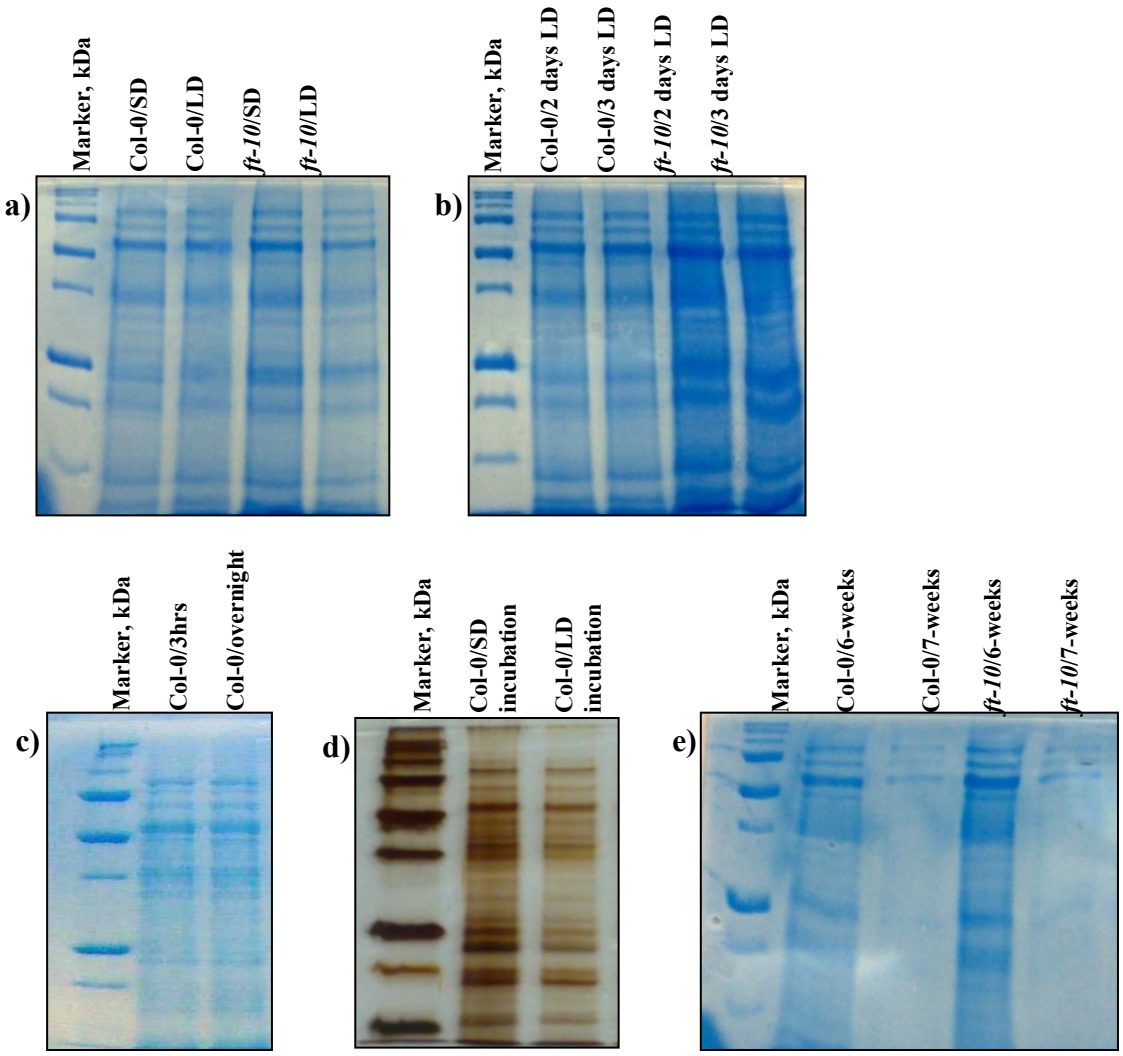


Figure 2.7: Exudation technique of *Arabidopsis* phloem sap protein was optimised under different parameters. (a) Protein from plants grown under full SD and plants exposed to LD 3 days prior to exudation, (b) protein from plants exposed to 2 days and 3 days LD prior to exudation, (c) protein from 3 hours and overnight incubation, (d) protein incubated in SD and LD condition and (e) protein from 6-weeks and 7-weeks old plants. 5 µg of protein from each sample was loaded onto 17% SDS-PAGE and stained with coomassie blue or silver for (d).

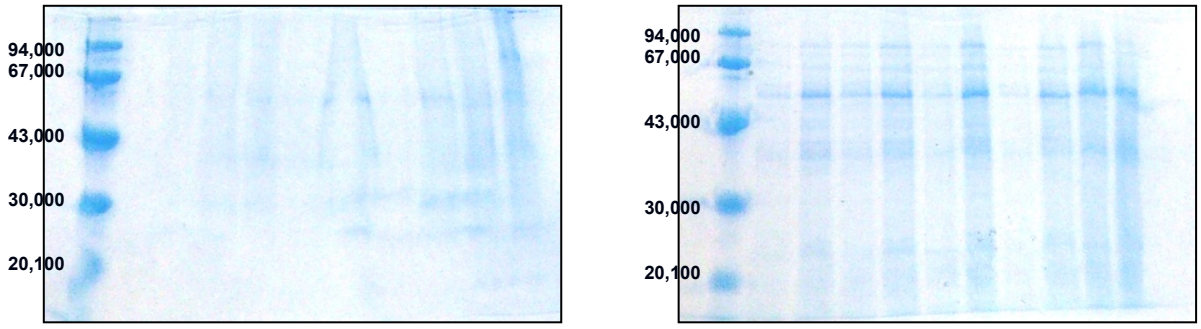


Figure 2.8: Purification of protein was carried out using acetone precipitation (left) and Amicon filter (right) where it can be seen from the SDS-PAGE that much of the samples were lost in the former method. Equal amount of proteins (5 µg) were loaded onto each well.

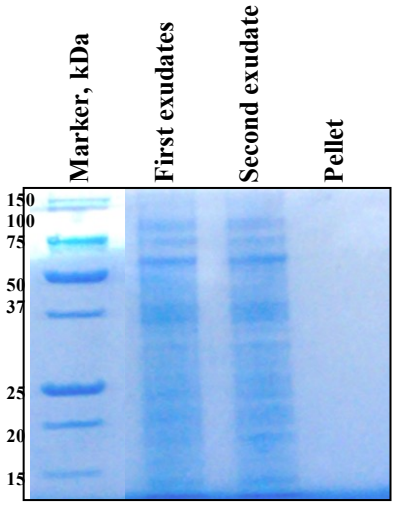


Figure 2.9: Phloem sap protein comparison from the first exudates collected right after wounding, the second exudates collected 30 minutes after wounding and overnight respectively. No protein was observed for the pellet spun down from both exudates.

2.3.4 Optimisation of 2D Gel Electrophoresis for *Arabidopsis* Phloem Sap Protein

Initial optimization for isoelectric focusing (IEF) was done using total leaf protein from wild type plants. Various IEF methods were applied and tested ranging from 100V to 8000V and from as short as 6 hours up to 26 hours duration. A method used in *Brassica napus* proteome study was found to be suitable for *Arabidopsis*. However, reproducibility of 2D spots using this method is not guaranteed as varying results had been obtained. In general, cleaned up protein when exposed to 100V of initial current and gradually increased to 2000V over a total period of approximately 26 hours produced the best spots on 2D gradient gel. A large amount of protein up to 1mg is required from phloem sap protein as compared to total leaf protein (Fig. 2.10).

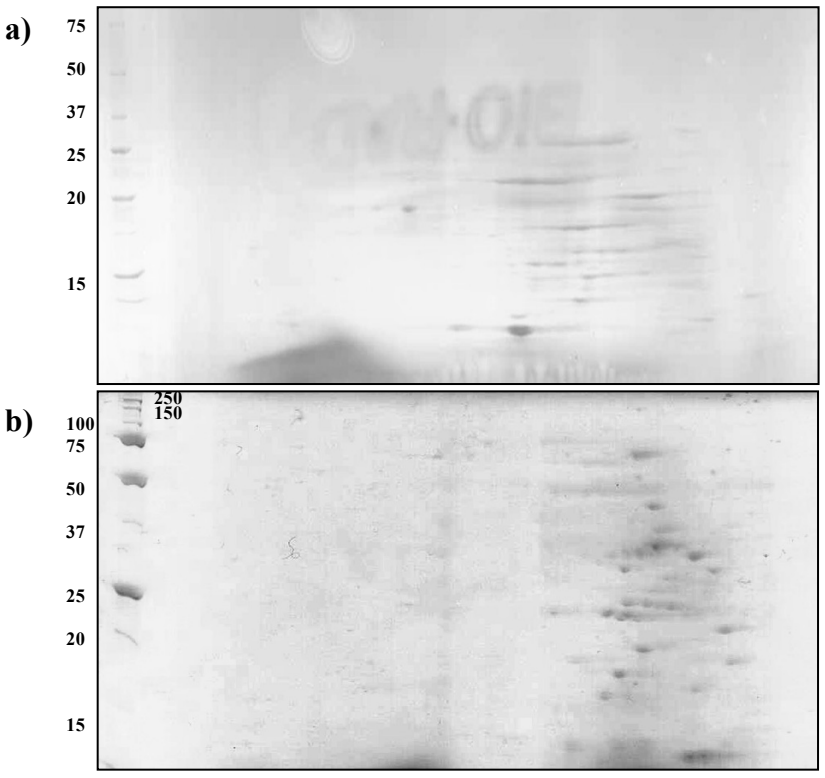


Figure 2.10: 2D gel electrophoresis of *Arabidopsis* protein. (a) Total leaf protein and (b) phloem sap protein. 200µg total leaf protein and 1mg phloem sap protein was used for IEF and IPG strip (pH3-10) was run on 8-16% SDS-PAGE gel and stained using coomassie blue. Huge difference in concentration of proteins used was based on optimisation of each type of proteins for IEF.

In certain gel especially earlier ones, horizontal streaking was sometimes observed (Fig 2.10 (a)). And, spots were mostly observed towards the basic range of the IPG strip which is obvious for protein from phloem sap (Fig 2.10(b)).

2.4 Discussion

2.4.1 Mutant Screening

Four flowering time mutants were confirmed. *co* produced late flowering plants with the highest number of rosette leaves followed by *soc1* from SAIL lines under LD. Meanwhile, *tsf* (promoter) remains as an early flowering mutant in both SD and LD conditions. A slight variation on the leaf count for the subsequent batch of the mutants as well as the wild type that saw a small increment in leaf number (Fig. 2.4) might be contributed by a larger seed number used than the initial screen and some fluctuation to the temperature and humidity in growth room settings.

Earlier in the screening process, there were two *tsf* mutants in which one contains a T-DNA insertion in the promoter while the other in the intron. From observation, both mutants flowered earlier than wild type. The difference between both mutants is that the former flowered earlier than the latter and the number of rosette leaf count produced (Fig. 2.3).

The same *tsf* (intron) has been used in a study by Michaels et al (2005) to induce early flowering lines by using activation-tagging vector carrying four copies of 35S cauliflower mosaic virus enhancer element. In an earlier study, overexpression of *TSF* via constitutive 35S promoter has been shown to lead to strong early-flowering phenotype (Kobayashi et al., 1999). Observation of the *tsf* (intron) in this study showed it flowered earlier than wild type though the number of rosette leaves produced by both plants was similar. As rosette leaf count has been in practice to determine flowering time, the *tsf* (intron) was concluded to have the same phenotypic characteristic as the wild type and was not chosen for proteomics work.

Meanwhile, the *tsf* (promoter) used in this study is not overexpressed with the 35S promoter but instead its promoter was disrupted with insertion of T-DNA. It is of great interest that the plant should flower early and may represent a new type of mutant. As a homolog to FT, TSF and its mRNA shared similar pattern and photoperiod response to FT (Yamaguchi et al., 2005). In the phloem, CO overexpression causes *FT* and *TSF* to induce early flowering by increasing their transcripts levels (Jang et al., 2009). In this study, the early flowering phenotype shown by *tsf* (promoter) may be due to T-DNA insertion at the promoter site. T-DNA insertion into promoter region can lead to 1) complete inactivation of the gene, 2) reduced expression of the gene and 3) increased expression of the gene (Cullis, 2004). Three bands were observed in the transcript of the mutant which may be due to increased expression or incorrect regulation of the gene. Studies done by Pouteau et al (2008) on early flowering mutants found that there is a correlation between early-flowering phenotypes and diversification in photoperiodic responses. Although it would be of interest to study the *tsf* (promoter) mutant in details, it may not be a good candidate for proteomics work. Early flowering mutants tend to have less and smaller leaves to normal plants hence would pose insufficiency of protein sample for proteomics work that depended heavily on leaf production. The main priority for conducting mutant screening experiment was to search for a better if not the best mutant for use in proteomics work; in this case a late-flowering homozygous mutant is desirable. Therefore the existing *tsf-1* mutant which is not early-flowering was chosen instead.

Both *soc1* mutants used in the study flowered late in SD. Under LD, *soc1* (SALK) flowered earlier than WT but *soc1* (SAIL) remained as late flowering (Fig. 2.4). Numerous studies using various *soc1* mutants documented late-flowering characteristics (Borner et al., 2000; Samach et al., 2000; Lee et al., 2000; Ferrandiz et al., 2000; Melzer et al., 2008). However, a mutation of *soc1* in the LD pathway resulted in delayed flowering time (Moon et al., 2005). This could be seen in the *soc1* (SAIL) plants. The *soc1* (SALK) that flowered earlier as in the wild type could be due to variation of flowering across plants. In 35S:FT plants, *soc1* mutant was found to slightly delay the strong early flowering characteristic (Moon et al., 2005). A marked increase in longevity and productivity was observed from the growth of many short determinate co-inflorescences which showed that loss of SOC1 activity is adequate to increase lifespan of plants independent of flowering time (Tepfer-Bamnolker and Samach, 2005; Yoo et al., 2005). In *Arabidopsis*, SOC1 functions to

regulate flowering time, floral patterning and floral meristem determinacy (Liu et al., 2007; Melzer et al., 2008). *SOC1* gene performed its role in the apical meristems by acting downstream of FT. However, the gene is also expressed in leaves during vegetative phase independently of FT. Coming from a MADS-box family encoding a conserved MADS box protein among angiosperms (Lee et al., 2000; Lee at al., 2008; Cseke et al., 2003; Ferrario et al., 2004; Nakamura et al., 2005) similar SOC1 characteristics are also reported in other plant species thus in plant development SOC1 is likely to play a role as a general regulator in organogenesis (Lee and Lee, 2010).

Flowering time of *co-1* under LD was reported with less than 15 leaves and wild type at approximately 10 leaves (Moon et al., 2005) which are more or less the same as found with the average leaf count of the *co* mutant from the present study (Fig. 2.3). However, as the growth conditions for both studies were not identical, further and complete comparison could not be done. While it has no effect on flowering in SD in this study, late flowering of *co* mutants has previously been documented under SD by Koornneef et al. (1998). The same team reported that under LD, native *CO* acts as the main regulator that elevates rate of flowering for plants. The *co* mutant in the present study showed late flowering characteristic which could be the result of the absence of *CO*. In addition, all the mutants used for screening experiment is in the Columbia ecotype. *co* in Columbia background is beneficial for proteomics work to ensure comparison against other mutant lines could be made without any bias. Further proof was obtained from statistical analysis via two-way ANOVA showing that when *co* was compared against WT there was a significant difference between photoperiod and the lines, $F(1, 56) = 26.35$, $p<0.001$.

2.4.2 Effect of EDTA on Exudation

The optimized conditions for phloem protein more or less correspond with previous methods applied in sunflower, mung bean, grape vine, orange, soybean, barley and oat (King and Zeevart, 1974; Friedman et al., 1986; Touraine and Astruc, 1990; Weibull et al., 1990) where phloem fluid was obtained either from detached leaves (Urquhart and Joy, 1981) or cut fruits (Fellows et al., 1978; Pate et al., 1974). EDTA is the main component for exudation in each plant. Apart from the use of Na_2-EDTA, K_2EDTA has also been used

to obtain protein from phloem exudate (Hoffmann-Benning et al., 2002). EGTA and citric acid had also been found to assist exudation process (King and Zeevart, 1974). However, in bean at least, EDTA was reported as better than citrate in improving exudation from detached leaves (Fellows and Zeevart, 1983).

EDTA is a chelating agent that had been found able to form stable complexes with divalent cations. Chabarek and Martell (1959) reported the efficiency of EDTA as a chelating agent decreased with decreasing pH when the pH was decreased from 7.0 to 4.4. EDTA binds divalent cations on an equimolar basis, resulting in freely available ions in the presence of a salt with greater concentration than EDTA. An excess of cation (Ca^{2+}) will lead to abolition (Chabarek and Martell, 1959). Thus, EDTA was thought to permit continuous exudation by maintaining the phloem in an unblocked state in the tissue near the base of the cut petiole (King and Zeevaart, 1974).

In this study various concentrations of EDTA ranging from 5 mM to 20mM have been used in exudation of phloem sap. Comparison was also made between two different concentration of EDTA; 10 and 20 mM. In terms of quantity, both concentrations produced similar amount of proteins from exudation based on quantification done using Bradford Assay (data not shown). Furthermore, the pattern of protein bands obtained via SDS-PAGE (data not shown) was identical indicating the quality was not affected by different concentrations of EDTA. In addition to EDTA concentration, exudation process itself is influenced by temperature, humidity and light intensity. Tully and Hanson (1979) reported that leaf exudates must be mainly of phloem origin since they found evidence that rate of exudation is inhibited by cyanide and Ca^{2+}.

Generally, exudation technique is a good method of obtaining phloem sap from *Arabidopsis*. It provides an alternative to stylet-cutting techniques (Yoshida et al., 1996) which may pose problems in terms of finding suitable insects, collecting sufficient amount of sample for analysis and especially during cutting of stylet. Furthermore, samples from stylet-cutting technique are only obtained from sieve elements accepted and fed by the aphids (Ponder et al., 2000; Wilkinson and Douglas, 2003). As to whether the phloem sap obtained and used for proteomics work in this study is pure or not could not be determined. To date there is yet a method to prove the purity of phloem sap or the degree of

contamination from surrounding cells during wounding and exudation. Zhang et al (2012) noted that the presence of hexoses in pumpking is severe as compared to small amount in cucumber. A non-mobile material, hexoses is being used by the team as an indicator for assessing contamination which may be released along with the exudates from the high pressure during wounding of the phloem. Another compound, Rubisco (Ribulose-1,5-bisphosphate carboxylase oxygenase) can also be used to measure the purity of phloem sap as it originates in the chloroplast (Zhang et al., 2012). Rubisco was among the proteins identified (Chapter 3, Table 3.2) hence the phloem sap protein obtained from the study might have somehow been compromised with contamination from ruptured chloroplast. Phloem is comprised of complex compartments of surrounding cells hence the degree of purity of phloem sap protein obtained could not be determined.

2.4.3 Correlation between Photoperiod and Leaf Morphology and Its Effect on Yield of Exudates

Photoperiod takes into account daily length of light and dark periods. It has been used to deliberately control flowering in many species and plays a major role in seasonal transition to flowering (Koornneef et al., 1998; Levy and Dean, 1998; Pineiro and Coupland, 1998). Being a species that responds to photoperiod, *Arabidopsis* is a facultative LD plant that promotes flowering under LD and delayed flowering under SD.

In many plant species photoperiod affects some of later stages of flower development (Thomas and Vince-Prue, 1997). Normally once floral primordia are induced, organization of flower meristem starts and floral organs are produced. Subsequent changes in photoperiod such as increment or decrement of light and dark periods can cause delayed floral primordia, floral reversion, male sterility and changes in sex expression (Thomas and Vince-Prue, 1997; Jeong and Clark, 2005).

However, in the present study plants for exudation should ideally produce more leaves rather than rapidly progress to flowering stage. The reason is because phloem sap is obtained from the cut petiole of mature leaves and the more leaves a plant produce the better is the chance of obtaining more sample. Based on observation, it was not a surprise

that *Arabidopsis* produced more leaves in SD as compared to LD. Under SD, leaves can develop further before moving on to flowering stage.

Even though leaves are produced in larger quantity in SD, length of light still play an important role. Plants grown in SD with 10 hours light for 6 weeks produced lower numbers of leaves as compared to those grown in 8 hours light (Corbesier et al., 1996). Hay and Delecolle (1989) suggested that response to daylength has an influence on final leaf number. However, plants grown in 8 hours light need at least 2 more weeks for leaves to be in a suitable condition for exudation. When allowed to grow further until flowering stage is reached, plants in SD with 10 hours light would flower within 8 weeks while those in 8 hours light took approximately 11 weeks. This confirms that SD did indeed delayed flowering in *Arabidopsis*. Adams et al (2008) reported short and compact plants were observed in SD condition while plants grown in LD (≥12 h light) tend to be tall where a more or less similar characteristics were observed in current study. However age also played a role in the phenotypes. When plants in SD finally flowered they have an advantage in height due their age and longer duration to light throughout its development. Yield of phloem sap protein obtained from either LD or SD did not differ much in this study. As equal number of leaves was used each time from both photoperiods, the amount of exudates and final protein concentration may not change much.

Reduced number of leaves is often linked to altered circadian rhythms. This was observed for plants grown in LD. Photoperiod has been known to have an effect on the final leaf number on the main stem of wheat and barley as it influence the time of flowering (Miglietta, 1989; Kirby, 1990; Hay and Kirby, 1991). In response to daylength, grass cultivars were reported to have different timing of response of final leaf number (Brooking et al., 1995). Though the number of leaves was reduced in LD, the leaves were generally bigger, longer and greener in appearance. Similar or longer leaves under LD have been implicated to be a result of morphogenetic effect of LD causing increment of leaf area (Adams et al., 2008). Increase in plant leaf area under continuous light was the result of increasing leaf blade dimension (length-wise) but did not increase leaf number or tiller (Heide et al., 1985). Increase daylength causing an increase in specific leaf area in a number of species has been reported. For example, under LD, petunia grew greener larger and thinner leaves while increased leaf growth and leaf greenness was reported in

impatiens (Adams et al., 2008). Understanding photoperiodic effect on leaf and its development may assist in future work to generate plants with increased number of leaves in response to LD photoperiod.

2.4.4 Plant Condition and Developmental Stage May Affect Protein Content of Phloem Sap

Changes in phloem characteristics may be a result of cessation of growth and maturity of leaf tissue. It had been found to restrict aphid performance on grasses (Caillaud and Niemeyer, 1996). This was thought to be the cause of longer phloem sap flow on stylets cut from young *Prunus padus* and *Hordeum vulgare* leaves (Sandstrom, 2000).

Younger and growing leaves contain higher amino acid variety and quantity than mature leaves. This preference was seen on willow aphids and many aphid species favour immature leaves over mature leaves as the former has high rates of nitrogen transport (Mittler, 1953). Although exudates are a representation of amino acid composition in phloem, it is not a useful quantification of concentration (Girousse et al., 1991; Sandstrom, 2000). In the present study no obvious difference in terms of yield obtained between the plants exposed to two different light treatments prior to exudation. However, phloem protein concentration was found to decrease with plant age (Fig 2.7 (e)) hence reducing protein detection in MS/MS spectra. Since amino acid content was not studied, the results obtained were inconclusive. In the meantime, various studies reported variation in free amino acids in phloem (Atkins et al., 1983; Chino et al., 1987; Douglas, 1993) according to leaf developmental stage.

Weibull et al (1990) concluded that free amino acids in EDTA method exudates come from phloem sap based on composition that is similar to those of phloem sap profile using aphid-stylet technique. While phloem sap from aphid stylet originates from single phloem vessels, phloem sap from exudation technique came from all phloem sap vessels in the cut-off leaf. The probability of getting proteins from other cells apart from SE could not be ruled out as the wound made would also injure neighbouring nucleate cells. To overcome this problem, an extra step was employed where collected exudates were ultracentrifuged to pellet down nuclear neighbour cell contents such as those from

membrane and apoplast. As the pellet from ultracentrifugation was waxy-like and difficult to dissolve, it may be possible that lipid or other protein mixture was present in the phloem sap. No band was detected on SDS-PAGE for the pellet after resuspension though the buffer used contained SDS for solubilisation therefore ultracentrifugation may not necessarily remove all contaminants.

2.4.5 Factors Affecting the Reproducibility of 2D Gel Electrophoresis of *Arabidopsis* Phloem Proteome

Attempts were made to optimize 2D gel electrophoresis procedures for *Arabidopsis* phloem proteome. In addition to the phloem sap protein, protein from leaf was also used as sample. However, using total leaf as starting material makes the protein susceptible to contamination from non-proteinaceous compounds such as nucleic acids, lipids, salts, phenolics and chlorophyll. These compounds can create artefacts during IEF.

Phenolic compounds have been known to be in abundant in certain plant and their presence in protein sample can interfere with protein separation during IEF (Rabilloud et al., 2007). On the other hand, nucleic acids at a high concentration have been found to blur the 2D electrophoresis pattern (Heizmann et al., 1980). There was not much of a problem with phloem sap protein obtained from using exudation technique and to test for contamination with either DNA or RNA the absorbance of the samples were measured with UV spectrophotometer. Absorbance at 230-, 260- and 280 nm that are specifics to DNA and RNA (260- and 280 nm for DNA and all three for RNA) gave no reading indicating the absence of those compounds in the protein sample. However, as Na_2-EDTA used for collection of exudation is a salt, it has to be purified further to desalt the protein.

Precipitation of protein at their pI sometimes occurs and could hamper IEF. Solubility of certain proteins may depend on salt requirement and though not preferable, a tolerable amount of salt is accepted at a total concentration of less than 40 mM. However, during initial stage of focusing, the salt will most definitely be lost due to high current employed. On the other hand, if salt concentration is too high it will interfere with focusing as more time is needed as a result of reduced current and making it hard to achieve a particular voltage. This can be seen on samples run using phloem sap protein where

voltage higher than 5000V is almost impossible to be reached which could probably be attributed to salt from EDTA and carrier ampholytes. Carrier ampholytes can be used as an alternative to overcome salt insufficiency to solubilize protein. But just like salt, high concentration of carrier ampholytes can slow down IEF by limiting the current (Garfin and Heerdt, 2004). All possible steps were carried out and extra care was applied during handling and sample preparation but results from 2D experiments were still not reproducible.

IEF is based on differences in the protein pI hence it is important for the protein to be soluble throughout this step as there is a high tendency for protein aggregation and precipitation. The solubilising agents used are similar to those in sample preparation but the choice is much more limited as compatibility of solutions with IEF must be taken into account. Weakly charged or uncharged compounds are preferred while the concentration of any charged compounds during sample preparation should be reduced. Concentration reduction of such compounds can be done either by substantial dilution, dialysis or using a commercial clean up kit (Berkelman, 2004). For this study, both total leaf and phloem sap proteins were subjected to clean up using commercial kits prior to IEF to minimize the concentration of any compound that may interfere with IEF.

For IEF, IPG strips pH 3 to 10 were used based on literature of previous study carried out in brassica and cucurbit. The chosen pH range was considered broad range and allows the display of most proteins in a single gel. Over extended focusing time, these strips provide gradient stability (Bjellqvist et al., 1982). In addition, the commercialized IPG strips facilitate IEF process with a choice of various pH ranges to suit many applications. The choice to use linear or nonlinear strip depends on research carried out. Linear strips work better with narrow pH range than broad range by displaying more spots per sample. However, nonlinear strip performed much better with broad range pH where protein can be better resolved in the middle of the pH range where it commonly tends to be focused.

The spots however, were not consistent or reproducible using both types of protein sample. Among the reasons thought to be the cause are amount of protein loaded and protein loading or sample application. Various amount of protein have been recommended

ranging from 50 µg to 750 µg from various plant species including cucurbit, cucumber, melon and pumpkin (Zhang et al., 2010; Di Carli et al., 2010; Malter and Wolf, 2011; Walz et al., 2004). Using total leaf protein, 200 up to 400 µg of protein was used with contradicting results as protein spots obtained were not consistent in terms of intensity and pattern. For phloem sap protein, 200 µg up to 1 mg of protein was used. There are factors associated with the decision of how much protein to be loaded such as purpose of the gel, abundance of proteins of interest, complexity of the mixture, pH range of IPG strip and subsequent analysis. A compromise load is needed for complex mixtures to avoid overshadowing of high abundance to low abundance proteins. If the purpose of the gel is to just get a good picture image, minimum amount of protein is needed just as in the case of broad range IPG strip. Meanwhile, if the purpose is to study low abundance proteins, larger protein load would be beneficial. Nevertheless, higher concentration or amount of protein did not necessarily produce better protein spots as observed in the study. Sometimes precipitation can occur during IEF. This comes back to sample application prior to IEF where the method used in the study is by cup loading. Rabilloud (1999) reported sample precipitation due to cup loading method could be avoided by applying the sample during rehydration instead. In addition, sample application during rehydration is advantageous as sample application is simple, provides shorter focusing time due to the availability of sample in the IPG strip and very large amount of protein is permitted. High amounts of total leaf protein (up to 400 µg) and phloem sap (up to 1 mg) were used in the current study. Reproducibility was still a problem and this indicates that higher amount of protein does not necessarily ensure the success of experiment. Since solubility was also thought to be a contributing factor to obtaining reproducible products, large amount of protein used may have precipitated during IEF as cup loading was the method used to load sample. In addition, sample was not applied to IPG strip during rehydration which was the suggested to avoid sample precipitation by cup loading.

A reproducible spot pattern for 2D gel electrophoresis is highly dependent on IEF. Consistent and reproducible focusing is required with appropriate time integral of voltage. Time needed for focusing depends on pH range of IPG strip, sample and buffer characteristics. Focusing voltages are usually programmed in stages to minimize ionic constituents in the sample and at the same time limiting the strips to electrical heat. A total of six IEF methods were applied during the study ranging from as little as 25,000 volt-

hours (V-hr) to 99,500 V-hr. The optimized focusing time was achieved using 25,000 V-hr for 25 hr and 40 min which was also used for phloem proteome study in brassica (Giavalisco et al., 2006). Reproducibility was still a problem with inconsistent result even when more than one IPG strip was run alongside using the same sample. It has been noted that individual strips will experience different conditions even when run at the same time. During focusing, total current limit is averaged over all strips in the tray although some strips may be exposed to more current than expected.

Once the first dimension (the IEF) is completed, the next step is the second dimension by running the product of IEF using 2D SDS-PAGE. Gel composition will determine resolution of sample protein, and gradient gel is preferable over single-percentage gel as was in the study. The decision lies with the fact that small proteins remain much longer in gradient gels giving the chance of both large and small molecules to be resolved in the same gel. Gradient gel is also a good choice for initial estimation experiment with complex mixtures where a single-percentage gel can later be used once interesting regions in the 2D array have been identified. In the study, 8-16% gradient gel was used that gave a good coverage of high and low molecular weight proteins.

Based on the principle of pI and IEF along with equilibration prior to second dimension separation, various protein spots can be seen across IPG strip pH 3 to pH 10 for proteins from both total leaf and phloem sap. The difference between spot patterns of the two is that more spots were obtained on the basic side of the pH gradient for phloem sap protein as compared to total leaf. The result may have a correlation with phloem content being basic in nature hence more spots towards alkaline pH. Protein spots from as large as 75 kDa and as small as below 15 kDa were observed especially from the phloem sap. The gels were stained using Coomassie blue for ease of handling during LC-MS/MS and it was possible that more protein spots can be detected using silver staining as it is well known for its sensitivity in detection. From the gels, selections of protein spots in varying molecular weight were processed to be identified. Only the brightest spots were chosen as they were more likely to contain protein in abundance. This would leave out many insoluble and low-copy proteins as they most probably could not be detected on the gel. Although 2D proteomics is a good tool for identifying protein it still has limitations in detecting and resolving some proteins which are large (>100 kDa), small (<10 kDa), basic

(pI > 10) and hydrophobic (Shaw and Riederer, 2003; Gorg et al., 2004; Rampitsch and Srinivasan, 2006).

In summary, optimisation carried out to establish EDTA exudation technique produced quantifiable protein samples for further proteomics work. Large amount of protein can be obtained by increasing the number of plants grown to increase leaf number and hence phloem sap protein. Similar pattern of bands were also reproduced on 1D SDS-PAGE which can be used for tryptic digest and protein identification via LC-MS/MS. However, optimisation method for 2D proteomics was unsuccessful and results were not reproducible. Therefore proteomics work to establish phloem proteome database from *Arabidopsis* was principally based on 1D proteomics utilising total phloem protein and excised gel bands.

CHAPTER 3
Arabidopsis PHLOEM EXUDATE PROTEOME

3.1 Introduction

Interest in the area of proteomics arose more than a decade ago. The invaluable information produced and wide application provides hope for better understanding of protein function. Several comprehensive proteome studies have been reported on varieties of plant parts or organs. The focus of this study is the phloem which is a major plant system for transport of important nutrients and signalling molecules for its development. Phloem proteomics has been studied in several species where plants such as *Brassica napus* (Giavalisco et al., 2006), pumpkin (*Cucurbita maxima*) (Walz et al., 2004; Lin et al., 2009), cucumber (*Cucumis sativus*) (Walz et al., 2004), melon (*Cucumis melo*) (Malter and Wolf, 2011), rice (*Oryza sativa*) (Aki et al., 2008), poplar (*Populus alba*) (Dafoe et al, 2009), and lupin (*Lupinus albus*) (Rodrigues-Medina et al., 2011) have provided platforms for phloem proteome databases to be set up. A wide range of proteomics studies in the model plant *Arabidopsis thaliana* has been reported for example those that derived from plasma membrane (Santoni et al., 1998; Kawamura and Eumura , 2003; Marmagne et al., 2004), mitochondria ((Kruft et al., 2001; Miller et al., 2001), cell wall (Chivasa et al., 2002; Bayer et al., 2006; Irshad et al., 2008), nucleus (Bae et al., 2003), chloroplast (Friso et al., 2004), vacuoles (Carter et al., 2004), organelle (Dunkley et al., 2006), root (Roth et al., 2006; 2008; De-la-Pena et al., 2010) and flower (Feng et al., 2009). The initial aim of this study was to provide steps towards providing guidelines for generating and annotating a comprehensive protein database from *Arabidopsis* phloem. However, recently, *Arabidopsis* phloem exudate protein studies (Guelette et al., 2012; Batailler et al., 2012) were reported. Hence, this study aimed to discover additional new phloem exudate proteins and compare them with the two similar studies. The phloem exudate proteome of *Arabidopsis* was initially investigated using 1D LC-MS/MS (1 dimensional liquid chromatography-tandem mass spectrometry) to identify proteins contained within the phloem. Utilising the exudation technique, comparison was made between total phloem sap protein and excised gel bands from SDS-PAGE. To validate the identified proteins obtained via the EDTA exudation method, a similar approach was used to obtain phloem sap proteins from *Brassica napus*. In addition, attempts to identify phloem sap protein

through 2D LC-MS/MS were also carried out as a comparison to 1D LC-MS/MS. As the major protein of interest in this project, FT had not previously been identified in *Arabidopsis* phloem proteome, a targeted, high sensitivity MRM (multiple reaction monitoring) method was employed to search for FT.

3.2 Materials and Methods

3.2.1 Phloem Sap Protein Collection via Exudation Technique

Optimised conditions were used and plants grown for up to 6 weeks in SD (10 h light, 14 h dark) followed by exposure to LD (16 h light, 8 h dark) three days prior to exudation as in 2.2.2.1. Exudates were further purified and concentrated using centrifuge filtration as in chapter 2 section 2.2.2.2.2. Quantity and quality of proteins obtained were checked via Bradford Assay and SDS-PAGE respectively (chapter 2 sections 2.2.2.3 and 2.2.2.4 respectively).

3.2.2 Trypsin Digestion

3.2.2.1 In-Gel Digestion

3.2.2.1.1 In-Gel Digestion of Coomassie Stained Gel

All visible bands (Fig 3.7) were cut from the gel and each of them was diced into smaller pieces of about 1-2mm^2. Approximately 100µL of 25mM NH_4HCO_3 (ammonium hydrogen carbonate)/50% (v/v) ACN (acetonitrile) was added to cover the diced gels and vortexed for 10 min. The supernatant was discarded and addition of 25mM NH_4HCO_3/50% ACN followed by vortexing was repeated twice before the gel pieces were dried completely using a centrifugal vacuum concentrator (Jouan) for about 20 min. 10 mM DTT (in 25 mM NH_4HCO_3) was added to cover the gel, vortexed, spun briefly and incubated at 56^{o}C for 1 hr. Following incubation, 55 mM iodoacetamide was added to the gel pieces, vortexed, spun briefly and incubated in the dark at room temperature for 45 min. After that, the supernatant was removed and the gels washed with 25mM NH_4HCO_3, vortexed for 10 min and spun briefly before the supernatant was removed. Gels were dehydrated with 25mM NH_4HCO_3/50% ACN, vortexed for 5 min and spun. The dehydration step was repeated once prior to complete drying of the gel with a speed

vacuum. Next, trypsin solution (12.5ng/μL Sequencing Grade trypsin (Promega) in 25mM NH_4HCO_3) was added to cover the gel pieces followed by rehydration of the gels on ice or at 4°C for 10 min. The samples were spun briefly and 25 mM NH_4HCO_3 was added to completely cover the gel pieces. After a brief spin, the samples were incubated overnight at 37°C.

The digest solution was transferred into a clean 1.5 mL tube and approximately 30μL 50% ACN/50% (v/v) formic acid was added to the gel pieces, vortexed 20-30 min, spun and sonicated for 5 min. The solution was transferred into the tube containing the overnight digest solution and the solution was vortexed and spun before the volume was reduced to approximately 10 μL with the speed vacuum. The samples were then cleaned with C18 Zip Tip (Millipore) by aspirating the tips 5 times; each in ACN, 1% formic acid, protein sample then washing in 0.1% (v/v) formic acid followed by elution in 10 μL 0.1% formic acid/50% (v/v) MeOH, before finally the solution was vacuum dried for 15 min. Prior to LC-MS, 10 μL of 3% (v/v) CH_3CN/0.1% formic acid was added to resuspend the samples.

3.2.2.1.2 In-Gel Digestion of Silver Stained Gel

To the diced bands, 30 μL 30mM $FeK_3(CN)_6$ (iron potassium cyanide): 100 mM $Na_2S_2O_3$ (sodium thiosulphate) (1:1) was added and the gels soaked in the solution for 10 min with vortexing in between. The solution can be decanted and the step repeated until gel bands are clear. After removal of the destaining solution, the gel pieces were washed three times with 100 μL water followed by vortexing for 15 min each. Once the water was removed, the gel pieces were washed 3X with 100 μL 25 mM NH_4HCO_3/50% ACN and vortexed for 15 min each. Next, the solution was removed and the gel pieces vacuum dried for about 15 min. 10 mM DTT (in 25mM NH_4HCO_3) was added to the dried gel pieces, vortexed, spun briefly and incubated for 1 hr at 56°C. The supernatant was removed and 55 mM iodoacetamide was added, vortexed, spun and incubated for 45 min at room temperature in the dark. After removal of the supernatant, the gel pieces were washed with 100 μL 25 mM NH_4HCO_3 and vortexed for 10 min. Following removal of supernatant, the gel pieces were dehydrated with 100 μL 25mM NH_4HCO_3/ACN, vortexed for 5 min, spun

briefly and the supernatant discarded. The last step was repeated once and dried completely. Trypsin solution (12.5ng/μL) was added to the samples, rehydrated on ice for 10 min, covered completely with 25 mM NH_4HCO_3 and incubated overnight at 37°C.

The solution from trypsin digestion was transferred to a new tube and 5% formic acid/50% ACN was added to the gel pieces. The gel pieces were vortexed 15 min, spun and sonicated for 15 min. The solution was then transferred to the same vial containing the trypsin digestion solution, vortexed, spun and the volume was reduced around 10-20 μL by vacuum drying. The C18 Zip Tip step proceeded as in section 3.2.2.1.1.

3.2.2.2 In-Solution Digestion

Protein solution volume containing 30 μg of protein was increased to 100 μL with 50mM NH_4HCO_3 and 5 μL 200 mM DTT (in 100mM NH_4HCO_3) was added to reduce the sample by boiling it for 10 min followed by incubation at room temperature for 45-60 min. Alkylation was done by adding 4 μL 1M iodoacetamide to the samples, vortexed, spun briefly and incubation at room temperature for 45-60 min. Neutralization of iodoacetamide proceeded by the addition of 20 μL 200 mM DTT, vortexed, spun and incubation at room temperature for 45-60 min. Trypsin digestion was carried out in a ratio of 1:50 to 1:20 of trypsin to sample where the samples were vortexed and spun briefly prior to overnight incubation at 37°C. 2 % formic acid was added to the sample until the pH reached pH 6.0 using a pH paper indicator. Sample cleanup with C18 Zip Tip followed as in section 3.2.2.1.1.

3.2.3 Mass Spectrometry

The second phase was to compare the phloem proteome profiles from total phloem sap protein and excised gel bands. Measurement was by direct LC-MS/MS analysis. Total proteins (20-50μg) and bands of proteins from PAGE analysis were subjected to in-solution and in-gel digestion respectively prior to LC-MS/MS. Gel bands were run using Agilent 1100 system nanoflow LC where 8μl injection was loaded onto a trap column. From there, the samples moved to C18 nanoflow analytical column (Agilent) which was

washed isocratically with 2.5% (v/v) acetonitrile and 0.1% formic acid. The gradient took approximately 59 min at a flow rate of 0.30 µL/min from 5% acetonitrile and 0.1 % (v/v) formic acid to 95% acetonitrile and 0.1% formic acid in water. The analysis was run at 40^{o}C. MS data were generated on a linear IT (ion trap) quadrupole mass spectrometer (QTrap, AB Sciex Instruments) with an acquisition program for the four most intense peaks set at *m/z* between 400 to 1500 and charge state +2 to +3 chosen for collision induced dissociation (MS/MS). Rolling collision energy was applied. For the total proteins, a similar procedure was employed but using capillary flow LC (Agilent) at a gradient system of 0% B to 40% B over 45 min at a flow rate of 10ul/min [(A) 94.9% H_2O, 5% CH3CN, 0.1% formic acid; (B) 94.9% (v/v) CH3CN, 5% (v/v) H_2O, 0.1% formic acid] and the analysis run at 60^{o}C.

3.2.4 Multiple Reaction Monitoring (MRM)

The samples were analysed on an Applied Biosystems QTrap MS coupled to an Agilent 1100 LC stack. The Agilent stack consisted of a Binary pump, Capillary pump, Well Plate autosampler and a column oven with integrated 6 port valve. The system was configured to load samples onto a trap column (Agilent Zorbax SB 5umx 0.3mm x 35mm) using the binary pump; the trap column was washed and then switched into the capillary flow; peptides were separated on a capillary column (Agilent SB 5um 0.5mm x 150mm column). The LC was interfaced to the MS with a Turbo Ion Spray Source. The loading/washing solvent was H_2O containing 0.2% (v/v) formic acid 0.02% (v/v) TFA at a flow rate 150ul/min and the resolving solvent was a gradient system of 0% B to 40% B over 45 min at a flow rate of 10ul/min [(A) 94.9% H_2O, 5% CH3CN, 0.1% formic acid; (B) 94.9% (v/v) CH3CN, 5% (v/v) H_2O, 0.1% formic acid]. The column oven was heated to 40^{o}C; the valve was switched to direct the flow from the trap into the resolving column after a 5min wash. Typically the MS parameters were set to - Curtain Gas 10psi, GS1 20psi, GS2 20psi, Interface heater on, TEM 150^{o}C, CAD 2, DP 65, EP 10, CXP 7, CE (collision energy) was set according to either empirically determined values or that estimated by MIDAS software (Applied Biosystems).

3.2.5 Database Searches

The MS/MS spectra were subjected to search against MASCOT (www.matrixscience.com) in general and *Arabidopsis thaliana* databases. For protein identification, search parameters were as follows: digestion enzyme-trypsin; fixed modification-carbamidomethylation at cysteine; variable modification-oxidation at methionine; peptide mass tolerance-1.2 Da; fragment mass tolerance-0.6 Da; 1 missed cleavage allowance. Default settings were used for other parameters.

ProteinPilot (Applied Biosystems/MDS Sciex) was also used as an alternative to MASCOT. Protein identification was based on the following parameters: cys alkylation – iodoacetamde, digestion – trypsin, special factors – gel-based ID, species – *Arabidopsis thaliana*, ID focus – biological modifications, database – UniProt, mode of search - thorough ID and detected protein threshold [unused protein score (confidence)] – 2.0 (95.0%). The results obtained from both search engines were pooled. Protein annotations according to GO (Gene Ontology) and MapMan was also carried out where protein were assigned according to functional categories on TAIR (http://www.arabidopsis.org) and BAR (http://bar.utoronto.ca/ntools/cgi-bin/ntools_classification_superviewer.cgi). Classification Superviewer program was used to determine frequencies of various MapMan functional categories (Provart and Zhu, 2003).

3.2.6 2D Proteomics

Phloem sap protein extracted from exudation technique as in 3.2.1 were used and cleaned up, subjected to isoelectric focusing (IEF) and run on 2D gel according to section 3.2.2.2 and 3.2.2.3.

Gel spots were cored out and trypsin digestion was performed as in 4.2.2.1.1 while LC-MS/MS analysis and database search were as in 4.2.3 and 4.2.5 respectively.

3.3 Results

3.3.1 General Comparison of Proteins Identified Across Different Genotypes

Independent samples of phloem sap exudates from five genotypes including wild type were collected using the optimised EDTA exudation technique and subjected to SDS-PAGE. Band patterns of proteins from all genotypes were compared and it was observed that the pattern did not differ substantially from one to another (Fig 3.1).

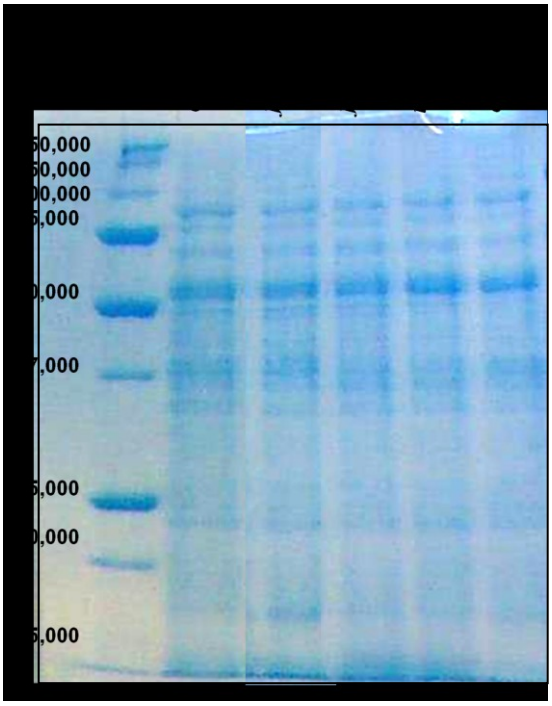


Figure 3.1: Phloem protein obtained from five different genotypes. Exudates were collected from plants after 6 weeks in SD and prior to exudation were exposed to LD for 3 days. Proteins (5 µg) were run on 17% SDS-PAGE gel and stained with Coomassie blue. *Col-0*=wild type and the mutants are *ft-10=flowering locus T, tsf=twin sister of FT, co=constans* and *ft-tsf*=double mutant.

From the gel, the bands were divided into 15 sections (Fig. 3.2) which were then used for protein identification via LC-MS/MS. Some gel pieces contain more than one band pooled together due to their close proximity and difficulty to separate them into individual band. For comparison purposes, phloem sap proteins from wild type plants were stained using either Coomassie or silver. However, excised gel bands from silver stained gel produced only a small number of spectra during LC-MS/MS and the spectra could not be identified as the peptides were too short. On the other hand, no problem was

encountered for excised gels from Coomassie stained gel and abundant spectra were observed and identified. An average of 119 identified proteins was obtained from excised gel bands in relation to sections of gel pieces and size as shown in Table 3.1. Nevertheless, the actual number included in Table 3.2 (Section 3.3.3) was reduced (60 identified proteins) after minimum scoring value was applied.

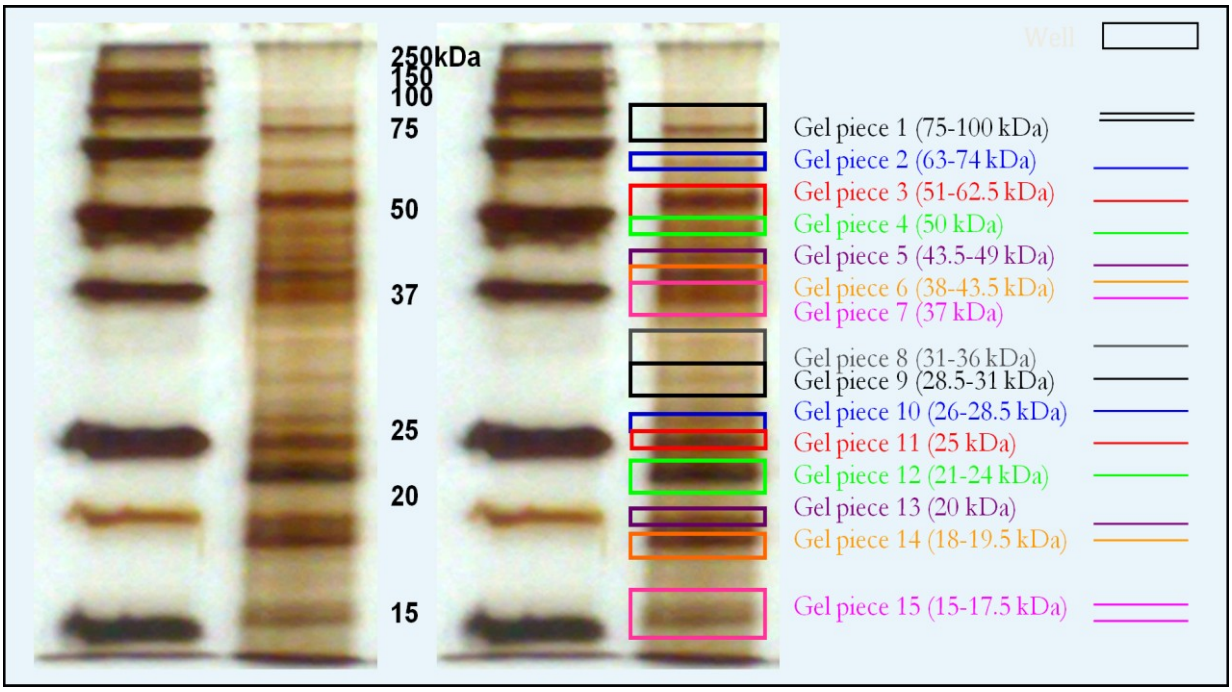


Figure 3.2: How gel bands were selected and categorized for LC-MS/MS analysis. Protein bands from SDS-PAGE gel (left) were divided into 15 sections (right) and excised for tryptic digestion prior to LCMS detection and data analysis.

Table 3.1: Average number of proteins identified from excised gel bands according to sections and size.

Gel pieces / section	Size (kDa)	Total number of identified protein	Average number of identified protein accepted based on minimum scoring value	
			Phloem protein	Cell wall protein and contaminants
1	75 – 100	8	4	1
2	63 – 74	8	3	1
3	51 – 62.5	14	6	1
4	50	7	1	0
5	43.5 – 49	8	5	1
6	38 – 43.5	14	6	4
7	37	8	1	1
8	31 – 36	9	6	1
9	28.5 – 31	7	2	1
10	26 – 28.5	9	3	0
11	25	2	0	1
12	21 – 24	3	1	0
13	20	6	4	0
14	18 – 19.5	3	3	0
15	12 – 17.5	4	3	0
	Total:	**110**	**48**	**12**

In addition to gel bands, total phloem proteins were also sent for LC-MS/MS by employing in-solution tryptic digestion (section 3.2.2.2). This was to ensure that all proteins in the phloem sap are accounted for as some may be left or missed out during gel band excision. A more robust approach but with lesser sensitivity using capillary LC-MS/MS was used instead of nano-capillary which was used for excised gel bands. As expected, more proteins were identified when total phloem protein is used as compared to excised gel bands (Table 3.2). In addition, the largest number of identified proteins came from wild type plants. Based on this, further work was carried out mostly on wild type plants. Nevertheless, it was observed that certain proteins detected in one mutant plant were not detected in another or wild type. In the end, all possible proteins across all five genotypes were combined and the results from LC-MS/MS analysis are shown in section 3.3.2.

3.3.2 Comparison between Different Treatments

Various treatments were employed in this study not only to identify but also to investigate variation of proteins detectable in each treatment. The first was to compare between plants treated using optimised condition (6 weeks in SD followed by 3 days in LD prior to exudation) and those grown solely in SD condition up until exudation period. A slightly lower number of proteins were identified in sample of plants grown solely in SD as compared to the sample exposed to LD for 2 or 3 days prior to exudation. This is expected based on the fact that *Arabidopsis* is a long day plant and flowered earlier. Genes that activate flowering for example would have been activated earlier hence producing more detectable protein compared to SD plants. In terms of quantity of protein based on Bradford Assay (data not shown), the two treatments produced similar amount of protein. Therefore for subsequent analysis only samples from plants grown in SD and prior to exudation were exposed for 3 days in LD were sent for LC-MS/MS analysis.

One of the treatments applied to the optimised exudation condition was ultra-centrifugation of exudates. This treatment was done based on assumption that as wound was made on the petiole it did not only cause rupture of the sieve elements (SEs), companion cells (CCs) and phloem parenchyma but a number of other cells nearby. Samples may therefore include the cell wall and membrane. By spinning at a high speed (100,000 *x g*), it is hoped that cell wall and membrane bound proteins would be pelleted down leaving only the proteins from ruptured SEs in the supernatant.

The last experiment was to collect the exudates in 2 stages. Right after wounding once the cut tip of the petiole had been washed with EDTA, the first exudates were collected within 30 min after wounding and the second followed immediately after with the normal duration (overnight). As observed in Figure 3.3, the concentration of protein from the second exudates was higher than the first due to time constraint of the first exudates which were collected in 30 min. From the total identified proteins (Table 3.1 in section 3.3.2), 8% came from the first exudates while 25.1% from the second exudates but no new protein was identified from both of these samples as compared to the normal method.

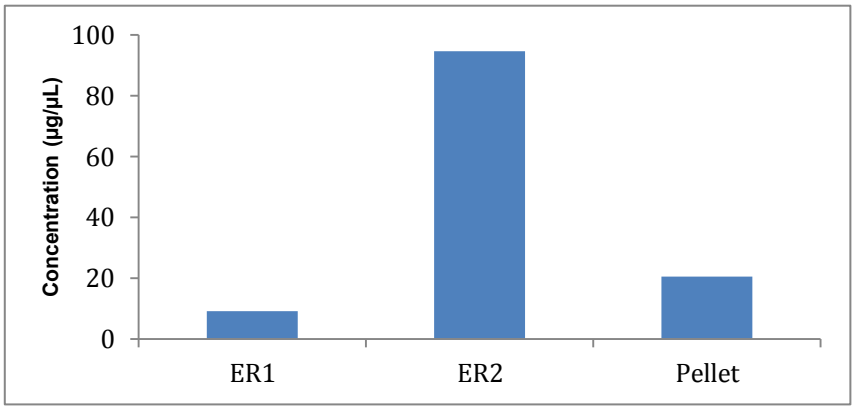


Figure 3.3: Concentration of protein from exudates collected 30 min after wounding, overnight from the same leaves and the pellet of both exudates after ultracentrifugation was measured using Bradford Assay, and compared against the standard curve performed on BSA. Experiments for each exudate were repeated two times

3.3.3 Data Analysis by MASCOT and ProteinPilot

Spectra obtained from LC-MS/MS were analyzed and searched against *Arabidopsis thaliana* using MASCOT initially and later on with ProteinPilot. Search parameters against MASCOT were based on the enzyme trypsin, monoisotopic mass value, unrestricted protein mass, fixed modification of carbamidomethyl at cysteine (C), variable modification is allowed for oxidation at methionine (M), peptide mass tolerance of ±2 Da, fragment mass tolerance of ± 0.8 Da and maximum missed cleavage of 1. The same parameters were also used for ProteinPilot but fixed modification, variable modification, peptide mass tolerance, fragment mass tolerance and maximum missed cleavage were not set as these were default settings of the program. In ProteinPilot a parameter not found in MASCOT, referred to as ID focus, was set on biological modifications.

Proteins from both total and excised gel bands were analyzed and compared. Initially, MS spectra were analysed using MASCOT specifying against *Arabidopsis thaliana* protein database. More than 200 proteins were identified but there were also a huge number of unidentified peptides (>1,000). The unidentified peptides were mostly short peptides comprising of at least 4 but up to 17 amino acids. Some of these peptides did come back with a hit when subjected to BLAST analysis using protein blast (blastp) in NCBI. However, for those with a hit the score values are very low (<30) hence they were

ruled out. Later on when ProteinPilot was adapted for analysis, the number of confirmed proteins increased. Different software applied different algorithms in its analysis and this is the case for ProteinPilot. Additionally, several spots from 2D gels were also excised and sent for analysis with LC-MS/MS.

When compared, both softwares produced different number and score level of proteins. There were proteins identified using one software but not the other and vice-versa. Nevertheless, the majority of high scored proteins were detected using both softwares. There was also the issue of protein redundancy with proteins from the same family or group identified more than once. Many proteins have only a single peptide with coverage limited to conserved sequences between family members. To overcome this problem, AGI (*Arabidopsis* Gene Index) was employed hence proteins with different AGI number were still counted even though they came from the same family or group.

During compilation or listing of proteins from the *Arabidopsis* phloem exudate, certain criteria were imposed to ensure the quality and confidence of the proteome database. Among the criteria were high score value given by the software used which was set at minimum 55 for MASCOT and every scoring for ProteinPilot ($p<0.05$) was accepted (Appendix A). Additionally, the number of unique peptides was among criteria for robust identification where proteins with more than one unique peptide are highly desirable.

Once the criteria were set, proteins from both softwares that has been analysed were combined. The list of all protein identified in *Arabidopsis* phloem proteome is given in Table 3.2. A total of more than 2000 different peptides were detected overall but the number was reduced down after elimination of proteins with low score value and unidentified peptides. A total of 221 proteins were reliably identified. The proteins were then compared against those in phloem proteome databases of Brassica (*Brassica napus*), Cucurbit (*Cucumis sativus* and *Cucurbita maxima*), Melon (*Cucumis melo*) and Rice (*Oryza sativa*). AGI number was used for comparison as well as annotations because rice proteins have their own unique identifiers while some of the proteins identified in brassica were based on *Brassica napus* and NCBI (National Centre for Biotechnology Information) accession number. In addition, the proteins in the list were also compared with proteins identified from phloem exudate of *Arabidopsis* via EDTA-facilitated method (Guelette et

al., 2012; Batailler et al., 2012;) (Table 3.2) and *Arabidopsis* phloem RNAs (Deeken et al., 2008) (Appendix A).

The list in Table 3.2 was obtained mostly based on scores from both MASCOT and ProteinPilot. Most importantly it is desirable that the proteins have more than one unique peptide and are found more than once throughout the repetitions of LC-MS/MS carried out on them (Fig. 3.4). In the list, 33.5% of the proteins have one unique peptide from its spectra but they are still taken into consideration based on their uniqueness, where some of them are considered new or not found or to be similar to any other proteins in other plant proteome databases and of course some have high similarity to proteins from other plant proteome databases. Observation was also made on influence of sample preparation and treatment whereby all of the proteins were identified from samples of non-ultracentrifuged samples including total phloem sap protein, excised gel bands and 2D spots; 28.3% from samples subjected to ultracentrifugation, 27.4% from excised gel bands and 4.9% from 2D spots. However, analysis of protein using LC-MS/MS for ultracentrifuged samples was only performed once as compared to up to 4 repetitions of earlier analysis of samples without ultracentrifugation. The ultracentrifugation experiment itself was applied rather late in the study and with time constraint repetitions could not be carried out. This could influence the number of proteins found as a different number of identified proteins, including new ones, was obtained with each run. Further experiments using 2D proteomics were discontinued due to the failure of obtaining reproducible gels (Appendix F) and relative ease of successful analysis of total protein and 1D analysis. Finally, the genotypes in which each protein was identified were also noted.

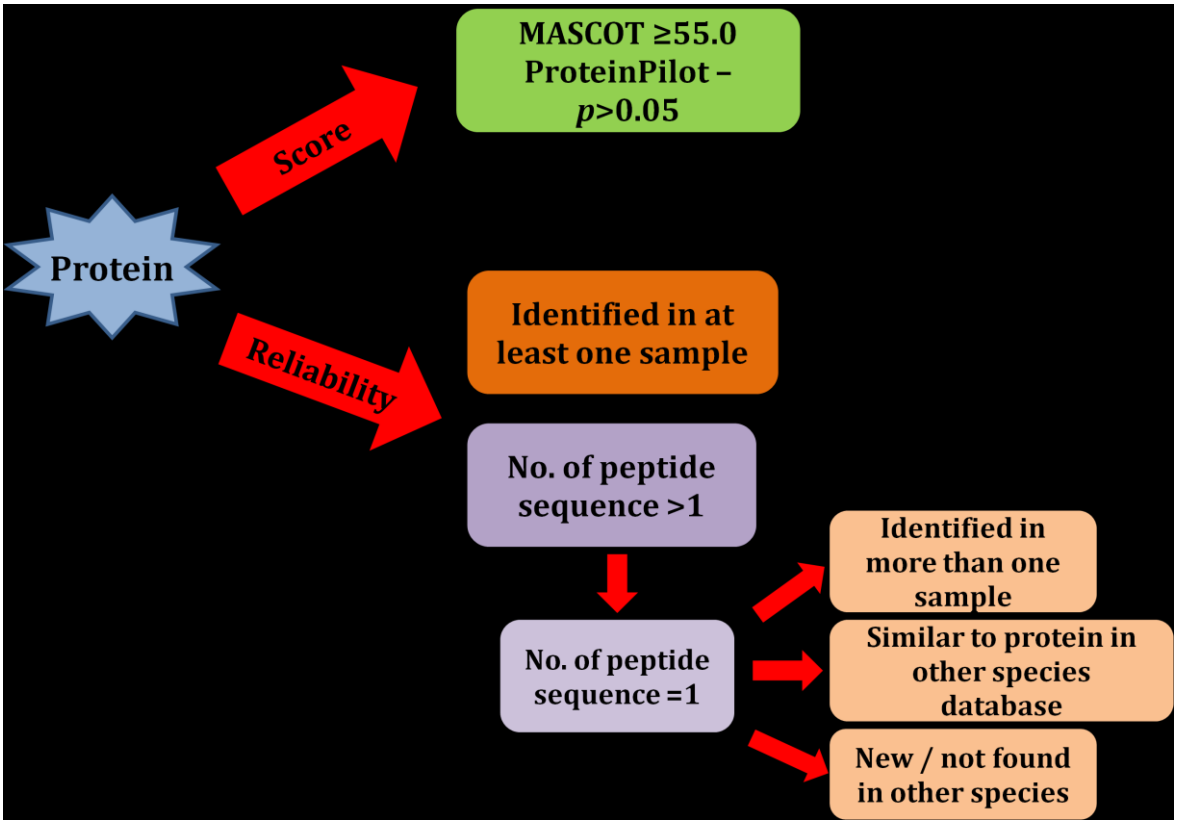


Figure 3.4: Flow diagram of selection of identified proteins following LC-MS/MS analysis.

Table 3.2: Complete list of identified proteins from *Arabidopsis* phloem exudate detected in various samples treatments and genotypes. Leaves were wounded to obtain the exudates however the proteins were not confirmed as wound-inducible phloem exudates protein. MS spectra were analysed using MASCOT and ProteinPilot against *Arabidopsis thaliana* database. Each protein is assigned under its functional class according to MapMan categories on TAIR. Identified proteins were compared against brassica, cucurbit and rice phloem proteome for homology. AGI no. was used for comparison with the exception of rice where homology was based on annotation. Annotation of similar protein were also noted for protein from brassica and cucurbit with unmatched AGI no. Comparison was also made between *Arabidopsis* phloem proteome studies by Batailler et al (2012) and Guelette et al (2012).

Represents all proteins identified by Batailler et al (2012) in the phloem exudate

Represents all proteins identified by Batailler et al (2012) in the full set of exudates

Proteins with shared homology in all four species are written in blue

∗ *Arabidopsis* phloem exudate proteins identified by Guelette et al (2012) via EDTA-facilitated method

	AGI	Protein ID	Mass observed (Da)	No. of sequence (ESI-MS)	Homology (other species)	Detected in different samples and preparation	Genotype
	Redox Reaction						
1	At1g07890	Ascorbate peroxidase 1 (APX16)	27413	4	B,C, R	NUC, UC	WT,c
2	At4g35090	Catalase 2 (CAT)	57308	1	C, R	NUC, UC,b	WT
3	At1g20620	Catalase 3 (CAT3)	55932	4	-	NUC, UC	c
4	At1g08830∗	Copper/zinc superoxide dismutase (CSD1) ∗	15259	5	C, R	NUC, UC	D,t,c

5	At2g28190*	Copper/zinc superoxide dismutase (CSD2) *	22322	2	-	NUC, UC	WT
6	At1g19570	Dehydroascorbate reductase (DHAR1)	23626	5	-	NUC	WT,f,t,c
7	At4g11600	Glutathione peroxidase 6 (GPX6)	18597	2	C, R	NUC	t
8	At2g24940	Membrane-associated progesterone binding protein 2 (MAPR2)	11002	4	-	NUC, UC	WT,f,D,t, c
9	At5g40370	Membrane-associated progesterone binding protein 2 (MAPR2)	11752	7	R	NUC, UC	WT,f,D,t, c
10	At3g52880*	Monodehydroascorbate reductase 1 (MDAR1) *	46500	7	B,C	NUC, b	WT
11	At1g63940	Monodehydroascorbate reductase 6 (MDAR6)	53299	3	B	NUC,b	WT
12	At5g03630	Pyridine nucleotide-disulphide oxidoreductase family protein (MDAR2)	46433	2	C	NUC, b	WT
13	At1g45145	Thioredoxin 3 (TRX3)	13227	3	B,C,R	NUC	WT,c
14	At3g02730	Thioredoxin F-type 1 (TRXF1)	19313	2	C	NUC, UC	WT,D,c
15	At5g16400	Thioredoxin F2 (TRXF2)	19966	1	-	NUC	c
16	At5g42980*	Thioredoxin H-type 3 (TRX-H-3) *	13112	13	B	NUC, UC	WT,f,D,t, c
17	At1g03680*	Thioredoxin M-type 1 (TRX-M1) *	19652	3	-	NUC, b	WT
18	At3g15360	Thioredoxin M-type 4 (TRX-M4)	21403	2	-	NUC	WT
19	At1g43560	Thioredoxin Y2 (Aty2_ty2)	18581	1	-	NUC	WT
20	At2g20270	Thioredoxin superfamily protein	19125	2	-	NUC	WT
21	At3g11630	Thioredoxin superfamily protein	29074	2	-	b	WT
22	At3g52960	Thioredoxin superfamily protein	24697	1	-	NUC	WT
23	At4g03520	Thioredoxin superfamily protein (ATHM2)	20314	5	-	NUC, b	WT,D
	Metabolic Pathway **A) Glycolysis**						
24	At2g24270	Aldehyde dehydrogenase 11A3 (ALDH11A3)	53022	1	C, R	NUC	WT
25	At2g36530*	Enolase (ENO2) / (LOS2) *	47719	3	B,C, R	NUC,b	WT

26	At5g42740	Glucose-6-phosphate isomerase, cytosolic (GPI) (phosphoglucose isomerase) (PGI)	61589	1	-	NUC	WT
27	At3g04120*	Glyceraldehyde-3-phosphate dehydrogenase C subunit 1 (GAPC1)*	36491	3	B,C,R	NUC,UC,b	WT
28	At3g55440*	Triosephosphate isomerase (TPI) *	27111	8	C, R	NUC,UC,b	WT,f
29	At3g03250*	UDP-glucose pyrophosphorylase 1 (UGP1) *	51706	9	B	NUC,UC,b	WT
30	At5g17310	UDP-glucose pyrophosphorylase 2 (UGP2)	51920	6	B	NUC,UC,b	WT
	B) Gluconeogenesis						
31	At5g09660	Peroxisomal NAD-malate dehydrogenase 2 (PMDH2)	37669	4	-	NUC	WT
	C) TCA						
32	At3g01500*	Carbonic anhydrase 1 (CA1) *	37450	5	-	NUC, 2D	WT
33	At5g14740	Carbonic anhydrase 2 (CA2)	36462	5	-	NUC,UC, 2D	WT
34	At1g04410	Lactate/malate dehydrogenase family protein (c-NAD-MDH1)	35548	4	B, R, C	NUC, UC	WT,t,c
35	At1g53240	Lactate/malate dehydrogenase family protein (mMDH1)	35782	4	B, C, R	NUC,UC,b	WT,f,D,t
36	At5g11670	NADP-malic enzyme 2 (NADP-ME2)	64333	2	-	b	WT
	Amino Acid Metabolism						
37	At4g14880	O-acetylserine (thiol) lyase (OAS-TL) isoform A1 (OASA1)	33805	3	C, R	b	WT
38	At4g13940	S-adenosyl-L-homocysteine hydrolase (SAHH1)	39994	14	B, C	NUC, b	WT
39	At3g23810	S-adenosyl-l-homocysteine (SAH) hydrolase 2 (SAHH2)	53159	3	-	NUC	WT
40	At1g02500	S-adenosylmethionine synthetase 1 (SAM1)	41453	1	B, C	NUC	WT
41	At4g31990	Aspartate aminotransferase 5 (ASP5)	49328	1	-	NUC	WT
42	At5g17920	Cobalamin-independent synthase family protein (ATCIMS)	84267	11	B, R	NUC, UC,b	WT,f,t,c
	Lipid Metabolim						
43	At1g31812	Acyl-CoA-binding protein 6 (ACBP6)	10466	3	C, R	NUC, UC	WT,f,D,t,

							c
44	At4g25050	Acyl carrier protein 4 (ACP4)	14536	2	-	NUC	WT,D,t
45	At5g16390	Chloroplastic acetylcoenzyme A carboxylase 1 (BCCP1 / CAC1A)	29613	2	-	NUC	WT
46	At1g27950	Glycosylphosphatidylinositol-anchored lipid protein transfer 1 (LTPG1)	19777	3	-	NUC, UC	WT
47	At2g38540	Lipid transfer protein 1 (LTP1)	11747	9	-	NUC, UC	WT,D
48	At2g38530	Lipid transfer protein 2 (LTP2)	11938	3	-	NUC	WT
49	At1g66970	SHV3-like 2 (SVL2)	83788	1	-	NUC	D
	Nucleotide Metabolism						
50	At3g09820	Adenosine kinase 1 (ADK1)	37812	7	-	NUC,UC,b	WT,f,D,t
51	At5g03300	Adenosine kinase 2 (ADK2)	31337	7	B,C, R	NUC, UC	WT
52	At4g09320	Nucleoside diphosphate kinase family protein (NDPK1)	15740	2	B, C	NUC	WT
53	At5g63310	Nucleoside diphosphate kinase 2 (NDPK2)	25463	1	-	NUC	WT
54	At5g09650*	Pyrophosphorylase 6 (PPa6) *	33373	3	-	NUC	WT,c
	Secondary Metabolism						
55	At3g24503	Aldehyde dehydrogenase 2C4 (ALDH1A / ALDH2C4)	54460	1	-	NUC	WT
56	At1g54040	Epithiospecifier protein (ESP)	37069	3	B	NUC	c
57	At3g44310	Nitrilase 1 (NIT1)	38152	1	R	NUC	WT
58	At5g26000*	Thioglucoside glucohydrolase 1 (TGG1) *	61133	10	B, C, R	NUC,UC,b	WT,D
	Metabolism – Miscellaneous A) **C1 Metabolism**						
59	At2g41530	S-formylglutathione hydrolase (SFGH)	30474	5	-	NUC, b	WT,f
60	At3g59970	Methylenetetrahydrofolate reductase 1 (MTHFR1)	66246	2	-	b	WT
	B) Hormone Metabolism						
61	At3g16430	Jacalin-related lectin 31 (JAL31)	32133	2	-	b	WT

62	At3g16470*	Mannose-binding lectin superfamily protein (JAL35_JR1) *	48698	13	B	NUC, UC	WT,D,t,
	C) Major CHO Metabolism						
63	At1g12240	Glycosyl hydrolases family 32 protein / Beta-fructosidase (ATBETAFRUCT4)	73465	3	-	NUC	WT
	D) Minor CHO Metabolism						
64	At5g37850	pfkB-like carbohydrate kinase family protein (SOS4)	34232	1	C, R	NUC, b	WT
65	At1g23740*	Oxidoreductase, zinc-binding dehydrogenase family protein *	40986	4	-	NUC	WT
66	At5g61130	Plasmodesmata callose-binding protein 1 (PDCB1)	20365	2	R	NUC	WT
	E) Nitrogen Metabolism						
67	At5g35630	Glutamine synthetase 2(GS2)	47256	6	B	2D	WT
	F) Tetrapyrrole Synthesis						
68	At5g08280*	Hydroxymethylbilane synthase (HEMC) *	39724	4	-	NUC	WT
69	At5g63570	Glutamate-1-semialdehyde-2,1-aminomutase (GSA1)	50222	2	-	NUC	WT
	G) Vitamin Metabolism						
70	At4g26500	Chloroplast sulfur E (SUFE1)	40888	1	C, R	NUC	WT
71	At5g54770*	Thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) *	17637	1	-	NUC	WT
	Photosynthesis A) **Calvin Cycle**						
72	At2g21330	Fructose-bisphosphate aldolase 1 (FBA1)	33302	4	B	NUC,b	WT
73	At4g38970*	Fructose-bisphosphate aldolase 2 (FBA2) *	42935	1	-	b	WT
74	At1g42970	Glyceraldehyde-3-phosphate dehydrogenase B subunit (GAPB)	47532	1	-	NUC	WT
75	At3g12780	Phosphoglycerate kinase 1 (PGK1)	51026	2	-	NUC	WT
76	At1g56190	Phosphoglycerate kinase family protein	49894	7	B, R	NUC,UC,b	WT,f,c
77	At2g21170	Plastid Isoform Triose Phosphate Isomerase (PDTPI)	33505	3	R	NUC	WT
78	At2g39730*	Rubisco activase (RCA) *	51999	30	B	NUC, UC	WT

79	AtCg00490	Ribulose-bisphosphate carboxylases (RBCL)	52669	24	B, C	NUC,UC, b, 2D	WT
80	At1g67090*	Ribulose bisphosphate carboxylase small chain 1A (RBCS1A) *	20003	13	B	NUC,UC,b	WT,D
81	At3g55800*	Sedoheptulose-bisphosphatase (SBPASE) *	42414	3	-	NUC,b	WT
82	At3g60750	Transketolase	79875	2	-	NUC	c
	B) Light Reaction						
83	At1g20340	Cupredoxin superfamily protein (DRT112_PETE2)	16988	6	-	NUC, UC	WT,t,c
84	At5g66190	Ferredoxin-NADP(+)-oxidoreductase 1 (FNR1)	40332	4	-	b	WT
85	At1g20020	Ferredoxin-NADP(+)-oxidoreductase 2 (FNR2)	41142	3	B	NUC,b	WT
86	At1g77090	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein	28501	1	-	NUC	WT
87	At1g06680	Photosystem II subunit P-1 (PSBP-1)	28003	13	-	NUC, UC	WT,f,t,c
88	At4g21280	Photosystem II subunit QA (PSBQ-1)	23888	4	-	NUC, UC	t,c
89	At4g05180	Photosystem II subunit Q-2 (PSBQ-2)	24611	10	-	NUC, UC	WT,f,t,c
90	At1g76100	Plastocyanin 1 (PETE1)	17589	1	-	NUC	WT
	C) Photorespiration						
91	At1g11860	Glycine cleavage T-protein family	44126	2	-	NUC	WT
92	At1g32470	Single hybrid motif superfamily protein	17886	2	-	NUC, UC	c
93	At3g14420	Aldolase-type TIM barrel family protein	40226	1	-	NUC	WT
	Stress **A) Abiotic Stress**						
94	At1g20440	Cold-regulated 47 (COR477)	33349	6	-	NUC	c
95	At1g20450	Dehydrin family protein (ERD10)	29425	5	-	NUC, UC	WT,f,D,t, c
96	At1g76180	Dehydrin family protein (ERD14)	20770	6	-	NUC,UC,b	WT,f,D
97	At1g70830	MLP-like protein 28 (MLP28)	27528	3	-	NUC	c
98	At1g70890*	MLP-like protein 43 (MLP43)	17695	5	B	NUC, UC	WT,f,D,t,

							c
99	At1g23130*	Polyketide cyclase/dehydrase and lipid transport superfamily protein (Bet v I allergen-like)	17848	3	R	NUC, b	WT,t,c
100	At4g23670	Polyketide cyclase/dehydrase and lipid transport superfamily protein	17529	2	-	NUC	D
	B) Biotic Stress						
101	At5g39730	AIG2-like (avirulence induced gene) family protein	20005	2	C	NUC, b	t,c
102	At3g04720	Pathogenesis-related 4 (PR4) / Hevein-like protein (HEL)	22937	4	-	NUC, UC	WT
103	At2g43530	Scorpion toxin-like knottin superfamily protein	9858	2	-	NUC	WT,D
	Cell Organization and Biogenesis						
104	At2g43530	Actin 7 (ACT7)	41774	5	B,C,R	NUC	WT
105	At5g59880	Actin-depolymerizing factor 3 (ADF3)	16080	3	R	NUC, UC	WT, t
106	At1g35720*	Annexin 1 (ANNAT1) *	35757	10	B,C, R	NUC, b	WT
107	At2g19760	Profilin 1 (PRO1)	14240	2	B, C	NUC	WT
108	At4g29350	Profilin 2 (PRO2)	13990	3	B	NUC	WT,t
109	At5g56600	Profilin 3 (PRO3)	14333	1	C, R	NUC	WT
110	At3g63190	Ribosome recycling factor, chloroplast precursor (RRF)	30403	2	-	NUC, UC	f
111	At4g38740	Rotamase CYP 1 (ROC1)	18333	7	R	NUC,UC,b	WT
112	At2g16600	Rotamase CYP 3 (ROC3)	18444	2	B,C, R	NUC	WT
113	At3g62030*	Rotamase CYP 4 (ROC4) *	28190	7	-	NUC,2D,b	WT,f,D,t,c
114	At4g34870	Rotamase CYP 5 (ROC5)	18366	6	-	NUC, UC	WT,D,t,c
115	At5g15970	Stress-responsive protein (KIN2) / stress-induced protein (KIN2) / cold-responsive protein (COR6.6) / cold-regulated protein (COR6.6) (AtCor6.6_COR6.6_KIN2)	6551	6	-	NUC, UC	WT
116	At5g62690	Tubulin beta-2/beta-3 chain (TUB2)	50003	1	-	NUC	WT
117	At4g14960	Tubulin/FtsZ family protein (TUA6)	49226	1	C	NUC	WT

	Signalling						
118	At2g27030	Calmodulin 5 (CAM5)	16652	1	C, R	NUC, UC	WT
119	At3g43810	Calmodulin-7 (CAM7)	16829	1	-	NUC, UC	WT
120	At5g38480	General regulatory factor 3 (GRF3_RCI1)	28600	1	-	NUC	WT
121	At4g09000	General regulatory factor 1 (GF14 CHI_GRF1)	29666	2	-	NUC	WT
122	At3g02520	General regulatory factor 7 (GF14 NU_GRF7)	29811	4	C, R	NUC	WT
123	At1g22300*	General regulatory factor 10 (14-3-3EPSILON_GF14 EPSILON_GRF10) *	28599	3	-	NUC,b	WT
124	At4g20260	Plasma-membrane associated cation-binding protein 1 (PCAP1)	24697	3	-	NUC	D,c
125	At3g59920	RAB GDP dissociation inhibitor 2 (GDI2)	46555	2	C, R	b	WT
	Protein Biosynthesis, Organization, Folding, Transport and Modification						
126	At4g04910	AAA-type ATPase family protein (NSF)	81652	1	C, R	NUC, b	WT
127	At5g20720	Chaperonin 20 (CPN20)	9021	4	-	NUC	WT
128	At5g03940	Chloroplast signal recognition particle 54 kDa subunit (CPSRP54P)	60231	1	-	NUC	WT
129	At5g50920	CLPC homologue 1 (CLPC1)	7883	1	-	NUC	t,c
130	At1g72730	DEA(D/H)-box RNA helicase family protein	46777	1	R	NUC	WT
131	At5g45680	FK506-binding protein 13 (FKBP13)	22039	1	-	NUC	WT,t
132	At5g02500	Heat shock cognate protein 70-1 (HSC70-11)	71358	2	B,C, R	NUC	WT
133	At5g60390	GTP binding Elongation factor Tu family protein	49402	2	B, C	NUC, UC	WT
134	At5g02240*	NAD(P)-binding Rossmann-fold superfamily protein *	27103	1	-	NUC	t
135	At3g12390	Nascent polypeptide-associated complex (NAC), alpha subunit family protein	21982	1	C, R	NUC	WT
136	At4g20360	RAB GTPase homolog E1B (ATRABE1B)	51398	2	-	NUC, UC	WT
137	At2g47170	Ras-related small GTP-binding family protein (ARF1A1C)	21646	1	R	NUC	WT

138	At5g58290	Regulatory particle triple-A ATPase 3 (RPT3)	45711	1	B[a] ,C, R	NUC	WT
139	At1g56070	Ribosomal protein S5/Elongation factor G/III/V family protein (LOS1)	92356	2	B, C	b	WT
140	At1g31340	Related to ubiquitin 1 (RUB1)	8525	1	B,C, R	NUC	c
141	At5g03240	Polyubiquitin (UBQ3)	28169	2	-	NUC	WT
	Developmental Process						
142	At3g16640	Translationally controlled tumor family protein (TCTP)	18872	2	B,C,R	NUC,UC,b	WT,f,t,c
	Transport						
143	At2g01320	ABC-2 type transporter family protein （ABCG7）	78599	1	-	NUC	WT
144	At1g72150	Patellin 1 (PATL1)	64046	1	-	NUC	WT
	Electron Transport						
145	At1g22840	Cytochrome C-1 (CYTC-1)	12390	3	-	NUC,UC,b	WT
	Biodegradation of Xenobiotics						
146	At1g67280	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein	39174	1	R	b	WT
147	At1g11840*	Glyoxalase I homolog (GLX1) *	31669	5	-	NUC	WT
	DNA						
148	At5g38470	Rad23 UV excision repair protein family (RAD23D)	40122	1	C	NUC	WT
	RNA						
149	At3g20390*	Endoribonuclease L-PSP family protein *	15151	3	-	NUC	WT
150	At2g21660*	Glycine-rich RNA-binding protein (GRP7) *	16596	4	B,C, R	NUC, UC	WT
151	At4g39260	Glycine-rich RNA-binding protein 8 (GRP8)	16304	2	B	NUC,UC,b	WT,f,c

152	At4g34110	Polyadenylate-binding protein 2 (PABP2)	67017	1	-	NUC	t
153	At1g49760	Poly(A) binding protein 8 (PABP8)	72733	3	C	NUC	WT
154	At2g23380	SET domain-containing protein (CLF / SET1)	84155	1	-	2D	WT
155	At3g61260	Remorin family protein	23130	1	-	NUC	f,t
156	At4g24770	31-kDa RNA binding protein (RBP31)	35787	1	-	NUC	t,c
	Miscellaneous						
157	At1g29660*	GDSL-like Lipase/Acylhydrolase superfamily protein *	40630	4	-	NUC, 2D	WT
158	At4g02520	Glutathione S-transferase PHI 2 (GSTF2)	24109	3	B[a]	NUC, UC	WT
159	At2g30870	Glutathione S-transferase PHI 10 (GSTF10)	24084	5	B, R	NUC,UC,b	WT
160	At3g16400	Nitrile specifier protein 1 (NSP1)	51670	8	B	NUC, UC	WT
161	At3g16410	Nitrile specifier protein 4 (NSP4)	67555	2	-	NUC	WT
162	At3g10720	Plant invertase/pectin methylesterase inhibitor superfamily	60013	6	-	NUC,b	WT
163	At1g78850	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain	49050	7	-	NUC,b	WT,f,D,t
164	At4g20830	FAD-binding Berberine family protein	63336	2	-	NUC, b	WT
	Unknown						
165	At3g05900	Neurofilament protein-related	72087	1	-	NUC	t,c
166	At1g09310	Protein of unknown function, DUF538	19934	3	B,R	NUC	WT,f,D,t,c
167	At1g55480	Protein containing PDZ domain, a K-box domain, and a TPR (ZKT) region	37387	1	-	NUC	t
168	At5g01750	Protein of unknown function (DUF567)	24295	2	-	NUC	c
169	At5g22580	Stress responsive A/B Barrel Domain	12349	1	C, R	NUC, UC	t
170	At3g02910	AIG2-like (avirulence induced gene) family protein	21333	1	R	NUC	WT
171	At5g19240	Glycoprotein membrane precursor GPI-anchored	21301	1	-	NUC	WT

172	At5g56170	LORELEI-LIKE-GPI-ANCHORED PROTEIN 1 (LLG1)	18904	1	-	NUC	D
173	At2g37660*	NAD(P)-binding Rossmann-fold superfamily protein *	34256	1	-	NUC, 2D	WT
174	At2g44920	Tetratricopeptide repeat (TPR)-like superfamily protein	23744	1	-	NUC	WT
175	At5g53490	Tetratricopeptide repeat (TPR)-like superfamily protein	25628	1	-	NUC	WT
	Cell Wall Protein and Contaminants (Plasma Membrane, Chloroplast, Plastid)						
176	At3g07390	Auxin-responsive family protein (AIR12)	25576	2	-	NUC, UC	WT,f,D,t
177	At5g64570	Beta-D-xylosidase 4 (XYL4)	84228	6	-	b	WT
178	At4g12420	Cupredoxin superfamily protein (SKU5)	65638	4	-	NUC	WT,f,D,c
179	At4g27520	Early nodulin-like protein 2 (ENODL2)	35064	2	-	NUC, UC	WT
180	At5g64570	Eukaryotic aspartyl protease family protein	45792	1	-	NUC, b	t
181	At5g55730	Fasciclin-like arabinogalactan-protein (FLA1)	44802	3	-	NUC	WT,f,D,t, c
182	At4g12730	Fasciclin-like arabinogalactan-protein (FLA2)	43465	6	-	NUC, b	c
183	At2g04780	Fasciclin-like arabinogalactan-protein (FLA7)	26759	1	-	NUC	f,D
184	At2g45470	Fasciclin-like arabinogalactan-protein (FLA8)	43125	3	-	NUC	t,c
185	At1g03870	Fasciclin-like arabinogalactan-protein (FLA9)	26129	1	-	NUC	c
186	At3g60900	Fasciclin-like arabinogalactan-protein (FLA10)	43596	3	-	NUC	WT
187	At5g44130	Fasciclin-like arabinogalactan-protein (FLA13)	26300	1	-	NUC, UC	WT
188	At1g29670	GDSL-like Lipase/Acylhydrolase superfamily protein	39863	8	-	NUC, b	WT, f, D, t c
189	At5g20630	Germin 3 (GER3)	21849	4	-	NUC, 2D	WT,c
190	At4g16260	Glycosyl hydrolase superfamily protein	37688	1	-	b	WT
191	At1g09750	Eukaryotic aspartyl protease family protein	39751	2	-	NUC, b	f,D
192	At5g20080	FAD/NAD(P)-binding oxidoreductase	35663	1	C, R	NUC	WT

193	At1g75040*	Pathogenesis-related protein 5 (PR-5) *	25236	12	-	NUC, UC, b, 2D	WT,f,D,t, c
194	At3g49120	Peroxidase CB (PRXCB)	38927	5	C	NUC, 2D,b	WT,f,D,t, c
195	At2g02850	Plantacyanin (ARPN)	14038	1	-	NUC	WT
196	At3g16420	PYK10-binding protein 1 (JAL30_PBP1)	32096	9	C	NUC, b	WT,f,t,c
197	At1g76160	SKU5 similar 5 (sks5)	60003	5	-	b	WT
198	At5g67360	Subtilase family protein (ARA12)	73556	1	-	NUC,b	WT
199	At4g37800	Xyloglucan endotransglucosylase/hydrolase 7 (XTH7)	33623	8	-	NUC	WT
200	At4g30270	Xyloglucan endotransglucosylase/hydrolase 24 (XTH24)	30576	7	-	NUC,UC,b	WT
	Ribosomal Protein						
201	At2g27710	60S acidic ribosomal protein family	11444	1	C, R	NUC	WT
202	At1g32990	Plastid ribosomal protein L11 (PRPL11)	23155	1	-	NUC	WT
203	At2g18020	Ribosomal protein L2 family	27843	1	-	NUC	WT
204	At3g25520	Ribosomal protein L5 (RPL5A)	34583	1	-	NUC	WT
205	At1g74050	Ribosomal protein L6 family protein	26111	1	-	NUC	WT
206	At3g62870	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	29000	2	-	NUC	WT
207	At2g44120	Ribosomal protein L30/L7 family protein	27933	1	-	NUC	WT
208	At3g09200	Ribosomal protein L10 family protein	34166	1	-	NUC	WT
209	At5g60670	Ribosomal protein L11 family protein	17704	1	B	NUC	f,t,c
210	At3g27850	Ribosomal protein L12-C (RPL12-C)	19670	2	-	NUC, UC	WT,f
211	At2g42740	Ribosomal protein large subunit 16A (RPL16A)	20956	2	C, R	NUC	WT
212	At3g55280	Ribosomal protein L23AB (RPL23A2_RPL23AB)	17299	1	-	NUC	WT
213	At3g53870	Ribosomal protein S3 family protein	27333	1	-	NUC	WT
214	At2g17360	Ribosomal protein S4 (RPS4A) family protein	29654	3	-	NUC	WT
215	At2g41840	Ribosomal protein S5 family protein	30888	2	B	NUC	WT
216	At2g37270	Ribosomal protein 5B (RPS5B)	22954	1	-	NUC	WT

217	At5g10360	Ribosomal protein S6e (RPS6B)	29111	2	-	NUC	WT
218	At5g20290	Ribosomal protein S8e family protein	24991	2	-	NUC	WT
219	At3g11510	Ribosomal protein S11 family protein	16263	2	B	NUC	f
220	At4g00100	Ribosomal protein S13A (RPS13A)	17006	1	-	NUC	WT
221	At1g22780	Ribosomal protein S13/S18 family (RPS18A)	17558	1	-	NUC	WT

NB:B=Brassica, C=Cucurbit, R=Rice, [a]=annotated, b=excised gel band, NUC=non-ultracentrifuged, UC=ultracentrifuged, 2D=2D proteomics gel spot, WT=wild type, f=ft-10, D=double mutant ft-tsf, t=tsf, c=co.

Table 3.2 shows a few notable comparisons. The first was to find homologous proteins shared by three other phloem proteome databases. From the information gathered, a Venn diagram was drawn to display overlapping and unique subsets of each plant against *Arabidopsis* (Fig 3.5 (a)). Among the four species, rice is the only monocot while the rest are dicots. It is possible that the mechanism of phloem trafficking might differ between monocot and dicot plants. There are only 16 proteins that all plants shared across all four proteomes. Therefore, the percentage of the homologous proteins detected from each species would vary depending on the total number of identified proteins. These current databases for each species with the exception of brassica, have a high number of unique proteins in their own phloem proteome. The percentage of overlapping proteins among all the three dicots (*Arabidopsis*, brassica and cucurbit) was rather small (3.6% in *Arabidopsis*, 7.8% in brassica and 1.3% in cucurbit). Among the three plants, *Arabidopsis* shared the highest overlapping proteins with cucurbit (23.5%) followed by rice (22.6%) and brassica (21.7%).

The second comparison was made against *Arabidopsis* phloem exudate database from Batailler et al. 2012. Both studies differed slightly in terms of the method from which the exudate was obtained: Batailler et al (2012) obtained the phloem exudates protein by means of EDTA-facilitated method whereby the final phloem exudates were collected in buffered EDTA. Additionally, they determined cell wall proteins and other cell wall contaminants including plasma membrane and excluded those from their list of identified phloem proteins. A larger number of proteins were identified from their study as shown in the Venn diagram (Fig 3.5 (b)). It was observed that proteins identified from the current study shared high similarity with the proteins from Batailler et al (41.6%), and 8.1% proteins from current study were confirmed as cell wall proteins or contaminants. Comparison with another phloem exudate protein by Guelette et al (2012) showed that only 14.5% proteins from the current study are similar to theirs (Fig. 3.5(c)). Guelette et al was the first to provide a database for phloem protein from *Arabidopsis* by means of EDTA-facilitated method where the final phloem exudates were collected in deionised water. They managed to identify a small number of proteins, the smallest among the three studies in *Arabidopsis*. A final comparison among the three proteins showed that 11.8% of identified proteins from the current study were also identified by Guelette et al and

Batailler et al (Fig. 3.5 (d)). From Fig. 3.5 (d) each of the three studies still has a considerate number of unique proteins only found in each respective study.

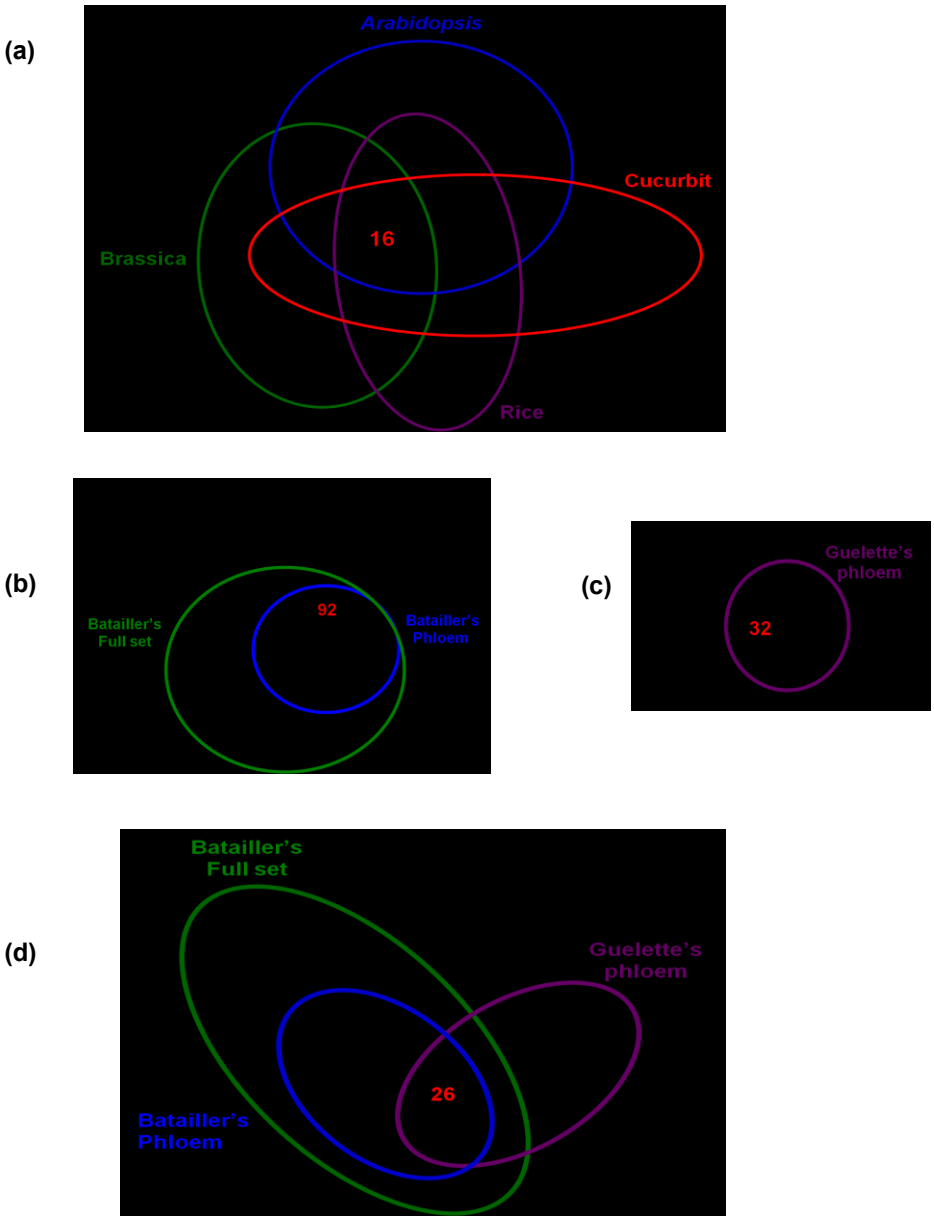


Figure 3.5: Comparisons of identified phloem exudate proteins are represented as Venn diagrams. (a) *Arabidopsis* phloem proteome was compared to three other species namely *Brassica napus* (rapeseed oil), *Cucurbita maxima* (pumpkin) and rice (*Oryza sativa*). The homologues that all four species shared are exhibited by the centre value in highlighted red. Identified proteins were also compared to *Arabidopsis* phloem proteome established by EDTA-facilitated method from (b) Batailler et al (2012) comprising of full set of proteins from phloem exudates and confirmed phloem protein without cell wall or

plasma membrane contaminant, and, from (c) Guelette et al (2012). (d) Identified proteins from the three *Arabidopsis* studies were then compared. The number of proteins shared between the studies in (b), (c) and (d) is highlighted in red. For all diagrams, values in brackets represent the total number of protein reported in each species or study. The rest of the values represent the subset of proteins for individual species or study, and, their correlation with other species or study in their respective diagram.

As shown in Table 3.2, each identified protein was assigned under their respective functional class based on MapMan categories. Apart from comparing the proteins for homology to phloem proteomes from other species and between phloem proteome studies in *Arabidopsis*, the proteins were also compared against *Arabidopsis* cell wall proteome, plasma membrane proteome and chloroplast proteome. Those with matched AGI number to any one of the proteome were noted. The proteins were checked again based on their function and from literature before being categorised under cell wall protein and contaminants from cell wall, plasma membrane, chloroplast and plastid. In total 25 proteins were found under this category. Interestingly most of the proteins were detected in samples where ultracentrifugation was not applied.

Finally, some ribosomal proteins (21 proteins) were also identified. Although it is thought that ribosomal proteins are not present in the phloem, some of these proteins were also found in the phloem proteome of brassica, cucurbit and rice. Findings by Zhang et al (2009) and Kragler (2010) are in agreement whereby the phloem exudates used for each study did not contain any translational activity regardless of the condition they were exposed to. Based on similar study using phloem exudates, very limited or no occurrence of protein synthesis was detected in the sieve tubes (Kragler, 2012). Similarly, all but one of these proteins was not detected in samples from the current study where ultracentrifugation was applied. When both proteins from the cell wall and contaminants and, ribosomal protein categories were excluded, the total number of proteins believed be *Arabidopsis* phloem exudates protein is finalised at 175 (Table 3.2).

3.3.4 Identification of Unique Proteins

Based on Figure 3.5 (d), there are many phloem exudate proteins that are unique to thecurrent study: 105 proteins from a total of 221 identified proteins. Of these, 7 proteins

were confirmed cell wall protein, plasma membrane, chloroplast and plastid. Furthermore, 21 proteins were confirmed ribosomal proteins. Therefore, in total the number of unique novel identified proteins from phloem exudate of the current study is 77. The list of these proteins is shown in Table 3.3. The proteins still have diverse functions with the majority involved in metabolism. Four of the proteins shared homologies to brassica, cucurbit and rice, increasing the likelihood of being phloem proteins common across species.

Table 3.3: List of identified proteins unique to the current study. Each protein is assigned under its functional class according to MapMan categories on TAIR. As previously, identified proteins were compared against brassica, cucurbit and rice phloem proteome for homology. Proteins with shared homology in all four species are written in blue.

	AGI	ID	Mass obser-ved (Da)	No. of sequence (ESI-MS)	Homology (other species)
	Redox Reaction				
1	At4g11600	Glutathione peroxidase 6 (GPX6)	18597	2	C, R
2	At2g24940	Membrane-associated progesterone binding protein 2 (MAPR2)	11002	4	-
3	At3g02730	Thioredoxin F-type 1 (TRXF1)	19313	2	C
4	At5g16400	Thioredoxin F2 (TRXF2)	19966	1	-
5	At3g15360	Thioredoxin M-type 4 (TRX-M4)	21403	2	-
6	At1g43560	Thioredoxin Y2 (Aty2_ty2)	18581	1	-
7	At2g20270	Thioredoxin superfamily protein	19125	2	-
8	At4g03520	Thioredoxin superfamily protein (ATHM2)	20314	5	-
	Metabolic Pathway **A) Glycolysis**				
9	At5g42740	Glucose-6-phosphate isomerase, cytosolic (GPI) (phosphoglucose isomerase) (PGI)	61589	1	-
	B) TCA				
10	At5g14740	Carbonic anhydrase 2 (CA2)	36462	5	-
	Amino Acid Metabolism				
11	At1g02500	S-adenosylmethionine synthetase 1 (SAM1)	41453	1	B, C
	Lipid Metabolism				

12	At1g31812	Acyl-CoA-binding protein 6 (ACBP6)	10466	3	C, R
13	At4g25050	Acyl carrier protein 4 (ACP4)	14536	2	-
14	At5g16390	Biotin carboxyl carrier protein of acetyl-CoA carboxylase precursor (BCCP)	29613	2	-
15	At1g27950	Glycosylphosphatidylinositol-anchored lipid protein transfer 1 (LTPG1)	19777	3	-
16	At2g38530	Lipid transfer protein 2 (LTP2)	11938	3	-
	Nucleotide Metabolism				
17	At3g09820	Adenosine kinase 1 (ADK1)	37812	7	-
18	**At5g03300**	**Adenosine kinase 2 (ADK2)**	**31337**	**7**	**B,C, R**
	Secondary Metabolism				
19	At1g54040	Epithiospecifier protein (ESP)	37069	3	B
20	At3g44310	Nitrilase 1 (NIT1)	38152	1	R
	Metabolism – Miscellaneous				
	A) C1 Metabolism				
21	At3g59970	Methylenetetrahydrofolate reductase 1 (MTHFR1)	66246	2	-
	B) Hormone Metabolism				
22	At3g16430	Jacalin-related lectin 31 (JAL31)	32133	2	-
	C) Major CHO Metabolism				
23	At1g12240	Glycosyl hydrolases family 32 protein / Beta-fructosidase (ATBETAFRUCT4)	73465	3	-
	D) Minor CHO Metabolism				
24	At5g61130	Plasmodesmata callose-binding protein 1 (PDCB1)	20365	2	R
	E) Nitrogen Metabolism				
25	At5g35630	Glutamine synthetase 2(GS2)	47256	6	B
	F) Vitamin Metabolism				
26	At4g26500	Chloroplast sulfur E (SUFE1)	40888	1	C, R
	Photosynthesis **A) Light Reaction**				
27	At1g77090	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein	28501	1	-
28	At4g21280	Photosystem II subunit QA (PSBQ-1)	23888	4	-
29	At4g05180	Photosystem II subunit Q-2 (PSBQ-2)	24611	10	-
30	At1g76100	Plastocyanin 1 (PETE1)	17589	1	-
	B) Photorespiration				
31	At1g32470	Single hybrid motif superfamily protein / Glycine cleavage system H protein	17886	2	-
32	At3g14420	Aldolase-type TIM barrel family protein	40226	1	-
	Stress **A) Abiotic Stress**				
33	At1g20440	Cold-regulated 47 (COR477)	33349	6	-
34	At1g20450	Dehydrin family protein (ERD10)	29425	5	-
35	At1g76180	Dehydrin family protein (ERD14)	20770	6	-
36	At1g70830	MLP-like protein 28 (MLP28)	27528	3	-

	B) Biotic Stress				
37	At2g43530	Scorpion toxin-like knottin superfamily protein	9858	2	-
	Cell Organization and Biogenesis				
38	**At5g09810**	**Actin 7 (ACT7)**	**41774**	**5**	**B,C,R**
39	At5g56600	Profilin 3 (PRO3)	14333	1	C, R
40	At3g63190	Ribosome recycling factor, chloroplast precursor (RRF)	30403	2	-
41	At5g15970	Stress-responsive protein (KIN2) / stress-induced protein (KIN2) / cold-responsive protein (COR6.6) / cold-regulated protein (COR6.6) (AtCor6.6_COR6.6_KIN2)	6551	6	-
42	At5g62690	Tubulin beta-2/beta-3 chain (TUB2)	50003	1	-
43	At4g14960	Tubulin/FtsZ family protein (TUA6)	49226	1	C
	Signalling				
44	At2g27030	Calmodulin 5 (CAM5)	16652	1	C, R
45	At3g43810	Calmodulin-7 (CAM7)	16829	1	-
46	At4g20260	Plasma-membrane associated cation-binding protein 1 (PCAP1)	24697	3	-
47	At3g59920	RAB GDP dissociation inhibitor 2 (GDI2)	46555	2	C, R
	Protein Biosynthesis, Organization and Modification				
48	At4g04910	AAA-type ATPase family protein (NSF)	81652	1	C, R
49	At5g20720	Chaperonin 20 (CPN20)	9021	4	-
50	At5g03940	Chloroplast signal recognition particle 54 kDa subunit (CPSRP54P)	60231	1	-
51	At5g50920	CLPC homologue 1 (CLPC1)	7883	1	-
52	At1g72730	DEA(D/H)-box RNA helicase family protein	46777	1	R
53	At5g45680	FK506-binding protein 13 (FKBP13)	22039	1	-
54	**At5g02500**	**Heat shock cognate protein 70-1 (HSC70-11)**	**71358**	**2**	**B,C, R**
55	At3g12390	Nascent polypeptide-associated complex (NAC), alpha subunit family protein	21982	1	C, R
56	At4g20360	RAB GTPase homolog E1B (ATRABE1B)	51398	2	-
57	At2g47170	Ras-related small GTP-binding family protein (ARF1A1C)	21646	1	R
58	**At5g58290**	**Regulatory particle triple-A ATPase 3 (RPT3)**	**45711**	**1**	**B[a] ,C, R**
59	**At1g31340**	**Related to ubiquitin 1 (RUB1)**	**8525**	**1**	**B,C, R**
60	At5g03240	Polyubiquitin (UBQ3)	28169	2	-
	Transport				
61	At2g01320	ABC-2 type transporter family protein （ABCG7)	78599	1	-
62	At1g72150	Patellin 1 (PATL1)	64046	1	-
	Electron Transport				
63	At1g22840	Cytochrome C-1 (CYTC-1)	12390	3	-
	DNA				
64	At5g38470	Rad23 UV excision repair protein family (RAD23D)	40122	1	C

	RNA				
65	At4g34110	Polyadenylate-binding protein 2 (PABP2)	67017	1	-
66	At1g49760	Poly(A) binding protein 8 (PABP8)	72733	3	C
67	At2g23380	SET domain-containing protein (CLF / SET1)	84155	1	-
68	At3g61260	Remorin family protein	23130	1	-
	Miscellaneous				
69	At3g16410	Nitrile specifier protein 4 (NSP4)	67555	2	-
70	At3g10720	Plant invertase/pectin methylesterase inhibitor superfamily	60013	6	-
	Unknown				
71	At3g05900	Neurofilament protein-related	72087	1	-
72	At1g55480	Protein containing PDZ domain, a K-box domain, and a TPR (ZKT)region	37387	1	-
73	At5g01750	Protein of unknown function (DUF567)	24295	2	-
74	At3g02910	AIG2-like (avirulence induced gene) family protein	21333	1	R
75	At5g56170	LORELEI-LIKE-GPI-ANCHORED PROTEIN 1 (LLG1)	18904	1	-
76	At2g44920	Tetratricopeptide repeat (TPR)-like superfamily protein	23744	1	-
77	At5g53490	Tetratricopeptide repeat (TPR)-like superfamily protein	25628	1	-

NB:B=Brassica, C=Cucurbit, R=Rice, [a]=annotated

3.3.5 Functional Categories of Identified Proteins

In order to determine the functions of proteins identified, they were assigned to GO (Gene Ontology) and MapMan categories based on the BAR (Bio-Array Resource for Plant Biology) (http://bar.utoronto.ca). Using their AGI numbers, each protein was then categorised according to their gene function. GO is basically used to assign genes and proteins based on their biological processes, molecular functions and cellular components. On the other hand, there are various functional categories under MapMan. A program called Classification SuperViewer is used to assign each protein by determining protein frequencies under MapMan functional categories (Provart and Zhu, 2003).

In this study, the main focus is in the biological process as the ability to distinguish the function of the proteins is of high interest and would be an advantage. The roles depicted by cellular components and molecular functions are equally important and are reported along with biological processes.

It was evident from GO analysis in the biological process functional category (Fig. 3.6 (a)) that a high percentage of proteins are stress-related proteins with response to abiotic or biotic stimulus. The normalised frequencies of these proteins are also significantly above the value for the *Arabidopsis* reference set (Fig. 3.6 (b)). On the other hand, transcription is under-represented among the identified protein which is anticipated in the SE of mature angiosperm. Biological processes associated with each of the identified protein from Table 3.3 was assigned and represented in Appendix B.

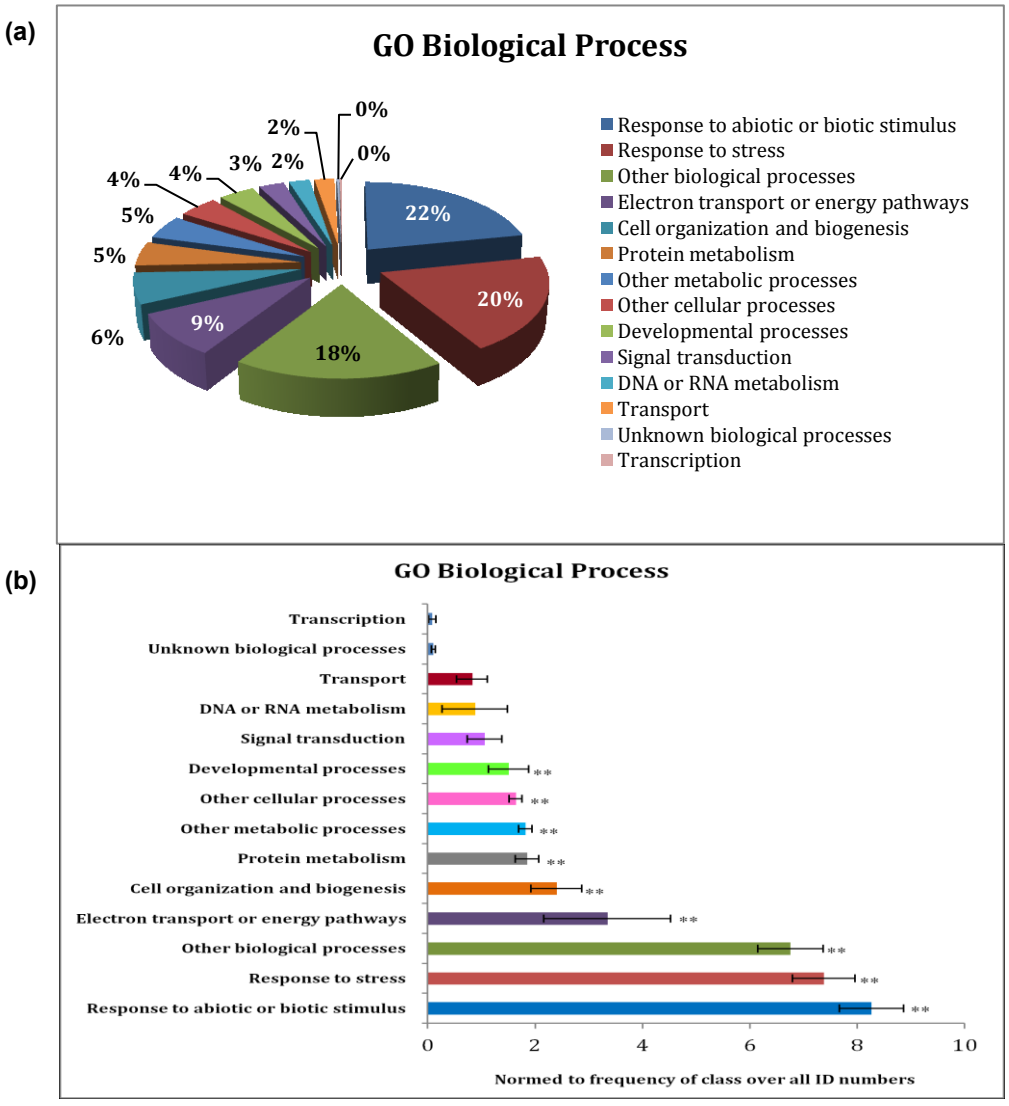


Figure 3.6: GO analysis of identified proteins from *Arabidopsis* specifying Biological Process. (a) Distribution of the differentially expressed genes across functional categories is represented in % of the normed frequencies for the proteins. (b) Normed frequencies (± *bootstrap StdDev*) for the proteins according to functional categories where ** indicated

frequencies significantly above the values for the *Arabidopsis* reference set ($p < 0.05$). Error bar showed ± bootstrap standard deviation calculate by BAR.

Analysis of GO on molecular function showed the proteins are largely involved in structural molecule activity (Fig. 3.7 (a)) and the normed frequency is significantly higher than the value for the *Arabidopsis* reference set (Fig. 3.7 (b)).

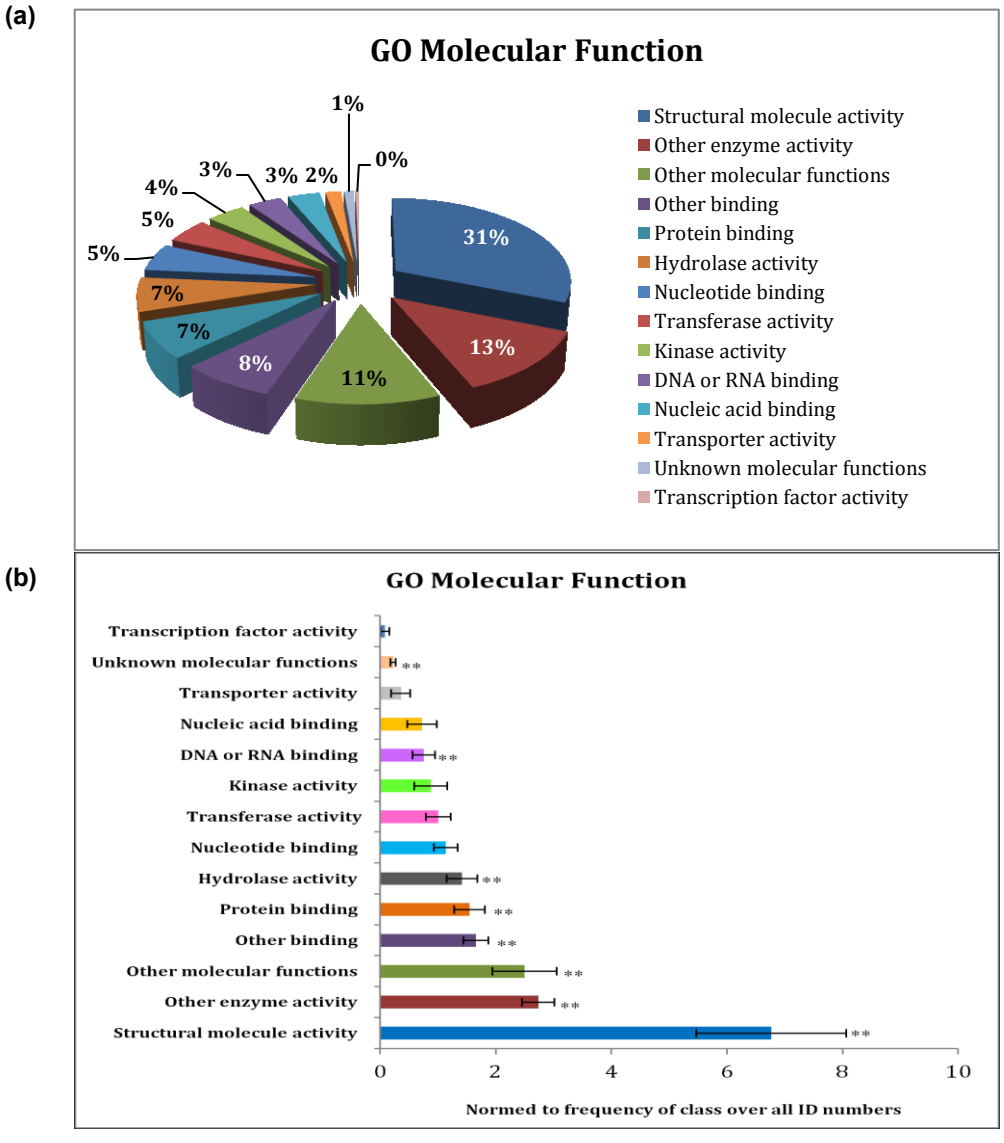
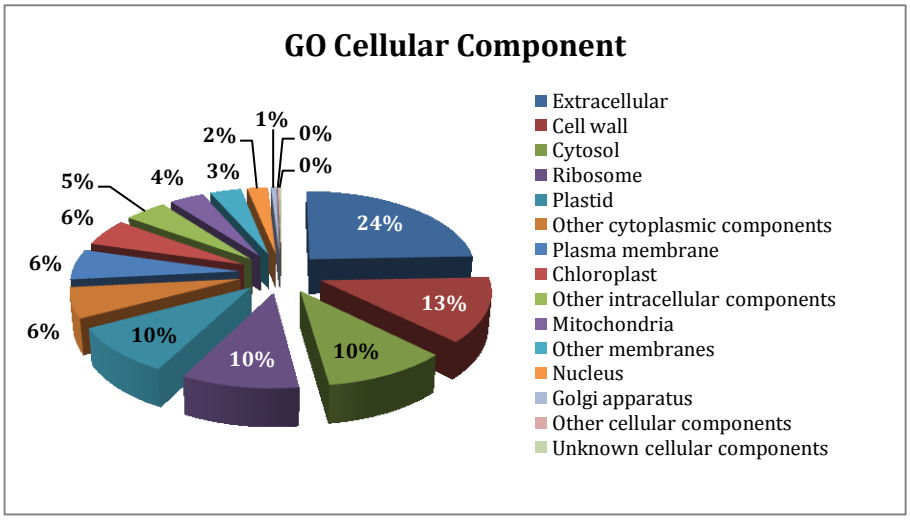


Figure 3.7: GO analysis of identified proteins from *Arabidopsis* specifying Molecular Function. (a) Distribution of the differentially expressed genes across functional categories is represented in % of the normed frequencies for the proteins. (b) Normed frequencies (± *bootstrap StdDev*) for the proteins according to functional categories where ** indicated frequencies significantly above the values for the *Arabidopsis* reference set ($p < 0.05$). Error bar showed ± bootstrap standard deviation calculate by BAR.

In the last functional category, the cellular component, the identified proteins were highly composed of extracellular components (Fig. 3.8 (a)) with the normed frequency significantly higher than the value for the *Arabidopsis* reference set (Fig. 3.8 (b)).

(a)



(b)

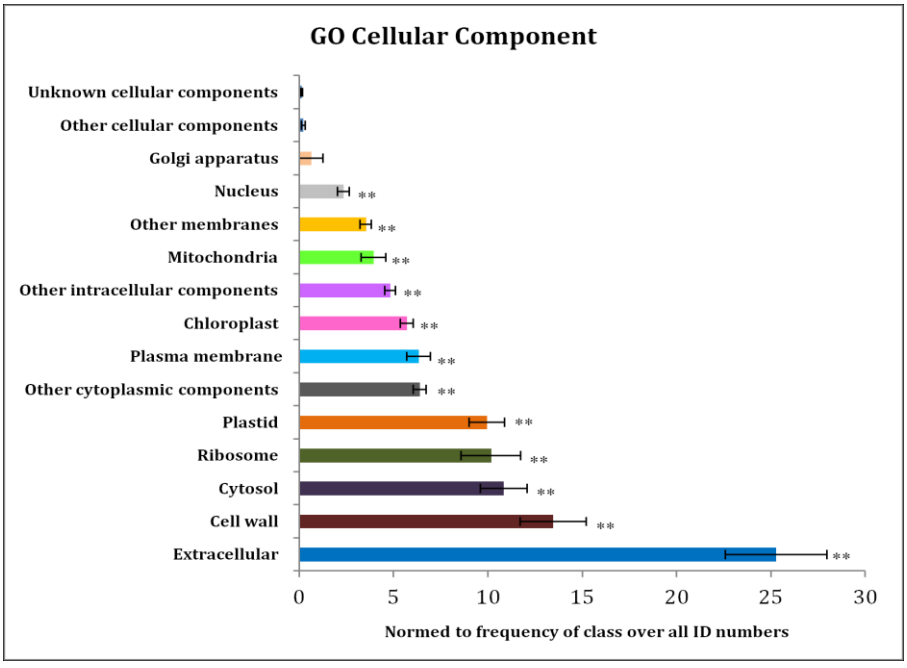


Figure 3.8: GO analysis of identified proteins from *Arabidopsis* specifying Cellular Component. (a) Distribution of the differentially expressed genes across functional categories is represented in % of the normed frequencies for the proteins. (b) Normed frequencies (± *bootstrap StdDev*) for the proteins according to functional categories where ** indicated frequencies significantly above the values for the *Arabidopsis* reference set (p <0.05). Error bar showed ± bootstrap standard deviation calculate by BAR.

A total of 125 GO terms that were enriched were identified from the identified proteins. A *p*-value cutoff of 0.05 were used in the analysis performed using AmiGO (www.geneontology.org). Characteristics of enrichment of certain GO terms were based on the contribution from other GO terms enriched at the lower hierarchy. A directed acyclic graph (DAG) was provided alongside the result of GO term enrichment (Appendix C-1 to C-5) to accommodate specifically enriched GO terms and how they affect other GO terms through upper hierarchies. For example in the aspect of biological process related to hydrogen peroxide catabolic process, this GO term was enriched at a rather low *p*-value cutoff (1.33e-04) (Appendix C-2). A large number of other GO terms at upper hierarchies were also enriched such as multi-organism process, immune system process, response to stimulus and cellular process (Appendix C-3 and C-4). In the aspect of molecular function ending with copper ion binding (*p*-value cutoff 5.41e-27), it was interesting to see that the rest of the GO terms were enriched but with lower probability (Appendix C-6). Finally, a mixture of *p*-value cutoff at the end of cellular component aspect was seen (Appendix C-5) and this functional category represented the most complex GO terms that are enriched.

According to MapMan classification of functional categories, a large number of the identified proteins from Table 3.2 are involved in metabolic processes. Figure 3.9 showed the normed frequencies of MapMan categories calculated using BAR Classification Superviewer. Some of the categories involved in metabolic processes (denoted by #) are significantly higher than the *Arabidopsis* reference set (p value $< 10^{-7}$). Proteins under the category of stress were found to have low normed frequency but are significant over the *Arabidopsis* reference set (p value < 0.03). Meanwhile, proteins involved in signalling have low normed frequency and is less significant over the *Arabidopsis* reference set.

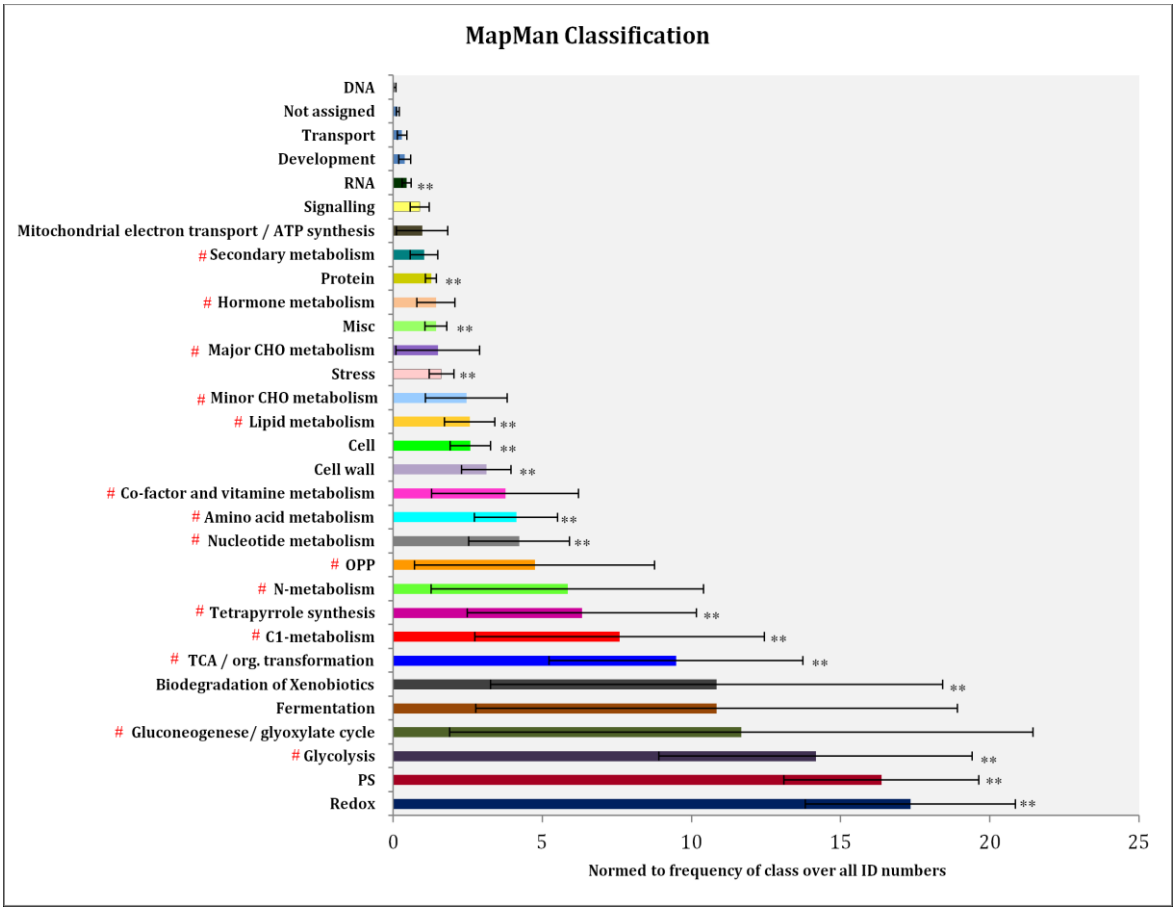


Figure 3.9: Normed frequencies of functional categories from MapMan classification for the identified proteins. ** indicated frequencies significantly above the values for the *Arabidopsis* reference set (p <0.05). # indicated categories involved in metabolic process. Error bar showed ± bootstrap standard deviation calculated by BAR. CHO=carbohydrate, OPP=oxidative pentose phosphate, PS=photosynthesis, TCA=tricarboxylic acid.

3.3.6 Validity of Proteins

In order to ensure the proteins obtained from the exudates come from phloem sap, the same exudation technique was employed on *Brassica napus*. This plant was chosen as Brassica's phloem proteome has been established and published (Giavalisco et al., 2006). Based on SDS-PAGE where the EDTA exudate sample was run alongside a pure phloem sap collected by needle wounding (Fig 3.10), it can be seen that the protein pattern did not differ between the two.

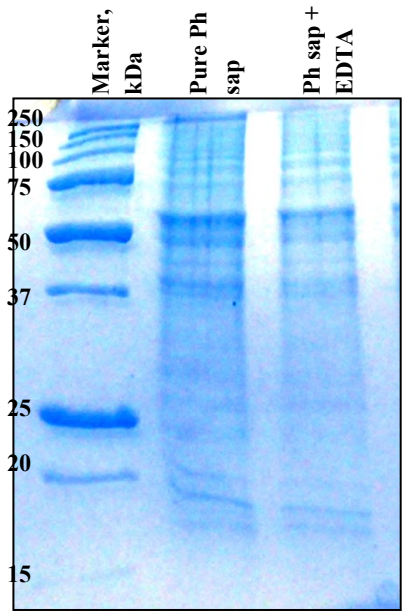


Figure 3.10: Comparison of phloem sap proteins from *Brassica napus* on SDS-PAGE between pure phloem sap obtained using needle wounding method and EDTA exudation technique. 5 µg of proteins were run on 15% SDS-PAGE and stained using Coomassie blue.

To further validate the proteins obtained via EDTA exudation technique, the sample was processed similarly with *Arabidopsis* and sent for LC-MS/MS. It was noted during the initial experiment that overnight incubation in EDTA causes the leaf to wilt and turned yellow the following day. Because of this, incubation time was reduced to a maximum of 5 hours. From analysis of spectra obtained, a relatively low number of proteins were identified, but 21% of those obtained were present in the published *Brassica* phloem proteome database. The proteins identified are presented in Table 3.4. When compared against protein list from *Arabidopsis* phloem proteome, a high percentage of homology (75%) was achieved.

Table 3.4: A list of proteins identified via LC-MS/MS from *Brassica napus* exudates collected using EDTA exudation technique.

	Protein	AGI no.	Found in *B.na* phloem proteome database	Found in *A.th* phloem proteome
1	Actin-7, ACT7	At5g09810	Yes	Yes
2	Adenosine kinase 1, AK1	At3g09820	No	Yes
3	Adenosylhomocysteinase 2, SAHH2	At3g23810	No	Yes
4	ADP-ribosylation factor 2, ARF2-A and ARF2-B	At1g23490 At1g70490	No	No
5	S-adenosylmethionine synthetase 1, SAM1	At1g02500	No	Yes
6	Cobalamin-independent methionine synthase isozyme, CIMS	At5g17920	Yes	Yes
7	Elongation factor 1-alpha (eEF-1A)	At1g07920	Yes	Yes
8	Elongation factor Tu, TUFA	At4g20360	No	Yes
9	Enolase (2-phosphoglycerate dehydratase)	At2g36530	Yes	Yes
10	Eukaryotic initiation factor 4A-3, TIF4A-3	At1g72730	No	Yes
11	Glyceraldehyde-3-phosphate dehydrogenase A, chloroplast precursor, GAPA	At3g26650	No	No
12	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic, GAPC	At3g04120	Yes	Yes
13	Glycine-rich RNA-binding protein 7, GRP7	At2g21660	Yes	Yes
14	Glycine-rich RNA-binding protein 8, GRP8	At4g39260	No	Yes
15	Malate dehydrogenase 1, mitochondrial precursor, mNAD-MDH 1	At1g53240	No	Yes
16	Myrosinase precursor (Sinigrinase) (Thioglucosidase)	At5g26000	No	Yes
17	Nucleoside diphosphate kinase 1, NDPK1	At4g09320	Yes	Yes
18	Oxygen-evolving enhancer protein 3-2, PSBQ2	At4g05180	No	Yes
19	Phosphoglycerate kinase, chloroplast precursor, PGK	At1g56190	No	Yes
20	Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial precursor, GPX6	At4g11600	No	Yes
21	Profilin-2, PRO2	At4g29350	Yes	Yes
22	Putative gamma-glutamylcyclotransferase	At3g02910	No	Yes
23	RubisCO large subunit, rbcL	AtCg00490	Yes	Yes
24	SufE-like protein, EMB1374	At4g26500	No	Yes
25	Translationally-controlled tumor protein homolog, TCTP	At3g16640	Yes	Yes
26	Tubulin beta-1 chain, TUBB1	At1g75780	No	No
27	Tubulin beta-2/beta-3 chain, TUBB2 and TUBB3	At5g62690 At5g62700	No	Yes

28	Tubulin beta-5 chain, TUBB5	At1g20010	No	No
29	Tubulin alpha-6 chain, TUBA6	At4g14960	No	Yes
30	Tubulin beta-9 chain, TUBB9	At4g20890	No	No
31	2-cys peroxiredoxin BAS1, chloroplast precursor	At3g11630	No	Yes
32	14-3-3-like protein GF14 nu (General regulatory factor 7), GRF7	At3g02520	No	Yes
33	26S protease regulatory subunit 6A homolog B, RPT5B	At1g09100	No	No
34	26S protease regulatory subunit 6B homolog (Regulatory particle triple-A ATPase subunit 3), RPT3	At5g58290	No	Yes
35	20 kDa chaperonin, chloroplast precursor (Protein Cpn21), CPN21	At5g20720	No	Yes
36	40S ribosomal protein S2-4, RPS2D	At3g57490	No	No
37	40S ribosomal protein S3-1, RPS3A	At2g31610	Yes	No
38	40S ribosomal protein S4-3, RPS4D	At5g58420	No	No
39	40S ribosomal protein S5-1, RPS5A	At2g37270	No	Yes
40	40S ribosomal protein S6-2, RPS6B	At5g10360	No	Yes
41	40S ribosomal protein S8-1, RPS8A	At5g20290	No	Yes
42	40S ribosomal protein S13-2, RPS13B	At4g00100	No	Yes
43	40S ribosomal protein S14-3, RPS14C	At3g52580	No	No
44	40S ribosomal protein S18, RPS18A, S18B, or, and S18C	At1g22780	No	Yes
45	60S acidic ribosomal protein P0-2, RPP0B	At3g09200	No	Yes
46	60S ribosomal protein L5-1, ATL5A	At3g25520	No	Yes
47	60S ribosomal protein L6-3, RPL6C	At1g74050	No	Yes
48	60S ribosomal protein L7a-2	At3g62870	No	Yes
49	60S ribosomal protein L7-3, RPL7C	At2g44120	No	Yes
50	60S ribosomal protein L8-3, RPL8C	AT4g36130	No	No
51	60S ribosomal protein L11-1, RPL11A	At2g42740	No	No
53	60S ribosomal protein L12-2, RPL12B	At3g53430	No	No
53	60S ribosomal protein L23a-2, RPL23AB	At3g55280	No	Yes

3.3.7 Detection of FT via Multiple Reaction Monitoring (MRM)

Even though FT protein was not found among the proteins identified using LC-MS/MS analysis, it is still believed to be present in the phloem of *Arabidopsis* based on previous study by grafting (Corbesier et al., 2007). It was later proven by Batailler et al (2012) where FT was detected in the phloem exudate of *Arabidopsis*. With that in mind, a higher sensitivity multiple reaction monitoring (MRM) method for detection of FT was developed using His-FT as a standard. Courtesy of David Charles (member of Turnbull lab), FT sequence was cloned into *E. coli*, grown in bulk after which the cells were

harvested and the protein purified. Based on SDS-PAGE gel, the 20 kDa band was cut out, subjected to tryptic digest and later used as a standard for MRM. Ten peptide sequences were obtained from His-FT and these were selected as external standard peptides for FT detection. The peptide sequences are QTYVAPGWR, VEIGGEDLR and VVGDVLDPFNR and the result is shown in Figure 3.11.

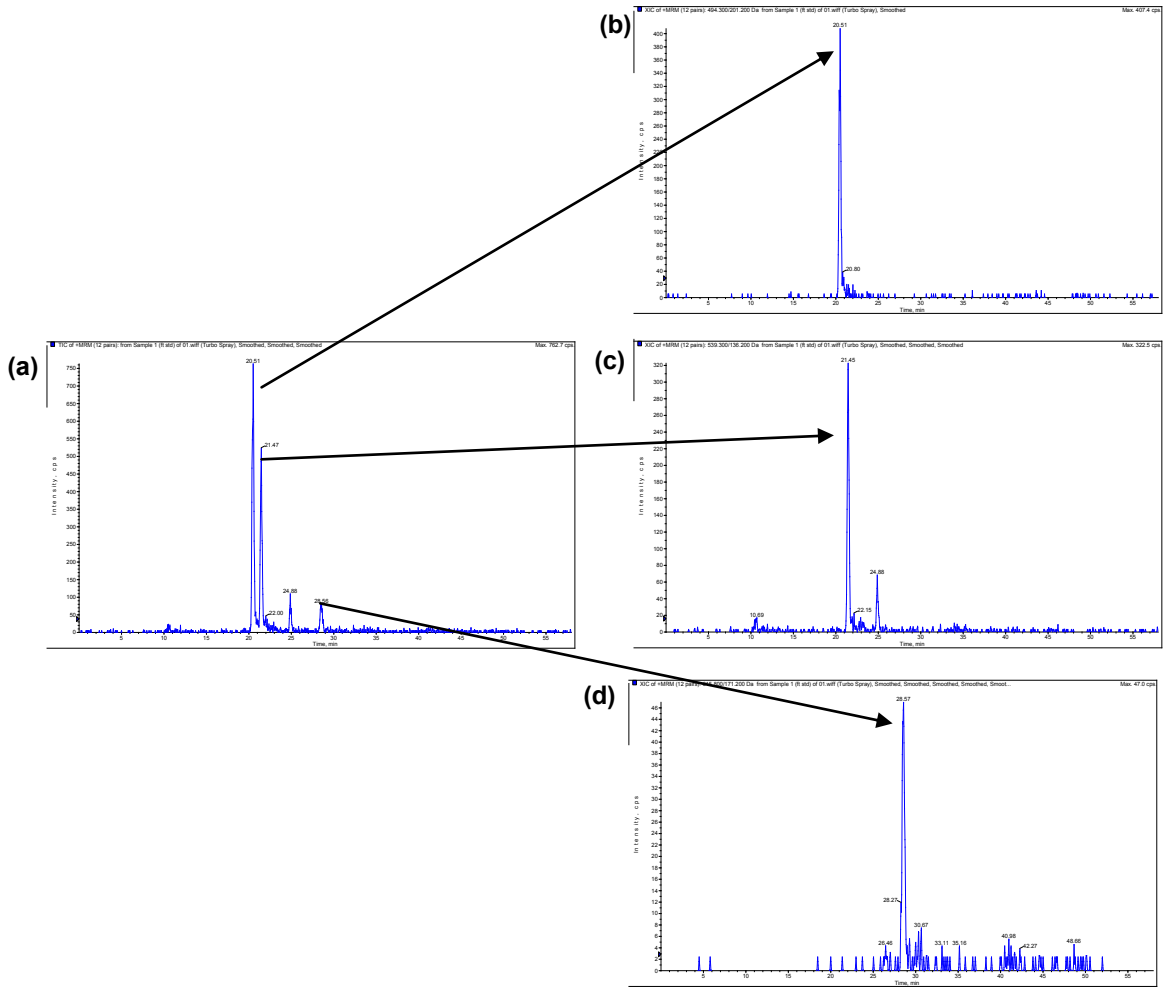


Figure 3.11: MRM was used to detect FT in the standard, His-FT using three chosen peptides where (a) all peptides were detected in the sample. Detection of individual peptide is also shown; (b) VIEGGEDLR, (c) VVGDVLDPFNR and (d) QTYVAPGWR.

Using His-FT as standard, the three peptides were used to detect FT in the samples. Criteria to validate detection of FT were determined where the peak area count should be more than 200 and at least two of the peptides met the first criteria. Results from MRM are as in Table 3.5. Two positive controls were used, 35S::FT:GFP which should generally

give positive FT signal throughout the whole plant tissues and SUC2::FT:GFP in which FT expression will only be in the phloem and major leaf veins. As expected, FT was found in abundance in SUC2::FT:GFP from phloem samples (peak area count of 1947). However, lower peak area count was obtained for 35S::FT:GFP (1080). As a negative control *ft-10* null mutant was used, and showed that FT was not detected. Likewise, FT was not detected in exudates from SD plants but FT was detected in WT plants exposed to LD for 3 days (peak area count 2077). The detectable FT peak for the positive control and LD sample including the absence of FT from SD sample and *ft-10* mutant is shown in Figure 3.12.

Table 3.5: Result of MRM carried out on the standard and other samples while setting up the procedure to detect FT.

Sample	Analyte peak name	Analyte peak area (counts)	Retention time (min)	Analyte expected retention time (min)
Standard	QTYVAPGWR3	4913	21.44	21.44
	VEIGGEDLR3	5897	20.52	20.52
	VVGDVLDPFNR3	1005	28.53	28.53
SUC2::FT:GFP	QTYVAPGWR3	526.3	21.47	21.44
	VEIGGEDLR3	1947	20.45	20.52
	VVGDVLDPFNR3	1916	28.53	28.53
35S::FT:GFP	QTYVAPGWR3	428.5	21.5	21.44
	VEIGGEDLR3	1080	20.44	20.52
	VVGDVLDPFNR3	1027	28.54	28.53
LD	QTYVAPGWR3	729.5	21.49	21.44
	VEIGGEDLR3	2077	20.48	20.52
	VVGDVLDPFNR3	1851	28.54	28.53
SD	QTYVAPGWR3	27.63	21.33	21.44
	VEIGGEDLR3	40.17	20.49	20.52
	VVGDVLDPFNR3	65.02	28.56	28.53
ft-10	QTYVAPGWR3	20.91	21.37	21.44
	VEIGGEDLR3	27	20.45	20.52
	VVGDVLDPFNR3	13.79	28.51	28.53

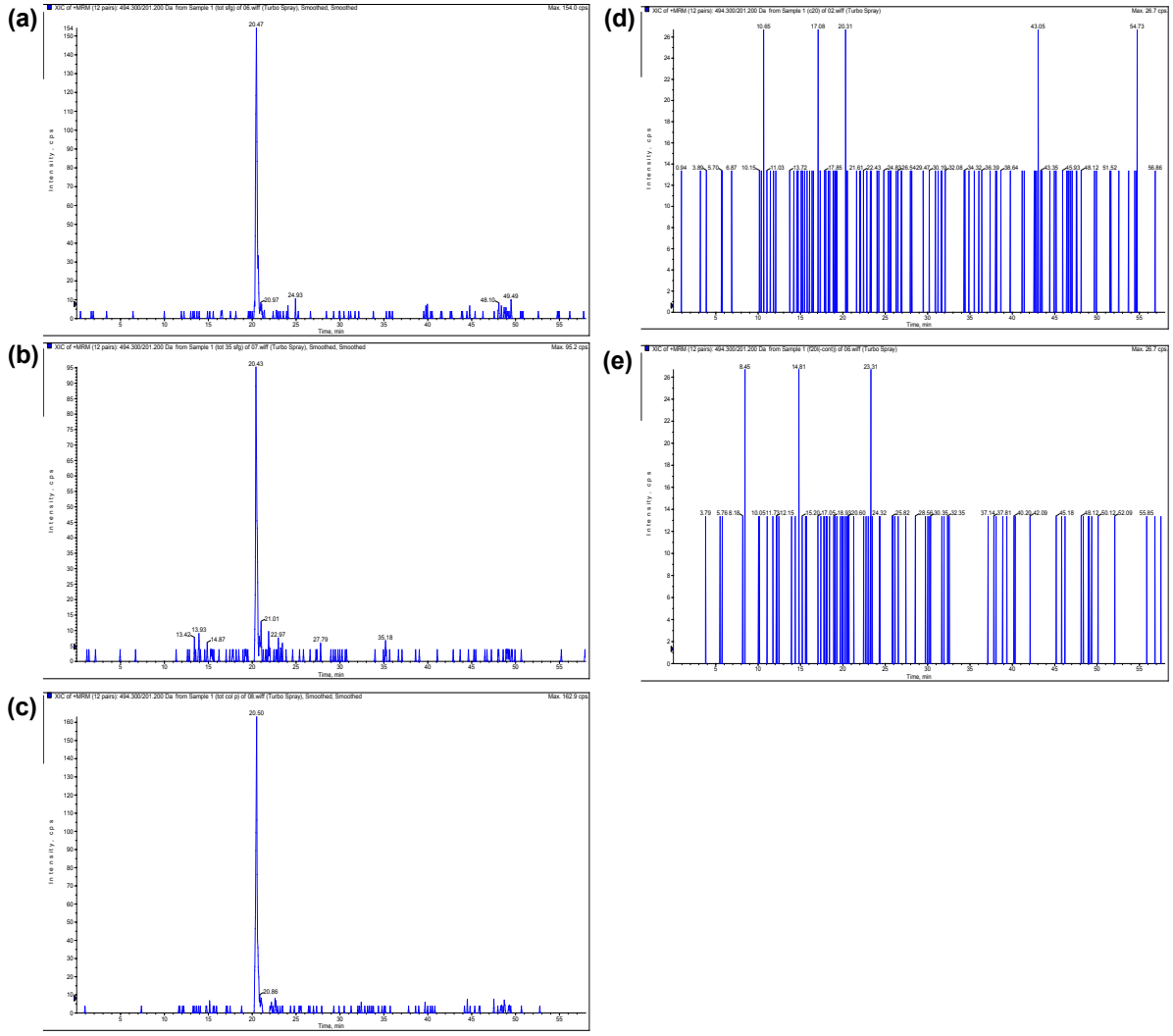


Figure 3.12: MS spectra of FT after MRM analysis using the peptide VEIGGEDLR. (a) SUC2::FT:GFP as positive control for phloem protein, (b) 35S::FT:GFP as general positive control, (c) phloem sap protein after 3 days exposure to LD, (d) WT grown only in SD and (e) null *ft-10*.

After the MRM method had been set up for FT, further samples were analysed from various treatments as in Table 3.6. Among the samples sent were those collected under different incubation periods during exudation optimization (3 hours and overnight), those exposed to ultracentrifugation and, exudates collected after the first 30 min followed by second exudates collected overnight using the same leaves and the pellet obtained after ultracentrifugation. However as the criteria for the presence of FT had been set, it was found that none of the samples met the requirement of having peak count of more than 200 at two peptides.

Table 3.6: Summary of MRM carried out on various treatments for detection of FT.

Sample	Analyte peak name	Analyte peak area (counts)	Retention time (min)	Analyte expected retention time (min)
Standard	QTYVAPGWR3	27560	19.51	19.51
	VEIGGEDLR3	267300	18.44	18.44
	VVGDVLDPFNR3	371700	27.04	27.04
3h incubation	QTYVAPGWR3	149	19.4	19.51
	VEIGGEDLR3	13.79	18.65	18.44
	VVGDVLDPFNR3	39.9	26.82	27.04
Overnight incubation	QTYVAPGWR3	1954	19.72	19.51
	VIEGGEDLR3	39.9	18.67	18.44
	VVGDVLDPFNR3	65.38	26.69	27.04
Collected after 30min	QTYVAPGWR3	3551	19.7	19.51
	VEIGGEDLR3	47.38	18.13	18.44
	VVGDVLDPFNR3	36.08	27.14	27.04
Collected after overnight	QTYVAPGWR3	470.1	19.75	19.51
	VIEGGEDLR3	39.41	18.19	18.44
	VVGDVLDPFNR3	14.33	26.94	27.04
Pellet	QTYVAPGWR3	122.1	19.69	19.51
	VEIGGEDLR3	26.6	18.55	18.44
	VVGDVLDPFNR3	33	27.35	27.04
Ultracentrifuged Sample	QTYVAPGWR3	140.5	19.3	19.51
	VEIGGEDLR3	13.79	18.46	18.44
	VVGDVLDPFNR3	26.6	26.92	27.04
Non-ultra-centrifuged sample	QTYVAPGWR3	59.56	19.61	19.51
	VEIGGEDLR3	45.54	18.85	18.44
	VVGDVLDPFNR3	81.42	27.15	27.04

3.3.8 Level of FT in Plant under Photoperiod Treatment

Based on FT mRNA expression results reported by Corbesier et al (2007), an attempt at FT quantification was constructed at approximately the same timeline. Six week old plants grown in SD were exposed up to 4 days in LD prior to exudation and the first exudate was collected 60 hours after exposure to LD. A 4 hour interval in between collection of exudates was carried out up until 104 hours where the last exudates was collected giving a total of 12 samples. LD light treatment began at 8 am and ended at 11:59 pm for 16 hours. The first collection of exudates was carried out at 12 pm and ended at 8 am two days later. Total phloem sap protein was used and after tryptic digestion, FT level in the samples were quantified via MRM using His-FT as standard and the three peptide sequences as in 3.3.5 with the same parameter set for detection of FT. Quantitative measurement of each sample was assessed through the ratio of transition ions derived from the endogenous peptides.

Summary of FT peptides detected in the 12 samples are represented in Table 3.7. Among the 12 samples, only one which was the third exudate collected after 68h did not meet the requirement where the peak area count on the two dominant peptides (VVGDVDLDPFNR and VEIGGEDLR) should be more than 200. FT was detected with various peak area counts across the other 11 samples where the highest peak area count was not restricted to any one single peptide. Identification of the VVGDVLDPFNR peptide was observed from five earlier samples giving the highest peak area count for each sample. The highest peak area count for the next six samples was equally identified using VEIGGEDLR and QTVYAPGWR peptides. It was observed that the highest peak area count was in the first sample (1391 in VVGDVLDPFNR) while the second highest peak area count was in the eighth sample at 88hour (1287 in VEIGGEDLR) (Figure 3.13).

Table 3.7: Summary of quantified FT in the standard (His-FT) and the 12 samples based on the main peptide VVGDVLPFNR. Retention times and peak area counts are noted.

Sample (actual collection time and the hours from the first exposure to LD)	Analyte peak name	Analyte peak area (counts)	Retention time (min)	Analyte expected retention time (min)
His-FT (standard)	QTYVAPGWR3	27560	19.51	19.51
	VEIGGEDLR3	267300	18.44	18.44
	VVGDVLDPFNR3	371700	27.04	27.04
W2-1 (1200, 60h)	QTYVAPGWR3	138	19.63	19.51
	VEIGGEDLR3	278.3	18.42	18.44
	VVGDVLDPFNR3	1391	26.8	27.04
W2-2 (1600, 64h)	QTYVAPGWR3	263.6	19.21	19.51
	VEIGGEDLR3	144.1	18.45	18.44
	VVGDVLDPFNR3	319.2	27.12	27.04
W2-3 (2000, 68h)	QTYVAPGWR3	142.9	19.66	19.51
	VEIGGEDLR3	193	18.52	18.44
	VVGDVLDPFNR3	661.6	27.24	27.04
W2-4 (0000, 72h)	QTYVAPGWR3	142.2	19.16	19.51
	VEIGGEDLR3	262.9	18.48	18.44
	VVGDVLDPFNR3	389.5	27.2	27.04
W2-5 (0400, 76h)	QTYVAPGWR3	63	19.6	19.51
	VEIGGEDLR3	210.7	18.51	18.44
	VVGDVLDPFNR3	340	28.84	27.04
W2-6 (0800, 80h)	QTYVAPGWR	113.6	19.58	19.51
	VEIGGEDLR3	596.6	18.56	18.44
	VVGDVLDPFNR3	636.9	27.27	27.04
W3-1 (1200, 84h)	QTYVAPGWR3	396.4	19.44	19.51
	VEIGGEDLR3	749.7	18.61	18.44
	VVGDVLDPFNR3	687.2	27.3	27.04
W3-2 (1600, 88h)	QTYVAPGWR3	158.4	19.4	19.51
	VEIGGEDLR3	1287	18.6	18.44
	VVGDVLDPFNR3	236.6	26.96	27.04
W3-3 (2000, 92h)	QTYVAPGWR3	870.5	19.44	19.51
	VEIGGEDLR3	692.2	18.69	18.44
	VVGDVLDPFNR3	282.6	27.4	27.04
W3-4 (0000, 96h)	QTYVAPGWR3	64.22	19.67	19.51
	VEIGGEDLR3	437.3	18.73	18.44
	VVGDVLDPFNR3	203.4	26.82	27.04
W3-5 (0400, 100h)	QTYVAPGWR3	733.5	19.54	19.51
	VEIGGEDLR3	636	18.8	18.44
	VVGDVLDPFNR3	275.8	27.02	27.04
W3-6 (0800, 104h)	QTYVAPGWR3	897.7	19.7	19.51
	VEIGGEDLR3	214.6	18.19	18.44
	VVGDVLDPFNR3	447.6	26.66	27.04

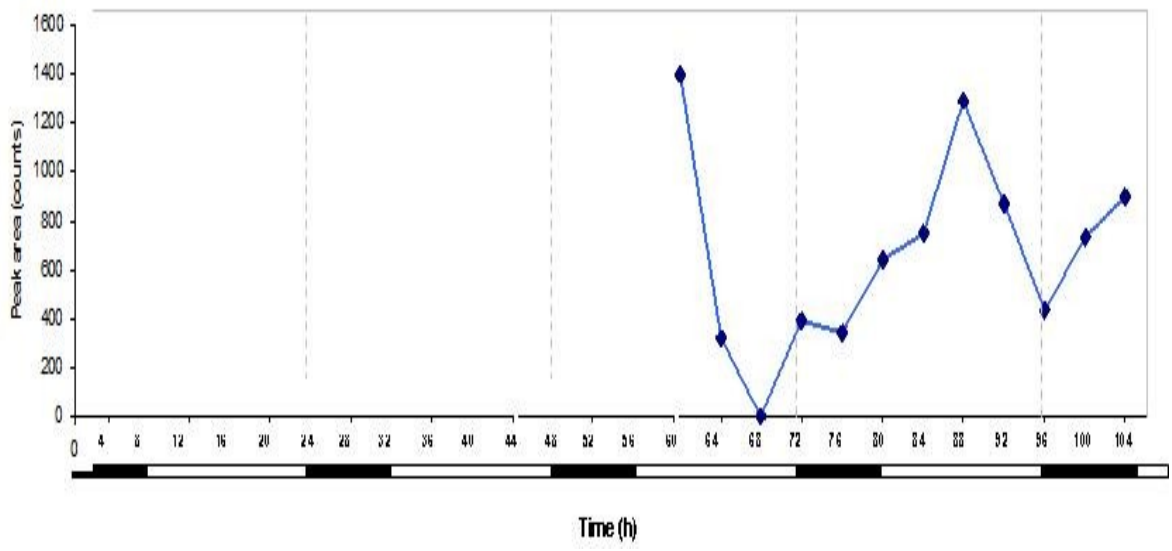


Figure 3.13: Quantification of FT protein via MRM was conducted under photoperiod induction based on the highest peak area count in any of the three peptides. Samples for quantification of FT was collected every 4 hours starting after two days of exposure to LD. Bars below the graph represented the duration of day (white) and night (black) and correspond to the time of collection.

3.4 Discussion

3.4.1 Data Analysis and Identification of Proteins from Phloem Exudate

Prior to protein identification, tryptic digestion is carried out followed by analysis using LC-MS/MS. Separation of cut peptides in LC-MS/MS is done through either reverse phase micro- or nano-capillary chromatography followed by ionization of eluted peptides. There are generally a lot of potential candidate ions to choose for subsequent MS/MS analysis, many of which are background ions associated to the LC eluent (Berg et al., 2006). To select the ion of interest IDA (Information Dependent Acquisition) was applied to obtain more information in the current study. Using IDA, no foreknowledge is required and clean MS/MS spectra could be obtained (Decaestecker et al., 2004). Specific ions can be focused on for increased productivity with IDA allowing simultaneous collection of single MS/MS data to maximise the information gained in a single injection. Once introduced into the MS the peptides are measured based on their masses and intensities (Berg et al., 2006). Information on their sequences is obtained through analysis of signal intensities of peptide fragment ions derived from tandem mass spectra that can later on be reconstructed into peptide

sequences and searched against reference databases to identify the protein (Berg et al., 2006; Fenyo, 2000).

Protein identification is based on its peptide sequence. After protein digestion, analysis through LC-MS/MS will give spectra representing one or more of the peptides. Identification of LC-MS with tandem mass spectra (MS/MS) is considered more specific than single LC-MS as MS/MS selection involves multiple steps. Tandem MS begins with ionisation of sample by ESI resulting in separation of precursor ion which is later fragmented to separate product ion for detection. This will further allow structural information contained within the sequence of the peptide to be extracted (Mann et al., 2001). From there a search is carried out by matching the experimental spectrum against a calculated spectrum for all peptides in a database using proteomics software. Among the commonly used software for proteomics work is MASCOT which provides sequence information of a given spectrum. Identification is based on a given score of calculated sequence against the tandem mass spectra. Another software package called ProteinPilot gave more detailed description of the sequence information of a spectrum including the exact location of the sequence and the occurrence of any modifications. As with MASCOT, scores were also given but this time based on the coverage of peptide sequence in the spectrum by providing percentage probability.

Using these softwares to identify protein was simple due to its automation, however, there are limitations especially for low scoring proteins. Often manual inspection is required for verification. During digestion, the sizes of peptides depend on the amino acid content as trypsin cuts on arginine (Arg, R) and lysine (Lys, K). Some of these peptides are quite short depending on the location of arginine and lysine. This could sometimes lead to difficulty in identifying unique peptides. Similar problems were encountered in the present study where quite a reasonable amount of peptides were not able to be identified using MASCOT or protein blast (blastp) in NCBI. On occasions when the peptide could be identified using blastp, the score level was too low to be taken as reliable. ProteinPilot could still identify some of these short peptides and most of the time they will be noted in the low confidence level category.

During analysis of excised gel bands from 1D gels, it was found that each band mostly contained more than one protein (Table 3.1). It is highly probable for one band to contain more than one polypeptide, as similar or sometimes identical molecular masses can be shared by different polypeptides (Mann et al., 2001). The identified proteins from excised gel bands correspond to the expected range of the size as shown in Table 3.8. At least five of the proteins shared homology to the other three plant phloem proteome. Additionally, high number of identified phloem protein was also found by Batailler et al (77.1%). The opposite was observed when compared to phloem proteins identified by Guelette et al (27.1%). Nevertheless, 22.9 % (or 11 of the identified proteins) were unique to the current study.

Table 3.8: Identified proteins from excised gel band based on gel sections.

Represents all proteins identified by Batailler et al (2012) in the phloem exudate

Represents all proteins identified by Batailler et al (2012) in the full set of exudates

Proteins with shared homology in all four species (*Arabidopsis*, brassica, cucurbit and rice) are written in blue

∗ *Arabidopsis* phloem exudate proteins identified by Guelette et al (2012) via EDTA-facilitated method

Identified protein									
		Phloem protein				Cell wall protein and contaminants			
Gel piece / section	Size (kDa)	No. of identified protein	AGI	ID	Mass Observed (kDa)	No. of identified protein	AGI	ID	Mass Observed (kDa)
1	75 – 100	4	At4g35090	Catalase 2 (CAT)	57.308	1	At5g64570	Beta-D-xylosidase 4 (XYL4)	84.228
			At5g17920	Cobalamin-independent synthase family protein (ATCIMS)	84.267				
			At4g04910	AAA-type ATPase family protein (NSF)	81.652				
			At1g56070	Ribosomal protein S5/Elongation factor G/III/V family protein (LOS1)	92.356				
2	63 – 74	3	At5g11670	NADP-malic enzyme 2 (NADP-ME2)	64.333	1	At5g67360	Subtilase family protein (ARA12)	73.556
			At3g59970	Methylenetetrahydrofolate reductase 1 (MTHFR1)	66.246				
			At4g20830	FAD-binding Berberine family protein	63.336				

3	51 – 62.5	6	At1g63940	Monodehydroascorbate reductase 6 (MDAR6)	53.299	1	At1g76160	SKU5 similar 5 (sks5)	60.003
			At3g03250	UDP-glucose pyrophosphorylase 1 (UGP1) *	51.706				
			At5g17310	UDP-glucose pyrophosphorylase 2 (UGP2)	51.92				
			At5g26000	Thioglucoside glucohydrolase 1 (TGG1) *	61.133				
			AtCg00490	Ribulose-bisphosphate carboxylases (RBCL)	52.669				
			At3g10720	Plant invertase/pectin methylesterase inhibitor superfamily	60.013				
4	50	1	At1g78850	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain	49.05	0			
5	43.5 – 49	5	At3g52880	Monodehydroascorbate reductase 1 (MDAR1) *	46.5	1	At1g03220	Eukaryotic aspartyl protease family protein	45.792
			At5g03630	Pyridine nucleotide-disulphide oxidoreductase family protein (MDAR2)	46.433				
			At2g36530	Enolase (ENO2) / (LOS2) *	47.719				
			At1g56190	Phosphoglycerate kinase family protein	49.894				
			At3g59920	RAB GDP dissociation inhibitor 2 (GDI2)	46.555				
6	38 – 43.5	6	At4g13940	S-adenosyl-L-homocysteine hydrolase (SAHH1)	39.994	4	At4g12730	Fasciclin-like arabinogalactan-protein (FLA2)	43.465
			At4g38970	Fructose-bisphosphate aldolase 2 (FBA2) *	42.935		At1g29670	GDSL-like Lipase/Acylhydrolase superfamily protein	39.863
			At3g55800	Sedoheptulose-bisphosphatase (SBPASE) *	42.414		At1g09750	Eukaryotic aspartyl protease family protein	39.751
			At5g66190	Ferredoxin-NADP(+)-oxidoreductase 1 (FNR1)	40.332		At3g49120	Peroxidase CB (PRXCB)	38.927

			At1g20020	Ferredoxin-NADP(+)-oxidoreductase 2 (FNR2)	41.142				
			At1g67280	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein	39.174				
7	37	1	At3g09820	Adenosine kinase 1 (ADK11)	37.812	1	At4g16260	Glycosyl hydrolase superfamily protein	37.688
8	31 – 36	6	At1g53240	Lactate/malate dehydrogenase family protein (mMDH1)	35.782	1	At3g16420	PYK10-binding protein 1 (JAL30_PBP1)	32.096
			At4g14880	O-acetylserine (thiol) lyase (OAS-TL) isoform A1 (OASA1)	33.805				
			At3g16430	Jacalin-related lectin 31 (JAL31)	32.133				
			At5g37850	pfkB-like carbohydrate kinase family protein (SOS4)	34.232				
			At2g21330	Fructose-bisphosphate aldolase 1(FBA1)	33.302				
			At1g35720	Annexin 1 (ANNAT1) *	35.757				
9	28.5 – 31	2	At3g11630	Thioredoxin superfamily protein	29.074	1	At4g30270	Xyloglucan endotransglucosylase/hydrolase 24 (XTH24)	30.576
			At2g41530	S-formylglutathione hydrolase (SFGH)	30.474				
10	26 – 28.5	3	At3g55440	Triosephosphate isomerase (TPI) *	27.111	0			
			At3g62030	Rotamase CYP 4 (ROC4) *	28.19				
			At1g22300	General regulatory factor 10 (14-3-3EPSILON_GF14 EPSILON_GRF10) *	28.599				
11	25	0				1	At1g75040	Pathogenesis-related protein 5 (PR-5) *	25.236
12	21 – 24	1	At2g30870	Glutathione S-transferase PHI 10 (GSTF10)	24.084	0			
			At4g03520	Thioredoxin superfamily protein	20.314				

13	20	4		(ATHM2)		0			
			At1g67090	Ribulose bisphosphate carboxylase small chain 1A (RBCS1A) *	20.003				
			At1g76180	Dehydrin family protein (ERD14)	20.77				
			At5g39730	AIG2-like (avirulence induced gene) family protein	20.005				
14	18 – 19.5	3	At1g03680	Thioredoxin M-type 1 (TRX-M1) *	19.652	0			
			At4g38740	Rotamase CYP 1 (ROC1)	18.333				
			At3g16640	Translationally controlled tumor family protein (TCTP)	18.872				
15	12 – 17.5	3	At1g23130	Polyketide cyclase/dehydrase and lipid transport superfamily protein (Bet v I allergen-like) *	17.848	0			
			At1g22840	Cytochrome C-1 (CYTC-1)	12.39				
			At4g39260	Glycine-rich RNA-binding protein 8 (GRP8)	16.304				
	Total:	**48**				**12**			

The criteria for protein identification in this study include the presence of more than one unique peptide for any given protein, a score of 55 and above from MASCOT or any scoring with high confidence level (p<0.05) from ProteinPilot. It was also desirable that each protein be identified from both softwares and that each was detected more than once from independent runs on a variety of samples. Most of these criteria are also used by other proteomics studies (Dafoe et al., 2009; Lin et al., 2009). The Proteomics Standards Initiative provided a guideline referred to as Minimum Information about a Proteomics Experiment (MIAPE) for reporting MS informatics (Binz et al., 2008). For identified proteins MIAPE outlined these elements where all or part of it should be covered in a report (Binz et al., 2008);

1) accession code in the queried database,
2) protein description,
3) protein scores,
4) validation status,
5) number of different peptide sequences assigned to the protein (without considering modifications),
6) percent peptide coverage of protein,
7) identity of supporting peptides and
8) in the case of PMF (peptide mass fingerprinting), the number of matched or unmatched peaks

The same guidelines by MIAPE were adopted for reporting the proteins identified from the current study. Accession code was provided in terms of *Arabidopsis* Gene Index (no. 1), proteins were described together with gene name (no. 2), their status were validated by comparison with database from other plant's phloem proteome as well as manual validation (no. 4), each protein identified by number of peptide(s) assigned to it (no. 5) and percentage of peptide coverage (no. 6) was based on ProteinPilot using the percentage given for peptides with high confidence (p<0.05). The last guideline (no. 8) was irrelevant to the present study. Meanwhile, protein scores (no. 3) were included in Appendix A with conflicting scoring for some proteins that were identified exclusively using ProteinPilot. Lastly, identity of supporting peptides (no.7) was also provided in Appendix A.

It was previously reported in the results section (3.3.3) that identified protein with more than one unique peptide is highly desirable. A total of 33.5% of the identified proteins in Table 3.2 have only one peptide where 23.9% came from phloem proteins, 5.9% from the ribosomal proteins and 3.6% from cell wall proteins and contaminants. On a careful look, the largest percentage of phloem proteins with one peptide came from the functional category of unknown protein (4%). All these proteins were included in the final list as they were identified at least more than once in at least three cycles of samples run for LC-MS/MS. Even with only one peptide, they were still considered as robust identified protein as detection of these proteins in more than one sample proved their existence. Additionally, 0.9% of the identified phloem protein was also identified in brassica, cucurbit and rice, and 8.1% was found by Batailler et al (2012) and 0.9% by Guelette et al (2012).

Once guidelines for identification had been set, identified proteins were compiled. From the list (Table 3.2), further information was added especially homology and biological function of each protein. Homology searching was done against three other plant species namely brassica (*Brassica napus*, Giavalisco et al., 2006), cucurbit (pumpkin, melon and cucumber, Walz et al., 2004; Lin et al., 2009) and rice (*Oryza sativa*, Aki et al., 2008). AGI number was used for the search against those that matched brassica and cucurbit while annotation was mostly used for homology search against rice. Some of the proteins identified were also matched by annotation against brassica or cucurbit. This was done as certain proteins from the brassica and cucurbit phloem database do not contain AGI number. A number of the proteins may have come from multigene family hence they may have been assigned to different AGI numbers based on their difference in structural arrangement and properties.

3.4.2 Number of Proteins Identified Differ According to Source of Protein and Treatments

Based on proteins identified, it was observed that more hits came from total phloem sample than excised gel bands. A few other factors may be influencing the result such as efficiency of tryptic digestion which can be compared against a standard where in the current study BSA was used as standard for comparison purpose, the use of different column types in LC-MS/MS (capillary and nano- in this study) and concentration of proteins prior to detection. As several proteins may be contained within a single band of 1D gel the relative amount of each individual protein may be reduced. Hence lower quantity may reduce the chance for identifying protein with high confidence.

An essential part of protein search depends on digestion of protein into smaller peptides using trypsin, an endopeptidase that cleaves interior peptide bonds of proteins and polypeptides. Trypsin is from the serine protease family that recognize target amino acids in a binding pocket and later cleave the C-terminal amide bond using a serine amino acid activated by a histidine and an aspartate. The Combination of serine-histidine-aspartate functions as a charge relay system in which a hydrogen atom is removed from serine making it more reactive upon attacking a target protein chain (Goodsell, 2003). Trypsin binding mechanism favours basic amino acids as it has a negatively charged aspartate at the bottom of its substrate binding pocket. Formation of the ionic bond requires the target amino acid to be positively charged with long side chains. Only arginine (Arg/R) and lysine (Lys/K) fit the criteria hence trypsin is considered a very specific protease. Because of its high cleavage specificity and stability under a variety of conditions, trypsin is chosen to degrade proteins to peptides for MS-based proteomics (Aebersold and Mann, 2003). Furthermore, preferred mass range for effective fragmentation by tandem mass spectrometry (MS/MS) is provided by cleaving C-terminal of basic amino acids to arginine or lysine (Fig. 3.14) (Olsen et al., 2004).

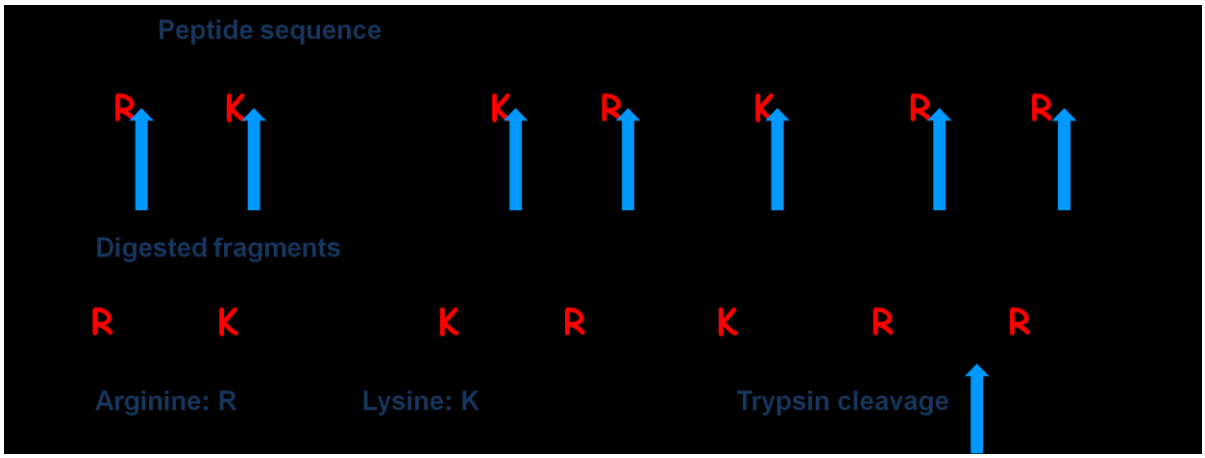


Figure 3.14: Proteolytic mechanism of trypsin that cleaves the C-terminal to arginine or lysine.

There is a range of different trypsin products on the market and those of sequencing grade were commonly used to assist peptide digestion. In general, digestion should be done with appropriate amount of trypsin for a certain period of time, usually overnight or from 4 up to 20 hr at 37^{o}C but it depends on the amount of starting protein used. Incomplete digestion may result from insufficient amount of trypsin or shorter incubation time or even from a low quality enzyme. On the other hand, artifacts have been reported from trypsin overdigestion (Srividhya and Schnell, 2006). This is due to the fact that trypsin also shared some chymotryptic activity in which peptides are cleaved after Ala/A, Tyr/Y, Trp/W and Leu/L. Transpeptidase activity that cleaves N-terminal addition of Lys or Arg has also been observed in tryptic-digested product by LC-MS/MS (Schaefer et al., 2005). In this study, in-gel digestion done on excised gel bands used smaller amount of trypsin as compared to in-solution digestion for total protein sample but in both cases trypsin was added in accordance with the protein concentration.

Even though trypsin has been proven to cleave fairly either excised gel pieces or total protein, more proteins were identified from total phloem exudate protein sample even after all the proteins from excised gel bands was combined. When all the proteins identified from both sample were compared some proteins were found to be missing from the list of excised gel bands. One reason excised gel pieces did not yield as much protein as the total phloem sap may be due to the cutting process where some

bands were not observed in some gels. In addition, keratin contamination was also observed in samples from the excised gel bands probably as a result of handling the samples especially during the cutting process as excised gel bands had to be further cut up to smaller pieces about 1-2mm to assist in better digestion. Samples might be contaminated with skin keratin, the most common contaminant seen in mass spectra. While it was not severe in the study, keratin contamination could sometimes interfere with the spectra by introducing vertical streaking parallel to the direction of migration on the gel.

The excised gel bands in this study originated from two staining methods prior to tryptic digest. While both methods differ in terms of the chemical component in the stain and its sensitivity in detecting protein, tryptic digests for both are similar in principle. The most important thing when excised gel bands are used in any proteomics work is to ensure the colour from the stain is completely removed before digestion with trypsin. This is because the chemicals contained in the stain will interfere with MS procedures. Between silver and Coomassie staining, the former is the more sensitive technique (Neuhoff et al., 1998). However, problems in nano ESI-MS/MS and reduced detection in peptide masses have been reported from the use of silver staining (Gevaert and Vandekerckhove, 2000; Scheler et al., 1998). The same problem was encountered in the current study where fewer peptides were observed in silver stained excised gel bands and the number of spectra produced following nano ESI-MS/MS was very low. It is thought that the presence of metals from the staining was not cleared thoroughly prior to digestion. Excised gel bands stained with Coomassie gave better results and many more proteins were successfully identified that matched those found in total phloem sap protein. Although Coomassie staining is preferred for its ability to produce results with better linearity, homogeneity and lesser interference with MS, its low sensitivity requires high amounts of proteins and analysis is commonly restricted to major proteins (Richert et al., 2004). That could be the reason why fewer proteins were identified in excised gel bands than total protein and most of the proteins identified were high abundance major proteins.

Many studies have attempted to improve compatibility of silver staining method with LCMS analysis. The sensitivity of silver staining is still its best

characteristic along with its simple visualization and stability over time . Earlier studies have utilised ferricyanide (Gharahdaghi et al., 1999) and hydrogen peroxide (Sumner et al., 2002) in the destaining procedures to remove the metals but inferiority in terms of peptide coverage as compared to Coomassie staining were obtained. Oxidation of silver and formaldehyde is another problem associated with destaining silver stained gel followed by complexation of silver ions. The former can be overcome with the use of either ferricyanide or hydrogen peroxide, however the latter can only be achieved using thiosulfate (Gharahdaghi et al., 1999) or ammonia (Sumner et al., 2002). With this knowledge it is possible to increase MS efficiency by designing destaining procedures of silver stained gel for better compatibility. Richert et al (2004) applied a formaldehyde-free silver staining method that proved to improve sequence coverage while still maintaining sensitivity and simplicity of silver staining. Due to time limitation, the formaldehyde-free silver staining was not tested in the present study.

Later experiments following method development (chapter 2, section 2.2.2.1) saw some changes incorporated in the hope of improving protein quality and eliminating other protein that might have cross-contaminate into the phloem during wounding and exudation. Data analysis from MS showed a reduction in peptide recovery of ultracentrifuged samples and a lower number of proteins was identified (Table 3.2). Normally ultracentrifugation is applied to total protein extract to separate soluble and insoluble fractions such as in pumpkin where Yoo et al. (2002) separated soluble and microsomal fractions from total vascular protein extraction via ultracentrifugation. As mentioned earlier in the results (section 3.3.3) a lot of the cell wall protein and its contaminants were not detected after ultracentrifugation was applied to samples prior to purification. However as this method was applied rather late in the study, further LC-MS/MS runs could not be performed to confirm the protein profile differences after ultracentrifugation. Regardless, ultracentrifugation may provide an alternative to remove cell wall proteins or other contaminants identified in Table 3.2. Cell wall proteins are mostly insoluble and they will be better separated by ultracentrifugation that has been proven to separate soluble and insoluble fractions.

Apart from applying ultracentrifugation, phloem exudates were also subjected to a different collection technique. The first collection was obtained 30 minutes after wounding the base of the petiole and the second collection was obtained after overnight incubation from the same sample but in a fresh EDTA solution. While the protein pattern on SDS-PAGE did not seem to differ (chapter 2, Fig 2.9), Bradford Assay done to determine protein concentration showed that the first collected exudates contained lower amounts of protein (Fig. 3.3). It was hypothesized that during wounding, the phloem protein that surged out will also be contaminated with protein from nearby cells where the wound was made and it was hoped that the first collection would eliminate any contaminant to leave the second exudates with only proteins from the phloem. However, the hypothesis remains unconfirmed as reduced number of proteins was identified from the second exudates as compared to previous experiments with single collection of exudates. Guelette et al (2012) utilised an EDTA-facilitated method whereby *Arabidopsis* leaves were incubated in EDTA for an hour followed by second exudation in water. They managed to identify 65 proteins. The small number of identified proteins is somewhat similar to what was encountered in the current study when using exudates from the second collection time. Nevertheless, the quality of proteins seemed to improve with reduced number of cell wall protein, contaminants and ribosomal protein.

When comparing the ease of experiment performed using total phloem protein and excised gel bands, the former was definitely much more convenient and faster. In-solution tryptic digest is employed for total phloem which takes less time than in-gel tryptic digest for excised gel bands though both protocols apply the same principle in which protein samples are brought up to pH 8 by ammonium hydrogen carbonate and reduced prior to tryptic digest. In terms of proteins identified from both methods, in theory there should not be any difference as gel bands run on the SDS-PAGE gel should contain the same protein as the total protein. However, technical error during excision of gel bands may cause one or more protein to be left out. Less visible bands may sometimes be left out from being excised for analysis or it might have been cut off accidently. It would be best to ensure that all possible bands, even those that are less visible to the eyes are excised by leaving no gap in between the gel pieces or sections. Both total exudate protein and excised gel band methods are suitable for

protein identification, each with their own advantage and disadvantage. Nevertheless, total exudate protein offers a better approach as it saves a lot of time without the need to run SDS-PAGE with shorter procedure in tryptic digestion and produced a larger number of identified proteins.

3.4.3 Phloem Exudate Protein Content from Various Genotypes

Protein content from phloem exudates was expected to be different from one genotype to another. Although the pattern of protein bands across the genotypes was similar on SDS-PAGE (Fig. 3.1), the absence or presence of certain protein(s) could not be reliably ascertained. Nevertheless, from analysis of MS/MS spectra, variation in protein content was detected across the five genotypes used initially in the study. Among the samples, wild type plants contain the most protein (85.52%) although it also lack certain protein found in some of the mutant as shown in Table 4.1. As four of the samples are mutant, loss of gene function in each mutant is expected where a gene is knocked out which may or may not have an effect on downstream genes. Furthermore, one of mutant, *ft-tsf* is a double mutant with double loss of flowering time genes. These could be the cause for the variations observed. However, as this study was not carried in depth to determine actual total list of protein content or quantitative variation from each genotype, no conclusive result could be reported. It would be beneficial if focus on number of samples analysed from each genotypes could be increased and possibly run at the same time for future study in this matter. Additionally, each genotype could also be subjected to photoperiod treatment for example, plants grown in SD versus plants grown in LD to compare if variation also occurs.

Being a long day (LD) plant, *Arabidopsis* flowers faster in LD as compared to SD. The key to induce flowering under LD in *Arabidopsis* is via production of CO. CO protein is only accumulated due to action of *phytochrome A (phy A*) and *cryptochrome 2 (cry 2*) and promoted by far red light in longer day length. In turn, CO will activate FT which will activate SOC1 in the floral meristem to signal flowering (Turck et al., 2008). However, from the result in 3.3.2 and list of identified proteins (Table 3.1), more proteins were identified from *co* as compared to the other

three mutants. The finding is unexpected in a way; for example lesser amount of proteins was identified from *ft* (15.93%) as compared to *co* (42.08%) from total phloem exudates protein list. CO activates FT and reduced expression of *FT* has been observed from *co* plants (Kardailsky et al., 1999) hence more protein is expected to be contained within *ft* mutant plants than *co*. However, as mentioned earlier in depth comparisons of genotypes were not carried out to determine the actual total proteome profile from each genotype. Hence, the % values are tentative since only two analyses of each genotype was conducted

Turck et al (2008) reported that CO may also activate TSF in addition to FT through inferred interconnection. While FT directly promotes flowering through SOC1 in the floral meristem, TSF is postulated to induce flowering by activating SOC1 in the inflorescence meristem. Jang et al. (2009) confirmed that *FT* and *TSF* are both partially redundant promoters of flowering regulated by *CO*. Recently cytokinin was also reported to activate TSF in addition to FD and SOC1 (D'Aloia et al., 2011). *tsf* mutants have not been classified as late flowering mutant where in most studies it was reported as having wild type phenotypic characteristics (Michaels et al., 2005; Yamaguchi et al., 2005. In both studies above, activation-tagging vector and overexpression respectively were performed on *tsf* mutant and early flowering was reported. The *tsf* used for proteomics study is different where the T-DNA insertion is in the intron region and the mutant was found to have wild type phenotypic characteristics. It is of no surprise that the number of proteins observed in this mutant is the highest among the other mutant plants used even though it did not exceed that of wild type.

The last mutant is *ft-tsf* double mutant. This is an extremely late flowering mutant that produced flowers later than all other mutants. As described previously that *tsf* mutant behaves almost as wild type (under LD) but evidently when *tsf* is coupled with another mutant, in this case *ft*, the result was a late-flowering mutant. This double mutant produced the second lowest number of proteins after *co*. As both *FT* and *TSF* have been reported to act redundantly with each other, late flowering characteristics exhibited by this double mutant is presumably related to delayed *SOC1* expression caused by the absence of *FT* and *TSF* activities (Jang et al., 2009). And,

this may also contribute to reduced number of identified protein though probably less severely than *co* mutant. The actions of both proteins are required for flowering under LD but not in SD as the double mutant flowered marginally later than WT as reported by Michaels et al. (2005) and Jang et al. (2009).

Even though the results obtained from MS analysis (Table 3.2) showed that WT contained the most identified proteins, this does not mean that there is a wide genotypic difference in proteomes. Most of the repeats were performed using WT (six repeats for WT versus one repeat for the mutants) hence identification of more proteins was of no surprise. In summary, the protein content of the five genotypes could not be compared conclusively as more samples needs to be analysed. Even though the absence of a gene upstream may affect the downstream genes, the effects on the proteome remain to be revealed.

3.4.4 *Arabidopsis* Phloem Exudate Proteins and Possible / Predicted Functions of the 16 Homologous Proteins across Four Phloem Proteomes

A substantial number (172) of proteins were identified from the phloem exudate of *Arabidopsis thaliana*. Proteins present in the phloem of a plant species most likely originated from immature SE or CC (Turgeon and Wolf, 2009). Upon maturity, SE lack nucleus and functioning ribosomes hence synthesis of phloem proteins are thought to occur in the CC (Dafoe et al., 2009). Nevertheless, the mechanism behind regulation of proteins transported into and out of the SE is largely unknown. Not all proteins in SE are transported over long-distances (Dafoe et al., 2009). There are immobile proteins or proteins that seemed to be immobile and there are those closely related to plasma membrane (van Bel, 2003). These immobile proteins are only released when SE is wounded.

Comparison of proteome conservation between *Arabidopsis*, brassica, cucurbit and rice was summarized in a Venn diagram (Fig. 3.5). All of the proteins shared similar AGI number and annotations. *Arabidopsis* shared the most proteins with cucurbit (23.5%) followed by rice (22.6%) and brassica (21.7%). Although both *Arabidopsis* and brassica came from the same family of Brassicaceae, both plant

species shared the least proteins in their proteomes. Higher overlapping proteins shared by *Arabidopsis* and cucurbit may be due to the established cucurbit phloem database which contains a large number of identified proteins. Accession number from different database may have an effect in the hit rate reported here along with the total size of phloem proteome reported for each species as they differ with one another. For instance, cucurbit has the largest database (>600 phloem proteins) compared to the other three species and even *Arabidopsis* from this study has uncovered twice as much protein than both brassica and rice. This opens up the possibility of many proteins that have yet to be identified from the phloem of each plant species. Generally, each of the four plant species including *Arabidopsis* in this study has their own unique proteins not shared with any other species that may have a specific function in the phloem of each species or may have yet to be discovered in other species. Identification of a wide variety of biochemically active proteins involved in various metabolism in this study suggested plant's systematic regulation of metabolism, development and defense responses (Nakamura et al., 1993; Balachandran et al., 1997; Ishiwatari et al., 1998; Kehr et al., 1999; Schobert et al., 2000; Aoki et al., 2002; Yoo et al., 2002; Walz et al., 2004; Aoki et al., 2005) with an organised and efficient genes, proteins and enzymes activities in each developmental stage.

From Table 3.2 and Fig. 3.5, 16 proteins were found to be conserved within all four species (Table 3.9), whereby two of them namely enolase and rotamase CYP3 were also identified in the phloem proteome of poplar (Dafoe et al., 2009).

Table 3.9: List of 16 proteins conserved within *Arabidopsis*, brassica, cucurbit and rice. Proteins highlighted in blue were similar to proteins identified by Batailler et al (2012) and * represented proteins identified by Guelette et al (2012).

	AGI	Protein ID	Mass observed (Da)	No. of sequence (ESI-MS)
	Redox Reaction			
1	At1g07890	Ascorbate peroxidase 1 (APX16)	27413	4
2	At1g45145	Thioredoxin 3 (TRX3)	13227	3
	Metabolic Pathway **A) Glycolysis**			
3	At2g36530*	Enolase (ENO2) / (LOS2) *	47719	3
4	At3g04120*	Glyceraldehyde-3-phosphate dehydrogenase C subunit 1 (GAPC1)*	36491	3
	TCA			
5	At1g04410	Lactate/malate dehydrogenase family protein (c-NAD-MDH1)	35548	4
6	At1g53240	Lactate/malate dehydrogenase family protein (mMDH1)	35782	4
	Nucleotide Metabolism			
7	At5g03300	Adenosine kinase 2 (ADK2)	31337	7
	Secondary Metabolism			
8	At5g26000*	Thioglucoside glucohydrolase 1 (TGG1) *	61133	10
	Cell Organization and Biogenesis			
9	At2g43530	Actin 7 (ACT7)	41774	5
10	At1g35720*	Annexin 1 (ANNAT1) *	35757	10
11	At2g16600	Rotamase CYP 3 (ROC3)	18444	2
	Protein Biosynthesis, Organization, Folding, Transport and Modification			
12	At5g02500	Heat shock cognate protein 70-1 (HSC70-11)	71358	2
13	At5g58290	Regulatory particle triple-A ATPase 3 (RPT3)	45711	1
14	At1g31340	Related to ubiquitin 1 (RUB1)	8525	1
	Developmental Process			
15	At3g16640	Translationally controlled tumor family protein (TCTP)	18872	2
	RNA			
16	At2g21660*	Glycine-rich RNA-binding protein (GRP7) *	16596	4

In general, the biochemistry occurring within the sieve elements could probably be portrayed by the presence of ADK2, malate dehydrogenase and ascorbate peroxidase along with other proteins involved in redox reaction and metabolism from Table 3.2. Within the highly reduced enucleate sieve elements, detection of ubiquitin, heat shock cognate 70-1 and actin-7 indicated the presence of basic cellular components (Lin et al., 2009). Each of these proteins may play multiple roles as discussed below.

3.4.4.1 Proteins Representing Basic Cellular Component in the Phloem

Plant actins function in root hair growth tip extension and streaming of organelles (Staiger et al., 1994; Bedinger et al., 1994; Thorn et al., 1997). ACT7 identified here is a vegetative actin (McDowell et al., 1996; Meagher et al., 1996) and a study on ACT7 isovariant in *Arabidopsis* showed that it is important during callus formation (Kandasamy et al., 2001). High levels of ACT7 were expressed in emerging rosette leaves and trichomes, young inflorescence stems, stomata and surrounding cells, and subsets of floral tissue (McDowell et al, 1996). ACT7 responds to the plant changing environment where wounding and treatment with exogenous hormones altered its expression patterns (McDowell et al., 1996; Kandasamy et al., 2001; Gilliland et al., 2003).

Various cell processes such as folding of newly synthesised protein, solubilising protein aggregate, disassembly of protein complexes, membrane translocation and controlling biological activities of regulatory proteins are aided by heat shock cognate 70-1 (HSC70.1), also described as a molecular chaperone. Its function in cellular components was exhibited during cold stress (Krenz et al., 2010) where it was found in the nucleus. HSC70.1 played a role in transport-related or glycosylation-related activities and DNA-replication activity as reported in tomato by Duck et al (1989). The protein might also be in involved in transporting photoassimilates in the phloem. Although it has yet to be proven, in the future it is worth to study this protein and its function in the phloem.

3.4.4.2 Proteins Involved in Maintaining the Biochemistry of Phloem

As one of the enzymes involved in the biochemistry of sieve elements, adenosine kinase (ADK) is essential for sustainability of methyl recycling. It phosphorylates adenosine and its nucleoside analogs into monophosphate derivatives and is ATP dependent. Being the housekeeping enzyme, ADK2 transcription level is highly expressed in the leaf and stem as compared to ADK1. Between the two ADK isoforms, there is high amino acid conservation (Moffatt et al., 2000). ADK protein level in various organs was monitored via immunobloting by Moffatt et al (2000) where all organs showed substantial level of ADK accumulation. This is especially true in silique, stem and mature flowers however lower amount of ADK was detected in leaves, roots and dry seeds of *Arabidopsis*.

Another enzyme involved in the biochemistry of phloem is the ascorbate peroxidase (APX) where in plants it functions to remove hydrogen peroxide in the chloroplast and cytosol. There are a variety of plant APX isoforms and they differ in molecular weight, optimal pH, stability and substrate specificity (Dabrowska et al. 2007). Found in several cellular compartments, the APX in *Arabidopsis* is found in the cytosol and under extreme light or heat stress conditions the cytosolic APX is unducible (Karpinsky et al., 1999). Davletova et al (2005) reported oxidation of chloroplastic proteins due to lack of cytosolic APX1 implementing that chloroplast protection is probably played by APX1. In addition, higher level of ROS (reactive oxygen species) and overexpression of genes associated with oxidative stresses was caused by disruption of the *APX1* gene (Pneuli et al., 2003)

Plant malate dehydrogenase (MDH) has various roles which depend on cellular localisations and cofactor specificities (Martinoia and Rentsch, 1994). Specifically it catalyses interconversion of malate and oxaloacetate (OAA) coupled to oxidation or reduction of NAD pool. Its function is also reversible but reduction of OAA is preferable. Isoforms of the enzyme can be found in mitochondria, chloroplast, peroxisome and cytosol. Due to wide range of compartments the enzyme is contained in reaction can occur across compartments provided by ready transport and, utilisation of malate, OAA and availability of NAD (Kromer, 1995). The exchange of reaction

between compartments was achieved through membrane transport but is controlled strictly as a result of redox differences in NAD(H) pool between compartments (Igamberdier and Gardestrom, 2003). In this study, two out of three of the identified malate dehydrogenase isoforms shared homology with the other three phloem proteomes and were from mitochondria and cytoplasmic isoforms. MDH from mitochondria (mMDH) operates in three different pathways; 1) TCA (tricarboxylic acid) cycle, 2) operate in reverse direction during conversion of G to S and, 3) provide supply of CO2 for fixation in bundle sheath chloroplast (Journet et al., 1981; Hatch and Osmond, 1976, Tomaz et al., 2010). In different plants, mMDH isoforms abundance differ between tissues and may indicate that the mMDH may not always expressed at same level but could be used as significant regulator of respiratory rate in plant (Tomaz et al., 2010). On the other hand, cytoplasmic MDH overexpression caused 4-fold increase of root organic aid in alfalfa (*Medicago sativa*) and high rates of organic acid exudation that increased aluminium tolerance through metal chelation in soil (Tesfaye et al., 2001). Malate is formed from reduction of OAA which oxidises NADH to NAD+ and malate can even enter mitochondria for reversible reaction by mMDH. Study by Tesfaye et al (2001) implies alteration in metabolic flux as a result of changing amount of MDH isoforms to form a range of organic acid which could have long-term effects on plant growth and development (Tomaz et al., 2010).

Another enzyme with shared homology to the other three phloem proteome is enolase, an enzyme involved in the final step of glycolysis. 2-phosphoglycerate (2-PGA) is catalyses by enolase or 2-phospho-D-glycerate hydrolyase into phosphoenolpyruvate (PEP) (Zhao et al., 2004). Enolases in plant can exist either as cytosolic or plastid isoforms and in *Arabidopsis* the enzyme was not imported into chloroplast based on the inability to detect the enzyme's activity and immunoassay in purified chloroplasts from the leaves (van Der Straeten et al., 1991). In eukaryotes, enolases with the ability to bind to polynucleotides and cytoskeletal proteins play important roles in biological and disease processes (Liu et al., 2010). In contrast to yeast enolase which is involved in growth control and thermal tolerance, and identical to heat shock protein 48 (HSP48) (Iida and Yahara, 1985), plant enolase is not a heat shock-inducible protein (van Der Straeten et al., 1991). However, plant enolases may function like heat shock protein (Liu et al., 2010).

Glyceraldehyde-3-phosphate dehydrogenase (GAPC) is also a glycolytic enzyme with multiple functions. In addition, the enzyme plays a role in DNA repair and as DNA-binding protein (Kim and Dayo, 2005). In *Arabidopsis thaliana*, GAPC had been correlated with other glycolytic enzyme in the cytoskeleton and mitochondria (Hoetgrawe et al., 2005). The cytosolic isoform is localised in the plant's nucleus cell (Anderson et al., 2004). Due to this, speculation was made that nuclear translocation and, or binding to DNA is mediated by modifications of GAPC itself (Holtgrefe et al., 2008). GAPC is activated upon oxidative stress hence it was suggested to play regulatory role in ROS signalling in plants (Hancock et al., 2005). During zygotic and, or early embryo development, GAPC activity is fundamental however GAPC protein was expressed at low level with reduced activity in leaves and flower via western blot analysis (Rius et al., 2008).

3.4.4.3 Stress Response Proteins

Some of the identified proteins have more than one function as shown in Appendix B. An example is HSC70-1 which apart from being a protein involved in basic cellular components of sieve elements is also a protein involved in stress response and defense mechanism. As a protein that responds to stress, HSC70-1 is able to bind substrates at low temperature such as during winter. Plants could be injured or killed by low temperature and their protein biogenesis is impaired. Not only that, unfolding of certain labile protein may occur and blocking the unfolding of certain nascent proteins after release from ribosome may result from cold temperature (Zhang and Guy, 2006). The presence of HSC70-1 with its chaperone activity would be advantageous to ensure proper folding of newly synthesised protein (Krunz et al., 2010).

Another identified protein, annexin (ANNAT), plays a role in apolar growth, growth regulation and stress response. Plant annexins are ubiquitous, soluble proteins (Mortimer et al., 2008). It is expressed throughout body and lifespan of plant including in the phloem sap (Barnes et al., 2004) but was mostly found in the cytosol (Baucher et al., 2011). This coincides with the other three plants for comparison as

well as the annexin found in the current study. Annexin is localised in the plasma membrane and its microfilaments, vacuoles, on the golgi and golgi-derived vehicles and nucleus. Expected to have various functions in plants, annexin is involved in oxidative stress response (in *Arabidopsis* and cotton), general environmental stress (alfalfa, *Medicago sativa*), mechanical stress response (*Bryonia dioica*), low temperature transduction (*Lavatera thuringaea* and *Triticum destivum*), cell volume or vacuole size regulation (*Nicotiana tabacum*) and exocytosis (*Zea mays*) (Hofmann, 2004). In oxidative stress response *Arabidopsis* Oxy5 (annexin) protein is able to restore *E.coli* OxyR growth in the presence of hydrogen peroxide (H_2O_2) where the expression was found to be increased in the presence of H_2O_2 or salicylic acid (Gidrol et al., 1996). Kush and Sabapathy, 2001 reported resistance against H_2O_2 in the by Oxy5 and decreased production of O_2. The same study concluded Oxy5 may act as a sensor of increased calcium influx associated with peroxidase stress. As plant annexins bind Ca^{2+}, in the presence of micromolar of Ca^{2+} it will bind to negatively charged phospholipids (Balasubramaniam et al., 2001). Fortunately this binding can be reversed by adding Ca^{2+} chelators (Mortimer et al., 2008) to allow phospholipid binding in Ca2+ independent manner (Baucher et al., 2011).

Glycine rich-RNA binding protein 7 (GRP7) is also one of the proteins shared by the four phloem proteomes including *Arabidopsis* with a possible involvement under cold stress condition both in monocot and dicot (Kim et al., 2010). Freezing tolerance of *Arabidopsis* plants was conferred by acceleration of seed germination and seedling growth under low temperature by AtGRP7 (Kim e t al., 2008). During cold adaptation process in *E.coli*, AtGRP7 has an RNA chaperone activity (Kim et al., 2007) and regulates mRNA export in guard cells to confer tolerance to cold (Kim et al., 2008). These provide evidence of a role of AtGRP7 and RNA chaperones during stress adaptation in plants and supply potential means for the development of stress-tolerant crops (Kim et al., 2010).

In addition, GRP7 may have possible function in mRNA stability and control as well as pathogen defense (van Nocker and Vierstra, 1993; Staigner et al., 2003). To regulate its own transcript, GRP7 is needed by binding to its own mRNA at the 3' UTR and second half of its intron. Woloshen et al (2011) reported production of

alternate transcript at elevated GRP7 protein level but it was degraded quickly with a half life of 30 min. They hypothesised that during circadian cycle, accumulation of GRP7 occurred and once reaching its threshold GRP7 will bind its own mRNA resulting in unstable alternative transcript causing a decline of functional protein.

The last common protein under stress response is thioredoxin 3 (TRX3). Thioredoxins are a big family comprising of subclass *f*, *m* and *h*, with all subclasses identified from the current study. Generally, plant thioredoxins interact through disulfide bridge with ascorbate peroxidise (no. 1 in Table 3.1) and secretory peroxidases, dehydroascorbate reductase (no. 6), superoxide dismutase (no. 4 and 5) and germin-like protein (no. 186) (Santos and Rey, 2006). TRX3 has a function during germination and early seedling development (Wong et al., 2002; Marx et al., 2003) and in barley, TRX3 regulate its germination (Li et al., 2009). The protein is found in abundance in the phloem sap of many monocots and dicots implying they may act as messenger protein (Ishiwatari et al., 1995; Schobert et al., 1998) with a role in DNA replication (Balmer et al., 2003; Barajaz-Lopez, et al., 2007). In *Arabidopsis*, AtTRX3 is strongly induced by biotic and abiotic stress. Its expression is closely related to wounding, abscission, senescence, pathogen attack and several different oxidative stress conditions (Laloi et al., 2004). High expression was deduced from GUS activity from vascular tissues surrounding the wounding area of green leaves or stems in PrTRX::GUS plants. During early stages of flower development no GUS activity was detected apart from in pollen and sepals of mature flowers (Reichheld et al., 2005). Gus staining was also detected at later stages of plant development in anthers, siliques, floral organs abscission zones and early senescence leaves (Laloi et al., 2004). The same group reported that only AtTRX is induced under oxidative stress condition and its specific involvement in resistance against oxidative stress among thioredoxin *h* family was proposed.

3.4.4.4 Defense Proteins

Many of the identified proteins are stress-related proteins and it is of no surprise that they may also have function in defense. However, the defense mechanism of some of the proteins is not known and is not supported by previous

work that can verify it. Myrosinase and thioredoxin are best known for the role they played in plant defense.

Myrosinases are a group of isoenzymes present in all Brassicaceae (Rodman et al., 1991). Known as a metabolic defense system in plants (Bones and Rossiter, 1996), the myrosinase system comprises of glucosinolates, myrosinase and thioglucosides. Furthermore, glucosinolates are synonymous with the Brassicaceae family where the hydrolysis products function in plant defense against microorganisms and insects. Not only do developmental stages affect the activity and specificity of many myrosinase isozymes, the chemical structure of glucosinolates and myrosinase-associated factors such as epithospecifier proteins (no. 56) and myrosinase-binding proteins (no. 61, 62, 161, 162) are also contributing factors. Similar findings were reported by Giavalisco et al (2006).

Thioglucoside glucohydrolase 1 (TGG1) is one of the myrosinase proteins found. Both *Arabidopsis* and *Brassica napus* are from the Brassicaceae family and they shared at least two known myrosinase isozymes (no. 56, 58 from Table 3.2). Hydrolysis of glucosinolates is catalysed by myrosinase to prduce degradation products Glc, sulphate and either isothiocyanates, nitriles, thiocyanates, amines, epithinitriles or oxozolidine-2-thiones (Reichelt et al., 2002). Plant myrosinase-glucosinolate system is a well known defense system against bacteria, pathogen and herbivores. The system includes myrosinase enzyme, interacting proteins such as epithiospecifier protein (no. 56), myrosinase-binding proteins (no. 61, 62, 161, 162), myrosinase associated proteins, nitrilases (no. 57) and, a total of more than 100 glucosinolate susbstrates (Bones and Rossiter, 1996).When plants are under attack or a tissue is injured, the system releases glucosinolates from the vacuole and myrosinase will hydrolysed it into various toxic small molecules (Wittock and Hallaer, 2002; Barth and Jander, 2006). Total myrosinase (or Thioglucosidase, TGG) activities differ between plant tissues and during leaf development. In *Arabidopsis*, expression of TGG1 was directed in the guard cells and phloem cells (Thangstad et al., 2004). The pattern of the TGG1 promoter activity shown by GUS activity was consistent with TGG activity detected in native polyacrylamide gel and northern blot analsis by the same group. With its expression localised in guard cells in every organ,

Arabidopsis TGG1 is the only gene with such expression. The presence of myrosinase in the guard cells is thought to be aprt of either the defense system activated after destruction of guard cells and, or is part of a system generating volatiles from degradation of glucosinolates. On the other hand, activities of TGG1 protein was detected in the pods, imbibed seeds, cotyledons, leaves, bracts, inflorescence stems and flower buds (Thangstad et al., 2004). A follow-up experiment by Alvarez et al (2008) showed that TGG1 is more highly expressed in hypocotyls than in suspension cells through western blot analysis utilising TGG-1- and TGG-2-specific antibodies. In addition, TGG1 seemed to partake in stimulating a stomatal response caused by wounding (Zhao et al., 2008).

As for thioredoxin 3, it defends plants through being induced during pathogen attack. TRX3 function in defense system is mediated by X-box cis element known for binding sites for WRKY plant specific transcriptional regulators controlling several types of plant stress responses usch as pathogen defense, wound response and senescence (Eulgem et al., 2000). Since the method of obtaining the protein was via wounding just as in brassica and cucurbit, identification of TRX3 and the myrosinase isozymes is of no surprise.

3.4.5.5 Protein-Folding and Destruction

As ubiquitous proteins, cyclophilins are involved in a variety of roles. One of them is *in vivo* protein folding (Mathouschek et al., 1995). Four cyclophilins were identified from the current study and one further protein, FK506-binding protein (no. 131) also exhibit rotamases activity. ROC3 mutant which is shared among the four phloem proteome have reduced mRNA expression (Romano et al., 2004). A small increment in cyclophilin mRNA expression was observed 2h and 8h after wounding, with highest expression in leaves (Chou and Gasser, 1997). The basic role played by ROC3 in protein folding may ensure that proteins produced are functional and hence the existence of this protein in the phloem may be beneficial.

On another level, plant responses to light, sucrose, developmental cues and pathogens are mediated by ubiquitin in plant signalling pathways (Devoto et al.,

2003). Ubiquitin's major function however is for protein destruction by targeting specific proteins (Rechsteiner, 2005). UBC 19 and 20 transcripts were expressed in the shoot apical meristem, leaf primordia and young leaves whereas in roots in situ hybridisation showed expression in the root meristem and emerging lateral roots. Meanwhile, UBC 19 protein fused to GFP and expressed in tobacco BY2 cells (*Nicotiana tabaccum* cv Bright Yellow 2) resulted in the protein being found in the cytoplasm and nucleus. It was hypothesised that mechanism by which UBC19 entered the nucleus is only by passive diffusion (Criqui et al., 2002).

3.4.4.6 Major Regulators of Cell Growth

One of the most interesting proteins found is a major regulator of cell growth in animals and fungi, TCTP (Berkowitz et al., 2008). Previously MacDonald et al (1995) characterised this protein as IgE-dependent histamine-releasing factor. Apart from the other 3 plants used for comparison, TCTP was also detected in the phloem sap of *Ricinus communis* (Barnes et al., 2004). While sharing the key GTPase binding surface property, there is a distinct sequence differences between plant TCTPs from non-plant homologs. Berkowitz et al (2008) reported stable expression throughout plant tissue and developmental stages with increased expression of *Arabidopsis* TCTP using GFP reporter lines. Earlier, Cans et al (2003) described interaction of TCTP with eEF1A (elongation factor Tu family) which is also one of the identified proteins in *Arabidopsis*, brassica and cucurbit, and, eEF1B which also function as guanine nucleotide dissociation inhibitor. However, only Aoki et al (2005) had so far described that TCTP might be involved in long-distance movement of phloem proteins.

3.4.4.7 Protein Catabolic Process

The 26S proteasome is a large ATP-dependent protease with multi-subunit where polyubiquitinated proteins are degraded (Voges et al., 1999). Although it possesses enzyme activity to hydrolyse all four nucleotide triphosphate, ATP and CTP are favourable over GTP and UTP (Rechsteiner, 2005). The identified protein from the current study was RPT3, one of the many isoforms from the large 26S

proteasome family which functions as an ATPase shared homology with the other three phloem proteome. The UPS system played an important role in hormone perception and response (Santner and Estelle, 2009) and the presence of RPT3 and UPS in the phloem may provide strong indication of their role in plant growth and development. In addition, many of the proteins identified from *Arabidopsis* proteome (1,3000 proteins or 5%) play a role in the UPS system (Vierstra, 2009).

It is highly likely that the 16 proteins with shared homology to the other three phloem proteomes are phloem sap protein. Certain proteins have been identified as phloem protein such as P protein (Balachandran et al., 1997) and lectin in cucurbit (Dinant et al., 2003) however many are still not listed and confirmed. In 2002 Walz et al listed a few examples of proteins from SE of *Cucurbita maxima* and *C. sativus* that was also found in the current study. They are peroxidase, superoxide dismutase (SOD) and dehydroascorbate reductase. Although these three proteins were not found in all four phloem proteomes, these proteins have been found in the phloem of many other species. Suggestion was also made that phloem sap protein may be involved in stress responses.

3.4.5 Proteins from the Phloem Exudate Have Diverse Function in Response to Leaf Wounding and Photoperiod

To date two other studies have been reported on phloem exudates of *Arabidopsis*. Guelette et al (2012) documented identification of 65 phloem exudate proteins via EDTA facilitated method whereas, Batailler et al (2012) described 287 proteins obtained by similar methods. Both summarised the proteins as having diverse functions. Previous studies likewise have shown that the phloem contains proteins with diverse function but mainly involved in metabolism, signalling and transport (Barnes et al., 2004; Walz et al., 2004; Giavalisco et al., 2006; Aki et al., 2008) as well as stress-related proteins (Hayashi et al., 2000; Walz et al., 2002; Giavalisco, 2006; Lin et al., 2009). Similar findings were also obtained from the current study with many proteins having more than one annotated biological function.

From GO analysis the identified proteins were categorised based on their biological process, molecular function and cellular compartment. In the biological process functional category, the highest percentage of genes was stress-related (Fig. 3.5) which is expected due to the means by which the phloem exudate was obtained. However, these proteins might also be localized within the SE as plant defense system. When the petiole was wounded, plant defense mechanism would likely trigger the expression of some of these proteins. However phloem being the vascular tissue for exchange of nutrients and many other compounds may also constitutively express some of these defense and stress-related protein that in companion cells and sieve elements. Although plants used in this study were healthy and unstressed, the transfer from SD to LD exposure could be a contributory factor. Dafoe et al (2009) also reported a large number of stress-related proteins from healthy and unstressed poplar phloem exudate and according to study done by Le Hir et al. (2008) expression of stress-related proteins in phloem tissues was found to be constitutive. Both studies in poplar and *Arabidopsis* utilised wounding of the leaf to acquire phloem sap.

Apart from stress response proteins discussed previously in section 3.4.4.3, other identified proteins from Table 3.2 have diverse functions in response to wounding and photoperiod. It is speculated that some of the unique proteins from Table 3.3 may be proteins that respond to photoperiod. The speculation was based on the fact that none were reported by Batailler et al and Guelette et al, both of which obtained their proteins by growing *Arabidopsis* solely in LD. Nevertheless, on a whole, most of the identified proteins have more than one function (Appendix B) however the possible roles of these proteins which has not been covered in section 3.4.4 will be discussed according to their known general function.

3.4.5.1 Redox Regulation, Anti-oxidant Response and Amino Acid Metabolism in Maintaining the Biochemistry of Phloem

A large number of enzymes were found within *Arabidopsis* phloem exudate similar to previously reports from other plants (Barnes, et al., 2004; Walz et al., 2004; Giavalisco et al., 2006; Aki et al., 2008; Lin et al., 2009; Dafoe et al., 2009). The broad range of enzymes found with roles as antioxidant and defense, and functions in

redox regulation, anti-oxidation responses and amino acid metabolism were consistent with previous reports (Dannenhoffer et al., 2001; Haebel and Kehr, 2001; Hoffmann-Benning et al., 2002; Walz et al., 2002; Barnes et al., 2004; Walz et al., 2004; Giavalisco et al., 2006; Aki et al., 2008; Lin et al., 2009; Dafoe et al., 2009). During their lifetime, functional SEs are maintained by proteins and metabolites with redox and antioxidant properties (Walz et al., 2002). Among proteins known to have these functions are monodehydroascorbate reductase (no. 10, 11, 12), ascorbate peroxidase (no. 1), GST (no. 158, 159), thioredoxin h (no. 13, 16), glutaredoxin (no. 9) and peroxiredoxin (no. 22). Reactive oxygen species (ROS) are generated when plants are exposed to abiotic or biotic stress and the proteins above are likely to be involved in removal of ROS or in regeneration of proteins inactivated by oxidative stress (Giavalisco et al., 2006). Ferredoxin (no. 84, 85), a metal binding protein may also be related to redox regulation and antioxidant defense.

The presence of antioxidant proteins such as glutathione S-transferase, and defense proteins like trypsin inhibitor and pathogenesis-related (PR) protein had been suggested to conserve the function of phloem proteins not only in *Arabidopsis* but also in other plants (Kehr, 2006). However, the presence of PR-5 protein in present study could also be associated with minor contamination during collection of exudate as this protein is known to be found in cell walls and apoplast (Dafoe et al., 2009; Bayer et al., 2006). In addition, Dafoe et al reported intracellular detection of PR-5 protein in organelle-like structure most likely to be plastids. Nonetheless, PR-5 protein was still identified in the sample after ultracentrifugation suggesting that probably an extra time is needed to pellet down the protein.

3.4.5.2 Signal Transduction

Calcium ions play an important role in signal transduction via the sieve tubes (Nakamura et al., 2006) where they function as a nutrient and second messenger. One of the proteins identified, annexin (no.106) functions as a target for Ca^{2+} and came from mutigene family of calcium-dependent membrane-binding protein (Clark et al., 2001). In response to certain stimuli, annexins undergo cellular redistribution and are involved in membrane trafficking, ion transport, mitotic signalling, cytoskeleton

rearrangement and DNA replication. Calmodulin (no. 118, 119) is another calcium-modulator protein that alters the activity of calmodulin-binding proteins such as kinase, glutamate decarboxylase and ATPases by binding to $Ca^{2+.}$ Transduction of a variety of abiotic and biotic signals to sink organs is also expected to be mediated by calmodulin (Nakamura et al., 2006).

3.4.5.3 Metabolism

Various metabolic pathways are known to be present in the phloem. A complete glycolytic pathway may be present in the phloem by the presence of enolase (no.25), UDP glucose pyrophosphorylase (no. 29, 30), glyceraldehyde-3-phosphate dehydrogenase (no. 27, 74), phosphoglycerate kinase (no. 75, 76), malate dehydrogenase (no. 31, 34, 35), fructose bisphosphate aldolase (no. 72, 73) and nucleoside diphosphate kinase (no. 52, 53). After wounding of SEs to get the phloem sap a product from glycolysis, UDP glucose pyrophosphorylase is crucial for callose synthesis for sealing the wound (Giavalisco et al., 2006).

Amino acid metabolism is regulated by adenosylhomocysteinase (no. 38, 39), methionine synthase (no. 42) and S-adenosylmethionine synthetase (no. 40) where methionine is synthesized, and homocysteine is converted into S-adenosylmethionine. Bourgis et al (1999) reported that S-adenosylmethionine synthetase as a major phloem sap component in many plants and recently Rodriguez-Medina et al (2011) suggested this protein supplies methyl group in the SEs and may be involved in post-translational modification. In addition, identification of glutamine synthetase (no. 67) and adenosine kinase (no. 50, 51) proved that more enzymes are involved in amino acid metabolism. Glutamine synthetase can produce glutamine from glutamate and ATP indicating that once loaded into SEs amino acids can be converted into bioactive compounds or transport anino acids. Meanwhile, adenosine monophosphates required for certain metabolic processes are provided by adenosine kinase.

3.4.5.4 Component of Cell Structure

Further common proteins found in the phloem proteome of other plant species are actin (no. 104), actin depolymerising factor (no. 105) and profilins (no. 107-109) (Giavalisco et al., 2006; Aki et al., 2008; Lin et al., 2009; Rodriguez-Medina et al., 2011). In castor bean, actin and profilin are phloem mobile and constantly delivered into SEs (Schobert et al., 2000). Actin polymerisation is regulated by profilin and inhibition could occur when microfilaments are prevented from forming by elevated profilin level (Schobert et al., 2000). Actin depolymerisation factor has the same function as profilin in which it influences the organisation state of actin and binding of phospholipids.

3.4.5.5 Protein Transport

Proteins with chaperone activity such as found in this study (no. 127) might have a role in cell-to-cell protein trafficking. Other proteins with such activity include heat shock protein (no. 132) and thylakoid luminal protein (no.86, 174, 175) though luminal proteins are better known for their function in electron transport (Peltier et al., 2000). In addition, 26S proteasome subunit (no.138) and rubisco activase (no. 78) could also be involved in protein assembly. They are classified as ATPases or AAA proteins that prevent aggregation of denatured protein by their chaperone activity (Ogura and Wilkinson, 2001). In castor bean, several chaperones identified from the phloem exudate (Schobert et al., 1995) are supposedly involved in symplastic movement from CCs to SEs by unfolding the proteins.

3.4.5.6 Nucleic Acid Binding

Nomata et al (2004) reported that some members of RNA binding proteins may have an effect on response to stress as mRNA levels changes in accordance to various stress responses. Two glycine rich binding proteins (GRP) (no. 150, 151) were identified in the present study and from previous phloem proteome studies in cucurbit, brassica, rice and lupin (Xoconostle-Cazares et al., 1999; Lin et al., 2009; Giavalisco et al., 2006; Aki et al., 2008; Rodriguez-Medina et al., 2011).

Interestingly, AtGRP7 has been documented to promote floral transition in *Arabidopsis* (Streiner et al., 2008).

3.4.5.7 Signalling Proteins

In eukaryotes, the 14-3-3 proteins or general regulatory factor proteins are known for their role in signalling pathways. They are potent regulators of signal throughput by affecting the activity of partner proteins (Paul et al., 2008). At least 12 isoforms have been reported in *Arabidopsis*. Four 14-3-3 proteins were identified from the present study (no. 120-123) and they are known components of plant hormone signalling pathways (Purwestri et al., 2009). Interactions of these proteins with *Arabidopsis* FT had been reported in tomato based on yeast two hybrid study (Pneuli et al., 1998) and 14-3-3 proteins were proposed to play a role downstream of flowering regulation pathway (Taoka et al., 2011). Furthermore, interaction between Hd3a, the rice ortholog of FT with 14-3-3 proteins was observed in the cells of the shoot apical implying 14-3-3 proteins play a role as intracellular receptor for florigen in rice. Additionally, late flowering was observed in rice from overexpression of GF14c, one of 14-3-3 proteins (Purwestri et al., 2009). Similar observation was also documented from tomato (Pneuli et al., 1998) indicating negative regulation of flowering. However in *Arabidopsis*, a slight delay in flowering is observed from loss of functions of 14-3-3 mu and 14-3-3 upsilon (Mayfield et al., 2007). In addition, Mayfield et al also reported interaction among CO and 14-3-3 mu and 14-3-3 upsilon. It would be of interest to study one of the 14-3-3 proteins identified in the present study to determine its interaction with FT since it was found in the phloem proteome.

In addition, comparisons were also made against *Arabidopsis* phloem mRNA based on study by Deeken et al (2008) (Appendix A) utilising phloem obtained using a technique referred to as laser microdissection coupled to laser pressure catapulting (LMPC) and phloem leaf exudate collection which later compared it to CC-EST collection by Ivashikina et al (2003). It was found that 46.2% of proteins from current study were also found in with Deekan et al with the highest percentage of homologs coming from exudates followed by LMPC. The presence of mRNAs has been reported in the sieve element (Sasaki et al., 1998; Doering-Saad et al., 2002;

Xoconostle-Cazares et al., 1999). The bulk flow of solution dragged from companion cell to sieve element may be the cause and mRNA may also be involved macromolecular trafficking as part of the long-distance transfer of information (Doering-Saad et al., 2006). The expression of gene(s) in the companion cell and long-distance movement of mRNA have been documented through grafted tissue (Ruiz-Medrano et al., 1999; Huang et al., 2005). Therefore, comparison of proteins from present study with mRNA of the phloem may be associated with transcript function (Asano et al., 2002; Nakazono et al., 2003; Vilaine et al., 2003) to accommodate information regarding gene expression in the companion cell. From the phloem of various plants, discovery of a range of RNA binding proteins have been documented (Xoconostle-Ca'zares et al., 1999; Owens et al., 2001; Yoo et al., 2004; Gomez et al., 2005), giving evidence of involvement of phloem proteins in RNA trafficking, including possible role of phloem proteins in RNA-based systemic signalling mechanisms (Aoki et al., 2005).

3.4.6 FT Was Detected Using MRM But Was Not Found Using Standard LC-MS/MS

As shown in the list of proteins (Table 3.1) identified via standard LC-MS/MS from both excised gel bands and total phloem protein, FT protein was not detected. This finding was in contrast to phloem proteome study in *Brassica napus* (Giavalisco et al., 2006) where FT and TSF homologs were identified by partial sequence determination by MS/MS from 2D spots. However, it was detected using MRM. MRM proved to be a better method for detection of specific protein as shown in this study due to increased sensitivity of detection of the targeted peptides selected , and this was aided by availability of recombinant standard FT protein for method development.

As the name implied, MRM allows detection of multiple peptides from specific proteins by selecting particular MS/MS signals within a sample. Using specific peptide for that protein, the peak or peaks for that protein is easily identified from highly complex samples. This is in contrast to IDA method where any and every peptide is taken into account. The downside of IDA is that signals from low

abundance proteins are often masked by peaks of other higher abundance proteins. The MS cycle time further limits detection of closely eluting peptides. Overall, detection of very low abundance proteins may still be hindered by current instrumentation. Baginsky and Gruissem (2006) cited some important contributing factors of failing to detect certain proteins or peptides from complex protein mixtures which include unusual peptide structure, unexpected post-translational modifications and the effect of 'undersampling'. Development of a statistical model suggested that to reach 95% coverage of all possible detectable peptides, as many as 10 independent tandem MS runs are needed (Liu et al., 2004). The same paper reported new peptide identification with each additional run conducted under the same conditions. Similar result was also found in present study in which the number of identified proteins varied with each repeated run of the same sample. Even for proteins with the same AGI number, sometimes new peptide not found previously was identified. The 'undersampling' effect can be minimised by using the new generation of ion trap MS/MS instruments. By increasing the scan speed to shorten the duty cycle of each scan, the scan rate is increased, permitting detection of low abundance peptides (Baginsky and Gruissem, 2006).

The most likely explanation for failing to detect FT protein by full scan MS-MS is its low abundance. Detection of its unique peptides might be overlapped with other higher abundance sequence (Lin et al., 2009). The same problem was also encountered by Lin et al, being unable to detect the CmFTL1 peptide due to its low abundance

3.4.7 Quantitative Analysis of FT in Relation to Photoperiod is Possible via MRM

The study of FT quantification in this study was established using MRM. Three peptides were used as internal standard. In the initial set up of MRM for detection of FT in *Arabidopsis*, the peptide VEIGGEDLR was found to give the best representation of optimum peak area count. However as time moved on, peptide VVGDVDLPFNR was found to be better as it was identified in more samples during quantification of FT in response to photoperiod. More importantly, in the standard it

was identified as having the highest peak area count as compared to the other two.The presence of Pro/P in the peptide sequence is a good indicator as proline has been found to increase EPI in MS. Furthermore the sequence contains no Cys/C, Met/M or Trp/W. Whenever possible; these three amino acids are avoided during peptide selection for MRM along with peptides that had modifications (Drabovich and Diamandis, 2010). Side reactions can occur when Cys, Met or Trp is used as the moeties of the amino acids could result in racemisation, oxidation and alkylation (Siedler et al., 1996; Huang and Rabenstein, 1999;Giraud et al., 1999; Jha, 2009). The third peptide sequence contained Pro and Trp making it the least reliable peptide for standard, though FT was identified from this peptide including three of the samples with higher peak areas than the other two peptides (Table 3.6). Between the other two, only the main peptide contained Pro. Fragmentation at Pro bond has been found to be predictable and if incorporated into peptide sequence algorithm it may be used to improve protein identification (Breci et al., 2003) hence the choice in this study.

Peptide selection is crucial to ensure the success of MRM experiment as tens to hundreds of peptides were produced by tryptic digestion from each targeted protein (Picotti et al., 2007). To determine its quantity only a few representative peptides are targeted per protein to ascertain the presence of a protein in a sample (Lange et al., 2008) hence the careful choice. In the current study, the full FT peptide sequence obtained from UniProt (http://www.uniprot.org) was run through MIDAS workflow and 14 peptides sequence with the best potential for detection were obtained. All given peptides were run through ESP-Predictor from GPM (Global Proteome Machine, http://www.thegpm.org). Fragmentation of peptides was also searched through GPM. As a comparison, EPI was also run on the peptides to determine fragmentation of peptides. Results from ESP-Predictor GPM fragmentation and EPI were compared and peptides with high ESP-Predictor value, lower basic value and have good fragmentation especially from EPI were chosen. From the 14 peptides, three fit the criteria and were chosen for FT measurements. Two of the peptides in the present study VEIGGEDLR and VVGDVDLPFNR are very similar to a peptide used in cucurbit to determine the presence of pumpkin FTL2 protein which is an FT homolog in cucurbit (Lin et al., 2007). SRM (single reaction monitoring) was applied

to investigate the movement of Cm-FTL2 in the scion of *Cucurbita moschata* from *Cucurbita maxima* stock as detected via VEIGGTDLR and VIGDVLDSFTK.

The present study was based on Corbesier et al (2007) using the same time point (3 d of LD induction) and we expect a varying level of FT protein as not all protein is translated almost immediately after mRNA is produced. This is based on finding that although the start of *CO* expression is marked approximately 12 hours after dawn where it remains high until close to next dawn, nuclear CO protein stabilises in the evening (Valverde et al., 2004). As reported in section 3.3.8 and represented by Table 3.6, level of FT quantified did not exactly match with those found on *FT* mRNA study (Corbesier et al., 2007). The highest level of FT was identified in the first time point when plants have been exposed to LD for only 2 days (60h), based on peak area. Among the 12 samples, one sample did not meet the criteria for detection of FT which was the third sample collected after 68h as it only had one peptide with peak area count of >200.

Although FT was detected in 11 out of the 12 samples subjected to photoperiod induction, varying level of FT protein observed was not considered sufficient to support firm conclusions on the actual production and accumulation of the protein. The experiment was only performed once and the concentration of proteins obtained from each time point was not consistent throughout the duration of collection. In addition, towards the end of the experiment some of the plants were found to be stressed probably due to growth room conditions and this may have affected the protein quality and quantity. A better approach to quantification is the iTRAQ method rather than MRM. iTRAQ modifies primary amino groups using isobaric tags (Ross et al., 2004; Wiese et al., 2007) that appear as single peaks in MS scans, providing relative quantitative protein information (Wiese et al., 2007). In addition, up to eight different tags are available for comparative analysis.

3.4.8 Reproducibility and Variation in Proteomics Data

Analysis of mass spectra obtained from LC-MS/MS of IDA and total protein across 5 genotypes in this study was done with at three independent repetitions. It is

not always easy analysing a large set of data done over different times, and in proteomics studies the final identified protein list was the most commonly reported (Berg et al., 2006). This is due to some issues that arise from collected tandem mass spectra causing complex or contradictory results. in the present study, the final identified proteins as outlined in Table 3.1, Appendix A and B were combined across more than three repetitions of MS from various samples obtained from different genotypes, and using different collection and purification techniques.

Mass spectrometer measures peptide masses and intensities. Any changes in peptide signal intensity in different experiments could result in variation of collected tandem mass spectra. Some spectra detected may be present in one experiment but not in others. In addition, the quality of each experiment differs, for example unexpected technical error may occur in one experiment but not the other and as a result certain protein may be identified in one experiment but not the other (Berg et al., 2006). Similar result was observed in the current study as the number of proteins identified between repetitions differs. Most major protein that are highly abundant and with high quality spectra are easily identifiable across all repetitions but the same could not be said about the rest of the proteins especially those present at low abundance.

Reproducibility in terms of identified major proteins was not much of a problem in IDA analysis. A majority of the proteins are reproducible across the repetitions though some of the proteins were only identified only once or twice (Appendix A). Reproducibility of the same peptide and spectra was not easy as a lot of parameters were involved especially during detection by MS. Sometimes the number of peptides assigned to certain proteins was reduced in one of the repetition or a longer peptide was obtained probably due to missed cleavage. In this study, all data obtained from all the repetitions were compiled and the best hit for each protein was chosen to represent the whole work.

However the same could not be said about 2D proteomics in this study. Reproducibility in terms of pattern of gel spots was the hardest. It was almost impossible to get the same spot pattern in two different gels even when run at the same time as shown in Fig. 4.8. Although phloem sap proteins were known to be

more basic in nature and the pattern on the gel showed spots accumulating towards the basic range after IEF (isoelectric focusing) using IPG strip pH 3 to 10, numbers of spots were not consistent each time. In addition, some of the spots were detected in one gel but not in another. Problems about 2D reproducibility have not been reported much due to increase in technology with the use of commercialised IPG strip of various range and better IEF equipment. Nevertheless in this study despite the identification of a few proteins from this method, 2D proteomics failed to be reproducible and relatively little data was obtained. Given enough time to further optimise IEF, this method may provide a better tool in proteomics work. Nevertheless, among the three sample types, total phloem protein remains the best sample for proteomics analysis.

CHAPTER 4
EXPRESSION OF FT FUSION PROTEIN USING DENDRA2

4.1 Introduction

Most studies relating to visualisation of protein expression in plants have made use of GFP (green fluorescent protein). Nowadays, there is a wider range of commercialised fluorescent proteins available. These fluorescent proteins were mostly derived from GFP and YFP (yellow fluorescent protein) but new features were added to enhance visualisation and detection. Among them is Dendra2 (Gurskaya et al., 2006), a newer version of Dendra which is a photoactivatable fluorescent protein (PAFP) with convertible detection at two different light emissions of 507 and 573 nm that emit green and red color respectively. The aim of this study was to transform FT fusion protein into wild type and late flowering *ft-10* mutant to study the expression of FT and its movement from the phloem through to the shoot apical meristem. The use of PAFP (photoactivatable fluorescent protein) such as Dendra2 that displays photoswitching ability with faster maturation and brighter fluorescence would allow for better detection of movement as compared to GFP. Along with its monomeric structure, Dendra2 provides high contrast photoconversion with fluorescence at the red spectral region, low phototoxic activation with 488 nm light and high photostability of the photoconverted state (Gurskaya et al., 2006; Zhang et al., 2007; Chudakov et al., 2007). For this study two promoters; SUC2 and GAS1 were used to determine FT expression. *Arabidopsis* SUCROSE TRANSPORTER2 (SUC2) acted as the main promoter in this experiment due to its active role in phloem companion cells (CCs) of both major and minor leaf veins (Imlau et al., 1999). On the other hand, Cucurbit GALACTINOL SYNTHASE1 (GAS1) was found to have an active function only in the phloem CCs of minor leaf veins (Haritatos et al., 2000). By fusing Dendra2 to FT cDNA carried by two different promoters that function in the phloem; one of which (SUC2) has been used in previous studies to detect FT expression (An et al., 2004; Corbesier et al., 2007; Notaguchi et al., 2008) the constructs were delivered into plants via floral-dip method (Clough and Bent, 1998) and selected to determine positive plants carrying the fusion protein genes. Transgenic plants were grown until T_2 generation before the transcript expression and morphology through phenotypic

observation were determined. Fluorescence microscopy was employed initially to check for any expression activity using SUC2::FT:GFP plants as positive control before moving on to confocal microscopy.

4.2 Materials and Methods

4.2.1. Fusion Protein Construction

4.2.1.1 Preparation of Promoters and Fluorescent Protein

AtSUC2::FT and AtGAS1::FT clones obtained from Germany (Prof G. Coupland, Max Planck Institute, Cologne) were used as the former contained SUC2 promoter specific to the phloem while the latter acts as a comparison against SUC2 promoter that functions mainly in the minor vein phloem of leaves. Five constructs were created for this purpose, all of which were fused to Dendra2 fluorescent protein (Evrogen). They are AtSUC2::FT::Dendra2, AtGAS1::FT::Dendra2, a true negative control from the vector of one of the clone fused to Dendra2 (without the promoter and FT) and, two FT-negative controls carrying only the promoters without the FT gene (AtSUC2::Dendra2 and AtGAS1::Dendra2).

Plasmids were extracted out from both clones and Dendra2 using Sigma GenElute® Plasmid Isolation Kit as per manufacturer's recommendation. Double-digestion was performed on plasmids from both clones using *Sac*I and *Stu*I as working construct, *Hind*III and *Sac*I as negative controls and AtSUC2::FT was cut with *Eco*RV and *Sac*I to create the true-negative control. The mixtures were incubated at 37^{o}C for 30min and run on 1% agarose gel in 1X TAE buffer for 60min at 50V. Bands corresponding to the expected size of each clone were cut out using scalpel and purified using SV PCR and Gel Purification Kit (Promega) as per manufacturer's recommendation.

For Dendra2, PCR amplification was performed using a mixture of 10 µL GoTaq buffer (Promega), 1 µL 10 mM dNTP, 0.25 µL *Taq* polymerase, 1 µL Forward Primer (10 mM), 1 µL Reverse Primer (10 mM), 1 µL DNA and 34.75 µL dH_2O to make up a 50 µL reaction. The PCR reaction began with initial denaturation

at 94°C for 5 min followed by 30 to 40 cycles of 94°C for 30 seconds, 59°C for 1 min (annealing temperature varies according to primer pair) and 72°C for 1 min followed by a final extension of 72°C for 10 min. All PCR products were run on 1% agarose gel (1% (w/v) agarose) in 1X TAE buffer (40mM Tris-Base, 40mM glacial acetic acid, 1mM EDTA) for 45 min at 80V. PCR was performed to amplify out only the Dendra2 region which is about 800bp in size (Fig.4.1). Once purified, the PCR product was amplified again using primer adaptors carrying restriction enzymes that complement the clones (Table 4.1). The amplified Dendra2 carrying matching restriction enzyme sites were then purified as previously and sequenced for verification

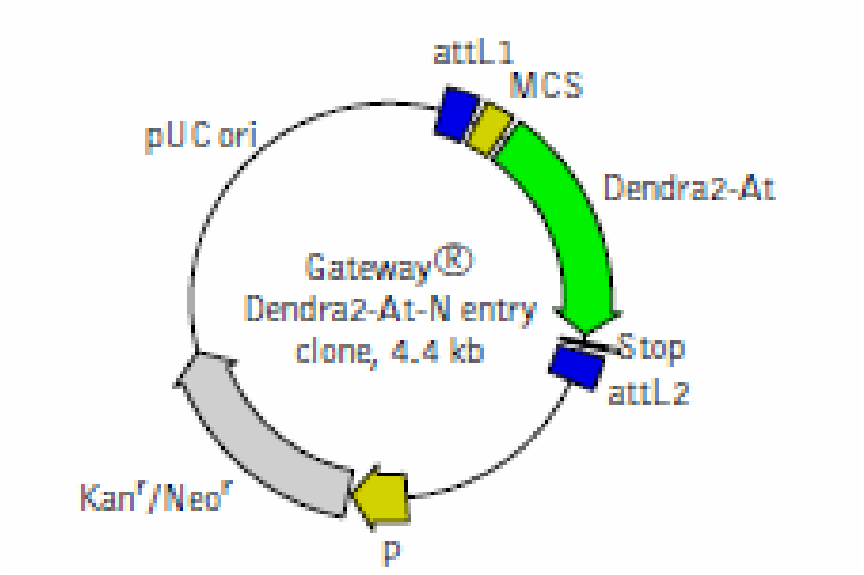


Figure 4.1: Map of Gateway® Dendra2-At-N used in the study.

Table 4.1: List of adaptor-primer used for construction of fusion protein.

	Long Primer	Short Primer
SacI-F	CAAGCTTCGAATTCTGCAGTCGACGGTACCG	ACTGAGCTCCTG
StuI-R	GATATCCACCTGAGTCTCCAGACCAAACC	TCTAGATATCAGT
HindIII-F	AAGCTTCGAATTCTGCAGTCGACGGTACCGCGGG	ACTAAGCCTCTG
SacI-R	GAGCTCACACCTGAGTCTCCAGACCAAACC	TCTGAGCTCCAGT
EcoRV-F	GATATCAAGCTTCGAATTCTGCAGTCGACGGTACC	ACTGATATCCTG

4.2.1.2 Preparation of GV3101 Electrocompetent Cell

From glycerol stock, GV3101 *Agrobacterium* cell was streaked onto LB plates (10% (w/v) tryptone, 5% (w/v) yeast extract and 10% (w/v) NaCl, 1.5% (w/v) agar) supplemented with 50µg/mL each of Rif and Gen and 2 days at 28°C. A colony was picked and grown in 5mL liquid LB supplemented with the same antibiotics at 28°C overnight with vigorous shaking at 180 rpm. The overnight culture was then transferred to 500mL LB (1/10 dilution) containing the exact antibiotics and let grown for up to 8h at 28°C at 180 rpm until OD_{600} reached 0.5 to 0.7. Cells were pelleted by centrifugation at 4°C for 20min at 4,000 rpm. Supernatant was discarded and the pellet resuspended in 10% (v/v) ice-cold glycerol. Wash the pellet by spinning it at 4,000 rpm at 4°C for 10min and the washing step was repeated twice. A final wash in the same solution for 10 min at the same spin and temperature was done followed by resuspension of the pellet in 10% ice-cold glycerol to a final OD_{600} of 70. 30µL of the cell was aliquoted into clean 1.5mL tube and the cell was snap-frozen in liquid nitrogen prior to storing at -80°C or use.

4.2.1.3 Ligation, Transformation, Selection and Confirmation via PCR

Double-digested products of all 5 constructs were ligated to Dendra2 in a 1:5 ratio using T4 DNA ligase (Promega) as recommended by manufacturer and incubated overnight at 16°C (16-18h). Electroporation was carried out the next day by mixing 5 to 7µL of ligated product into GV3101 electrocompetent cells and the mixture electroporated using Micropulser Electroporator (Bio-Rad) utilising manufacturer's instruction. Immediately after electroporation, 500µL of LB was added and the mixture was incubated for 2 to 4h at 28°C with gentle agitation. At the end of incubation period, the cells were pelleted by centrifugation at 13,000 rpm for 1min at RT and 200µL of the supernatant was spread onto LB plate supplemented with 50µg/mL each of Rif, Gen and Kan. The plate was incubated at 28°C for 2 to 3 days and up to 10 colonies were then picked to be streaked onto fresh LB plate containing the same antibiotics as before. PCR was performed on the colony to verify the insertion of Dendra2 into the promoters. Those positively identified to carry the Dendra2 insertion were analysed by PCR again with the promoters and universal

primers to ascertain correct orientation. Bands from positive transformants were purified and sent for sequencing for further confirmation.

4.2.2 Transformation via Floral-Dip

Wild type plants and *ft-10* mutants were grown until flowers started to develop. The first inflorescence from the plants was cut off to induce branching. Flowers need to be at a right stage for dipping where a few flowers with a lot of unopened buds are good candidates.

From glycerol stocks, all constructs were streaked onto individual plates of LB containing 50µg/mL each of Gen and Kan and incubated overnight at 28^{o}C. From the plate, single colony was chosen to start a 5 mL overnight culture at 28^{o}C in liquid LB supplemented with the same selection of antibiotics as before. The overnight starter culture was then added to 500 mL of LB and let grown for 18 to 24h at 28^{o}C. Cells were harvested via centrifugation at 4,000 rpm for 20min at RT. Following that, the pellets were resuspended in 5% (w/v) sucrose at a final OD_{600} of 0.8. Just before dipping, 0.05% Silwet L-77 was added to the mixture. Floral tissues were then immersed into the solution for 30s to 1min and the beaker containing the solution is rotated to prevent precipitation of *Agrobacterium* cells. A bag is used to cover the plants after that to maintain humidity where the plants were left at room temperature overnight. The following day, the cover was removed and the plants transferred to LD for further growth and seed collection.

4.2.3 Selection of Transformants using Basta

Seeds from transformed plants were grown on soil and were treated with Basta via spraying at a concentration of 30 mg/L after 5 days. Subsequent spraying was done with 4 days interval for a total of 4 times. Healthy surviving plants with green leaves were transferred into individual pot for further growth. In the meantime, genomic DNA extraction was carried out on young leaves of surviving plants and PCR was performed to determine the presence of transgenes. A section of the leaf and a portion of the root tip were also taken and viewed under fluorescent microscope to

detect Dendra2. Plants that were identified positive via PCR were selected for further growth through to seed.

4.2.4 Verification of Fusion Protein via PCR

4.2.4.1 Genomic DNA Extraction

Young leaves from T_1 and T_2 plants were used for genomic DNA extraction. 50 to 100 mg of young leaves were ground to a fine powder in mortar and pestle using liquid nitrogen and transferred to a 1.5 mL tube. 400 µL of extraction buffer (200 mM Tris-HCl pH 7.5, 25mM EDTA, 250 mM NaCl, 0.5 % (w/v) SDS) was added to the tube and the mixture was vortexed rigorously for about 10s. Centrifugation at 13,000 rpm proceeded for 1 min. Approximately 300 µL of supernatant was transferred to a fresh tube where 300 µL cold isopropanol was added and the mixture mixed thoroughly. The mixture was spun at 13,000 rpm for 5 min and this time the supernatant was discarded. The pellet was added with 1 mL of 70% (v/v) cold ethanol and the sample was spun at 13,000 rpm for 1 min. The supernatant was discarded and the sample is briefly spun at 13,000 rpm for approximately 10s. Residual supernatant was removed by pipetting and the pellet dried using a vacuum drier for 2 to 5 min. 100 µL TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA) was added to redissolve the pellet by gentle pipetting. The sample was then stored at -20^{o}C prior to use.

4.2.4.2 RNA Extraction

Young leaves from T_2 plants were used for this step. RNA extraction was performed using TRI Reagent (Sigma). 50-100 mg tissue was frozen and ground to a fine powder. 1 mL of Tri Reagent was added and the sample was let stand for 5 min at room temperature followed by addition of 0.2 mL chloroform and vigorous shaking for 15 s before letting it stand at room temperature for 15 min. The mixture was then centrifuged at 12,000 x *g* for 15 min at 4^{o}C. Three phases separated the mixture after centrifugation comprising of upper aqueous phase (containing RNA), an interphase (Containing DNA) and a red organic phase at the bottom (containing protein). The aqueous phase was transferred to a fresh tube and 0.5 mL of isopropanol was added and mixed. The mixture was allowed to stand for 10 min at room temperature before

centrifuging it at 4°C for 10 min at 12,000 x *g*. The supernatant was removed and the RNA which was precipitated into a pellet form was washed with 1 mL of 75% (v/v) cold ethanol. The sample was vortexed briefly and centrifuged at 7,500 x *g* for 5 min at 4°C. After discarding the supernatant, the RNA pellet was air-dried at room temperature for 5-10min but not to a complete dry. The RNA pellet was dissolved in an appropriate amount (about 40-50 µL) of TE buffer (10mM Tris-Base pH8.0, 1mM EDTA) by gentle pipetting. The RNA was then stored at -80°C prior to use.

4.2.4.3 Polymerase Chain Reaction (PCR) and Reverse-Transciptase PCR (rt-PCR)

PCR amplification was performed using a mixture of 10 µL GoTaq buffer (Promega), 1 µL 10 mM dNTP, 0.25 µL *Taq* polymerase, 1 µL Forward Primer (10 mM), 1 µL Reverse Primer (10 mM), 1 µL DNA and 34.75 µL dH_2O to make up a 50 µL reaction. The PCR reaction began with initial denaturation at 94°C for 5 min followed by 30 to 40 cycles of 94°C for 30 seconds, 53°C for 1 min (annealing temperature varies according to primer pair) and 72°C for 1 min followed by a final extension of 72°C for 10 min. All PCR products were run on 1% agarose gel (1% (w/v) agarose) in 1X TAE buffer (40mM Tris-Base, 40mM glacial acetic acid, 1mM EDTA) for 45 min at 80V.

Reverse-transcriptase PCR was achieved by preparing a mixture containing 10 µL GoTaq Flexi buffer (Promega), 1 µL 10mM dNTP, 3 µL 25mM MgCL2, 0.5 µL Taq polymerase (5U/µL), 1 µL Forward Primer (10mM), 1 µL Reverse Primer (10mM), 2 µL cDNA and 33.5 µL dH_2O. PCR cycle condition was as followed; 94°C for 4 min, 40 cycles of 94°C for 30s, 55°C for 1 min (annealing temperature varies), 72°C for 1 min 20s and a final extension at 72°C for 7 min. Products were run for 45 min at 80V on 1% agarose gel.

4.2.5 Phenotypic Observation

Seeds from T_1 plants that were positive for carrying the constructs were used for growing in soil and placed in both LD and SD condition for phenotypic

observation. Seeds were grown on soil to vermiculite ratio of 4:1 and kept at 4°C for 2-3 days. Subsequent growth was done in growth chamber in SD condition at 23°C with 10h light (160µmol) and 60% humidity. Plants that need to be grown in LD condition were exposed to light for 16 h with the same intensity and humidity. Rosette leaf count of the T_2 plants was carried out to determine flowering time of each line.

4.2.6. Tracking Protein Movement of FT/Dendra2 Fusion Construct

Seeds from T_2 plants were sterilized in 75% (v/v) ethanol for 30s followed by 3% (v/v) Clorox for 10 to 15 min with occasional shaking. Next seeds were washed for 5x in ddH2O for 1 min each to rinse off excess Clorox. The seeds were dried on a sterilized filter paper before they were sowed onto ½ Murashige and Skoog culture plate (2.1 g/L Murashige and Skoog medium, basal salt mix by Duchefa Biochemie, with 0.7% (w/v) phytagel) supplemented with 1% (w/v) sucrose. Plates were stratified in the cold room in a dark condition for 2 days and later transferred to LD for further growth. Healthy plants with good root growth were checked under fluorescent microscope and those that showed positive result were chosen for confocal work.

4.3 Results

4.3.1 Construction of FT:Dendra2 Fusion Proteins

A total of five fusion proteins were constructed for this study. The fusion proteins were constructed using two original constructs obtained from collaborator (Germany), each containing FT cDNA in SUC2 and GAS1 promoter.

After successful digestion with appropriate restriction enzymes to cut out the GFP region from the original constructs, both constructs were fused to Dendra2 which was previously amplified out and added with adaptors to complement ligation process (Fig. 4.2 (a)). Sequencing was carried out to verify insertion of adaptors into Dendra2 and later on the full constructs to ensure in frame insertion. Vector control was created from one of the original construct by eliminating the promoter and FT regions so it carries only the vector followed by fusion to Dendra2 (Fig. 4.2 (c) and designated as Vector:Dendra2. Bacterial transformation into *Agrobacterium* cell

GV3101 proceeded and positive transformants were confirmed by PCR using primers corresponding to the promoters, FT cDNA and the Dendra2 regions. The constructs were then designated as SUC2::FT:Dendra2, GAS1::FT:Dendra2 and vector:Dendra2. Two extra constructs acting as FT negative control for the promoters were later made from SUC2::FT:Dendra2 and GAS1::FT:Dendra2 by cutting out the FT region (Fig. 4.2 (b) and the two new FT control constructs were designated SUC2::Dendra2 and GAS1::Dendra2 respectively.

Once fusion proteins were ready, a bulk stock was made for transformation into plants and glycerol stocks were made for long term storage.

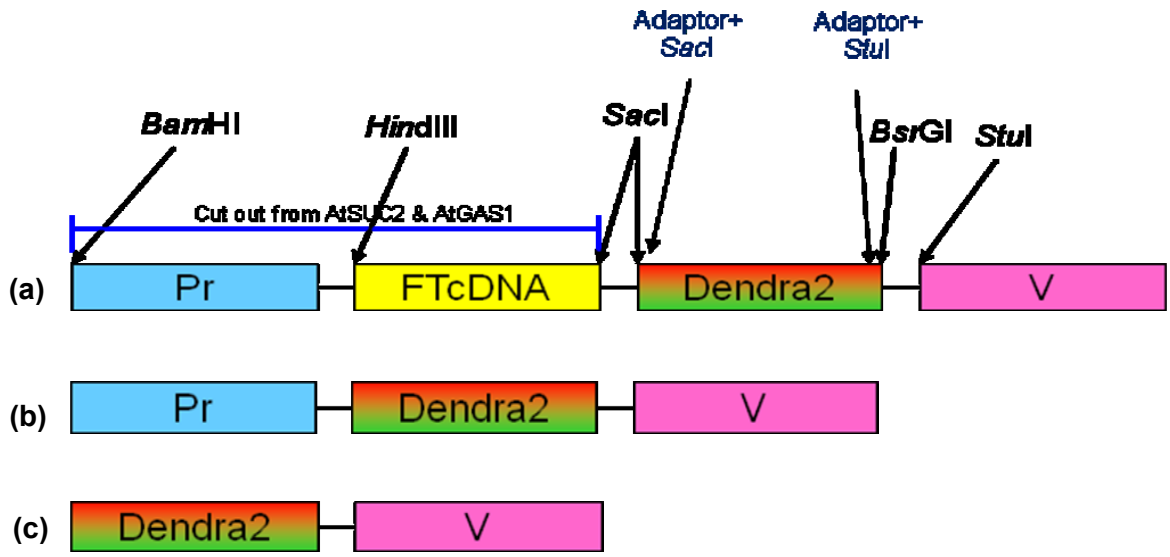


Figure 4.2: Construction of the five fusion proteins for FT accumulation study. (a) SUC2::FT:Dendra2 and GAS1::FT:Dendra2 were constructed by digesting original AtSUC2 and AtGAS1 with *Sac*I And *Stu*I followed by ligation with Dendra2 which previously was amplified our via PCR and attached with adaptors carrying *Sac*I And *Stu*I sites for complementation. (b) FT negative control for both promoters were constructed by eliminating FTcDNA region via double digestion and (c) true negative control was made by eliminating the promoter region to carry only the vector and Dendra2. Pr=promoter, V=vector

4.3.2 Transformation of Fusion Proteins via Floral Dip and Analysis of Fusion Proteins in *Arabidopsis* T_1 Plants

For transformation purposes, both wild type (Col-0) and mutant (*ft-10*) plants were grown in soil in LD condition until earlier stage of flowering was observed. Plants with lots of buds with unopened flowers were preferred for transformation and as such, *ft-10* plants were grown 2 weeks earlier than wild type since the mutant is a late flowering mutant. Preparation of fusion proteins for dipping was then made and transformation of the constructs into both wild type and *ft-10* plants proceeded.

Once transformation via floral dip took place, up to 6 weeks were needed to set the seeds from T_0 plants. Collected seeds were then subjected to Basta treatment to select for positive transformants carrying the fused protein. Plants were exposed to 4 Basta treatments and spraying was done at 4 days intervals. After the last treatment, plants that survive and were healthy with green leaves (Fig. 4.3) were chosen and transferred to individual pots for further growth. A week after that genomic DNA was extracted from young leaves of all surviving plants and PCR was carried out to determine if the plants carry the transgene.

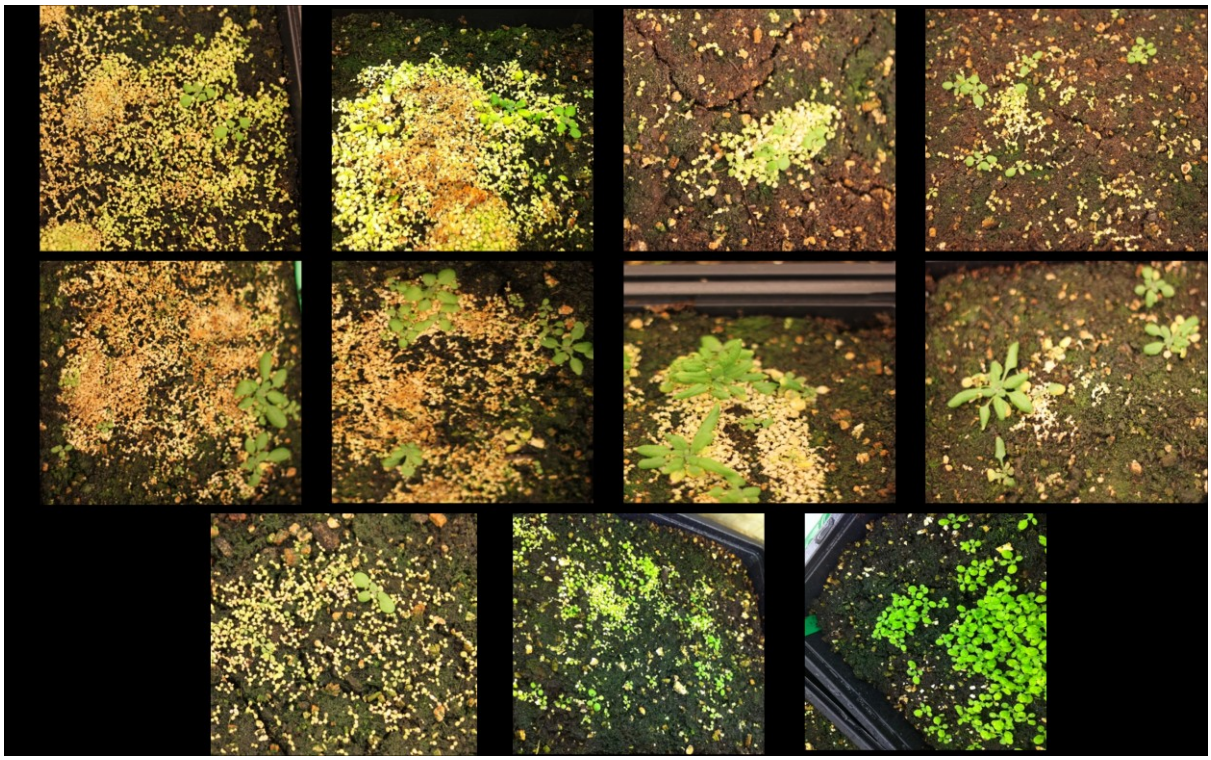


Figure 4.3: Basta-treated transformed plants after the last spray. Two days after last treatment was employed (plants were 3 weeks old), approximately 5 to 20% surviving plants were observed in (a) SUC2::FT:Dendra2 plants in WT ecotype, (b) SUC2::FT:Dendra2 plants in *ft-10*, (c) GAS1::FT:Dendra2 plants in WT and (d) GAS1::FT:Dendra2 plants in *ft-10*. Plants that grew healthily 2 weeks later (5 weeks old plants) such as (e) SUC2::FT:Dendra2 plants in WT , (f) SUC2::FT:Dendra2 plants in *ft-10*, (g) GAS1::FT:Dendra2 plants in WT and (h) GAS1::FT:Dendra2 plants in *ft-10* were chosen and transferred into individual pots for further growth. Control plants (i) WT (negative control), (j) ft-10 (negative control) and (k) SUC2::FT:GFP (positive control) were also exposed to equal treatment of Basta (all control plants were 3 weeks old).

Initially, 14 plants for each WT and mutant were used for transformation of SUC2::FT:Dendra2 and GAS1::FT:Dendra2, and 10 plants each ecotype for SUC2:Dendra2, GAS1:Dendra2 and vector:Dendra2. The number of plants used for T_2 experimentation decreased after genotyping as shown in Table 4.2. Seeds from these plants were divided into two for T_2 generation where half was used for growth in soil for phenotypic observation and the other half was grown on tissue culture plate for expression analysis. Figure 4.4 shows results of PCR that was performed on genomic DNA of surviving T_2 transformed plants using primers to confirm for Dendra2 (a), SUC2 promoter (b), GAS1 promoter (c) and FTcDNA regions (d).

SUC2::FT:Dendra2 and GAS1::FT:Dendra2 plants that were positive for all primer pairs, SUC2::Dendra2 and GAS1:Dendra2 plants positive for Dendra2 and promoter(s) primer pair(s), and Vector:Dendra2 plants positive for Dendra2 and vector primer pairs (Fig 5.3 (e)) were chosen for further growth. From PCR results, 6 plants positive for fusion protein construct SUC2::FT:Dendra2 in WT (designated as SFW), 5 for SUC2::FT:Dendra2 in ft-10 (SFf), 3 each for GAS1::FT:Dendra2 in WT and ft-10 (GFW and GFf), 2 each for Vector:Dendra2 in both ecotype (VW and Vf), 2 each for SUC2:Dendra2 as well as GAS1:Dendra2 in WT (SW and GW), and finally 1 each for SUC2:Dendra2 and GAS1:Dendra2 in ft-10 (Sf and Gf).

Table 4.2: Analysis of T_1 plants via PCR to determine incorporation of transgene. Plants for each construct were designated with numbers and numbers that are not listed in the table did not produce amplicon from PCR. Primer pairs used for confirming presence of SUC2 is specific for native SUC2, GAS1 is specific for cucurbit GAS1 and FT cDNA is specific to native FT cDNA.

Primer pair	**SUC2::FT: Dendra2**		**GAS1::FT: Dendra2**		**SUC2: Dendra2**		**GAS1: Dendra2**		**Vector: Dendra2**	
	WT	***ft-10***	**WT**	***ft-10***	**WT**	***ft-10***	**WT**	***ft-10***	**WT**	***ft-10***
Dendra 2	All plants	All plants	All plants	All plants	All plants	All plants	All plants	All plants	All plants	All plants
SUC2	SFW 1, 2, 3, 5, 6, 9	SFf 3, 5, 6, 7, 8			SW	Sf				
GAS1			GFW 1, 3, 4	GFf 1, 3, 4			GW	Gf		
FT cDNA	All plants	All plants	All plants	All plants						
Vector	All plants	All plants	All plants	All plants	All plants	All plants	All plants	All plants	All plants	All plants

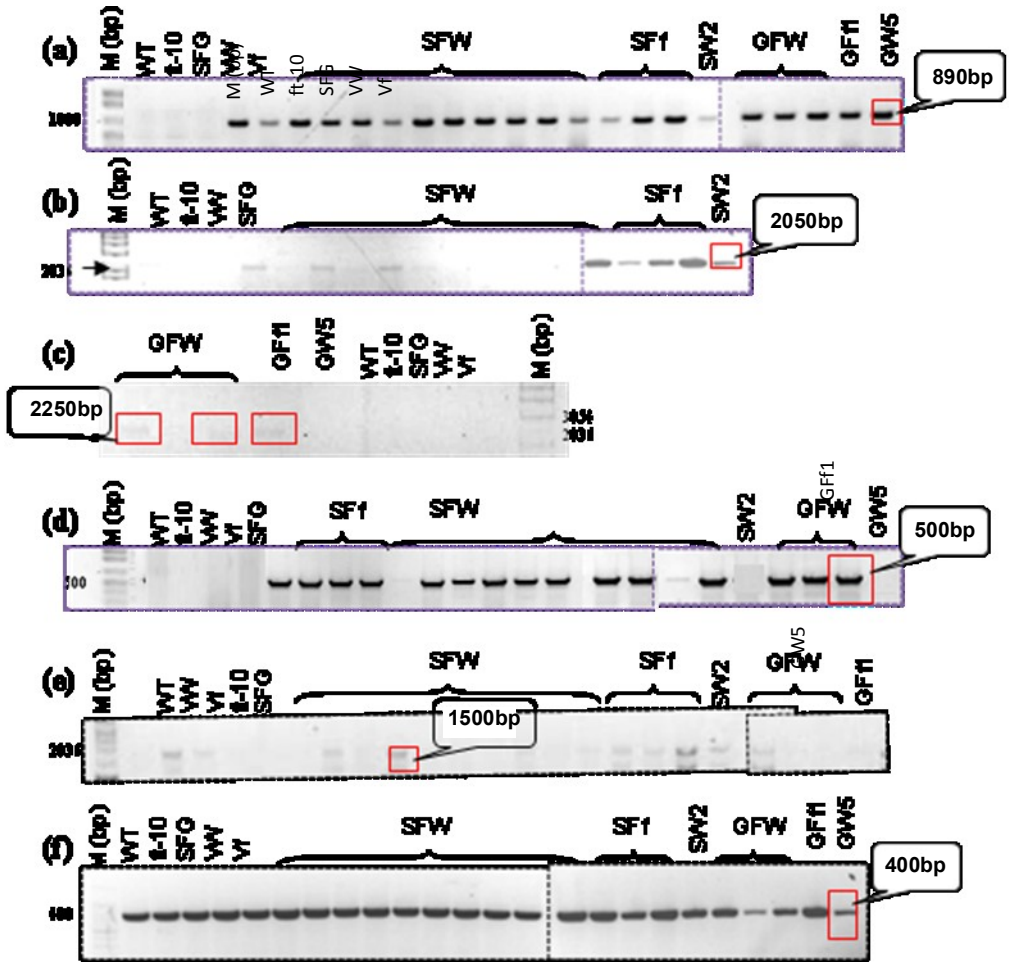


Figure 4.4: Presence of fused transgene were confirmed using PCR to amplify out (a) Dendra2 fluorescent protein, (b) SUC2 promoter, (c) GAS1 promoter, (d) FTcDNA region and (e) pGreen vector from genomic DNA of each lines and controls. Plants that had been successfully transformed with either one of the promoter should show a band with primer pairs of (a), (b or c), (d) and (e). Transformants were also tested against Actin primer as control to ascertain quality of genomic DNA (f). M=Marker, WT=Wild type, VW=vector:Dendra2 in WT background, Vf=Vector:Dendra2 in ft-10 background, SFG=SUC2::FT:GFP, SFW=SUC2::FT:Dendra2 in WT, SFf=SUC2::FT:Dendra2 in ft-10, SW=SUC2:Dendra2 in WT, GFW=GAS1::FT:Dendra2 in WT, GFf=GAS1::FT:Dendra2 in ft-10, GW=GAS1:Dendra2 in WT.

4.3.3 Phenotypic Observation of Plants Expressing Fusion Protein Genes

Phenotypic observation was carried out by planting seeds of T2 plants in soil and exposing them to LD and SD photoperiod. Flowering time was determined by rosette leaf count and compared against untransformed and negative control WT and mutant plants. From Fig. 4.5 almost all T2 plants from SUC2::FT:Dendra2 in the WT background flowered earlier or at about the same time as WT. Plants from GAS1::FT:Dendra2 flowered more or less about the same time as WT but not earlier. In the case of plants in mutant background none of the plants flowered later than *ft-10* from both SUC2::FT:Dendra2 and GAS1::FT:Dendra2 lines which was expected.

Apart from the standard calculations of standard deviation and SE (standard error), one way and two way ANOVA were also performed. Using the one way ANOVA flowering time in each LD and SD were found to be significantly different, $F(28, 289) = 96.30$, $p<0.0001$ and $F(28, 289) = 218.2$, $p<0.0001$ respectively (Fig. 4.5). A post-hoc test was also performed using Bonferroni's multiple comparison test where individual transgenic lines in each LD and SD were compared against the vector control plants in WT and *ft-10* background (Appendix E). As shown in Appendix E, there are variations in the significance level between each line against the vector control plants both in LD and SD. In a two-way ANOVA analysis, the main effect of photoperiod was $F(1, 522) = 40392.61$, $p<0.001$ giving an indication that all transgenic lines including the controls under LD flowered earlier than SD (Fig. 4.6). Another main effect is between different lines which is extremely significant at $F(28,522) = 286.19$, $p<0.001$ and it showed that lines in the WT background generally flowered earlier than those in the mutant background. Finally interaction between photoperiod and transgenic lines was extremely significant at $F(28, 522) = 50.97$, $p<0.001$.

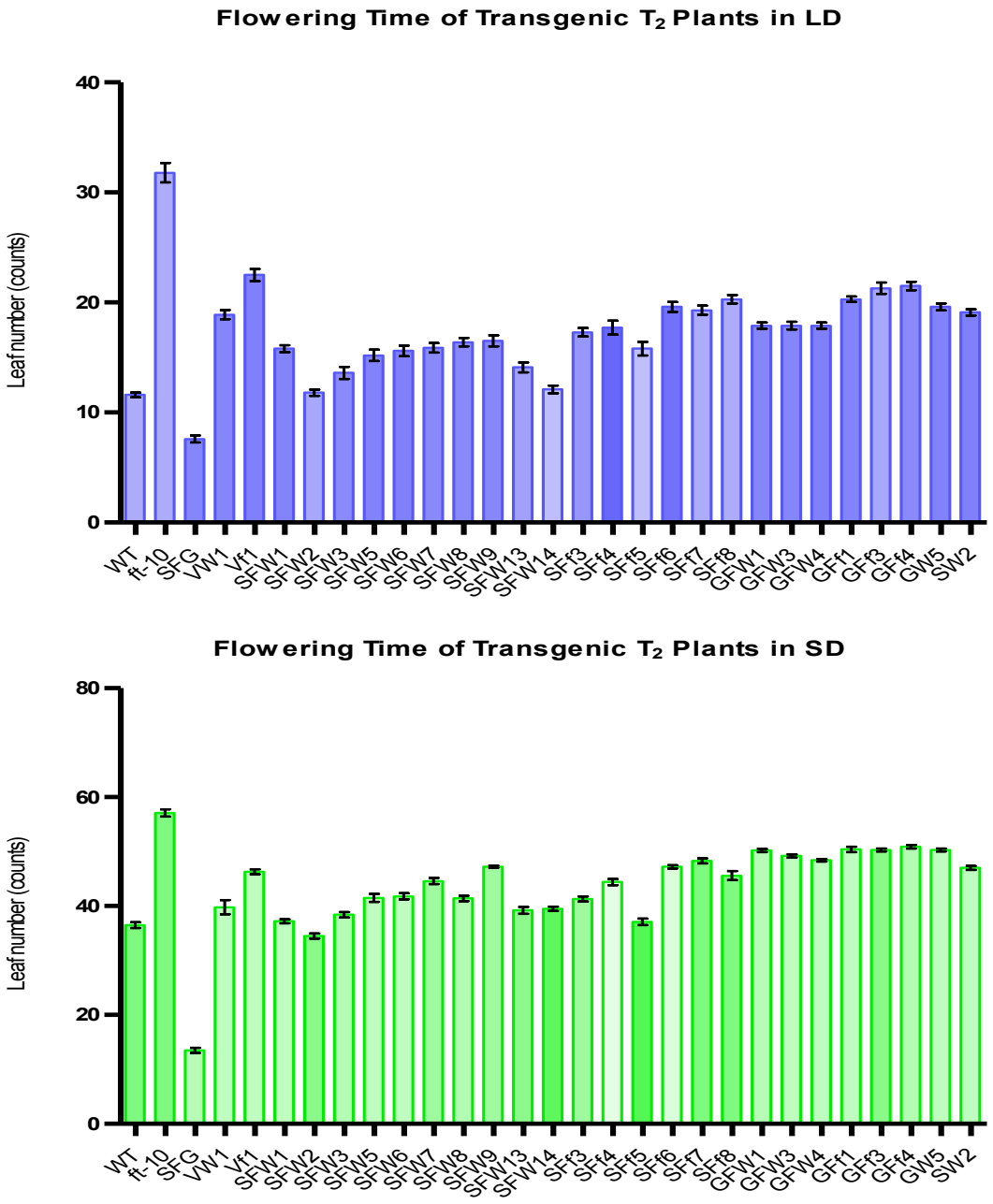


Figure 4.5: Flowering time of T_2 plants by rosette leaf count for plants both in LD (above) and SD (below). Each value represents mean ± SEM (n=10). One Way ANOVA was performed and post-test using Bonferroni's multiple comparison test is as shown in Table 5.3. LD=long day, SD=short day, SEM=standard error of the median, ANOVA=analysis of variance, WT=wild type, *ft-10=ft* null mutant, VW=vector:Dendra2 in WT, Vf= vector:Dendra2 in *ft* mutant, SFW= SUC2::FT:Dendra2 in WT, SFf= SUC2::FT:Dendra2 promoter in *ft* mutant, SW=SUC2:Dendra2 in WT, GFW= GAS1::FT:Dendra2 in WT, GFf= GAS1::FT:Dendra2 promoter in *ft* mutant, GW=GAS1:Dendra2 in WT. Graph with columns in blue colour represent transgenic plants grown in LD while green columns represent transgenic plants grown in SD.

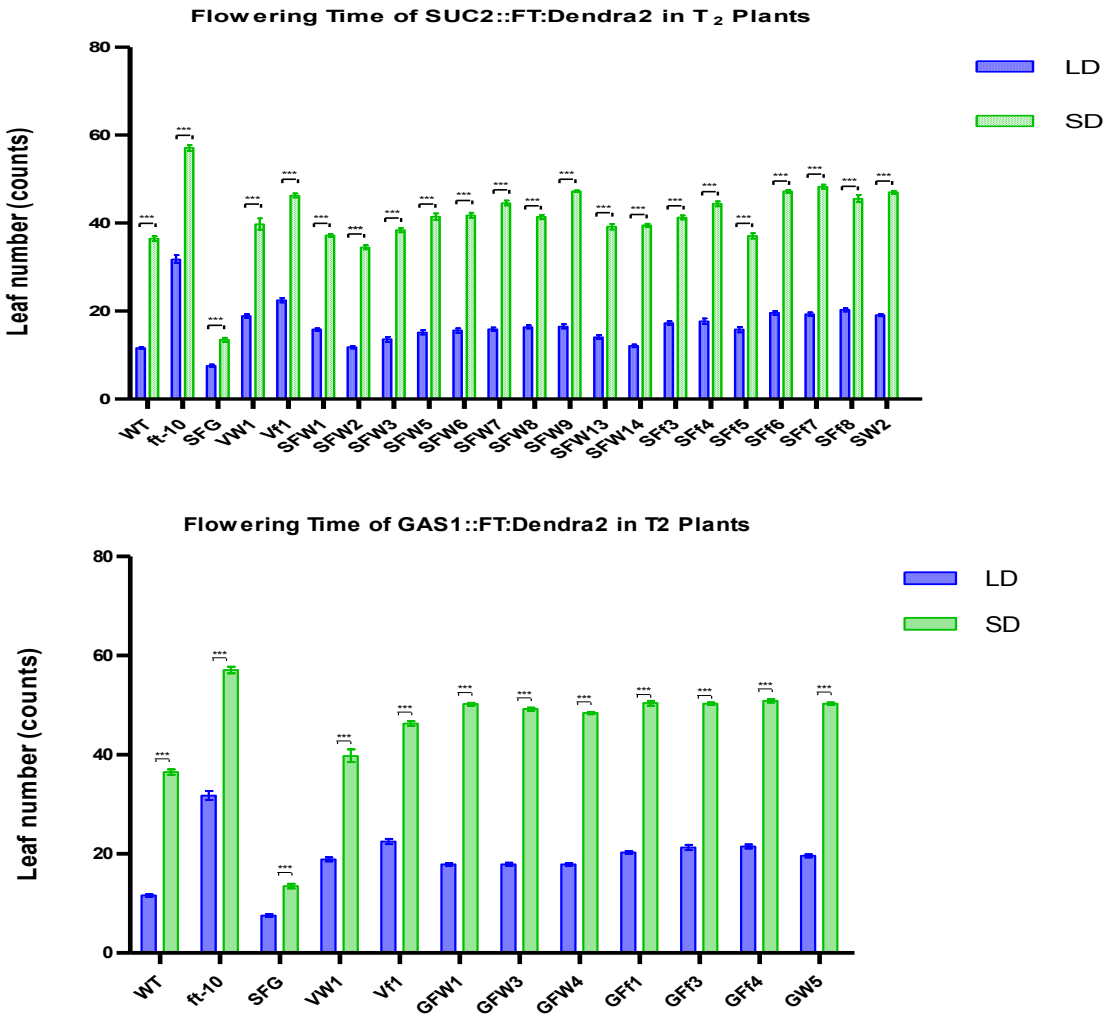


Figure 4.6: Flowering time of T_2 plants were determined by rosette leaf count for plants both in LD and SD for (a) SUC2::FT:Dendra2 line and (b) GAS1::FT:Dendra2 line. Each value represents mean ± SEM (n=10). Two Way ANOVA was performed with non repeated measurement. Significance was determined as *p<0.05, **p<0.01, ***<0.001.

LD=long day, SD=short day, SEM=standard error of the median, ANOVA=analysis of variance, WT=wild type, *ft-10*=*ft* null mutant, VW=vector:Dendra2 in WT, Vf= vector:Dendra2 in *ft* mutant, SFW= SUC2::FT:Dendra2 in WT, SFf= SUC2::FT:Dendra2 promoter in *ft* mutant, SW=SUC2:Dendra2 in WT, GFW= GAS1::FT:Dendra2 in WT, GFf= GAS1::FT:Dendra2 promoter in *ft* mutant, GW=GAS1:Dendra2 in WT.

4.3.4 Expression Analysis of SUC2::FT:Dendra2 in Wild Type and *ft-10* Plants

Plants from T_2 generation were used to further validate fusion protein construct integration into wild type and *ft-10* mutant plants. PCR was carried out on genomic DNA extracted from these plants showed that not all of the selected plants from T_2 carried the complete construct (Fig. 4.7). Only those confirmed to be carrying the full construct were chosen for expression analysis via reverse transcription PCR (rt-PCR) to determine transcript levels. Design of the primers for the promoters is as shown in Figure 4.8.

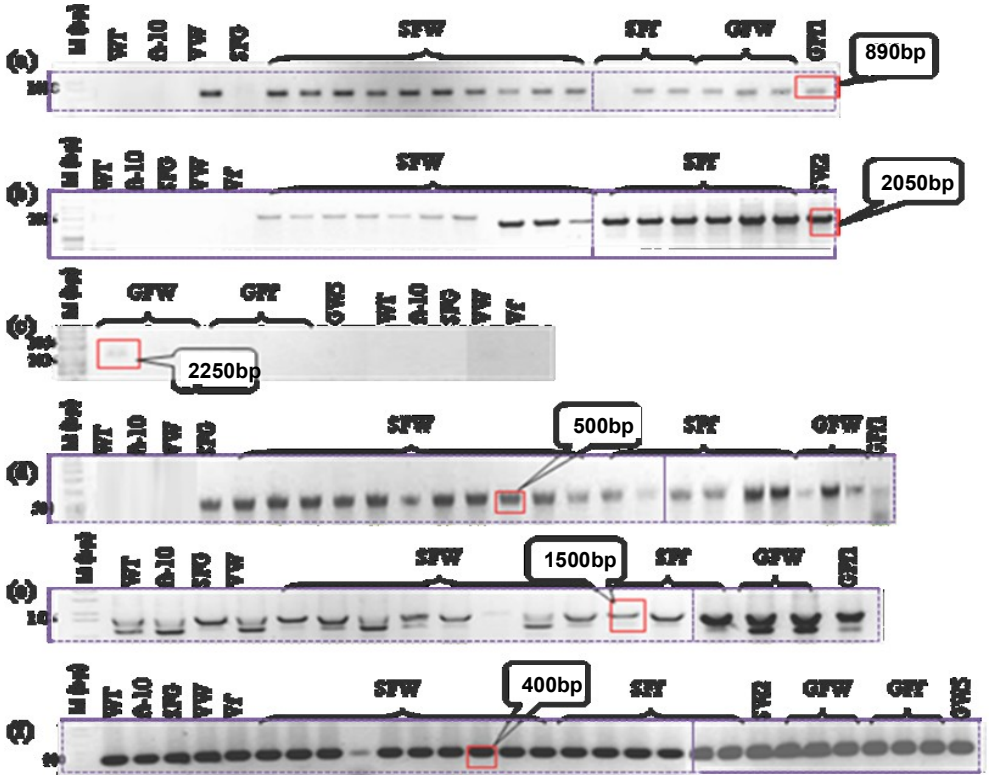


Figure 4.7: Further verification of T_2 plants carrying the full construct was done on genomic DNA using primer pairs to confirm the presence of (a) Dendra2, (b) SUC2 promoter, (c) GAS1 promoter, (d) FTcDNA and (e) M13. Actin was used as control (f). M=Marker, WT=Wild type, VW=vector:Dendra2 in WT background, Vf=Vector:Dendra2 in ft-10 background, SFG=SUC2::FT:GFP, SFW=SUC2::FT:Dendra2 in WT, SFf=SUC2::FT:Dendra2 in ft-10, SW=SUC2:Dendra2 in WT, GFW=GAS1::FT:Dendra2 in WT, GFf=GAS1::FT:Dendra2 in ft-10, GW=GAS1:Dendra2 in WT.

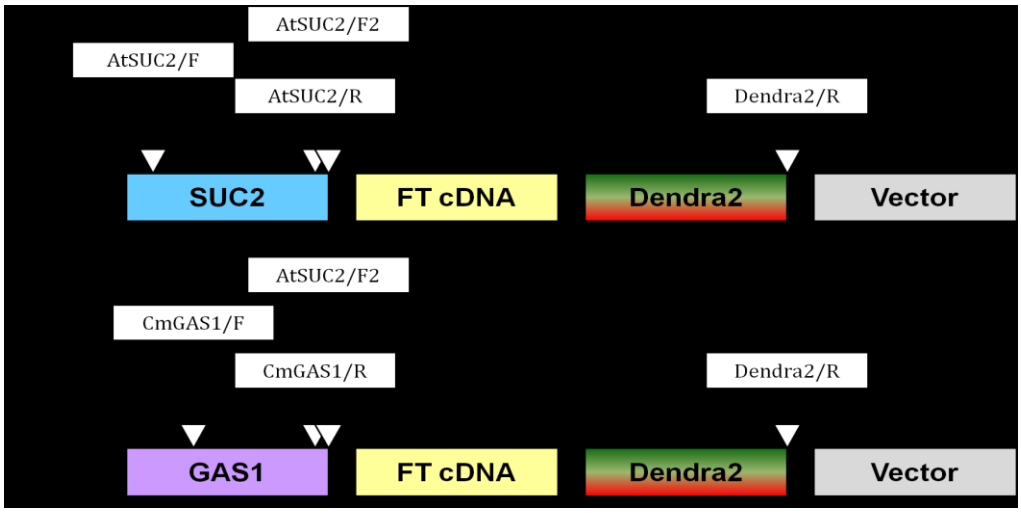


Figure 4.8: Schematic representation of the location where the primers for the promoters were designed on (a) SUC2::FT:Dendra2 and (b) GAS1::FT:GFP. For rt-PCR the presence of FT:Dendra2 transcripts were determined by using the second forward primers of each promoter (AtSUC2/F2 and CmGAS1/F2) with reverse primer of Dendra2.

From rt-PCR analysis, it was apparent that some of those thought to be carrying the full construct sequence did not produce any transcript at all. Level of transcript abundance in all plants varied for each of the promoters, Dendra2 and FT. All of the plants were positive or produced the transcripts of Dendra2 and FT (Fig 4.9 (a) and (d)) with the exception of the two lines (SW2 and GW5) that did not contain any FT cDNA in their construct.

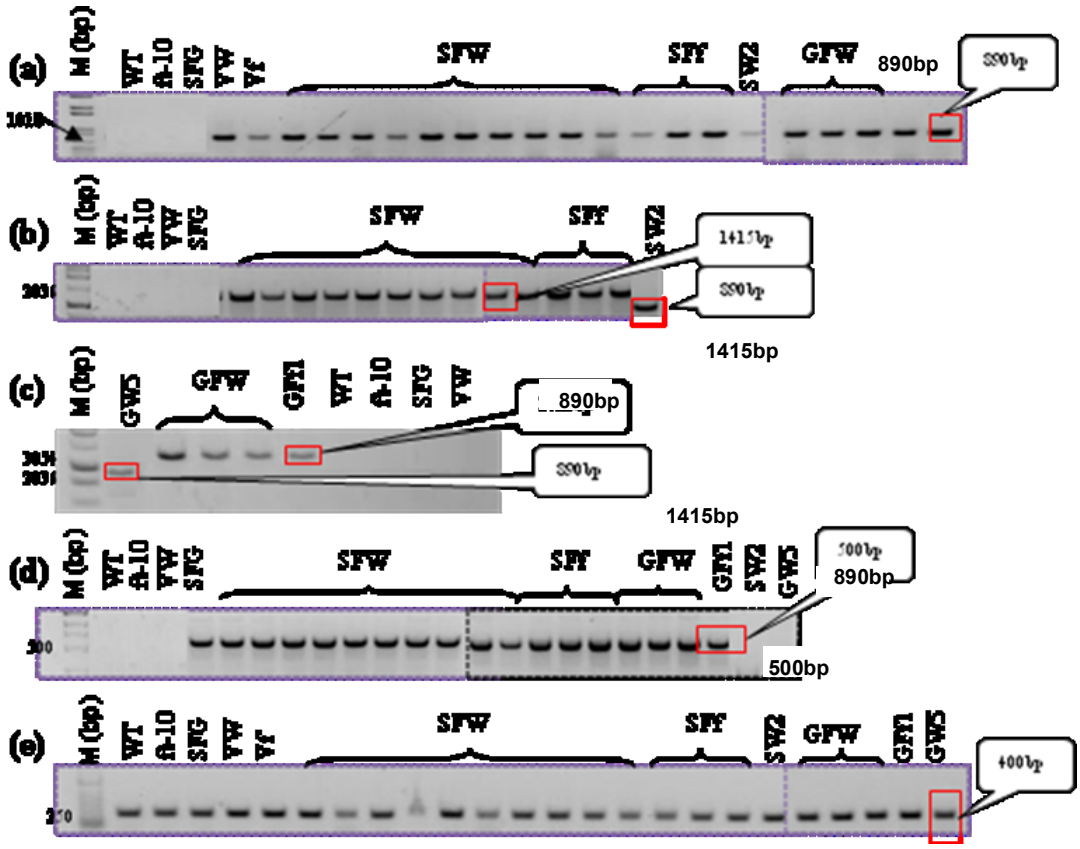


Figure 4.9: rt-PCR carried out on T_2 plants to determine transcription level of (a) Dendra2, (b) FT:Dendra2 in lines mediated by SUC2, (c) FT:Dendra2 in lines mediated by GAS1 and (d) FT in the fusion proteins. The same primer pairs as in genomic DNA PCR were used with the exception of M13. Actin was used as control as shown in (e). Negative control (-ve) used was the product of first strand cDNA synthesis using RNA from WT without the addition of reverse transcriptase enzyme. M=Marker, WT=Wild type, -ve=negative control, VW=vector:Dendra2 in WT background, Vf=Vector:Dendra2 in ft-10 background, SFG=SUC2::FT:GFP, SFW=SUC2::FT:Dendra2 in WT, SFf=SUC2::FT:Dendra2 in ft-10, SW=SUC2:Dendra2 in WT, GFW=GAS1::FT:Dendra2 in WT, GFf=GAS1::FT:Dendra2 in ft-10, GW=GAS1:Dendra2 in WT.

To determine the presence of FT:Dendra2 transcripts in the transgenic lines, a second forward primer was designed from the end of each promoter (Fig. 4.8) and paired with the reverse primer of Dendra2. Combination of the primer pairs will give a product size of 1415bp for the transgenic lines carrying full construct and only the Dendra2 transcript for the two lines without any FT cDNA in their construct (Fig. 4.9 (b) and (c)). These plants along with the controls were then used in an attempt to

study FT protein expression level via Dendra2. Under the fluorescent microscopy, weak expression of FT was observed (Fig. 4.10) but further analysis using confocal was unsuccessful.

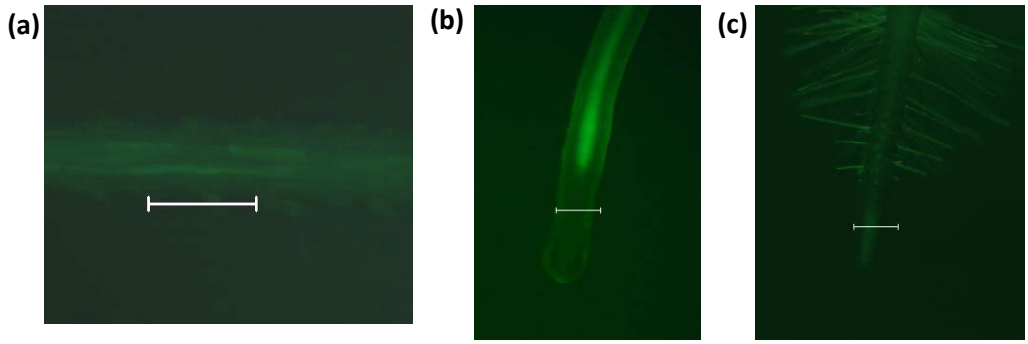


Figure 4.10: Preliminary analysis of FT:Dendra expression. Via fluorescence microscopy, (a) weak expression of FT through Dendra2 was observed in the phloem of the roots of T_2 SUC2::FT:GFP plants but not the root tip as compared to (b) SUC2::FT:GFP and no expression was observed on (c) WT plant. Bar scale showed the size of 0.2mm.

4.4 Discussion

4.4.1 Comparison of Flowering Time and Transcript Level of Expression of SUC2::FT:Dendra2 and GAS1::FT:Dendra2 Lines

Flowering time for T_2 plants was determined both in LD and SD as represented in Fig. 5.4. Based on the graph alone it was apparent that a large number of plants for both lines flowered more or less at about the same time as wild type, even for plants in *ft-10* background. Moreover, all of those in the mutant background flowered earlier than the mutant itself which gave an indirect indication of insertion and expression of FT to induce flowering.

However, it was observed that the control plants (WT and mutant plants incorporated with just the vector without any promoter or FT) showed a rather unexpected flowering time. While both did not flower earlier than WT plant, control plants in mutant background flowered earlier both in LD and SD. In theory both of these control plants should behave as the original WT and *ft-10* mutant plants as they contain neither SUC2 promoter that is very much involved in phloem transport to

increase gene expression or FT protein. The vector used is a common pGreenII vector which has no specific function in plant or in any other organism apart from being used as a plasmid to facilitate generation of transgenes.

In comparison, transgenic plants from the SUC2::FT:Dendra2 lines flowered earlier than GAS1::FT:Dendra2 both in SD and LD (Fig. 4.5) Nevertheless, promoter-control plants (SUC2:Dendra2 and GAS1:Dendra2) did not show much difference in flowering time between each other. However, the emergence of the first flower bud was also noted. Based on observation it was found that all of the SUC2::FT:Dendra2 line in WT background and two of the six SUC2::FT:Dendra2 lines in *ft-10* background actually flowered earlier than WT in both LD and SD. However, based on Bonferroni's post-test of one way ANOVA, some of these lines were not significantly different (Table 5.3). In addition, at least one of the GAS1::FT:Dendra2 line (GFW3) in WT background flowered a day earlier and comparison of this line with WT was highly significant in both LD and SD. This finding was in contradiction with the flowering time based on rosette leaf count and it was not known why the transgenic lines produced more leaves. On the other hand, in nature, changes in the environmental conditions influence the switching on and off of genes that govern leaf and flower development (Tooke et al., 2005). In a controlled environment such as with the current study, this may not likely be the case. It is safe to say that the transgenic plants did not go through floral reversion whereby after a certain period of flower development plant return back to produce leaf or to an earlier phase of development. A study done on *Arabidopsis thaliana* had seen reversion from flower to inflorescence development but not back to leaf production (Tooke et al., 2005).

In another comparison, this time through expression of transcript based on rt-pcr, although most SUC2::FT:Dendra2 plants showed positive result with required primer pairs for complete construct, expression level could not be determined accurately as semi-quantitative Real time PCR was not carried out due to time constraints. On the other hand, only 2 of the GAS1::FT:Dendra2 plants carried the full construct which are GFW1 and GFf1 but nothing conclusive can be made from their flowering time. Similar level of expression of FT transcripts observed between SUC2::FT:Dendra2 and the other lines were unexpected. Furthermore, weak

expression was observed in only a few lines under fluorescent microscope while none under confocal which was in contrast with the expression of transcripts. Previous study using FT:GFP fusion protein in SUC2::FT:GFP transgenic lines proved that the fusion protein is highly expressed and could be detected without a problem under fluorescent microscope (Corbesier et al, 2007).

In this study, the promoter in addition to the integrated FT protein inserted into the transgenic plants played a role in determining and regulation of flowering time. SUC2 and GAS1 have both been used in phloem studies as a promoter. Though both function mainly in the leaf, each played a separate role and on different organ or tissue in the leaf. SUC2 is a phloem specific promoter that functions mainly in the companion cells of major veins of leaves and in stems and roots hence its wide use as a positive control for complementation in phloem studies (An et al., 2004; Corbesier et al., 2007; Notaguchi et al., 2008). In FT study where the protein was observed to move from the leaf to the shoot apical meristem (Corbesier et al., 2007), FT protein movement was mediated by SUC2 promoter. On the other hand, GAS1 functions in the minor leaf veins, thus the promoter can be used as a comparison to SUC2 in phloem transport study. However GAS1 functions mainly in the minor leaf veins but not in stems and roots, and this makes it a comparable negative promoter alongside SUC2. Corbesier et al (2007) demonstrated GAS1:FT caused early flowering in the *ft-7* mutant plants, much earlier than WT but GAS1:FT:GFP have no effect on transgenic plants and flowered as late as *ft-7* mutants.

The SUC2 promoter used originated from AtSUC2, a predominant Suc-proton (Suc/H^+) symporter catalysing phloem loading. During phloem transport AtSUC2 expression has been reported to express strongly and implication had been made of its efflux and retrieval role to and from flanking tissues through the apoplast (Srivastava et al., 2008). They reported that growth of *Atsuc2* mutant plant was restored via the expression of AtSUC2 gene supporting a role in retrieval. From the current study, lines carrying the SUC2 promoter and FT in *ft* mutant background were restored to nearly WT level in terms of flowering time and accelerated growth was observed for transgenic plants in the WT background. Driven by SUC2 promoter, expression of free GFP was documented in the leaves of *Arabidopsis thaliana* followed by

transportation to roots (Stadler et al., 2005). Based on study by Stadler et al (2005) and also previous studies on FT it is assumed that FT was transported to the shoot apical meristem in lines with mutant background to initiate flowering.

Meanwhile, GAS1 promoter in this study was derived from *Cucumis melo (CmGAS1p*) which is only expressed in the minor vein of mature leaves. GAS1 synthesised galactinol in the phloem minor vein but long-distance transportation of galactinol was limited (Srivastava et al., 2008). Leakage of galactinol from the phloem was speculated to be redistributed in the lamina and apoplast of mature leaves by xylem transport (Ayre et al., 2003). Srivastava et al (2008) reported mutant plants expressing *AtSUC2* cDNA incorporated with this GAS1 promoter lack AtSUC2 activity in the transport phloem with intermediate growth and normal morphology. Almost similar observation was obtained in the current study whereby all of the transgenic plants in both WT and *ft-10* mutant backgrounds containing this promoter either alone or in combination with FT had intermediate growth with normal morphology.

4.4.2 Could FT and Its Movement Be Tracked Using Dendra2 Fluorescent Protein?

Being a photoswitchable fluorescent protein, using Dendra2 to track FT protein movement or localisation would provide better visualisation. Dendra2 is an improved version of its predecessor Dendra that displays faster maturation and brighter fluorescence before and after photoswitching. Compared to common GFP that only gives green fluorescence, the ability of Dendra 2 to photoswitch from green emission when excited at 490 nm to red emission when excited at 553 nm along with ability to photoconvert using UV at 405 nm or blue light at 488nm (www.evrogen.com) are an added bonus. Coupled with its small size of 26kDa and monomeric sturcture, Dendra2 offers a better tracking option for studying protein.

FT protein was tracked by Corbesier et al (2007) in the leaf and shoot apical meristem of grafted *Arabidopsis* using FT fused to GFP. With a size of 27kDa, GFP does not differ much with Dendra2 in terms of size. Hence fusion of FT which is

approximately 20kDa to Dendra2 should be able to be tracked in a given cell. However, other factors such as vector and promoter used have to be taken into account to ensure successful tracking. In addition, the cell where the protein is to be tracked also played an important role.

Driven by SUC2 which is a phloem specific promoter, FT can be delivered efficiently from the leaf to the shoot apical meristem via phloem. Previous study has proven this mechanism utilising grafting process (Corbesier et al., 2007) and FT protein from the scion of SUC2::FT:GFP plant was later detected in the shoot apical meristem of *ft-7* mutant rootstock plant (Notaguchi et al., 2008). Early flowering was reported in single ft mutant or double mutant by the expression of FT from promoters specific for phloem companion cells (An et al., 2004; Abe et al., 2005; Corbesier et al., 2007; Jang et al., 2009).

Though Dendra2 has been mostly used in animal cells, it is gaining popularity in plants. It was used to track SYNV (*Sonchus yellow net virus*) protein in the nuclei of *Nicotiana bentamiana* (Martin et al., 2009) and was found to have similar localisation pattern to GFP fusions. Martin et al also reported the possibility of selective photoconverting subnuclear regions of interest from green to red without disturbing other nuclei in the adjacent cells. The green fluorescence of RSZp22-Dendra2 fusion protein, a vital nuclear-localised splicing factor from serine/arginine-rich (SR) proteins expressed in tobacco leaf cells was detected before photoactivation within the nucleus before it gradually decreased with continuous photoactivation and increasing the red fluorescence proportionally. However, the same fusion protein was not detectable in *Arabidopsis* using the 488 nm blue light or even intense UV light irradiation (Rausin et al., 2010). Not much else has been done in other plants and even an attempt in this study to track FT movement was not successful. Weak expression was detected under fluorescent microscope but nothing could be detected under confocal microscopy. Teerawanichpan et al (2007) reported that sometimes fluorescent protein that is suited to animal or bacterial cells may not suit well in plant cells. Since Dendra2 had been successfully used and its expression detected in the tobacco family, it is not totally ineffective as a tool in plants and further study needs

to be done to investigate the use of this photoswitchable fluorescent protein in *Arabidopsis*.

4.4.3 Stability of Dendra2 in Fusion Protein to Detect and Track Movement of FT

Based on studies by Martin et al (2009) and Rausin et al. (2010) who had successfully use Dendra2 in *Nicotiana benthamiana* and tobacco, both from the same plant family, a similar result could be predicted for other species. However, since in the current study FT protein fused to Dendra2 was not successfully tracked, it may come down the question of stability of Dendra2 in conjunction with FT.

Many factors could be involved that may hinder the successful use of Dendra2 in *Arabidopsis*. These factors include the choice of vector used as a carrier for the fusion protein, restriction site modifications and gene copy number after transformation. Various vectors are available for this purpose and it all comes down to choosing the correct vector that suits each work. This is because most of the PAFP could only be cloned into Gateway-compatible binary vectors. Certain vectors may have restriction to its usage and hence may deem unsuitable for fusion protein expression. Vectors should be chosen based on their specific advantages, for example the pSITE-BiFC and pSITEII vectors used by Martin et al (2009) which offer wide range of functionality of autofluorescent protein (AFPs). pSITEII vectors has modified restriction sites to facilitate replacement and testing of novel AFPs and had been proved to be efficient in plant-based assays (Merzlyak et al., 2007; Teerawanichpan et al., 2007; Schenkel et al., 2008; Shaner et al., 2008; Subach et al., 2008). In comparison, the fusion protein in the current study was constructed by attaching adaptors to the Dendra2 before ligating it to the restriction site from the MCS (multiple cloning sites) of pGreenII.

Though not as strong as the 35S promoter, SUC2 has been used in a lot of FT studies and known for its role in the major vein of the phloem. Bronsius (1984) reported that selection of cells with reduced promoter-gene dosage could be the result of gene products accumulation from strong promoters detrimental to cell growth

hence reducing replication efficiency of plasmid DNA. In addition, gene copy number also played an important role to ensure stability of transgene. Reduced transcript from rt-pcr performed using primers designed for SUC2 and GAS1 and also FT may be due to reduced or low gene copy number which had been implied to influence plant activity (Srivastava et al., 2008).

Although strong Dendra2 transcripts were not seen across all the lines, FT cDNA level of expression was generally strong apart from promoter control plants of SW2 and GW5 (Fig. 5.7 (a) and (d) respectively). Strong level of expression for FT:Dendra2 transcripts were also observed across the SUC2::FT:Dendra2 lines (Fig. 5.7 (b)) but variation in level of expression was observed in GAS1::FT:Dendra2 lines. However, when viewed under confocal microscopy no signal was observed. Only weak expression was detectable in some plants under fluorescent microscope. The results obtained from rt-PCR were in contrast with microscopy observation. Most probably the transcripts of the fusion protein were not stable for the conversion to protein. This could be the cause to the inconsistency of transcripts observed in the transgenic plants in this study that needs to be investigated further.

Besides, problem in obtaining stable fusion protein expressing Dendra2 plants in *Arabidopsis* had been reported by Rausin et al. (2010) but did not give any reason for the failure. Utilization of Dendra2 as a fusion protein in stably transformed *Arabidopsis* has never been reported prior to a study done by Wu et al. (2011). Theirs is the first to document Dendra2 fusion protein with stable expression in *Arabidopsis*. Following that, Nagawa et al (2012) also reported transient photoconversion expression of Dendra2 fusion protein in tobacco and *Arabidopsis* leaves. The methods used for photoconversion by Nagawa et al (2012) did not differ much from what was recommended by manufacturers or those previously used in animal cells. By contrast, Wu et al (2011) did some modifications for photoconversion of Dendra2 by optimizing confocal settings. Additionally the team noted that existing construct originally containing GFP should not be directly substituted with Dendra2 though Nagawa et al (2012) had no problem with direct replacement of construct already containing GFP with Dendra2.

As transgenic plants in the current study were only grown up until T_3 generation, probably homozygous stage of the fused protein had yet to be reached for certain lines. Even though flowering time of all independent lines for a particular transgene showed a more or less similar count across repetitions (data not shown) it may still be the transcript that influences the protein expression the most. In summary, it is still inconclusive to differentiate and determine successfully transformed plants from plants that can express the FT:Dendra2 fusion protein. Direct replacement of the original constructs already containing GFP with Dendra2 may also contribute to the failure of producing fluorescent fusion protein.

CHAPTER 5
GENERAL DISCUSSION

5.1 General Discussion

The main aim of this thesis is to provide a list of phloem-mobile proteins, in other words an *Arabidopsis* phloem exudate proteome, especially in relation to regulation of flowering time. Interest in the phloem was originally based on the fact that it plays an important role in transportation of nutrients (Lucas et al., 1995). Over time the phloem has progressively been identified to protect plants against pathogens and stress, is involved in long-distance signalling and regulates plant growth and development (Ruiz-Medrano et al., 2001; Wu et al., 2003; van Bel and Gaupels, 2004; Haywood et al., 2005; Lough and Lucas, 2006).

Many studies have led to the discovery of various proteins in the phloem of many plant species such as brassica, cucurbit, rice, poplar and lupin (Giavalisco et al., 2006; Walz et al., 2004; Lin et al., 2009; Aki et al., 2008; Dafoe et al., 2009; Rodrigues-Medina et al., 2011). Some of these proteins are reported to be mobile based on their identification from phloem exudate.

At the start of this study no detailed phloem proteomics study was reported in *Arabidopsis*. Hence, method development (chapter 2) was incorporated to establish suitable exudation methods for good quality phloem exudate for identification of protein. Incorporation of method development is necessary to optimise all possible parameters including scientifically justified and logical step-by-step experimental approaches. This ensured that reliable results were obtained in the main study. A suitable exudation method utilising EDTA to study *Arabidopsis* phloem proteome was developed by optimising previous protocols used in other plants (King and Zeevaart, 1974; Hoffman-Benning et al., 2002). Optimisation of the chelating agent concentration is important due to exudation technique employed that required collection of phloem sap from wounded leaves. In the phloem, sieve tubes are equipped with defense mechanisms and once wound is inflicted, protein clogging and callose sealing are activated. However, as both mechanisms are calcium dependent

(Will and van Bel, 2006), the presence of calcium chelating agents such as ethylenediaminetetraacetate (EDTA) or ethylene-glycol-bis-(b-aminoethyl-ether) N, N9-tetraactetate (EGTA) decreases the abundance of free Ca^{2+} in the sieve elements (Knoblauch et al., 2001) allowing collection of phloem exudate from excised stems (King and Zeevaart, 1974). In brief, the optimised method for obtaining sufficient phloem exudate protein from *Arabidopsis* is using 6-weeks old plants by collecting the phloem exudate overnight in 20 mM EDTA.

A total of 221 phloem exudate proteins were found within the *Arabidopsis* phloem proteome via EDTA exudation method. Comparisons made against phloem proteome databases from three other plant species indicated that only 16 of the identified proteins (7.2%) were homologous to phloem proteins from brassica (Giavalisco et al., 2006), cucurbit (Lin et al., 2009) and rice (Aki et al., 2008). This was in accordance to previous phloem proteome studies in other species that a large number of proteins are contained within the phloem, many of which are stress-related and involved in defense system.

Two other similar phloem proteome studies were reported in *Arabidopsis* earlier this year (Guelette et al., 2012; Batailler et al., 2012). Each study including the current study managed to identify a number of proteins with Batailler et al finding the most (377 proteins) and a lower number documented by Guelette et al (65 proteins). The difference in number of identified protein from each study may be due to the exudation method employed. While all studies utilised the use of EDTA for exudation of phloem, the buffer used for final collection of exudates in each study is different from to another. These will be discussed further in the next section (5.2).

In the current study, proteins identified were mostly from both total phloem protein and excised gel band of 1D SDS-PAGE. Although protein content from both types of sample should contain similar proteins, more proteins were identified from total phloem than excised gel bands. During gel-cutting less visible bands on the gel might have been left out hence reducing the number of identified proteins. And, in terms of ease of handling the materials for experiments, the use of total phloem protein fits the requirement where the sample need not be run on SDS-PAGE prior to

tryptic digest, and the digestion process itself could be done in a shorter amount of time than when using excised gel bands. This is also one feature that is different from the current study as compared to Guelette et al and Batailler et al as both studies only used excised gel bands for MS analysis. Additionally, 2D was performed in the current study and some of the spots sent for MS analysis were among the identified proteins (Table 3.2, Chapter 3).

FT was not identified from the current study, similar to result obtained by Guelette et al. but in contrast to the study by Batailler et al (2012). Furthermore, FT homologs were also found in brassica, and rice (Giavalisco et al., 2006; Aki et al., 2008). Nevertheless, FT was identified in the current study using MRM methods targeting FT peptides for detection, and a similar finding was also reported in cucurbit (Lin et al., 2007). Although FT protein was identified in the phloem proteome of *Arabidopsis* from another group and not in the current main proteomics study, there could be several reasons leading to the event. Batailler et al conducted their study by growing the plants in LD from the beginning. Plants from current study were grown in SD and were only exposed to LD for three days prior to exudation. FT mRNA level was quantifiable as reported by Corbesier et al (2007) using plants grown in similar condition hence identification of FT protein was anticipated. As FT protein was identified using MRM, it showed that FT was present in *Arabidopsis* plants exposed to inductive LD photoperiod treatment. However, as FT is one of the low abundance proteins, which hinders its detection and identification via non-targeted LC-MS/MS.

Detection of FT using MRM proved to be a useful method not only to identify targeted peptides from a protein but may be applied for quantitative analysis as shown in this study. Using three FT peptides as standards, an attempt to determine level of FT protein in response to photoperiod was conducted. After 60 hours of exposure to LD, FT was identified from each sample collected at 4 hourly intervals. This experimental design was based on an earlier study of mRNA level of expression by Corbesier et al (2007). As reported in section 3.3.8 (Table 3.7), attempts to quantify level of FT protein using MRM did not exactly match with results from *FT* mRNA study. Instead, varying level of FT protein was observed (Fig. 3.13). This was perhaps expected as not all protein is translated almost immediately after production of

mRNA. In summary, findings in this study demonstrate that various proteins involved in a range of biological processes are contained within *Arabidopsis* phloem proteome and that FT protein level in response to photoperiod is quantifiable using suitable method.

In the last part of this study, an attempt to investigate the expression, movement, and localisation of FT was conducted by constructing FT fusion protein genes driven by SUC2 and GAS1 promoters, in an attempt to provide further detail on FT expression and movement route. Fusion protein constructs were transformed into WT and *ft-10* mutant background for comparison. From flowering time determined by rosette leaf count of transgenic lines, the lines carrying FT:Dendra2 constructs were observed to flower more or less at the same time as WT in both LD and SD. Additionally, all plants flowered earlier than *ft-10* in LD and SD. Statistical analysis performed on data from LD and SD plants showed different significance levels (Table 4.3 in section 4.3.3). Expression level of FT:Dendra2 in lines carrying SUC2 promoter showed strong transcript levels in both WT and *ft-10* background while lines carrying GAS1 promoter showed a slight variation (Fig. 4.7, section 4.3.4). Observation of FT protein movement by Corbesier et al (2007) through grafting was mediated by SUC2 that functions mainly in the companion cells of major veins of leaves in addition to stems and roots (An et al., 2004; Corbesier et al., 2007; Notaguchi et al., 2008). On the other hand, GAS1 mainly functions in the minor veins of leaves and not in the stems or roots, so it is often used as a comparable promoter alongside SUC2 since both promoters have function in the phloem. Observation of FT:Dendra2 expression mediated by SUC2 in this study was unsuccessful where weak expression was observed in a few plants (Fig. 4.10) but the fusion protein could not be detected under confocal microscopy. Among factors thought to contribute to the failure was that homozygous state might not yet have been achieved in all transgenic lines. Further generations are needed to select homozygous lines expressing Dendra2 for better detection. Another factor to consider is the choice of vector used for protein expression. The contrasting results from RT-PCR and microscopy observation could be caused by unstable transcripts, preventing conversion to protein. Rausin et al (2010) reported similar problem in obtaining stable fusion protein expressing Dendra2 in *Arabidopsis.* Nevertheless, Wu et al (2011) and

Nagawa et al (2012) respectively documented stable expression of Dendra2 fusion protein in *Arabidopsis* and transient photoconversion of Dendra2 fusion protein in tobacco and *Arabidopsis* leaves. Previously, Dendra2 has been successfully used to detect fusion protein in *Nicotiana benthamiana* (Martin et al., 2009). Wu et al suggested substituting original construct already in GFP to complement replacement with Dendra2. Interestingly, Nagawa et al did no such substitution and directly replaced the original construct already containing GFP with Dendra2. Results from both of these studies proved that Dendra2 could be used to replace GFP in fusion proteins for expression in *Arabidopsis*. However, further work needs to be done to ensure compatibility of Dendra2 with chosen fusion protein constructs followed by obtaining stable transformed plants and lastly modification of confocal settings for photoconversion of Dendra2.

5.2 Establishing a Phloem Proteome Database for *Arabidopsis thaliana*

A wide variety of biochemically active proteins in the phloem sap of various plant species has been documented where they are involved especially in the coordination of metabolism, development and defense response (Nakamura et al., 1993; Balachandran et al., 1997; Ishiwatari et al., 1998; Kehr et al., 1999; Schobert et al., 2000; Aoki et al., 2002; Yoo et al., 2002; Walz et al., 2004; Giavalisco et al., 2006; Aki et al., 2008; Lin et al., 2009; Rodriguez-Medina et al., 2011). As mentioned previously, extensive studies of phloem proteome from several plant species such as brassica, cucurbit, lupin, poplar, rice and melon have been reported (Giavalisco et al., 2006; Walz et al., 2004; Lin et al., 2009; Rodriguez-Medina et al., 2011; Dafoe et al., 2009; Aki et al., 2008; Malter and Wolf, 2010).

Until recently, proteomics studies in *Arabidopsis thaliana* have been reported for chloroplasts, mitochondria, peroxisome, root, organelle and seeds (Peltier et al., 2000; Fukao et al., 2002; Dunkley et al., 2006; Gallardo et al., 2001). Just before the submission of this thesis, two studies on phloem proteome in *Arabidopsis* were released (Guelette et al., 2012; Batailler et al., 2012). Earlier, Deeken et al (2008) described transcriptomic study of *Arabidopsis* phloem. The list of identified proteins from the phloem provided from current study in addition to the previous two studies

all contribute to the establishment of a comprehensive *Arabidopsis* phloem proteome database.

In total and across all studies, there are 397 proteins identified in the sieve element of *Arabidopsis thaliana* (Appendix D). The identified proteins have diverse functions in stress response, redox regulation, anti-oxidant response, various metabolic processes, signalling, protein folding, turnover and transport, cell organisation and nucleic acid binding. As shown in Figure 3.5 (d) (Chapter 3), there are 25 proteins that were identified in all three *Arabidopsis* studies. Another 78 proteins overlapped between combinations of two studies. Additionally, each study has its own unique proteins that were identified exclusively. Guelette et al and Batailler et al each have 26 and 189 unique proteins respectively. All the proteins were confirmed proteins from the SE and free from contaminants. In current study, 78 of the identified proteins were unique, and the reduced number from what was reported in Fig. 3.5 (d) was the result of elimination of cell wall protein, contaminants and ribosomal protein.

The 25 proteins identified in all studies are involved mostly in metabolic processes (glycolysis, TCA, secondary metabolism, hormone metabolism and carbohydrate metabolism) followed by redox reaction. A lot of enzymes were identified hence their involvement in redox reaction is hardly surprising. The rest of the proteins play roles in photosynthesis, stress-related, cell organisation, signalling, protein biosynthesis, biodegradation of xenobiotics and RNA-binding. Only one protein involved in signalling was found by all studies, general regulatory factor 10 epsilon (GRF10). Nevertheless, each individual study managed to identify a few other GRF family members where Batailler et al identified the most GRF protein. Additionally, they also identified FT. So far, only FT is known to be mobile and involved in systemic signalling inducing flowering. However, 14-3-3 proteins or general regulatory (GRF) proteins may be involved in interaction with FT where they were proposed to have roles downstream of flowering regulation pathway and in *Arabidopsis*, loss of function of two of GRF isoforms resulted in slight delay in flowering (Mayfield et al., 2007). The GRF proteins found from this study differ from the ones reported by Mayfield et al and their function remains to be investigated.

Although not one of the proteins found in all three studies, TCTP has been previously reported to be involved in long-distance movement of phloem proteins (Aoki et al., 2005). This protein was identified by both Batailler et al and present study, and it would be interesting to study this protein's function in the phloem. Finally, among the 25 proteins, two of them were under the category of miscellaneous and unknown. The fact that these two proteins were identified in all three studies may be an indication that they may serve a purpose but their roles have yet to be determined, or they could be just high abundance protein in the *Arabidopsis* phloem exudate.

Another interesting finding from the current study is the identification of ribosomal proteins (no. 201 to 221 in Table 3.2, chapter 3). In the sieve tubes of flowering plants, it is a known fact that no translation could occur yet the presence of ribosomal proteins and ribosomal RNA (rRNA) was documented in these enucleated cells (Zhang et al., 2009; Lin et al., 2009; Buhtz et al., 2007). Kragler (2010) suggested existence in conducting phloem tissue of the principal components necessary to facilitate mRNA translation. Although the idea was fascinating, it would still be impossible for translation to occur in the phloem in the absence of functional ribosomes. The ribosomal proteins identified from current study were most probably caused by translocation occurring in the immature SE then matured by eliminating most organelles (van Bel and Knoblauch, 2000) and the proteins may have been stored in immature SE. As a result of degradation of organelles, the content of mature SE comprise of the plasma membrane and cytoplasm that includes endoplasmic reticulum, phloem-specific plastids and mitochondria (van Bel and Knoblauch, 2000). From the current study, plastid proteins were also identified and since the plant was wounded, the sieve element plastids were likely to be ruptured (van Bel et al., 2002) resulting in the flow of such proteins in the exuded phloem sap. However, sucrose transporter protein was not identified in all three *Arabidopsis* phloem proteome studies nor has it been identified in other plant phloem proteome most probably due to its expression in the CC (Sauer and Stolz, 1994; Truernit and Sauer, 1995; Stadler and Sauer, 1996).

The list in Apendix D was compiled from three studies which differ in the exudation method used. Guelette et al used EDTA-facilitated exudation method

whereby they induced phloem exudation in EDTA solution for an hour, but then collect the exudates in deionised water. Similarly Batailler et al also used EDTA-facilitated exudation for their study however the exudates were in collected in phosphate buffer with EDTA. Meanwhile, the exudates in the present study were collected in EDTA solution. Additionally, the exudation periods were different in each study. In terms of plant growth conditions, both Batailler et al and Guelette et al used plants grown in LD. This is the most obvious difference with the current study where plants were generally grown in SD and were transferred to LD 3 days prior to exudation. *Arabidopsis* is a LD plant that flowers when exposed to long period of light therefore it might contribute to Batailler et al's greater number of identified proteins. This could also be the reason why FT was identified in their study. Though FT was identified in current study by MRM, FT might be a low abundance protein. Accumulation of FT over time could contribute to its identification by Batailler et al as compared to current study where plants were only exposed to LD for three days. Likewise FT was not identified by Guelette et al and they reported the least amount of identified protein. A similar problem was encountered in the current study when EDTA wash was applied as the exudates collected after the first wash contained less amount of protein (Chapter 3). This could be the reason for low number of identified protein by Guelette et al.

Despite the many differences, the similarities shared by all three studies were the use of excised gel bands and 1D LC-MS/MS for protein analysis. Most phloem proteome studies employed 2D analysis. In the middle of the current study I tried to incorporate 2D proteomics to better identify the *Arabidopsis* phloem exudates protein. However, due to failed attempts at getting reproducible gels and time constraint, the method was discontinued. As proved by Batailler et al, large numbers of proteins could still be identified by 1D LC-MS/MS although 1D SDS-PAGE is less specific as compared to 2D SDS-PAGE. Although all three studies were based on EDTA for phloem exudation, excised gel bands from 1D SDS-PAGE and 1D LC-MS/MS for analysis, each study has their own uniqueness and advantages. Of the three, current study is the one with a lot of differences as mentioned previously. In addition to excised gel bands, I also used total phloem exudates and a few of 2D excised gel spots for protein identification.

In summary, the phloem exudate proteome established from *Arabidopsis thaliana* contained various proteins with diverse function, and major abundant proteins were highly reproducible. Identification of many novel proteins along with the rest of the total list provides an insight that they have functions in the phloem that needs to be further explored.

5.3 Future Work

Earlier phloem proteomics studies were performed using 2D gels, 2D LC-MS or 1D LC-MS, or, a combination of more than one of the methods. 2D gel provides better coverage than 1D gel where one spot from 2D gel represents one protein instead of more than one protein contained within a single band of 1D gel. Although 2D proteomics was attempted in this study and the results were not reproducible, further optimisation needs to be explored. A change in solution for phloem exudate collection may increase IEF sensitivity. Previously, solution used to collect exudates contained EDTA and salt is known to interfere with IEF by creating artefacts (Rabilloud et al., 2007). The EDTA-facilitated exudation methods by Batailler et al (2012) and Guelette et al (2012) could be used as alternatives. It would definitely be interesting to try out the method employed by Batailler et al since they managed to identify a large amount of protein using just 1D SDS-PAGE. Another suggestion would be to apply aphid stylectomy (Doering-Saad et al., 2002). Although this method can be tedious since *Arabidopsis* is a small plant, it is not impossible to be done and the phloem obtained will be pure without the need to concentrate it further as with EDTA-exudation technique. If these could be carried out and reproducible results could be obtained, it would facilitate extension of the *Arabidopsis thaliana* phloem exudate proteome database.

It would be interesting to study further the differences in phloem protein content from various genotypes. In this study, five genotypes including wild type were used but definitive work to determine the genotypic differences in proteome was not carried out. Four mutants were used in this study namely *co*, *ft*, *tsf* and *ft-tsf* where a loss of gene function occurred in each mutant and in the case of *ft-tsf*, a double loss

of function. The dysfunctional gene or genes have an effect on the photoperiodic pathway in relation to flowering. It is predicted that there will be variation in the proteins contained within each genotype. From the four pathways of photoperiodic flowering of *Arabidopsis*, gibberellic acid (GA) signalling pathway is able to induce flowering directly by activating *SOC1* under SD (Moon et al., 2003). When plants are transferred from SD to LD, bolting was the result observed due to an increment in active GA concentration (Xu et al., 1997) in addition to activation of FT. Blazquez et al (2002) reported stem elongation from cell expansion due to increased concentration of GA under LD. Hence, the mutants not only regulate genes and proteins involved in flowering time but also plant growth in general.

Further quantitative analysis is needed before level of FT protein in response to photoperiod can be verified. Although MRM is a good and cost-effective technique for protein quantification, an alternative approach would be to use iTRAQ (isobaric tags for relative and absolute quantification) utilising stable isotope labelling. The N-terminus of protein digests will be tagged chemically using iTRAQ (Applied Biosystems). The labeled mass tags can then be subjected to LC-MS/MS analysis. Currently 8 tags are available for labeling hence eight different conditions (such as photoperiod condition, plant genotype, etc.) can be applied and later multiplexed together in one experiment. Once combined, the samples will be fractionated using cation exchange, separated by reverse phase nano-LC and analysed by tandem MS. Identification of the labeled peptides from the fragmentation of the tag attached to the peptides generates a low molecular mass reporter ion unique to each individual tag used for labeling protein digests. Quantification of the peptides in each digest is done by measuring the intensity of the reporter ions. An unambiguous quantitative comparison for each peptide from the MS/MS scan is obtained after the peptide is sequenced.

The attempted quantification of FT performed in this study involved plants in response to LD photoperiod. Once the quantification has been obtained, FT quantifications can be further explored in a variety of condition such as LD versus SD, different period of exudation collection (30 min, 3 hours and overnight) and plants age. These studies would provide better understanding of FT and its function in

flowering where it may be used to enhance and accelerate flowering in certain plant species.

Finally, the yield of an expression system is correlated with gene copy number whereby increasing the gene copy number via gene amplification may lead to enhanced protein production. This knowledge can be applied for future fusion protein work expressing Dendra2. Real-Time PCR assay could be performed to test out gene copy number suitable for protein production. Absolute or relative quantification can be carried out to enable data analysis using copy number and relative quantification study via real-time PCR. Another suggestion for future fusion protein work expressing Dendra2 would be exploring the choice of vector. Martin et al (2009) suggested using vectors based on their specific advantages especially if they offer wide range of functionality of autofluorescent protein. One of the vectors proven to be efficient in plant-based assays with modified restriction sites that facilitates replacement and testing of autofluorescent protein is pSITEII (Merzlyak et al., 2007; Teerawanichpan et al., 2007; Schenkel et al., 2008; Shaner et al., 2008; Subach et al., 2008) which coincidentally was also used by Martin et al to successfully detect protein fused to Dendra2 in *Nicotiana benthamiana*. Initial constructs of FT protein used in current study were within pGreenII vector and although the presence of multiple cloning sites (MCS) on the vector was to facilitate recombination, it may not be suitable for insertion of Dendra2. Using another vector which had been tested and proved to facilitate protein expression in place of existing vector such as those used by Wu et al (2011) and Nagawa et al (2012) may improve the chance to study movement of FT protein in *Arabidopsis* and hence provide a better understanding of how FT regulates flowering. Additionally, further verification of transformation and expression of fusion protein should be incorporated. Sequencing was carried out on plants that showed positive result carrying the transformed fusion protein from PCR and rt-PCR but was only done on T_1 plants. Further sequencing was not carried out on T_2 and T_3 plants. Sequencing should be incorporated all the way through to ensure no changes or mutation has occurred to the fusion protein. Furthermore, plants should also be grown in both LD and SD to determine the expression of FT and for comparison purposes.

In summary, a combination of proposed future studies can further enhance knowledge of the *Arabidopsis* phloem proteome and FT in the model plant. Although FT remains the only florigen to date, some of the identified proteins particularly the 14-3-3 (GRF) proteins which has been reported to have direct interaction with FT (Purwestri et al., 2009; Taoka et al., 2011) could be explored further to determine their function in long-distance systemic signalling and flowering. Their interaction with FT could implicate their importance and thus possible regulatory role in flowering. Meanwhile, the choice of various commercial PAFP would accommodate construction of improved FT fusion proteins. Driven by phloem-specific promoters, study of FT fusion proteins would provide a better insight of FT movement from the phloem to the shoot apical meristem, an essential component of exploration of the regulation of FT and its homologs.

BIBLIOGRAPHY

Abe M, Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., Araki, T. (2005) FD, a bZIP Protein Moderating Signals from the Floral Pathway Integrator FT at the Shoot Apex. *Science* 309:1052-1056.

Adam V, Nienhaus, K., Bourgeois, D., Nienhaus, G.U. (2009) Structural Basis of Enhanced Photoconversion Yield in Green Fluorescent Protein-Like Protein Dendra2. *Biochemistry* 48(22):4905-4915.

Adams SR, Valdes, V.M., Langton, F.A. (2008) Why Does Low Intensity, Long-Day Lighting Promote Growth in *Petunia, Impatiens*, and Tomato. *J. Hort. Sc. & Biotech.* 83(5):609-615.

Aebersold R, Mann, M. (2003) Mass Spectrometry-Based Proteomics. *Nature* 422:198-207.

Aki T, Shigyo, M., Nakano, R., Yoneyama, T., Shuichi, Y. (2008) Nano Scale Proteomics Revealed The Presence Of Regulatory Proteins Including Three FT-Like Proteins In Phloem And Xylem Saps From Rice. *Plant Cell Physiol.* 49(5):767-790.

Alvarez S, He, Y., Chen, S. (2008) Comparative Investigations of the Glucosinolate-Myrosinase System in *Arabidopsis* Suspension Cells and Hypocotyls. *Plant Cell Physiol.* 49(3):324-333.

An H, Roussot, C., Suarez-Lopez, P., Corbesier, L., Vincent, C., Pineiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C. (2004) *CONSTANS* Acts in the Phloem to Regulate a Systemic Signal That Induces Photoperiodic Flowering of *Arabidopsis*. *Development* 131:3615-3626.

Anderson L, Hunter, C.L. (2006) Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Molecular & Cellular Proteomics* 5:573-588.

Anderson LE, Ringenberg, M.R., Carol, A.A. (2004) Cytosolic Glyceraldehyde-3-P Dehydrogenase and the B Subunit of the Chloroplast Enzyme are Present in the Pea Leaf Nucleus. *Protoplasma* 223:33-43.

Ando R, Hama, H., Yamamoto-Hino, M., Mizuno, H., Miyawaki, A. (2002) An Optical marker Based on the UV-Induced Gree-to-Red Photoconversion of a Fluorescent Protein. *PNAS* 99(20):12651-12656.

Aoki K, Kragler, F., Xonoxostle-Cazares, B., Lucas, W.J. (2002) A Subclass of Plant Heat Shock Cognate 70 Chaperones Carries a Motif that Facilitates Trafficking Through Plasmodesmata. *PNAS* 99:16342-16347.

Aoki K, Suzui, N., Fujimaki, S., Dohmae, N., Yonekura-Sabakibara, K., Fujiwara, T., Hayashi, H., Yamaya, T., Sakakibara, H. (2005) Destination-Selective Long-Distance Movement of Phloem Proteins. *Plant Cell* 17:1801-1814.

Aoki N, Hirose, T., Scofield, G.N., Whitfield, P.R., Furbank, R.T. (2003) The Sucrose Transporter Gene Family in Rice. *Plant Cell Physiol.* 44:223-232.

Arrenberg AB, Del Bene, F., Baier, H. (2009) Optical Control of Zebrafish Behavior with Halorhodopsin. *PNAS* 160(42):17968-17973.

Asano T, Masumura, T., Kusano, H., Kikuchi, S., Kurita, A., SHimada, H., Kadowaki, K.I. (2002) Construction of a Specialized cDNA Library from Plant Cells Isolated by Laser Capture Microdissection, Toward Comprehensive Analysis of the Genes Expressed in the Rice Phloem. *Plant J.* 32:401-408.

Atkins CA, Pate, J.S., Peoples, M.B., Joy, K.W. (1983) Amino Acid Transport and Metabolism in Relation to the Nitrogen Economy of Legume Leaf. *Plant Physiol.* 71:841-848.

Ayre BG, Blair, J.E., Turgeon, R. (2003) Functional and Phylogenetic Analyses of a Conserved Regulatory Program in the Phloem of Minor Veins. *Plant Physiol.* 133:1229-1239.

Bae MS, Cho, E.J., ChoiE-Y., Park, O.K. (2003) Analysis of the *Arabidopsis* Nuclear Proteome and Its Response to Cold Stress. *Plant J.* 36:652-663.

Baerenfaller K, Grossman, J., Grobei, M. A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., Baginsky, S. (2008) Genome-Scale Proteomics Reveals *Arabidopsis thaliana* Gene Models and Proteome Dynamics. *Science* 320:938-941.

Baginsky S, Gruissem, W. (2006) *Arabidopsis thaliana* Proteomics: From Proteome to Genome. . *J. Exp. Bot.* 57(7):1485-1491.

Balachandran S, Xiang, Y., Schobert, C., Thompson, G.A., Lucas, W.J. (1997) Phloem Sap Proteins from *Cucurbita maxima* and *Ricinus communis* Have the Capacity to Traffic Cell to Cell Through Plasmodesmata. *PNAS* 94:14150-14155.

Balasubramanian K, Bevers, E.M., Willems, G.M., Schroit, A.J. (2001) Binding of Annexin 5 to Membrane Products of Lipid Peroxidation. *Biochemistry* 40:8672-8676.

Balmer Y, Koller, A., del Val, G., Manieri, W., Schürmann, P., Buchanan, B.B. (2003) Proteomics Gives Insight into the Regulatory Function of Chloroplast Thioredoxins. *PNAS* 100:370-375.

Barajas-Lo´pez JdD, Serrato, A.J., Olmedilla, A. Chueca, A., Sahrawy, M. (2007) Localization in Roots and Flowers of Pea Chloroplastic Thioredoxin f and Thioredoxin m Proteins Reveals New Roles in Nonphotosynthetic Organs. *Plant Physiol.* 145(3):946-960.

Barnes A, Bale, J., Constantinidou, C., Ashton, P., Jones, A., Pritchard, J. (2004) Determining Protein Identity from Sieve Element Sap in *Ricinus communis* L. by Quadrupole Time of Flight (Q-TOF) Mass Spectrometry. *J. Exp. Bot.* 55(402):1473-1481.

Barth C, Jander, G. (2006) *Arabidopsis* Myrosinases TGG1 and TGG2 have Redundant Function in Glucosinolate Breakdown and Insect Defense. *Plant J.* 46:549-562.

Batailler B, Lemaitre, T., Vilaine, F., Sanchez, C., Renard, D., Cayla, T., Beneteau, J., Dinant, S. (2012) Soluble and Filamentous Proteins in *Arabidopsis* Sieve Elements. *Plant, Cell and Environment.*

Baucher M, Lowe,Y.O., Vandeputte, O.M., Bopopi, J.M., Moussawi, J., Vermeersch, Mol, A., El Jaziri, M., Homblé, F., Pérez-Morga, D. (2011) Ntann12 Annexin Expression is Induced by Auxin in Tobacco Roots. *J Exp. Bot.* 62(11):4055-4065.

Baud S, Wuilleme, S., Lenoine, R., Kronenberger, J., Caboche, M., Lepiniee, L., Rochat, C. (2005) The AtSUC5 Sucrose Transporter Specifically Expressed in the Endosperm is Involved in Early Seed Development in *Arabidopsis*. *Plant J.* 43:824-836.

Baurle I, Dean, C. (2006) The Timing of Developmental Transitions in Plants. *Cell* 125:655-664.

Bayer EM, Bottrill, A.R., Walshaw, J., Vigouroux, M., Naldrett, N.J., Thomas, C.L., Maule, A.J. (2006) *Arabidopsis* Cell Wall Proteome Defined Using Multidimensional Protein Identification Technology. *Proteomics* 6(1):301-311.

Bedinger PA, Hardeman, K.J., Loukides, C.A. (1994) Travelling in Style: The Cell Biology of Pollen. *Trends in Cell Biol.* 4(4):132-138.

Berg M, Parbel., A., Pettersen, H., Fenyo, D., Bjorkertes, L. (2006) Reproducibility of LC-MS-Based protein Identification. *J Exp Bot.* 57(7):1509-1514.

Berkelman T, Brubacher, M.G., Chang, H. (2004) Important factors Influencing Protein Solubility for 2-D Electrophoresis. *BioRadiations* 114:30-32.

Berkowitz O, Jost, R., Pollmann, S., Masle, J. (2008) Characterization of TCTP, the Translationally Controlled Tumor Protein, from *Arabidopsis thaliana*. *Plant Cell* 20:3430-3447.

Bernier G (1988) The Control of Floral Evocation and Morphogenesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 39:175-219.

Bernier G, Perilleux, C. (2005) A Physiological Overview of the Genetics of Flowering Time Control. . *Plant Biotech Journal* 3:3-16.

Bick JA, Neelam, A,, Smith, E., Nelson, S.J., Hall, J.L., Williams, L.E. (1998) Expression Analysis of a Sucrose Carrier in the Germinating Seedling of *Ricinnus communis*. *Plant Mol. Biol.* 38:425-435.

Biddulph O, Cory, R. (1957) An Analysis of translocation in the Phloem of the Bean Plant Using THO, P^{32} and C^{14}. *Plant Physiol.* 32:608-619.

Binz P-A, Barkovich, R., Beavis, R.C., Creasy, D., Horn, D.M., Julian, R.K., Seymour, S.L., Taylor, C.F., Vandenbrouck, Y. (2008) MIAPE:Mass Spectrometry Informatics. In *Minimal Informations About a Proteomics Experiment (MIAPE)*.

Bjellqvist B, Ek, K., Righetti, P.G., Gianazza, E., Gorg, A., Westermeier, R., Postel, W. (1982) Isoelectric Focusing in Immobilized pH Gradients:

Principle, Methodology and Some Applications. *J Biochem Biophys Methods* 6:317-339.

Blazquez MA, Trenor, M., Weigel, D. (2002) Independent Control of Gibberellin Biosynthesis and Flowering Time by the Circadian Clock in *Arabidopsis*. *Plant Physiol.* 130:1770-1775.

Blázquez MA, Weigel, D. (2000) Integration of Floral Inductive Signals in *Arabidopsis*. *Nature* 404:889-892.

Bones AM, Rossiter, J.T. (2006) The Myrosinase-Glycosinolate SYstem, Its Organisation and Biochemistry. *Physiologia Plantarum* 97:194-208.

Borner R, Kampmann, G., Chandler, J., Gleibner, R., Wisman, E., Apel., K. & Melzer, S. (2000) A MADS Domain Gene Involved in the Transition to Flowering in *Arabidopsis*. *Plant J.* 24(5):591-599.

Boss PK, Bastow, R.M., Mylne, J.S., Dean, C. (2004) Multiple Pathways in the Decision to Flower: Enabling, Promoting and Resetting. . *Plant Cell.* 16:S18-S31.

Bourgis F, Roje, S., Nuccio, M.I., Fisher, D.B., Tarczynski, M.C., Li, C., Herschbach, C., Rennenberg, H., Pimenta, M.J., Shen, T.L. (1999) S-Methylmethionine Plays A Major Role in Phloem Sulfur Transport and is Synthesized by a Novel Type of Methyltransferase. *Plant Cell.* 11:1485-1498.

Breaux J, Jones, K., Boulas, P. (2003) Analytical Method Development and Validation: Pharmaceutical Technology.

Breci LA, Tabb, D.L., Yates, J.R., Wysocki, V.H. (2003) Cleavage N-Terminal to Proline: Analsyis of a Database of Peptide Tandem Mass Spectra. *Anal. Chem.* 75:1963-1971.

Bronsius J (1984) Plasmid Vectors for the Selection of Promoters. *Gene* 27:1510160.

Brooking IR, Jamieson, P.D., Porter, J.R. (1995) The Influence of Daylength on Final Leaf Number in Spring Wheat. *Field Crops Research* 41:155-165.

Buhtz A, Springer, F., Chappell, L., Baulcombe, D.C. and Kehr, J. (2007) Identification and Characterization of Small RNAs From the Phloem of *Brassica napus*. *Plant J.* 53:739-749.

Burkle L, Hibberd, J.M., Quick, W.P., Kuhn, C., Hirner, B., Frommer, W.B. (1998) The H^+-Sucrose Cotransporter NtSUT1 is Essential for Sugar Export from Tobacco Leaves. *Plant Physiol.* 118:59-68.

Caillaud CM, Niemeyer, H.M. (1996) Possible Involvement of the Phloem Sealing System in the Acceptance of a Plant as Host by an Aphid. *Experientia.* 52:927-931.

Cans C, Passer, B.J., Shalak, V., Nancy-Portebois, V., Crible, V., Amzallag, N., Allanic, D., Tufino, R., Argentini, M., Moras, D., Fiucci, G., Goud, B., Mirande, M., Amson, R., Telerman, A. (2003) Translationally Controlled Tumor Protein Acts as a Guanine Nucleotide Dissociation Inhibitor on the Translation Elongation Factor eEF1A. *PNAS* 100:13892-13897.

Carter C, Pan, S., Zouhar, J., Avila, E.L., Girke, T., Raikhel, N.V. (2004) The Vegetative Vacuole Proteome of *Arabidopsis thaliana* Reveals Predicted and Unexpected Proteins. *Plant Cell.* 16:3285-3303.

Chabarek S, Martell., A.E. (1959) *Organic Sequestering Agents.* (John Wiley).

Chailakhyan MK (1975) Forty Years of Research on the Hormonal Basis of Plant Development - Some Personal Reflections. . *Botanical Review* 41:1-29.

Chan J, Mao, G., & Lloyd, C. (2003) Proteomic Analysis of Plant microtubule-Associated Proteins. *Cell Biol. Intl.* 27:181.

Chandran D, Reinders, A., Ward, J.M. (2003) Substrate Specificity of the *Arabidopsis thaliana* Sucrose Transporter AtSUC2. . *J. Biol. Chem.* 278:44320-44325.

Chevalier F (2010) Highlights on the Capacity of "Gel-Based" Proteomics. *Proteome Science* 8:1-10.

Chevalier F, Martin, O., Rofidal, V., Devauchelle, A-D., Barteau, S., Sommerer, N., Rossignol, M. (2004) Proteomic Investigation of Natural Variation Between *Arabidopsis* Ecotypes. *Proteomics* 4:1372-1381.

Chino M, Hayashi, H., Fukumorita, T. (1987) Chemical Composition of Rice Phloem Sap and Its Function. *J. Plant Nutrition* 10:1651-1661.

Chivasa S, Ndimba, B.K., SImon, W.J., Robertson, D., Yu, X., Know, J.P., Bolwell, P., Slabas, A.R. (2002) Proteomic Analysis of the *Arabidopsis thaliana* Cell Wall. *Electrophoresis* 23:1754-1765.

Chou IT, Gasser, C.S. (1997) Characterisation of the Cyclophilin Gene Family of *Arabidopsis thaliana* and Phylogenetic Analysis of Known Cyclophilin Proteins. *Plant Mol. Biol.* 35:873-892.

Christeller JT, Farley, P.C., Ramsay, R.J., Sullivan, P.A., Laing ,W.A. (1998) Purification, Characterization and Cloning of an Aspartic Proteinase Inhibitor from Squash Phloem Exudate. *Eur J Biochem* 254:160-167.

Chudakov DM, Verkhusha, V.V., Staroverov, D.B., Souslova, E.A., Lukyanov, S. & Lukyanov, K.A. (2004) Photoswitchable Cyan Fluorescent Protein for Protein Tracking. *Nat. Biotech.* 22(1435-1439):1435.

Chudakov DM, Lukyanov, S. & Lukyanov, K. A. (2007) Using Photoactivable Fluorescent Protein Dendra2 to Track Protein Movement. *Biotechniques* 42(5):553-563.

Clark GB, Sessions, A., Eastburn, D.J., Roux, S.J. (2001) Differential Expression of Members of the Annexin Multigene Family in *Arabidopsis*. *Plant Physiol.* 126:1072-1084.

Clough SJ, Bent, A.F. (1998) Floral Dip: a Simplified Method for *Agrobacterium*-Mediated Transformation of *Arabidopsis thaliana*. *Plant J.* 16(6):735-743.

Colasanti J, Sundaresan, V. (2000) Florigen Enters the Molecular Age: Long Distance Signals that Cause Plants to Flower. *Trends in Biochemical Sciences* 25:236-240.

Colombani A, Djerbi, S., Bessueille, L., Blomqvist, K., Ohlsson, A., Berglund, T., Teeri, T.T., Bulone, V. (2004) *In vitro* Synthesis of (1/3)-b-d-Glucan

(Callose) and Cellulose by Detergent Extracts of Membranes from Cell Suspension Cultures of Hybrid Aspen. *Cellulose* 11:313-327.

Concenco G, Galon, L. (2011) Plasmodesmata: Symplastic Transport of Herbicides within the Plant. *Herbicides, Theory and Applications.*, ed Larramendy PM (InTech).

Cooper B, Eckert, D., Andon, N.L., Yates III, J.R. & Haynes, P.A. (2003) Investigative Proteomics: Identification of an Unknown Plant Virus from Infected Plants using Mass Spectrometry. *J. of the American Society for Mass Spectrometry* 14:736-741.

Corbesier L, Gadisseu, I., Silvestre, G., Jacqmard, A., Bernier, G. (1996) Design in *Arabidopsis thaliana* of a Synchronous System of Floral Induction by One Long Day. *Plant J.* 9(6):947-952.

Corbesier L, Prinsen, E., Jacqmard, A., Lejeune, P., Van Onckelen, H., Perilleux, C., Bernier, G. (2003) Cytokinin Levels in Leaves, Leaf Exudate and Shoot Apical Meristem of *Arabidopsis thaliana* During Floral Transition. *J Exp. Bot.* 54:2511-2517.

Corbesier L, Coupland, G. (2006) The Quest for Florigen: A Review of Recent Progress. *J. Exp. Bot.* 57(13):3395-3403.

Corbesier L, Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I.,Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., Coupland, G. (2007) FT Protein Movement Contributes To Long-Distance Signalling In Floral Induction of *Arabidopsis*. *Science* 316(5827):1030-1033.

Covington MF, Maloof, J.N., Starume, M., Kay, S.A., Harmer, S.L. (2008) Global Transcriptome Analysis Reveals Regulations of Key Pathways in Plant Growth and Development. *Genome Biology* 9:R130.

Cox DM, Zhong, F., Du, M., Duchoslav, E., Sakuma, T., McDermott, J.C. (2005) Multiple Reaction Monitoring as a Method for Identifying Protein Posttranslational Modifications. *J Biomolecular Technique* 16(2):83-90.

Crafts AS, Crisp, C.E. ed (1971) *Phloem Transport in Plants.* (*W.H. Freeman and Company*).

Cresti M, van Went, J.L. (1976) Callose Deposition and Plug Formation in Petunia Pollen Tubes. *In Situ*. *Planta* 133:35-40.

Criqui MC, de Almeida Engler, J., Camasses, A., Capron, A., Parmentier, Y., Inze´, D., Genschik, P. (2002) Molecular Characterization of Plant UbiquitinConjugating Enzymes Belonging to the UbcP4/E2-C/UBCx/UbcH10 Gene Family. *Plant Physiol.* 130(3):1230-1240.

Cronshaw J (1981) Phloem Structure and Function. *Ann. Rev. Plant Physiol.* 32:465-484.

Crosthwaite SK, Jenkins, G.I. (1993) The Role of Leaves in the Perception of Vernalizing Temperatures in Sugar-Beet. *J Exp Bot.* 44:801-806.

Cseke LJ, Zheng, J. & Podila, G.K. (2003) Characterization of PTM5 in Aspen Trees: A MADS-Box Gene Expressed During Woody Vascular Development. *Gene* 318:55-67.

Cullis C (2004) *Plant Genomics and Proteomics.* (John Wiley and Sons).

D'Aloia M, Bonhomme, D., Bouche´, F., Tamseddak, K., Ormenese, S., Torti, S., Coupland, G., Pe´rilleux, C. (2011) Cytokinin Promotes Fowering of *Arabidopsis* via Transcriptional Activation of the FT Paralogue TSF. *Plant J.* 65:972-979.

Dabrowska G, Kata, A., Goc, A., Szechynska-Hebda, M., Skrzypek, E. (2007) Characteristics of the Plant Ascorbate Peroxidase Family. *Acta Biologica Cracoviensia* 49(1):7-17.

Dafoe NJ, Zamani, A., Ekramoddoullah, A.K.M., Lippert, D., Bohlmann, J. & Constabel, C.P. (2009) Analysis of the Poplar Proteome and Its Response to Leaf Wounding. *J of Proteome Research* 8:2341-2350.

Dannenhoffer JM, Suhr, R.C., Thompson, G.A. (2001) Phloem-Specific Expression of the Pumpkin Fruit Trypsin Inhibitor. *Planta* 212:155-162.

Dass, C. (2007) *Fundamentals of Contemporary Mass Spectrometry. (Wiley-Interscience)*.

Davletova S, Rizhsky, L., Liang, H., Shengqiang, Z., Oliver, D.J., Coutu, J., Shulaev, V., Schlauch, K., Mittler, R. (2005) Cytosolic Ascorbate Peroxidase 1 is a Central Component of the Reactive Oxygen Gene Network of *Arabidopsis* . *Plant Cell.* 17:268-281.

de Wet LR, Botha, C.E.J. (2007) Resistance or Tolerance: An examination of Aphid (*Sitobion yakini)* Phloem Feeding on Betta and Betta-Dn Wheat (*Triticum aestivum*). . *South African J. Bot.* 73:35-39.

Decaestecker TN, Vande Casteele, S. R., Wallemacq, P.E., Van Peteghem, C.H., Defore, D.L., Van Bocxlaer, J.F. (2004) Information-Dependent Acquisition-Mediated LC-MS/MS Screening Procedure with Semiquantitative Potential. *Anal. Chem.* 76:6365-6373.

Deeken R, Ache, P., Kajahn, I., Klinkenberg, J., Bringmann, G., Hedrich, R. (2008) Identification of *Arabidopsis thaliana* Phloem RNAs Provides a Criteria for Phloem-Based Transcripts Hidden in Complex Datasets of Microarray Experiments. *Plant J.* 55:746-759.

De-la-Pena C, Badri, D.V., Lei, Z., Watson, B.S., Brandao, M.M., Siva-Filho, M.C., Sumner, L.W., Vivanco, J.M. (2010) Root Secretion of Defense-Related Proteins is Development-Dependent and Correlated with Flowering Time. *J. Biol. Chem. (JBC).* 285:30654-30665.

Devoto A, Muskett, P.R., Shirasu, K. (2003) Role of Ubiquitination in the Regulation of Plant Defence Against Pathogens. *Curr Opin in Plant Biology* 6:307-311.

Di Carli M, Villani, M.E., Bianco, L. (2010) Proteomic Analysis of the Plant Virus In Cucumber Mosaic Virus (CMV) Resistant Transgenic Tomato. *J. Proteome Res.* 9:5684-5696.

Dinant S, Clark, A.M., Zhu, Y., Vilaine, F., Palauqui, J-C., Kusiak, C., Thompson, G.A. (2003) Diversity of the Superfamily of Phloem Lectins (Phloem Protein 2) in Angiosperms. *Plant Physiol.* 131(1):114-128.

Doering-Saad C, Newbury, H.J., Bale, J.S., Pritchard, J. (2002) Use of Aphid Stylectomy and RT-PCR for the Detection of Transporter mRNAs in Sieve Elements. *J Exp Bot.* 53:631-637.

Doering-Saad C, Newbury, H.J., Couldridge, C.E., Bale, J.S., Pritchard, J. (2006) A Phloem-Enriched cDNA Library from Ricinus: Insights into Phloem Function. *J Exp Bot.* 57(3183-3193):3183.

Douglas AE (1993) The Nutritional Quality of Phloem Sap Utilised by Natural Aphid Populations. *Ecology Entomology* 18:31.

Drabovich AP, Diamandis, E.P. (2010) Combinatorial Peptide Libraries Facilitate Development of Multiple Reaction Monitoring Assays for Low-Abundance Proteins. *J of Proteome Research* 9:1236-1245.

Duck N, McCormick, S., Winter, J. (1989) Heat Shock Protein hsp70 cognate Gene Expression in Vegetative and Reproductive Organs of *Lycopersicon esculentum*. *PNAS* 86:3674-3678.

Duckett CM, Oparka, K.J., Prior, D.A.M., Dolan, L., Roberts, K. (1994) Dye-Coupling in the Root Epidermis of *Arabidopsis* is Progressively Reduced During Development. . *Development* 120:3247-3255.

Dunkley TPJ, Hester, S., Shadforth, I.P., Runions, J., Weimar, T., Hantom, S.L., Griffin, J.L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R.B., Dupree, P., Lilley, K.S. (2006) Mapping the *Arabidopsis* Organelle Proteome. *PNAS.* 103(17):6518-6523.

Erwee MG, Goodwin, P.B. (1985) Symplast Domains in Extrastellar Tissues of *Egeria densa* Planch. *Planta* 163:9-19.

Erwee MG, Goodwin, P.B. (1983) Characterization of the *Egeria densa* Planch. Leaf Symplast. *Planta* 158(320-328):320.

Erwin JE (2006) Flower Induction of Ornamental Plants. *Flower Breeding and Genetics: Issues, Challenges.*, ed Anderson N (Springer), pp 7-48.

Eulgem T, Rushton, P.J., Robatzek, S., Somssich, I.E. (2000) The WRKY Superfamily of Plant Transcription Factors. *Trends in Plant Sci.* 5(5):199-206.

Evert RF (1977) Phloem Structure and Histochemistry. *Ann. Rev. Plant Physiol.* 28:199-222.

Fellows RJ, Egli, D.B., Leggett, J.E. (1978) A Pod Leakage Technique for Phloem Translocation Studies in Soybean (*Glycine max* (L.) Merr.). *Plant Physiol.* 62:812-814.

Fellows RJ, Zeevart, J.A.D. (1983) Comparison of Ethylenediaminetetraacetate-Enhanced Exudation from Detached and Translocation from Attached Bean Leaves. *Plant Physiol.* 71:716-718.

Feng B, Li, L., Zhou, X., Stanley, B., Ma, H. (2009) Analysis of the *Arabidopsis* Floral Proteome: Detection of over 2 000 Proteins and Evidence for Posttranslational Modifications. *Journal of Integrative Plant Biology.* 51(2):207-223.

Fenyo D (2000) Identifying the Proteome:Software Tools. *Curr. Opinion in Biotech.* 11:391-395.

Ferrandiz C, Gu, Q., Martienssen, R., Yanofsky, M.F. (2000) Redundant Regulation of Meristem Identity and Plant Architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development* 127:725-734.

Ferrario S, Busscher, J., Franken, J., Gerats, T., Vandenbussche, M., Angenent, G.C. & Immink, R.G. (2004) Ectopic Expression of the Petunia MADS Boz Gene *UNSHAVEN* Accelerates Flowering and Confers Leaf-Like Characteristics to Floral Organs in a Dominant-Negative Manner. *Plant Cell* 16:1490-1505.

Fisher DB, Frame, J.M. (1984) A Guide to the use of Exuding-Stylet Technique in Phloem Physiology. *Planta* 161:385-393.

Foster TM, Lough, T.J., Emerson, S.J., Lee, R.H., Bowman, J.L.,Forster, R.L.S., Lucas, W.J. (2002) A Surveillance System Regulates Selective Entry of RNA into the Shoot Apex. *Plant Cell* 14:1497-1508.

Franks SJ, Sim, S., Weis, A.E. (2007) Rapid Evolution of Flowering Time by Annual Plant in Resposne to a Climate Fluctuation. *PNAS* 104:1278-1282.

Friedman R, Levin, N., Altman , A. (1986) Presence and Identification of Polyamines in Xylem and Phloem Exudates of Plants. *Plant Physiol.* 82:1154-1157.

Friso G, Giacomelli, L., Ytterberg, A.J., Peltier, J-B., Rudella, An., Sun, Q., van Wijk, K.J. (2004) In-Depth Analysis of Thylakoid Membrane Proteome of *Arabidopsis thaliana* Chloroplast: New Proteins, New Functions and a Plastid Proteome Database. *Plant Cell.* 16:478-499.

Fukao Y, Hayashi, M., Nishimura, M. (2002) Proteomic Analysis of Leaf Peroxisomal Proteins in Greening Cotyledons of *Arabidopsis thaliana*. *Plant Cell Physiol.* 43(7):689-696.

Furch ACU, Hafke, J.B., Schulz, A., van Bel, A.J.E. (2007) Ca^{2+}-Mediated Remote Control of Reversible Sieve Tube Occlusion in *Vicia faba*. . *J Exp. Bot.* 58:2827-2838.

Furch ACU, Zimmermann, M.R., Will, T., Hafke, J.B., van Bel, A.J.E. (2010) Remote-Controlled Stop of Phloem Mass Fow by Biphasic Occlusion in *Cucurbita maxima*. *J Exp. Bot.* 61(13):3697-3708.

Gallardo K, Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J., Job, D. (2001) Proteomic Analysis of *Arabidopsis thaliana* Seed Germination and Priming. *Plant Physiol.* 126:835-849.

Galloway LF, Burgess, K.S. (2012) Artificial Selection on Flowering Time: Influence on Reproductive Phenology Across Natural Light Environments. *Journal of Ecology*.

Garfin D, Heerdt, L. (2004) *2-D Electrophoresis for Proteomics: A Methods and Product Manual* (BioRad Laboratories, Inc.).

Geiger D (2011) Plant Sucrose Transporters from a Biophysical Point of View. *Mol. Plant* 4(3):395-406.

Gessler A, Rennenberg, H., Keitel, C. (2004) Stable Isoptope Composition of Organic Compounds Transported in the Phloem of European Beech -

Evaluation of Different Methods of Phloem Sap Collection and Assessment of Gradients in Carbon Isotope Composition during Leaf-to Stem Transport. *Plant Biol.* 6:721-729.

Gevaert K, Vandekerckhove, J. (2000) Protein Identification Methods in Proteomics. *Electrophoresis* 21:1145-1154.

Gharahdaghi F, Weinberg, C.R., Maegher, D.A., Imai, B.S., Mische, S.M. (1999) Mass Spectrometric Identification of Proteins from Silver-Stained Polyacrylamide Gel: A Method for the Removal of Silver Ions to Enhance Sensitivity. *Electrophoresis* 20:601-605.

Giavalisco P, Kapitza, K., Kolasa, A., Buhtz, A., Kehr, J. (2006) Towards The Proteome of *Brassica napus* Phloem Sap. *Proteomics* 6(3):896-909.

Gidrol X, Sabelli, P.A., Fern, Y.S., Kush, A.K. (1996) Annexinlike Protein from *Arabidopsis thaliana* Rescues ΔoxyR Mutant of *Escherichia coli* from H_2O_2 Stress. *PNAS* 93:11268-11273.

Gilliland LU, Pawloski, L.C., Kandasamy, M.K., Meagher, R.B. (2003) *Arabidopsis* Actin Gene ACT7 Plays an Essential Role in Germination and Root Growth. *Plant J.* 33:319-328.

Giraud M, Cavelier, F., Martinez, J. (1999) A Side-Reaction in the SPPS of Trp-Containing Peptides. *J. Pept. Sci.* 5:457-461.

Girousse C, Bonnemain, J.L., Delrot, S., Bournoville, R. (1991) Sugar and Amino Acid Composition of Phloem Sap of *Medicago sativa*: A Comparative Study of Two Collecting Methods. *Plant Physiol. Biochem.* 29:41-48.

Glover B ed (2007) *Understanding Flowers and Flowering: An Integrated Approach.* (*Oxford University Press*).

Golecki B, Schulz, A., Thompson, G.A. (1999) Translocation of Structural P Proteins in the Phloem. *Plant Cell.* 11:127-140.

Gomez G, Torres, H., Pallas, H. (2005) Identification of Translocatable RNA-Binding Phloem Proteins from Melon, Potential Components of the Long-Distance RNA Transport System. *Plant J.* 41(1):107-116.

Goodsell DS (2003) Trypsin.

Gorg A (2004) 2-D Electrophoresis: Principles and Methods. (GE Healthcare).

Gottwald JR, Krysan, P.J., Young, J.C., Evert, R.F., Sussman, M.R. (2000) Genetic Evidence for the *In Planta* Role of Phloem Specific Plasma Membrane Sucrose Transporters. *PNAS* 97:13979-13984.

Greenbaum D, Colangelo, C. Williams, K., Gerstein, M. (2003) Comparing Protein Abundance and mRNA Expression Levels on a Genomic Scale. . *Genome Biology* 4:117.

Guelette BS, Benning, U.F., Hoffmann-Benning, S. (2012) Identification of lipids and lipid-binding proteins in phloem exudates from *Arabidopsis thaliana. J Exp. Bot.*

Gurskaya NG, Verkhusha, V.V., Scheglov, A.S., Staroverov, D.B., Chepurnykh, T.V., Fradkov, A.F., Lukyanov, S., Lukyanov, K.A. (2006) Engineering of

a Monomeric Green-to-Red Photoactivatanle Fluorescent Protein Induced by Blue Light. *Nat. Biotech.* 24:461-465.

Haebel S, Kehr, J. (2001) Matrix-Assisted Laser Desorption / Ionization Time of Flight Mass Spectrometry Peptide Mass Fingerprints and Post Source Decay: A Tool for the Identification and Analysis of Phloem Protein from *Cucurbita maxima* Duch. Separated by Two-Dimensional Polyacrylamide Gel Electrophoresis. *Planta* 213:586-593.

Hancock JT, Henson, D., Nyirenda, M., Desikan, R., Harrison, J., Lewis, M., Hughes, J., Neill, S.J. (2005) Proteomic Identification of Glyceraldehyde 3-Phosphate Dehydrogenase as an Inhibitory Target of Hydrogen Peroxide in *Arabidopsis*. *Plant Physiol. Biochem.* 43:828-835.

Haritatos E, Ayre, B.G., Turgeon, R. (2000) Identification of Phleom Involved in Assimilate Loading in Leaves by the Activity of the Galactinol Synthase Promoter. *Plant Physiol.* 123(3):929-937.

Harmer CL (2009) The Circadian System in Higher Plants. *Annu. Rev. Plant Biol.* 60:357-377.

Hay RKM, Delecolle, R. (1989) Leaf Appearance and Extension in Field-Grown Winter Wheat Plants:The Importance of Soil Temperature During Vegetative Growth. *Annals of Applied Biol.* 115:333-341.

Hay RKM, Kirby, E.J.M. (1991) Convergence and Synchrony - A Review of the Coordination of Development in Wheat. *Aust. J. Agric. Res.* 42:661-700.

Hayashi H, Chino, M. (1986) Collection of Pure Phloem Sap from Wheat and Its Chemical Composition. *Plant Cell Physiol.* 27:1387-1393.

Hayashi H, Fukuda, A., Suzui, N., Fujimaki, S. (2000) Protein in the Sieve Element-Companion Cell Complexes: Their Detection, Localisation and Possible Functions. *Aust. J. Plant Physiol.* 27:489-496.

Haywood V, Yu, T-S., Huang, N-C. & Lucas, W.J. (2005) Phloem Long-Distance Trafficking of *GIBBERELLIN ACID-INSENSITIVE* RNA Regulates Leaf Development. *Plant J.* 42(1):49-68.

Heath MC (2002) Cellular Interactions between Biotrophic Fungal Pathogens and Host or Nonhost Plants. . *Can. J. Plant Pathol.* 24:259-264.

Heide OM, Hay, R.K.M., Baugerod, H. (1985) Specific Daylength Effcets on Leaf Growth and Dry-Matter Production in High-Latitude Grasses. *Annals of Botany* 55:579-586.

Heim R, Prasher, D.C., Tsien, R.Y. (1994) Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *PNAS* 19:12501-12504.

Heizmann CW, Arnold, E.M., Kuenzle, C.C. (1980) Fluctuations of Non-Histone Chromosomal Proteins in Differentiating Brain Cortex and Cerebellar Neurons. *J Biol Chem* 255:11504-11511.

Helden MV, Tjallingii, W.F., van Beck, T.A. (1994) Phloem Sap Collection from Lettuce (*Lactuca sativa* L.): Chemical Comparison Among Methods. *J. Chemical Ecology* 20:3191-3206.

Henderson IR, Dean, C. (2004) Control of *Arabidopsis* Flowering: The Chill Before the Bloom. *Development* 131:3829-3838.

Herbert BR, Molloy, M.P., Gooley, A.A., Walsh, W.G., Williams, K.L. (1998) Improved Protein Solubility in Two-Dimensional Electrophoresis Using Tributyl Phosphine as Reducing Agent. *Electrophoresis* 19:845-851.

Hoffmann-Benning S, Gage, D.A., McIntosh, L., Kende, H. and Zeevaart, J.A.D. (2002) Comparison of Peptides in the Phloem Sap of Flowering and Non-Flowering Perilla and Lupine Plants using Microbore HPLC Followed by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. . *Planta* 216:140-147.

Hofmann A (2004) Annexins in the Plant Kingdom: Perspectives and Potentials. *Annexins* 1:51-61.

Holtgrefe S, Gohlke, J., Starmann, J., Druce, S., Klocke, S., Altmann, B., Wojtera, J., Lindermayr, C., Scheibe, R. (2008) Regulation of Plant Cytosolic Glyceraldehyde 3-Phosphate Dehydrogenase Isoforms by Thiol Modifications. *Plant Physiol.* 133:211-228.

Hong Z, Zhang, Z., Olson, J.M., Verma, D.P.S. (2001) A Novel UDP-Glucose Transferase is Part of the Callose Synthase Complex and Interacts with Phragmoplastin at the Forming Cell Plate. *Plant Cell.* 13:769-779.

Huang H, Rabenstein, D.L. (1999) A Cleavage Cocktail for Methionine-Containing Peptides. *J. Peptide Res.* 53:548-553.

Huang T, Bohlenius, H., Eriksson, S., Parcy, F., Nilsson, O. (2005) The mRNA of the *Arabidopsis* Gene FT Moves from Leaf to Shoot Apex and Induces Flowering. *Science* 309:1694-1696.

Hunter TC, Andon, N.L., Koller, A., Yates, J.R. III, Haynes, P.A. (2002) The Functtional Proteomics Tool Box: Methods and Applications. *J Chromatography B* 782:165-182.

Hurkman WJ, Tanaka, C.K. (2007) High-Resolution Two-Dimensional Gel Electrophoresis: A Cornerstone of Plant Proteomics. *Plant Proteomics, Eds*:15.

Ideker T, Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., & Bumgarner R, Goodlett, D.R., Aebersold, R., Hood, L. (2001) Integrated Genomic and Proteomic Analysis of a Systematically Pertubed Metabolic Network. *South African J. Bot.* 292:929-934.

Iida H, Yahara, I. (1985) Yeast Heat-Shock Protein of M, 48,000 is an Isoprotein of Enolase. *Nature* 315:688-690.

Imlau A, Truemit, E., Sauer, N. (1999) Cell-to-Cell and Long-Distance Trafficking of the Green Fluorescent Protein in the Phloem and Synmplastic Unloading of the Protein into Sink Tissues. . *Plant Cell* 11:309-322.

Irshad M, Canut, H., Borderies, G., Pont-Lezica, R., Jamet, E. (2008) A New Picture of Cell Wall Protein Dynamics in Elongating Cells of *Arabidopsis thaliana*: Confirmed Actors and Newcomers. *BMC Plant Biology* 8:94.

Ishiwatari Y, Honda, C., Kawashima, I., Nakamura, S., Hirano, H., Mori, S., Fujiwara, T., Hayashi, H., Chino, M. (1995) Thioredoxin h is One of the Major Proteins in Rice Phloem Sap. *Planta* 195:456-463.

Ishiwatari Y, Fujiwara, T., McFarland, K.C., Nemoto, K., Hayashi, H., Chino, M., Lucas, W.J. (1998) Rice Phloem Thioredoxin h Has the Capacity to Mediate Its Own Cell-to-Cell Transport Through Plasmodesmata. *Planta* 205:12-22.

Ivashikina N, Deeken, R., Ache, P., Kranz, E., Pommerrenig, B., Sauer, N., Hedrich, R. (2003) Isolation of AtSUC2 Promoter–GFP-Marked Companion Cells for Patch–Clamp Studies and Expression Profiling. *Plant J.* 36:931-945.

Jacobs AK, Lipka, V., Burton, R.A., Panstruga, R., Strizhov, N., Schulze-Lefert, P. (2003) An *Arabidopsis* Callose Synthase, GSL5, is Required for Wound and Papillary Callose Formation. . *Plant Cell.* 15:2503-2513.

Jaeger KE, Wigge, P. A. (2007) FT Protein Acts as a Long-Range Signal in *Arabidopsis*. . *Curr. Biol.* 17:1050-1054.

Jang S, Torti, S., Coupland, G. (2009) Genetic and Spatial Interactions Between *FT, TSF* and *SVP* During the Early Stages of Floral Induction in *Arabidopsis*. *Plant J.* 60:614-625.

Jeong S, Clark, S.E. (2005) Photoperiod Regulates Flower Meristem development in *Arabidopsis thaliana*. *Genetics* 169:907-915.

Jha D (2009) Development of a New Cell Penetrating Peptide: Design, Synthesis and Applications. PhD (University of Tubingen).

Johnson RPC (1978) The Microscopy of P-protein Filaments in Freeze-etched Sieve Pores. *Planta* 143(2):191-205.

Kandasamy MK, Gilliland, L.U., McKinney, E.C., Meagher, R.B. (2001) One Plant Actin Isovariant, ACT7, Is Induced by Auxin and Required for Normal Callus Formation. *Plant Cell.* 13:1541-1554.

Kardailsky I, Shukla, V., Ahn, J., Dagenais, N., Christensen, S., Nguyen, J., Chory, J, Harrison, M., Weigel, D. (1999) Activation Tagging of the Floral Inducer *FT*. . *Science* 261:230-232.

Karpinski S, Reynolds, H., Karpinska, B., Wingsle, G., Creissen, G., Mullineaux, P. (1999) Systemic Signaling and Acclimation in Response to Excess Excitation Energy in *Arabidopsis*. *Science* 284:654-657.

Kawamura Y, Uemura, M. (2003) Mass Spectrometric Approach for Identifying Putative Plasma Membrane Protein of *Arabidopsis* leaves Associated with Cold Acclimation. *Plant J.* 36:141-154.

Kehr J, Haebel, S., Blechschmidt-Schneider, S., Willmitzer, L., Steup, M., Fisahn, J. (1999) Analysis of Phloem Protein Patterns from Different Organs of *Cucurbita maxima* Duch. by Matrix-Assisted Laser Desoption/Ionization Time of Flight Mass Spectrometry Combined with a Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis. *Planta* **207:612-619.**

Kehr J (2006) Phloem Sap proteins: Their Identities and Potential Roles in the Interaction **between Plants and Phloem-Feeding Insects. . *J. Exp. Bot.* 57(4):767-774.**

Kim I, Zambryski, P.C. (2005) Cell-to-Cell Communication Via Plasmodesmata During *Arabidopsis* Embryogenesis. *Curr. opinion in Plant Biol.* 8:593-599.

Kim JS, Park, S.J., Kwak, K.J., Kim, Y.O., Kim, J.Y., Song, J., Jang, B., Jung, C-H., Kang, H. (2007) Cold Shock Domain Proteins and Glycine-Rich RNA-Binding Proteins from *Arabidopsis thaliana* Can Promote the Cold Adaptation Process in *E. coli*. *Nucleic Acids Research* 35:506-516.

Kim JS, Jung, H.J., Lee, H.J., Kim, K.A., Goh, C-H., Woo, Y., Oh, S.H., Han, Y.S., Kang, H. (2008) Glycine-Rich RNA-Binding Protein7 Affects Abiotic Stress Responses by Regulating Stomata Opening and Closing in *Arabidopsis thaliana*. *Plant J.* 55:455-466.

Kim JY, Kim, W.Y., Kwak, K.J., Oh, S.H., Han, Y.S., Kang, H. (2010) Glycine-Rich RNA-Binding Proteins are Functionally Conserved in *Arabidopsis thaliana* and *Oryza sativa* During Cold Adaptation Process. *J Exp. Bot.* 61(9):2317-2325.

Kim M, Canio, W., Kessler, S., Sinha, N. (2001) Developmental Changes Due to Long-Distance Movement of a Homeobox Fusion Transcript in Tomato. *Science* 293:287-289.

King RW, Zeevart, A.D. (1974) Enhancement of Phloem Exudation from Cut Petioles by Chelating Agents. *Plant Physiol.* 53:96-103.

Kirby EJM (1990) Co-Ordination of Leaf Emergence and Leaf and Spikelet Primordium Initiation in Wheat. *Field Crops Research* 25:253-264.

Kitteringham NR, Jenkins, R.E., Lane, C.S., Elliott, V.L., Kevin Park, B. (2009) Multiple Reaction Monitoring for Quantitative Biomarker Analysis in Proteomics and Metabolomics. *J Chromatography B* 877:1229-1239.

Kleffmann T, Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W., Baginsky, S. (2004) The *Arabidopsis thaliana* Chloroplast Proteome Revelas Pathway Abundance and Novel Protein Functions. . *Curr. Biol.* 14:362-375.

Knoblauch M, van Bel., A.J.E. (1998) Sieve Tube in Action. *Plant Cell* 10(1):35-50.

Knoblauch M, Peters, W.S., Ehlers, K., van Bel, A.J.E. (2001) Reversible Calcium-Regulated Stopcocks in Legume Sieve Tubes. . *Plant Cell.* 13:1221-1230.

Knoblauch M, Noll, G.A., Müller, T., Prüfer, D., Schneider-Hüther, I., & Scharner D, van Bel, A.J.E., Peters, W.S. (2003) ATP Independent Contractile Proteins from Plants. *Nature Materials* 2(600-603):600.

Kobayashi Y, Kaya, H., Goto, K., Iwabuchi, M., Araki, T. (1999) A Pair of Related Genes with Antagonistic Roles in Mediating Flowering Signals. *Science* 286:1960-1962.

Kobayashi Y, & Weigel, D. (2007) Move On Up, It's Time For Change Mobile Signals Controlling Photoperiod-Dependent Flowering. . *Genes & Dev.* 21:2371-2384.

Koornneef M, Alonso-Blanco, C., Peeters, A.J.M., Soppe, W. (1998) Genetic Control of Flowering Time in *Arabidopsis*. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 499:345-370.

Kragler F (2010) RNA in the Phloem: A Crisis Or A Return on Investment? *Plant Sci.* 178:99-104.

Kragler, F (2012) *Short and Long Distance Signaling*. (Springer).

Krenz B, Windeisen, V., Wege, C., Jeske, H., Kleinow, T. (2010) A Plastid-Targeted Heat Shock Cognate 70 kDa Protein Interacts with the Abutilon Mosaic Virus Movement Protein. *Virology* 401(1):6-17.

Krömer S (1995) Respiration During Photosynthesis. *Ann Rev Plant Physiol Plant Mol Biol* 46:45-70.

Kruft V, Eubel., H., Jansch, L., Werhahn, W., Braun, H.P. (2001) Proteomic Approach to Identify Novel Mitochondrial Proteins in *Arabidopsis*. *Plant Physiol.* 127:1694-1710.

Labas YA, Gurskaya, N.G., Yanushevich, Y.G., Fradkov, A.F., Lukyanov, K.A., Lukyanov, S.A., Matz, M.V. (2002) Diversity and the Evolution of the Green Fluorescent Protein Family. *PNAS* 99:4256-4261.

Lacey EP (1986) Onset of Reproduction in Plants: Size- Versus Age- Dependency. *Trends Ecol. Evol.* 1:72-75.

Laloi C, Mestres-Ortega, D., Marco, Y., Meyer, Y., Reichheld, J.P. (2004) The *Arabidopsis* Cytosolic Thioredoxin h5 Gene Induction by Oxidative Stress and Its W-Box-Mediated Response to Pathogen Elicitor. *Plant Physiol.* 134:1006-1016.

Lalonde S, Tegeder, M., Throne-Holst, M., Frommer, W.B., Patrick, J.W. (2003) Phloem Loading and Unloading of Sugars and Amino Acids. . *Plant, Cell & Environment* 26:37-56.

Lange V, Picotti, p., Domon, B., Aebersold, R. (2008) Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial. *Molecular Systems Biol.* 4:1-14.

Le Hir R, Beneteau, J., Bellini, C., Vilaine, F., Dinant, S. (2008) Gene Expression Profiling: Keys for Investigating Phloem Functions. *Trends in Plant Sci.* 13:273-280.

Lee H, Su., S-S., Park, E., Cho, E., Ahn, J.H., Kim, S-G., Lee, J.S., Kwon, Y.M., Lee, I. (2000) The AGAMOUS-LIKE 20 MADS Domain Protein Integrates Floral Inductive Pathways in *Arabidopsis. Genes & Dev.* 14:2366-2376.

Lee J, Oh, M., Park, H., Lee, I. (2008) SOC1 Translocated to the Nucleus by Interaction with AGL24 Directly Regulates LEAFY. *Plant J.* 55:832-843.

Lee J, Lee, I. (2010) Regulation and Function of SOC1, a Flowering Pathway Integrator. *J. Exp. Bot.* 61(9):2247-2254.

Lemoine R, Kuhn, C., Thiele, N., Delrot, S., Frommer, W.B. (1996) Antisense Inhibition of the Sucrose Transporter in Potato: Effects on Amount and Activity. *Plant, Cell and Environment.* 19:1124-1131.

Lemoine R, Burkle, L., Barker, L., Sakr, S., Kuhn, C., Regnacq, M., Gaillard, C., Delrot, S., Frommer, B.W. (1999) Identification of A Pollen-Specific Sucrose Transporter-Like Protein NtSUT3 from Tobacco. *FEBS Letters.* 454:325-330.

Levy A, Guenoune-Gelbart, D., Epel, B.L. (2007) b-1,3-Glucanases. Plasmodesmal Gate Keepers for Intercellular Communication. . *Plant, Signaling and Behavior* 5:288-290.

Levy YY, Dean, C. (1998) The Transition to Flowering. *Plant Cell* 10:1973-1989.

Li Y, Ren, J., Cho, M.-J., Zhou, S., Kim, Y.B., Guo, H., Wong, J.H., Niu, H., Kim, H.-K., Morigasaki, S., Lemaux, P.G., Frick, O.L., Yin, J., Buchanan, B.B. (2009) The Level of Expression of Thioredoxin is Linked to Fundamental Properties and Applications of Wheat Seeds. *Molecular Plant* 2:430-441.

Lifschitz E, Eviatar, T., Rozman, A., Shalit, A., Goldshmidt, A., Amsellem, Z., ALvarez, J.P., Eshed, Y. (2006) The Tomato *FT* Ortholog TriggersSystemic Signals that Regulate Growth and Flowering and Substitute for Diverse Environmental Stimuli. *PNAS* 103:6398-6403.

Lin M, Belanger, H, Lee YJ, Varkonyi-Gasic, E, Taoka, KI, Miura, E, Xoconostle-Ca´ zares, B, Gendler, K, Jorgensen, RA, Phinney, B. (2007) *FLOWERING LOCUS T* Protein May Act As The Long Distance Florigenic Signal In The Cucurbits. *Plant Cell* 19:1488-1506.

Lin M-K, Lee, Y-J., Lough, T.J., Phinney, B.S. & Lucas, W.J. (2009) Analysis of the Pumpkin Phloem Proteome Provides Insights into Angiosperm Sieve Tube Function. *Molecular & Cellular Proteomics* 8(2):343-356.

Lippincott-Schwartz J, Patterson, G.H. (2003) Development and Use of Fluorescent Protein Markers in Living Cells. *Science* 300:87-91.

Lippincott-Schwartz J, Patterson, G.H. (2009) Photoactivatable Fluorescent Proteins for Diffraction-Limited and Super-Resolution Imaging. *Trends in Cell Biology* 19(11):555-565.

Liu C, Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T., Yu, H. (2007) Specificatin of *Arabidopsis* Floral Meristem identity br Repression of Flowering Time Genes. *Development* 134:1901-1910.

Liu H, Sadygov R.G., Yates, J.R., III, A. (2004) Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* 76:4193-4201.

Liu YQ, Li, Y. P., Wang, ., Xia, R.X., Li, X.S., Wan, H.L., Qin, L., Jiang, D.F., Lu, C., Xiang, Z. H. (2010) cDNA Cloning and Expression Pattern of Two Enolase Genes from the Chinese Oak Silkworm, *Antheraea pernyi*. *Acta Biochem. Biophys. Sin.* 42:816-826.

Lough TJ, Lucas, W.J. (2006) Integrative Plant Biology: Role of Phloem Long-Distance Macromolecular Trafficking. *Ann. Rev. Plant Physiol.* 57:203-232.

Lucas WJ, Bouche-Pillon, S., Jackson, D.P., Nguyen, L., Baker, L., Ding, B., Hake, S. (1995) Selective Trafficking of KNOTTED1 Homeodomain Protein and Its mRNA Through Plasmodesmata. *Science* 270:1980-1983.

Lukyanov KA, Chudakov, D.M., Lukyanov, S., Verkhusha, V.V. (2005) Photoactivatable Fluorescent Proteins. *Molecular & Cell Biology* 6(885-891):885.

MacDonald SM, Rafnar, T., Langdon, J., Lichtenstein, L. M. (1995) Molecular Identification of an IgE-Iependent Histamine-Releasing Factor. *Science* 269:688-690.

Malter D, Wolf, S. (2011) Melon Phloem-Sap Proteome: Developmental Control and Response to Viral Infection. *Protoplasma* 248:217-224.

Mann M, Hendrickson, R.C., Pandey, A. (2001) Analysis of Proteins and Proteomes by Mass Spectrometry. . *Ann. Rev. Biochem.* 70:437-473.

Marmagne A, Rouet, M.A., Ferro, M., Rolland, N., Alcon, C., Joyard, J., Garin, J., Barbier-Brygoo, H., Ephritikhine, G. (2004) Identification of New Intrinsic Proteins in *Arabidopsis* Plasma Membrane Proteome. *Mol Cell Proteomics* 3(7):675-691.

Martin K, Kopperud, K., Chakrabarty, R., Banerjee, R., Brooks, R., Goodin M.M. (2009) Transient Expression in *Nicotiana benthamiana* Fluorescent Marker Lines Provides Enhanced Definition of Protein Localization, Movement and Interactions *in planta*. *Plant J.* 59:150-162.

Martinoia E, Rentsch, D. (1994) Malate Compartmentation - Responses to a Complex Metabolism. *Ann Rev Plant Physiol Plant Mol Biol* 45:447-467.

Marx C, Wong, J.H., Buchanan, B.B. (2003) Thioredoxin and Germinating Barley: Targets and Protein Redox Changes. *Planta* 216:454-460.

Mathieu J, Warthmann, N., Kuttner, F., Schmid, M. (2007) Export of FT Protein from Phloem Companion Cell is Sufficient for Floral Induction in *Arabidopsis*. *Curr. Biol.* 17:1055-1060.

Matz MV, Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L. & Lukyanov, S.A. (1999) Fluorescent Proteins from Nonbioluminescent *Anthozoa* Species. . *Nat. Biotech.* 17(969-973):969.

Mayfield JD, Folta, K.M., Paul, A-L., Ferl, R.J. (2007) The 14-3-3 Proteins m and u Influence Transition to Flowering and Early Phytochrome Response. *Plant Physiol.* 145:1692-1702.

McClung CR (2006) Plant Circadian Rhythms. *Plant Cell* 18:792-803.

McDowell JM, An, Y.-Q., McKinney, E.C., Huang, S., Meagher, R.B. (1996) The *Arabidopsis* ACT7 Actin Gene is Expressed in Rapidly Developing Tissues and Responds to Several External Stimuli. *Plant Physiol.* 111:699-711.

Meagher RB, McKinney, E.C., Vitale, A.V. (1999) The Evolution of New Structures: Clues from Plant Cytoskeletal Genes. *Trends Genet.* 15:278-284.

Melzer S, Lens, F., Gennen, J., Vanneste, S., Rohde, A., Beeckman, T. (2008) Flowering-Time Genes Modulate Meristem Determinacy and Growth Form in *Arabidopsis thaliana*. *Nature Genetics* 40(12):1489-1492.

Merzlyak EM, Goedhart, J., Shcherbo, D. (2007) Bright Monomeric Red Fuorescent Protein With An Extended Fuorescence Lifetime. *Nat. Methods.* 4:555-557.

Metzger JD (1988) Localization of the Site of Perception of thermoinductive Temperatures in Thlaspi-Arvense L. *Plant Physiol.* 88:424-428.

Michaels SD, Himelblau, E., Kim, S.Y., Schomburg, Amasino, R.M. (2005) Integration of Flowering Signals in Winter-Anual *Arabidopsis*. *Plant Physiol.* 137:149-156.

Miglietta F (1989) Effect of Photoperiod and Temperature on Leaf Initiation Rates in Wheat (*Triticum* spp.). *Field Crops Research* 21:121-130.

Miller AH, Sweetlove, L.J., Geige, P., Leaver, C.J. (2001) Analysis of *Arabidopsis* Mitochondrial Proteome. *Plant Physiol.* 127:1711-1727.

Mittler TE (1953) Amino Acids in Phloem Sap and Their Excretion by Aphids. *Nature* 172(4370):207.

Mizuno H, Mal, T.K., Tong, K.I., Ando, R., Furuta, T., Ikura, M., Miyawaki, A. (2003) Photo-Induced Peptide Cleavage in the Green-to-Red Conversion of a Fluorescent Protein. *Mol. Cell* 12:1051-1058.

Mo M, Tse, Y.C., Jiang, L. (2003) Organelle Identification and Proteomics in Plant Cells. *Trends in Biotech.* 21(8):331-332.

Mocz G (2007) Fluorescent Proteins and Their use in Marine Biosciences, Biotechnology and Proteomics. . *Marine Biotech.* 9:305-328.

Moffatt BA, Wang, L., Allen, M.S., Stevens, Y.Y., Qin, W., Snider, J., von Schwartzenberg, K. (2000) Adenosine Kinase of *Arabidopsis*. Kinetic Properties and Gene Expression. *Plant Physiol.* 12(4):1775-1785.

Moon J, Suh, S.S., Lee, H., Choi, K.R., Hong, C.B., Paek, N.C., Kim, S.G., Lee, I. (2003) The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in *Arabidopsis*. *Plant J.* 35(5):613-623.

Moon J, Lee, H., Kim, M., Lee, I. (2005) Analysis of Flowering Pathway Integrators in *Arabidopsis*. *Plant Cell Physiol.* 46(2):292-299.

Mortimer. J.C. L, A., Macpherson, N., Webb, A., Brownlee, C., Battey, N.H., Davies, J.M. (2008) Annexins: Multifunctional Components of Growth and Adaptation. *J Exp. Bot.* 59(3):533-544.

Munier-Jolain N, Salon, C. (2003) Can Sucrose Content in the Phloem Sap Reaching Field Pea seeds (Pisum sativum L.) be an Accurate Indicator of Seed Growth Potential? *J. Exp. Bot.* 54:2457-2465.

Nagawa S, Xu, T., Lin, D., Dhonukshe, P., Zhang, X., Friml, J., Scheres, B., Fu, Y.,Yang, Z. (2012) ROP GTPase-Dependent Actin Microfilaments Promote PIN1 Polarization by Localized Inhibition of Clathrin-Dependent Endocytosis. *PLoS Biol* 10(4).

Najla S, Vercambre, G., Genard, M. (2010) Improvement of the Enhanced Phloem Exudation Technique to Estimate Phloem Concentration and Turgor Pressure in Tomato. *Plant Science* 179:316-324.

Nakamura S, Hayashi, H., Mori, S., Chino, M. (1993) Protein Phosphorylation in the Sieve Tubes of Rice Plants. *Plant Cell Physiol.* 34:927-933.

Nakamura S, Hayashi, H., Mori, S., Chino, M. (2006) Detection of Calmodulin and Calmodulin-Binding Proteins in Pure Phloem Sap of Rice Plants. *Soil Sci. & Plant Nutrition* 52:195-202.

Nakamura T, Song, I.J., Fukuda, T., Yokohama, J., Maki, M., Ochiai, T., Kameya, T., Kanno, A. (2005) Characterization of *TrcMADS1* Gene of *Trillium camtschatcense* (Trilliaceae) Revelas Functional Evolution of the *SOC1/TM3*-Like Gene Family. *Journal of Plant Research* 118:229-234.

Nakashima J, Laosinchai, W., Cui, X., Brown, R.M. Jr. (2003) New Insight into the Mechanism of Cellulose and Callose Biosynthesis: Proteases May Regulate Callose Biosynthesis Upon Wounding. . *Cellulose* 10:369-389.

Nakazono M, Qiu, F., Borsuk, L.A., Schnable, P.S. (2003) Laser Capture Microdissection, a Tool for the Global Analysis of Gene Expression in Specific Plant Cell Types, Identification of Genes Expressed Differentially in Epidermal Cells or Vascular Tissues of Maize. *Plant Cell* 15:583-596.

Neuhoff V, Arold N, Taube D, Ehrhardt W. (1998) Improved Staining of Proteins in Polyacrylamide Gels Including Isoelectric Focusing Gels with Clear Background at Nanogram Sensitivity Using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9:255-262.

Newton RP, Brenton, G.A., Smith, C.J., Dudley, E. (2004) Plant Proteome Analysis by Mass Spectrometry: Principles, Problems, Pitfalls and Recent Developments. *Phytochemistry* 65:1449-1485.

Nienhaus K, Nienhaus, G.U., Wiedenmann, J., Nar, H. (2005) Structural Basis for Photo-Induced Protein Cleavage and Green-to-Red Conversion of Fluorescent Protein EosFP. *PNAS* 102(26):9156-9159.

Nomata T, Kabeya, Y., Sato, N. (2004) Cloning and Characterisation of Glycine-Rich RNA-Binding Protein cDNAs in the Moss *Physcomitrella patens*. *Plant Physiol.* 45:48-56.

Notaguchi M, Abe, M., Kimura, T., Daimon, Y.,Kobayashi, T., Yamaguchi, A., Tomita, Y., Dohi, K., Mori, M., Araki, T. (2008) Long Distance Graft-Transmissible Action of *Arabidopsis* FLOWERING LOCUS T Protein to Promote Flowering. *Plant Cell Physiol.* 49:1645-1658.

Ogura T, Wilkinson, A.J (2001) AAA+ Superfamily ATPases: Common Structure--Diverse Function. *Genes Cells* 6(7):575-597.

Olsen JV, Ong, S-E., Mann, M. (2004) Trypsin Cleaves Exclusively C-Terminal to Arginine and Lysine Residues. *Mol. & Cell. Proteomics* 3.6:608-614.

Omid A, Keilin, T., Glass, A., Leshkowitz, D., Wolf, S. (2007) Characterization of Phloem-Sap Transcription Profile in Melon Plants. *J Exp Bot.* 58:3645-3656.

Oparka K, Duckett, C., Prior, D., Fisher, D. (1994) Real-Time Imaging of Phloem Unloading in the Root Tip of *Arabidopsis*. . *Plant J.* 6:759-766.

Oparka KJ, Roberts, A. (2001) Plasmodesmata. A Not So Open-and-Shut Case. *Plant Physiol.* 125:123-126.

Ormo M, Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. & Remington, S.J. (1996) Crystal Structure of the *Aequorea Victoria* Green Fluorescent Protein. . *Science* 273:1392-1395.

Owens RA, Blackburn, M., Ding, B. (2001) Possible Involvement of the Phloem Lectin in Long-Distance Viroid Movement. *Mol. Plant-Microbe Interact.* 14:905-909.

Pandey A, Mann, M. (2000) Proteomics to Study Genes and Genomes. . *Nature* 405:837-846.

Pate JS, Sharkey, P.J., Lewis, O.A.M. (1974) Phloem Bleeding from Legume Fruits - a Technique for Study of Fruit Nutrition. *Planta* 120:229-243.

Patterson GH, Lippincott-Schwartz, J. (2002) A Photoactivatable GFP for Selective Photolabeling of Proteins and Cells. *Science* 19(11):555-565.

Paul A-L, Folta, K.M., Ferl, R.J. (2008) 14-3-3 Proteins, Red Light and Photoperiodic Flowering: A Point of Connection? *Plant Signaling and Behavior.* 3:511-515.

Peltier JB, Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., van Wijk, K.J. (2000) Proteomics of the Chloroplast: Systematic Identification and Targeting Analysis of Lumenal and Peripheral Thylakoid Proteins. . *Plant Cell.* 12:319-341.

Phinney B, Thelen, J.J. (2005) Protein Characterization of Triton X-100 Insoluble Fraction from Chloroplasts defines a Novel group of proteins Associated with Macromolecular Structures. *J of Proteome Research* 4:447-450.

Picotti P, Aebersold, R., Domon, B. (2007) The Implications of Proteolytic Background for Shotgun Proteomics. *Mol. & Cell. Proteomics* 6:1589-1598.

Pineiro M, Coupland, G. (1998) The Control of Flowering Time and Floral Identity in *Arabidopsis*. *Plant Physiol.* 117:1-8.

Pneuli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., ron, N., Adir, N., Lifschitz, E. (2001) Tomato SP-Interacting Proteins Define a Conserved Signaling System that Regulates Shoot Architecture and Flowering. *Plant Cell* 13:2687-2702.

Pnueli L, Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganal, M., Zamir, D., Lifschitz, E. (1998) The *SELF-PRUNING* Gene of Tomato Regulates Vegetative to Reproductive Switching of Sympodial Meristems and is the Ortholog of *CEN* and *TFL1*. *Development* 125:1979-1989.

Pnueli L, Liang, H., Rozenberg, M., Mittler, R. (2003) Growth Suppression, Altered Stomatal Responses, and Augmented Induction of Heat Shock Proteins in Cytosolic Ascorbate Proxidase (Apx1)-Deficient *Arabidopsis* Plants. *Plant J.* 34(2):187-203.

Ponder KL, Pritchard, J., Harrington, R., Bale, J.S. (2000) Difficulties in Location and Acceptance of Phloem Sap Combined with Reduced Concentration of Phloem Amino Acids Explained Lowered Performance of the Aphid *Rhopalosiphum padi* on Nitrogen Deficient Barley (*Hordeum vulgare*) Seedlings. *Entomologia Experimentalis et Applicata* 97:203-210.

Pouteau S, Carre, I., Gaudin, V., Ferret, V., Lefebvre, D., Wilson, M. (2008) Diversification of Photoperiodic Response Patterns in a Collection of Early-Flowering Mutants in *Arabidopsis*. *Plant Physiol.* 148:1465-1473.

Pritchard J (1996) Aphid Stylectomy Reveals An Osmotic Step Between Sieve Element and Cortical Cells in Barley Roots. *J Exp. Bot.* 47:1519-1524.

Provart N and Zhu, T. (2003) A Browser-Based Functional Classification SuperViewer for *Arabidopsis* Genomics. *Currents in Computational Molecular Biology* 2003:271-272.

Purwestri YA, Ogaki, Y., Tamaki, S., Tsuji, H., Shimamoto, K. (2009) The 14-3-3 Protein GF14c Acts as a Negative Regulator of Flowering in Rice by Interacting with the Florigen Hd3a. *Plant Cell Physiol.* 50:429-438.

Putterill J, Robson, F., Lee, K., Simon, R., Coupland, G. (1995) The CONSTANS Gene of *Arabidopsis* Promotes Flowering and Encodes a Protein Showing Similarities to Zinc Finger Transcription Factors. *Cell* 80:847-857.

Putterill J, Rebecca, L., Macknight, R. (2004) It's Time to Flower: The Genetic Control of Flowering Time. *Bioassays* 23:363-373.

Rabilloud T (1996) Solubilization of Proteins for Electrophoretic Analysis. *Electrophoresis* 17:813-829.

Rabilloud T (1999) Solubilization of Proteins in 2-D Electrophoresis. *Methods Mol Biol* 112:9-19.

Rabilloud T, Luche, S., Santoni, V., Chevallet, M. ed (2007) *Plant Proteomics: Methods and Protocols* (Humana Press, USA.), pp 111-120.

Rampitsch C, Srinivasan, M. (2006) The Application of Proteomics to Plant Biology: A Review. *Canadian J. Bot.* 4:883-892.

Rausin G, Tillemans, V., Stankovic, N., Hanikenne, M., Motte, P. (2010) Dynamic Nucleocytoplasmic Shuttling of an *Arabidopsis* SR Splicing Factor: Role of the RNA-Binding Domains. *Plant Physiol.* 153:273-284.

Rechsteiner M, Hill, C.P. (2005) Mobilizing the Proteolytic Machine: Cell Biological Roles of Proteasome Activators and Inhibitors. *Trends in Cell Biol.* 15:27-33. 1.

Reichelt M, Brown, P.D., Schneider, B., Oldham, N.J., Stauber, E., Tokuhisa, J., Kliebenstein, D.J., Mitchell-Olds, T., Gershenzon, J. (2002) Benzoic Acid Glucosinolate Esters and Other Glucosinolates from *Arabidopsis thaliana*. *Phytochemistry* 59:663-671.

Reichheld J-P, Meyer, E., Khafif, M., Bonnard, G., and Meyer, Y. (2005) AtNTRB is The Major Mitochondrial Thioredoxin Reductase in *Arabidopsis thaliana*. *FEBS Letters.* 579:337-342.

Richert S, Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., Rabilloud, T. (2004) About the Mechanism of Interferernce of Silver Staining with Peptide Mass Spectrometry. *Proteomics* 4:906-916.

Rinne PLH, van der Schoot, C. (1998) Symplasmic Fields in the Tunica of the Shoot Apical Meristem Coordinate Morphogenetic Events. *Development* 1215:1477-1485.

Rius SP, Casati, P., Iglesias, A.A., Gomez-Casati, D.F. (2008) Characterization of *Arabidopsis* Lines Deficient in GAPC-1, a Cytosolic NAD-Dependent Glyceraldehyde-3-Phosphate Dehydrogenase. *Plant Physiol.* 148:1655-1667.

Rodman JE (1991) A Taxonomic Analysis of Glucosinolate-Producing Plants. Part 1: Phenetics. . *Syst Bot* 16:598-618.

Rodriguez-Medina C, Atkins, C.A., Mann, A.J., Jordan, M.E., Smith, P.M.C. (2011) Macromolecular Composition of Phloem Exudate from White Lupin (*Lupinus albus* L.). *BioMed Central* 11:1-19.

Romano PGN, Horton, P., Gray, J.E. (2004) The *Arabidopsis* Cyclophilin Gene Family. *Plant Physiol.* 134(4):1268-1282.

Ross PL, Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S. Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J. (2004) Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol. & Cell. Proteomics* 3:1154-1169.

Roth U, von Roepenack Lahaye, E., Clemens, S. (2006) Proteome changes in *Arabidopsis thaliana* roots upon exposure to Cd2+. *J Exp. Bot.* 15:4003-4015.

Ruan YL, Llewellyn, D.J., Furbank, R.T. (2001) The Control of Single-Celled Cotton Fibre Elongation by Developmentally Reversible Gating of Plamsodesmata and Coordinated Expression of Sucrose and K^+ Transporters and Expansin. *Plant Cell.* 13:47-60.

Ruiz-Medrano R, Xoconostle-Cázares, B., Lucas, W.J. (1999) Phloem Long-Distance Transport of CmNACP mRNA: Implications for Supracellular Regulation in Plants. *Development* 126:4405-4419.

Ruiz-Medrano R, Xoconostle-Ca´ zares, B., Lucas, W.J. (2001) The Phloem as Conduit for Inter-Organ Communication. . *Curr Opin in Plant Biology* 4:202-209.

Rutschow HL, Baskin, T.I., Kramer, E.M. (2011) Regulation of Solute Flux through Plasmodesmata in the Root Meristem. *Plant Physiol.* 155:1817-1826.

Samach A, Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., Coupland, G. (2000) Distinct Roles of CONSTANS Target Genes in Reproductive Development of *Arabidopsis*. *Science* 288(5349-5361):5349.

Sandstrom J (2000) Nutritional Quality of Phloem Sap in Relation to Host Plant-Alteration in the Bird Cherry-Oat Aphid. *Chemeocology.* 10:17-24.

Santner A, Estelle, M. (2009) Recent Advances and Emerging Trends in Plant Hormone Signalling. *Nature* 459:1071-1078.

Santoni V, Rouquie, D., Doumas, P., Mansion, M., Boutry, M., Degand, H., Dupree, P., Packman, L., Sherrier, J., Prime, T., Bauw, G., Posada, E., Rouze, P., Dehais, P., Sahnoun, I., Barlier, I., Rossignol, M. (1998) Use of Proteome Strategy for Tagging Proteins Present at the Plasma Membrane. *Plant J.* 16:633-641.

Santos CVD, Rey, P. (2006) Plant Thioredoxins Are Key Actors in the Oxidative Stress Response. *Trends in Plant Sci.* 11(7):329-334.

Sasaki T, Chino, M., Hayashi, H., Fujiwara, T. (1998) Detection of Several mRNA Species in Rice Phloem Sap. *Plant Cell Physiol.* 39:895-897.

Sauer N, Stolz, J. (1994) SUC1 and SUC2: Two SucroseTtransporters from *Arabidopsis thaliana*; Expression andCcharacterization in Baker's Yeast and Identification of the Histidine-Tagged Protein. *Plant J.* 6(1):67-77.

Schaefer H, Chamrad, D.C., Marcus, K., Reidegeld, K.A., Bluggel, M., Meyer, H.E. (2005) Tryptic Transpeptidase Products Observed in Proteome Analysis by Liquid Chromatography-Tandem Mass Spectrometry. *Proteomics* 5(4):846-852.

Scheler C, Lamer, S., Pan, Z., Li, X.P., Salnikow, J., Jungblut, P. (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* 19:918-927.

Schenkel M, Sinclair, A.M., Johnstone, D., Bewley, J.D., Mathur, J. (2008) Visualizing the Actin Cytoskeleton in Living Plant Cells Using a Photo-Convertible mEos::FABD-mTn Fuorescent Fusion Protein. *Plant Methods.* 19:21.

Schobert C, Grobmann, P., Gottschalk, M., Komor, E., Pecsvaradi, A., Nieden, U.Z. (1995) Sieve-Tube Exudate from *Ricinus communis* L. Seedlings Contains Ubiquitin and Chaperones. *Planta* 196:205-210.

Schobert C, Baker, L., Szederkenyi, J., Grossmann, P., Komor, E., Hayashi, H., Chino, M., Lucas, W.J. (1998) Identification of Immunologically Related Proteins in Sieve-Tube Exudate Collected from Monocotyledonous and Dicotyledonous Plants. *Planta* 206:245-252.

Schobert C, Gottschalk, M., Kovar, D.R., Satiger, C.J., Yoo, B.C., Lucas, W.J. (2000) Characterization of *Ricinum communis* Phloem Profilin, RcPRO1. *Plant Mol. Biol.* 42:719-730.

Schulz A (1998) The phloem. Structure Related to Function. *Progress in Botany.* 59:429-475.

Searle I, He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasinoe, R.A., Coupland, G. (2006) The Transcription factor FLC Confers a Flowerin Response to Vernalization by Repressing the Meristem Competence and Systemic Signaling in *Arabidopsis*. *Genes & Development* 20:898-912.

Shaner NC, Patterson, G.H., Davidson, M.W. (2007) Advances in Fluorescent Protein Technology. *J. Cell Science* 120(24):4247-4260.

Shaner NC, Lin, M.Z., McKeown, M.R., Steinbach, P.A., Hazelwood, K.L., Davidson, M.W., Tsien, R.Y. (2008) Improving the Photostability of Bright Monomeric Orange and Red Fuorescent Proteins. *Nat. Methods.* 5:545-551.

Shaw MM, Riederer, B.M. (2003) Sample Preparation for Two-Dimensional Gel Electrophoresis. *Proteomics* 3:1408-1417.

Shiratake K (2007) Genetics of Sucrose Transporter in Plants. *Genes, Genomes and Genomics* 1(1):73-80.

Shrividhya J, Schnell, S. (2006) Why Substrate Depletion Has Apparent First-Order Kinetics in Enzymatic Digestion. *Comp. Biol. Chem.* 30(3):209-214.

Siedler F, Weyher, E., Moroder, L. (1996) Cysteine Racemization in Peptide Synthesis: A New and Easy Detection Method. *J. Pept. Sci.* 2:271-275.

Sjölund RD (1997) The Phloem Sieve Element: A River RunsTthrough It. . *Plant Cell.* 9:1137-1146.

Soloneski S, Larramendy, M.L. ed (2011) *Herbicides, Theory and Applications.* (InTech).

Srivastava AC, Ganesan, S., Ismail, I.O., Ayre, B.G (2008) Functional Characterization of the *Arabidopsis* AtSUC2 Sucrose/H^+ Symporter by the Tissue-Specific Complementation reveals an Essential Role in Phloem Loading But Not in Long-Distance Transport. *Plant Physiol.* 148:200-211.

Stadler R, Sauer, N. (1996) The *Arabidopsis thaliana* AtSUC2 Gene is Specifically Expressed in Companion Cells. *Botanica Acta* 109:299-306.

Stadler R, Wright, K.M., Lauterbach, C., Amon, G., Gahrtz, M., Feuerstein, A., Oparka, K.J. & Sauer, N. (2005) Expression of GFP-Fusions in *Arabidopsis* Companion Cells Reveals Nonspecific Protein Trafficking into Sieve Elements and Identifies a Novel Post-Phloem Domain in Roots. *Plant J.* 41:319-331.

Staiger CJ, Yuan, M., Valenta, R., Shaw, P.J., Warn, R.M., Lloyd, C.W. (1994) Microinjected Profilin Affects Cytoplasmic Streaming in Plant Cells by Rapidly Depolymerizing Actin Microfilaments. . *Curr. Biol.* 4:215-219.

Staigner D, Zecca, L., Wieczorek, D.A., Apel, K., Eckstein, L. (2003) The Clock Regulated RNA-Binding Protein AtGRP7 Autoregulates Its Expression by Influencing Alternative Splicing of Its Own Pre-mRNA. *Plant J.* 33:361-371.

Streiner C, Danisman, S., Wehrle, L., Schoning, J.C., Alfano, J.R., Staigner, D. (2008) The Small Glycine-Rich RNA Binding Protein AtGRP7 Promotes Floral Transition in *Arabidopsis thaliana*. *Plant J.* 56:239-250.

Suarez-Lopez P, Wheatley, K., Robson, F., Onouchi, H., Valverde, F., Coupland, G. (2001) CONSTANS Mediates Between the Circadian Clock and the Control of Flowering in *Arabidopsis*. *Nature* 410:1116-1120.

Subach OM, Gundorov, I.S., Yoshimura, M., Subach, F.V., Zhang, J., Gru¨ enwald, D., Souslova, E.A., Chudakov, D.M., Verkhusha, V.V. (2008) Conversion of Red Fuorescent Protein into a Bright Blue Probe. . *Chem. Biol.* 15:1116-1124.

Sumner LW, Wolf-Sumner, B., White, S.P., Asirvatham, V.S. (2002) Silver Stain Removal Using H_2O_2 for Enhanced Peptide Mass Mapping by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* 16:160-168.

Takada S, Goto, K. (2003) TERMINAL FLOWER2, an *Arabidopsis* Homolog of HETEROCHROMATIN PROTEIN1, Counteracts the Activation of *FLOWERING LOCUS T* by CONSTANS in the Vascular Tissues of Leaves to Regulate Flowering Time. *Plant Cell* 15:2856-2865.

Taoka K, Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., Yamaguchi, M., Nakashima, C., Purwestri, Y.A., Tamaki, S., Ogaki, Y., Shimada, C., Nakagawa, A., Kojima, C., Shimamoto, K. (2011) 14-3-3 Proteins Act as Intracellular Receptors for Rice Hd3a Florigen. *Nature* 476:332-338.

Teerawanichpan P, Hoffman, T., Ashe, P., Datla, R., Selvaraj, G. (2007) Investigations of combinations of Mutations in the Jellyfish Green Fuorescent Protein (GFP) that Afford Brighter Fuorescence, and Use of a Version (VisGreen) in Plant, Bacterial, and Animal Cells. *Biochim. Biophys. Acta* 1770:13601368.

Tepfer-Bamnolker P, Samach, A. (2005) The Flowering Integrator FT Regulates SEPALLATA3 and FRUITFULL Accumulation in *Arabidopsis* Leaves. *Plant Cell* 17:2662-2675.

Tesfaye M, Temple, S.J., Allan, D.L., Vance, C.P., Samac, D.A. (2001) Overexpression of Malate Dehydrogenase in Transgenic Alfalfa Enhances Organic Acid Synthesis and Confers Tolerance to Aluminum. *Plant Physiol.* 127(4):1836-1844.

Thangstad OP, Gilde, B., Chadchawan, S., Seem, M., Husebye, H., Bradley, D., Bones, A.M. (2004) Cell Specific, Cross-Species Expression of Myrosinases in *Brassica napus, Arabidopsis thaliana* and *Nicotiana tabacum*. *Plant Mol. Biol.* 54:597-611.

Thelen JJ, Peck, S.C. (2007) Quantitative Proteomics in Plants: Choices in Abundance. *Plant Cell.* 19:3339-3346.

Thomas B, Vince-Prue, D. (1997) Photoperiodism in Plants. (Academic Press, New York).

Thorn KS, Christensen, H.E., Shigeta, R., Huddler, D., Shalaby, L., Lindberg, U., Chua, N,H., Schutt, C.E. (1997) The Crystal Structure of a Major Allergen from Plants. *Structure* 5:19-32.

Tomaz T, Bagard, M., Pracharoenwattana, I., Lindén, P., Lee, C.P. Carroll, A.J., Ströher, E., Smith, S.M. Gardeström, P., Millar, A.H. (2010) Mitochondrial Malate Dehydrogenase Lowers Leaf Respiration and Alters Photorespiration and Plant Growth in *Arabidopsis*. *Plant Physiol.* 154(3):1143-1157.

Tooke F, Ordidge, M., Chiurugwi, T., Battey, N. (2005) Mechanisms and Function of Flower and Inflorescence Reversion. . *J Exp. Bot.* 56:2587-2599.

Touraine B, Astruc, S. (1990) Purification and dual-column HPLC determination of carboxylic acids in tissues, phloem and xylem saps of soybean plants. . *Chromatographia* 30:388-392.

Truernit E, Sauer, N. (1995) The Promoter of the *Arabidopsis thaliana* SUC2 Sucrose-H^+ Sumporter Gene Directs Expression of b-Glucuronidase to the Phloem: Evidence for Phloem Loading and Unloading by SUC2. *Planta* 196:564-570.

Truernit E (2001) Plant Physiology: The Importance of Sucrose Transporters. *Curr. Biol.* 11(5):R169-R171.

Tsien RY (1998) The Green Fluorescent Protein. *Ann. Rev. Biochem.* 67:509-544.

Tsutsui H, Karasawa, S., Shimizu, H., Nukina, N., Miyawaki, A. (2005) Semi-Rational Engineering of a Coral Fluorescent Protein into an Efficient Highlighter. *EMBO Rep.* 6:233-238.

Tully, R.E., Hanson, A.D. (1979) Amino Acids Translocated From Turgid and Water-Stressed Barley Leaves. I. Phloem Exudation Studies. *Plant Physiol.* 64:460-466.

Turck F, Fomana, F., Coupland, G. (2008) Regulation and Identity of Florigen: FLOWERING LOCUS T Moves Center Stage. *Annu. Rev. Plant Biol.* 59:573-594.

Turgeon R, Wolf, S. (2009) Phloem Transport: Cellular Pathways and Molecular Trafficking. *Annu. Rev. Plant Biol.* 60:207-221.

Tuteja N, Umate, P., van Bel, A.J.E. (2010) Forisomes: Calcium-Powered Protein Complexes with Potential as 'Smart' Biomaterials. . *Trends Biotechnol.* 28:102-110.

Urquhart AA, Joy, K.W. (1981) Use of Phloem Exudate technique in the Study of Amino Acid Transport in Pea Plants. *Plant Physiol.* 68:750-754.

Valverde F, Mouradov, A., Soppe, W., Ravencroft, D., Samach, A. & Coupland, G. (2004) Photoreceptor Regulation of CONSTANS Protein in Photoperiodic Flowering. *Science* 303:1003-1006.

van Bel AJE, Ehlers, K., Knoblauch, M. (2002) Sieve Elements Caught in the Act. *Trends Plant Sci.* 7(3):126-132.

van Bel AJE, Gaupels, F. (2004) Pathogen-Induced Resistance and Alarm Signals in the Phloem. *Molecular Plant Pathology* 5(5):495-504.

van Bel AJE (2003) The Phloem, A Miracle of Ingenuity. . *Plant, Cell & Environment* 26:125-149.

Van Der Straeten D, Rodrigues-Pousada, R.A., Goodman, H.M., Van Montagu, M. (1991) Plant Enolase: Gene Structure, Expression, and Evolution. *Plant Cell.* 3:719-735.

van Nocker S, Vierstra, R. (1993) Two cDNAs from *Arabidopsis thaliana* Encode Putative RNA Binding Proteins Containing Glycine-Rich Domains. *Plant Mol. Biol.* 21:695-699.

van Wijk KJ (2001) Challenges and Prospects of Plant Proteomics. *Plant Physiol.* 126:501-508.

Verma DPS, Hong, Z. (2001) Plant Callose Synthase Complexes. . *Plant Mol. Biol.* 47:693-701.

Vierstra RD (2009) The Ubiquitin/26S Proteasome System at the Nexus of Plant Development. *Nat. Rev. Mol. Cell Biol.* 10:385-397.

Vilaine F, Palauqui, J.C., Amselem, J., Kusiak, C., Lemoine, R., Dinant, S. (2003) Towards Deciphering Phloem: a Transcriptome Analysis of the Phloem of *Apium graveolens*. *Plant J.* 36:67-81.

Voges D, Zwicki, P., Baumeister, W. (1999) The 26S Proteasome: a Molecular Machine Designed for Controlled Proteolysis. *Ann. Rev. Biochem.* 68:1015-1068.

Walsh MA, Melaragno, J.E. (1981) Structural Evidence for Plastid Inclusions as a Possible 'Sealing' Mechanism in the Phloem of Monocotyledons. . *J. Exp. Bot.* 32:311-320.

Walz C, Juenger, M., Schad, M., Kehr, J. (2002) Evidence for The Presence And Activity Of A Complete Antioxidant Defence System In Mature Sieve Tubes. . *Plant J.* 31:189-197.

Walz C, Giavalisco, P., Schad, M., Juenger, M., Klose, J., Kehr, J. (2004) Proteomics of Cucurbit Phloem Exudates Reveals a Network of Defence Proteins. *Phytochemistry* 65:1795-1804.

Webb JA, Gorham, P.R. (1964) Translocation of Photosynthetically Assimilated C^{14} in Straight-Necked Squash. *Plant Physiol.* 39:663-672.

Weibull J, Ronquist, F. & Brishammar, S. (1990) Free Amino Acid Composition of Leaf Exudates and Phloem Sap. *Plant Physiol.* 92:222-226.

Wiese S, Reidegeld, K.A., Meyer, H.E., Warscheid, B. (2007) Protein Labeling by iTRAQ: A New Tool for Quantitative Mass Spectrometry in Proteome Research. *Proteomics* 7:340-350.

Wigge PA, Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., Weigel, D. (2005) Integration of Spatial and Temporal Information During Floral Induction in *Arabidopsis*. *Science* 309:1056-1059.

Wilkinson TL, Douglas, A.E. (2003) Phloem Amino Acids and the Host Plant Range of the Polyphagous Aphid, *Aphis fabae*. *Entomologia Experimentalis et Applicata* 106:103-113.

Will T, van Bel, A.J.E. (2006) Physical and Chemical Interactions Between Aphids and Plants. . *J Exp. Bot.* 57:729-737.

Williams LE, Lemoine, R., Sauer, N. (2000) Sugar Transporters in Higher Plants: A Diversity of Roles and Complex Regulation. *Trends in Plant Sci.* 5:283-290.

Willmitzer AG (2006) Spatial temporal Analysis of Florigenic Signals in *Arabidopsis thaliana*, *Sinapsis alba* and *Brassica napus*. Doctor of Philosophy (Universitat Potsdam, Potsdam).

Wittstock U, Halkier, B.A. (2002) Glucosinolate Research in the *Arabidopsis* Era. *Trends in Plant Sci.* 7:263-270.

Woloshen V, Huang, S., Li, X (2011) RNA-Binding Proteins in Plant Immunity. *Journal of Pathogens*.

Wong JH, Kim, Y.B., Ren, P.H., Cai, N., Cho, M.J., Hedden, P., Lemaux, P.G., Buchanan, B.B. (2002) Transgenic Barley Grain Overexpressing Thioredoxin Shows Evidence that the Starchy Endosperm Communicates with the Embryo and the Aleurone. *PNAS* 99:16325-16330.

Wu S, Koizumi, K., MacRae-Crerar, A., Gallagher, K.L. (2011) Assessing the Utility of Photoswitchable Fluorescent Proteins for Tracking Intercellular Protein Movement in the *Arabidopsis* Root. *PLoS ONE* 6(11).

Wu X, Dinneny, J.R., Crawford, K.M., Rhee, Y., Citovsky, V., Zambryski, P.C., Weigel, D. (2003) Modes of Intercellular Transcription Factor Movement in the *Arabidopsis* Apex. *Development* 130:3735-3745.

Xie B, Hong, Z. (2011) Unplugging the Callose Plug from Sieve Pores. *Plant Signaling and Behavior.* 6(4):491-493.

Xoconostle-Cázares B, Xiang, Y., Ruiz-Medrano, R., Wang, H.L., Monzer, J., Yoo, B.C., McFarland K.C., Franceschi, V.R., Lucas, W.J. (1999) Plant Paralog to Viral Movement Protein that Potentiates Transport of mRNA into the Phloem. *Science* 283:94-98.

Xoconostle-Cázares B, Ruiz-Medrano, R., Lucas, W.J. (2000). Plant J. 24, 735–747. (2000) Proteolytic Processing of CmPP36, a Protein from the Cytochrome b(5) Reductase Family, Is Rrequired for Entry into the Phloem Translocation Pathway. *Plant J.* 24:735-747.

Xu YL, Gage, D.A., Zeevaart, J.A. (1997) Gibberellins and Stem Growth in *Arabidopsis thaliana*. Effects of Photoperiod on Expression of the GA4 and GA5 loci. *Plant Physiol.* 114:1471-1476.

Yamada K, Lim, J., Dale, J.M. (2003) Empirical Analysis of Tracscriptional Activity in the *Arabidopsis* Genome. . *Science* 302:842-846.

Yamaguchi A, Kobayashi, Y. Goto, K., Abe, M., Araki, T. (2005) TWIN SISTER OF FT (TSF) Acts as a Floral Pathway Integrator Redundantly with FT. *Plant Cell Physiol.* 46:1175-1189.

Yang Y, Klejnot, J., Yu, X., Liu, X., Lin, C. (2007) Florigen (II):It is a Mobile Protein. *J of Integrative Plant Biology* 49(12):1665-1669.

Yeung K, Seitz, T., Li, S. (1999) Suppression of Raf-1 Kinase Activity and MAP Kinase Signalling by RKIP. *Nature* 401:173-177.

Yoo B-C, Lee, J-Y., Lucas, W.J. (2002) Analysis of the Complexity of Protein Kinases within the Phloem Sieve Tube System. *J Biol Chem* 277(18):15325-15332.

Yoo B-C, Kragler, F., Varkonyi-Gasic, E., haywood, V., Archer-Evans, S., Lee, Y.M., Lough, T.J., Lucas, W.J. (2004) A Systemic Small RNA signalling System in Plants. *Plant Cell* 16:1979-2000.

Yoo SK, Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S., Ahn, J.H. (2005) *CONSTANS* Activates *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* Through *FLOWERING LOCUS T* to Promote Flowering in *Arabidopsis*. *Plant Physiol.* 139:770-778.

Yoshida H, Tsumuki, H., Kawada, K., Kanehisa, K. (1996) Gramine and Resistance of Barley to Aphids: Analysis of EDTA Exudates from Barley Leaves. *Bull. Res. Inst. Bioresour. Okayama Univ.* 4:73-78.

Zabotin AI, Barysheva, T.S., Trofimova, O.I., Lozovaya, V.V., Widholm, J. (2002) Regulation of Callose Metabolism in Higher Plant Cells *In Vitro*. . *Russian Journal of Plant Physiology.* 49:792-798.

Zhang C, Guy, C.L. (2006) *In Vitro* Evidence of Hsc70 Functioning as a Molecular Chaperone During Cold Stress. *Plant Physiol. Biochem.* 14:844-850.

Zhang C, Yu, X., Ayre, B.G., Turgeon, R. (2012) The Origin and Composition of Cucurbit "Phloem" Exudate. *Plant Physiol.* 158(4):1873-1882.

Zhang B, Tolstikov, V., Turnbull, C., Hicks, L.M., Fiehn, O. (2010) Divergent Metabolome and Proteome Suggest Functional Independence of Dual Phloem Transport Systems in Cucurbits. *PNAS*:1-6.

Zhang L, Gurskaya, N.G., Merzlyak, E.M., Staroverov, D.B., Mudrik, N.N., Samarkina, O.N., Vinokurov, L.M., Lukyanov, S., Lukyanov, K.A. (2007) Method for Real-Time Monitoring of Protein Degradation at the Single Cell Level. *Biotechniques* 42:446-450.

Zhang S, Li., S., Kragler, F. (2009) The Phloem-Delivered RNA Pool Contains Small Noncoding RNAs and Interferes with Translation. *Plant Physiol.* 150(1):378-387.

Zhao J-Y, Zuo, K-J., Qin, J., Tang, K-X. (2004) Molecular Cloning and Characterization of Enolase from Oilseed Rape (*Brassica napus*). *Acta Botanica Sinica* 46(10):1226-1233.

Zhao Z, Zhang, W., Stanley, B.A., Assmann, S.M. (2008) Functional Proteomics of *Arabidopsis thaliana* Guard Cells Uncovers New Stomatal Signaling Pathways. *Plant Cell.* 20:3210-3226.

Appendix A

Complete list of identified proteins from *Arabidopsis* phloem sap. MS spectra were analysed using MASCOT and ProteinPilot against *Arabidopsis thaliana* database. A minimum score of 55 and above was applied for MASCOT and any scoring from the ProteinPilot that with $p<0.05$ was accepted. Partial sequence was provided for each protein with high confidence ($p<0.05$) and peptide sequence with $p<0.1$ was represented in red. The frequency of proteins found in each run was noted. Additionally, similar findings between the identified proteins and *Arabidopsis* phloem mRNA by AGI was provided.

Represents all proteins identified by Batailler et al (2012) in the phloem exudate

Represents all proteins identified by Batailler et al (2012) in the full set of exudates

Proteins with shared homology in all four species are written in blue

∗ Protein found in translatome study by Mustroph et al (2009)

Peptide sequence with $p<0.1$ was represented in red and *

No	AGI No.	Protein	Partial sequence	Frequency found	Identified using MASCOT (M) or ProteinPilot (PP)	Score	Similarity to Ath phloem mRNA (Deeken et al., 2008)
1	At1g07890	Ascorbate peroxidase 1 (APX16) ∗	EGLLQLVSDK ALLDDPVFRPLVEK ALLDDPVFRP	2	M, PP	126, 13.2	EX

			LSELGFADA				
2	At4g35090	Catalase 2 (CAT) *	APGVQTPVIVR	3	M	63	-
3	At1g20620	Catalase 3 (CAT3) *	LGPNYLQLPVNAPK VPTPTNSYTGIR APGVQTPVIVR DEEINYYPSK	3	M, PP	63, 5.3	L, EX
4	At1g08830	Copper/zinc superoxide dismutase (CSD1) *	AVVVHADPDDLGK GGHELSLATGNAGGR ALGDTTNGCMSTGPHFNPDGK VACGIIGLQG VVHADPDDLGK VVVHADPDDLGK	2	M, PP	115, 38.8	L, EX
5	At2g28190	Copper/zinc superoxide dismutase (CSD2) *	GGHELSLTTGNAGGR LACGVIGLTPL LACGVIGLT	2	PP	12	L, EX
6	At1g19570	Dehydroascorbate reductase (DHAR1) *	YVISGWAPK VSAVDLSLAPK TLFSLDSFEK YPDPPLK SHDGPFIAGER	1	M	169	L, EX
7	At4g11600	Glutathione peroxidase 6 (GPX6) *	FLVDKDGNVVDR FAPTTSPLSIEK	1	M	59	EX, EST
8	At2g24940	Membrane-associated progesterone binding protein 2 (MAPR2) *	VFDVTTGK EINTLNDWETK NEEDVSPSLEGLTEK SFYGSGGDYSMFAGK	2	M	185	EX, EST
9	At5g40370	Membrane-associated progesterone binding protein 2 (MAPR2) *	TYCPYCVR LVPLLTEAGAIAGK ITGQYTVPNVFIGGK TTVAENPVVVYSK	2	M, PP	102, 41.4	L, EX

			VKELLQQLGAK PLLTEAGAIAGK PNVFIGGNHIGGCDATSNLHK				
10	At3g52880	Monodehydroascorbate reductase 1 (MDAR1) *	EFANQGVQPGELAVISK VVGAFMEGGSGDENK PYERPALSK GTVASGFTAQPNGEVK PSAESLDELVK SLVSATGDVFK QGISFAAK	2	M, PP	75, 13.6	EX
11	At1g63940	Monodehydroascorbate reductase 6 (MDAR6) *	QGISFAAK VVGAFMEGGSGDENK GASINNLEAGSDGR	2	M, PP	75, 2.8	EX
12	At5g03630	Pyridine nucleotide-disulphide oxidoreductase family protein (MDAR2) *	LSDFGVPGADAK YQTLLAATGSSVIR	3	M, PP	105, 61	EX
13	At1g45145	Thioredoxin 3 (TRX3) *	VEAMPTFVFMK IDVDELQAVAQEFK FIAPVFAEMAK	1	M, PP	108, 21.2	L, EX
14	At3g02730	Thioredoxin F-type 1 (TRXF1) *	YDDVVFLK YDDLVAAIETAR LDCNPDNRPLAK	2	M, PP	107, 6.7	-
15	At5g16400	Thioredoxin F2 (TRXF2) *	YEDLLAAIEAAR	1	M, PP	90, 6.5	-
16	At5g42980	Thioredoxin H-type 3 (TRX-H-3) *	FIAPVFADLAK FIAPVFADLAKK VQAMPTFIFMK EGEIKETVVGAAK VDVDELNTVAEEFK HLDVVFFK	2	M, PP	256, 72	L, EX, EST

			AAEGEVIACHTVEDWTEK APVFADLAK APVFADLAKK EEIIANLEK GEVIACHTVEDWTEK PVFADLAK VFADLAK				
17	At1g03680	Thioredoxin M-type 1 (TRX-M1)	LNTDESPATPGQYGVR MIDPIVNELAQK ADEPVFVDFWAPWCGPCK	2	M, PP	67, 25.7	EX
18	At3g15360	Thioredoxin M-type 4 (TRX-M4) *	VLESDVPVLVEFWAPWCGPCR PTVIIFK	1	PP	14.5	-
19	At1g43560	Thioredoxin Y2 (Aty2_ty2) *	FEGALPANQLVER	1	M	55	-
20	At2g20270	Thioredoxin superfamily protein *	ITGQYTVPNVFIGGK TTVAENPVVVYSK	1	PP	15.6	-
21	At3g11630	Thioredoxin superfamily protein *	AQADDLPLVGNK SGGLGDLNYPLISDVTK	2	M, PP	132, 10.9	EX
22	At3g52960	Thioredoxin superfamily protein *	YAILADDGVVK	1	M	67	-
23	At4g03520	Thioredoxin superfamily protein (ATHM2) *	LNTDESPNTPGQYGVR MIDPLVNDLAQHYTGK ATGPVVVDFWAPWCGPCK DPLVNDLAQHYTGK DTIIGAVPK	3	M, PP	119, 31.7	EX
24	At2g24270	Aldehyde dehydrogenase 11A3 (ALDH11A3) *	FLLSDSFPGNDR	1	PP	2.4	L, EX
25	At2g36530	Enolase (ENO2) / (LOS2) *	AGAVVSGIPLYK AVGNVNNIIGPALIGK VNQIGSVTESIEAVK ATITVVK	4	M, PP	37, 9.7	EX

26	At5g42740	Glucose-6-phosphate isomerase, cytosolic (GPI) (phosphoglucose isomerase) (PGI) *	FLANIDPVDVAR	1	M, PP	70, 2.1	-
27	At3g04120	Glyceraldehyde-3-phosphate dehydrogenase C subunit 1 (GAPC1) *	AASFNIIPSSTGAAK TLLFGEKPVTVF IGINGFGR	3	M, PP	55, 8	L, EX
28	At3g55440	Triosephosphate isomerase (TPI) *	NVSADVAATTR IIYGGSVNGGNCK AILNESSEFVGDK VIACVGETLEER VAYALAQGLK EAGSTMDVVAAQTK SADVAATTR VASPAQAQEVHDELR YGGSVNGGNCK	2	M, PP	142, 28.7	-
29	At3g03250	UDP-glucose pyrophosphorylase 1 (UGP1) *	SIPSIVELDSLK FFDNAIGVNVPR IFNTNNLWVNLK SAVDGLTEMSESEK TNPSNPSIELGPEFK MTPVSQDVAETK VVADEFVPWPSK IQTPTDEIVVPYEK SGFISLVSR	4	M, PP	274, 13.4	-
30	At5g17310	UDP-glucose pyrophosphorylase 2 (UGP2) *	SIPSIVELDSLK FFDNAIGVNVPR IFNTNNLWVNLK IQTPTDEIVVPYDK AATATEKLPQLK	3	M, PP	170, 13.4	-

			PSIVELDSLK TNPTNPAIELGPEFK VVADEFVPWPSK SIPSIVELDSLK				
31	At5g09660	Peroxisomal NAD-malate dehydrogenase 2 (PMDH2) *	LLGVTTLDVAR AGAGSATLSMAYAAAK TGAEEVYQLGPLNEYER IQNGGTEVVEAK	1	M, PP	80, 11	-
32	At3g01500	Carbonic anhydrase 1 (CA1) *	NIANMVPPFDK VCPSHVLDFQPGDAFVVR VISELGDSAFEDQCGR YETNPALYGELAK YMVFACSDSR VENIVVIGHSACGGIK	2	PP	19.6	-
33	At5g14740	Carbonic anhydrase 2 (CA2) *	ITAELQAASSSDSK NIANMVPPFDK ITAELQAASSSDSK VLAESESSAFEDQCGR YETNPALYGELAK	2	M, PP	87, 24.7	-
34	At1g04410	Lactate/malate dehydrogenase family protein (c-NAD-MDH1) *	MELIDAAFPLLK NGDWSIVQGLPIDEVSR VLVVANPANTNALILK ALGQISER LSVPVSDVK	2	M, PP	136, 16	L, EX
35	At1g53240	Lactate/malate dehydrogenase family protein (mMDH1) *	SEVVGYMGDDNLAK ALEGADLVIIPAGVPR DDLFNINAGIVK VAILGAAGGIGQPLALLMK	2	M, PP	62, 17.9	-
36	At5g11670	NADP-malic enzyme 2 (NADP-ME2) *	YMALMDLQER	1	M	113	EX

			TYDLGLASNLPR				
37	At4g14880	O-acetylserine (thiol) lyase (OAS-TL) isoform A1 (OASA1) *	IGFSMISDAEK IDGFVSGIGTGGTITGAGK YLSTVLFDATR	1	M, PP	112, 9.3	EX
38	At4g13940	S-adenosyl-L-homocysteine hydrolase (SAHH1) *	WVFPETK DIIMVDHMR TEFGPSQPFK AGCYASNVVIQR FNVEYIAGGATQNSIK SGCTYPEKPDFN SLIANLSAANCYK TTVITQGADPVVVAEDGK GETLQEYWWCTER LVGVSEETTTGVK VAVICGYGDVGK PQTDR AGIIVLAEGR DMSQADFGR	3	M, PP	154, 8.9	L, EX
39	At3g23810	S-adenosyl-l-homocysteine (SAH) hydrolase 2 (SAHH2) *	GVSEETTTGVKR VAVICGYGDVGK SGIIVLAEGR	2	PP	5	EX
40	At1g02500	S-adenosylmethionine synthetase 1 (SAM1) *	FVIGGPHGDAGLTGR	1	PP	3.8	EX, EST
41	At4g31990	Aspartate aminotransferase 5 (ASP5) *	VATIQGLSGTGSLR	1	PP	3.1	-
42	At5g17920	Cobalamin-independent synthase family protein (ATCIMS)	VVEVNALAK YLFAGVVDGR IPSSEEIADR GNASVPAMEMTK DEALFSANAAALASR	4	M, PP	133, 10.6	EX

			FALESFWDGK GGIGVIQIDEAALR ASHIVGYPR PAMEMTK YGAGIGPGVYDIHSPR PSSEEIADR				
43	At1g31812	Acyl-CoA-binding protein 6 (ACBP6) *	SSEEAMNDYITK QLLEVAASK VNTLTELPSNEDLLILYGLYK	2	M, PP	95, 45.7	L, EX
44	At4g25050	Acyl carrier protein 4 (ACP4) *	FPENFTGCQDLAK EQLALAADVPLTAESK	1	M	89	EX, EST
45	At5g16390	Biotin carboxyl carrier protein of acetyl-CoA carboxylase precursor (BCCP) *	SPAPGEPPFIK SPMAGTFYR	1	M, PP	55, 3.9	-
46	At1g27950	Glycosylphosphatidylinositol-anchored lipid protein transfer 1 (LTPG1) *	DLGVQEDK APAAPGGALADECNQDFQK CLCFVIQQAK	2	PP	19.2	-
47	At2g38540	Lipid transfer protein 1 (LTP1) *	ACGVNIPYK ACGVNIPYKISTSTNCKTVR TTPDRQQACNCIQGAAR ALGSGLNAGR QQACNCIQGAAR ALSCGSVNSNLAACIGYVLQGGVI PPAC-CSGVK ALGSGLNAGRAAGIPK PDRQQACNCIQGAAR GSGLNAGRAAGIPK ACGVNIPYKISTSTNCK	3	M, PP	108, 72.9	EX
48	At2g38530	Lipid transfer protein 2 (LTP2) *	GAPLTQGCCNGVTNLK CLQSAAKAVGPGLNTAR	1	M, PP	86, 46.6	EX

			PSACKVNIPYKISASTNCNTVR				
49	At1g66970	SHV3-like 2 (SVL2) *	WQTLNGDAPLVIAR	2	M, PP	92, 1.8	EX
50	At3g09820	Adenosine kinase 1 (ADK11) *	YPVIPLPK AGCYASNVVIQR FNVEYIAGGATQNSIK TTVITQGADPVVVAEDGK LNNAILAEDK NAILAEDK SGCTYPEKPDFN	4	M, PP	201, 16.3	EX
51	At5g03300	Adenosine kinase 2 (ADK2) *	LNNAILAEDK AGCYASNVVIQR FNVEYIAGGATQNSIK SGCTYPEKPDFN SLIANLSAANCYK TTVITQGADPVVVAEDGK TTVITQGADPVVVAEDGKVK	3	PP	24.1	EX
52	At4g09320	Nucleoside diphosphate kinase family protein (NDPK1) *	IIGATNPAASEPGTIR GDFAIDIGR	1	M, PP	96, 16.8	L, EX
53	At5g63310	Nucleoside diphosphate kinase 2 (NDPK2) *	TDPLQAEPGTIR	1	PP	5.2	-
54	At5g09650	Pyrophosphorylase 6 (PPa6) *	AHLVNDVEDVEK VQEEGPAESLDYR MEVATDEDFTPIK	2	M, PP	124, 8.3	L
55	At3g24503	Aldehyde dehydrogenase 2C4 (ALDH1A / ALDH2C4) *	SPLLIFNDADIDK	1	PP	2.6	EX
56	At1g54040	Epithiospecifier protein (ESP) *	LGEEGAPAIPR APTLQGQWIK WAQLPDPGDNFEK	1	M	96	-
57	At3g44310	Nitrilase 1 (NIT1)	GAELVLFPEGFIGGYPR	1	PP	4.9	EX
58	At5g26000	Thioglucoside glucohydrolase 1 (TGG1)	GYALGTDAPGR	5	M, PP	181, 17.2	EX, EST

			LPEFSETEAALVK YYNGLIDGLVAK GFIFGVASSAYQVEGGR GMIGPVMITR NWITINQLYTVPTR GIYYVMDYFK FSIAWSR LPEFSETEAALVK LFNSGNFEK IGPVMITR				
59	At2g41530	S-formylglutathione hydrolase (SFGH) *	LLSENFSQLDTTK ASGLSEIGSTK AASTHGIALVAPDTSPR AFTNYLGDNK MYDYVVK	2	M, PP	104, 13.4	-
60	At3g59970	Methylenetetrahydrofolate reductase 1 (MTHFR1)	LQQEWVVPLK EVIQPTIVDPASFK	1	M	55	L, EX
61	At3g16430	Jacalin-related lectin 31 (JAL31) *	IFGSDGSVITMLR VYVGQAQDGISAVK	1	M	145	EX
62	At3g16470	Mannose-binding lectin superfamily protein (JAL35_JR1) *	LGVNVAPIAK VYVGQGDSGVVYVK GGDVWDDGGAYDNVK IVGFYGQAGEYLYK GGEEWDDGGAYENVK TSQPFGLTSGEEAELGGGK IFSVDAPAIVSLK IYVGQGDSCVTYFK	4	M, PP	289, 34.1	EX

			IVGFYGQAGEYLYK GLTSGEEAELGGGK PAQGGDGGVAWDDGVHDSVK PFGLTSGEEAELGGGK TSQPFGLTSGEEAELGGGK				
63	At1g12240	Glycosyl hydrolases family 32 protein / Beta-fructosidase (ATBETAFRUCT4) *	TAGNGLDTSVNGPDVK NLVQWPVEEIK TGISLVYDTIDFK	1	M	84	-
64	At5g37850	pfkB-like carbohydrate kinase family protein (SOS4) *	LIQSQEDIR	2	M, PP	60, 2.9	EX
65	At1g23740	Oxidoreductase, zinc-binding dehydrogenase family protein *	VADAFSYLETNHATGK VVAAALNPVDAK EGDEVYANVSEK FVVTSNGDVLK	2	M, PP	65, 7.3	-
66	At5g61130	Plasmodesmata callose-binding protein 1 (PDCB1) *	QSCFNPDNVR LYGPDPGALAALR	2	PP	5	-
67	At5g08280	Hydroxymethylbilane synthase (HEMC) *	AAEIFSNK EHISAYGEGNER HETASIDQFSWGVANR SMREEGGFEVIKK TIEKPVEDPSELPK HKEHISAYGEGNER	1	PP	15.1	-
68	At5g08280	Glutamate-1-semialdehyde-2,1-aminomutase (GSA1) *	GSPLALAQAYETR ILSQPLADIGGK AGPGFFGN TPIAGYASK	2	M, PP	86, 6.6	-
69	At5g63570	Glutamate-1-semialdehyde 2,1-aminomutase 1, GSA 1	AGSGVATLGLPDSPGVPK IIGGGLPVGAYGGR	1	PP	6.8	EX
70	At4g26500	Chloroplast sulfur E (SUFE1) *	YEQLMFYGK	1	PP	2.4	-

71	At5g54770	Thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) *	ALDMNTAEDAIVR	1	M	59	L, EX
72	At2g21330	Fructose-bisphosphate aldolase 1 (FBA1) *	NLNAMNQLK GILAADESTGTIGK LASINVENVETNR ATPEQVASYTLK	1	M	168	EX
73	At4g38970	Fructose-bisphosphate aldolase 2 (FBA2) *	GILAMDESNATCGK	1	M	62	EX
74	At1g42970	Glyceraldehyde-3-phosphate dehydrogenase B subunit (GAPB) *	AAALNIVPTSTGAAK	1	PP	3.4	-
75	At3g12780	Phosphoglycerate kinase 1 (PGK1) *	SVGDLTSADLK GVTTIIGGGDSVAAVEK	1	M	162	EX
76	At1g56190	Phosphoglycerate kinase family protein *	TFSEALDTTK ELDYLVGAVANPK GVTTIIGGGDSVAAVEK FAAGTEAIANK FSLAPLVPR GVSLLLPTDVVVADK LAELSEKGVTTIIGGGDSVAAVEK	4	M, PP	124, 15.1	-
77	At2g21170	Plastid Isoform Triose Phosphate Isomerase (PDTPI) *	IDISGQNSWVGK GPEFATIVNSVTSK IIYGGSVNGGNSAELAK	1	M, PP	68, 13.7	EX
78	At2g39730	Rubisco activase (RCA) *	AALGDANADAIGR AETYLSQAALGDANADAIGR GLAYDTSDDQQDITR LAETYLSQAALGDANADAIGR ALGDANADAIGR GAQQVNLPVPEGCTDPVAENFDPTAR KGAQQVNLPVPEGCTDPVAENFDPTAR LGDANADAIGR	3	M, PP	73, 26.9	EX

			LSQAALGDANADAIGR MSAGELESGNAGEPAK NIADNPTNVQLPGMYNK PEGCTDPVAENFDPTAR PVAENFDPTAR CYHIEPVPGEETQFIAY DDENVNSQPFMR DLAVEGNEIIR DTDILAAFR ESTLGFVDLLR GGDHIHAGTVVGK GGLDFTKDDENVNSQPF GPPHGIQVER LSGGDHIHAGTVVGK LTYYTPEYETK PELAAACEVWK QGPPHGIQVER SGGDHIHAGTVVGK TFQGPPHGIQVER VALEACVQAR WSPELAAACEVWK DNGLLLHIHR GGLDFTK				
79	AtCg00490	Ribulose-bisphosphate carboxylases (RBCL)	DLAVEGNEIIR ESTLGFVDLLR LTYYTPEYETK LSGGDHIHAGTVVGK TFQGPPHGIQVER	4	M, PP	157, 18.8	-

			FEFAKVDTIDEEGK EHGNSPGYYDGR FETLSYLPDLTDSELAK FGCTDSAQVLK IAYKPPSFTG IIGFDNTR LPLFGCTDSAQVLK PCVEFELEHGFVYR LTDSELAK SYLPDLTDSELAK VALEACVQAR DTDILAAFR GPPHGIQVER PELAAACEVWK QGPPHGIQVER VALEACVQAR WSPELAAACEVWK PQTETKASVGFK TFQGPPHGIQVER GGLDFTKDDENVNSQPFMR				
80	At1g67090	Ribulose bisphosphate carboxylase small chain 1A (RBCS1A) *	IIGFDNTR MQVWPPIGK LPLFGCTDSAQVLK EHGNSPGYYDGR FETLSYLPDLTDSELAK FGCTDSAQVLK IAYKPPSFTG	3	M, PP	56, 47.8	L, EX, EST

			KFETLSYLPDLTDSELAK PCVEFELEHGFVYR LTDSELAK SYLPDLTDSELAK WIPCVEFELEHGFVYR KKFETLSYLPDLTDSELAK				
81	At3g55800	Sedoheptulose-bisphosphatase (SBPASE) *	GIFTNVTSPTAK YTGGMVPDVNQIIVK FEETLYGTSR TGGMVPDVNQIIVK	2	M, PP	65, 6.4	-
82	At3g60750	Transketolase *	ALPTYTPESPGDATR SIITGELPAGWEK	1	M	75	EX
83	At1g20340	Cupredoxin superfamily protein (DRT112_PETE2) *	NNAGYPHNVVFDEDEIPSGVDVAK VFDEDEIPSGVDVAK IEVLLGGGDGSLAFIPNDFSIAK PGTYSFYCAPHQGAGMVGK PSGVDVAK IVFKNNAGYPHNVVFDEDEIPSGVDVAK	2	M, PP	56, 41.9	L, EX
84	At5g66190	Ferredoxin-NADP(+)-oxidoreductase 1 (FNR1) *	LVYTNDGGEIVK EGQSIGVIPEGIDK DNTFVYMCGLK LYSIASSAIGDFGDSK	1	M	93	EX
85	At1g20020	Ferredoxin-NADP(+)-oxidoreductase 2 (FNR2) *	NEEGVIVNR LVYTNDQGETVK EGQSVGVIADGIDK	1	M	106	-
86	At1g77090	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein *	VINAFGPEVIGENVEGK	1	PP	6.2	-

87	At1g06680	Photosystem II subunit P-1 (PSBP-1) *	EIEYPGQVLR TNTDFLPYNGDGFK AYGEAANVFGKPK HQLITATVNGGK ILESSSQEVGGK PYYYLSVLTR TADGDEGGKHQLITATVNGGK TNTDFLPYNGDGFK TNTDFLPYNGDGFKVQVPAK HQLITATVNGGK PSKEIEYPGQVLR SITDYGSPEEFLSQVNYLLGK GEAANVFGKPK	2	M, PP	136, 41.8	L, EX
88	At4g21280	Photosystem II subunit QA (PSBQ-1) *	VGPPPAPSGGLPGTDNSDQAR FYLQPLPPTEAAAR LFDTIDNLDYAAK YYAETVSALNEVLAK	1	M, PP	109, 28.3	EX
89	At4g05180	Photosystem II subunit Q-2 (PSBQ-2) *	AWPYVQNDLR YYSETVSSLNNVLAK VGGPPLPSGGLPGTDNSDQAR LFQTIDNLDYAAR FYIQPLSPTEAAAR FQTIDNLDYAAR DLTAKLFQTIDNLDYAAR GPPLPSGGLPGTDNSDQAR EAIPIKVGGPPLPSGGLPGTDNSDQAR PSGGLPGTDNSDQAR	3	M, PP	284, 32.2	EX

90	At1g76100	Plastocyanin 1 (PETE1) ∗	NNAGFPHNVVFDEDEIPSGVDASK	1	PP	14	L, EX
91	At1g11860	Glycine cleavage T-protein family ∗	IGEITSGGFSPNLK VGFFSSGPPAR	2	M, PP	126, 6.1	L
92	At1g32470	Single hybrid motif superfamily protein / Glycine cleavage system H protein ∗	VKPSSPAELEALMGPK LTESPGLINSSPYEDGWMIK	2	M, PP	88, 21.7	-
93	At3g14420	Aldolase-type TIM barrel family protein	AIALTVDTPR	1	PP	2.7	-
94	At1g20440	Cold-regulated 47 (COR477) ∗	GLFDFLGKK ITLLEELQEK EGLLQLVSDK ALLDDPVFRPLVEK AEDSPAVTSTPLVVTEHPVEPTTEL PVE-HPEEK AEEYKNNVPEHETPTVATEESPAT TTEV-TDR	1	M, PP	90, 27.9	EX
95	At1g20450	Dehydrin family protein (ERD10) ∗	VATEESSAPEIK TQISEPESFVAK AEEYKNTVPEQETPK AEEYKNTVPEQETPKVATEESSAP EIK TVEEENQGVMDR	2	M, PP	137, 19.6	L, EX
96	At1g76180	Dehydrin family protein (ERD14) ∗	VATEESSAEVTDR VHISEPEPEVK AEEIKNVPEQEVPK AEEIKNVPEQEVPKVATEESSAEVT DR HESLLEK PPPVEEAHPVEK	3	M, PP	126, 30.8	-
97	At1g70830	MLP-like protein 28 (MLP28) ∗	LETDVEIK TPETPSLVGK ADVATKHPMEDEVK	2	M, PP	91, 7.2	EX

			ASADKFHHMFAGKP IEAVEPDKNLITFR				
98	At1g70890	MLP-like protein 43 (MLP43) ∗	VLEGDLMNEYK VGSIVIWK AEASSLVGK SFAFTLQVTPK EIDEHLLAEE	2	M, PP	89, 24.7	EX
99	At1g23130	Polyketide cyclase/dehydrase and lipid transport superfamily protein (Bet v I allergen-like) ∗	LLELFAK SFLVTIQATPK VGSVIFWNYAIDGQPK	3	M	94	EST
100	At4g23670	Polyketide cyclase/dehydrase and lipid transport superfamily protein ∗	SWDYTYDGK NDDFPEPSGYMK	1	M	120	EX
101	At5g39730	AIG2-like (avirulence induced gene) family protein ∗	TVEVVLTDTSEK IPEIVSATLPGFK	2	M	88	L, EX
102	At3g04720	Pathogenesis-related 4 (PR4) / Hevein-like protein (HEL) ∗	TNAAVTVR ATYHFYNPAQNNWDLR AVSAYCSTWDADKPYAWR YGWTAFCGPAGPR	2	M, PP	111, 25.9	EX
103	At2g43530	Scorpion toxin-like knottin superfamily protein ∗	CFCDFCSDKP GYIDAPLSYCMAR	2	M, PP	79, 27.1	-
104	At5g09810	Actin 7 (ACT7) ∗	EITALAPSSMK GYMFTTTAER NYELPDGQVITIGAER PEEHPVLLTEAPLNPK VAPEEHPVLLTEAPLNPK	1	PP	14.6	L, EX
105	At5g59880	Actin-depolymerizing factor 3 (ADF3) ∗	IFFVAWSPDTAR ANAASGMAVHDDCK IGEPGQTHEDLAASLPADECR	1	M, PP	85, 33.8	L, EX

106	At1g35720	Annexin 1 (ANNAT1) *	AQINATFNR YEGDEVNMTLAK YQDDHGEEILK TLDKELSNDFER VSDSVPAPSDDAEQLR DALLANEATKR VIGEEYQR TSTQLLHAR HYNDEDVIR AILLWTLEPGER	2	M	288	EX
107	At2g19760	Profilin 1 (PRO1) *	YMVIQGEQGAVIR DFEEPGFLAPTGLFLGGEK	2	M, PP	84, 24.4	L, EX
108	At4g29350	Profilin 2 (PRO2) *	DFEEAGHLAPTGLFLGGEK LGDYLIESGL YMVVQGEAGAVIR	2	M, PP	72, 32.1	L, EX
109	At5g56600	Profilin 3 (PRO3) *	YMVIQGEPNAVIR	1	M, PP	71, 9.9	-
110	At3g63190	Ribosome recycling factor, chloroplast precursor (RRF) *	AIVNSDLGVTPNNDGDVIR ATIEEIEAEK	2	M, PP	57, 10.6	-
111	At4g38740	Rotamase CYP 1 (ROC1) *	FEDENFER IVMELYTDK VGSSSGKPTKPVVVADCGQ VIPNFMCQGGDFTAGNGTGGESIY GSK VYFDMTIDGQPAGR MELYTDKTPR HVVFGQVVEGLDVVK	4	M, PP	68, 41.9	EX, EST
112	At2g16600	Rotamase CYP 3 (ROC3) *	TSKPVVIADCGQIS YADTTPETAENFR	1	PP	15.6	L, EX, EST

113	At3g62030	Rotamase CYP 4 (ROC4) *	FEDENFTLK IVMGLFGEVVPK AAEEEEVIEPQAK HVVFGQVIEGMK TLESQETR AFDVPK DENFTLK	5	M, PP	77, 20.8	-
114	At4g34870	Rotamase CYP 5 (ROC5) *	HVVFGQVVK VIPGFMCQGGDFTAK IEMELFADTTPNTAENFR PGFMCQGGDFTAK VIPGFMCQGGDFTAK NGTGGESIYGAK	2	M, PP	140, 24.4	L, EST
115	At5g15970	Stress-responsive protein (KIN2) / stress-induced protein (KIN2) / cold-responsive protein (COR6.6) / cold-regulated protein (COR6.6) (AtCor6.6_COR6.6_KIN2)	DAAAAAGASAQQAGK SETNKNAFQAGQAAGK SETNKNAFQAGQAAGKAEEK NAFQAGQAAGK SISDAAVGGVN SISDAAVGGVNFVK	2	PP	74.2	EX
116	At5g62690	Tubulin beta-2/beta-3 chain (TUB2) *	VSEQFTAMFR YTGDSDLQLER	1	PP	2.2	L, EX
117	At4g14960	Tubulin/FtsZ family protein (TUA6) *	TVGGGDDAFNTFFSETGAGK	1	PP	4.4	-
118	At2g27030	Calmodulin 5 (CAM5) *	ADQLTDDQISEFK	2	PP	8.7	L, EX, EST
119	At3g43810	Calmodulin-7 (CAM7) *	ADQLTDDQISEFK	2	PP	8.7	-
120	At5g38480	General regulatory factor 3 (GRF3_RCI1) *	TVDVEELSVEER	1	PP	8.6	EX
121	At4g09000	General regulatory factor 1 (GF14 CHI_GRF1) *	YEEMVEFMEK GASSARDEFVYMAK	1	PP	9	L, EX

122	At3g02520	General regulatory factor 7 (GF14 NU_GRF7) *	EAAESTLVAYK NLLSVAYK SREENVYLAK TVDTDELTVEER	1	PP	15.5	EX
123	At1g22300	General regulatory factor 10 (14-3-3EPSILON_GF14 EPSILON_GRF10) *	ILSSIEQK YLAEFSSGAER YDEMVEAMK	3	M, PP	65, 7.8	EX
124	At4g20260	Plasma-membrane associated cation-binding protein 1 (PCAP1) *	VSVFLPEEVK TKEIPVEEVK DETPETAVVEEK	1	M	156	EX
125	At3g59920	RAB GDP dissociation inhibitor 2 (GDI2)	VIGVTSEGETAK AVDGSYVFVKGK	1	M	72	EX
126	At4g04910	AAA-type ATPase family protein (NSF) *	VSTGDMVSVSR	2	M,PP	82, 1.5	-
127	At5g20720	Chaperonin 20 (CPN20) *	EKPSIGTVIAVGPGSLDEEGK ITPLPVSTGSTVLYSK TLGGILLPSTAQSKPQGGEVVAVG EGRDGSNYIALR	2	M, PP	70, 25.3	L, EX
128	At5g03940	Chloroplast signal recognition particle 54 kDa subunit (CPSRP54P) *	IVHDELVKLMGGEVSELQFAK	1	PP	3.7	-
129	At5g50920	CLPC homologue 1 (CLPC1) *	VLENLGADPSNIR	1	M	55	EX
130	At1g72730	DEA(D/H)-box RNA helicase family protein *	GLDVIQQAQSGTGK	1	PP	3.4	-
131	At5g45680	FK506-binding protein 13 (FKBP13) *	IPPELAYGDR	1	M	60	-
132	At5g02500	Heat shock cognate protein 70-1 (HSC70-11) *	NALENYAYNMR FELSGIPPAPR	1	PP	3.4	L, EX
133	At5g60390	GTP binding Elongation factor Tu family protein *	IGGIGTVPVGR LPLQDVYK RSFKYAWVLDK*	2	PP	4.3	-
134	At5g02240	NAD(P)-binding Rossmann-fold superfamily protein *	LGNGNILVWK	2	M, PP	62, 4	EX

			ANLPTVLVTGASGR*				
135	At3g12390	Nascent polypeptide-associated complex (NAC), alpha subunit family protein *	APDLSNVISK	1	PP	4.9	EX
136	At4g20360	RAB GTPase homolog E1B (ATRABE1B) *	ILDEALAGDNVGLLLR VGETVDLVGLR	3	PP	3.4	EX
137	At2g47170	Ras-related small GTP-binding family protein (ARF1A1C)*	QDLPNAMNAAEITDK	1	PP	8.3	L, EX
138	At5g58290	Regulatory particle triple-A ATPase 3 (RPT3) *	GVLLYGPPGTGK	1	PP	2.9	-
139	At1g56070	Ribosomal protein S5/Elongation factor G/III/V family protein (LOS1)	GVQYLNEIK EGPLAEENMR	1	M	55	EX
140	At1g31340	Related to ubiquitin 1 (RUB1) *	IQDKEGIPPDQQR DQDKEGIPPDQQR TLADYNIQK ADYNIQK	1	M, PP	73, 11.8	EX
141	At5g03240	Polyubiquitin (UBQ3) *	TLADYNIQK IQDKEGIPPDQQR	1	M	55	EX, EST
142	At3g16640	Translationally controlled tumor family protein (TCTP) *	VVDIVDTFR LSEEDQAVFK	4	M, PP	102, 11.9	L, EX
143	At2g01320	ABC-2 type transporter family protein （ABCG7) *	MGKSQTSIQDR	1	PP	1.5	-
144	At1g72150	PATELLIN 1 (PATL1) *	IGSTDEPVITDSFK	1	PP	2.4	EX
145	At1g22840	CYTOCHROME C-1 (CYTC-1) *	QSGTTAGYSYSAANK QGPNLNGLFGR ASFDEAPPGNAKAGEK	3	M, PP	42, 36.9	L, EX
146	At1g67280	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein *	GPTPEPLCQVMLR	1	M	62	-
147	At1g11840	Glyoxalase I homolog (GLX1) *	GGGSVIAFVK EAGPLPGLGTK TVLVDNKDFLK DPDGYTFELIQR GPTPEPFCQVMLR	1	M	314	L, EX

148	At5g38470	Rad23 UV excision repair protein family (RAD23D) *	PQAVTVTPEER	1	PP	2.9	-
149	At3g20390	Endoribonuclease L-PSP family protein *	APAALGPYSQAIK FVSESVEDQTEQVLK STYQVAALPLNAK	1	M	163	L, EX
150	At2g21660	Glycine-rich RNA-binding protein (GRP7) *	DAIEGMNGQDLDGR ALETAFAQYGDVIDSK SITVNEAQSR GFGFVTFKDEK	3	PP	22.7	-
151	At4g39260	Glycine-rich RNA-binding protein 8 (GRP8) *	VITVNEAQSR TFSQFGDVIDSK	3	M, PP	73, 13	-
152	At4g34110	Polyadenylate-binding protein 2 (PABP2) *	GFGFVNFENADDAAR	1	M, PP	73, 2.4	L, EX
153	At1g49760	Poly(A) binding protein 8 (PABP8) *	GSGFVAFSTPEEATR GFGFVNFENSDDAAR AITEMNGKMIVTKPLYVALAQR	1	M	60	-
154	At2g23380	SET domain-containing protein (CLF / SET1) *	VAERVPR	1	PP	0.8	-
155	At3g61260	Remorin family protein *	IALESESPAK	1	M	69	EX
156	At4g24770	31-kDa RNA binding protein (RBP31) *	VYVGNLPWDVDNGR	1	M, PP	96, 4.3	EX
157	At1g29660	GDSL-like Lipase/Acylhydrolase superfamily protein *	ITFSGQVENYK DQLNALYNYGAR QYTPEQYADDLISR ADYFPYGIDFGGPTGR	1	M	289	-
158	At4g02520	Glutathione S-transferase PHI 2 (GSTF2)	VNEWVAEITK NPFGQVPAFEDGDLK NPFGQVPAFEDGDLKL	2	PP	12.3	EX
159	At2g30870	Glutathione S-transferase PHI 10 (GSTF10) *	SQGPDLLGK QPEYLAIQPFGK LAEVLDVYEAQLSK QPEYLAIQPFGK	3	M, PP	84, 10.2	L, EX

			IPVLVDGDYK SQGPDLLGK				
160	At3g16400	Nitrile specifier protein 1 (NSP1)	LLTPVEEGPTPR FEYVNGSQVVVGDEHGK FGGEEETPSSR GGAGLEVVQGK IYSFGGEFTPNQPIDK LNTLDSYNIVDK QYNGFYSFDTTTNEWK VYVGQAQYGIAFVK	2	PP	23.2	EX
161	At3g16410	Nitrile specifier protein 4 (NSP4)	LLTPVEEGPTPR IYSFGGEFTPNQPIDK	1	M	140	EX
162	At3g10720	Plant invertase/pectin methylesterase inhibitor superfamily *	VYTPSTSLR DYYVVVSSR LPLPDGILINGR QYLGQQLYLR FTSNVLTTTGIFR YAVNSVSFKPADTPLK	2	M	55	-
163	At1g78850	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain	WVWEANR VPVDDQFR IEGADSFMTK LVLYYTTNK TGDSSLVAYVK ILENGNMVIYDSSGK VVNEGGYTDYSPIEYNPDVR	2	M	252	EX
164	At4g20830	FAD-binding Berberine family protein *	IGLVFNPYGGK NEQSIPTVLSK	2	M, PP	74, 1.9	-
165	At3g05900	Neurofilament protein-related *	AVPTNQDLDTEPK	1	M	62	-
166	At1g09310	Protein of unknown function, DUF538*	EISMPNGLLPLK	1	M	125	EX

			TFPVTAFIVPEEPAK GLVTEEVR				
167	At1g55480	Protein containing PDZ domain, a K-box domain, and a TPR (ZKT)region *	IGPLLMQMEK	1	M	57	-
168	At5g01750	Protein of unknown function (DUF567) *	VLLDGSGTPVVTLR YCAPYPIDMAIVR	1	M, PP	63, 12.4	EX, EST
169	At5g22580	Stress responsive A/B Barrel Domain *	GLENLVSQIDTVK	2	M, PP	55, 11.7	EX
170	At3g02910	AIG2-like (avirulence induced gene) family protein *	LDELEGISR	1	PP	4.9	-
171	At5g19240	Glycoprotein membrane precursor GPI-anchored *	NTDGENLLFTVFNQYR	1	PP	8	EX
172	At5g56170	LORELEI-LIKE-GPI-ANCHORED PROTEIN 1 (LLG1) *	YPPGLFANQCK	1	M	55	-
173	At2g37660	NAD(P)-binding Rossmann-fold superfamily protein *	AATTEPLTVLVTGAGGR	2	PP	5.2	-
174	At2g44920	Tetratricopeptide repeat (TPR)-like superfamily protein *	VADGVNATTGNATR	1	PP	6.3	-
175	At5g53490	Tetratricopeptide repeat (TPR)-like superfamily protein *	FCDYTNDQTNLK	1	PP	5.1	EX
176	At5g64570	Beta-D-xylosidase 4 (XYL4) *	LGFFDGNPK GQETPGEDPLLASK LPMTWYPQSYVEK NTGCLPLSPK AISNNFLTLMR TMIGNYEGTPCK	1	M	97	EX
177	At1g03220	Extracellular dermal glycoprotein, putative, EDGP	DQSTLQYTTVINQR	1	M	83	EX
178	At5g55730	Fasciclin-like arabinogalactan-protein (FLA1) *	VFAETLLTNPGASK VGFGPDGGDLSSFFVK IVDTLIDEQPLAIYATDK	1	M, PP	145, 11.3	-
179	At4g12730	Fasciclin-like arabinogalactan-protein (FLA2) *	VAFGVQDDDSK SGNGAVNTLATEGNNK DQEPLIVYK QGCKAFSDILK STGADKTFQDTVDGGLTVFCPSDS AVG-K	2	M, PP	97 14.4	-

			TALVLYHGMPVYQSLQMLR AFSDILK				
180	At2g04780	Fasciclin-like arabinogalactan-protein (FLA7) *	VSSSVFSTDPVAVYQVNR	1	M	81	-
181	At2g45470	Fasciclin-like arabinogalactan-protein (FLA8) *	TFANLLVSSGVLK GLTVFAPSDEAFK NNISTLATNGAGK	1	M, PP	71, 6.2	-
182	At1g03870	Fasciclin-like arabinogalactan-protein (FLA9) *	YYSMDDLLSVSNPVR	1	M, PP	102, 6.1	EX
183	At3g60900	Fasciclin-like arabinogalactan-protein (FLA10) *	TFANLLVSSGVIK GLTVFAPSDEAFK IFFVAWSPDTAR	1	M, PP	69, 6.2	-
184	At5g44130	Fasciclin-like arabinogalactan-protein (FLA13) *	GGQFVTLIR	2	PP	3.6	EX
185	At1g29670	GDSL-like Lipase/Acylhydrolase superfamily protein *	ISFSGQVR INSANQIFNNK QILSGVNYASAAAGIR YSTQLNALYNYGAR QFTPEQYANDLISR SYNAQSASDAYPMDISR SLVDQLNNNHPDAK SNYFPYGIDFGGPTGR	1	M	457	EX
186	At5g20630	Germin 3 (GER3)	NPDQVTENDFAFTGLGK GPQSPSGYSCK AAVTPAFAPAYAGINGLGVSLAR SVQDFCVADPK	2	M, PP	89, 21.3	-
187	At4g16260	Glycosyl hydrolase superfamily protein *	LYDPNQAALNALR	1	M	91	EX
188	At1g09750	Eukaryotic aspartyl protease family protein *	ILFDVPNSR FAQPVYEAIR	2	M	87	-
189	At5g20080	FAD/NAD(P)-binding oxidoreductase *	IFYTVDNPTK	1	PP	3	-
190	At1g75040	Pathogenesis-related protein 5 (PR-5) *	QLTAPAGWSGR	5	M, PP	376, 57.7	EX

			LGDGGFELTPGASR FNTDQYCCR VMDQNNVVACK GANDKPETCPPTDYSR NNCPTTVWAGTLAGQGPK YAGCVSDLNAACPDMLK DFYDVSLVDGYNVK TGCNFDASGNGR CVTGDCGGLR TDFTLRNNCPTTVWAGTLAGQGPK YAGCVSDLNAACPDMLK				
191	At1g76160	SKU5 similar 5 (sks5) *	VYTPSTSLR DYYVVVSSR LPLPDGILINGR QYLGQQLYLR FTSNVLTTTGIFR	1	M	171	EX
192	At3g49120	Peroxidase CB (PRXCB) *	MGNITPTTGTQGQIR TPTVFDNK GLDRPSDLVALSGGHTFGK NVGLDRPSDLVALSGGHTFGK NQCQFILDR	5	M, PP	119, 15	EX
193	At4g37800	Xyloglucan endotransglucosylase/hydrolase 7 (XTH7) *	TEYLFGK IILTVDDTPIR LVPGNSAGTVTTFYLK LDPSSGCGFASK PSSGCGFASK SRFPVPPPECSAGI	3	M, PP	163, 8.9	-

			APFYAYYK VNLWFDPSR				
194	At4g30270	Xyloglucan endotransglucosylase/hydrolase 24 (XTH24) *	ILNNGQLLTLSLDK NYESLGVLFPK MYASLWNADDWATR LVPGNSAGTVTTFYLK PGNSAGTVTTFYLK IILTVDDTPIR TEYLFGK	3	M, PP	163, 14.5	EX
195	At2g02850	Plantacyanin (ARPN) *	AGDVLVFNYNPR	1	PP	9.3	-
196	At3g16420	PYK10-binding protein 1 (JAL30_PBP1)	GGAGLEVVQGK VYVGQAQYGIAFVK WTQVETFGVRPSER QYNGFYSFDTTTNEWK LGVNVAPIAK IFSVDAPAIVSLK TSQPFGLTSGEEAELGGGK GANLWDDGSTHDAVTK FVYDKSPEEVTGEEHGK	1	M	149	EX
197	At4g12420	Cupredoxin superfamily protein (SKU5) *	ASGGFGSFVVNPR GQLPPGPQDEFDK PPGPQDEFDK PPVTISGK	1	M, PP	87, 5.8	-
198	At5g67360	Subtilase family protein (ARA12) *	TVTSVGGAGTYSVK	2	M, PP	60, 1.8	-
199	At3g07390	Auxin-responsive family protein (AIR12) *	MAGSQAFLAYR LAFDFWNLR	2	M, PP	91, 7.9	-
200	At4g27520	Early nodulin-like protein 2 (ENODL2) *	VDDGDSEISLDR YGPFYFISGNEDNCK	2	PP	7.7	EX, EST

201	At2g27710	60S acidic ribosomal protein family *	ASPTSADIK GKDLAELIAAGR	1	PP	7.8	L, EX
202	At1g32990	Plastid ribosomal protein l11 (PRPL11) *	ATPAPPVGPALGSK	1	PP	6.3	-
203	At2g18020	Ribosomal protein L2 family *	GVVTEIIHDPGR	1	PP	4.7	L, EX
204	At3g25520	Ribosomal protein L5 (RPL5A) *	ALLDVGLIR	1	PP	3	EX
205	At1g74050	Ribosomal protein L6 family protein	VDISGVTLDK	1	PP	4.3	L, EX
206	At3g62870	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein *	VTNPLFER ANFNDKYEEYR LKVPPALNQFTK	1	PP	7.4	L, EX
207	At2g44120	Ribosomal protein L30/L7 family protein *	VVVPESVLK	1	PP	3.7	EX
208	At3g09200	Ribosomal protein L10 family protein *	GTVEIITPVELIK	1	PP	4.1	L, EX, EST
209	At5g60670	Ribosomal protein L11 family protein *	VTGGEVGAASSLAPK	2	M, PP	89,9	L
210	At3g27850	Ribosomal protein L12-C (RPL12-C) *	IGSEISSLTLEEAR ILVDYLQDK	2	M, PP	122, 12.3	EX
211	At2g42740	Ribosomal protein large subunit 16A (RPL16A) *	AMQLLESGLK VLEQLSGQTPVFSK	1	PP	13.2	-
212	At3g55280	Ribosomal protein L23AB (RPL23A2_RPL23AB) *	LTPDYDALDVANK (88%)	1	PP	0	L, EX
213	At3g53870	Ribosomal protein S3 family protein	ELAEDGYSGVEVR	1	PP	5.2	L, EX
214	At2g17360	Ribosomal protein S4 (RPS4A) family protein *	EVISILMQR FDVGNVVMVTGGR LTIIEEAR	1	PP	11.5	L, EX
215	At2g41840	Ribosomal protein S5 family protein *	ALSTSKPDPVVEDQA IEQIYLHSLPVK	2	PP	5.3	EX
216	At2g37270	Ribosomal protein 5B (RPS5B) *	VNQAIFLLTTGAR	1	PP	6.3	L, EX
217	At5g10360	Ribosomal protein S6e (RPS6B) *	LVTPLTLQR PGLTDTEKPR	1	PP	7.6	L, EX
218	At5g20290	Ribosomal protein S8e family protein *	SAIVQVDAAPFK SLDSHIEDQFASGR	1	PP	11.7	EX

219	At3g11510	Ribosomal protein S11 family protein ∗	IEDVTPIPTDSTR DVTPIPTDSTR	2	M, PP	71, 8.7	L, EX
220	At4g00100	Ribosomal protein S13A (RPS13A) ∗	GISASALPYK	1	PP	6.6	L, EX
221	At1g22780	Ribosomal protein S13/S18 family (RPS18A) ∗	YSQVVSNALDMK	1	PP	7.9	L, EX

NB: L=Laser, EX=exudate, EST=expressed sequence tag, M=MASCOT, PP=ProteinPilot

Appendix B

Identified proteins from *Arabidopsis* phloem sap with biological function according to GO analysis of the Biological Process.

Represents all proteins identified by Batailler et al (2012) in the phloem exudate
Represents all proteins identified by Batailler et al (2012) in the full set of exudates
Proteins with shared homology in all four species are written in blue
∗ Protein found in translatome study by Mustroph et al (2009)

No	Protein	AGI No.	Cell Organisation and Biogenesis	Developmental Processes	DNA/RNA Metabolism	Electron transport / Energy Pathway	Protein Metabolism	Other Biological Processes	Other Cellular processes	Other Metabolic Processes	Response to Abiotic / Biotic Stress	Response to Stress	Signal Transduction	Transport	Unknown Biological Processes	Unknown
1	At1g07890	Ascorbate peroxidase 1 (APX16) ∗										√				
2	At4g35090	Catalase 2 (CAT) ∗										√				
3	At1g20620	Catalase 3 (CAT3) ∗										√				
4	At1g08830	Copper/zinc superoxide dismutase (CSD1) ∗						√	√	√		√				
5	At2g28190	Copper/zinc superoxide dismutase (CSD2) ∗						√	√	√		√				
6	At1g19570	Dehydroascorbate reductase (DHAR1) ∗												√		
7	At4g11600	Glutathione peroxidase 6 (GPX6) ∗												√		

8	At2g24940	Membrane-associated progesterone binding protein 2 (MAPR2) *													√		
9	At5g40370	Membrane-associated progesterone binding protein 2 (MAPR2) *							√								
10	At3g52880	Monodehydroascorbate reductase 1 (MDAR1) *					√		√	√		√					
11	At1g63940	Monodehydroascorbate reductase 6 (MDAR6) *						√				√					
12	At5g03630	Pyridine nucleotide-disulphide oxidoreductase family protein (MDAR2) *						√				√					
13	At1g45145	Thioredoxin 3 (TRX3)*										√					
14	At5g16400	Thioredoxin F2 (TRXF2) *										√					
15	At3g02730	Thioredoxin F-type 1 (TRXF1) *										√					
16	At5g42980	Thioredoxin H-type 3 (TRX-H-3) *										√					
17	At1g03680	Thioredoxin M-type 1 (TRX-M1)										√					
18	At3g15360	Thioredoxin M-type 4 (TRX-M4) *										√					
19	At1g43560	Thioredoxin Y2 (Aty2_ty2) *										√					
20	At2g20270	Thioredoxin superfamily protein *							√								
21	At3g11630	Thioredoxin superfamily protein *				√						√					
22	At3g52960	Thioredoxin superfamily protein)*					√		√	√		√					
23	At4g03520	Thioredoxin superfamily protein (ATHM2) *										√					
24	At2g24270	Aldehyde dehydrogenase 11A3 (ALDH11A3) *								√							
25	At2g36530	Enolase (ENO2) / (LOS2) *						√				√					
26	At5g42740	Glucose-6-phosphate isomerase, cytosolic (GPI) (phosphoglucose isomerase) (PGI) *				√		√	√	√		√					
27	At3g04120	Glyceraldehyde-3-phosphate dehydrogenase C subunit 1 (GAPC1) *				√		√	√	√		√					
28	At3g55440	Triosephosphate isomerase (TPI) *				√	√	√	√	√		√					
29	At3g03250	UDP-glucose pyrophosphorylase 1 (UGP1) *							√	√		√					
30	At5g17310	UDP-glucose pyrophosphorylase 2 (UGP2) *						√	√	√		√					
31	At5g09660	Peroxisomal NAD-malate dehydrogenase 2 (PMDH2) *										√					

32	At3g01500	Carbonic anhydrase 1 (CA1) *						√	√			√				
33	At5g14740	Carbonic anhydrase 2 (CA2) *						√				√				
34	At1g04410	Lactate/malate dehydrogenase family protein (c-NAD-MDH1) *										√				
35	At1g53240	Lactate/malate dehydrogenase family protein (mMDH1) *										√				
36	At5g11670	NADP-malic enzyme 2 (NADP-ME2) *					√			√		√				
37	At4g14880	O-acetylserine (thiol) lyase (OAS-TL) isoform A1 (OASA1) *										√				
38	At4g13940	S-adenosyl-L-homocysteine hydrolase (SAHH1) *							√	√		√				
39	At3g23810	S-adenosyl-l-homocysteine (SAH) hydrolase 2 (SAHH2) *							√	√						
40	At1g02500	S-adenosylmethionine synthetase 1 (SAM1) *						√	√	√		√				
41	At4g31990	Aspartate aminotransferase 5 (ASP5) *						√				√				
42	At5g17920	Cobalamin-independent synthase family protein (ATCIMS)						√	√	√		√				
43	At1g31812	Acyl-CoA-binding protein 6 (ACBP6) *										√		√		
44	At4g25050	Acyl carrier protein 4 (ACP4) *												√		
45	At5g16390	Biotin carboxyl carrier protein of acetyl-CoA carboxylase precursor (BCCP) *							√	√						
46	At1g27950	Glycosylphosphatidylinositol-anchored lipid protein transfer 1 (LTPG1) *												√		
47	At2g38540	Lipid transfer protein 1 (LTP1) *												√		
48	At2g38530	Lipid transfer protein 2 (LTP2) *	√	√					√			√		√		
49	At1g66970	SHV3-like 2 (SVL2) *								√						
50	At3g09820	Adenosine kinase 1 (ADK11) *										√				
51	At5g03300	Adenosine kinase 2 (ADK2) *										√				
52	At4g09320	Nucleoside diphosphate kinase family protein (NDPK1) *						√				√				
53	At5g63310	Nucleoside diphosphate kinase 2 (NDPK2) *	√		√					√	√					
54	At5g09650	Pyrophosphorylase 6 (PPa6) *												√		

55	At3g24503	Aldehyde dehydrogenase 2C4 (ALDH1A / ALDH2C4) *							√	√						
56	At1g54040	Epithiospecifier protein (ESP) *							√	√		√				
57	At3g44310	Nitrilase 1 (NIT1)						√								
58	At5g26000	Thioglucoside glucohydrolase 1 (TGG1)							√	√						
59	At2g41530	S-formylglutathione hydrolase (SFGH) *												√		
60	At3g59970	Methylenetetrahydrofolate reductase 1 (MTHFR1)							√	√						
61	At3g16430	Jacalin-related lectin 31 (JAL31) *														√
62	At3g16470	Mannose-binding lectin superfamily protein (JAL35_JR1) *						√				√				
63	At1g12240	Glycosyl hydrolases family 32 protein / Beta-fructosidase (ATBETAFRUCT4) *							√	√						
64	At5g37850	pfkB-like carbohydrate kinase family protein (SOS4) *										√				
65	At1g23740	Oxidoreductase, zinc-binding dehydrogenase family protein *								√		√				
66	At5g61130	Plasmodesmata callose-binding protein 1 (PDCB1) *														√
67	At5g35630	Glutamine synthetase 2(GS2) *										√				
68	At5g08280	Hydroxymethylbilane synthase (HEMC) *							√	√		√				
69	At5g63570	Glutamate-1-semialdehyde-2,1-aminomutase (GSA1) *							√	√		√				
70	At4g26500	Chloroplast sulfur E (SUFE1) *							√	√		√				
71	At5g54770	Thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) *										√				
72	At2g21330	Fructose-bisphosphate aldolase 1 (FBA1) *						√	√	√						
73	At4g38970	Fructose-bisphosphate aldolase 2 (FBA2) *								√		√				
74	At1g42970	Glyceraldehyde-3-phosphate dehydrogenase B subunit (GAPB) *				√		√	√	√		√				
75	At3g12780	Phosphoglycerate kinase 1 (PGK1) *								√		√				
76	At1g56190	Phosphoglycerate kinase family protein *										√				

77	At2g21170	Plastid Isoform Triose Phosphate Isomerase (PDTPI) *	√					√	√	√		√				
78	At2g39730	Rubisco activase (RCA) *										√				
79	AtCg00490	Ribulose-bisphosphate carboxylases (RBCL)				√										
80	At1g67090	Ribulose bisphosphate carboxylase small chain 1A (RBCS1A) *				√					√					
81	At3g55800	Sedoheptulose-bisphosphatase (SBPASE) *							√	√		√				
82	At3g60750	Transketolase *						√				√				
83	At1g20340	Cupredoxin superfamily protein (DRT112_PETE2) *		√			√	√	√	√		√				
84	At5g66190	Ferredoxin-NADP(+)-oxidoreductase 1 (FNR1) *				√								√		
85	At1g20020	Ferredoxin-NADP(+)-oxidoreductase 2 (FNR2) *												√		
86	At1g77090	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein *												√		
87	At1g06680	Photosystem II subunit P-1 (PSBP-1) *		√							√					
88	At4g21280	Photosystem II subunit QA (PSBQ-1) *									√					
89	At4g05180	Photosystem II subunit Q-2 (PSBQ-2) *									√					
90	At1g76100	Plastocyanin 1 (PETE1) *														√
91	At1g11860	Glycine cleavage T-protein family *						√								
92	At1g32470	Single hybrid motif superfamily protein / Glycine cleavage system H protein *							√	√						
93	At3g14420	Aldolase-type TIM barrel family protein								√						
94	At1g20440	Cold-regulated 47 (COR477) *										√				
95	At1g20450	Dehydrin family protein (ERD10) *										√				
96	At1g76180	Dehydrin family protein (ERD14) *										√				
97	At1g70830	MLP-like protein 28 (MLP28) *										√				
98	At1g70890	MLP-like protein 43 (MLP43) *										√				
99	At1g23130	Polyketide cyclase/dehydrase and lipid transport superfamily protein (Bet v I allergen-like) *										√				
100	At4g23670	Polyketide cyclase/dehydrase and lipid transport superfamily protein *		√								√				

101	At5g39730	AIG2-like (avirulence induced gene) family protein *										√				
102	At3g04720	Pathogenesis-related 4 (PR4) / Hevein-like protein (HEL) *						√				√				
103	At2g43530	Scorpion toxin-like knottin superfamily protein *												√		
104	At5g09810	Actin 7 (ACT7) *	√	√					√			√				
105	At5g59880	Actin-depolymerizing factor 3 (ADF3) *		√				√				√				
106	At1g35720	Annexin 1 (ANNAT1) *										√				
107	At2g19760	Profilin 1 (PRO1) *	√						√							
108	At4g29350	Profilin 2 (PRO2) *	√						√							
109	At5g56600	Profilin 3 (PRO3) *	√						√							
110	At3g63190	Ribosome recycling factor, chloroplast precursor (RRF) *		√			√		√			√				
111	At4g38740	Rotamase CYP 1 (ROC1) *					√		√	√			√			
112	At2g16600	Rotamase CYP 3 (ROC3) *					√		√	√			√			
113	At3g62030	Rotamase CYP 4 (ROC4) *					√	√	√	√		√	√			
114	At4g34870	Rotamase CYP 5 (ROC5) *					√		√	√			√			
115	At5g15970	Stress-responsive protein (KIN2) / stress-induced protein (KIN2) / cold-responsive protein (COR6.6) / cold-regulated protein (COR6.6) (AtCor6.6_COR6.6_KIN2)										√				
116	At5g62690	Tubulin beta-2/beta-3 chain (TUB2) *						√				√				
117	At4g14960	Tubulin/FtsZ family protein (TUA6) *										√				
118	At2g27030	Calmodulin 5 (CAM5) *										√				
119	At3g43810	Calmodulin-7 (CAM7) *										√				
120	At5g38480	General regulatory factor 3 (GRF3_RCI1) *														√
121	At4g09000	General regulatory factor 1 (GF14 CHI_GRF1) *						√								
122	At3g02520	General regulatory factor 7 (GF14 NU_GRF7) *														√
123	At1g22300	General regulatory factor 10 (14-3-3EPSILON_GF14 EPSILON_GRF10) *						√	√				√			
124	At4g20260	Plasma-membrane associated cation-binding protein 1 (PCAP1) *										√				

125	At3g59920	RAB GDP dissociation inhibitor 2 (GDI2)								√				√		
126	At4g04910	AAA-type ATPase family protein (NSF) *														√
127	At5g20720	Chaperonin 20 (CPN20) *												√		
128	At5g03940	Chloroplast signal recognition particle 54 kDa subunit (CPSRP54P) *	√						√					√		
129	At5g50920	CLPC homologue 1 (CLPC1) *							√					√		
130	At1g72730	DEA(D/H)-box RNA helicase family protein *													√	
131	At5g45680	FK506-binding protein 13 (FKBP13) *					√		√	√						
132	At5g02500	Heat shock cognate protein 70-1 (HSC70-11) *						√	√	√		√				
133	At5g60390	GTP binding Elongation factor Tu family protein *					√		√	√						
134	At5g02240	NAD(P)-binding Rossmann-fold superfamily protein *						√								
135	At3g12390	Nascent polypeptide-associated complex (NAC), alpha subunit family protein *										√		√		
136	At4g20360	RAB GTPase homolog E1B (ATRABE1B) *					√		√	√						
137	At2g47170	Ras-related small GTP-binding family protein (ARF1A1C) *					√		√	√						
138	At5g58290	Regulatory particle triple-A ATPase 3 (RPT3) *					√		√	√						
139	At1g56070	Ribosomal protein S5/Elongation factor G/III/V family protein (LOS1)										√				
140	At1g31340	Related to ubiquitin 1 (RUB1) *					√		√	√						
141	At5g03240	Polyubiquitin (UBQ3) *		√			√		√	√		√				
142	At3g16640	Translationally controlled tumor family protein (TCTP) *		√				√	√			√				
143	At2g01320	ABC-2 type transporter family protein （ABCG7) *												√		
144	At1g72150	PATELLIN 1 (PATL1) *										√		√		
145	At1g22840	CYTOCHROME C-1 (CYTC-1) *												√		
146	At1g67280	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein *								√		√				
147	At1g11840	Glyoxalase I homolog (GLX1) *								√						
148	At5g38470	Rad23 UV excision repair protein family (RAD23D) *		√								√				

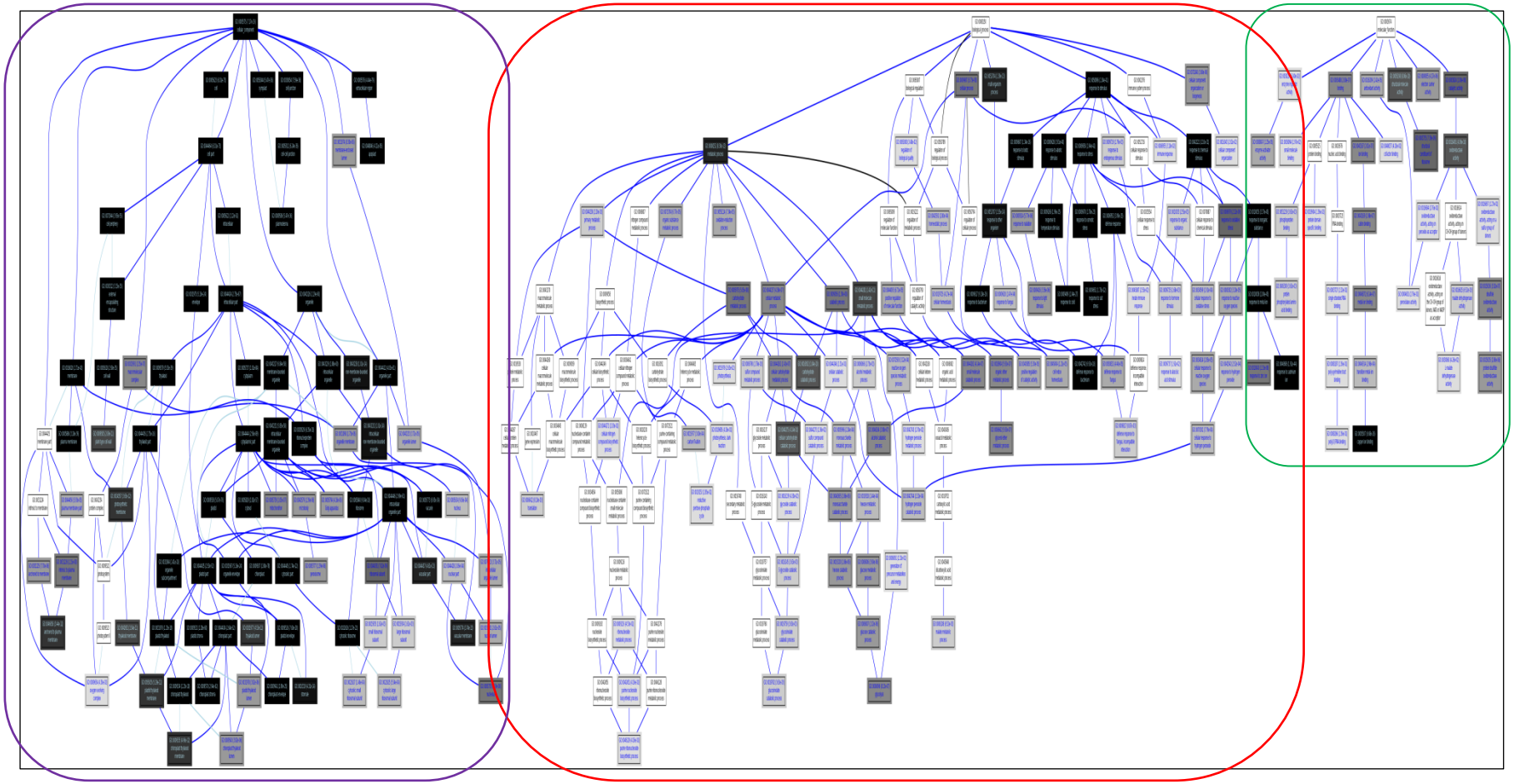
149	At3g20390	Endoribonuclease L-PSP family protein ∗						√								
150	At2g21660	Glycine-rich RNA-binding protein (GRP7) ∗		√	√			√	√	√		√		√		
151	At4g39260	Glycine-rich RNA-binding protein 8 (GRP8) ∗		√				√				√				
152	At4g34110	Polyadenylate-binding protein 2 (PABP2) ∗		√			√		√	√		√				
153	At1g49760	Poly(A) binding protein 8 (PABP8) ∗						√								
154	At2g23380	SET domain-containing protein (CLF / SET1) ∗		√								√				
155	At3g61260	Remorin family protein ∗													√	
156	At4g24770	31-kDa RNA binding protein (RBP31) ∗										√			√	
157	At1g29660	GDSL-like Lipase/Acylhydrolase superfamily protein ∗								√						
158	At4g02520	Glutathione S-transferase PHI 2 (GSTF2)												√		
159	At2g30870	Glutathione S-transferase PHI 10 (GSTF10) ∗												√		
160	At3g16400	Nitrile specifier protein 1 (NSP1)							√	√		√				
161	At3g16410	Nitrile specifier protein 4 (NSP4)								√						
162	At3g10720	Plant invertase/pectin methylesterase inhibitor superfamily ∗						√								
163	At1g78850	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain													√	
164	At4g20830	FAD-binding Berberine family protein ∗										√				
165	At3g05900	Neurofilament protein-related ∗													√	
166	At1g09310	Protein of unknown function, DUF538∗													√	
167	At1g55480	Protein containing PDZ domain, a K-box domain, and a TPR (ZKT)region ∗										√				
168	At5g01750	Protein of unknown function (DUF567) ∗													√	
169	At5g22580	Stress responsive A/B Barrel Domain ∗													√	
170	At3g02910	AIG2-like (avirulence induced gene) family protein ∗										√				
171	At5g19240	Glycoprotein membrane precursor GPI-anchored ∗													√	
172	At5g56170	LORELEI-LIKE-GPI-ANCHORED PROTEIN 1 (LLG1) ∗													√	
173	At2g37660	NAD(P)-binding Rossmann-fold superfamily protein ∗										√				

174	At2g44920	Tetratricopeptide repeat (TPR)-like superfamily protein *											√		
175	At5g53490	Tetratricopeptide repeat (TPR)-like superfamily protein *											√		
176	At5g64570	Beta-D-xylosidase 4 (XYL4) *							√						
177	At1g03220	Eukaryotic aspartyl protease family protein *									√				
178	At5g55730	Fasciclin-like arabinogalactan-protein (FLA1) *													√
179	At4g12730	Fasciclin-like arabinogalactan-protein (FLA2) *													√
180	At2g04780	Fasciclin-like arabinogalactan-protein (FLA7) *													√
181	At2g45470	Fasciclin-like arabinogalactan-protein (FLA8) *													√
182	At1g03870	Fasciclin-like arabinogalactan-protein (FLA9) *													√
183	At3g60900	Fasciclin-like arabinogalactan-protein (FLA10) *													√
184	At5g44130	Fasciclin-like arabinogalactan-protein (FLA13) *													√
185	At1g29670	GDSL-like Lipase/Acylhydrolase superfamily protein *							√						
186	At5g20630	Germin 3 (GER3)					√		√	√		√			
187	At4g16260	Glycosyl hydrolase superfamily protein *									√				
188	At1g09750	Eukaryotic aspartyl protease family protein *					√								
189	At5g20080	FAD/NAD(P)-binding oxidoreductase *											√		
190	At1g75040	Pathogenesis-related protein 5 (PR-5) *						√	√	√		√			
191	At1g76160	SKU5 similar 5 (sks5) *				√									
192	At3g49120	Peroxidase CB (PRXCB) *	√						√	√		√			
193	At4g37800	Xyloglucan endotransglucosylase/hydrolase 7 (XTH7) *							√	√					
194	At4g30270	Xyloglucan endotransglucosylase/hydrolase 24 (XTH24) *						√	√						
195	At2g02850	Plantacyanin (ARPN) *									√		√		
196	At3g16420	PYK10-binding protein 1 (JAL30_PBP1)							√	√					
197	At4g12420	Cupredoxin superfamily protein (SKU5) *				√									
198	At5g67360	Subtilase family protein (ARA12) *					√		√	√		√			
199	At3g07390	Auxin-responsive family protein (AIR12) *		√				√	√			√			
200	At4g27520	Early nodulin-like protein 2 (ENODL2) *												√	

201	At2g27710	60S acidic ribosomal protein family ∗					√		√	√		√				
202	At1g32990	Plastid ribosomal protein l11 (PRPL11) ∗					√		√	√						
203	At2g18020	Ribosomal protein L2 family ∗					√		√	√		√				
204	At3g25520	Ribosomal protein L5 (RPL5A) ∗					√		√	√						
205	At1g74050	Ribosomal protein L6 family protein					√		√	√						
206	At3g62870	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein ∗					√		√	√						
207	At2g44120	Ribosomal protein L30/L7 family protein ∗					√		√	√						
208	At3g09200	Ribosomal protein L10 family protein ∗					√		√	√		√				
209	At5g60670	Ribosomal protein L11 family protein ∗					√		√	√						
210	At3g27850	Ribosomal protein L12-C (RPL12-C) ∗					√		√	√		√				
211	At2g42740	Ribosomal protein large subunit 16A (RPL16A) ∗					√		√	√						
212	At3g55280	Ribosomal protein L23AB (RPL23A2_RPL23AB) ∗					√		√	√		√				
213	At3g53870	Ribosomal protein S3 family protein					√		√	√						
214	At2g17360	Ribosomal protein S4 (RPS4A) family protein ∗					√		√	√						
215	At2g41840	Ribosomal protein S5 family protein ∗					√		√	√						
216	At2g37270	Ribosomal protein 5B (RPS5B) ∗					√		√	√						
217	At5g10360	Ribosomal protein S6e (RPS6B) ∗					√		√	√						
218	At5g20290	Ribosomal protein S8e family protein ∗					√		√	√						
219	At3g11510	Ribosomal protein S11 family protein ∗					√		√	√						
220	At4g00100	Ribosomal protein S13A (RPS13A) ∗					√		√	√						
221	At1g22780	Ribosomal protein S13/S18 family (RPS18A) ∗					√		√	√						

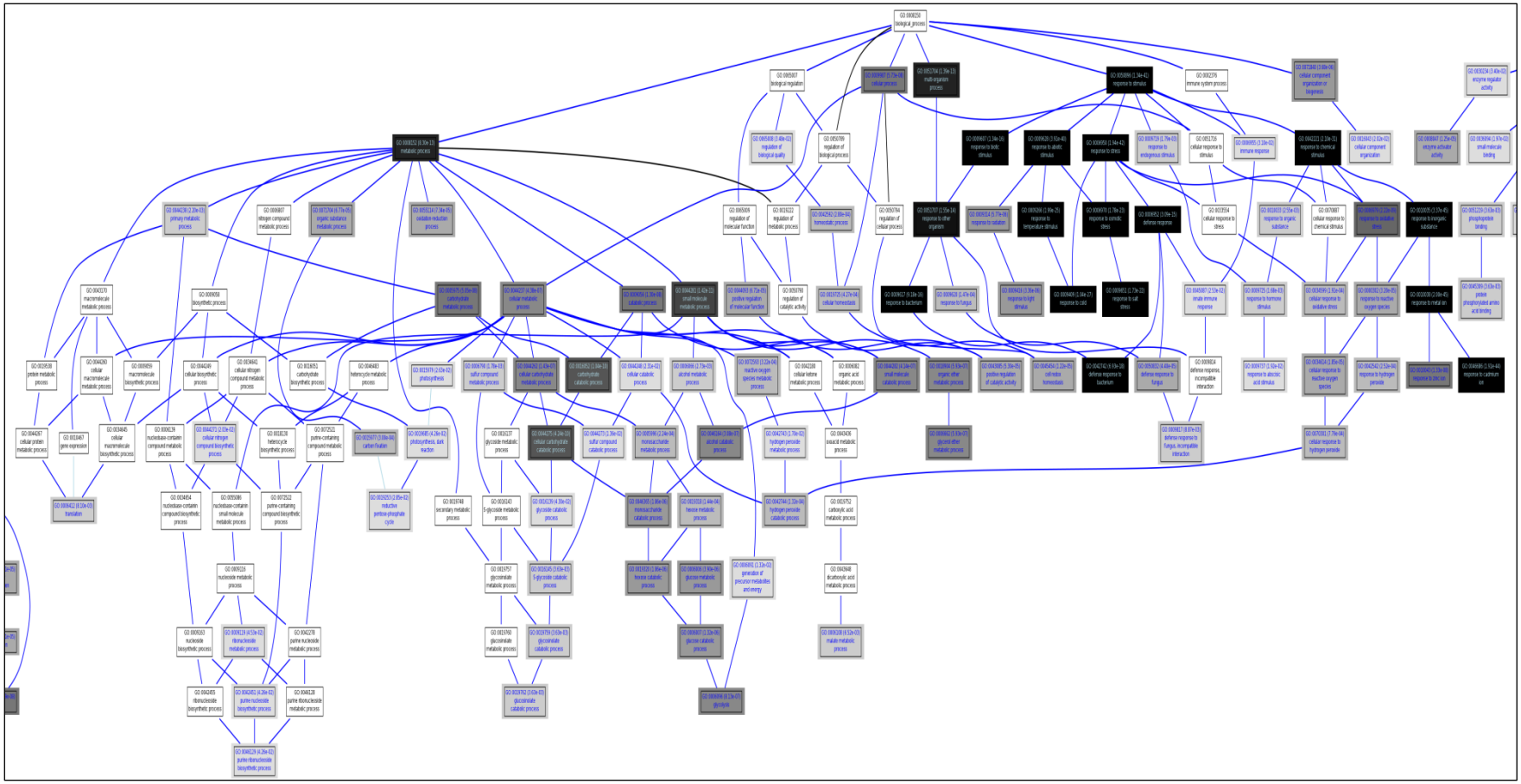
Appendix C-1

Overview of GO Term Enrichment against TAIR database using p-value cutoff of 0.05 with two minimum number of gene product. Close-up of all three individual categories will be shown subsequently. Green rectangle represents molecular function, red represents biological process and purple represents cellular components.



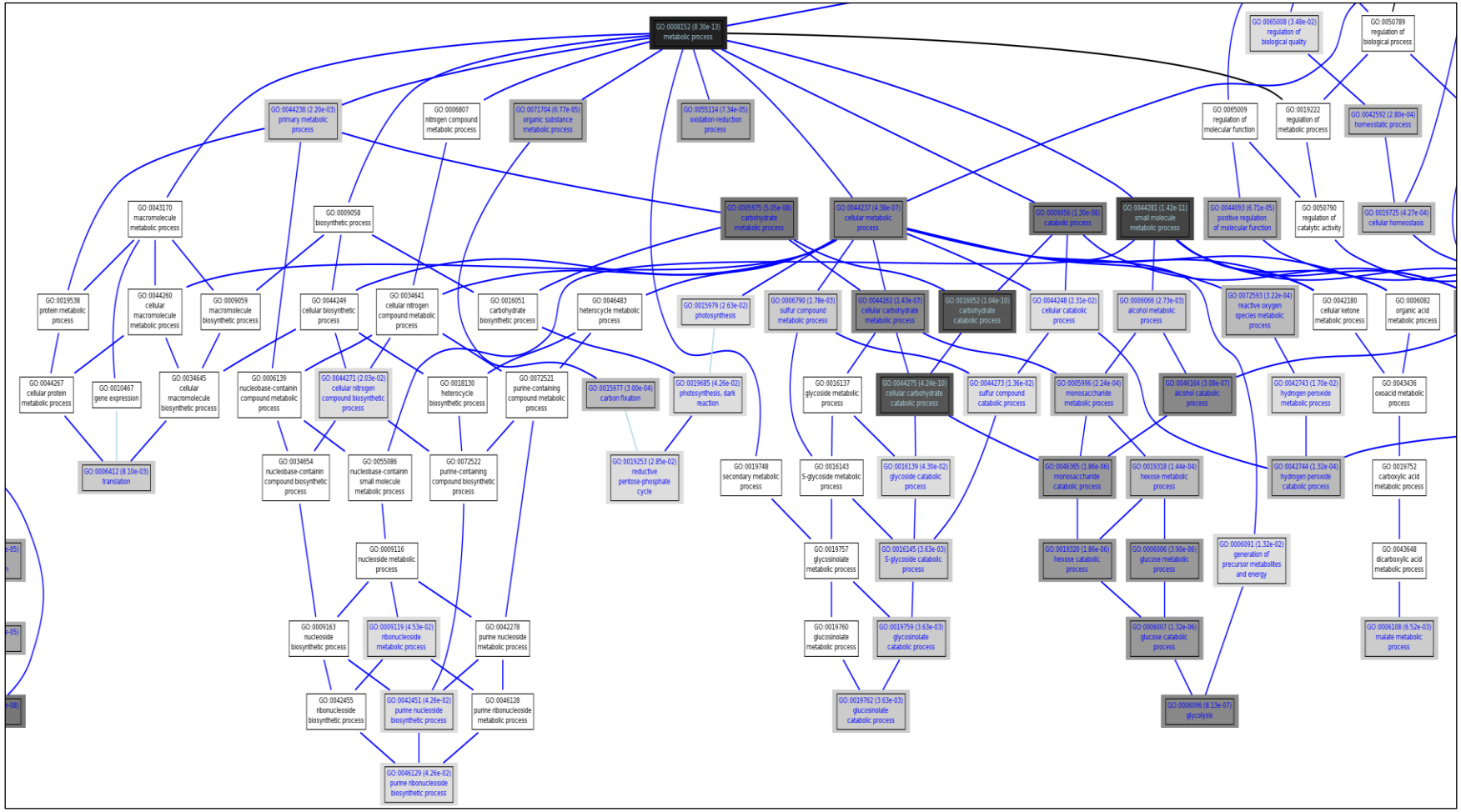
Appendix C-2

GO term enrichment for Biological Processes category.



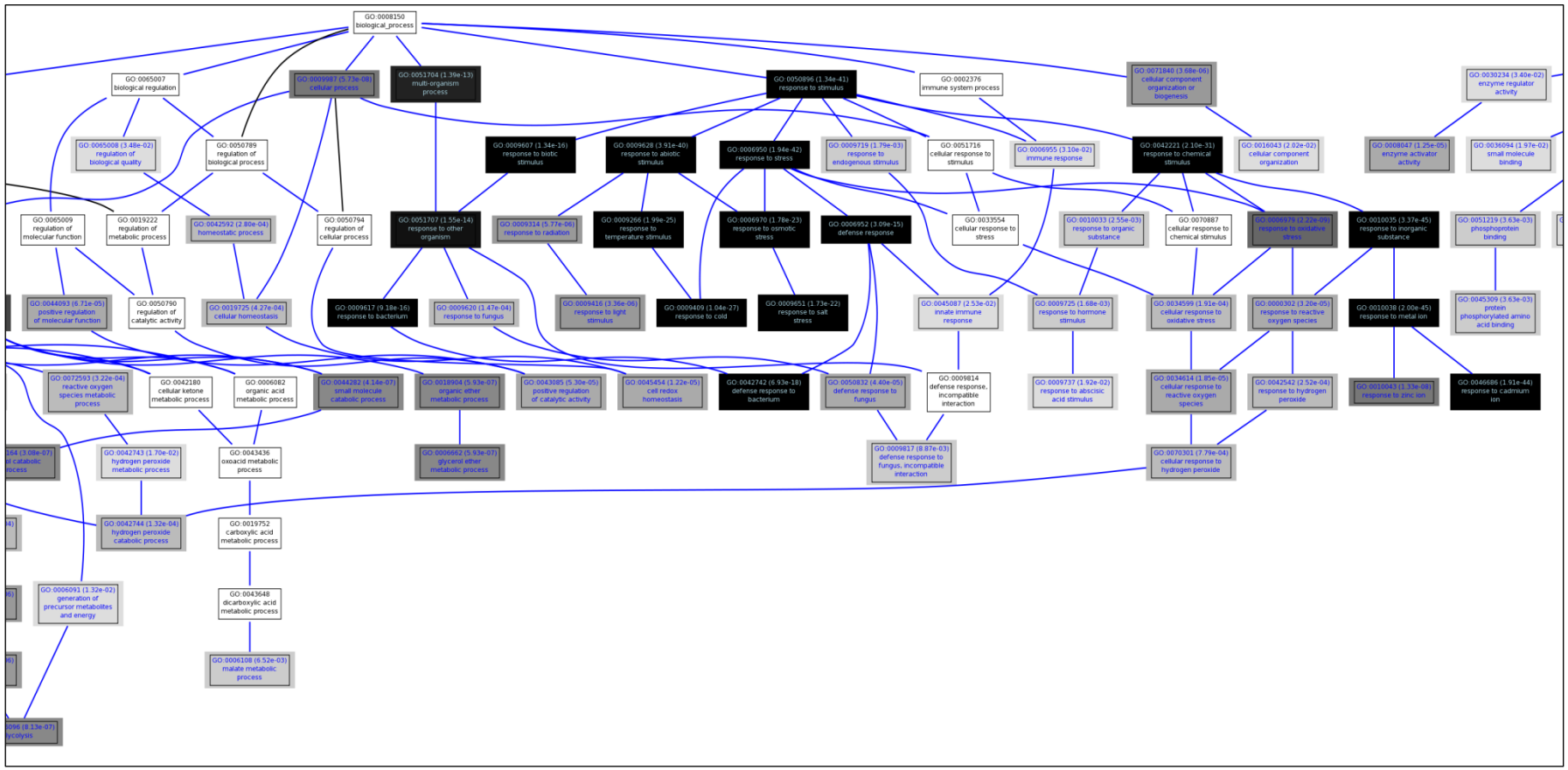
Appendix C-3

GO term enrichment for Biological Processes category which is spread out over two pages with the first focusing on metabolic process



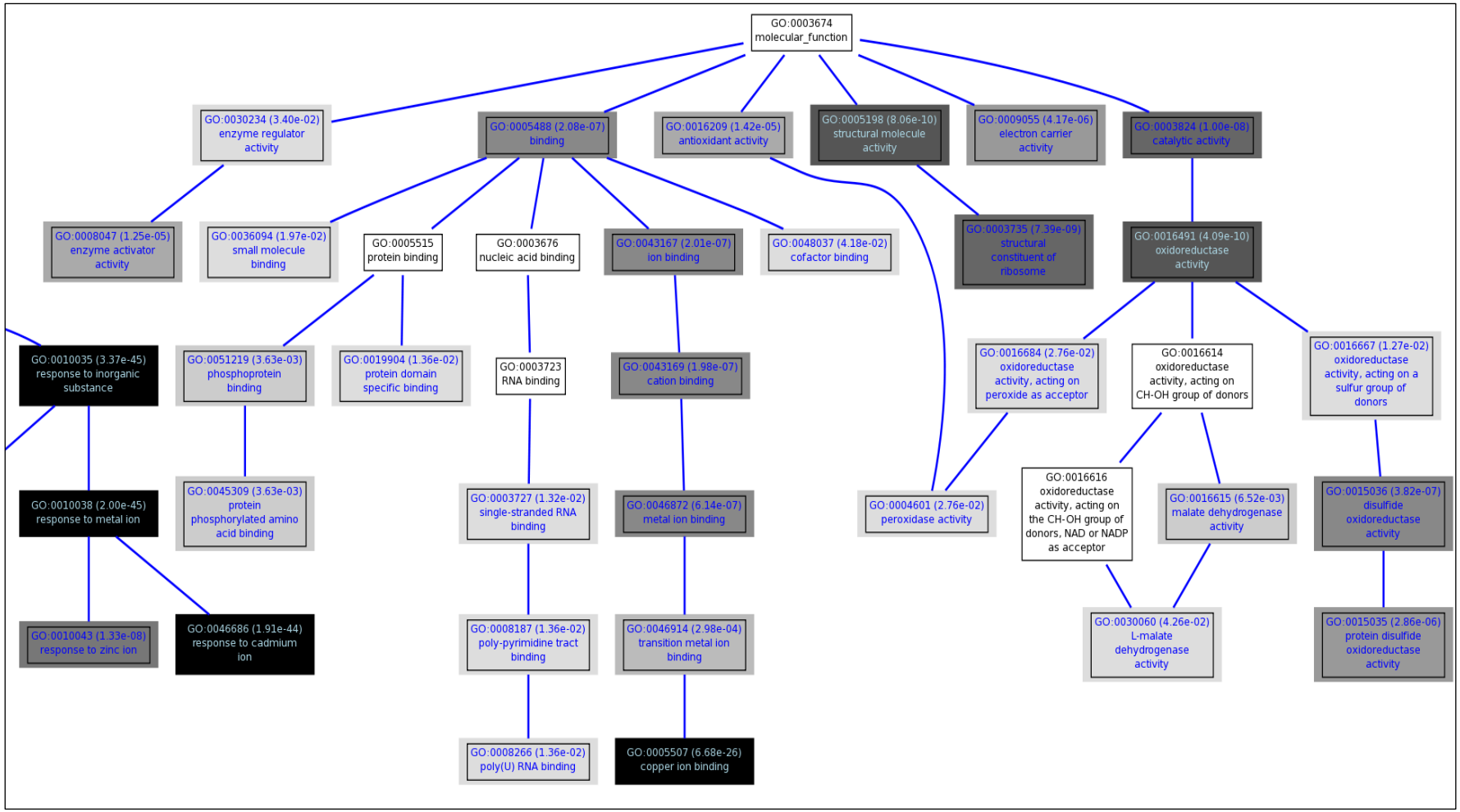
Appendix C-4

GO term enrichment for Biological Processes category which is spread out over two pages which covers cellular processes and was linked to stress-related and defences gene (Appendix C-3).



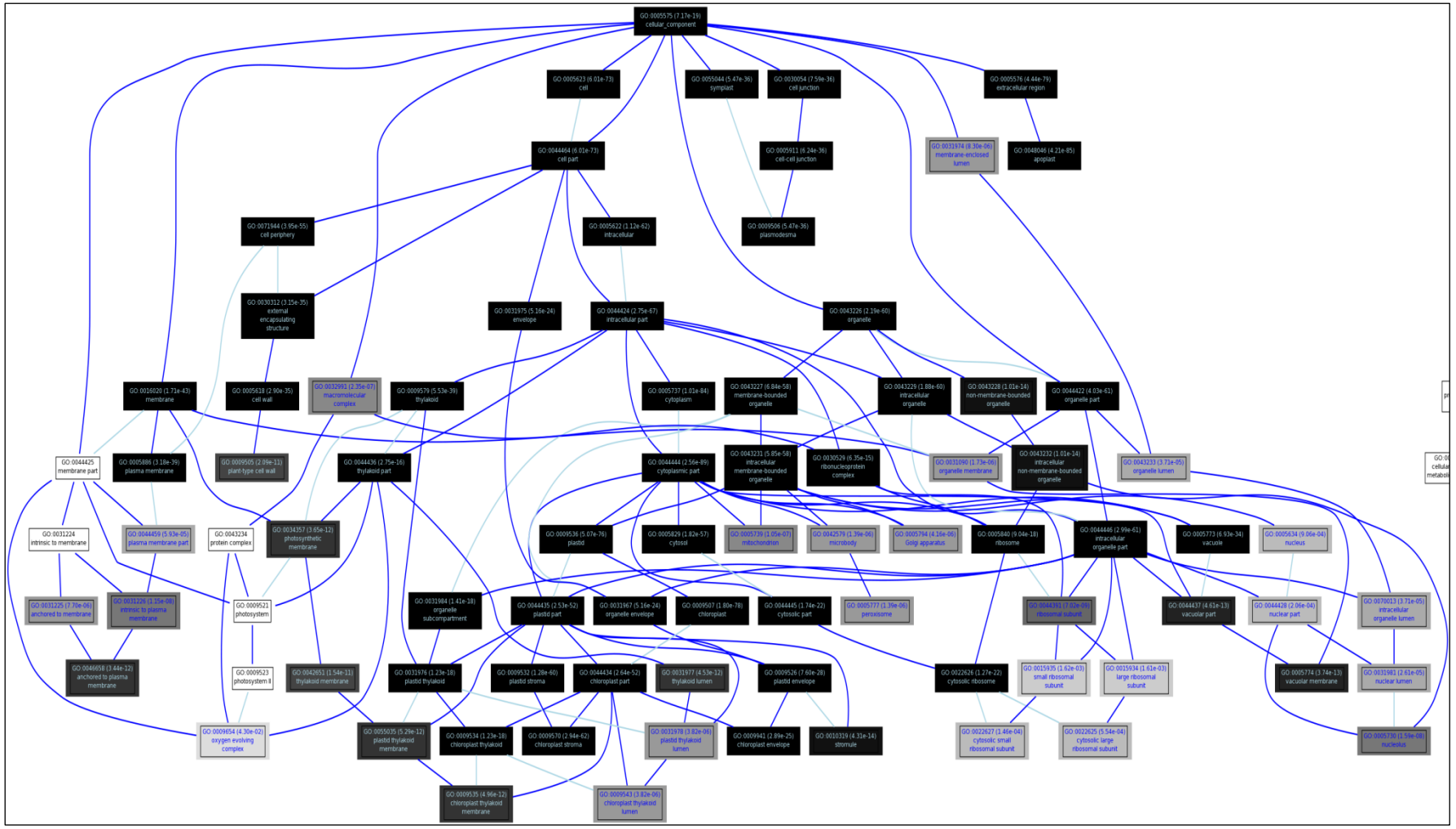
Appendix C-5

GO term enrichment for Molecular Function category.



Appendix C-6

GO term enrichment for Cellular Component Category.



Appendix D

Complete list of *Arabidopsis* phloem exudate protein identified from the current study, Batailler et al and Guelette et al.

Proteins identified in the current study are written in black

Proteins identified by Batailler et al are written in blue

Proteins identified by Guelette et al are written in red

Represents proteins identified by all three studies

Represents proteins identified by Batailler et al and current study

Represents proteins identified by Guelette et al and current study

Represents proteins identified by Batailler et al and Guelette et al

	AGI	Protein	Mass (Da)	No. of sequence (ESI-MS)
	Redox Reaction			
1	At1g07890	Ascorbate peroxidase 1 (APX16)	27413	4
2	At4g35090	Catalase 2 (CAT)	57308	1
3	At1g20620	Catalase 3 (CAT3)	55932	4
4	At1g08830	Copper/zinc superoxide dismutase (CSD1)	15259	5
5	At2g28190	Copper/zinc superoxide dismutase (CSD2)	22322	2
6	At1g19570	Dehydroascorbate reductase (DHAR1)	23626	5
7	At4g11600	Glutathione peroxidase 6 (GPX6)	18597	2
8	At2g24940	Membrane-associated progesterone binding protein 2 (MAPR2)	11002	4
9	At5g40370	Membrane-associated progesterone binding protein 2 (MAPR2)	11752	7
10	At3g52880	Monodehydroascorbate reductase 1 (MDAR1)	46500	7
11	At1g63940	Monodehydroascorbate reductase 6 (MDAR6)	53299	3
12	At5g03630	Pyridine nucleotide-disulphide oxidoreductase family protein (MDAR2)	46433	2
13	At1g45145	Thioredoxin 3 (TRX3)	13227	3
14	At3g02730	Thioredoxin F-type 1 (TRXF1)	19313	2
15	At5g16400	Thioredoxin F2 (TRXF2)	19966	1
16	At5g42980	Thioredoxin H-type 3 (TRX-H-3)	13112	13
17	At1g03680	Thioredoxin M-type 1 (TRX-M1)	19652	3
18	At3g15360	Thioredoxin M-type 4 (TRX-M4)	21403	2
19	At1g43560	Thioredoxin Y2 (Aty2_ty2)	18581	1
20	At2g20270	Thioredoxin superfamily protein	19125	2
21	At3g11630	Thioredoxin superfamily protein	29074	2

22	At3g52960	Thioredoxin superfamily protein	24697	1
23	At4g03520	Thioredoxin superfamily protein (ATHM2)	20314	5
24	At3g10920	MEE33_MSD1 (manganese superoxide dismutase 1); superoxide dismutase	25300	2
25	At5g06290	2CPB; 2-cys peroxiredoxin B	29700	6
26	At4g20840	FAD-binding domain-containing protein	60000	3
27	At4g20860	FAD-binding domain-containing protein	60300	2
28	At5g54500	FQR1; flavodoxin-like quinine reductase 1	21700	4
29	At4g19880	GST; glutathione S-transferase	40500	3
30	At1g78380	GST8; glutathione S-transferase TAU 19	25600	7
31	At1g02920	GST11; glutathione S-transferase 11	23500	11
32	At1g27130	GST12; glutathione S-transferase TAU 13	25000	3
33	At2g02930	GST16: glutathione S-transferase 16	24000	10
34	At2g02390	GST18; glutathione S-transferase ZETA 1	24800	3
35	At2g29420	GST25; glutathione S-transferase TAU 7	26000	3
36	At2g47730	GSTF5; glutathione S-transferase PHI 8	29100	15
37	At1g02930	GSTF6; glutathione S-transferase 6	23400	11
38	At2g30860	GSTF7; glutathione S-transferase PHI 9	24100	21
39	At1g10360	GST29; glutathione S-transferase	25800	2
40	At2g29450	GSTU1; glutathione S-transferase TAU 5	25900	6
41	At1g78370	GSTU20; glutathione S-transferase TAU 20	24900	8
42	At1g78360	GSTU21; glutathione S-transferas TAU 21	76800	14
43	At5g02790	In2-1; glutathione S-transferase	27000	6
44	At5g43940	ADH2; S-(hydroxymethyl)glutathione dehydrogenase/ S-nitrosoglutathione reductase	40600	2
45	At3g24170	GR1; glutathione-disulfide reductase/ oxidoreductase	53700	3
46	At1g75270	DHAR2; dehydroascorbate reductase 2	23300	3
47	At5g05340	peroxidase	34100	6
48	At2g37130	PRXR5: peroxidase 21	36600	5
	Metabolic Pathway			
	A) Glycolysis			
49	At2g24270	Aldehyde dehydrogenase 11A3 (ALDH11A3)	53022	1
50	At2g36530	Enolase (ENO2) / (LOS2)	47719	3
51	At5g42740	Glucose-6-phosphate isomerase, cytosolic (GPI) (phosphoglucose isomerase) (PGI)	61589	1
52	At3g04120	Glyceraldehyde-3-phosphate dehydrogenase C subunit 1 (GAPC1)	36491	3
53	At3g55440	Triosephosphate isomerase (TPI)	27111	8
54	At3g03250	UDP-glucose pyrophosphorylase 1 (UGP1)	51706	9
55	At5g17310	UDP-glucose pyrophosphorylase 2 (UGP2)	51920	6
	B) Gluconeogenesis			
56	At5g09660	Peroxisomal NAD-malate dehydrogenase 2 (PMDH2)	37669	4
	C) TCA			
57	At3g01500	Carbonic anhydrase 1 (CA1)	37450	5
58	At5g14740	Carbonic anhydrase 2 (CA2)	36462	5

59	At1g04410	Lactate/malate dehydrogenase family protein (c-NAD-MDH1)	35548	4
60	At1g53240	Lactate/malate dehydrogenase family protein (mMDH1)	35782	4
61	At5g11670	NADP-malic enzyme 2 (NADP-ME2)	64333	2
62	At5g50850	[8.1.1.1] MAB1 (MACCI-BOU)_PDHE1-B (PDH2); catalytic/ pyruvate dehydrogenase	39100	4
63	At1g48030	[8.1.1.3] mtLPD1; mitochondrial lipoamide dehydrogenase 1	53900	7
64	At3g17240	[8.1.1.3] mtLPD2; lipoamide dehydrogenase 2	53900	7
65	At4g35830	[8.1.3] ACO1; cytoplasmic aconitate hydratase, citrate hydro-lyase, aconitase	98000	11
66	At1g65930	[8.1.4] CICDH; cytosolic NADP+ isocitrate dehydrogenase	45600	8
67	At2g20420	[8.1.6] succinyl-CoA ligase (GDP-forming)	45200	5
68	At5g08300	[8.1.6] succinyl-CoA ligase (GDP-forming)	36000	3
69	At3g15020	[8.1.9] malate dehydrogenase (NAD), mitochondrial	35800	8
70	At5g43330	[8.2.9] NAD-dependent malate dehydrogenase, cytosolic,	35600	7
71	At3g47520	[8.2.9] MDH (malate dehydrogenase); NAD-dependent L-malate dehydrogenase, cytosolic	42300	5
	Amino Acid Metabolism			
72	At4g14880	O-acetylserine (thiol) lyase (OAS-TL) isoform A1 (OASA1)	33805	3
73	At4g13940	S-adenosyl-L-homocysteine hydrolase (SAHH1)	39994	14
74	At3g23810	S-adenosyl-l-homocysteine (SAH) hydrolase 2 (SAHH2)	53159	3
75	At1g02500	S-adenosylmethionine synthetase 1 (SAM1)	41453	1
76	At4g31990	Aspartate aminotransferase 5 (ASP5)	49328	1
77	At5g17920	Cobalamin-independent synthase family protein (ATCIMS)	84267	11
78	At5g04140	[12.2.1.1] GLU1 1; FD-GOGAT_ferredoxin-dependent glutamate synthase 1 [NADH]	176500	23
79	At5g07440	[12.3.1] GDH2: glutamate dehydrogenase 2 [NAD(P)+]	44600	3
80	At2g30970	[13.1.1.2.1] ASP1; aspartate aminotransferase 1 , L-aspartate:2-oxoglutarate aminotransferase	47600	8
81	At5g19550	[13.1.1.2.1] ASP2 _AAT2; aspartate aminotransferase 2, L-aspartate:2-oxoglutarate aminotransferase	44100	15
82	At5g11520	[13.1.1.2.1] ASP3_YLS4; aspartate aminotransferase 3, L-aspartate:2-oxoglutarate aminotransferase	48800	6
83	At1g23310	[13.1.1.3.1] AOAT1; L-alanine:2-oxoglutarate aminotransferase	53200	5
84	At1g17290	[13.1.1.3.1] AlaAT1_AOAT4; L-alanine:2-oxoglutarate aminotransferase	59700	2
85	At1g80600	[13.1.2.3.4] WIN1; N2-acetyl-L-ornithine:2-oxoglutarate 5-aminotransferase	48700	4
86	At2g19940	[13.1.2.3.1004] N-acetyl-gamma-glutamyl-phosphate reductase, NAD or NADH binding	39500	2
87	At1g75330	[13.1.2.3.21] OTC; ornithine carbamoyltransferase	40900	2
88	At2g43910	[13.1.3.4.13] HOL1; thiopurine methyltransferase, methionine S-methyltransferase	25200	5
89	At1g14810	[13.1.3.6.1.2] semialdehyde dehydrogenase	40600	2
90	At4g35630	[13.1.5.1.2] PSAT; phosphoserine aminotransferase	47200	2
91	At3g61440	[13.1.5.3.1] OASTL_CYSC1; cysteine synthase C1	39800	5
92	At3g04940	[13.1.5.3.1] OASTL_CYSD1; cysteine synthase D1	34200	4
93	At2g43750	[13.1.5.3.1] OASTL_OASB; O-acetylserine (thiol)lyase B, cysteine synthase	41500	8
94	At4g13360	[13.2.6.3], [11.9.4.3] crotonase, enoyl-CoA hydratase, isomerase 3-hydroxyisobutyryl-CoA hydrolase	46100	5

95	At3g17390	[15.2] MAT4_SAMS3; methionine adenosyltransferase , S-adenosylmethionine synthase 3	42700	3
96	At2g36880	[15.2], [13.1.3.4.1003] MAT3; methionine adenosyltransferase 3	42400	4
97	At4g13430	[16.5.1.1.1.3],[8.2.3] IPMI; isopropylmalate dehydratase, large subunit	54900	2
98	At3g19710	[16.5.1.1.1.1], [13.1.4.1.4] BCAT4 (branched-chain aminotransferase 4); methionine-oxo-acid transaminase	38900	4
99	At1g31180	[16.5.1.1.1.5], [13.1.4.4.3] IMD3; isopropylmalate dehydrogenase,	43700	2
	Lipid Metabolism			
100	At1g31812	Acyl-CoA-binding protein 6 (ACBP6)	10466	3
101	At4g25050	Acyl carrier protein 4 (ACP4)	14536	2
102	At5g16390	Chloroplastic acetylcoenzyme A carboxylase 1 (BCCP1 / CAC1A)	29613	2
103	At1g27950	Glycosylphosphatidylinositol-anchored lipid protein transfer 1 (LTPG1)	19777	3
104	At2g38540	Lipid transfer protein 1 (LTP1)	11747	9
105	At2g38530	Lipid transfer protein 2 (LTP2)	11938	3
106	At1g66970	SHV3-like 2 (SVL2)	83788	1
107	At3g48990	[11.1.8] AMP-dependent synthetase and ligase family protein; acyl coA ligase	55400	9
108	At4g04460	Aspartic protease	56000	1
	Nucleotide Metabolism			
109	At3g09820	Adenosine kinase 1 (ADK1)	37812	7
110	At5g03300	Adenosine kinase 2 (ADK2)	31337	7
111	At4g09320	Nucleoside diphosphate kinase family protein (NDPK1)	15740	2
112	At5g63310	Nucleoside diphosphate kinase 2 (NDPK2)	25463	1
113	At5g09650	Pyrophosphorylase 6 (PPa6)	33373	3
114	At3g57610	[23.1.2.20] ADSS; adenylosuccinate synthase	52800	2
115	At1g27450	[23.3.1.1] APT1; adenine phosphoribosyltransferase 1	19600	7
116	At4g29680	[23.4.99] type I phosphodiesterase/nucleotide pyrophosphatase	54500	2
	Secondary Metabolism			
117	At3g24503	Aldehyde dehydrogenase 2C4 (ALDH1A / ALDH2C4)	54460	1
118	At1g54040	Epithiospecifier protein (ESP)	37069	3
119	At3g44310	Nitrilase 1 (NIT1)	38152	1
120	At5g26000*	Thioglucoside glucohydrolase 1 (TGG1) *	61133	10
121	At5g16440	[16.1.2.7] IPP1 (isopentenyl diphosphate isomerase 1); isopentenyl-diphosphate delta-isomerase	33100	5
122	At1g74920	[16.4.2.1] ALDH10A8; aldehyde dehydrogenase 10A8, betaine-aldehyde dehydrogenase	54300	2
123	At3g23490	[16.4.3.1] CYN (cyanase); DNA binding / cyanate hydratase, hydro-lyase	18500	2
124	At2g20610	[16.5.1.1.4.2], [13.2.6.2] ALF1_SUR1 (Superroot 1); S-alkylthiohydroximate lyase, tyrosine transaminase	51000	3
125	At1g75280	[16.8.5] isoflavone reductase	33600	4
126	At4g30530	[23.1.99] GGP1; gamma-glutamyl peptidase	28300	7

	Metabolism – Miscellaneous			
127	At5g58330	[8.2.9] putative malate dehydrogenase	48200	4
128	At4g13930	[25.1] SHM4 (serine hydroxymethyltransferase 4); glycine hydroxymethyltransferase	51600	5
129	At3g51000	[26.1] epoxide hydrolase,	36600	3
130	At5g19440	[26.11.1] CAD; cinnamyl-alcohol dehydrogenase,	35500	3
131	At5g34850	[26.13] PAP26 (Purple acid phosphatase 26); acid protein serine/threonine phosphatase	54900	4
132	At2g03980	[26.28] GDSL-motif lipase/hydrolase family protein	40600	2
133	At5g12040	[35.1.1023] carbon-nitrogen hydrolase	40200	2
	A) C1 Metabolism			
134	At2g41530	S-formylglutathione hydrolase (SFGH)	30474	5
135	At3g59970	Methylenetetrahydrofolate reductase 1 (MTHFR1)	66246	2
	B) Hormone Metabolism			
136	At3g16430	Jacalin-related lectin 31 (JAL31)	32133	2
137	At3g16470	Mannose-binding lectin superfamily protein (JAL35_JR1)	48698	13
138	At1g24100	[16.5.1.1.4.3], [17.2.1.1001] UGT74B1 (UDP-glucosyl transferase 74B1); thiohydroximate beta-D-glucosyltransferase	50900	2
139	At3g44300	[16.5.1.3.3], [17.2.1.1001] NIT2 (nitrilase 2); indole-3-acetonitrile nitrilase	37000	16
140	At2g23600	[17] MES2_ACL (acetone-cyanohydrin lyase); methyl indole-3-acetate esterase	29600	3
141	At1g08980	[26.8] AMI1_ATAMI1 (amidase-like protein 1); indoleacetamide hydrolase, amidase nitrile lyase	44900	3
142	At2g23610	[26.8] MES3 (methylesterase 3); methyl indole-3-acetate methylesterase	29700	2
	C) Carbohydrate Metabolism			
143	At1g12240	Glycosyl hydrolases family 32 protein / Beta-fructosidase (ATBETAFRUCT4)	73465	3
144	At5g37850	pfkB-like carbohydrate kinase family protein (SOS4)	34232	1
145	At1g23740	Oxidoreductase, zinc-binding dehydrogenase family protein	40986	4
146	At5g61130	Plasmodesmata callose-binding protein 1 (PDCB1)	20365	2
147	At1g06020	[2.2.1.1] pfkB-type carbohydrate kinase family protein; fructokinase	37500	4
148	At4g25000	[2.2.2.1] AMY1; alpha-amylase	47300	5
149	At4g15210	[2.2.2.1] BMY1; beta-amylase 5	55900	8
150	At2g37760	[3.5] aldo/keto reductase family protein	34600	3
151	At3g17940	[3.5] aldose 1-epimerase family protein	37100	3
152	At3g47800	[3.5] aldose 1-epimerase family protein	39600	4
153	At1g08200	[10.1.3] AXS2 (UDP-D-apiose/UDP-D-xylose synthase 2); UDP-glucuronate decarboxylase	43700	3
154	At5g49650	[3.7] xylulose kinase	61200	6
	D) Nitrogen Metabolism			
155	At5g35630	Glutamine synthetase 2(GS2)	47256	6
	E) Tetrapyrrole Synthesis			

156	At5g08280	Hydroxymethylbilane synthase (HEMC)	39724	4
157	At5g63570	Glutamate-1-semialdehyde-2,1-aminomutase (GSA1)	50222	2
	F) Vitamin Metabolism			
158	At4g26500	Chloroplast sulfur E (SUFE1)	40888	1
159	At5g54770	Thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1)	17637	1
160	At3g14990	[18.2] 4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein (thiamin biosynthesis)	41700	6
161	At2g20690	[18.3.2] riboflavin synthase	29500	5
162	At3g12290	[25.5] tetrahydrofolate dehydrogenase/cyclohydrolase	31500	3
163	At5g57850	[26.26.1] aminodeoxychorismate lyase	40900	2
164	At1g78680	[29.5] GGH2 (gamma-glutamyl hydrolase 2); folylpolyglutamate gamma-glutamyl hydrolase	38500	3
165	At3g48420	[35.1.1023] riboflavin kinase (hydrolase)	34100	6
	Oxidative Pentose Phosphate Pathway			
166	At3g04790	[7.2.4],[1.3.10] ribose 5-phosphate isomerase-related	29200	2
167	At3g02360	[7.1.3] 6-phosphogluconate dehydrogenase	53400	3
168	At5g41670	[7.1.3] 6-phosphogluconate dehydrogenase	53200	2
169	At1g64190	[7.1.3] 6-phosphogluconate dehydrogenase	53200	4
170	At5g13420	[7.2.2] transaldolase	47600	6
	Photosynthesis			
	A) Calvin Cycle			
171	At2g21330	Fructose-bisphosphate aldolase 1 (FBA1)	33302	4
172	At4g38970	Fructose-bisphosphate aldolase 2 (FBA2)	42935	1
173	At1g42970	Glyceraldehyde-3-phosphate dehydrogenase B subunit (GAPB)	47532	1
174	At3g12780	Phosphoglycerate kinase 1 (PGK1)	51026	2
175	At1g56190	Phosphoglycerate kinase family protein	49894	7
176	At2g21170	Plastid Isoform Triose Phosphate Isomerase (PDTPI)	33505	3
177	At2g39730	Rubisco activase (RCA)	51999	30
178	AtCg00490	Ribulose-bisphosphate carboxylases (RBCL)	52669	24
179	At1g67090	Ribulose bisphosphate carboxylase small chain 1A (RBCS1A)	20003	13
180	At3g55800	Sedoheptulose-bisphosphatase (SBPASE)	42414	3
181	At3g60750	Transketolase	79875	2
182	At1g32060	[1.3.12] PRK (phosphoribulokinase); ATP binding / phosphoribulokinase	44300	9
183	At3g26650	[1.3.4] GAPA-1; glyceraldehyde-3-phosphate dehydrogenase A subunit	42400	21
184	At1g12900	[1.3.4] GAPA-2; glyceraldehyde-3-phosphate dehydrogenase A subunit 2	42700	19
185	At1g16300	[1.3.4] GAPCP-2; NAD or NADH binding, glyceraldehyde-3-phosphate dehydrogenase	44700	4
186	At1g79550	[4.1.12] PGK; 3-phosphoglycerate kinase	42000	23
187	At1g13440	[4.1.9] GAPC-2; NAD or NADH binding, glyceraldehyde-3-phosphate dehydrogenase	36800	37
188	At4g24620	[4.2.4] PGI1; phosphoglucose isomerase	62200	4
189	NP_051067	RubisCo LSU	53000	11
190	At2g47400	Putative CP12 protein	13000	2

	B) Light Reaction			
191	At1g20340	Cupredoxin superfamily protein (DRT112_PETE2)	16988	6
192	At5g66190	Ferredoxin-NADP(+)-oxidoreductase 1 (FNR1)	40332	4
193	At1g20020	Ferredoxin-NADP(+)-oxidoreductase 2 (FNR2)	41142	3
194	At1g77090	Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein	28501	1
195	At1g06680	Photosystem II subunit P-1 (PSBP-1)	28003	13
196	At4g21280	Photosystem II subunit QA (PSBQ-1)	23888	4
197	At4g05180	Photosystem II subunit Q-2 (PSBQ-2)	24611	10
198	At1g76100	Plastocyanin 1 (PETE1)	17589	1
199	At3g50820	PSBO-2 (photosystem II subunit O-2); photosystem II subunit, poly(U) binding	34900	2
200	AtMg00280	ORF110A; ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	12600	7
201	At5g38430	RBCS-1B: chloroplast ribulose bisphosphate carboxylase small chain 1B	20200	16
202	At1g29910	CAB3; chlorophyll a-b binding protein	28100	6
203	At4g10340	CP26; chlorophyll a-b binding protein	30100	4
204	At5g66570	Oxygen-evolving enhancer protein	35000	5
205	At2g34420	Chlorophyll a/b binding protein	28000	5
	C) Photorespiration			
206	At1g11860	Glycine cleavage T-protein family	44126	2
207	At1g32470	Single hybrid motif superfamily protein	17886	2
208	At3g14420	Aldolase-type TIM barrel family protein	40226	1
209	At5g36700	[1.2.1] PGLP1 (2-phosphoglycolate phosphatase 1); phosphoglycolate phosphatase	39600	8
210	At2g13360	[1.2.3], [13.1.5.2.3] AGT1; alanine:glyoxylate aminotransferase	44100	9
211	At4g33010	[1.2.4.1], [25] GLDP1; glycine decarboxylase P-protein 1	112700	3
212	At1g68010	[1.2.6], [13.2.5.2] HPR1; hydroxypyruvate reductase, glycerate dehydrogenase	42100	10
	Stress			
	A) Abiotic Stress			
213	At1g20440	Cold-regulated 47 (COR477)	33349	6
214	At1g20450	Dehydrin family protein (ERD10)	29425	5
215	At1g76180	Dehydrin family protein (ERD14)	20770	6
216	At1g70830	MLP-like protein 28 (MLP28)	27528	3
217	At1g70890	MLP-like protein 43 (MLP43)	17695	5
218	At1g23130	Polyketide cyclase/dehydrase and lipid transport superfamily protein (Bet v I allergen-like)	17848	3
219	At4g23670	Polyketide cyclase/dehydrase and lipid transport superfamily protein	17529	2
	B) Biotic Stress			
220	At5g39730	AIG2-like (avirulence induced gene) family protein	20005	2
221	At3g04720	Pathogenesis-related 4 (PR4) / Hevein-like protein (HEL)	22937	4
222	At2g43530	Scorpion toxin-like knottin superfamily protein	9858	2
	Cell Organization and Biogenesis			

223	At2g43530	Actin 7 (ACT7)	41774	5
224	At5g59880	Actin-depolymerizing factor 3 (ADF3)	16080	3
225	At1g35720*	Annexin 1 (ANNAT1) *	35757	10
226	At2g19760	Profilin 1 (PRO1)	14240	2
227	At4g29350	Profilin 2 (PRO2)	13990	3
228	At5g56600	Profilin 3 (PRO3)	14333	1
229	At3g63190	Ribosome recycling factor, chloroplast precursor (RRF)	30403	2
230	At4g38740	Rotamase CYP 1 (ROC1)	18333	7
231	At2g16600	Rotamase CYP 3 (ROC3)	18444	2
232	At3g62030*	Rotamase CYP 4 (ROC4) *	28190	7
233	At4g34870	Rotamase CYP 5 (ROC5)	18366	6
234	At5g15970	Stress-responsive protein (KIN2) / stress-induced protein (KIN2) / cold-responsive protein (COR6.6) / cold-regulated protein (COR6.6) (AtCor6.6_COR6.6_KIN2)	6551	6
235	At5g62690	Tubulin beta-2/beta-3 chain (TUB2)	50003	1
236	At4g14960	Tubulin/FtsZ family protein (TUA6)	49226	1
237	At3g12110	ACT11 (actin-11); structural constituent of cytoskeleton	41600	3
238	At2g21130	CYP2 (cyclophilin)	18000	3
	Signalling			
239	At2g27030	Calmodulin 5 (CAM5)	16652	1
240	At3g43810	Calmodulin-7 (CAM7)	16829	1
241	At5g38480	General regulatory factor 3 (GRF3_RCI1)	28600	1
242	At4g09000	General regulatory factor 1 (GF14 CHI_GRF1)	29666	2
243	At3g02520	General regulatory factor 7 (GF14 NU_GRF7)	29811	4
244	At1g22300	General regulatory factor 10 (14-3-3EPSILON_GF14 EPSILON_GRF10)	28599	3
245	At4g20260	Plasma-membrane associated cation-binding protein 1 (PCAP1)	24697	3
246	At3g59920	RAB GDP dissociation inhibitor 2 (GDI2)	46555	2
247	At5g65020	ANNAT2; annexin 2	36200	2
248	At1g56340	CRT1A; calreticulin-1	48400	2
249	At1g09210	CRT1B; calreticulin-2	48000	2
250	At1g65480	FT (flowering locus T)	19700	3
251	At1g78300	GRF2 (G-box binding factor 2); 14-3-3-like protein GF14 omega	29100	9
252	At1g35160	GRF4 (G-box binding factor 4); 14-3-3-like protein GF14 PHI chain	30100	9
253	At5g10450	GRF6 (G-box binding factor 6); 14-3-3 lambda_AFT1	27900	10
254	At5g65430	GRF8 (G-box binding factor 8); 14-3-3-like protein general regulatory factor GF14 kappa	27900	7
255	At2g42590	GRF9 (G-box binding factor 9); 14-3-3-like protein GF14 mu	29400	8
256	At1g33610	LRR leucine-rich repeat protein	99500	7
257	At1g18080	RACK1A (receptor for activated C kinase 1A); versatile scaffold protein	35600	12
258	At2g26740	SEH (soluble epoxide hydrolase); epoxide hydrolase	36300	6
259	At2g10450	14-3-3-like protein GF14 psi	9000	1
	Defense			

260	At3g28940	AIG2-like; avirulence-responsive protein	19400	4
261	At4g37000	RCCR_ACD2 (accelerated cell death 2); red chlorophyll catabolite reductase	36300	2
262	At4g23170	CRK9; cysteine-rich receptor-like kinase (RLK) 9	29600	3
263	At4g35220	cyclase	29900	3
264	At4g14040	EDA38_SBP2 (selenium binding protein 2); selenium binding protein	53800	4
265	At3g14210	ESM1; epithiospecifier modifier 1	43900	12
266	At3g12500	PR-3; basic chitinase (CHI-B)	34500	4
267	At2g43570	CHI; chitinase, glycoside hydrolase, family 19	29700	6
268	At2g43590	chitinase; glycoside hydrolase, family 19	28300	4
269	At2g43620	chitinase; glycoside hydrolase, family 19	30300	6
	Protein Biosynthesis, Organization, Folding, Transport and Modification			
270	At4g04910	AAA-type ATPase family protein (NSF)	81652	1
271	At5g20720	Chaperonin 20 (CPN20)	9021	4
272	At5g03940	Chloroplast signal recognition particle 54 kDa subunit (CPSRP54P)	60231	1
273	At5g50920	CLPC homologue 1 (CLPC1)	7883	1
274	At1g72730	DEA(D/H)-box RNA helicase family protein	46777	1
275	At5g45680	FK506-binding protein 13 (FKBP13)	22039	1
276	At5g02500	Heat shock cognate protein 70-1 (HSC70-11)	71358	2
277	At5g60390	GTP binding Elongation factor Tu family protein	49402	2
278	At5g02240	NAD(P)-binding Rossmann-fold superfamily protein	27103	1
279	At3g12390	Nascent polypeptide-associated complex (NAC), alpha subunit family protein	21982	1
280	At4g20360	RAB GTPase homolog E1B (ATRABE1B)	51398	2
281	At2g47170	Ras-related small GTP-binding family protein (ARF1A1C)	21646	1
282	At5g58290	Regulatory particle triple-A ATPase 3 (RPT3)	45711	1
283	At1g56070	Ribosomal protein S5/Elongation factor G/III/V family protein (LOS1)	92356	2
284	At1g31340	Related to ubiquitin 1 (RUB1)	8525	1
285	At5g03240	Polyubiquitin (UBQ3)	28169	2
286	At1g27970	NTF2B (nuclear transport factor 2B); Ran GTPase binding / protein transporter	13900	2
287	At5g28540	BIP1; luminal binding protein BiP, member of the HSP70 family	73500	2
288	At3g01480	CYP38 (cyclophilin 38); peptidyl-prolyl cis-trans isomerase	47900	2
289	At3g55260	HEXO1; Beta-N-acetylhexosaminidase-like protein	63100	6
290	At4g38830	Cysteine-rich receptor kinase	74000	2
291	At3g17840	Putative receptor kinase (rice)	72000	4
292	At5g58350	Tetratricopeptide repeat protein	-	1
293	At4g01270	RING zinc-finger protein	56000	2
294	At2g16750	Putative kinase	53000	3
	Protein Turn-over			
295	At4g33090	APM1; aminopeptidase M1	98000	2
296	At2g24200	aminopeptidase	54400	13
297	At1g63770	aminopeptidase M1	103300	5

298	At5g10760	aspartyl protease	49400	10
299	At1g06260	cysteine proteinase	37700	2
300	At3g14067	subtilase: serine endopeptidase	81700	8
301	At3g45010	scpl48; serine carboxypeptidase	56800	4
302	At4g16500	cysteine protease inhibitor (cystatin)	12500	6
303	At3g22110	PAC1 (20S proteasome alpha subunit C1); threonine-type endopeptidase	27400	2
304	At1g53850	PAE1 (20S proteasome alpha subunit E1); threonine-type endopeptidase	25900	4
305	At3g60820	PBF1 (20S proteasome beta subunit PBF1); threonine-type endopeptidase	24500	2
306	At1g16470	PAB1 (20S proteasome subunit PAB1); endopeptidase/ threonine-type endopeptidase	25600	2
307	At3g51260	PAD1 (20S proteasome alpha subunit PAD1); endopeptidase/ threonine-type endopeptidase	27200	2
308	At1g56450	PBG1 (20S proteasome beta subunit); Proteasome subunit beta type-4	27600	3
309	At1g13060	PBE1 (20S proteasome beta subunit); Proteasome subunit beta type-5-A	29600	2
310	At3g18060	CUL4 RING ubiquitin ligase complex; transducin	65900	7
311	At3g49600	Ubiquitin-specific protease 26	120000	4
	Developmental Process			
312	At3g16640	Translationally controlled tumor family protein (TCTP)	18872	2
	Transport			
313	At2g01320	ABC-2 type transporter family protein （ABCG7）	78599	1
314	At1g72150	PATELLIN 1 (PATL1)	64046	1
	Electron Transport			
315	At1g22840	CYTOCHROME C-1 (CYTC-1)	12390	3
316	At5g08670	ATP synthase subunit beta-1	59500	6
317	AtMg01190	ATP synthase subunit alpha	54900	3
318	At1g76030	V-type proton ATPase subunit B1	54000	5
	Methylglyoxal Degradation			
319	At1g53580	[24.1] GLX2-3_GLY3 ((glyoxylase II 3); hydroxyacylglutathione hydrolase	32200	3
320	At1g08110	[24.2] lactoylglutathione lyase, glyoxalase I	20800	2
321	At2g32090	[24.2] lactoylglutathione lyase, glyoxalase I	14900	2
	Biodegradation of Xenobiotics			
322	At1g67280	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein	39174	1
323	At1g11840	Glyoxalase I homolog (GLX1)	31669	5
324	At1g76680	[17.7.1.5] OPR1; 12-oxophytodienoate reductase (degradation TNT)	41100	3
325	At2g32520	[26.1] dienelactone hydrolase	25800	3
326	At3g23570	[26.1] dienelactone hydrolase	26400	2
327	At3g23600	[26.1] dienelactone hydrolase (degradation of chlorochatechol)	25800	10

	DNA			
328	At5g38470	Rad23 UV excision repair protein family (RAD23D)	40122	1
329	At5g63120	Ethylene-responsive RNA helicase	29000	2
330	At2g21530	Protein with FHA domain UPP6	26000	2
331	At5g43470	Resistance protein homolog RPP-WsB RPP8	141000	3
	RNA			
332	At3g20390	Endoribonuclease L-PSP family protein	15151	3
333	At2g21660	Glycine-rich RNA-binding protein (GRP7)	16596	4
334	At4g39260	Glycine-rich RNA-binding protein 8 (GRP8)	16304	2
335	At4g34110	Polyadenylate-binding protein 2 (PABP2)	67017	1
336	At1g49760	Poly(A) binding protein 8 (PABP8)	72733	3
337	At2g23380	SET domain-containing protein (CLF / SET1)	84155	1
338	At3g61260	Remorin family protein	23130	1
339	At4g24770	31-kDa RNA binding protein (RBP31)	35787	1
340	At2g37220	29 kDa ribonucleoprotein, RNA-binding protein cp29	30600	3
341	At1g09340	CSP41B_CRB (chloroplast RNA binding)	42500	11
342	At5g16450	dimethylmenaquinone methyltransferase family protein, ribonuclease inhibitor activity	17700	2
343	At3g63140	CSP41A (chloroplast stem-loop binding protein of 41 KDA); mRNA binding protein	43800	3
344	At5g07530	Glycine-rich protein (oleosin) GRP17	53000	3
	Storage Protein			
345	At1g07750	cupin	38200	4
346	At2g28680	cupin	38400	3
347	At3g05930	GLP8 (germin-like protein 8); manganese ion binding / nutrient reservoir	22900	3
348	At5g24780	VSP1 (vegetative storage protein 1); acid phosphatase and vegetative storage proteins	30200	3
	Miscellaneous			
349	At1g29660	GDSL-like Lipase/Acylhydrolase superfamily protein	40630	4
350	At4g02520	Glutathione S-transferase PHI 2 (GSTF2)	24109	3
351	At2g30870	Glutathione S-transferase PHI 10 (GSTF10)	24084	5
352	At3g16400	Nitrile specifier protein 1 (NSP1)	51670	8
353	At3g16410	Nitrile specifier protein 4 (NSP4)	67555	2
354	At3g10720	Plant invertase/pectin methylesterase inhibitor superfamily	60013	6
355	At1g78850	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain	49050	7
356	At4g20830	FAD-binding Berberine family protein	63336	2
357	At1g11930	alanine racemase; pyridoxal phosphate-dependent enzyme	27900	2
358	At2g44100	GDI1: guanosine nucleotide diphophate dissociation inhitor 1; RAB GDP-dissociation inhibitor	49700	3
359	At1g12570	glucose-methanol-choline (GMC) oxidoreductase family protein	61900	2
360	At4g14030	SBP1 (selenium-binding protein 1); selenium binding, contains Cytochrome cd1-nitrite reductase-like domain	53900	4
361	At1g53070	legume lectin family protein	30300	3

362	At4g20900	MS5	51000	2
363	At3g26680	SNM1	50000	2
364	At2g39435	Putative PIG-P	50000	Ed
365	At4g33030	Sqd1	44000	4
366	At3g01310	Putative acid phosphatase / Phosphoglycerate mutase-like	119000	3
	Unknown			
367	At3g05900	Neurofilament protein-related	72087	1
368	At1g09310	Protein of unknown function, DUF538	19934	3
369	At1g55480	Protein containing PDZ domain, a K-box domain, and a TPR (ZKT) region	37387	1
370	At5g01750	Protein of unknown function (DUF567)	24295	2
371	At5g22580	Stress responsive A/B Barrel Domain	12349	1
372	At3g02910	AIG2-like (avirulence induced gene) family protein	21333	1
373	At5g19240	Glycoprotein membrane precursor GPI-anchored	21301	1
374	At5g56170	LORELEI-LIKE-GPI-ANCHORED PROTEIN 1 (LLG1)	18904	1
375	At2g37660	NAD(P)-binding Rossmann-fold superfamily protein	34256	1
376	At2g44920	Tetratricopeptide repeat (TPR)-like superfamily protein	23744	1
377	At5g53490	Tetratricopeptide repeat (TPR)-like superfamily protein	25628	1
378	At2g40600	appr-1-p processing enzyme family protein	27400	2
379	At1g70810	C2 calcium/lipid-binding region, CaLB	18600	2
380	At1g44790	ChaC protein [cation transport (chaC)]	22100	4
381	At1g53280	DJ-1 putative peptidase	46900	4
382	At3g54600	DJ-1 putative peptidase	43000	2
383	At5g48540	DUF26 protein	28900	3
384	At1g19130	DUF985 protein	21200	2
385	At5g14780	FDH; formate deshydrogenase	42300	7
386	At3g03990	hydrolase	29500	2
387	At3g24420	hydrolase	29900	2
388	At4g39730	lipase/lipooxygenase	20000	4
389	At2g31670	stress responsive alpha-beta barrel	28800	6
390	At1g16080	unknown protein	34000	7
391	At1g29050	unknown protein	42600	3
392	At2g35810	unknown protein	21900	2
393	At4g21580	oxidoreductase, zinc-binding dehydrogenase	34400	5
394	At5g48850	Hypothetical protein UPP2	63000	2
395	NP_001005208	HAS-like protein UPP1	69000	2
396	At1g33020	Hypothetical protein UPP3	61000	2
397	At4g16820	Lipase class 3 family protein	58000	2

Appendix E

Result of one way ANOVA performed based on leaf number differences in transgenic lines grown in LD (left) and SD (right). Bonferroni's multiple comparison test was carried out where WT and vector control plants in WT were paired with lines in WT background and compared, and ft-10 and vector control plant in ft-10 were paired with lines in ft-10 background and compared. *p<0.05, **p<0.01, ***p<0.001. ns=not significant

LD=long day, SD=short day, ANOVA=analysis of variance, WT=wild type, *ft-10=ft* null mutant, VW=vector:Dendra2 in WT, Vf= vector:Dendra2 in *ft* mutant, SFW= SUC2::FT:Dendra2 in WT, SFf= SUC2::FT:Dendra2 promoter in *ft* mutant, SW=SUC2:Dendra2 in WT, GFW= GAS1::FT:Dendra2 in WT, GFf= GAS1::FT:Dendra2 promoter in *ft* mutant, GW=GAS1:Dendra2 in WT.

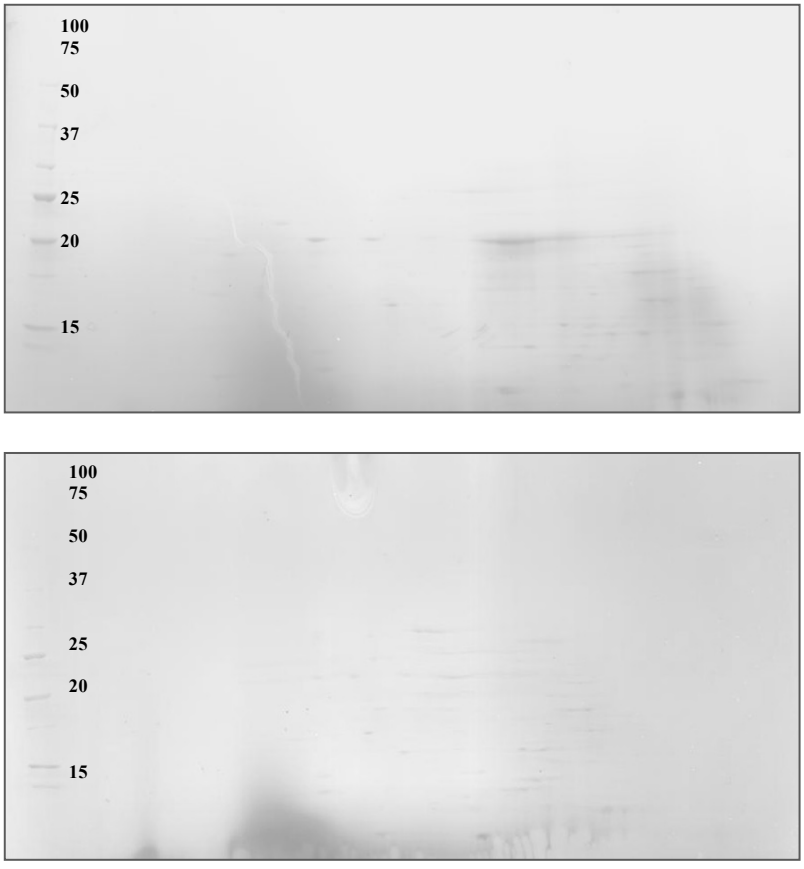
LD			
Bonferroni's Multiple Comparison Test	**Mean Diff.**	**Significant? P < 0.05?**	**Summary**
WT vs SFW1	-4.200	Yes	***
WT vs SFW2	-0.200	No	ns
WT vs SFW3	-2.000	No	ns
WT vs SFW5	-3.600	Yes	***
WT vs SFW6	-4.000	Yes	***
WT vs SFW7	-4.300	Yes	***
WT vs SFW8	-4.800	Yes	***
WT vs SFW9	-4.900	Yes	***
WT vs SFW13	-2.500	Yes	*
WT vs SFW14	-0.500	No	ns
WT vs SW2	-7.500	Yes	***
VW1 vs SFW1	3.100	Yes	***
VW1 vs SFW2	7.100	Yes	***
VW1 vs SFW3	5.300	Yes	***
VW1 vs SFW5	3.700	Yes	***
VW1 vs SFW6	3.300	Yes	***
VW1 vs SFW7	3.000	Yes	**
VW1 vs SFW8	2.500	Yes	*
VW1 vs SFW9	2.400	No	ns
VW1 vs SFW13	4.800	Yes	***
VW1 vs SFW14	6.800	Yes	***
ft-10 vs SFf3	14.50	Yes	***
ft-10 vs SFf4	14.10	Yes	***
ft-10 vs SFf5	16.00	Yes	***
ft-10 vs SFf6	12.20	Yes	***
ft-10 vs SFf7	12.50	Yes	***
ft-10 vs SFf8	11.50	Yes	***
Vf1 vs SFf3	5.200	Yes	***
Vf1 vs SFf4	4.800	Yes	***
Vf1 vs SFf5	6.700	Yes	***

SD			
Bonferroni's Multiple Comparison Test	**Mean Diff.**	**Significant? P < 0.05?**	**Summary**
WT vs SFW1	-0.700	No	ns
WT vs SFW2	2.000	No	ns
WT vs SFW3	-1.900	No	ns
WT vs SFW5	-5.000	Yes	***
WT vs SFW6	-5.300	Yes	***
WT vs SFW7	-8.100	Yes	***
WT vs SFW8	-4.900	Yes	***
WT vs SFW9	-10.70	Yes	***
WT vs SFW13	-2.700	No	ns
WT vs SFW14	-3.000	Yes	*
WT vs SW2	-10.50	Yes	***
VW1 vs SFW1	2.600	No	ns
VW1 vs SFW2	5.300	Yes	***
VW1 vs SFW3	1.400	No	ns
VW1 vs SFW5	-1.700	No	ns
VW1 vs SFW6	-2.000	No	ns
VW1 vs SFW7	-4.800	Yes	***
VW1 vs SFW8	-1.600	No	ns
VW1 vs SFW9	-7.400	Yes	***
VW1 vs SFW13	0.6000	No	ns
VW1 vs SFW14	0.3000	No	ns
ft-10 vs SFf3	15.80	Yes	***
ft-10 vs SFf4	12.70	Yes	***
ft-10 vs SFf5	20.00	Yes	***
ft-10 vs SFf6	9.900	Yes	***
ft-10 vs SFf7	8.800	Yes	***
ft-10 vs SFf8	11.50	Yes	***
Vf1 vs SFf3	5.000	Yes	***
Vf1 vs SFf4	1.900	No	ns
Vf1 vs SFf5	9.200	Yes	***

Vf1 vs SFf6	2.900	Yes	**
Vf1 vs SFf7	3.200	Yes	***
Vf1 vs SFf8	2.200	No	ns
WT vs GFW1	-6.300	Yes	***
WT vs GFW3	-6.300	Yes	***
WT vs GFW4	-6.300	Yes	***
WT vs GW5	-8.000	Yes	***
VW1 vs GFW1	1.000	No	ns
VW1 vs GFW3	1.000	No	ns
VW1 vs GFW4	1.000	No	ns
VW1 vs GW5	-0.700	No	ns
ft-10 vs GFf1	11.50	Yes	***
ft-10 vs GFf4	10.50	Yes	***
ft-10 vs GFf4	10.30	Yes	***
Vf1 vs GFf1	2.200	No	ns
Vf1 vs GFf3	1.200	No	ns
Vf1 vs GFf4	1.000	No	ns

Vf1 vs SFf6	-0.900	No	ns
Vf1 vs SFf7	-2.000	No	ns
Vf1 vs SFf8	0.7000	No	ns
WT vs GFW1	-13.7	Yes	***
WT vs GFW3	-12.70	Yes	***
WT vs GFW4	-11.90	Yes	***
WT vs GW5	-13.80	Yes	***
VW1 vs GFW1	-10.40	Yes	***
VW1 vs GFW3	-9.400	Yes	***
VW1 vs GFW4	-8.600	Yes	***
VW1 vs GW5	-10.50	Yes	***
ft-10 vs GFf1	6.700	Yes	***
ft-10 vs GFf4	6.800	Yes	***
ft-10 vs GFf4	6.200	Yes	***
Vf1 vs GFf1	-4.100	Yes	***
Vf1 vs GFf3	-4.000	Yes	***
Vf1 vs GFf4	-4.600	Yes	***

Appendix F



Spots obtained from 2D SDS-PAGE gels from the same sample were not reproducible and varies in pattern. Sample from phloem sap protein were subjected to IEF at the same time and second dimension gels were run simultaneously.

Appendix G

List of primers used for PCR of genomic DNA of transgene and rt-PCR for expression of transcript in Chapter 4.

	Forward Primer	**ReversePrimer**
SUC2 for genomic PCR	AAAATCTGGTTTCATATTAATTTCA	ATTTGACAAACCAAGAAAGTAAGA
SUC2 for rt-PCR	GCGGCAGCGGTAATGGTGT	
GAS1 for genomic PCR	GGTGAGATATGTGGCCGTTGGCT	TGGGGCAGCTGCAGGAGACAT
GAS1 for rt-PCR	ATGTCTCCTGCAGCTGCCCCA	
FTcDNA	GGGAAGGCCGAGATTGTAGAT	TAGTAAGCAGAGTTGTTGGAGACG
Dendra2	GCTCAAGCTTCGAATTCTGCAGTCG AC	ATCTACACCTGAGTCTCCAGACCAA AC
Actin	CGCCAGTGGTCGTACAACCGG	CCCGCTCTGCTGTTGTGGTGA
M13	GTTGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACCATG