

Pectin remodelling enzymes of flax and their roles in fiber development

by

David Pinzón Latorre

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

Department of Biological Sciences  
University of Alberta

© David Pinzon Latorre, 2014

## Abstract

*Linum usitatissimum* (flax) is an annual eudicot of which two types are cultivated: linseed and fiber flax. The stem fibers of linseed are not generally used commercially because they are of lower quality and yield than those obtained from fiber flax. Moreover, the extraction of fibers by dew-retting is not possible in the climate of Canada. Our goal was to study the pectin composition in the cell wall of the fibers and surrounding cells and find a set of candidate pectin-modifying genes with roles in fiber development. Flax phloem fibers elongate intrusively by diffuse growth, so, they need to penetrate between adjacent cells during elongation through the middle lamella. This is hypothesized to require dissolution of the middle lamella and sufficient rigidity of the fiber to allow penetration, while maintaining flexibility for cell wall expansion. The degree and pattern of methylesterification of galacturonic acid (GalA) residues in homogalacturonan (HG) influences the rigidity of the middle lamella and cell wall. Pectin methylesterases (PME) mediate the demethylesterification of GalA *in muro*, in either a block-wise fashion (resulting in rigidification), or random fashion (resulting in wall loosening via the subsequent action of polygalacturonases (PG) or pectate lyase like proteins (PLL)). Through immunohistochemistry, I defined some of the modifications that occur in pectin during fiber elongation, and generated a model of fiber development, in which low methylesterification of elongating fibers is associated with abundant galactan side chains that help to regulate interactions between pectins and prevent premature rigidification of fiber cell wall during its growth. I characterized the PME, the pectin methylesterase inhibitors (PMEI) and PLL gene families in flax using the recently sequenced genome, and using transcript profiling assays on nine different stages of development I defined a

set of candidate genes with roles in fiber development. I expressed one of these PMEIs in *Escherichia coli* and demonstrated that it was able to inhibit most of the native PME activity in the upper portion of the flax stem. Together, these results clarify the role of pectin modification during bast fiber development and identify targets for crop improvement.

## **Preface**

Chapter 3 of this thesis has been published as Pinzon-Latorre D, Deyholos MK, “Characterization and transcript profiling of the pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) gene families in flax (*Linum usitatissimum*)”, BMC Genomics 2013, vol. 14:742. I was responsible for the concept formation, acquisition of the data, analysis, and manuscript preparation. Deyholos MK was the supervisory author and was involved with concept formation and manuscript revisions.

## **Dedication**

To my wife, for walking side by side every day of my life, your love, support and company means everything to me. Having you to share my most difficult moments and the most memorable ones made this goal easier to achieve. Next to you, everything is possible; and I am already looking forward for our next adventure. To my parents and sister, you taught me, through your love, example, and sacrifice, the essential pillars of my life that have allowed me to reach this point. To the rest of my family, extended family, and friends, for always being there and allow me to be part of your life.

To my supervisor, Dr. Michael Deyholos, for allowing me to build this project with your unconditional support and guidance on every idea I had, which, I think, made a huge difference in my goal of being a great scientist. You are a great human being and I am sure you will be as successful in all the new projects that are about to come.

## **Acknowledgments**

I would like to thank my supervisors, Dr. Michael Deyholos, for helping me out every time that I needed him, and for providing me advice and feedback in an amazingly fast and effective manner. I would also like to thank my supervisory committee, Dr. Janice Cooke and Dr. Stephen Strelkov, for their kind words of support and dedication. Special thanks to Arlene Oatway for her amazing help in the biological sciences microscopy unit.

The past and current members of my lab made of this an amazing experience. I learnt from every one of you, I hope I have been able to contribute to your life as well. Special thanks to Mary De Pauw, Neil Hobson, and Leonardo Galindo for valuable scientific and personal conversations.

I would also like to thank Genome Canada and NSERC for funding this project.

## Table of Contents

1	Chapter 1: Introduction .....	1
1.1	Flax and fiber development .....	1
1.1.1	Flax .....	1
1.1.2	Flax fiber development .....	2
1.2	Cell wall and fiber development.....	4
1.2.1	Cell wall composition.....	4
1.2.2	Pectic Polysaccharides .....	5
1.2.3	Cell wall modifying enzymes .....	7
1.2.3.1	Pectin methylesterases.....	7
1.2.3.2	Pectate lyases and polygalacturonases .....	10
1.3	Present Study .....	11
2	Chapter 2 Characterization of pectin methylation patterns and PME activity in developing flax fibers.....	14
2.1	Introduction.....	14
2.2	Materials and Methods.....	16
2.2.1	Plant material .....	16
2.2.2	Pectin methylesterase activity.....	17
2.2.2.1	Protein extraction.....	17
2.2.2.2	Radial activity assay .....	17

2.2.3	Pectin Profiling by Immunohistochemistry .....	18
2.2.3.1	Cell wall polysaccharide extraction.....	18
2.2.3.2	Quantification of polysaccharides .....	18
2.2.3.3	Immunodots .....	19
2.2.3.4	Embedding, sectioning and labeling for fluorescent microscopy.....	20
2.2.3.5	Embedding, sectioning and labeling for transmission electron microscopy .....	21
2.2.4	Calcium subcellular localization .....	22
2.3	Results .....	23
2.3.1	Tissues corresponding to the different stages of development.....	23
2.3.2	The PME activity is lower at the top of the plant.....	26
2.3.3	The degree of methylesterification and RG-I structure change along the stem .....	30
2.3.3.1	Methylesterification state at the different developmental stages.....	30
2.3.3.2	(1-4)- $\beta$ -D-galactans are present in higher amounts in the fibers since coordinated growth starts. ....	39
2.3.3.3	Calcium subcellular localization.....	40
2.4	Discussion .....	44
2.4.1	Methylesterification state of the fibers.....	44
2.4.2	Methylesterification and calcium cross linking.....	44

2.4.3	Presence of galactan in developing fibers.....	46
2.4.4	Model for implication of the pectin polysaccharides in the development of fibers in flax .....	47
2.5	Conclusion .....	48
3	Chapter 3: Characterization and transcript profiling of the pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) gene families in flax ( <i>Linum usitatissimum</i> ).....	49
3.1	Introduction.....	49
3.2	Materials and methods .....	51
3.2.1	Annotation of PME and PMEI domain in flax and other species .....	51
3.2.2	PMEs and PMEIs in other plants.....	52
3.2.3	Primer design for qRT-PCR .....	52
3.2.4	Tissues for Quantitative Real Time PCR using a 96.96 dynamic array .....	53
3.2.5	Analysis of 96.96 dynamic array results.....	55
3.2.6	EST and RNAseq data mapping.....	55
3.2.7	Signal peptide, transmembrane domain, and protein subcellular localization predictions.....	55
3.2.8	Cleavage site prediction .....	55
3.2.9	Isoelectric point .....	56
3.2.10	Phylogenetic analysis .....	56

3.2.11	Conserved residues.....	57
3.3	Results and Discussion.....	57
3.3.1	Annotation of LuPMEs and LuPMEIs.....	57
3.3.2	Detection of PMEs and PMEIs in other plants.....	60
3.3.3	Transcript expression profiling.....	60
3.3.4	qRT-PCR using a 96.96 dynamic array .....	61
3.3.5	Transcript expression patterns .....	73
3.3.6	Protein subcellular localization .....	78
3.3.7	Cleavage site.....	79
3.3.8	Isoelectric point .....	80
3.3.9	Phylogenetic analysis .....	84
3.3.10	Conserved residues in PMEs .....	92
3.3.11	Conserved residues PMEI.....	95
3.3.12	Gene expression and conserved residues.....	96
3.4	Conclusion .....	98
4	Chapter 4: Characterization of pectinmethylesterases (PME) and pectinmethylesterase inhibitors (PMEI ) enriched during flax fiber development in ( <i>Linum usitatissimum</i> ) .....	99
4.1	Introduction.....	99
4.2	Materials and Methods.....	100

4.2.1	Plant material .....	100
4.2.2	RNA extraction.....	101
4.2.3	cDNA synthesis and quantitative real time PCR.....	101
4.2.4	Gene clustering based on expression .....	102
4.2.5	Heterologous expression .....	102
4.2.6	LC MS/MS.....	104
4.2.7	PME activity .....	104
4.2.8	PMEI inhibitory activity .....	104
4.3	Results .....	104
4.3.1	Selection of candidate genes.....	104
4.3.2	Transcript expression profiling of selected LuPMEs and LuPMEIs.....	105
4.3.3	Clustering of transcript expression data .....	113
4.3.4	Heterologous expression .....	128
4.4	Discussion .....	137
4.4.1	Genes enriched in fiber containing tissues.....	137
4.4.1.1	Genes enriched during fiber elongation .....	137
4.4.1.2	Genes enriched below the snap point .....	139
4.4.2	Genes enriched in the xylem.....	140
4.4.3	PMEI inhibitory activity .....	142
4.5	Conclusion .....	142

5	Chapter 5: A pectate lyase enriched during flax intrusive growth development.....	144
5.1	Introduction.....	144
5.2	Materials and Methods.....	145
5.2.1	Annotation of PLL in flax.....	145
5.2.2	Phylogenetic analysis.....	145
5.2.3	Microarray analysis.....	146
5.2.4	Plant growth and qRT-PCR.....	146
5.2.5	Heterologous expression.....	146
5.2.6	LC MS/MS.....	147
5.2.7	Pectate lyase assay.....	147
5.3	Results.....	147
5.3.1	Pectate lyase-like proteins of flax.....	147
5.3.2	Phylogenetic relations of PLL proteins.....	148
5.3.3	Pectin degrading enzymes from microarray data.....	152
5.3.4	Transcription profiling of a PLL gene in whole stem and stem peel tissues... .....	155
5.3.5	Heterologous expression of PLL4.....	155
5.3.6	Pectate lyase activity assay.....	160
5.4	Discussion.....	162

5.4.1	The putative <i>PLL4</i> may be important for fiber growth.....	162
5.4.2	PLL4 could aid in the intrusive growth of fibers .....	163
5.5	Conclusion .....	163
6	Chapter 6: General Discussion .....	165
6.1	The methylesterification state of the cell walls and middle lamella changes during fiber development.....	165
6.2	Galactans side chains of RG-I impede primary cell wall strengthening during elongation.....	167
6.3	The PME and PMEIs in flax play diverse roles in the fibers and the plant development in general .....	170
6.4	Pectin degrading enzymes are enriched during fiber elongation.....	176
6.5	Conclusion .....	177
7	Bibliography .....	178
8	Appendix.....	198
8.1	Microarray analysis.....	198

## List of Tables

### Chapter 2

Table 2 - 1 Tukey's multiple comparisons test of Pectin Methylesterase activity along the flax stem. ....	29
---	----

### Chapter 3

Table 3 - 1 Important features of type-1 LuPMEs .....	64
Table 3 - 2 Important features of type-2 LuPMEs .....	67
Table 3 - 3 Important features of LuPMEIs.....	69
Table 3 - 4 Genes putatively associated with fiber development during elongation (*), thickening (¥), and thickening and maturation (§) processes.....	90

### Chapter 4

Table 4 - 1 Selected LuPMEs and LuPMEIs for expression profiling at different fiber developmental stages. ....	106
Table 4 - 2 Maximum fold change of expression between two tissues of selected PME and PMEIs. ....	110
Table 4 - 3 Tissue enrichment expression of selected flax PMEs and PMEIs in whole stem compared to stem peels.....	111
Table 4 - 4 Tukey's multiple comparisons test of dCT expression values between whole stem tissues. ....	123
Table 4 - 5. Tukey's multiple comparisons test of dCT expression values between stem peel tissues.....	127
Table 4 - 6 LC MS/MS analysis results of bands marked with an asterisk in figures 4-4 and 4-5. ....	135

## **Chapter 5**

Table 5 - 1. Main features of the predicted PLL proteins in flax..... 150

Table 5 - 2 Tukey's multiple comparisons test for the expression of different genes  
between the five different tissues used in flax stem microarray..... 154

Table 5 - 3 Tukey's multiple comparisons test for the expression of PLL4 in whole stem  
and stem peel tissues..... 158

Table 5 - 4 LC MS/MS analysis results of band marked with an asterisk in figures 5-3.159

## **Appendix**

Table A3 - 1 Primers used for qRT-PCR, indicating TaqMan probe. .... 199

Table A3 - 2 Genetic distance between possible paralogs in the LuPMEIs gene family 207

## List of Figures

### Chapter 2

Figure 2 - 1 Location in the flax stem of the tissues used for the experiments used in this study.....	24
Figure 2 - 2 Cross sections of flax showing the tissues that are obtained when the stem is peeled at the different points used. ....	25
Figure 2 - 3 PME activity of native flax proteins detected as the decrease in methylesterification of highly esterified pectin measured by the binding of ruthenium red. ....	28
Figure 2 - 4 Immunodot of flax stem peel tissues polysaccharides from different stages of fiber development. ....	31
Figure 2 - 5 Immunodot of flax whole stem tissues polysaccharides from different stages of fiber development.....	32
Figure 2 - 6 Epifluorescence labeling with LM19 antibody of flax stem cross sections at different stages of fiber development. ....	35
Figure 2 - 7 Epifluorescence labeling with LM20 antibody of flax stem cross sections at different stages of fiber development. ....	36
Figure 2 - 8 Transmission electron microscopy of fibers at stages B and D labeled by immunogold. ....	38
Figure 2 - 9 Epifluorescence labeling with LM5 antibody of flax stem cross sections at different stages of fiber development. ....	41
Figure 2 - 10 Calcium subcellular localization in flax fibers along the stem using potassium antimonate .....	42

### Chapter 3

Figure 3 - 1 Representatives of the three types of proteins classified in this report .....	59
Figure 3 - 2 Heat map of transcript abundance of PMEs (A) and PMEIs (B) at different tissues.....	63
Figure 3 - 3 Percentage of genes detected by qRT-PCR per tissue.....	76
Figure 3 - 4 LuPME transcript expression in various tissues .....	76
Figure 3 - 5 LuPMEI transcript expression in various tissues.....	77
Figure 3 - 6 Subcellular localization of PMEs and PMEIs.....	82
Figure 3 - 7 Size of mature PMEs and PMEIs proteins. ....	82
Figure 3 - 8 Isoelectric point of mature PMEs and PMEIs proteins. ....	83
Figure 3 - 9 Maximum likelihood dendrogram of PMEs in flax and related species .....	86
Figure 3 - 10 Maximum likelihood dendrogram of PMEIs in flax and related species.....	87
Figure 3 - 11 Tertiary structures showing conserved residues of PMEs and PMEIs and their level of conservation in flax. ....	94

### Chapter 4

Figure 4 - 1 Transcript expression of PMEs and PMEIs of interest from whole stem and stem peel tissues.....	109
Figure 4 - 2 Clusters of transcript expression patterns in segments of whole stem tissues. ....	115
Figure 4 - 3 Clusters of transcript expression patterns in segments of stem peels.....	118
Figure 4 - 4 Purification of LuPME67 and LuPME79 expressed in <i>E. coli</i> . ....	131
Figure 4 - 5 Purification of LuPMEI45 expressed in <i>E. coli</i> . ....	132
Figure 4 - 6 Inhibitory capacity of LuPMEI45 expressed in <i>E. coli</i> . ....	132

Figure 4 - 7 Radial assay of inhibitory capacity of LuPMEI45. ....	133
Figure 4 - 8. LuPMEI45 inhibitory activity on flax proteins extracted from whole stem and stem peel tissues at different stages of development of the fiber. ....	134
<b>Chapter 5</b>	
Figure 5 - 1 Distribution of size of the predicted PLL proteins in flax. ....	149
Figure 5 - 2. Maximum likelihood dendrogram of PLL in flax .....	151
Figure 5 - 3 qRT-PCR validation of genes of interest using same tissues as those used by To [125] in the microarray. ....	153
Figure 5 - 4. Expression of PLL4 in whole stem and stem peel tissues. ....	156
Figure 5 - 5. Purification of pectate lyase PLL4 expressed in <i>E. coli</i> . ....	157
Figure 5 - 6 Activity of PLL4 measured as the production of unsaturated products from the hydrolysis of PolGalA. ....	161
<b>Chapter 6</b>	
Figure 6 - 1 Model showing the effect of galactan side chains on calcium cross linking during fiber elongation. ....	169
<b>Appendix</b>	
Figure A2 - 1 Standard curve of PME activity by radial assay. ....	209
Figure A3 - 1 Percentage of PMEs and PMEIs, respect to the total number of proteins, in Embryophyta plants with available full genomes in Phytozome (version 9.1).....	210

## **List of Abbreviations**

PME: Pectin Methylesterase

PMEI: Pectin Methylesterase Inhibitor

PG : Polygalacturonases

PL: Pectate Lyases

PLL: Pectate Lyases like

HG: Homogalacturonan

RG-I: Rhamnogalacturonan I

RG-II: Rhamnogalacturonan II

$\alpha$ -L-Rha:  $\alpha$ -L-rhamnose

$\alpha$ -D-GalA:  $\alpha$ -D-galacturonic acid

CML: Compound middle lamella

# 1 Chapter 1: Introduction

## 1.1 Flax and fiber development

### 1.1.1 Flax

Flax (linseed, *Linum usitatissimum*) is an annual eudicot grown predominantly in temperate climates. It belongs to the order Malpighiales, which includes other economically important species that have been sequenced, such as cassava (*Manihot esculenta*), castor (*Ricinus communis*), and poplar (*Populus* spp.). The family Linaceae encompasses 14 genera, with more than 270 species widespread around the world [1]. The genus *Linum* consists of over 180 species [2]. Flax fibers have been identified at archeological sites from the Upper Paleolithic (30,000 years ago) in the Caucasus region of Georgia [3]. The first archeological evidence for the use of linseed dates to 8000 BC in Syria [4]. During the last century, demand for flax fibers and linseed oil decreased as cotton and petroleum became more widely available [2]. More recently, interest in flax has increased as seed components such as lignans and  $\alpha$ -linolenic acids have been reported to have potential health benefits [2], and because of efforts to develop flax fibers as a replacement for glass fibers in composite materials [5].

Cultivated flax is presumed to have originated from a wild ancestor with similarity to *L. bienne* [6]. *L. usitatissimum* is a self-pollinating diploid ( $n=15$ ), with a genome size of 373 Mb, comprising 43,384 predicted protein-coding genes [1]. Two types of flax are cultivated: fiber flax, and linseed. Fiber flax is grown for its phloem fibers, which are valued in textile manufacturing because of their unusually high length and strength.

Linseed is grown for the oil contained in its seeds [4, 7]. Only linseed is grown in Canada. Its fibers are not used commercially, because they are of lower quality and yield than those obtained from fiber flax. More importantly, the climate of Canada does not permit dew-retting, in which naturally occurring microbes help to separate phloem fibers from surrounding tissues [8]. The time from seeding to harvesting of linseed flax ranges between 90 and 125 days [9]. Improving the efficiency of fiber extraction from Canadian-grown linseed would be a major contribution in the development of dual purpose flax, which is a long-term goal of the flax industry.

### 1.1.2 Flax fiber development

Fibers provide structural support and develop in the parts of the plant that have stopped elongating [10]. Flax contains both xylem fibers and primary phloem fibers. Xylem fibers of flax have a typical secondary cell wall that is lignified, while flax phloem fibers are very long (average 2-5 cm [11]), are rich in cellulose (60-70% [12]), with high crystallinity, and have very low lignin (1.5-4.2% [13] reported to be as low as 0.4% [14]). These phloem fibers are known also as bast fibers and they will be referred to simply as fibers throughout the rest of this document. Flax fibers are of a gelatinous-type, which is also found in reaction wood [15]. Two layers can be distinguished in the secondary cell wall of developing flax fibers [16]. The innermost layer is a loosely packed, galactan-rich matrix (galactan layer, Gn-layer) [16]. The outermost layer is the mature, gelatinous layer (G-layer), in which crystalline cellulose has replaced the galactan-rich matrix of the Gn-layer. As the fiber matures, the G-layer gradually replaces the entire Gn-layer [16].

Flax fibers originate exclusively from the primary meristems of the shoot [17]. Elongation begins approximately 400  $\mu\text{m}$  from the shoot apex [18]. It occurs first by

symplastic growth, through which fibers can reach lengths of 70 to 100  $\mu\text{m}$ . During symplastic growth, fibers elongate at the same rate as the surrounding cell-types. While the surrounding cells are still elongating (at approximately 1 to 2 mm from the shoot apex), fibers begin diffusive-intrusive growth [18]. Intrusive growth is characterized by faster elongation of the fibers relative to surrounding cells, a hypothesized increase in turgor pressure, and the formation of knee-like structures at the fiber ends [19]. Intrusive elongation requires that fibers extend themselves through the middle lamella that separates hundreds of neighboring cells, even destroying plasmodesmata in the process. Growth is diffusive, as the entire cell (and not only its tips) expands [18]. Once intrusive growth ceases, the fibers start the deposition of secondary cell wall material.

Gorshkova and collaborators [20] identified a mechanically definable region in the flax stem, which they named the snap point. Above this point, fibers grow intrusively and do not deposit secondary cell walls, so the stem is easily broken. Below this point, the stem gets stronger as the fibers acquire a secondary cell wall. Within each fiber bundle, secondary thickening begins in fibers that are closest to the epidermis, and continues centripetally until secondary wall deposition has been initiated also in the cambial side of the bundle. Later, further wall thickening occurs throughout the whole bundle [20]. At maturity, the fiber cell wall is 5-15  $\mu\text{m}$  thick [11], and the connections between fibers within a bundle are stronger than the connections between fibers and other cell-types [21]. This is the basis of dew-retting [8], which degrades the middle lamella surrounding the fiber bundle, but largely leaves the bundle intact.

## 1.2 Cell wall and fiber development

The cell wall confers protection and mechanical strength to the plant. Plant cells have a primary cell wall, which is similar in composition throughout the plant, and has the ability to expand. Each facet of the cell wall is generated in an independent cell division. When division takes place, a cell plate is formed at the center of the cell and it extends until it reaches the mother cell wall, generating two halves, at this point callose is removed, and cellulose is deposited at both sides of the plate, without crossing from one side to the other, generating two distinguishable cell walls separated by a middle lamella, which is rich in pectic polysaccharides. The middle lamella provides a contact interface between neighboring cells, and is responsible for adhesion between cells. The primary cell wall of the mother cell is locally dissolved to allow the junction of the newly generated middle lamella with the middle lamella of the mother cell wall, creating a junction rich in polysaccharides [22, 23].

The pectic polysaccharides are mainly present in the middle lamella and the primary cell wall, so the distribution and methylesterification state directly affects the physical properties of the cell and how it adheres to other cells. The following section explores the composition of the cell wall and some of the modifications to it that can occur.

### 1.2.1 Cell wall composition

The cell wall is made up of cellulose, pectic polysaccharides, hemicellulose, and proteins. Cellulose is a linear polysaccharide of  $\beta$  (1 $\rightarrow$ 4) linked D-glucose units. These units have the capacity to generate intra- and intermolecular hydrogen bonds, making up the microfibrils, which are cross-linked via hydrogen bonds with hemicellulose, a diverse

group of polysaccharides characterized by the presence of  $\beta(1-4)$ -linked backbones (e.g. xyloglucan) [24]. The hemicellulose forms an interface between the cellulose and the rest of the components of the cell wall. The orientation and organization of the microfibrils directly affects the shape and growth of the cell, as well as its mechanical properties [25]. In flax fibers, cellulose microfibrils in the G-layer are oriented in parallel to the longitudinal axis of the fiber, and interact laterally with one another, entrapping polysaccharides during the crystallization [26], specifically pectic galactan [27]. These pectic polysaccharides are the main determinants of the porosity of the cell wall [23], as will be explained in the following section.

### 1.2.2 Pectic Polysaccharides

Pectins are a family of polysaccharides containing Galacturonic Acid (GalA), which makes up ~70% of the pectin in the cell wall. Pectins comprise ~35% of the primary cell wall in dicots, and are especially abundant in the primary cell wall of growing and dividing cells, and in the cell corners and compound middle lamella [28]. The main types of pectin are homogalacturonan (HG), rhamnogalacturonan II (RG-II), and rhamnogalacturonan I (RG-I) [29].

HG is the most abundant pectin, making up between 55 and 70% of the pectic polysaccharides [30]. It is made of (1-4)-linked  $\alpha$ -D-GalA residues; it may be O-acetylated at O-2 or O-3, and it is methylesterified in the medial cisternae in the Golgi apparatus [31] by pectin methyltransferases, which transfer a methyl group from S-adenosyl-methionine to the C-6 of the GalA residue [32]. HG is then exported to the cell wall in a highly methylesterified form [33], where it can be demethylesterified by pectin methylesterases.

RG-I makes up 20 to 35% of pectins [28]. Its backbone consists of alternating  $\alpha$ -L-rhamnose ( $\alpha$ -L-Rha) and  $\alpha$ -D-GalA ( $-\alpha$ -D-GalA-1,2- $\alpha$ -L-Rha-1-4-]<sub>n</sub>. The GalA residues have an O-acetyl group at C2 or C3. RG-I can have side chains of various sizes, attached to the C4 of the rhamnosyl residues. The side chains are composed mostly of arabinosyl and galactosyl residues, and in lower amounts, fucosyl and glucosyluronic acid residues, including  $\alpha$ -1,5-linked L-arabinan and  $\beta$ -1,4-linked D-galactans [28].

RG-II makes up 1-4% of the pectin in the primary cell wall of angiosperms and gymnosperms [34]. Twelve different monosaccharides have been identified in RG-II, with  $\alpha$ -L-Rha and  $\alpha$ -D-GalA as the main components. Its structure is highly conserved among plants. Its backbone is made of at least eight (1-4)-linked  $\alpha$ -D-GalA residues, which can be methylesterified, and are modified by four complex side chains. RG-II is often present as a dimer, which structure appears to be critical for the normal growth and development of the plant [28]

The GalA backbones of HG have the capacity to form complexes by calcium cross linking (in the case of de-esterified GalA), and through hydrogen bonding and hydrophobic interactions (in the case of esterified GalA). When a minimum of nine contiguous residues of GalA are non-esterified, their negative charges ionically interact with calcium ions, forming calcium bridges between GalA backbones and strengthening the cell wall. This is known as the “egg-box” model [35], which accounts for 70% of the pectic gel in the plant cell wall [36]. HG, RG-I and RG-II are believed to be linked via their backbones [28]. HG domains need to be in close proximity for calcium cross linking to occur. Consequently the side chains of RG-I and the dimerization of RG-II are postulated to interfere with the formation of calcium bridges, which would be responsible

for the generation of pores in the cell wall in the space between cellulose microfibrils [23]. Furthermore, it has been found that a decline in wall expansion is correlated with a reduction in the arabinan and galactan side chains of RG-I, and an increase in demethylation of HG, accompanied by an increase in cross linked HG [37], which further demonstrates that the side chains of RG-I are a physical obstruction to the formation of calcium cross linking.

Consequently, the control of the level of methylesterification in the cell wall plays a key role in the presence and size of pores, in the adhesion forces between cells, and in the strength of the cell wall. The following section will explore the enzymes that modify the methylesterification state of the cell, and those enzymes that depend on the methylesterification state that can contribute to the loosening of the cell wall.

### 1.2.3 Cell wall modifying enzymes

#### 1.2.3.1 Pectin methylesterases

Once the highly methylesterified HG is secreted into the cell wall, the GalA residues can be demethylesterified by pectin methylesterases (PME), generating negatively charged carboxyl groups. The PMEs are a multigene family, first described by Richard and collaborators [38]. They are associated with the cell wall by ionic interactions [39], and are classified in CAZy (carbohydrate-active enzymes) as class 8 of the carbohydrate esterases (EC 3.1.1.11) [40]. There are multiple PME isoforms in a given plant; they have a variety of optimal pHs and are expressed in various tissue-specific patterns and at various stages of development. In Arabidopsis, 66 PMEs have been annotated [41]. They are subdivided into two categories, depending on the presence of a pro-region, which is

similar to the PMEI domain (see below): Type-1 PMEs, which possess the pro-region; and Type-2 PMEs which do not. PMEs have two modes of action: (i) blockwise demethylesterification, which leads to the demethylesterification of contiguous GalA residues, and thus facilitates calcium cross linking and rigidifying of the cell wall; and (ii) random demethylesterification, which is the demethylesterification of non-contiguous GalA residues, which can lead to cell wall loosening as the molecule becomes a substrate for pectin degrading enzymes, including pectate lyases and polygalacturonases [42, 43]. The isoelectric point of respective PMEs has been correlated with their mode of action: PMEs with an acidic pI are likely to randomly demethylesterify, while PMEs with an alkaline pI tend to have blockwise activity [39]. The activity of the PMEs is tightly regulated by pectin methylesterase inhibitors (PMEI), another multigene family [44], discovered in kiwi (*Actinidia chinensis*) [45]. PMEIs bind to the active site of the PME, generating a 1:1 complex in a stable, non-covalent, reversible interaction [46, 47].

Although the PMEs are all very similar in sequence, it has been demonstrated that different isoforms have non-redundant functions. Jiang and collaborators [48] showed that when the knockout of *VANGUARD1* (an Arabidopsis PME) was complemented with a closely related PME, the wild type phenotype was re-established. However, a less similar isoform was not able to complement the mutation. These results show that not all PMEs are functionally redundant. It is therefore necessary to characterize the complete set of PMEs in a species to identify genes associated with specific physiological functions and stages of development.

Although a large number of predicted PMEs and PMEIs have been identified through DNA sequencing projects, relatively few studies have demonstrated their activity. The

first functional characterization of a PME was accomplished for *PMEU1*, a tomato PME, by Gaffe and collaborators [49]. In flax, only three PMEs have been described, *LuPME1*, *LuPME3*, and *LuPME5* [50-53]. *LuPME5* was found to have a higher expression in elongating parts of the hypocotyl, so it was proposed to be involved in cell wall stiffening after elongation [52]. Two Arabidopsis PMEs, *VANGUARD1* [48] and *AtPPME1* [54], were found by mutant analysis to be required for pollen tube growth. *QUARTET1* (*QUA1*, another Arabidopsis PME) is responsible for the separation of pollen tetrads through demethylesterification of the pollen mother cell wall [55]. Furthermore, *QUA1* works upstream of a polygalacturonase, *QUARTET3*, which allows the degradation of the mother cell wall [56]. Siedlecka and collaborators [57] found a PME with a possible role in intrusive growth of fibers in hybrid aspen (*Populus tremula* x *tremuloides*): *PttPME1* had higher expression in developing wood than in non-growing tissue. When *PttPME1* was downregulated, it stimulated fiber elongation. It was therefore concluded that *PttPME1* activity hinders intrusive growth of the fibers, presumably by strengthening cellular adhesion between the cells that the fibers would penetrate.

Balestrieri and collaborators [44] were the first ones to report the inhibitory activity of a PME (from kiwi fruit), which was later purified and confirmed to interact with a PME [58]. In another example, *BoPMEII* (from *Brassica oleracea*) was demonstrated to have a role in pollen tube growth. When the ortholog of this gene was silenced in Arabidopsis by expression of antisense *BoPMEII*, there was a reduction in pollen tube growth, suggesting a higher PME activity. If this activity was random, it may have led to loosening of the pollen tube cell wall, resulting in shorter pollen tubes [59]. This result contrasts with the outcome of the *PttPME1* silencing experiment described above [57],

as PttPME1 activity affects the middle lamella, while *BoPME11* silencing affected the activity of PMEs in the primary cell wall [57, 59], leading to two different outcomes. Also, based on transcription data, Cooke and collaborators [60] described a PME1 from hybrid poplar (*Populus balsamifera* ssp. *trichocarpa* × *deltoides*) that could be involved in cell wall maturation during leaf development.

#### 1.2.3.2 Pectate lyases and polygalacturonases

Demethylesterification of GalA residues of HG may increase their susceptibility to degradation by enzymes such as polygalacturonases (PG, EC 3.2.1.15) and pectate lyases (PL, EC 4.2.2.2 [40]). Pathogens are one source of these enzymes and seek to weaken cell walls of a host plant during infection [23]. PG and PL are also encoded by plants, indicating additional roles for pectin degradation outside of pathogenesis [61]. PGs hydrolyze demethylesterified HG and require as many as four consecutive demethylesterified GalA residues for substrate recognition. PLs require calcium ions as cofactors and catalyze  $\beta$ -elimination cleavage of demethylesterified HG [61]. PL can only cleave glycosidic linkages attached to the O-4 of unesterified D-galacturonic acid, so they potentially can only cleave HG and RG-I, however, RG-I is sterically inaccessible to them. PLs differ from pectin lyases (PELs, EC 4.2.2.10) in the fact that the latter can only cleave methylesterified D-galacturonic acid [23]. *GhPEL* is a PL from cotton (that was incorrectly named a PEL rather than a PL for historical reasons) that was determined to loosen cell walls. It was proposed that GhPEL activity loosens the cell wall to facilitate the rapid growth of the seed trichomes (i.e. cotton fibers) [62].

### 1.3 Present Study

The main objective of this project is to determine the role of pectin modifying enzymes such as PME, PMEI, and PL, in flax fiber development, through the discovery and characterization of genes. It is expected that this information will support efforts to develop flax varieties with improved phloem (bast) fiber extractability.

One of the main obstacles to the utilization of linseed flax fibers produced in Canada is the extraction of the fibers. In warmer climates, this is usually done by dew-retting. Other methods, such as enzymatic retting, are not cost-effective, or contaminate large portions of water, as water-retting. The best alternative to this problem is to be able to generate plants in which the pectins that glue the fibers from surrounding tissues are weakened, so the extraction of the fibers is facilitated. This may also increase fiber length. Both of the length and extractability of fibers are presumed to be dependent on the activity of PMEs, PMEIs, PGs, and PLs.

As described above, when a fiber is growing intrusively, it needs to penetrate the middle lamella of surrounding tissues. This is likely to require partial digestion of the middle lamella [19]. I hypothesize that in order to do so, two conditions would need to be met. First, for PG and PL to act in the middle lamella, random demethylesterification must occur. Second, the fiber must gain rigidity to penetrate between adjacent cells, and withstand the pressure from surrounding cells. This involves blockwise demethylesterification and calcium cross linking. Also, when the fibers elongate through the surrounding cells, a new contact interface is generated, the rigidity of this newly generated middle lamella directly influences the extractability of the fibers, so the activity

of PME's will also likely affect the rigidity of this new connection. Which can also be affected by the borate mediated dimerization of RGII and the side chains of the RGI [36]. It is also possible that ferulate esters cause cross linking of polysaccharides, as is observed in grasses [63], however, little is known about their presence in flax.

To determine the presence and distributions of pectic polysaccharides in different stages of fiber development, I used immunodots, epifluorescent microscopy, and immunogold microscopy (Chapter 2). I used a set of antibodies LM5, LM19, LM20, and 2F4 [64-66] that recognize different epitopes in the pectic polysaccharides, to establish how methylesterification status, the presence of galactans, and calcium cross linking varies at different stages of fiber development.

I used bioinformatics to characterize PME's, PMEIs and PLs in the recently sequenced flax genome [1], and to determine their expression based on EST databases and microarray data (Chapter 3). This, combined with phylogenetic analysis and expression profiling in different tissues/stages-of-development in flax, allowed me to propose a role for specific genes in fiber development for specific genes.

Once candidate genes were established, the expression of these genes was verified in nine stages of fiber development, in both whole stem and cortical peels (Chapter 4). The genes that showed the greatest change in expression from the different stages of development were heterologously expressed in *E. coli* to confirm their enzymatic activity and their mode of action.

Finally, I analyzed an unpublished flax stem microarray data set to identify PLs and PGs that are differentially expressed during and fiber development in the stem. I found

PLs and a PG that were enriched during intrusive growth (Chapter 5), and I assessed the expression of one of the PLs in nine stages of fiber development and heterologously expressed it in *E. coli*.

## **2 Chapter 2 Characterization of pectin methylation patterns and PME activity in developing flax fibers**

### **2.1 Introduction**

Flax phloem fibers are long (2 - 5 cm), multinucleate cells [67], and are arranged in bundles of 12 to 36 fibers [11]. Their length, strength, and composition (high content of crystalline cellulose [68]) makes them useful in production of textiles, paper, and composites [5]. The term fiber will be used here to refer exclusively to phloem fibers, which are distinct from xylem fibers.

Flax fibers start elongating by symplastic growth at 400  $\mu\text{m}$  from the shoot apex. The mode of growth gradually becomes diffusive-intrusive at 1 to 2 mm from the shoot apex [18]. Once intrusive growth ceases, just above the so called “snap point” [20], secondary cell wall deposition begins within each bundle in the fibers closest to the epidermis. This is followed by a simultaneous thickening and maturation of the secondary cell wall within the whole bundle [20]. During the maturation of the secondary cell wall, two layers are visible: a loosely packed galactan rich matrix (the Gn-layer) which, as it matures, is gradually converted into the gelatinous G-layer, which is rich in crystalline cellulose [16].

Flax fibers contain two kinds of galactan: (i) typical  $\beta$ -1,4-linked D-galactan side chains of the rhamnogalacturonan I (RG-I), which are fixed within the cell wall [12, 16], and (ii) buffer soluble  $\beta$ -1,4-linked D-galactan, which is a highly branched RG-I-like molecule with mainly short side chains of  $\beta$ -1,4-linked D-galactans [12, 15]. Gorshkova and collaborators [12] determined that the stem peels bearing thickening fibers contained

a tissue specific, soluble galactan, which was absent from both the shoot tip (top 0.5 cm) and the xylem, but was present below the snap point, and was characterized by a high molecular mass. Almost all of the rhamnose residues in RG-I were substituted with this soluble galactan, which was later found to be localized to secondary cell walls of flax phloem fibers, below the snap point [16], and is degraded as the fibers mature [16, 69], by  $\beta$ -galactosidases [70, 71]. Several immunohistochemical studies have demonstrated the presence of galactans in Gn-layers of flax fibers below the snap point, [17, 72, 73], but not above it. Although galactan epitopes have also been found in primary cell walls [73], this was also in cells below the snap point.

The pectins present in the fibers and surrounding cells are believed to play an important role during fiber elongation. Pectins are constantly modified. Homogalacturonan (HG), a pectic polysaccharide made of (1-4)-linked  $\alpha$ -D-GalA residues, is deposited in the cell wall in a highly methylesterified form [32], and is selectively demethylesterified by pectin methylesterases (PMEs) [33, 38]. Depending on the mode of action, PME mediated demethylesterification can lead to either strengthening of the cell wall (blockwise demethylesterification, followed by calcium cross linking of homogalacturonan domains), or loosening (random demethylesterification followed by cell wall degrading enzymes) [33, 42]. Consequently, the methylesterification state of the cell wall affects fiber properties at many stages of development. During intrusive growth, fibers grow between cells, which implies that the middle lamella would need to be digested [19, 57], this implies a random demethylesterification leading to cleavage by polygalacturonases (PGs) or pectate lyases (PLs) [42]. Furthermore, as fibers penetrate between cells, they must resist pressure from surrounding tissues, which requires that the

cell wall be stiff but not so rigid to prevent further cell wall expansion [25, 74-76].

Snegireva and collaborators [19] have previously demonstrated that above the snap point, detection of the LM19 epitope (low methylesterified HG), is higher in the fibers than in surrounding tissues. The same researchers also showed that LM20 antibody (which recognizes high methylesterified HG) weakly binds to all cell-types above the snap point.

Here we explore the state of the pectic polysaccharides of the cell wall of flax by using specific antibodies to label fibers and surrounding tissues at successive stages of development, including elongation and secondary cell wall deposition. We also use antimonate staining and transmission electron microscopy to identify calcium accumulation patterns in developing fibers. This will help to define the dynamics of pectin methylation (and by inference cell rigidity) that occur during fiber development.

## **2.2 Materials and Methods**

### **2.2.1 Plant material**

Plants were grown in a growth chamber at 22°C, with 16 hours day length, and were fertilized with 3 g/L of a 20-20-20 water soluble fertilizer (Plant-Prod) every two weeks. The soil was left to almost dry before watering the plants again.

Tissue was collected when plants reached between 46 and 48 cm, which occurred approximately five weeks after germination. At the time of harvest, the snap point was at an average distance of 7.1 cm from the shoot apex. In all cases, the leaves were removed. Sections 1-cm long were collected from positions along the stem as either whole stem, or as stem peels. Sections were collected at 9 positions along the stem, based on the stage of development of the fiber [20], as follows: 0 to 1 cm (SA), 1 to 2 cm (1-2), 2 to 3 cm (2-3),

3 to 4 cm (3-4), 4 to 5 cm (A), 11.5 to 12.5 cm (B), 18 to 19 cm (C), 32 to 33 cm (D), and 44 to 45 cm (E). A biological replicate consisted of the 1-cm fragments obtained from ten different plants that reached 46 to 48 cm the same day.

## 2.2.2 Pectin methylesterase activity

### 2.2.2.1 Protein extraction

Proteins were extracted from three biological replicates according to the protocol of Hongo and collaborators [77]. Seven fragments (1 cm length each) obtained from equivalent positions along stems of different individuals were pooled for each extraction. Tissues were ground in liquid nitrogen, and 1 mL of extraction buffer, containing 12.5 mM citric acid, 50 mM phosphate buffer pH 7.0, with 1 M NaCl plus one tablet per 10 mL of cComplete ULTRA protease inhibitor (Roche), was added. The sample was incubated at 4°C for 4 hours on a rocker and was then centrifuged at 15,000 rcf for 15 min, and the supernatant was collected. The protein concentration was determined using Qubit® Fluorometric Quantitation (Life Technologies).

### 2.2.2.2 Radial activity assay

The radial assay was done with three biological replicates and three technical replicates as described by Downie and collaborators [78], with modifications [77]. 2% (w/v) agarose was dissolved in McIlvaine buffer with pH adjusted to 6.0 and 7.0 and was autoclaved, after which 0.1% (w/v) of highly methylesterified pectin (~85% methyl esterified) (Sigma-Aldrich, P9561) was added and dissolved. 13 mL of this mixture was poured into 90 mm Petri dishes. After cooling, wells with a diameter of 4 mm were punched in the agarose using a micropipette tip. 10 µL of freshly extracted protein (396

$\mu\text{g/mL}$ ) plus 10  $\mu\text{L}$  of 50 mM Tris HCl 300 mM NaCl buffer were dispensed into each well. This was incubated for 18 h at 28°C, and the gel was stained with an aqueous solution of 0.05% (w/v) ruthenium red for 1 h and washed with distilled water. The plates were photographed immediately and the area of the halo was measured using ImageJ [79].

### 2.2.3 Pectin Profiling by Immunohistochemistry

#### 2.2.3.1 Cell wall polysaccharide extraction

Polysaccharides were extracted from three biological replicates as described by Lionetti and collaborators [80]. Seven fragments (1 cm length each) obtained from equivalent positions along stems of different individuals were pooled for each extraction. The tissues were ground and the pellet was washed (vortex and centrifugation at 10,000 rcf for 10 min) twice with 70% ethanol, then twice with chloroform:methanol (1:1, vol/vol), and twice with 80% acetone, and then allowed to dry at ambient temperature. The pellet was then homogenized for 10 min at 80°C in 50 mM Tris-HCl and 50 mM trans-1,2-cyclohexanediaminetetraacetic acid (CDTA), pH 7.2. This solution was centrifuged at 10,000 rcf for 10 min, and the supernatant obtained, which was dialyzed with water in an Amicon Ultra 3K centrifugal filter unit (Millipore). The samples were then freeze dried and dissolved in water.

#### 2.2.3.2 Quantification of polysaccharides

Polysaccharides were quantified according to Dubois and collaborators [81] with an extra boiling step. In a glass vial, we mixed 200  $\mu\text{L}$  of sample and 200  $\mu\text{L}$  of 5% v/v phenol, to which we added 1 mL of sulfuric acid. This mixture was left at room

temperature for 10 min and then boiled for 15 min, and brought to room temperature in a water bath. 300  $\mu$ L of this was transferred to a flat bottom ELISA plate, and absorbance was measured at 480 nm. Galacturonic acid (Sigma) was used for a standard curve. The concentration of all the samples was adjusted to 415  $\mu$ g/mL for the immunodot assays.

### 2.2.3.3 Immunodots

Immunodots were done using three biological replicates, and three technical replicates, following the protocol described by Willats and Knox [82]. The results shown are representative of the three biological replicates. The pectin antibodies LM5, LM19, LM20, and 2F4 [64-66] were used for the assay. 1  $\mu$ L samples of the diluted polysaccharides (415  $\mu$ g/mL) were applied in triplicate to nitrocellulose membranes. The membranes were air dried for 30 min, and blocked overnight in PBS (phosphate-buffered saline pH 7.2) containing 5% milk powder. The membranes were then incubated for 2 h in primary antibody diluted 1 in 10, except 2F4, which was diluted 1:100, in PBS containing 1% milk powder. Membranes were washed extensively with municipal water and then with PBST (PBS containing 0.1% Tween 20). Following washes, the membranes were incubated for 1.5 h in secondary antibody (anti-rat horseradish peroxidase conjugate, GenScript) diluted 1:1,000 in MPBS (1%) for all the antibodies, except 2F4, which was incubated with anti-mouse IgG–alkaline phosphatase antibody (Sigma) diluted 1:25,000. The membranes were washed as above, and the anti-rat secondary antibody was detected with TMB substrate kit for peroxidase (Vector Labs. Cat. # SK-4400), while the anti-mouse secondary antibody was detected with Immun-Star™ AP Substrate Pack (Bio-Rad).

#### 2.2.3.4 Embedding, sectioning and labeling for fluorescent microscopy

Three biological replicates were analyzed. The plant segments SA, A, B, C, D and, E, described in section 2.2.1 were used. The 1 cm fragments were fixed under vacuum in 4% p-formaldehyde in PME buffer (50 mM PIPES, 5 mM MgSO<sub>4</sub>, 5 mM EGTA) and washed with PME buffer. Then, the fragments of segments A, B, C, D, and E were embedded in 8% agar, and 100 µm cross sections were obtained using a vibratome in a PME buffer bath. For the SA, the 1 cm section were wax-embedded in a Leica TP1020 tissue processor, and 10 µm sections were obtained at the shoot apical meristem, and at 100, 200, 300, 400, 500, 700 and 900 µm from the shoot apical meristem using a microtome (Leica RM2125 RTS).

For the antibody labeling, 100 µm sections were transferred to different wells in a flat bottom ELISA plate containing 100 µL of 5% BSA in TBST buffer (10mM Tris-HCl pH 7.0, 0.25 M NaCl, 0.1% Tween-20) and blocked for at least one hour. The 10 µm sections from the first centimeter were de-waxed with toluene and washed with decreased concentrations of ethanol, and finally distilled water. Then, they were blocked and treated on the slides as follows. Sections were then washed three times with TBST buffer and 30 µL of the different primary antibodies diluted in TBST buffer were added (LM5, LM19 and LM20 diluted 1:10), and incubated for 1 h at room temperature. The tissues were then washed three times in TBST buffer, and 30 µL of a 1:100 dilution in TBST buffer of anti-rat IgG FITC was added and this was incubated for 1 h at room temperature, then washed again three times in TBST buffer. The sections were mounted on slides and covered with Mowiol coverslip mounting solution, and were kept in dark and allowed to polymerize overnight, followed by observation on an Olympus BX51 microscope equipped with

epifluorescence irradiation (Filter excitation/emission 480 nm/535 nm). The contrast and brightness of the images was adjusted in the same way in all the images to facilitate the visualization. For the bright field images the color was also adjusted to enhance the contrast.

#### 2.2.3.5 Embedding, sectioning and labeling for transmission electron microscopy

Fragments of 1 cm length from positions B and D were fixed in 4% p-formaldehyde in PME buffer (50mM PIPES, 5mM MgSO<sub>4</sub>, 5mM EGTA) and washed with PME buffer. Then the sections were cut in three parts and were dehydrated and embedded as in His and collaborators [72]. Briefly, tissues were dehydrated in ethanol series of 10, 20, 35, 40, 50, 60, 70, 80, 90, and 100%, with incubation for 30 min at each step. Sections were washed twice with propylene oxide and infiltrated gradually with LR white resin in propylene oxide for 24 h each at 10, 30, 50, 70, and 90% and then 100% LR white resin for 30 days, with the resin replaced daily, and doing the first two infiltrations at 100% LR white under vacuum. An ultramicrotome with a diamond knife was used to obtain 90 nm thick cross sections, which were collected on 300-mesh coated gold grids.

Immunolabeling was done according to His and collaborators [72], with modifications in the wash stringency. The grids with the sections were placed on Parafilm M, and then were blocked for non-specific binding for 30 min using 30  $\mu$ L of 3% milk powder in PBST. Three washes (30  $\mu$ L of PBST), were followed by incubation overnight at 4 °C, with 30  $\mu$ L of the appropriate antibody (LM19, LM20 and LM5 diluted in PBST 1 in 10, or just PBST for the negative control). The grids were then washed thrice with 0.5% PBST and were subsequently incubated overnight at 4 °C in 30  $\mu$ L of a 1:100 dilution of the Anti-Rat IgG–Gold antibody (Sigma). The grids were placed in a Chien Staining pad

(Ted Pella Inc) and were jet washed with 5 mL of 0.5% PBST using a syringe with a 0.45 µm filter (Millipore), and were washed again, with 5 mL of Milli-Q water in a syringe with a 0.45µm filter (Millipore). The sections were then contrasted with 4% uranyl acetate for 20 min and with Reynolds' lead citrate for 7 min [83] and were washed with Milli-Q water. They were observed using a transmission electron microscope (Philips – FEI Model = Morgagni 268) operating at 80 kV.

#### 2.2.4 Calcium subcellular localization

The potassium antimonate technique was used to localize calcium in sections SA, A, B, C, D, and E, following the protocol of Slocum and Roux [84] with modifications by Plavcová and Hacke [85]. Three biological replicates were used for this experiment. The 1 cm sections were cut in three parts and fixed in 2% GA, 2.5% FA and 0.1% tannic acid in 0.1 M potassium phosphate buffer plus 2% potassium antimonate, pH 7.6, for 6 h at 21 °C in darkness. Sections were then washed with the potassium phosphate buffer plus antimonate twice for 15 min and postfixed in 1% OsO<sub>4</sub> in antimonate phosphate buffer, for 2 h. Then, three 10 min washes in potassium phosphate buffer plus antimonate were done, followed by washing for 30 min in 0.01M potassium phosphate buffer. The tissues were then dehydrated in an ethanol series and washed with propylene oxide and progressively embedded in Spurr resin and polymerized at 70 °C for 16 h. An ultramicrotome was used to obtain 90 nm thick cross sections, which were collected on 300-mesh copper grids. The sections were then contrasted with 4% uranyl acetate for 20 min and with Reynolds' lead citrate for 7 min [83] and washed with Milli-Q water, and observed using a transmission electron microscope (Philips – FEI Model = Morgagni 268 Operating at 80 kV).

## 2.3 Results

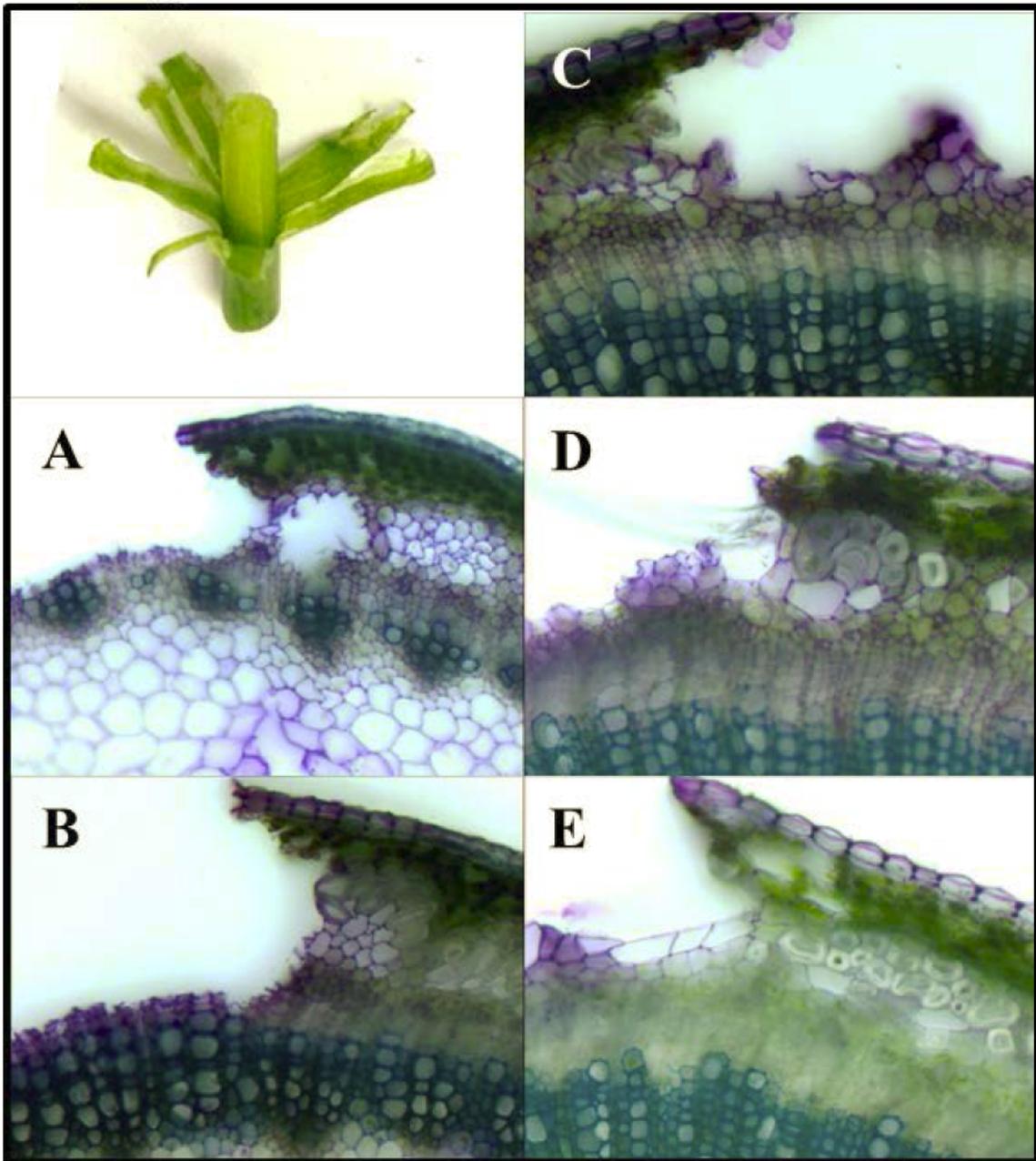
### 2.3.1 Tissues corresponding to the different stages of development

Gorshkova and collaborators [20] defined different stages of development of flax fibers relative to a mechanically defined “snap point” on the stem. In general, fiber specification and elongation occur apically to the snap point, and fiber cell wall thickening occurs basally. With this frame of reference, we examined the stem anatomy of linseed flax (variety CDC Bethune), in plants 46 to 48 cm long, ~5 weeks after germination, just before flowering. Based on our observations, and with reference to the precedent established by Gorshkova, we identified nine positions along the stem that represented progressive stages of fiber development (Figures 2-1 and 2-2). Five of these positions (points SA to A) were apical to the snap point, and four positions (B through E) were basal to the snap point. A 1 cm segment of whole stem was harvested at each of the nine positions. Additional 1 cm segments were obtained from positions A through E, and these were peeled to obtain only the outer tissues of the stem (epidermis, cortex, phloem, and some cambial zone cells), while excluding xylem (Figure 2-2). Segments SA to 3-4 were too delicate to effectively peel.



**Figure 2 - 1 Location in the flax stem of the tissues used for the experiments used in this study.**

Plants were 5 weeks old, in vegetative stage, and their height from the hypocotyls was between 46 to 48 cm. Tissues were harvested and rapidly placed into the fixative for microscopy, or were frozen with liquid nitrogen for all other applications.



**Figure 2 - 2 Cross sections of flax showing the tissues that were obtained when the stem was peeled at the different points used.**

Flax stems were manually peeled to obtain the stem peel, which included the tissues outside of the vascular cambium. Panels A to E represent the different points in the stem shown in Figure 2-1. Top left panel shows a piece of stem peeled halfway.

### 2.3.2 The PME activity is lower at the top of the plant

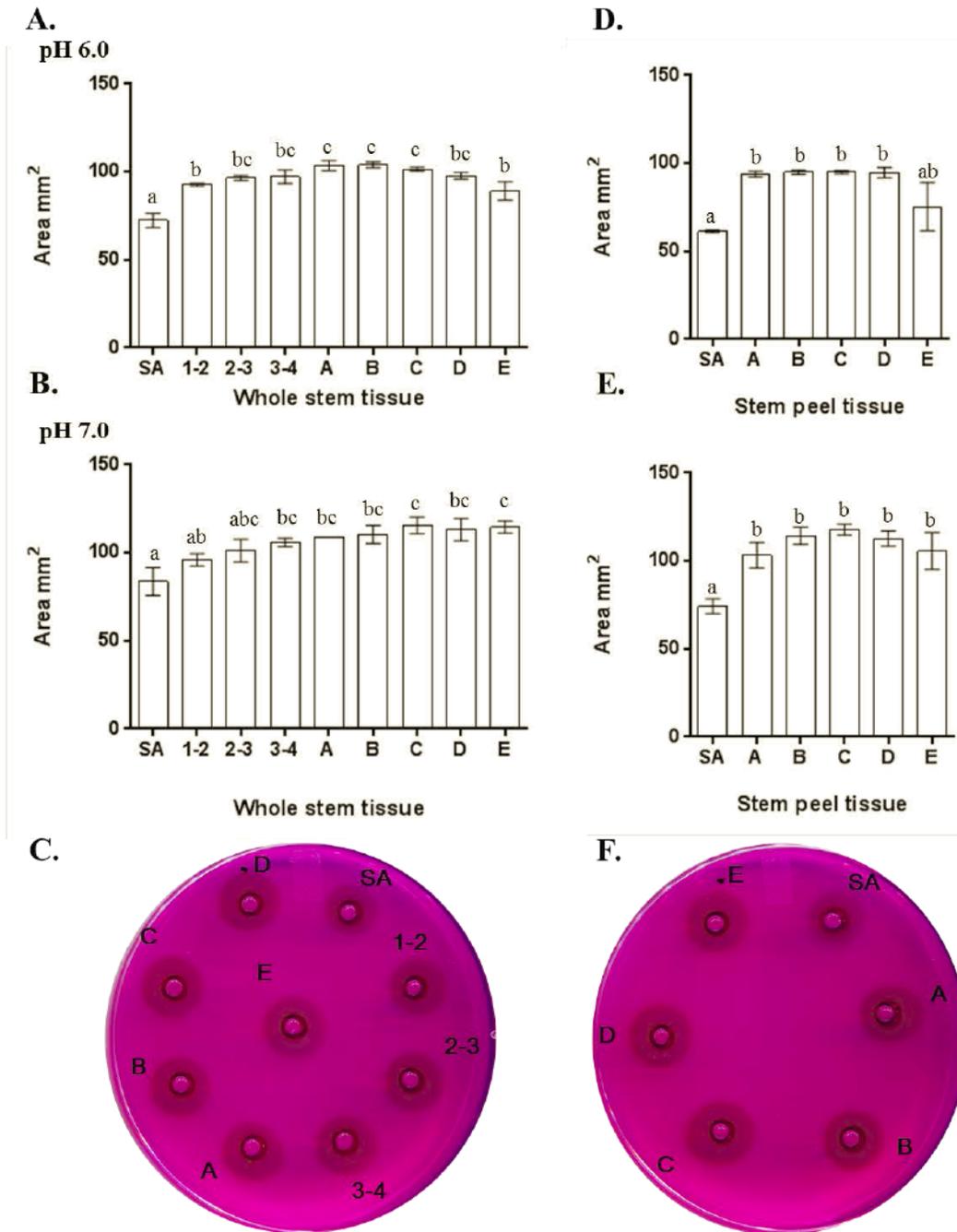
PME activity in the nine different segments of whole stem and five segments of stem peels was assessed, using three biological replicates, which were each measured in three technical replicates. In this assay, proteins extracted from stem segments were allowed to radially diffuse from a well into an agarose gel containing pectin and ruthenium red, and PME activity was detected by the development of a dark halo around the well.

Measurement of the area of the halo allowed for a semi-quantitative estimate of PME activity

We first established a standard curve to determine whether the area of the halo was directly proportional with the concentration of protein. We found a positive correlation with an R-square value of 0.9598 (Appendix, Figure A2-1), which supports the use of the radial assay to quantify the activity of the PMEs.

We assayed PME activity at both pH 7.0 and pH 6.0. These pH values were chosen based on the results of a pilot study of flax stem PME activity at pHs 5.0, 6.0, 7.0 and 8.0, which showed maximum activity at pH was 7.0 (data not shown). We also conducted the full assay at pH 6.0, since this is representative of the pH of the natural cell wall. In whole stem tissues at either pH 6.0 or 7.0, PME activity was significantly lower ( $p < 0.05$ ) at position SA relative to almost all other tissues (Figure 2-3, Table 2-1). This was also true in the stem peel, where PME activity was lower in SA than in any other tissues tested. Furthermore, the activity of SA (whole stem) was significantly lower ( $p < 0.05$ ) compared to the stages A to E of the stem peel tissues (Figure 2-3 D to F). Thus, PME activity as a proportion of the total proteins extracted) appeared to be highest in tissues below the

apical-most 1 cm of the stem, with a peak around position A, and was higher in stem peels than in whole stem.



**Figure 2 - 3 PME activity of native flax proteins detected as a decrease in methylesterification of highly esterified pectin measured by the binding of ruthenium red.**

Proteins extracted from whole stems or stem peels at different developmental stages (as defined in Figure 2-1) were assayed for PME activity using a radial diffusion assay. In this semi-quantitative assay, the area of the halo formed was proportional to PME activity. The bar graphs show results of an ANOVA followed by Tukey's multiple comparisons test from three technical replicates of each of three biological replicates. The plates show results of a representative diffusion assay. Panels A to C correspond to the activity of proteins extracted from the whole stem. Panels D to F correspond to proteins extracted from the stem peel, plus stage SA of whole stem, for comparison purposes.

Table 2 - 1 Tukey's multiple comparisons test of Pectin Methylesterase activity along the flax stem.

An ANOVA test was followed by a Tukey's multiple comparisons test using GraphPad Prism version 6.00 for Windows. The asterisks denote the p-value as follows. \* 0.01-0.05; \*\*0.001-0.01, \*\*\*0.0001-0.001; \*\*\*\*<0.0001. ns: non-significant difference (p>0.05)

Whole stem tissues			Stem Peel		
	pH 6	pH 7		pH 6	pH 7
SA vs. 1-2	****	ns	SA vs. A	****	****
SA vs. 2-3	****	ns	SA vs. B	****	****
SA vs. 3-4	****	**	SA vs. C	****	****
SA vs. A	****	**	SA vs. D	****	****
SA vs. B	****	***	SA vs. E	ns	****
SA vs. C	****	****	A vs. B	ns	ns
SA vs. D	****	***	A vs. C	ns	ns
SA vs. E	****	****	A vs. D	ns	ns
1-2 vs. 2-3	ns	ns	A vs. E	**	ns
1-2 vs. 3-4	ns	ns	B vs. C	ns	ns
1-2 vs. A	**	ns	B vs. D	ns	ns
1-2 vs. B	**	ns	B vs. E	**	ns
1-2 vs. C	*	*	C vs. D	ns	ns
1-2 vs. D	ns	ns	C vs. E	**	ns
1-2 vs. E	ns	*	D vs. E	**	ns
2-3 vs. 3-4	ns	ns			
2-3 vs. A	ns	ns			
2-3 vs. B	ns	ns			
2-3 vs. C	ns	ns			
2-3 vs. D	ns	ns			
2-3 vs. E	ns	ns			
3-4 vs. A	ns	ns			
3-4 vs. B	ns	ns			
3-4 vs. C	ns	ns			
3-4 vs. D	ns	ns			
3-4 vs. E	ns	ns			
A vs. B	ns	ns			
A vs. C	ns	ns			
A vs. D	ns	ns			
A vs. E	***	ns			
B vs. C	ns	ns			
B vs. D	ns	ns			
B vs. E	***	ns			
C vs. D	ns	ns			
C vs. E	**	ns			
D vs. E	ns	ns			

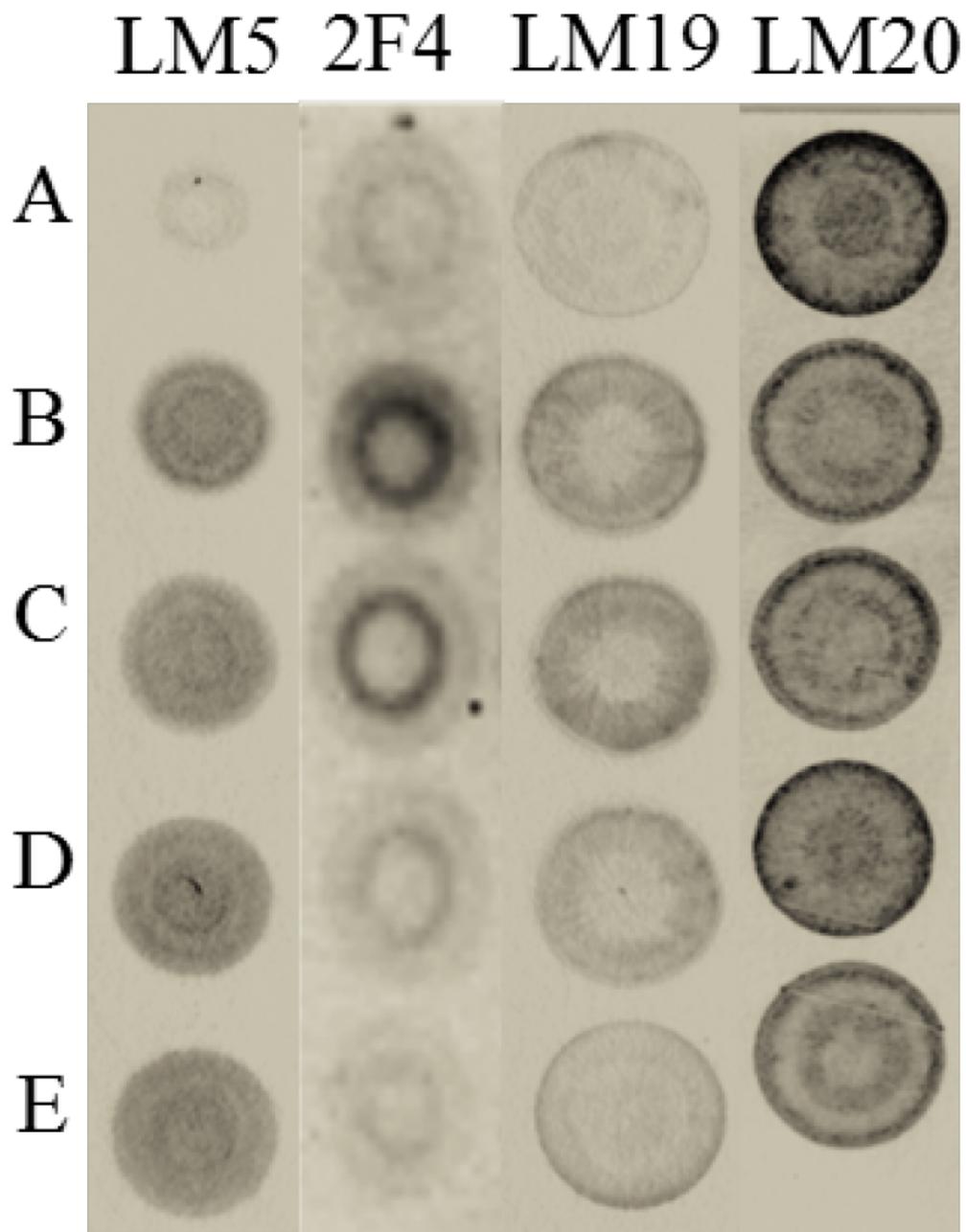
### 2.3.3 The degree of methylesterification and RG-I structure change along the stem

The degree of methylesterification in transverse stem sections was evaluated by immunohistochemistry of equal amounts of polysaccharides extracted from the cell wall. We used antibodies LM19, LM20, and 2F4 to assess the methylesterification of the HG residues, and LM5 antibody to assess the presence of 1,4-linked  $\beta$ -galactans (the galactosyl containing chains of RG-I), in the fiber cell walls before and during thickening.

#### 2.3.3.1 Methylesterification state at the different developmental stages

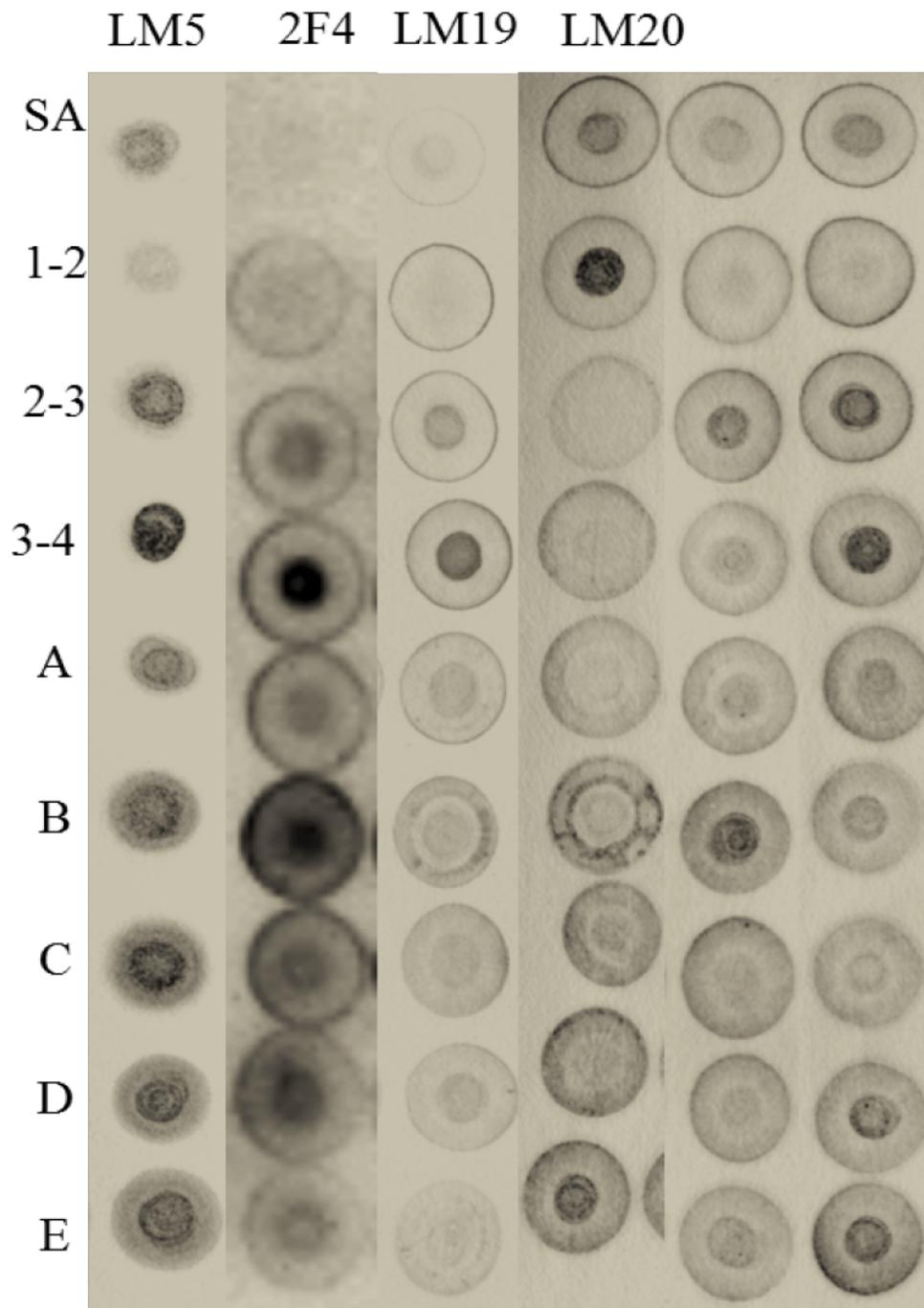
We assessed the methylesterification and calcium cross linking state of the fibers using immunodot assays of cell wall-extracted polysaccharides, from whole stem and stem peel tissues.

We also assayed the binding of 2F4 [66] to detect calcium cross-linked HG, which is indicative of rigidification of the cell wall. In both whole stems and stem peels, we found that 2F4 had very weak hybridization in the shoot apex and in mature fibers. Weak labeling was observed from points SA to A in the whole stem and in point A in the stem peel. Then, a peak in 2F4 binding was observed below the snap point, B, where secondary cell wall deposition had started, and then decreased towards E (Figure 2-4, Figure 2-5).



**Figure 2 - 4 Immunodot of flax stem peel tissues polysaccharides from different stages of fiber development.**

1  $\mu$ L samples of same amounts of polysaccharides were applied in triplicate to nitrocellulose membranes and labeled with the corresponding antibodies. The results shown are representative of the three biological replicates.



**Figure 2 - 5 Immunodot of flax whole stem tissues polysaccharides from different stages of fiber development.**

1  $\mu$ L samples of same amounts of polysaccharides were applied in triplicate to nitrocellulose membranes and labeled with the corresponding antibodies. The results shown are representative of the three biological replicates. For LM20 the three replicates are shown, as no consensus pattern was observed

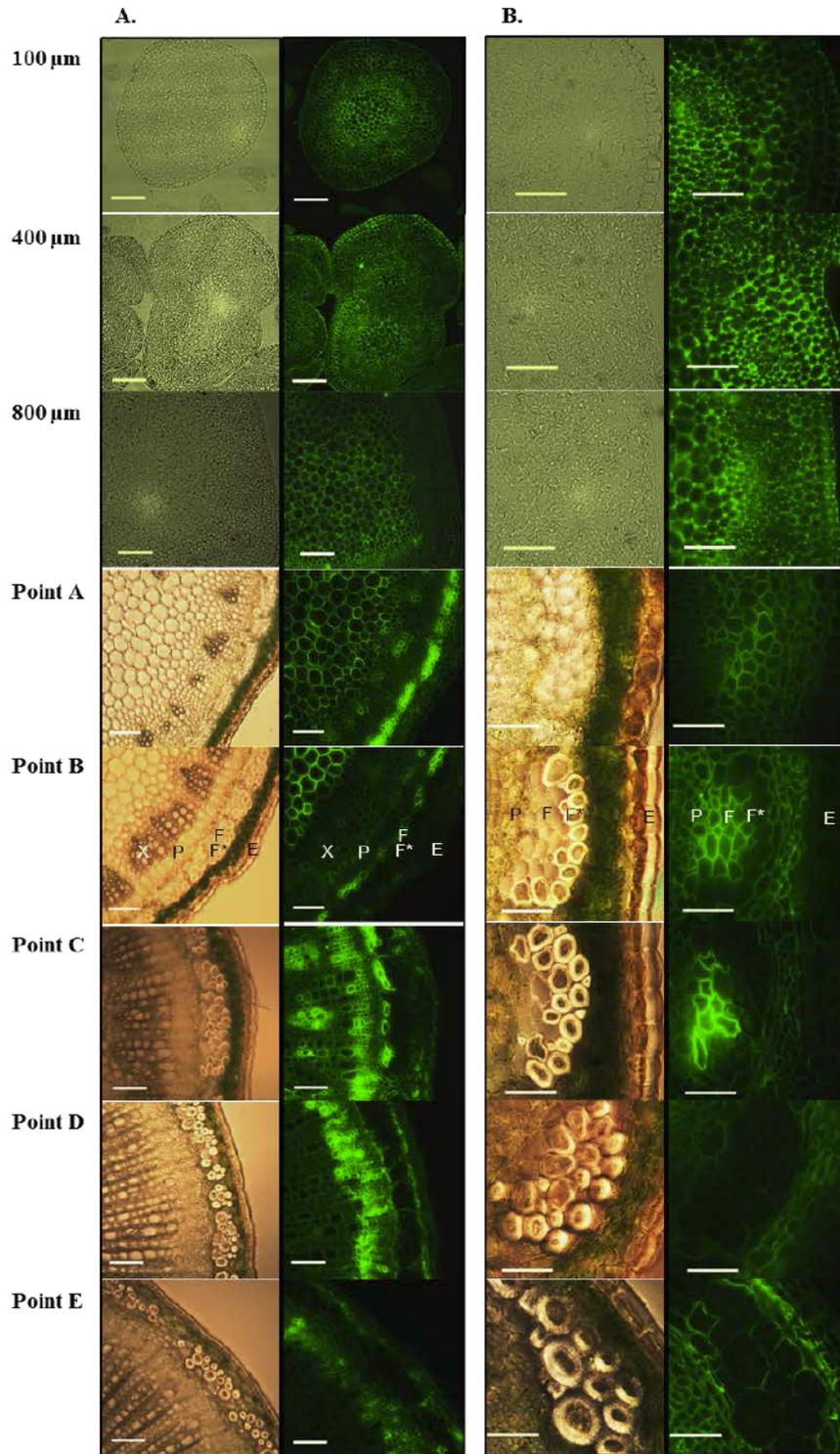
LM19, which recognizes demethylesterified HG [64], had a similar binding pattern to the one observed with 2F4 in the immunodot assay (Figure 2-4, Figure 2-5), which is consistent with the expectation that blockwise demethylesterification is associated with calcium bridging. LM20, which binds to highly methylesterified homogalacturonan [64], had an overall even distribution in the stem peel tissues. However, there was not a defined pattern along the stem in whole stem tissues (Figure 2-5).

To further analyze the distribution of pectic epitopes recognized by LM19 and LM20 antibodies during stem and fiber development, we performed immunohistochemical staining in conjunction with epifluorescent microscopy, and immunogold localization using TEM.

For epifluorescence, we examined transverse sections of whole stems at 14 positions along the primary vegetative stem. Nine of these positions were within the apical-most 1000  $\mu\text{m}$ , and the remaining five points were located basal to these, as already defined (positions A, B, C, D, and E; Figure 2-2).

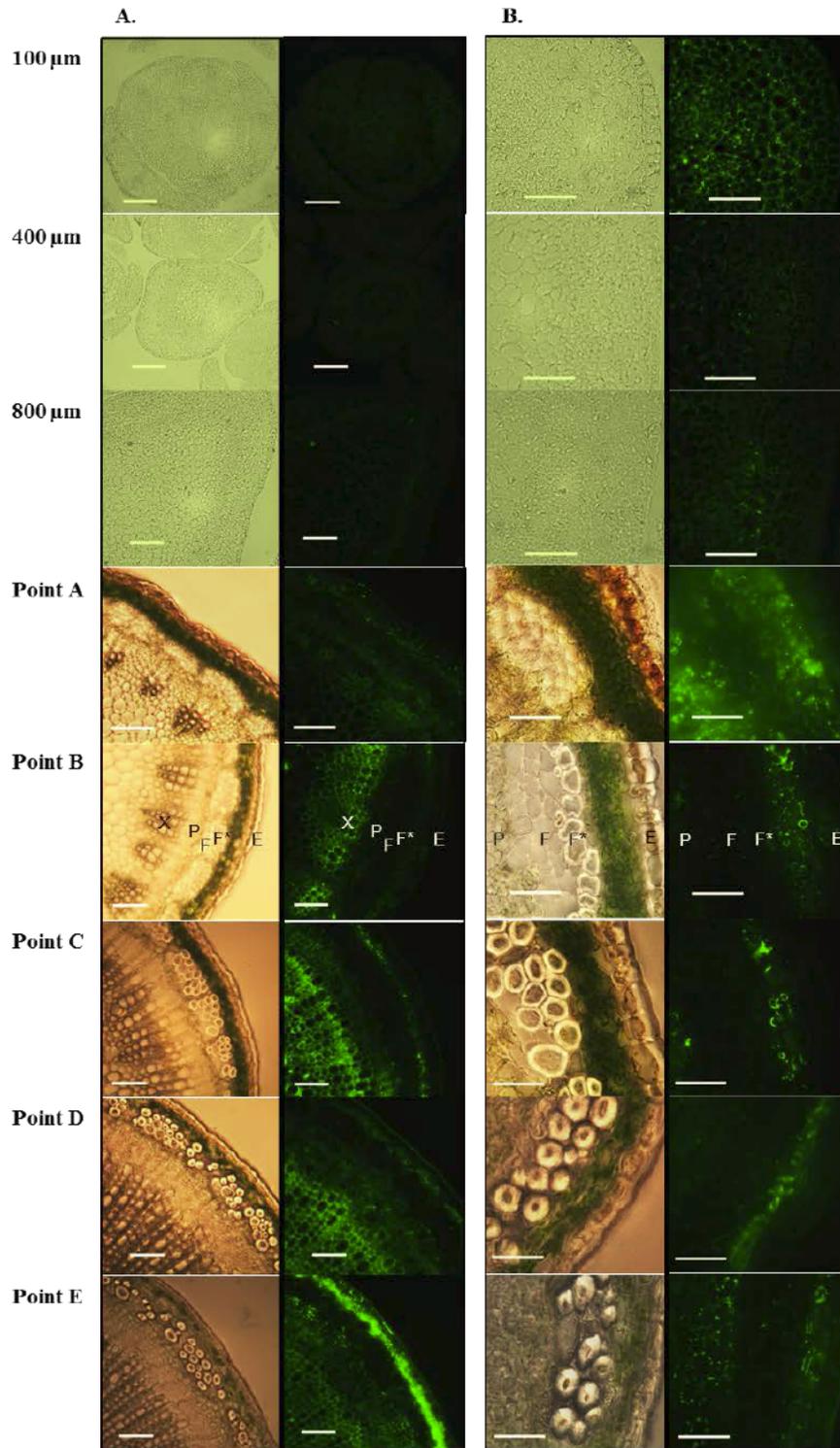
As shown in Figure 2-6, LM19 labeled primary walls of almost all cells within the apical-most 1000  $\mu\text{m}$  of the flax stem. However, starting at position A, LM19 labeled phloem fibers and components of the xylem with more intensity than surrounding cells. Within each phloem fiber, LM19 labeling diminished as secondary walls became visible, and this loss of labeling happened first with the outermost fibers (i.e. fibers closest to the epidermis). In summary, starting at point A, LM19 labeling of phloem fibers was notably more intense than in surrounding cells, meaning that high amounts of unesterified HG

were present. However, very little calcium cross-linked HG was detected in this region, according to the 2F4 immunodot results (Figures 2-4 and 2-5).



**Figure 2 - 6 Epifluorescence labeling with LM19 antibody of flax stem cross sections at different stages of fiber development.**

Same specimens observed using a 20X (A) or 60X (B) objective lens. Bars: A. 100  $\mu$ m; B: 50  $\mu$ m. The images on the left on each panel are bright field images and on the right the corresponding image under fluorescence. E: Epidermis; F: Fiber without secondary cell wall; F\*: Fiber with secondary cell wall; P: Phloem X: Xylem.



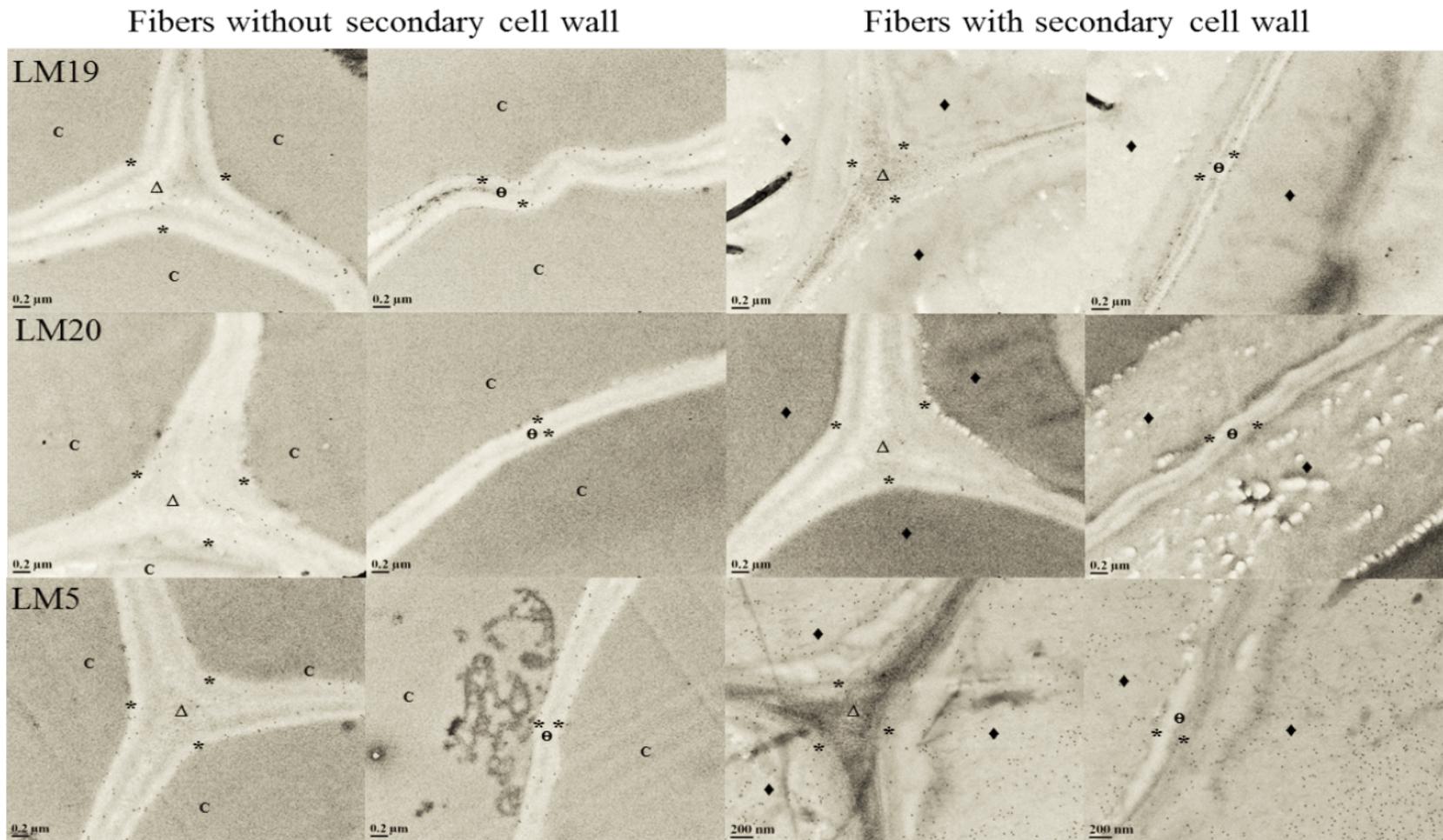
**Figure 2 - 7 Epifluorescence labeling with LM20 antibody of flax stem cross sections at different stages of fiber development.**

Same specimens observed using a 20X (A) or 60X (B) objective lens. Bars: A. 100  $\mu\text{m}$ ; B: 50  $\mu\text{m}$ . The images on the left on each panel are bright field images and on the right the corresponding image under fluorescence. E: Epidermis; F: Fiber without secondary cell wall; F\*: Fiber with secondary cell wall; P: Phloem X: Xylem.

Both above and below the snap-point, LM20 (high methylesterification) only weakly labelled the phloem fibers compared to surrounding cells (Figure 2-7). Below the snap point, but not above at point A, LM20 labeling was detected in the cell corners of fibers that lacked a secondary cell wall. Secondary xylem, in contrast, was intensely labeled by LM 20 at most stages of development,

To investigate the localization of the LM19 and LM20 epitopes at higher resolution, we used immunogold labeling of LR white-embedded tissues at stages B and D. These positions were selected because they showed a range of fiber developmental stages that were differentially labeled by LM19 (Figure 2-6). The TEM showed that LM19 labeled the primary wall and middle lamella at both stages, with the most intense labeling in the cell corners (Figure 2-8). Furthermore, we observed that LM19 epitopes were more abundant in the cell wall and middle lamella between immature fibers (stage B), than between thickening fibers (i.e. fibers with a secondary cell wall, stage B and D). In both stages, the secondary cell wall was not labeled by LM19.

Immunogold labeling with LM20 showed that it bound to the cell corners, with higher abundance in the corners between young fibers than between thickening fibers. No labeling was detected in the compound middle lamella (CML) or the primary cell wall interfaces between fibers at either stage B or D (Figure 2-8).



**Figure 2 - 8 Transmission electron microscopy of fibers at stages B and D labeled by immunogold.**

The images showing the fibers without secondary cell wall are from point B. The images showing fibers with secondary cell wall are from either point B or D. c: Cytoplasm; \*: Primary cell wall; ◆: Secondary cell wall; ϑ: Compound middle lamella; Δ: Cell corner middle lamella

### 2.3.3.2 (1-4)- $\beta$ -D-galactans are present in higher amounts in the fibers since coordinated growth starts.

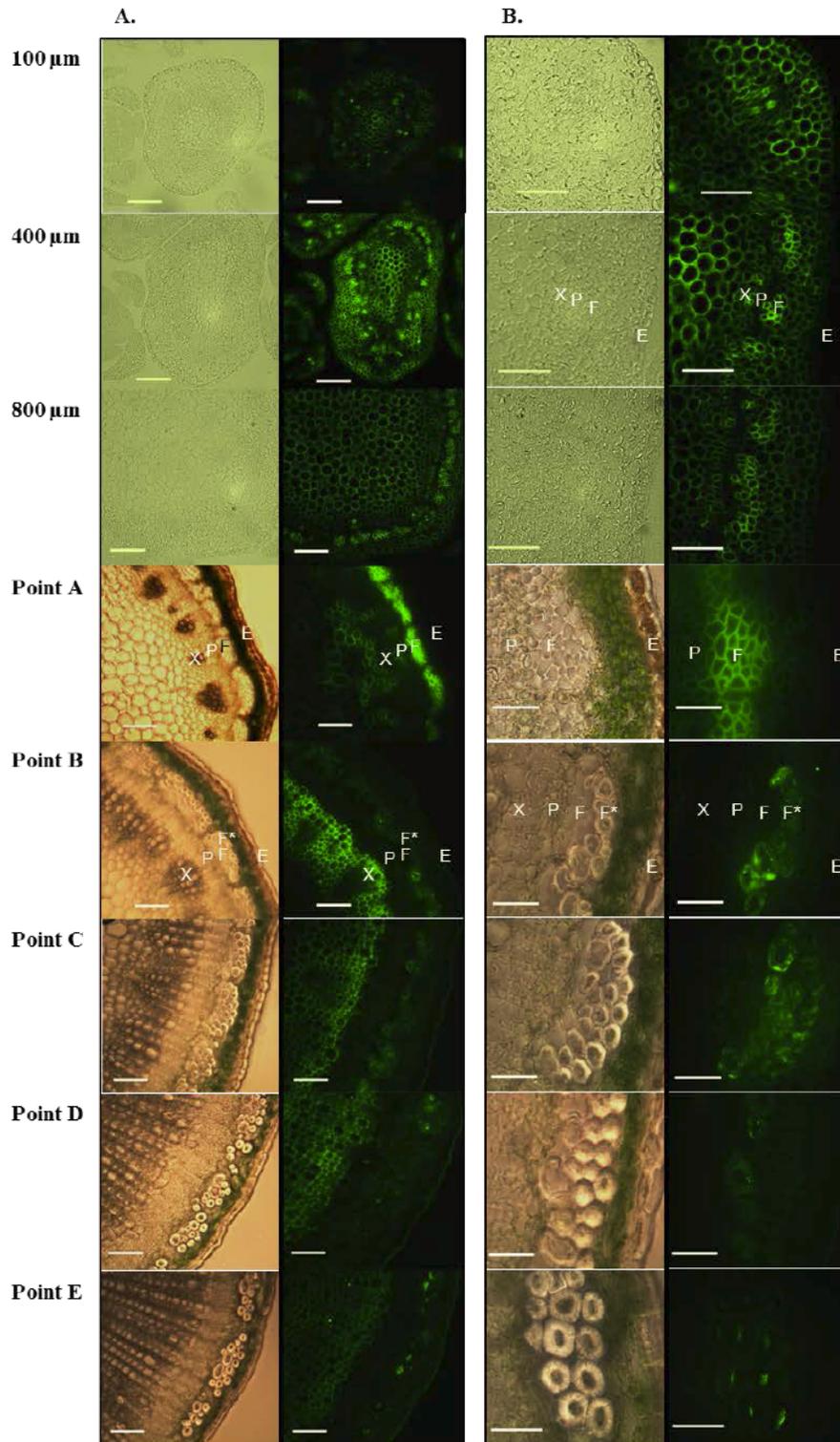
Because it has been found that galactans play an important role during fiber maturation [15, 16, 70], and as they had not been studied in detail above the snap point, we used an immunodot assay to investigate the distribution of galactan epitopes in developing stem and fibers using the LM5 antibody, which recognizes a linear tetrasaccharide in (1-4)- $\beta$ -D-galactans [65]. We observed two patterns of immunodot morphology. In polysaccharides extracted from above the snap point, LM5 stained a compact dot. In contrast, for extracts obtained below the snap-point, LM5 stained a much more dispersed dot (Figure 2-4 and 2-5).

When LM5 was used to label transverse stem sections, we observed distinct labeling of the primary cell wall of fibers, but not the middle lamella (Figure 2-9) starting at 400  $\mu$ m from the apex to position A. This was different from LM19 and LM20, which showed no specific labeling of fibers in the apical-most 1000  $\mu$ m of the stem. Then, when secondary cell wall deposition had started in the bundle, starting with the fibers at the outer periphery (point B), the presence of the LM5 epitope in the fibers with only primary cell walls drastically decreased in comparison with point A (Figure 2-9), and the epitope was rather detected in the fibers that were thickening, as was observed from point B to E. Also, in point E, the labeling of LM5 was higher in the internal part of the secondary cell wall, the Gn-layer (Figure 2-9), than in the more mature, G-layer. These findings were consistent also with results from immunogold labeling (Figure 2-8). Thus in summary, there was intense labeling by LM5 at the primary cell wall of the fibers during their

coordinate and intrusive elongation, which drastically diminishes when elongation ceases, but increases again in the secondary cell wall of the thickening fibers.

#### 2.3.3.3 Calcium subcellular localization

As calcium is involved in the cross linking between HG domains, we studied the subcellular localization of calcium using potassium antimonate [84, 85]. Higher amounts of calcium in the cell junctions of fibers were observed below the snap point (Figure 2-10). In stages SA and A, there were lower amounts of calcium, compared to B through E. Meanwhile the presence of calcium in the compound middle lamella and primary cell wall between the fibers was constant along the stem, in fibers with and without secondary cell wall. The calcium was mainly present in the primary cell wall and middle lamella, and it was particularly abundant in the cell corners, where higher amounts were detected in the middle lamella. Also, some accumulations of calcium were observed at the plasma membrane of thickening fibers, in stages B through E (Figure 2-10).



**Figure 2 - 9 Epifluorescence labeling with LM5 antibody of flax stem cross sections at different stages of fiber development.**

Same specimens observed using a 20X (A) or 60X (B) objective lens. Bars: A. 100  $\mu$ m; B: 50  $\mu$ m. The images on the left on each panel are bright field images and on the right the corresponding image under fluorescence. E: Epidermis; F: Fiber without secondary cell wall; F\*: Fiber with secondary cell wall; P: Phloem X: Xylem.

a.

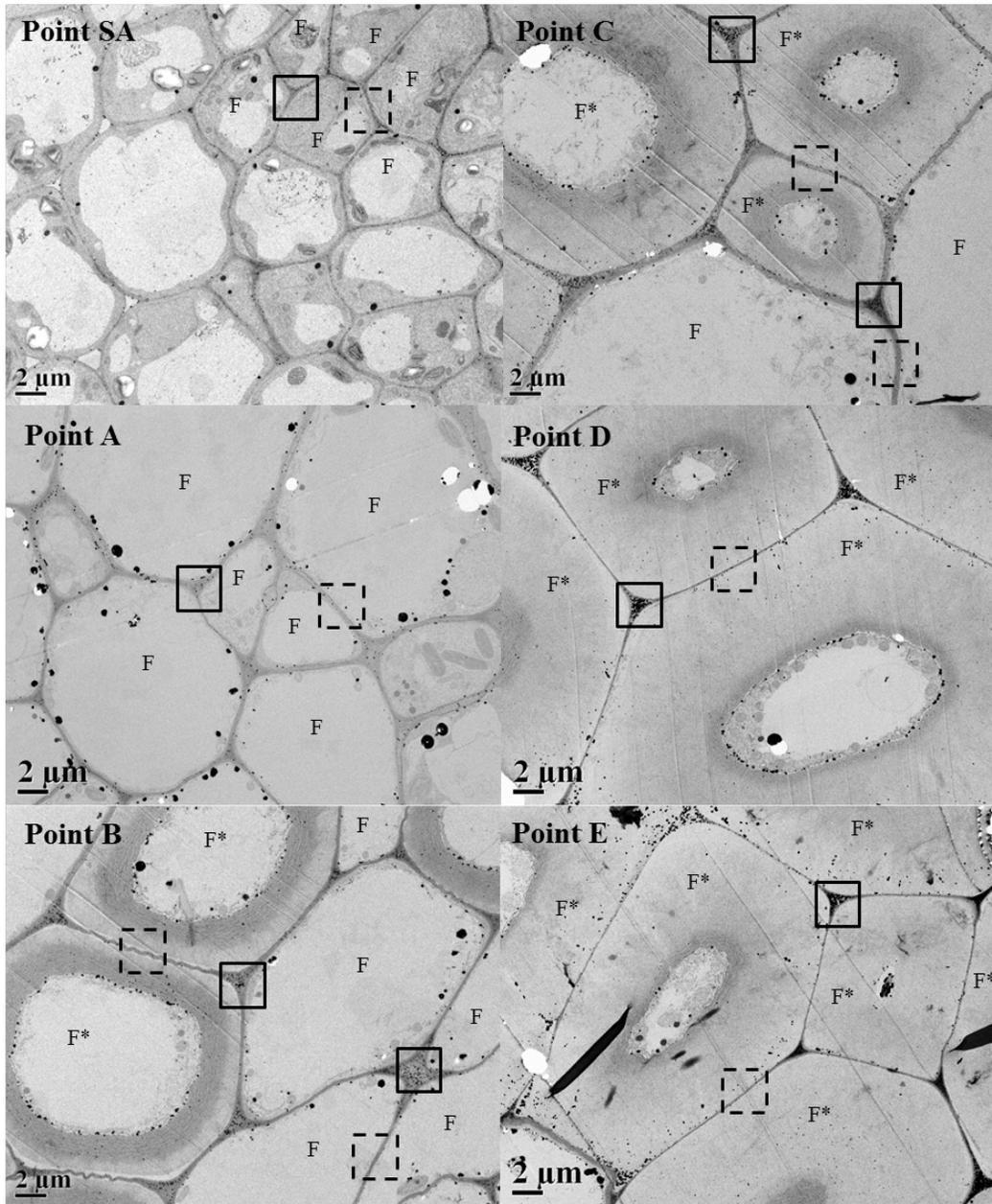
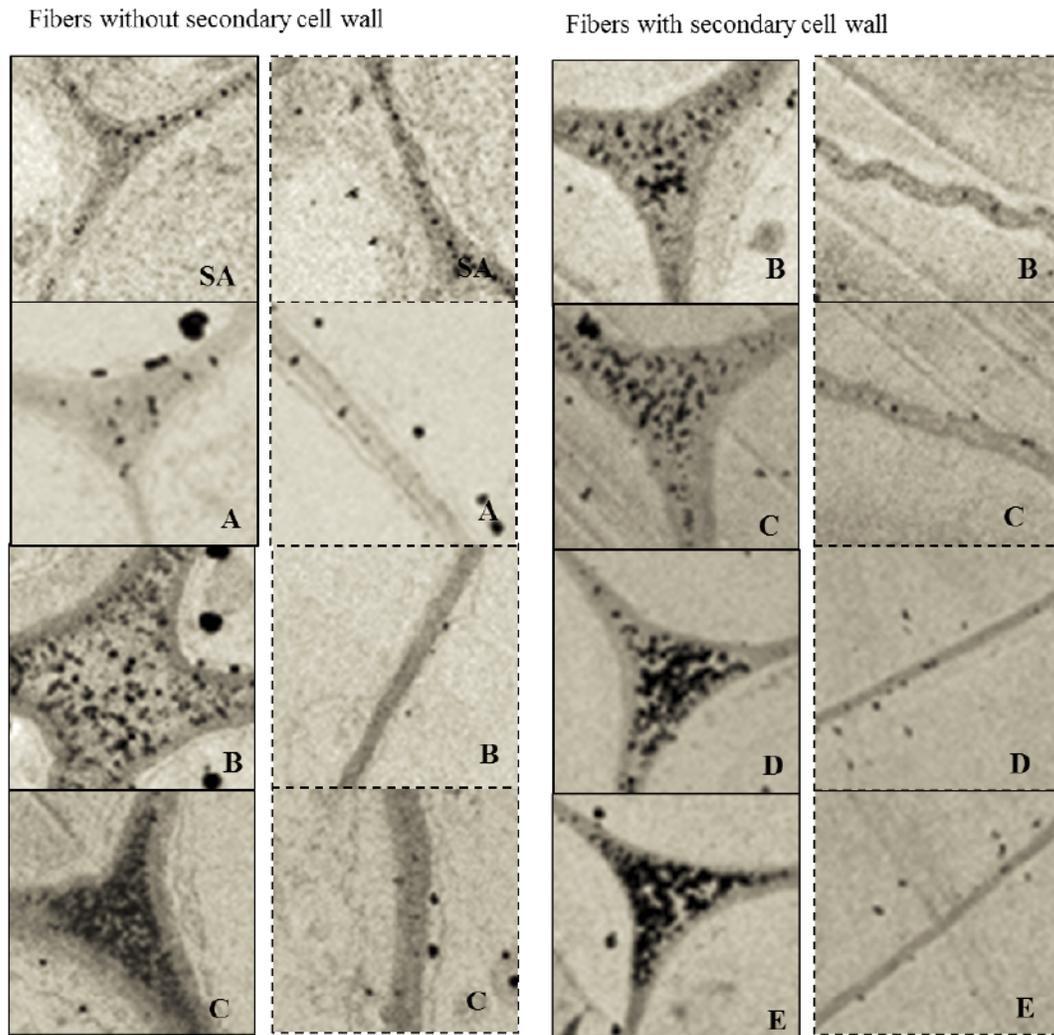


Figure 2 - 10 Calcium subcellular localization in flax fibers along the stem using potassium antimonate. Continued on next page.

**b.**



**Figure 2 - 10 Calcium subcellular localization in flax fibers along the stem using potassium antimonate. Continued.**

Tissues were treated with potassium antimonate and embedded in Spurr resin and observed using transmission electron microscopy. Image in panel b shows the zoom up of the sections indicated in panel a. F: Fiber without secondary cell wall. F\*: Fiber with secondary cell wall.

## 2.4 Discussion

### 2.4.1 Methylesterification state of the fibers

The LM19 antibody, which binds demethylesterified pectin, labeled more intensely than surrounding tissues the primary cell wall and middle lamella of phloem fibers above the snap point, and below the snap point in the non-thickening fibers (Figure 2-6). Once secondary cell wall thickening could be detected, the binding was drastically reduced, which was consistent with other observations that calcium cross linking diminished from point B to E, as was observed with the immunodot assay (Figures 2-4 and 2-5).

Meanwhile, the presence of the LM20 epitope was barely detectable in fibers in any stage of development (Figure 2-7), meaning that once the HG was deposited in the wall, it was rapidly demethylesterified to such a level that the amounts of highly methylesterified HG, detected by LM20, were very low. This is in agreement with the findings of Snegireva and collaborators [19], who found that the presence of the LM19 epitope (i.e. low methylesterification) in the fibers is higher in intrusively growing cells relative to the rest of the tissues, and the LM20 epitope is barely visible in all the cells above the snap point.

### 2.4.2 Methylesterification and calcium cross linking

As stated above, according to immunohistochemistry, the primary cell wall and middle lamella of non-thickening fibers had relatively low methylesterification. We also found, as seen on the immunodot assay with 2F4, that the calcium cross linking of HG drastically increased in point B, below the snap point (Figures 2-4 and 2-5).

Consequently, because calcium cross-linking plays a major role in the bonding of pectins within the cell wall and in adhesion with surrounding cells [86], we can interpret these

results in two non-exclusive ways. First, adhesion of fibers to adjacent cells, via the middle lamella, is expected to begin when elongation has stopped (i.e. below the snap point), and this would be associated with increased calcium cross linking observed, and may result in rigidification of the middle lamella between fibers and surrounding tissues. The calcium subcellular localization results allowed us to confirm the presence of higher amounts of calcium in fibers below the snap point (Figure 2-10). The vast majority of calcium was observed in the cell corners, specifically in the middle lamella. This supports the interpretation that the higher presence of calcium leads to an establishment of the connections between the fibers once they stop elongating. Second, the other possible interpretation is that because the fibers need to maintain their morphology once they stop growing, and as the galactan of the primary cell walls appears to be degraded, strengthening the cell wall by calcium cross linking may be necessary. Similar stiffening occurs in the sub-apical regions of growing pollen tube (which grows by intrusive tip growth) demonstrated with mutants of PMEs in Arabidopsis. Jiang and collaborators [48] found that the mutant of the gene *VANGUARD1*, a PME gene, led to a reduction of the cell wall strength in the pollen tube, based on the morphology of the pollen tube in the mutant compared to the wild type, this was hypothesized to be due to a decrease in the formation of calcium cross linking, based on the blockwise mode of action expected for this PME. Similar results with a similar approach were found for the Arabidopsis PME gene *AtPPME1* [54]. The thickening of the secondary wall basal to point B may be sufficient to support the wall from pressure from surrounding tissues so that rigidification by calcium is no longer required, which could explain the decreased in 2F4 binding basal to point B.

### 2.4.3 Presence of galactan in developing fibers

Based on the mobility patterns observed in the immunodot [82], two kinds of RG-I molecules were present in these tissues (Figures 2-4 and 2-5). Above the snap point, a RG-I with long branches, which reduces the dispersion of the pectins in the nitrocellulose so only a central dot is observed, and below the snap point a RG-I with shorter side chains allowing more mobility in the nitrocellulose, which coincides with the description of the soluble galactan that is deposited in the secondary cell wall of the fibers, below the snap point, in which most of the side chains are short [12, 15, 16]. The presence of RG-I with long chains above the snap point may lead to the generations of pores in the cell wall to allow the ingress or exit of cell wall modifying proteins, such as polygalacturonases, or expansins [87], necessary for the fiber expansion. More importantly, the presence of long side chains in RG-I has been hypothesized to prevent the formation of calcium cross linking [88], preventing the rigidification of the cell wall, which could be detrimental for fiber elongation during intrusive growth [25, 74, 75].

The discovery of a differential labelling of fibers by LM5 as close as 400  $\mu\text{m}$  to the shoot apex constitutes a useful marker of early fiber identity (Figure 2-9). This position is approximately correlated with the onset of coordinated growth [18]. The presence of the LM5 epitope in the growing fibers was maintained during intrusive growth (point A), but disappeared below the snap point in non-thickening fibers and appeared in the secondary cell wall of thickening fiber, which is the soluble galactan previously reported [16, 69]. As previously reported [16, 89], as the fibers matured the galactan in the fibers secondary cell wall was degraded and was more visible in the internal parts of the secondary cell wall. Together this means that the presence of the galactan residues has an effect on flax

fiber elongation. A similar conclusion was obtained when an exponential increase of LM5 (1-4)- $\beta$ -galactan was observed when cultured carrot cells were induced from proliferation to elongation [87], and also with the findings of Stolle-Smits and collaborators [37], who determined that during cell expansion in the pods of *Phaseolus vulgaris* the amount of RG-I side chains, containing galactosyl and arabinosyl residues, increased.

#### 2.4.4 Model for implication of the pectin polysaccharides in the development of fibers in flax

Taking together these results, we propose that the presence of (1-4)- $\beta$ -galactan might be important for the onset and maintenance of the fibers' coordinate and intrusive elongation, and for control of cell wall rigidity during growth. The fibers have lower levels of methylesterification than surrounding tissues in the middle lamella, and especially in the primary cell wall, however, despite the fact that demethylesterification is a pre-requisite for calcium cross linking, relatively low levels of calcium cross linking were detected above the snap point, which might be explained by the presence of the long side chains in the RG-I in the primary cell wall of the fibers, which impede the formation of calcium cross linked homogalacturonan [88], and prevent low methylesterification from limiting elongation of growth [25, 74, 75]. The reason why the methylesterification diminished from SA to point A (observed as an increment in the labeling of LM19 in the immunodot), could be that as the fibers grew intrusively, the number of fibers per segment increased [20], so a higher number of cells in the given tissue were demethylesterified. This increase in demethylesterification, which followed an increasing basipetal gradient away from the shoot apex, was also observed with the PME activity radial assay.

## 2.5 Conclusion

In conclusion, a lower methylesterification in the elongating fibers might aid in a controlled rigidification, by the presence of the (1-4)- $\beta$ -galactan side chains, of the cell wall of the fibers, avoiding excessive compression from surrounding tissues and the fibers in the bundle, and at the same time conferring sufficient flexibility to the cell wall. The high amounts of (1-4)- $\beta$ -galactan observed in the primary cell wall of the fibers above the snap point drastically diminished below the snap point, where fiber elongation had stopped. This could allow the generation of calcium cross linking at the middle lamella, strengthening the adhesion between cells, and at the primary cell wall rigidifying the cell wall. Further genetic and biochemical analyses could be used to test this hypothesis.

### **3 Chapter 3: Characterization and transcript profiling of the pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) gene families in flax (*Linum usitatissimum*).**

A version of this chapter has been published: Pinzon-Latorre and Deyholos, BMC Genomics (2013) 14:742.

#### **3.1 Introduction**

Pectins are complex polysaccharides present in the plant cell wall and in the middle lamella and are dynamically modified by pectin methylesterases (PMEs). The PME gene family was first described by Richard and collaborators [38], and later classified in the Carbohydrate Active Enzymes database (CAZy) as class 8 of the carbohydrate esterases (EC 3.1.1.11) [40]. In current models, pectins are synthesized in the Golgi complex as highly methylesterified polymers (e.g. homogalacturonan, HG) that are secreted to the cell wall. Once in the cell wall, PMEs catalyze the demethylesterification of HG, which generates negatively charged carboxyl groups. If demethylesterification occurs on contiguous sugar residues (i.e. blockwise demethylesterification),  $\text{Ca}^{2+}$  bonds can form between pectin molecules, thereby rigidifying the cell wall. Conversely, if the demethylesterification occurs on non-contiguous sugars (i.e. random demethylesterification), the molecule becomes a substrate for pectin degrading enzymes, leading to cell wall loosening [42, 90]. The activity of the PMEs is regulated by pectin methylesterase inhibitors (PMEIs) [44], which bind to the active site of the PME, generating a 1:1 complex [46, 47]. PMEs are classified as either Type-1 PMEs (i.e. those

with a pro-region, similar to the PME1 domain), or Type-2 PMEs (no pro-region). In Type-1 PMEs, the pro-region and the PME domain are translated as part of the same protein and then, in the Golgi complex, as a pre-requisite for secretion to the cell wall, the pro region is removed by a subtilisin-like protease [91].

The bast (phloem) fibers of flax (*Linum usitatissimum*) are valued industrially for their length and strength. Extraction of high quality fibers requires retting, a process by which stems are exposed to the action of microbes that degrade the middle lamella and so, facilitate separation of fibers from surrounding tissues. Flax fibers grow from the shoot apex intrusively after a very short period of coordinated growth [18]. During intrusive elongation, fibers first penetrate the middle lamella between adjacent cells, and subsequently generate new contact interfaces. Both of these processes presumably influence fiber length and the efficiency of retting, and are dependent on the activity of PMEs.

Different varieties of flax are grown for either fibers or for seeds (i.e. linseed) [4]. Although stems of linseed varieties contain fiber, these fibers are not harvested, because of relatively low fiber yield and the difficulty of retting in the environments where linseed is typically grown. A better understanding of PMEs is therefore important to the development of dual-purpose flax, in which both fibers and seeds can be utilized from a single variety.

Three PMEs have been previously characterized in flax: *LuPME1*, *LuPME3*, and *LuPME5* [50-53]. These are all Type-1 PMEs. Al-Qsous and collaborators [52] reported that in 2 dpg (days post germination) hypocotyls, transcript abundance of *LuPME5* is

higher in the apical region, while *LuPME3* transcript abundance is higher in the basal region. Also, *LuPME5* has the highest transcript abundance of the three characterized LuPMEs in hypocotyls. The highest transcript abundance of *LuPME3* in seedlings is in the roots [50]. Here, we expand on these studies and present an analysis of the complete family of PME and PMEIs in flax, based on the recently assembled whole genome sequence of the linseed variety CDC Bethune [1]. A specific objective of this research is to identify PMEs that are expressed during stages of fiber development that are likely to influence the industrially relevant properties of flax bast fibers.

## **3.2 Materials and methods**

### **3.2.1 Annotation of PME and PMEI domain in flax and other species**

Predicted proteins that contained PME (PF01095) and/or PMEI (PF04043) domains were identified from the whole genome shotgun (WGS) assembly of flax [1] (version 1.0) using default parameters in hmmsearch/PfamScan [92]. The predicted proteins from the flax WGS assembly were also aligned to previously described PMEs and PMEIs from Arabidopsis obtained from TAIR [93], using BLASTp. All of the LuPMEs/LuPMEIs that were identified by BLASTp to Arabidopsis were also identified by HMM-alignment to the PFAM domains. Predicted flax proteins that had both a PMEI and PME domain were designated Type-1 PMEs, and proteins with a PME domain (but no PMEI domain) were designated Type-2 PMEs. Genes with questionable PFAM annotations (i.e. significant PME and/or PMEI domain but low e-value; low coverage of the domain; more than one PME or PMEI domain; an extra domain different from PME or PMEI), and genes that were adjacent on scaffolds of the WGS assembly were manually curated, which included

reanalysis of their predicted gene structures by submitting their genomic sequence (i.e. the predicted gene plus 1000 bp up and downstream) to the Augustus web server [94]. The Augustus Arabidopsis gene model parameters were used for the gene reannotation, in combination with any ESTs that aligned to the prediction region (95% identity and 90% coverage) [95] as well as unpublished RNAseq reads (<http://www.onekp.com/>, version April 25 2013).

### 3.2.2 PMEs and PMEIs in other plants

An hmmsearch using PFAM domains PF01095 (Pectinesterase) and PF04043 (PMEI) was conducted with default parameters on transcripts deposited in Phytozome (version 9.1). To determine the statistical significance of the presence of the PME domain, all the protein sequences that had the domain were retrieved, and these were searched again against PFAM using batch search. For the putative PMEIs, protein sequences that had a PME domain but not a PME domain were obtained and then searched again on PFAM to establish the statistical significance of the predicted domains and confirm the absence of a PME domain.

### 3.2.3 Primer design for qRT-PCR

The Universal ProbeLibrary Assay Design Center (Roche) was used to design specific primers and probes for each gene. Groups of 10 closely related genes were submitted in batches for the design of specific primers and Roche UPL probes. The specificity of primers was evaluated by BLASTn alignment of the primers against the complete predicted transcriptome and the entire genome assembly. All primer pairs were designed so at least one primer of each pair had three or more mismatches to any off-target gene, near the 3' of the primer. For those genes for which a specific primer could not be

designed, a primer common to two PME or PME1 genes was used. The list of primers can be accessed in Appendix Table A3-1.

#### 3.2.4 Tissues for Quantitative Real Time PCR using a 96.96 dynamic array

RNA was obtained from 12 different tissues from three biological replicates. Each biological replicate was assayed independently. Five of the tissues/organs (shoot apex (SA), leaves (L), roots (R), early cortical peel (ECP), and early fibers (EF)) were collected from vegetative stage plants 1 month after germination; the other seven tissues (senescent leaves (SL), xylem (X), late cortical peel (LCP), late fibers (LF), flower buds (FB), flowers (F), and green bolls (B)) were collected from plants 2 months after germination, at the green capsule stage. The cortical peel, xylem, and fiber tissues were obtained from the first 15 cm of the plant from the hypocotyls to the top. The shoot apex tissue corresponds to the top 2 cm of the plant. A phenol/chloroform based method was used for extraction of RNA, with subsequent treatment with DNase. 5 ng of RNA were used to synthesize the cDNA for the 96.96 dynamic array (Fluidigm Corporation, CA, USA). The cDNA was tested for genomic DNA contamination by PCR using a set of primers flanking an intron.

A total of 12.5 ng of cDNA were used for the pre-amplification reaction containing 50 nm of each primer pair (a pool consisting of 89 primers for the PME and/or PME1 genes plus 3 endogenous controls GAPDH, ETIF1, ETIF5A [96]), and 1x TaqMan PreAmp Master Mix (Applied Biosystems) in a final volume of 10  $\mu$ L. The following thermal cycles were followed: 1 cycle: 95°C 10 min; 14 cycles: 95°C 15 seconds, 60°C 4 min. The pre-amplified product was diluted 1:5 and the pre-amplification reaction was tested with a

pass/fail test with GAPDH endogenous control primers, verifying that the  $C_T$  value was close to 20.

Primer and Roche UPL probe mix (“promer”) was prepared by mixing 2  $\mu$ l of a 20  $\mu$ M mix of both primers for each 1  $\mu$ L of the respective probe (10  $\mu$ M stock). The “promer” was tested with an equimolar mixture of cDNA from all the tissues (except fiber, which are included in the cortical peels), if it did not work new primers were designed, if the new primers did not work, it was presumed to be not expressed and the primers were used regardless, and run in the 96.96 dynamic array (Fluidigm Corporation, CA, USA).

Fluidigm 96.96 control line fluid was used to prime the fluidics arrays with the 136x chip prime script. Then the appropriate inlets were loaded with the different assays and sample mixes. The three biological replicates for each tissue were each placed in three different positions on the array for three technical replicates each.

The manufacturer’s protocol was followed to prepare the assay and sample mixes. The assay inlets contained a 6.5  $\mu$ L assay mixture containing 1x DA assay loading reagent and 2  $\mu$ L of the respective “promer” (primer + probe) mixture for each inlet. The sample inlets contained 1x TaqMan® Universal PCR Master Mix, No AmpErase® UNG (Applied Biosystems PN 4324018), 1x DA sample loading reagent (Fluidigm PN 85000735), and 2.5 $\mu$ L of the respective preamplified sample.

Once the samples and assays inlets were loaded, the 136x load mix script was executed to load the samples and assays. Once loaded, the chip was moved to the Biomark instrument and the following thermal cycles were executed: 1 cycle: 95°C 10 min; 40 cycles: 95°C 15 seconds, 60°C 1 min.

### 3.2.5 Analysis of 96.96 dynamic array results

Only those wells with a quality score of  $\geq 0.65$  were used in further analyses. The mean of the technical replicates was calculated. Then the delta- $C_T$  was obtained by calculating the geometric mean of the endogenous controls for the given tissue/biological-replicate, and subtracting that value to the  $C_T$  of the gene at that tissue/biological-replicate. Subsequently, the mean and the standard error of the delta- $C_T$  of the three biological replicates were calculated.

### 3.2.6 EST and RNAseq data mapping

The EST and assembled RNAseq data (Deyholos and collaborators, manuscript in preparation) were mapped against the PME and PME1 CDS sequences, using the read mapping tool, on the CLC Genomics Workbench 6.0.1, with the default parameters, except for the length fraction (0.8), and the similarity fraction (0.9 or 0.98).

### 3.2.7 Signal peptide, transmembrane domain, and protein subcellular localization predictions

SignalP 4.0 was used to search for signal peptides [97] Transmembrane domains were predicted using TMHMM v.2.0 [98]. The protein subcellular localization was predicted using WoLF PSORT and Plant-mPLoc [99, 100].

### 3.2.8 Cleavage site prediction

Proteolytic cleavage sites were predicted using a protease recognition pattern described by Pelloux and collaborators and Wolf and collaborators [90, 91]. The motif [RKQ][RKEHLN][LDMI][LMAKR] was searched in the Type-1 PME proteins using “Protein Pattern Find” at <http://www.bioinformatics.org/>. These sites were also identified

visually on a ClustalW multiple alignment of the protein sequences of sequences in the same phylogenetic group. This allowed us to confirm the motifs found with the web tool, and also to identify possible novel cleavage recognition motifs, by comparison of the aligned sequences with known motifs.

### 3.2.9 Isoelectric point

The predicted isoelectric point of the complete and the mature proteins (i.e. after signal peptide and/or cleavage site removal) was calculated using Vector NTI 10 [101].

### 3.2.10 Phylogenetic analysis

Phylogenetic relationships among the PMEs and PMEIs from flax, *Manihot esculenta*, *Ricinus communis*, *Populus trichocarpa*, and *Arabidopsis thaliana* were inferred as follows. The PME and PMEI protein sequences from *M. esculenta*, *R. communis*, *P. trichocarpa*, and *Arabidopsis* were extracted from Phytozome (version 9.1) using the PFAM identifiers PF04043 (for PMEI) and PF01095 (for PME) in a keyword-based ontology search. Alignments for the complete PMEs and PMEIs proteins of these four species plus flax, and for the LuPMEs and LuPMEIs proteins of flax alone, were constructed using MUSCLE [102]. The alignments were used to first determine the substitution model that best described the evolutionary process of each set of proteins, using ProtTest [103], and then these models were used to construct maximum likelihood trees using GARLI [104] under the CIPRES web interface [105], with 100 bootstraps and 2 search-replicates. The result of the analysis in ProtTest showed that the model of evolution that best fit the set of genes for LuPMEs was WAG+I+G+F, and the same model was obtained for the LuPMEIs. For the analysis of the PMEs and PMEIs in all the analysed species, the best model was WAG+G+F. To estimate the divergence time of

presumptive paralogs (Ks) we aligned the nucleotide sequences of the LuPMEs and the LuPMEIs, and then we used MEGA5 [106] to determine genetic distance, for which we used the Kimura 2-parameter model [107] with the pairwise deletion option, and then calculated the divergence time using  $t = K/2r$ , where  $t$  is time,  $K$  is the genetic distance, and  $r$  is the substitution rate, either  $1.5 \times 10^{-8}$  [108] or  $8.1 \times 10^{-9}$  [109].

### 3.2.11 Conserved residues

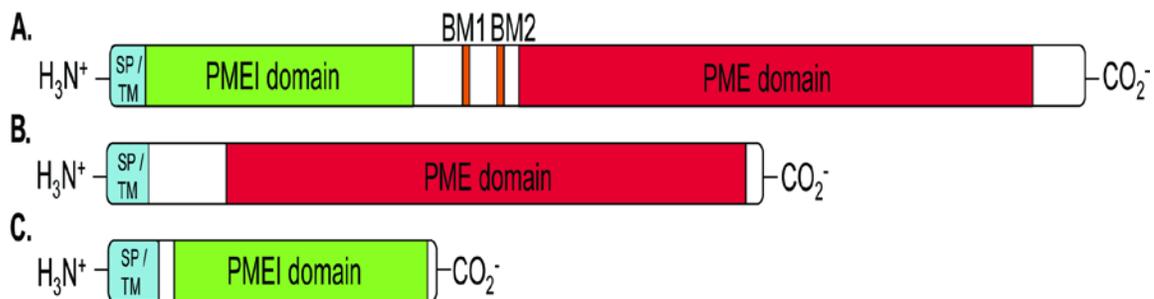
The presence of the most important residues for the protein activity was established based on the structural analysis done for the PME [47] and PMEI [46, 47]. For PME, 11 important residues were searched: six active sites with conserved aromatic residues, three catalytic residues, and two protein stabilizers. For PMEI, 33 important residues were analysed, including residues interacting with the active sites of the PME, residues responsible for disulfide bridging, and several residues responsible for maintaining the structure of the protein.

## 3.3 Results and Discussion

### 3.3.1 Annotation of LuPMEs and LuPMEIs

We identified 105 putative LuPMEs and 95 putative LuPMEIs (The FASTA formatted sequences can be accessed at <http://www.biomedcentral.com/1471-2164/14/742/additional>) by searching predicted transcripts of the flax whole-genome assembly (version 1.0) [1] for the PFAM domains Pectinesterase (PF01095) and PMEI (PF04043) [92]. Independent alignment of the *Arabidopsis thaliana* PME and PMEI families [110] to the flax genome did not identify any additional flax genes other than those identified by the PFAM domain alignment. Among the predicted LuPMEs, 60 were

Type-1 (i.e. encoding both a PMEI (PF04043) and PME (PF01095) domain [42]), and 45 were Type-2 (i.e. encoding a PME domain, but no PMEI domain [42]; Figure 3-1). Only one of the genes (*LuPME89*) contained an additional PFAM domain other than a PME or PMEI domain. This was a zf-RING\_2 domain (PF13639).



**Figure 3 - 1 Representatives of the three types of proteins classified in this report**

A. Examples of Type-1 PME: LuPME1; B.Type-2 PME: LuPME10; C. PMEI: LuPMEI1. SP: Signal peptide. TM: Transmembrane domain. BM: Binding motif.

### 3.3.2 Detection of PMEs and PMEIs in other plants

To identify putative PMEs and PMEIs in species other than flax, primary transcripts in the Phytozome v9.1 database were searched for the presence of a PME or PMEI domain. The number of predicted PMEs and PMEIs in each species was compared as a proportion of all proteins predicted for each species (Appendix Figure A3-1). The proportion of PMEs (0.25%) and PMEIs (0.22%) in flax was similar to the average proportion in the Malpighiales species sampled, i.e. 0.23% and 0.16%, respectively. Among the plants analyzed, *Mimulus guttatus*, followed by *Capsella rubella* had the highest proportion of PMEs 0.30% and 0.29%, respectively. Meanwhile *C. rubella* and *Arabidopsis thaliana* had the highest proportion of PMEIs 0.31% and 0.27%, respectively. The proportion of PMEs was diminished significantly in the grasses as compared to other angiosperms (t-test  $p < 0.01$ ). On average the angiosperms (not including grasses) had 0.22% PMEs as a proportion of the total predicted gene number, while grasses had 0.12%.

### 3.3.3 Transcript expression profiling

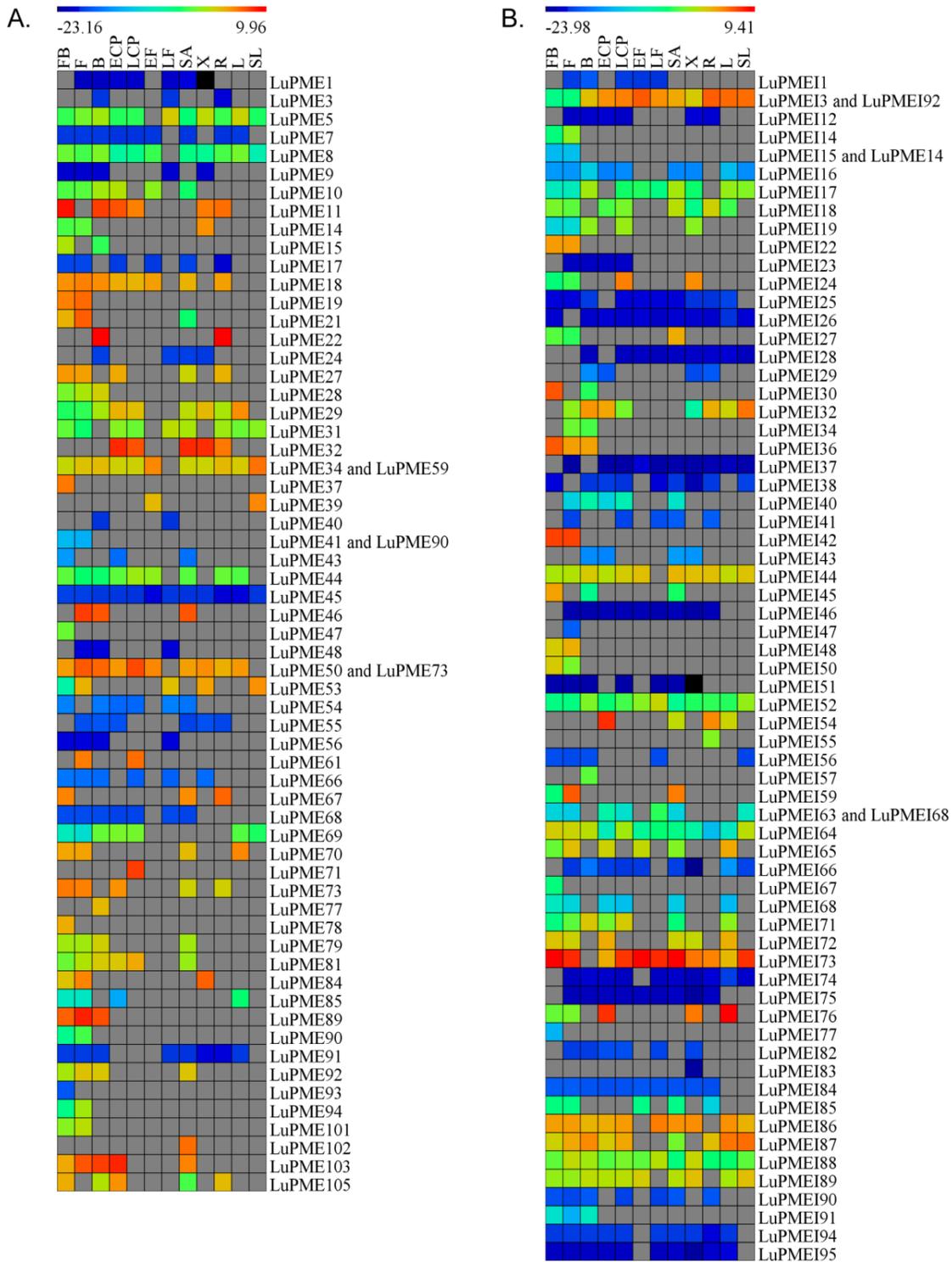
We examined transcript expression data to determine whether each predicted LuPME and LuPMEI was expressed, and if so, under what circumstances. Data sources for this analysis included qRT-PCR experiments (described below), published microarray data (Fenart and collaborators [111]), published flax ESTs (Venglat and collaborators [112]; NCBI), and unpublished Illumina RNAseq read data from the flax shoot apical meristem (Deyholos and collaborators, manuscript in preparation), and from the developing flax stem (One Thousand Plants Consortium, manuscript in preparation).

### 3.3.4 qRT-PCR using a 96.96 dynamic array

We used a Fluidigm 96.96 microfluidic array to conduct qRT-PCR on 12 different tissues of flax. Five of the tissues (shoot apex (SA), leaves (L), roots (R), early cortical peel (ECP), and early fibers (EF)) were collected from vegetative stage plants 1 month after germination; the other seven tissues (senescent leaves (SL), xylem (X), late cortical peel (LCP), late fibers (LF), flower buds (FB), flowers (F), and green bolls (B)) were collected from plants 2 months after germination, at the green capsule stage. We were able to design gene-specific primers for 102 out of the 105 predicted PMEs. Transcripts of 62/102 PME genes (60.8%) were detected in at least one of the tissues (Figure 3-2A), with a minimum Fluidigm 0.65 quality score and in at least 2 out of 3 biological replicates. 40/102 predicted PMEs (39.2%) could not be detected in any of the tissues assayed by qRT-PCR. However, transcripts of 6/40 of these genes could be found among public ESTs collections (80% coverage and 98% identity), and an additional 8/40 genes could be identified among reads from RNAseq transcript profiling experiments (manuscript in preparation) (Tables 3-1 and 3-2). Finally, one of the three predicted PMEs for which qRT-PCR was not attempted was also detected by RNAseq. The unigenes assembled by Fenart and collaborators [111] from ESTs of a flax fiber variety were also queried. Out of the 20 unigenes (16 PMEs and 4 PMEIs) that mapped with 80% coverage and 98% identity to the predicted LuPMEs listed in Tables 3-1 and 3-2, 14 PMEs and 3 PMEIs were detected within the Fluidigm qRT-PCR array, while all were identified by either the EST collections or the RNAseq data. In total we were therefore able to find experimental evidence for the transcription of 77/105 predicted PMEs (Tables 3-1 and 3-2).

We used the same Fluidigm qRT-PCR array system to assay transcription of 94 out of 95 predicted PMEIs. 66/94 genes (70.2%) were detected in one or more tissues (Figure 3-2B), and 28/94 (29.8%) PMEIs were not detected in any of the tissues. However, 17/29 of the predicted PMEIs that were not detected or assayed by qRT-PCR were identified among either public ESTs collections or in RNAseq data from developing stems (Table3-3). Together, these data provide evidence that at least 83/95 (87.4%) of the predicted PMEIs are transcribed.

Using qRT-PCR and by querying previously published and unpublished transcript databases, we were able to confirm that 77/105 and 83/95 of the predicted LuPMEs and LuPMEIs, respectively, are transcribed. The remaining genes might also be transcribed but under conditions different from those assayed to date. We note, for example, that none of the tissues surveyed to date are from plants subjected to stress, which is likely to induce PMEs that may not be otherwise transcribed. Among genes that are known to be transcribed, we found transcripts expressed in fibers and fiber bearing tissues, during either elongation (7 PMEs, 3 PMEIs), thickening (16 PMEs, 10 PMEIs), or maturation and thickening (19 PMEs, 24 PMEIs) (Table3-4). These genes are primary targets for manipulation by reverse genetics, in order to develop flax feedstocks with modified fiber properties.



**Figure 3 - 2 Heat map of transcript abundance of PMEs (A) and PMEIs (B) at different tissues**

Delta- $C_T$  ( $C_T$  of gene minus  $C_T$  of geometric mean of the endogenous controls). The color of the cell represents transcript abundance. Gray cells indicate no transcripts were detected. When two different genes appear in the same row it means one set of primers was used as a common assay for both genes. FB: Flower buds; F: Flowers; B: Green boll; ECP: Early cortical peels; EF: Early fibers; LF: Late fibers; SA: Shoot apical meristem; X: Xylem; R: Roots; L: Leaves; SL: Senescent leaves.

**Table 3 - 1 Important features of type-1 LuPMEs**

LuPME	Mature protein			Binding motif for cleavage			Subcellular localization				Conserved residues			Tree	Expression						
	AA	Kda	pl	BM1	BM2	positions	sp	tm	WoLF PSORT	Plant-mPloc	as	st	cr		Group	oe	fl	uf	ev	en	rs
1	333	34.9	8.82	RKLL	RKLL	203-223	+	-	ch, m	cw	6	2	3	D	+	+				S3	+
3	320	34.4	9.77		RRLL	235	+	+	ch, v, ex	cw	6	2	3	D	+	+					+
5	319	34.6	9.53		RRLL	234	-	+	pl, g_pl, er, g	cw	6	2	3	D	+	+	1	S, SP, F	5	All	+
6	322	35.5	7.18	RKLL	RRVL	202-217	+	-	ch, n, pl	cw	5	2	3	C	+	-					+
7	322	35	6.6	RKVA	RRLL	235-259	+	-	ch, v, n	cw	6	2	3	D	+	+	1	S	2	All	+
8	364	40.1	8.66	RKLL	RRLL	263-286	-	+	n, cy, v, er, m, pl	cw	6	2	3	B	+	+	1	S	2	S3 and 4	+
9	322	35.6	9.6	RRLL	RKLL	224-242	+	+	ex, er, v, g, ch, n, cy	cw	5	2	3	C	+	+		F	34		+
11	593	64.1	8.46				-	+	ch, m	cw	6	2	3	A	+	+	1	ES, HE, TE, GS	4	All	+
13	323	35.3	10		RRLL	137	-	-	ch	cw	6	2	3	D	-	-					
14	321	35.3	9.47	RRLL	RKLL	226-248	-	+	er, pe, ex, g, ch	cw	5	2	3	C	+	+		F	1	SA	+
17	322	35.4	8.89	RRLW	RRLL	197-221	+	+	v, ch, ex, er, g	cw	6	2	3	D	+	+				S3 and 4	+
18	393	43.1	8.39	RKLR		207	+	-	ch, ex, g, n, cy, pl, v, er	cw	5	2	3	D	+	+				S3 and 4	+
19	331	36.4	8.33	RKLL	RKLL	163-187	-	-	cy, n	cw	6	2	3	D	+	+					
22	324	35.8	8.8		RRKL	613	-	-	n, cy, v	cw	6	2	3	D	+	+	2	GE	1	all	+
23	334	36.8	7.06				+	+	ch, n, ex, m, pl, v	cw	6	2	3	D	-	-					
30	291	32	9	RRLW	RRLL	232-258	-	+	v, ch, g, n, pl	cw	5	1	2	D	+	-			4	All	+
31	323	35.6	8.21	RKLK		195	+	+	ch, ex, n, pl, v, er	cw	6	2	3	D	+	+	1			S3 and 4	
32	386	41.2	9.51				-	+	m, ch	cw	1	0	0	A	+	+				S3 and 4	+
36	534	58.9	8.97				+	-	v, ch, n, g	cw	6	2	3	D	+	-				SA	+
37	398	45.2	5.93	RRLL	RRLL	246-283	-	+	ch, n, ex, v, er, g	cw	6	2	3	C	+	+					
38	473	52.2	6.41				+	-	ex, ch, v, n, cy, pl	cw	6	2	3	D	+	-	1	CE, ME	3		
43	297	32.6	9.49	RRML	RKLL	241-273	-	+	er, ch, cy, v, ex	cw	6	2	3	C	+	+					
44	414	46	7.74		RKLL	103	-	-	n, cy	cw	6	2	3	D	+	+		TS	1	all	+
45	318	33.5	8.6		RELL	195	+	+	ch, cy, v, pe	cw	5	2	3	D	+	+		TS	2	S3 and 4	+
46	324	35.8	9.67	RRLL	RRLL	263-292	-	+	n, er, cy, ct, ch, m, v	cw	6	2	3	D	+	+				all	+
47	326	35.6	9.17	RRML	RKLL	241-273	-	+	v, g, ch	cw	6	2	3	C	+	+					
49	525	58.4	9.29				+	-	ex, ch, v, er	cw	6	2	3	D	+	-	2			all	+
50	427	46	9.47				+	-	ch, n	cw	4	1	1	D	+	+	1			SA and S3	+
51	359	39.8	5.21	RRLL		406	-	-	cy, ch, n, pl	cw	2	1	0	C	-	-					

**Table 3-1 Important features of type-1 LuPMEs (continued)**

53	322	35.4	9.05	RRLL	RRLL	213-239	-	-	cy, ch	cw	5	2	3	C	+	+	F	5			
54	533	58.9	9.79	RRLL	RRLL	253-281	-	+	n, cy, er	cw, n	6	2	3	D	+	+	1	ES, S, F	3	all	+
56	322	35.6	9.61	RRLL	RKLL	222-240	+	+	ex, er, v, g, ch, n	cw	5	2	3	C	+	+	F	37			
62	334	36	9.94		RRLL	217	+	+	ch, n, ex, v, cy	cw	6	2	3	D	-	-					
63	518	57.6	9.54				+	+	ch, ex, er, n, cy, m	cw	6	2	3	D	-	-					
64	322	35.9	10.32		RKVL	358	-	-	ch, n	cw	6	2	3	D	-	-					
65	319	34.3	9.01		RRLL	187	-	-	n, ch, cy, v	cw	6	2	3	D	-	-					
66	348	38.9	5.87	RRLL	RRML	265-289	-	+	cy, cy_pe, pe, er, pl	cw	5	2	3	C	+	+	F	4			
70	387	42.7	7.71	RKLR		204	+	-	ch, n	cw	6	2	3	D	+	+			all	+	
71	529	58.6	5.92				+	-	ch, n, ex, v, m	cw	6	2	3	D	+	+				+	
72	407	46.1	5.57	RRLL	RRLL	318-350	-	-	n, cy, ch	cw	6	2	3	C	+	-	F	10		+	
73	223	24.6	9.03		RKLL	250	+	-	ch, n	cw	6	2	3	D	+	+	1		all	+	
74	526	58.6	9.29				+	-	ch, ex, v, er, g	cw	6	2	3	D	+	-			all	+	
75	327	36.1	9.92		GRLL		+	+	ch, v, n	cw	6	2	3	D	-	-					
76	335	36.8	8.76		RRLL	224	+	-	ex, ch, v	cw	6	1	3	D	-	-					
78	400	45.4	6.12	RRLL	RRLL	227-258	+	+	ch, ex, v, g, n	cw	6	2	3	C	+	+					
79	536	59	9.04				+	-	v, ch, n	cw	6	2	3	D	+	+			all	+	
80	327	36.1	9.68				+	+	ch, v	cw	6	2	3	D	-	-					
81	327	36	9.13		GRLL		+	-	ex, er, ch, v	cw	6	2	3	D	+	+					
82	334	37.2	9.49	RRLL	REYL	242-253	-	+	cy, cy_n, ch, n	cw	6	2	3	B	-	-					
83	322	35.4	8.95	RRLL	RKLL	237-260	-	+	er, ex, g, ch, cy, v	cw	5	2	3	C	+	-	F	16			
84	547	60.7	5.25				-	+	ch, pl, n	cw	6	2	3	A	+	+					
85	318	33.4	7.84		RKLL	203	+	+	ch, cy, cy_n, v, pe	cw	5	2	3	D	+	+			S3 and 4		
86	330	37	9.52	RRLL	REYL	241-252	-	+	cy, cy_n, n, ch, pl, v	cw	5	2	3	B	-	-					
91	321	35.2	9	RKLL	RRLL	274-301	-	+	n, cy, cy_pe, v, er, ct	cw	6	2	3	B	+	+	1	F	1	all	+
92	331	36.3	7.82		RRLL	267	+	+	pl, er, v, n	cw	6	2	3	D	+	+	S	1	all	+	
93	347	37.7	5.16	RRLL	RRFL	246-272	-	+	cy, ex, pe	cw	3	2	3	C	+	+	F	1		+	
95	323	34.8	9.47		RRLL	214	+	+	ex, v, ch, g, pl	cw	6	2	3	D	-	-					

**Table 3-1 Important features of type-1 LuPMEs (continued)**

96	327	36.6	10.22		RKVL	154	-	-	n, cy, pl, ct_pl, ch, ct	cw	6	2	3	D	-	-		
97	402	45.1	10.48		RRVL	228	+	+	v, g, n, pl, ex	cw	6	2	3	D	-	-		
99	324	36.2	8.24	RRLL	RRML	277-301	-	+	cy, er, ch, pl, v	cw	5	2	3	C	+	-	F	10

LuPME8, LuPME51, LuPME97, and LuPME22 have two PME domains. (SP): Presence of signal peptide; (TM): Presence of transmembrane domain. Subcellular localization: ch: chloroplast; cw: cell wall; cy: cytosol; er: endoplasmic reticulum; ex: extracellular/cell wall; g: Golgi apparatus; m: mitochondria; n: nuclear; pl: plasma membrane; v: vacuolar membrane; pe: peroxisome; ct: cytoskeleton. (AS): Number of conserved residues at active site (out of 6); (ST): Number of conserved stabilizer residues (out of 2). CR: Number of conserved catalytic residues (out of 3). Expression: A gene was reported as positive if the coverage with the EST or assembled RNAseq sequence was higher than 80% and the identity higher than 98%. (OE) Overall transcript expression based on all the methods assessed: (+) expression was detected with at least one of the methods; (-) expression not detected in any of the methods. (FL) Gene expression based on qRT-PCR (Fluidigm): (+) it was expressed in at least one tissue; (-) no expression; (NA) no assay done. (UF): Number of genes aligning with unigenes reported by Fenart and collaborators [111], 80% coverage and 98% identity. (EV) Venglat and collaborators [112] ESTs, 80% coverage and 98%identity: (F) flower; (S) stem; (SP) stem peel; (ES) etiolated seedling; (L) leaf; (GE) globular embryo; (HE) heart embryo; (TE) torpedo embryo; (ME) mature embryo; (TS) torpedo seed coat; (GS) globular seed coat; (EP) endosperm pooled. (EN) Number of ESTs from NCBI as of April 2013. (RS): Alignment with RNAseq assembled sequences data (80% coverage and 98%identity) obtained from one kp project at different positions in the stem: (SA) shoot apical meristem; (S3): stem 3; (S4) stem 4. (RSA): RNAseq shoot apical meristem.

**Table 3 - 2 Important features of type-2 LuPMEs**

LuPME	Mature protein			Subcellular localization				# of Conserved residues			Tree Group	Expression						
	AA	Kda	pl	SP	TM	Wolf PSORT	Plant-mPLOC	AS	ST	CR		OE	FL	UF	EV	EN	RS	RSA
2	192	21	4.8	+	-	ch, ex, v	ow	1	1	2	E	-	-		0			
4	305	34.4	6.5	-	-	n, cy, m, ch, ex	ow	5	1	3	E	-	-		0			
10	348	38.2	8.4	+	+	ch, ex, v, m	ow	4	2	3	E	+	+	ES, L, GE, HE, TE, ME	12	All	+	
12	223	23.1	8.7	-	-	ch, n, m	ow	2	0	2	C	+	-		0		+	
15	267	29.9	5.7	-	-	n, cy, pl	ow	3	1	2	E	+	+		0			
16	329	37.1	9.4	+	-	ch	ow	5	1	3	E	+	-		0		+	
20	252	27.6	8.9	-	-	cy, ch, ex, n, m	ow	5	1	3	D	-	-		0			
21	352	39.1	8.5	-	+	ex, er, ch, cy, m, v	ow	6	2	3	E	+	+		0	All	+	
24	414	45.4	8.3	-	+	ex, ch, v, er	ow	5	2	3	E	+	+	1	0	All		
25	319	34.7	8.5	+	-	ch, ex, m, v	ow	4	1	3	E	+	-		0		+	
26	322	35.4	9.5	+	-	ch	ow	5	1	3	E	+	NA		0	all	+	
27	363	39.7	8.9	+	+	ch, v, n, pl	ow	6	2	3	E	+	+		0	SA and S4	+	
28	74	8.5	11.2	-	-	cy, ch, n, pl	ow, ch	0	0	0	D	+	+		0			
29	356	39.3	9.1	+	+	ch, v, n, pl	ow	6	2	3	E	+	+	F	2	all	+	
33	219	24.8	4.9	-	-	cy, ch, n, pe	ow	5	0	2	A	-	-		0			
34	317	35.5	7.7	-	-	cy, n, ch	ow	6	2	3	E	+	+	1	0		+	
35	145	15.9	9.1	-	-	n, ch, cy	ow	3	0	1	D	-	-		0			
39	348	38.8	7.8	+	+	v, cy, ch, m	ow	4	2	3	E	+	+		0			
40	316	35.8	9	+	-	ch, n, m	ow	5	1	3	E	+	+	TSC	1			
41	340	37.3	6.3	+	-	ex, v, er, ch, n	ow	3	2	3	E	+	+		0			
42	216	24.4	8.4	-	-	m, ch, v, n	ow, n	1	0	0	E	-	NA		0			
48	379	40.2	8.7	-	-	ch, m, n	ow	3	0	2	C	+	+		0		+	
52	336	37.6	6	-	-	cy, er, n	ow	6	1	3	A	+	-		0		+	
55	330	36.2	9	+	-	ch, v, ex, n	ow	3	1	3	E	+	+		0			
57	260	28.8	6.8	-	-	n, cy, ct, ex	ow	2	1	3	E	-	-		0			
58	123	14.4	9.4	-	-	cy, n	ow	3	0	1	C	+	-	F	1			
59	318	35.6	7.7	-	-	cy, n, ct, ct_pl	ow	6	2	3	E	+	+		0	SA	+	
60	325	35.2	8.4	+	+	ch, ex	ow	4	1	3	E	-	-		0			
61	302	31.9	9.4	-	-	cy, ch, m, n	ow	6	2	3	D	+	+	1	0	S3 and 4		
67	418	45.9	5.6	-	+	v, g, ex, ch	ow	4	2	3	E	+	+		0	All		
68	316	35.1	9.4	-	-	m, ch, n	ow	6	2	3	E	+	+	ES	1	SA	+	
69	220	24.5	7.9	-	-	ch, n, cy	ow	5	2	3	D	+	+		0			

**Table 3-2 Important features of type-2 LuPMEs (continued)**

77	318	35.2	7.7	+	-	ch, n	cw	5	1	3	E	+	+	0	+
87	348	38.5	7.3	-	-	cy, ct, n	cw	3	2	3	E	-	-	0	
88	140	15.5	8.9	-	-	cy, ch, m	cw	1	0	0	E	-	-	0	
89	311	33.6	5.2	-	-	n, ch, cy, ex	n	3	2	2	E	+	+	0	+
90	339	37.1	6.3	+	-	n, er, er_pl, m, pl, ch, cy	cw	3	2	3	E	+	+	0	+
94	332	37.6	8.5	+	-	ex, v, er, cy	cw	3	1	2	E	+	+	0	
98	354	40.1	8.1	+	-	ch, n, ex, cy	cw	5	1	3	E	-	-	0	
100	289	32.3	7	-	-	cy, n, ct, ch	cw	1	0	2	E	-	NA	0	
101	318	36.2	7.2	+	-	ex, v, er, ch, cy	cw	4	1	2	E	+	+	0	F
102	317	35.1	5.5	-	-	cy, n, ct	cw	5	2	3	E	+	+	0	All
103	359	39.8	8.9	+	-	ch, cy, n, m, ex	cw	6	2	3	E	+	+	0	all
104	349	38.4	9.3	-	-	m, ch_m, ch	cw	4	1	3	E	-	-	0	
105	343	37.6	9.1	-	-	ch, cy	cw	6	2	3	E	+	+	0	SA

LuPME8, LuPME51, LuPME97, and LuPME22 have two PME domains. (SP): Presence of signal peptide; (TM): Presence of transmembrane domain. Subcellular localization: ch: chloroplast; cw: cell wall; cy: cytosol; er: endoplasmic reticulum; ex: extracellular/cell wall; g: Golgi apparatus; m: mitochondria; n: nuclear; pl: plasma membrane; v: vacuolar membrane; pe: peroxisome; ct: cytoskeleton. (AS): Number of conserved residues at active site (out of 6); (ST): Number of conserved stabilizer residues (out of 2). CR: Number of conserved catalytic residues (out of 3). Expression: A gene was reported as positive if the coverage with the EST or assembled RNAseq sequence was higher than 80% and the identity higher than 98%. (OE) Overall expression based on all the methods assessed: (+) expression was detected with at least one of the methods; (-) expression not detected in any of the methods. (FL) Gene expression based on qRT-PCR (Fluidigm): (+) it was expressed in at least one tissue; (-) no expression; (NA) no assay done. (UF): Number of genes aligning with unigenes reported by Fenart and collaborators [111], 80% coverage and 98% identity. (EV) Venglat and collaborators [112] ESTs, 80% coverage and 98% identity: (F) flower; (S) stem; (SP) stem peel; (ES) etiolated seedling; (L) leaf; (GE) globular embryo; (HE) heart embryo; (TE) torpedo embryo; (ME) mature embryo; (TS) torpedo seed coat; (GS) globular seed coat; (EP) endosperm pooled. (EN) Number of ESTs from NCBI as of April 2013. (RS): Alignment with RNAseq assembled sequences data (80% coverage and 98% identity) obtained from the OneKP project at different positions in the stem: (SA) shoot apical meristem; (S3): stem 3; (S4) stem 4. (RSA): RNAseq shoot apical meristem.

Table 3 - 3 Important features of LuPMEIs

LuPMEI	Mature prot.			Subcellular localization				Conserved residues							Expression						
	AA	Kda	pI	SP	TM	Wolf PSORT	Plant-mPLoc	*	\$	□	£2	£3	¥	P55	OE	FL	UF	EV	EN	RS	RSA
1	155	16.1	6.8	+	-	ex, m, v, ch	pl	10	4	4	1	0	4	N	+	+		EP	1		
2	174	18.9	8.6	-	-	ch, cy, m, n	pl, ch	8	2	2	0	2	2	A	-	-			0		
3	162	17.9	6.1	+	-	ex, ch	pl	6	3	2	1	2	4	G	+	+			0		+
4	113	12.6	4.9	-	-	ch, n, m, pl, g_pl	pl, n	2	1	4	1	2	2	-	-	-			0		
5	159	17	4.9	+	-	ex, ch, m	pl	9	5	4	3	1	4	P	+	-		HE	1		
6	177	18.7	8.8	+	+	v, pl, ex	pl	11	5	4	2	1	4	A	+	-		GS	1		
7	155	16.8	5.6	+	+	ex, v, ch, n	pl	11	4	3	1	3	4	A	-	-			0		
8	167	17.8	5.3	-	+	ex, ch, v	pl	7	2	3	2	1	4	T	-	-			0		
9	192	21.5	10	+	+	ch	pl	5	4	4	1	1	4	T	-	-			0		
10	199	21.2	5.1	+	-	ex, ch, m, v	pl	7	2	4	0	0	4	P	-	-			0		
11	296	33.7	4.7	-	-	n	n	6	3	4	1	2	4	H	-	-			0		
12	155	16.7	8.4	+	-	ex, ch	pl	8	4	2	1	0	3	A	+	+		ES	1		+
13	160	16.8	8.3	+	-	ex, ch, v	pl	11	4	2	1	0	4	S	+	-			0	S4	+
14	159	17.4	5.1	+	-	ch, v, ex	pl	11	3	4	2	1	4	S	+	+			0		
15	159	17.3	5.1	+	-	pl, ch, er, ex	pl	11	3	4	2	1	4	S	+	+			0		
16	198	21.3	5.1	+	-	ex	pl	7	5	4	3	1	4	I	+	+		GS	1	SA and S3	+
17	187	20.3	4.5	+	+	ch	pl, n	8	4	3	1	2	4	A	+	+		HE, F	2	S4	+
18	181	19.5	8.3	+	-	ch, m, v	pl	8	3	3	3	1	4	A	+	+		TS	1	SA	+
19	159	17	5	+	-	ex, ch, m	pl	9	5	4	3	1	4	P	+	+		GE, F	4		
20	239	26	5.6	-	+	ch, ch_m, n, m, cy	pl	9	4	4	2	1	4	S	+	-			0	S3 and 4	+
21	161	17.2	7.9	+	-	ch, ex, cy	pl	11	3	4	1	2	4	A	-	-			0		
22	483	55.4	4.6	+	-	ex, v, g, n	n	5	3	4	1	2	4	P	+	+			0		+
23	198	20.8	4.9	+	-	ch, ex, cy	pl	6	3	3	1	1	4	P	+	+		F	5		
24	180	19.6	8.9	+	-	v, ch, ex	pl	9	3	3	0	0	4	P	+	+		F	3		
25	172	18.6	4.9	+	-	ex, ch	pl	11	5	3	3	1	4	K	+	+			0		
26	154	16.2	5.5	+	-	ex, m, ch	pl	10	4	2	1	1	4	A	+	+		L	1	SA and S3	+
27	149	16.4	4.7	+	-	ex, v, ch	pl	10	3	4	1	2	4	N	+	+			0		
28	166	17.1	8.3	+	-	ex, pl, ch, m, v	pl	11	4	2	1	0	4	S	+	+			0	SA	+
29	149	16.7	4.8	+	-	v, ch, ex, er	pl	10	2	4	1	1	4	N	+	+		EP, HE	3		
30	149	16.4	4.6	+	+	ex, v, cy, m	pl	10	2	4	1	0	4	T	+	+		EP,GS	24		

Table 3-3 Important features of LuPMEIs (continued)

31	165	17.8	5.5	+	-	pl, g, ex, er	pl	11	4	2	1	1	3	K	-	-		0		
32	155	16.5	8.4	+	+	ch, v, cy, n, m	pl	10	4	3	1	0	4	T	+	+		0		
33	152	15.9	5.1	+	-	ex, ch	pl	9	4	3	1	2	4	A	+	-	GS	1		
34	152	16.2	4.7	+	+	v, pl	pl	11	4	4	2	2	4	A	+	+		0		
35	154	16.1	4.6	+	-	ex, ch, n	pl	10	4	4	2	2	4	A	+	-		0		+
36	152	16.3	4.7	+	-	ex, ch, cy, m	pl	11	5	4	3	2	4	A	+	+		0		
37	167	17.9	5.4	-	-	ch, cy, n, ct_n	pl	8	4	4	2	1	4	P	+	+		0		+
38	163	17.8	7.8	+	-	ex, ch, pl	pl	9	5	3	1	2	4	P	+	+		0	S3	+
39	158	17.4	4.5	-	-	ch, ex, n	pl	10	5	2	3	1	4	K	+	-		0		+
40	164	17.6	5.7	+	+	v, ex, pl, m, er_pl, n	pl	9	5	3	2	2	4	A	+	+	1	HE	1	+
41	151	16.1	5.3	+	-	ch, v, n, m, ex	pl	11	5	2	1	2	4	M	+	+	HE, GS	2		
42	159	17.2	6.1	+	-	ex, ch, v, m	pl	10	3	4	1	1	4	S	+	+		0		
43	181	19.6	9.5	+	-	ch, v	pl	7	4	4	1	1	4	I	+	+		0	SA and S3	+
44	187	19.9	6.8	+	+	ex, v	pl	7	5	4	2	1	4	I	+	+		0	All	+
45	159	17.3	6.5	+	-	ex, ch, n, cy	pl	7	3	4	3	1	4	I	+	+		0	SA	+
46	389	45	5.1	-	-	n	n	5	3	4	1	2	4	P	+	+		0		+
47	188	20.2	8.4	+	-	ex, v, ch, m	pl, n	7	3	4	0	1	4	P	+	+		0		
48	193	21.2	7.7	+	+	n, ex, ch, cy, v	pl, n	7	3	4	1	0	4	P	+	+	F	2		
49	197	21.5	6.9	+	+	ex, ch, n, m, v	pl, n	8	3	4	1	0	4	P	-	NA		0		
50	467	50.6	6.5	-	-	n, ch, pl, m	pl, n	7	4	3	1	3	4	A	+	+		0		
51	368	37.4	4.3	+	+	ex, ch, v, n	n	8	4	2	1	2	4	P	+	+		0		+
52	65	7.3	4.5	-	-	cy, n, ch, pl, ex	pl, n	5	1	1	0	1	1	-	+	+		0		
53	161	17.8	9	+	+	ch, ex	pl	7	5	4	3	1	4	I	+	-	S	1		+
54	175	18.7	6.1	+	-	ch, m, ex	pl	7	5	4	3	1	4	I	+	+		0	All	+
55	141	15.3	5.9	-	-	ch, m	cw	4	1	4	2	1	3	-	+	+		0		
56	167	18.8	8.3	+	-	ch, ex, cy, m, pl	pl, n	7	4	3	1	2	4	D	+	+		0		+
57	154	16	6.4	+	+	ex, ch, cy, m	pl	9	4	4	2	0	4	N	+	+	GE	1		
58	195	21.1	5.2	+	-	ex, ch	pl	7	5	4	3	1	4	I	+	-		0	all	
59	173	18.9	4.5	+	-	ex, ch, n	pl	10	5	2	3	1	4	K	+	+		0		
60	195	21.7	9.9	+	-	ch	pl	6	4	4	1	1	4	T	+	-		0		+

Table 3-3 Important features of LuPMEIs (continued)

61	468	51.7	8.7	-	-	n, ch, cy	pl, n	7	4	3	3	1	3	G	+	-		0	All	+
62	118	13.2	6.9	-	+	ch, cy, n, ex	pl	4	1	3	2	3	2	-	-	-		0		
63	170	18.3	5.9	+	-	ch, v, ex, m	pl	7	5	4	2	2	4	V	+	+	TE	1		+
64	177	19.1	10	+	-	ch	pl	7	4	4	3	1	4	I	+	+		0	SA	+
65	180	19.1	5.3	+	-	ch, m	pl	6	5	4	3	1	4	I	+	+	1	0	all	+
66	160	17.8	7.8	+	-	ex, ch, n, cy	pl	7	3	4	0	2	4	T	+	+	TS, EP	4	SA	
67	174	18.8	4.6	+	-	ex, ch	pl	11	5	4	3	1	4	K	+	+		0		
68	177	19	5.5	+	-	ch, ex, v, m	pl	7	5	4	3	2	4	V	+	+		0		+
69	179	19.3	9.2	+	-	ch	pl	7	5	4	3	2	4	I	+	-		0		+
70	177	19.2	10.1	+	-	ch	pl	7	4	4	3	1	4	I	+	-		0	All	+
71	183	19.6	5.7	+	-	ch, m	pl	6	5	4	3	1	4	I	+	+		0		+
72	183	20	9.4	+	-	ch, n	pl	7	5	4	1	1	4	I	+	+		0		+
73	212	22.6	7.7	+	-	ex, m, v	pl	7	5	4	2	1	4	I	+	+		0	SA and S4	+
74	159	17.4	5.5	+	-	ch, ex, n, cy, m	pl	6	3	4	3	1	3	I	+	+		0	S3 and 4	+
75	170	19.1	9	+	-	ch, pl	pl, n	7	5	3	1	2	4	D	+	+		0	S4	+
76	228	25	9	+	-	ch, v, ex	n	9	3	2	0	0	4	P	+	+	F	1		
77	220	23	4.6	-	-	ch, n	pl	9	3	3	1	1	4	P	+	+		0		
78	163	17.7	5.7	+	-	pl, g, ex, ch	pl	11	4	2	1	1	4	K	+	-		0	SA and S3	
79	155	16.6	6.9	+	+	er, ex, er_pl, ch, m, pl, cy	pl	10	4	3	1	0	4	T	+	-		0	S3 and 4	
80	229	24.9	5.7	+	-	ch, ex	pl, n	7	4	3	2	2	4	A	-	-		0		
81	184	19.3	4.8	-	-	ch, m, n	pl	5	2	4	2	1	3	P	+	-	1	0	S3	+
82	334	34.4	4.4	+	+	ex, ch, v, n	n	8	4	2	1	2	4	P	+	+		0		
83	249	26.9	4.7	+	-	ch, ex, v	pl, n	7	4	3	1	2	4	A	+	+		0		
84	193	20.5	4.7	+	-	ex, ch, v	pl	7	5	4	2	1	4	V	+	+		0	S3	+
85	156	16.9	8.4	+	-	ex, v, pl	pl	8	3	3	0	0	4	R	+	+		0	All	+
86	150	16.2	9.2	+	+	v, ex, pl	pl	10	2	4	0	1	4	A	+	+		0	All	+
87	162	17.7	9	+	-	ch	pl	7	5	4	3	1	4	I	+	+		0		+
88	172	18.3	5.4	+	+	ch	pl	7	5	4	3	1	4	I	+	+	TS	1	All	+
89	155	16.6	8.8	+	+	ex, pl, v, ch, cy, m	pl	10	2	4	0	1	4	A	+	+		0	S3 and 4	+
90	188	19.9	5.7	-	-	ch, m, cy	pl, cy	6	3	3	2	1	2	P	+	+	F	1		+

**Table 3-3 Important features of LuPMEIs (continued)**

91	327	36.6	9.2	-	+	pl, ch, m, er	pl	9	4	3	2	3	4	T	+	+		0			
92	162	17.9	7.8	+	-	ex, ch	pl	6	3	2	1	2	4	G	+	+		0	all	+	
93	156	16.5	4.4	+	+	n, ex, m, er_pl, pl, er	pl	11	5	4	2	2	4	G	+	-	EP	1			
94	331	35.9	6.3	-	+	n, g, pl, er_pl, cy	cy	9	4	4	2	1	4	S	+	+	1	TE	1	SA and S4	+
95	157	17.3	5.3	-	-	cy, ch, n, ex	pl	8	4	3	1	2	3	P	+	+		0		+	

(SP): Presence of signal peptide; (TM): Presence of transmembrane domain. Subcellular localization: ch: chloroplast; cw: cell wall; cy: cytosol; er: endoplasmic reticulum; ex: extracellular/cell wall; g: Golgi apparatus; m: mitochondria; n: nuclear; pl: plasma membrane; v: vacuolar membrane; pe: peroxisome; ct: cytoskeleton. Number of conserved residues at each one of the following: (\*) non polar bundle-hairpin interface (out of 12); (§) polar bundle-hairpin interface (out of 5); (¥) Polar interacting with PME aromatic residues (out of 5); (£2) acidic patch on  $\alpha$ 2 helix (out of 3); (£3) acidic patch on  $\alpha$ 3 helix (out of 3); (¥) disulphide bridge (out of 4). Expression: A gene was reported as positive if the coverage with the EST or assembled RNAseq sequence was higher than 80% and the identity higher than 98%. (OE) Overall expression based on all the methods assessed: (+) expression was detected with at least one of the methods; (-) expression not detected in any of the methods. (FL) Gene expression based on qRT-PCR (Fluidigm): (+) it was expressed in at least one tissue; (-) no expression; (NA) no assay done. (UF): Number of genes aligning with unigenes reported by Fenart and collaborators [111], 80% coverage and 98% identity. (EV) Venglat and collaborators [112] ESTs, 80% coverage and 98%identity: (F) flower; (S) stem; (SP) stem peel; (ES) etiolated seedling; (L) leaf; (GE) globular embryo; (HE) heart embryo; (TE) torpedo embryo; (ME) mature embryo; (TS) torpedo seed coat; (GS) globular seed coat; (EP) endosperm pooled. (EN) Number of ESTs from NCBI as of April 2013. (RS): Alignment with RNAseq assembled sequences data (80% coverage and 98%identity) obtained from one kp project at different positions in the stem: (SA) shoot apical meristem; (S3): stem 3; (S4) stem 4. (RSA): RNAseq shoot apical meristem.

### 3.3.5 Transcript expression patterns

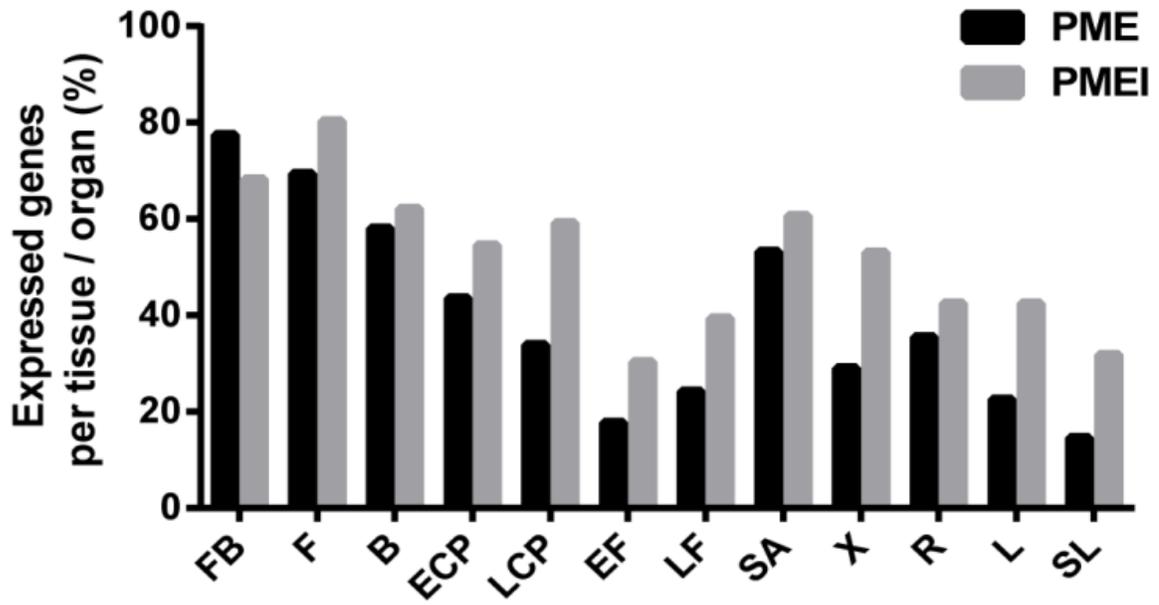
The tissue in which the largest number of expressed PMEs (48/62; 77.4%) was detected was the flower bud. Conversely, the tissue in which the lowest number of PMEs was detected was senescent leaves (9/62; 14.5%). The highest number of PMEIs detected was also in a reproductive tissue (flowers; 53/66 (80.3%). Conversely, the tissue with the fewest detectable PMEIs was early fibers, with only 20/66 (30.3%) (Figure 3-3).

We also identified PMEs and PMEIs whose transcript abundance was correlated with phloem fiber development. The transcript expression of 11 PMEs and 20 PMEIs was detected in EF, while 15 PMEs and 26 PMEIs were expressed in LF (Figure 3-2). 15 PMEIs were expressed in both EF and LF while only one PME was expressed in both of these stages. Nine PMEs and five PMEIs were detected in EF but not LF, and conversely 13 PME and 12 PMEIs were detected in LF and not EF. In general there were more PMEIs expressed in the fibers. Specifically there were more PMEIs expressed in the LF than in the EF, which might indicate that the inhibitory activity of the PMEIs is low at early stages of fiber development (i.e. EF stage), when fibers actively synthesize secondary cell walls, and demethylesterification of the newly synthesized homogalacturonan is required. However, when the cell wall deposition ceases, in the late fiber stage, PMEIs are expressed, and so the PME activity diminishes. Seven PMEs and three PMEIs were expressed in the shoot apex (SA), but not in any other of the stem vascular tissues. Moreover, nine PMEs and six PMEIs were expressed in the early cortical (ECP) peel, but not the late cortical peel (LCP), and three PMEs and nine PMEIs were expressed in late cortical peel but not early cortical peel (Figure 3-4A and 3-5A). 13 PMEs and 14 PMEIs

were found only in reproductive tissues; and three PME and two PME1 were found only in vascular tissues (Figure 3-4B and 3-5B). Seven PMEs and six PME1s showed specific transcript expression in only one tissue/organ; these transcripts were detected in flower buds (four PMEs and two PME1s), flowers (one PME1), bolls (one of each), xylem (one PME1), roots (one PME1), late cortical peel (one PME), and shoot apex (one PME). Two of these might be important for phloem fiber development: *LuPME71*, which was detected only in LCP, a fiber containing tissue where secondary cell wall deposition and maturation is taking place, and *LuPME102*, only detected in the SA, where intrusive growth takes place (Figure 3-2).

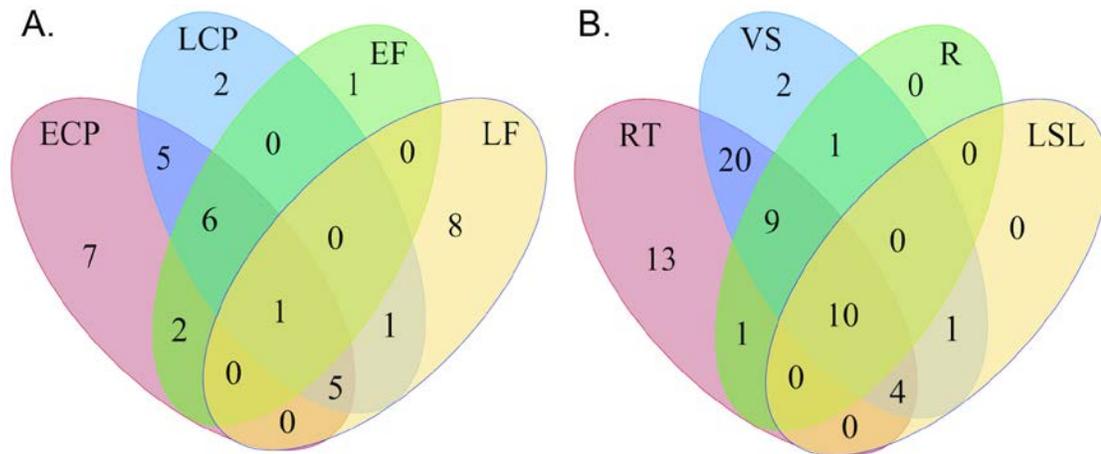
Previous publications have reported transcript expression patterns of specific LuPMEs. Al-Qsous and collaborators [52] found that the transcript abundance of *LuPME5* is higher than either *LuPME1* or *LuPME3* in hypocotyls. Our results were consistent with these observations: The calculated delta- $C_T$  of *LuPME5* was higher than *LuPME1* or *LuPME3* in all the tissues tested (Figure 3-2). Our transcript abundance data also showed that *LuPME5* was expressed in the shoot apex, while *LuPME3* was not, which could be correlated with the findings that showed that *LuPME5* transcript abundance was higher in the upper parts of the hypocotyl after two days of growth, while *LuPME3* was higher in the bottom of the hypocotyl [52]. Mareck and collaborators [53] found a very high transcript abundance of *LuPME3* in roots, as observed with the promoter fusion in tobacco [50], in which a GUS construct using *LuPME3* promoter was used to detect its expression in stems, roots and leaves. The expression was observed in the vascular tissues of roots, shoots and young leaves. This correlates with our results, as we detected *LuPME3* transcript expression only in roots, late fibers and the boll. Our study used

mature leaves, rather than young leaves, which may explain why we failed to detect transcript expression in this tissue, in contrast to Mareck and collaborators.



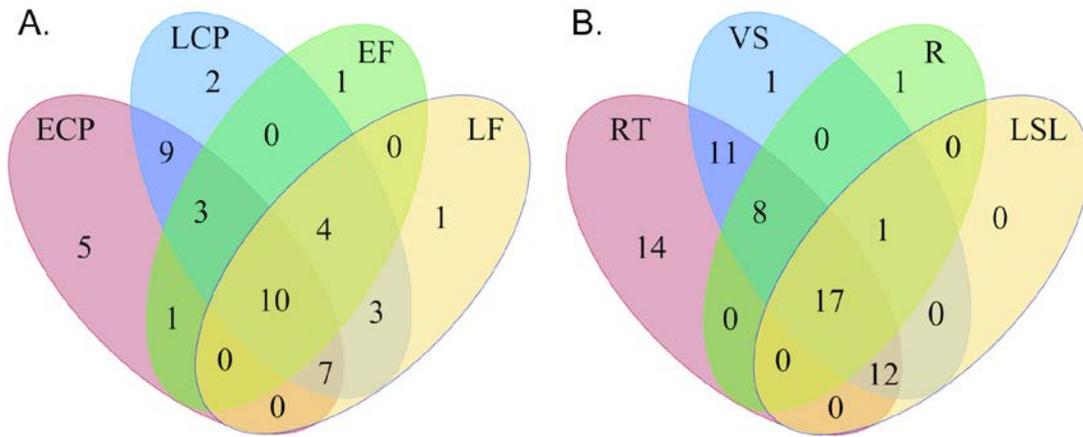
**Figure 3 - 3 Percentage of genes detected by qRT-PCR per tissue**

The percentage was calculated based on those genes that showed transcript expression in at least one tissue. FB: Flower buds; F: Flowers; B: Green boll; ECP: Early cortical peels; EF: Early fibers; LF: Late fibers; SA: Shoot apical meristem; X: Xylem; R: Roots; L: Leaves; SL: Senescent leaves.



**Figure 3 - 4 LuPME transcript expression in various tissues**

Venn diagram showing the number of LuPMEs detected in phloem-fiber containing tissues (A) and in tissue systems (B) ECP: early cortical peels. LCP: late cortical peels. EF: early fibers. LF: late fibers. RT: reproductive tissues. VS: Vascular tissues at shoot. R: Root. LSL: Leaves and senescent leaves.



**Figure 3 - 5 LuPMEI transcript expression in various tissues**

Venn diagram showing the number of LuPMEIs detected in phloem-fiber containing tissues (A) and in tissue systems (B) ECP: early cortical peels. LCP: late cortical peels. EF: early fibers. LF: late fibers. RT: reproductive tissues. VS: Vascular tissues at shoot. R: Root . LSL: Leaves and senescent leaves.

### 3.3.6 Protein subcellular localization

To be secreted to the cell wall via the Golgi apparatus and secretory pathway, PMEs and PMEIs require an N-terminal signal peptide or a transmembrane domain [90]. As shown in Figure 3-6, we found that 71/105 LuPMEs had a predicted transmembrane domain and/or signal peptide, and 81/95 LuPMEIs had a transmembrane domain and/or signal peptide (Tables 1 to 3). To further investigate subcellular localization, we used WoLF PSORT and Plant-mPLoc [99, 100]. Using Wolf PSORT we found 56/105 LuPMEs, and 71/95 LuPMEIs that were predicted to be secreted to either the cell wall or to the plasma membrane. Plant-mPLoc predicted 104/105 LuPMEs to be secreted to the cell wall, and 88/95 LuPMEIs were predicted to be secreted to the plasma membrane or the cell wall. In total all the LuPMEs and 93/95 LuPMEIs were predicted to be extracellular, based on protein subcellular localization software prediction and/or the presence of a signal peptide and/or transmembrane domain. The two LuPMEIs, LuPMEI11 and LuPMEI46, that were not predicted to be extracellular were predicted by both software tools to be localized to the nucleus.

### 3.3.7 Cleavage site

During their maturation, most Type-1 PMEs are proteolytically cleaved at one of two possible sites between the PMEI domain and the PME domain, before exiting the Golgi apparatus. These sites are designated binding motif 1 (BM1), and binding motif 2 (BM2) (Figure 3-1), and are separated by between 11–32 amino acids in Arabidopsis [91]. A conserved cleavage site consisting of four residues with the pattern [RKQ][RKEHLN][LDMI][LMAKR] was previously defined by analysis of Arabidopsis PMEs [90, 91]. We identified this pattern at a single site in each of 25/60 of the predicted flax Type-1 PMEs. Moreover, 19/60 flax Type-1 PMEs had two sites that matched the previously defined pattern, and these were separated by between 14 and 33 residues. In 3/60 of the LuPMEs, a novel tetrapeptide motif (RRKL or GRLL) was found in the place of the conserved pattern in BM2. Other novel motifs were also found, but these were all accompanied by a conserved motif in the other binding site, we found RKVA and RRLW in BM1, and REYL and RRFL in BM2. In 13/60 Type-1 PMEs, a cleavage site motif was not found (Table 3-1). Figure 3-7 shows the distribution of sizes of the mature proteins of PMEs and PMEIs (after signal peptide and/or pro-region removal, if present).

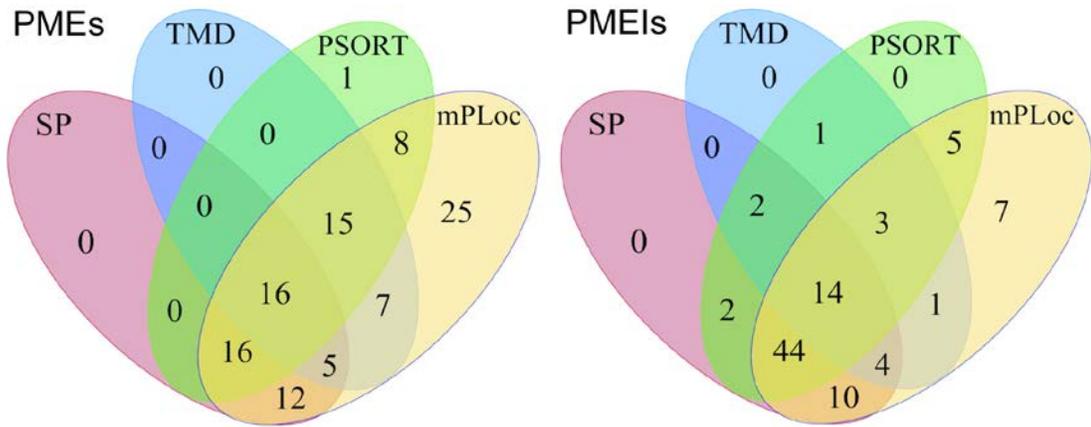
There was no significant difference (Fisher's Exact Test  $p > 0.05$ ) in the distribution of the cleavage site features between the 45/60 predicted PMEs for which evidence of transcription has been found, and the 15/60 predicted PMEs for which no evidence has been found. The inability to cleave out the PMEI domain (pro-region) would presumably prevent the export of the PMEs to the cell wall, according to Wolf and collaborators [91], who showed that unprocessed Type-1 PMEs are retained in the Golgi apparatus. Nevertheless, LuPME5 was detected in both the unprocessed and processed forms in the

cell walls of flax cell cultures [52], and LuPME3 was only detected in the unprocessed form in flax seedlings and callus [53]. This raises the possibility that the processing of Type-1 LuPMEs might be dispensable for the proper functioning or at least localization of the protein.

### 3.3.8 Isoelectric point

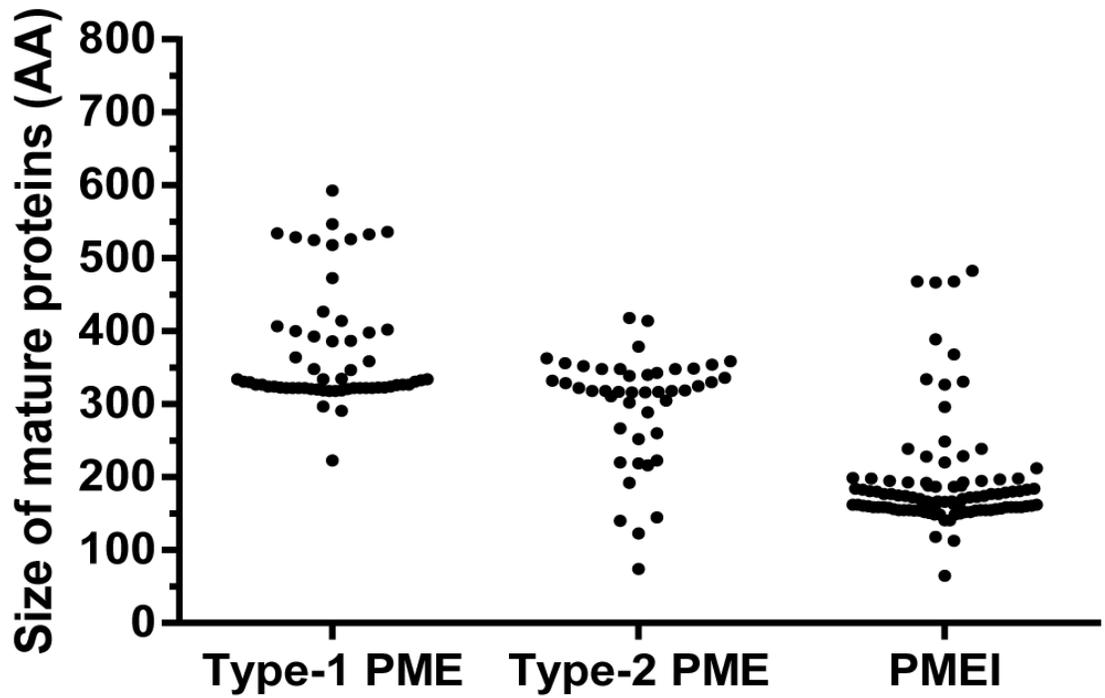
The isoelectric point is one of the factors that influences the action of PMEs (i.e. random, acidic pI, or blockwise, alkaline pI [33]) and so can facilitate either stiffening or loosening of the cell wall. Consequently, the prediction of the pI of the proteins contributes to the definition of their physiological role in the plant. We calculated the pI for the mature proteins (i.e. with any signal peptide or pro-region removed) (Figure 3-8). Most of the PMEIs (51) had an acidic pI, while 26 had a basic pI (pH  $\bar{x}$ : 6.48,  $\sigma_x$ : 1.77). On the other hand, most of the PMEs (70) had a basic (above pH8.0) pI (pH  $\bar{x}$ : 8.26,  $\sigma_x$ : 1.46.), while just a few (13) had an acidic pI, below 6: Out of these, four showed tissue specificity: Two in FB: *LuPME93* and *LuPME37*, and two in fiber containing tissues: LCP (*LuPME71*), and SA (*LuPME102*). As an acidic pI would lead to random demethylesterification, which ultimately could lead to cell wall loosening; it could be expected that when *LuPME71* is expressed in the LCP, it decreases the rigidity of the connections between fibers, while *LuPME102* might enhance fiber growth as it could loosen the connections between parenchyma cells to facilitate fiber intrusive growth. Only one of the genes whose transcript expression was detected in the EF or LF had an alkaline pI, this was LuPME66 (pI 5.87), which was expressed in reproductive tissues, xylem, LCP and LF. This could have a role similar to the described for *LuPME71*.

The wide range of predicted isoelectric points for the mature PME<sub>s</sub> (pH 4.75 to 11.25; Tables 1, 2) is consistent with previous reports from flax. Gaffe and collaborators [113] tested the PME activity from cell walls of flax calli; they found isoforms at pI 5.5, 7, 7.3, 7.8, 8.8, and 10. Mareck and collaborators [114] found a similar range of isoforms in flax calli; they found PME<sub>s</sub> with pI<sub>s</sub> 4.3, 4.8, 6.3, 7.1, 7.6, and 9.6. Alexandre and collaborators [115] found 2 PME<sub>s</sub> in hypocotyls, at pH 8 to 9, and at pH 9.5 to 10. Al-Qsous and collaborators [52] observed PME activity in the hypocotyls at 5 different isoelectric points, from pH 7.0 to 10.0. Finally, Mareck and collaborators [53] also found a similar pattern of the isoforms in epicotyls, cotyledons, hypocotyls and roots; they found two neutral, four basic and one strongly basic PME isoform.



**Figure 3 - 6 Subcellular localization of PMEs and PMEIs**

Number of PMEs and PMEIs with transmembrane domains, signal peptides, and predicted to be secreted to the cell wall or plasma membrane using Wolf PSORT and Plant-mPLoc.



**Figure 3 - 7 Size of mature PMEs and PMEIs proteins.**

The signal peptide and/or cleavage site, if present, were removed.

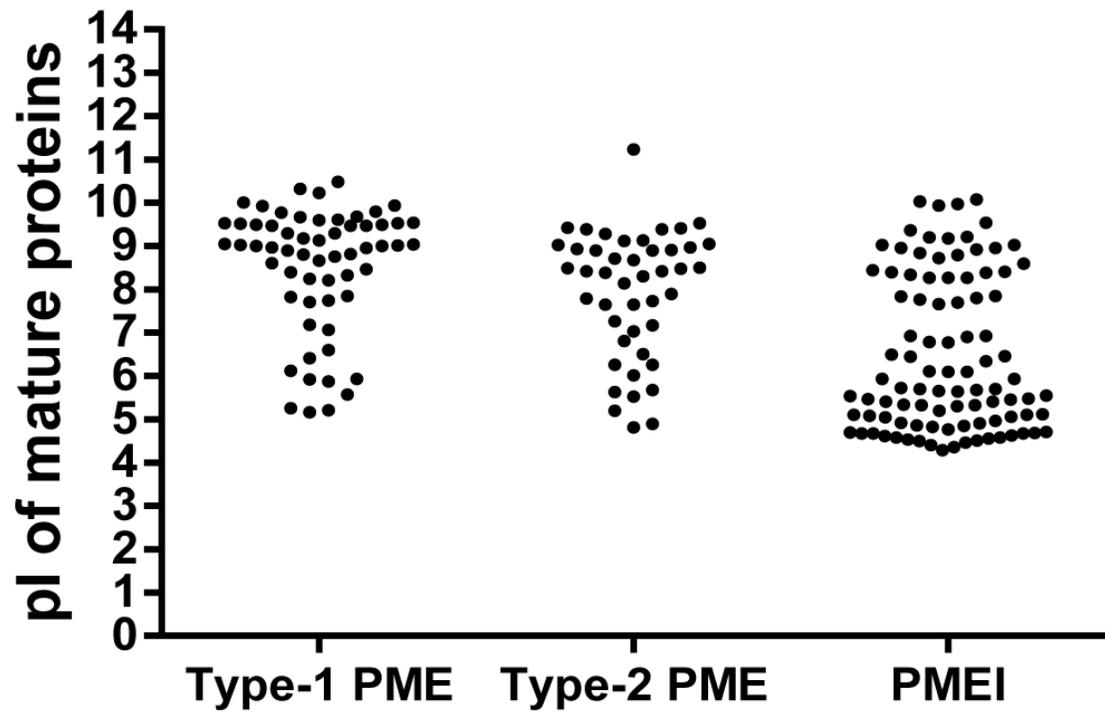


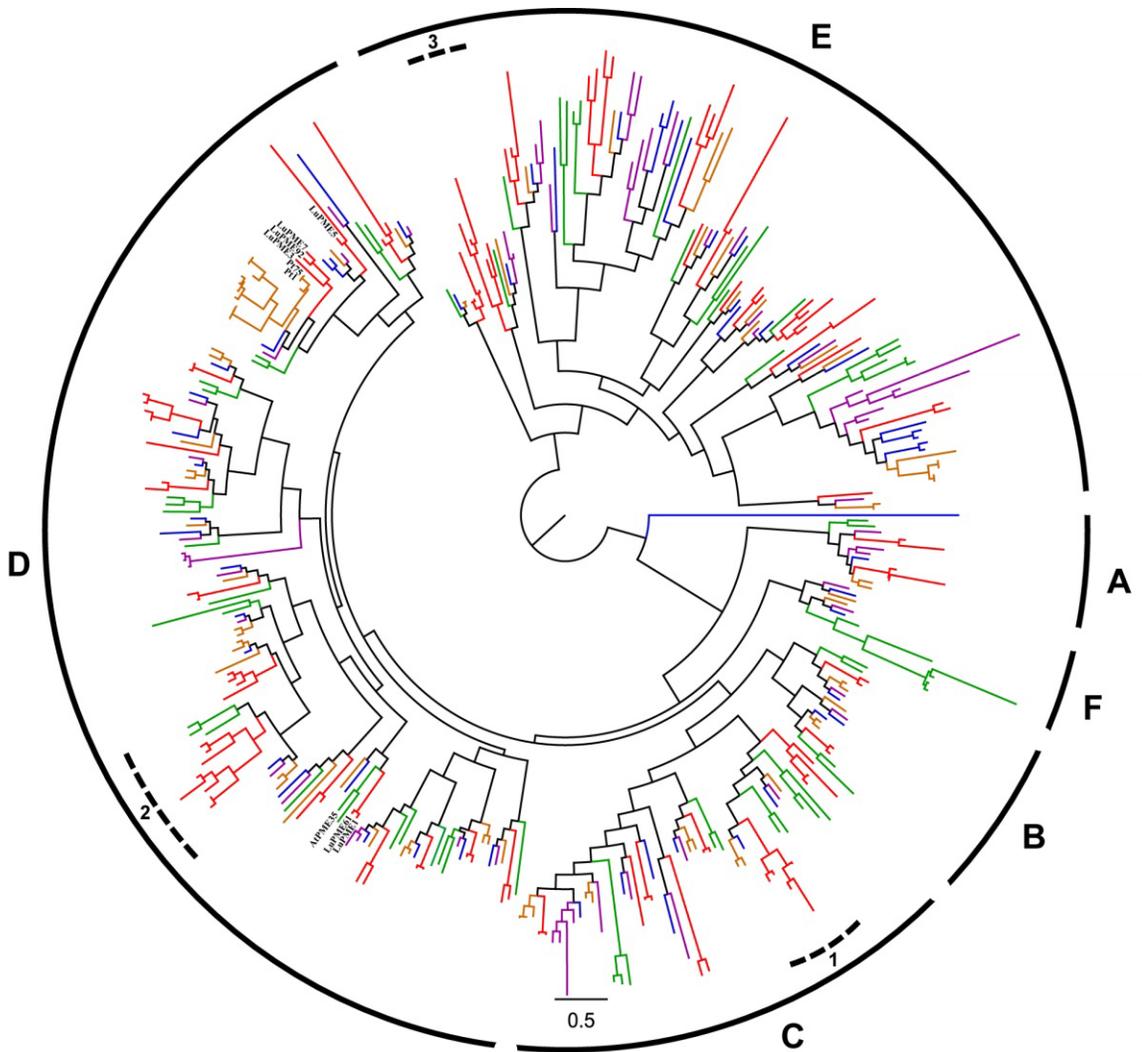
Figure 3 - 8 Isoelectric point of mature PMEs and PMEIs proteins.

The signal peptide and/or cleavage site, if present, were removed.

### 3.3.9 Phylogenetic analysis

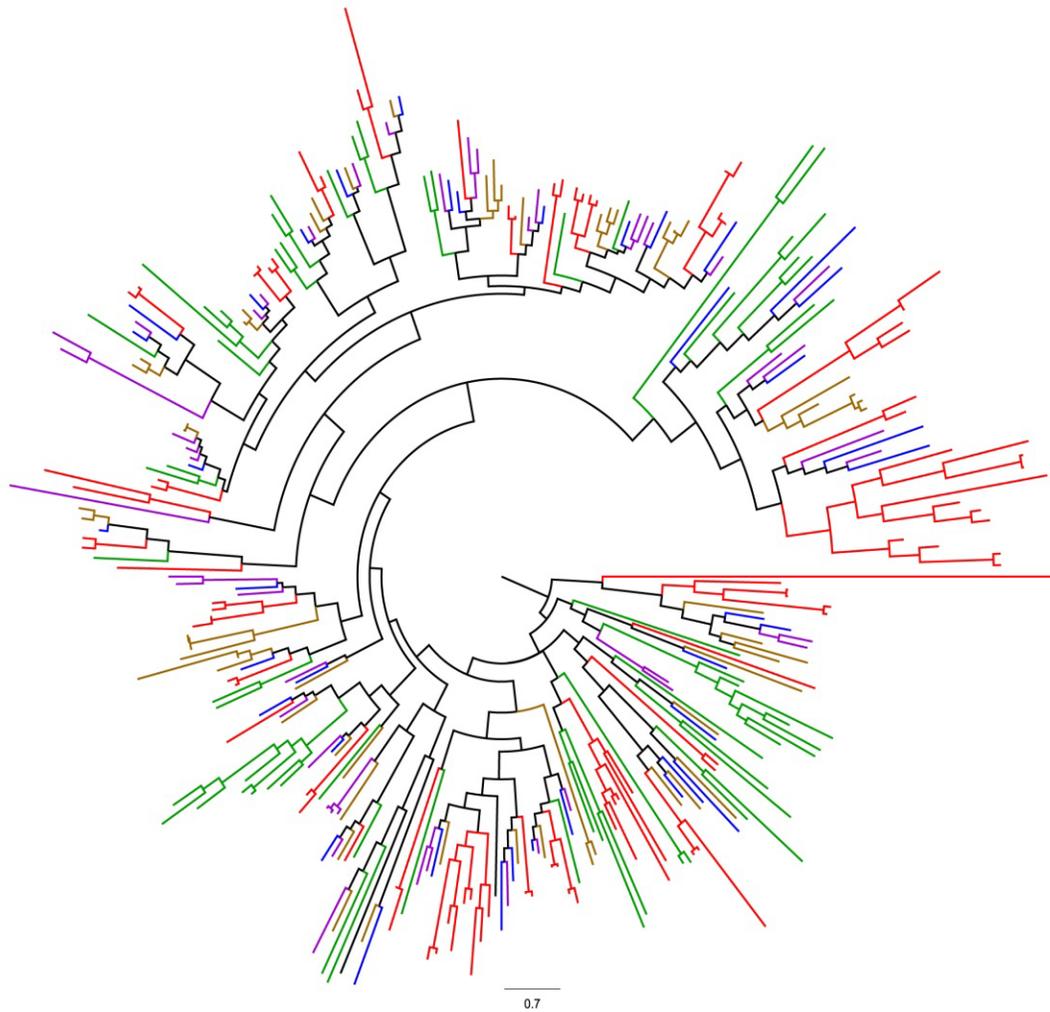
To classify the predicted LuPMEs and LuPMEIs on the basis of amino acid sequence similarity and inferred evolutionary relationships, we aligned their amino acid sequences with predicted PME and PMEIs from four other angiosperms: cassava (*Manihot esculenta*), castor (*Ricinus communis*), poplar (*Populus trichocarpa*), and Arabidopsis. These species were chosen because Arabidopsis is a well-characterized model organism, and castor, cassava, and poplar are in the same taxonomic order (Malpighiales) as flax, and whole-genome assemblies are available for each of these species. Following alignment, maximum likelihood phylogenetic trees for PMEs (Figure 3-9) and PMEIs (Figure 3-10) were constructed. Based on the groups defined by Louvet and collaborators [116] for Arabidopsis PMEs, the branch length, and the bootstrap values (online at <http://www.biomedcentral.com/content/supplementary/1471-2164-14-742-s4.jpeg>), six major monophyletic groups of PME could be defined (A, B, C, D, E, and F which correspond to groups 3, 1, 2, 1, 4, and 1 respectively in Louvet *et al* [116]). Group A included five LuPMEs; three of them were Type-1 PME, and none of them had a cleavage recognition site. The PMEs in all the organisms of group B were Type-1 PMEs, and all the LuPMEs in this group had a cleavage recognition site, and a transmembrane domain, but no predicted signal peptide. Group C was composed of Type-1 and Type-2 PMEs. All of the Type-1 PMEs had a cleavage recognition site, and none of the Type-2 PMEs had either a signal peptide or transmembrane domain. Group D contained the previously described LuPME1, LuPME3, and LuPME5, we did not find any characteristic defining this group based on the parameters we described above (e.g. Table 3-1). The PMEs of all the species in group E were Type-2 PMEs. Finally group F contained PMEs of all the species analyzed except flax.

In the PMEI phylogenetic tree (Figure 3-10), groups were distinguished by very low bootstrap values in the base nodes (Online at <http://www.biomedcentral.com/content/supplementary/1471-2164-14-742-s5.jpeg>), making sub-classification of PMEIs ambiguous. Furthermore, we did not find any common sequence features that distinguished subgroups of PMEIs from each other.



**Figure 3 - 9 Maximum likelihood dendrogram of PMEs in flax and related species**

The main groups and some subgroups are shown. The previously reported LuPMEs and the homologous LuPMEs to PttPME1 and AtPME35 are labeled. Red: *Linum usitatissimum*; Purple: *Manihot esculenta*; Blue: *Ricinus communis*; Orange: *Populus trichocarpa*; Green: *Arabidopsis thaliana*. 100 bootstraps and 2 search-replicates (three with bootstrap values and IDs available online at <http://www.biomedcentral.com/content/supplementary/1471-2164-14-742-s4.jpeg>)



**Figure 3 - 10 Maximum likelihood dendrogram of PMEIs in flax and related species**

Red: *Linum usitatissimum*; Purple: *Manihot esculenta*; Blue: *Ricinus communis*; Orange: *Populus trichocarpa*; Green: *Arabidopsis thaliana*. 100 bootstraps and 2 search-replicates (tree with bootstrap values and IDs available online at <http://www.biomedcentral.com/content/supplementary/1471-2164-14-742-s5.jpeg>).

Phylogenetic analysis of LuPMEs and LuPMEIs grouped 43 pairs of LuPMEs (out of 105 genes in total) and 39 pairs of LuPMEIs (out of 95 genes in total) at the terminal nodes of the tree (Figures 3-9 and 3-10). The remaining genes, 19 PMEIs and 17 PMEs, did not have obvious paralogs. The pairs of genes were confirmed by reciprocal BLASTn and BLASTp to test if they were the best BLAST match. 38/43 pair of LuPMEs and 38/39 pairs of LuPMEIs were found to be the best match to one another. This suggests these probably originated from a recent whole genome duplication event believed to have occurred 5 to 9 MYA [1]. Indeed, for the LuPMEIs, the estimated time of divergence of presumptive paralogs was calculated to be 4.5-8.4 MYA (Appendix Table A3-2), and for the LuPMEs estimate was 6.4 -11.9 MYA (Appendix Table A3-3). Lineage-specific expansion of groups of PMEs or PMEIs may indicate that selection had occurred for particular functions in flax. Expansion of at least three sub-groups (C1, D2, and E3) of PMEs was evident in the ML tree (Figure 3-9).

We identified PMEs genes that have been associated with stem development in previous studies in other species, and found their presumptive homologs in flax. Siedlecka and collaborators [57] found that when the transcript abundance of PttPME1 (accession no. AJ277547) in hybrid aspen (*Populus tremula* × *tremuloides*) increased, the fiber elongation decreased, and conversely, when the transcript abundance of the gene was low it stimulated fiber elongation. The closest PMEs in *P. trichocarpa* for PttPME1 are Pt1 (POPTR\_0001s16250.1) and Pt75 (POPTR\_0214s00200.1). Based on the phylogenetic tree, we found three LuPMEs that were closely related to Pt1 and Pt75; they were LuPME7, LuPME92, and LuPME3 (Figure 3-9; Table 3-1), which were all type-1 PME, as PttPME1. The study of these genes will be important in future studies as they

may regulate fiber elongation the same way as in poplar, as the fibers of both plants elongate intrusively. Hongo and collaborators [77] found that the type-1 PME *AtPME35* (*At3g59010*), has a role in strengthening the inflorescence stem of *Arabidopsis*. By mediating the demethylesterification of the primary cell wall of cortical cells and interfascicular fibers, this gene was suggested to have a blockwise demethylesterification action. We inferred that the common ancestor of *LuPME61* and *LuPME1* is the likely ortholog to *AtPME35* (Figure 3-9; Table 3-4). Both *LuPME61* and *LuPME1* have basic pIs (9.42 and 8.82, respectively), similar to the pI of the mature protein of *AtPME35* (pI 8.70), so it is possible that they also have a blockwise demethylesterification activity similar to *AtPME35*, which leads to stiffening of the cell wall. The study of loss-of-function mutants for these genes in flax might identify informative phenotypes related to stem development.

**Table 3 - 4 Genes putatively associated with fiber development during elongation (\*), thickening (¥), and thickening and maturation (§) processes**

	<b>Gene</b>	<b>Expression pattern</b>	<b>Homologous to</b>
§	LuPME1	L-E	AtPME35
§	LuPME3	O-X; L-E	PttPME1
§	LuPME5	L-E	
¥	LuPME7	O-X; E-L	PttPME1
¥	LuPME8	E-L	
§	LuPME9	L-E	
¥	LuPME10	O-X; E-L; EC-LC	
¥	LuPME17	O-X; E-L; EC-LC	
¥	LuPME18	O-X; E-L	
*	<u>LuPME21</u>	SA-S	
§	LuPME24	L-E	
¥	LuPME27	O-X; EC-LC	
§	LuPME31	O-X; L-E	
§	LuPME32	OV	
¥	LuPME34 and LuPME59	E-L	
¥	LuPME39	O-X; E-L	
§	LuPME40	O-X; L-E	
¥	LuPME43	O-X; EC-LC	
¥	LuPME44	O-X; E-L	
*	<u>LuPME46</u>	SA-S	
§	LuPME48	O-X; L-E	
¥	LuPME50 and LuPME73	E-L	
§	LuPME53	L-E	
§	LuPME54	O-X; L-E	
¥	LuPME55	EC-LC	
§	LuPME56	O-X; L-E	
§	LuPME61	O-X; LC-EC	AtPME35
§	LuPME66	L-E; LC-EC	
*	<u>LuPME67</u>	SA-S	
§	LuPME68	O-X; L-E	
§	LuPME69	O-X	
*	<u>LuPME70</u>	SA-S	
§	LuPME71	OV; O-X; LC-EC; LCP	
¥	LuPME73	O-X; EC-LC	
*	<u>LuPME79</u>	SA-S	
§	LuPME81	O-X	
¥	LuPME85	O-X; EC-LC	
§	LuPME91	L-E	
*	<u>LuPME92</u>	SA-S	PttPME1
*	<u>LuPME102</u>	OV; SA-S	
¥	<u>LuPME103</u>	O-X; EC-LC	
¥	<u>LuPME105</u>	O-X; EC-LC	

**Table 3-4. Genes putatively associated with fiber development during elongation (\*), thickening (¥), and thickening and maturation (§) processes. Continued**

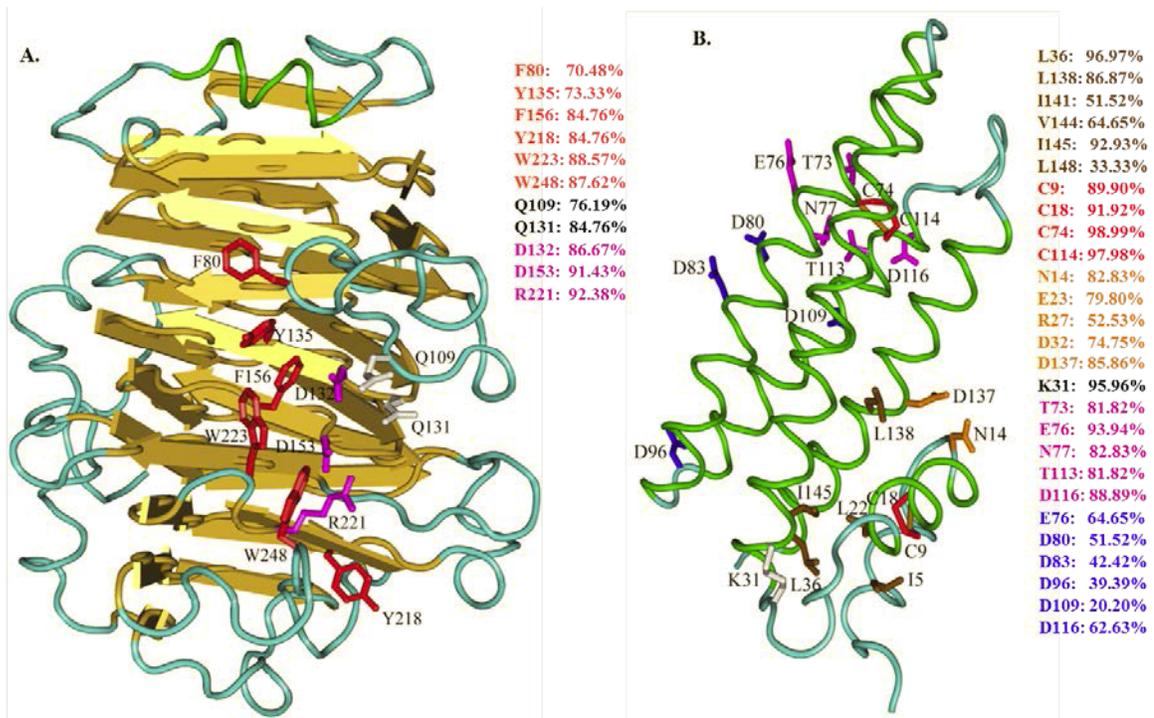
	<b>Gene</b>	<b>Expression pattern</b>	<b>Homologous to</b>
§	LuPMEI1	O-X; LC-EC	
§	LuPMEI17	LC-EC	
§	LuPMEI19	LC-EC	
§	LuPMEI23	O-X	
§	LuPMEI24	LC-EC	
§	LuPMEI25	LC-EC	
*	LuPMEI27	SA-S	
§	LuPMEI28	LC-EC	
¥	LuPMEI29	EC-LC	
§	LuPMEI38	L-E	
§	LuPMEI40	O-X	
§	LuPMEI41	O-X; L-E; LC-EC	
¥	LuPMEI43	EC-LC	
¥	LuPMEI44	E-L	
*	LuPMEI45	SA-S	
§	LuPMEI51	L-E; LC-EC	
¥	LuPMEI54	O-X; EC-LC	
§	LuPMEI55	OV	
§	LuPMEI56	O-X; L-E	
*	LuPMEI59	SA-S	
§	LuPMEI63 and LuPMEI68	O-X; L-E	
¥	LuPMEI65	O-X; E-L; EC-LC	
¥	LuPMEI66	E-L	
§	LuPMEI68	O-X	
§	LuPMEI71	O-X	
¥	LuPMEI72	EC-LC	
§	LuPMEI74	L-E	
¥	LuPMEI76	EC-LC	
§	LuPMEI82	L-E	
§	LuPMEI83	OV	
¥	LuPMEI85	O-X; E-L	
§	LuPMEI86	L-E	
§	LuPMEI87	O-X	
¥	LuPMEI89	E-L	
§	LuPMEI90	O-X; L-E; LC-EC	
§	LuPMEI94	L-E	
§	LuPMEI95	L-E	

If two genes are shown in one line it means the primers used amplified both genes. (OV) expressed only in vascular tissues; (SA-S) present in SA but not in the rest of the stem (EF, LF, ECP, LCP); (O-X) present in outer tissues (cortical peel and fibers) but not xylem; (L-E) present in LF but not EF; (E-L) present in EF but not LF; (EC-LC) present in ECP but not LCP; (LC-EC) present in LCP but not ECP; (SA) only present in SA; (LCP) only present in LCP. The genes underlined were found to be "SA-S" using the qRT-PCR data, however, according to RNAseq assembled reads they are present in other parts of the stem in addition to SA. As they were only detected in SA with qRT-PCR it is expected that their expression is higher in the SA.

### 3.3.10 Conserved residues in PMEs

We searched the predicted LuPMEs for conserved amino acid residues previously reported to be important for PME function. These amino acids were identified in a Type-1 PME from tomato (*Solanum lycopersicum*), (PME1\_SOLLC, SwissProt P14280), and the positions listed here refer to that sequence [47]. Three residues are proposed to be catalytic residues: D132, D153, and R221. Two residues, Q109 and Q131, are believed to stabilize the intermediate that is formed after nucleophilic attack on the carboxymethyl group. Finally, six aromatic amino acids at conserved positions are required for substrate binding (F80, Y135, F156, Y218, W223, and W248), and of these F80, Y135, and W223 are possible targets of the PMEIs. We searched the predicted LuPMEs for all eleven of the residues that have been proposed to be critical to PME function (Table 3-1, 3-2). The most highly conserved residues were the catalytic residues: D132, D153, and R221, which could be identified in (91/105), (96/105), and (97/105) of the predicted LuPMEs, respectively (Figure 3-11A). The aromatic residues responsible for substrate binding were also highly conserved, where any aromatic residue is considered to be a conserved residue in comparison to the substrate binding positions defined in PME1\_SOLLC (i.e. F80, Y135, F156, Y218, W223, and W248). The most highly conserved of these were W223 (93/105) and W248 (92/105), and the conservation of other substrate binding residues was also high: F80 (74/105), Y135 (77/105), F156 (89/105), and Y218 (89/105). Among these aromatic residues there were three positions (F156, W223 and W248) in which the identity (and not merely similarity) of the aromatic amino acids was also highly conserved: at F156, F was found in 88/89 of the LuPMEs that had an aromatic residues at that position; at W223, 90/93 aromatic residues were W, and in W248 91/92 aromatic residues were W. F80, Y135, and W223 are responsible of substrate binding of the PME

and also interact with the PME1. F80 generated the highest number of contacts (17 in total) with the PME1 [47]. Out of all the aromatic residues, position F80 was the least conserved, followed by Y135, meanwhile W223 showed the highest conservation. This might imply that F80 and Y135 are not fundamental for binding to the substrate, as it might be W223, and on the other hand, lacking these residues may limit inhibition by the PMEIs.



**Figure 3 - 11 Tertiary structures showing conserved residues of PMEs and PMEIs and their level of conservation in flax.**

Structures shown are for PME1\_SOLLIC (A) and PME1\_ACTDE (B), PDB: 1XG2. For the residues involved in bundle-hairpin interface in PME1\_ACTDE only those residues with conservation higher than 80% are labeled in the structure.

### 3.3.11 Conserved residues PMEI

PMEIs have four antiparallel alpha-helices ( $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  and  $\alpha_4$ ) arranged in an up and down topology, and three short alpha-helices ( $\alpha_a$ ,  $\alpha_b$ , and  $\alpha_c$ ) at the N-terminus [47]. Two groups have studied PMEI structure in detail. Di Matteo and collaborators [47] identified residues with important roles in the structure and activity of a PMEI from kiwi (*Actinidia deliciosa*; PMEI\_ACTDE, SwissProt accession P83326), focusing on the interaction with a PME. They identified cysteine residues that generated disulfide bridges connecting helices  $\alpha_a$  and  $\alpha_b$  (C9 and C18), and helices  $\alpha_2$  and  $\alpha_3$  (C74 and C114). Furthermore, PMEI residues T73, E76, N77, T113, and D116 were found to allow interaction of the PMEI with three of the aromatic residues of the PME (F80, Y135, and W223). An acidic patch was formed by three conserved residues on both the  $\alpha_2$  helix (E76, D80, and D83) and on the  $\alpha_3$  helix (D96, D109, and D116). Finally, salt bridges occurred between the PMEI residues D116 and E76 and PME residues K224 and R81. Hothorn and collaborators [46] also studied the important residues for the PMEI activity and structure, using Arabidopsis PMEI-1 (PMEI1\_ARATH, SwissProt accession number Q9LNF2) as a model. They identified a disulfide bridge connecting helices 5 and 6 ( $\alpha_2$  and  $\alpha_3$  in Di Matteo and collaborators[47]), formed by residues C71 and C111. They also identified a residue responsible for the N-terminal orientation (P28) that is located between the three N-terminal  $\alpha$ -helices and the four  $\alpha$ -helices towards the C-terminal; and several residues contributing to the bundle-hairpin interface.

We searched the predicted PMEIs of flax for the critical residues identified by Di Matteo and collaborators [47] and Hothorn and collaborators [46]. Our results (Table 3-3) are presented using as reference positions the mature PMEI\_ACTDE protein. The

residues in LuPMEIs with the highest conservation were C74 and C114 (both 94/95), which generate a disulfide bridge (Figure 3-11B). The conservation of the other two cysteines, C9 and C18, which stabilize the protein by hydrophobic interactions, was slightly lower, (86/95) and (88/95), respectively. The conservation (i.e. similarity, not necessarily identity) in the polar PMEI residues that interact with the aromatic PME residues, F80, Y135, and W223, was higher than 80% in all the cases: T73 (78/95), E76 (89/95), N77 (78/95), T113 (78/95), and D116 (85/95). 87.4% (83/95) of the PMEIs had conservation of at least 3 out of the 5 polar residues. On the other hand, the Aspartic acid (D), and Glutamic acid (E) residues that are predicted to be important in the generation of an acidic patch on alpha helices 2 and 3, had a low conservation (i.e. similarity, not necessarily identity), 47.5% on average. The conservation of the residues contributing to the bundle-hairpin interface was also low. Out of the 12 non-polar residues analyzed, only five were conserved (i.e. similarity, not necessarily identity) in more than 80% of the LuPMEIs, those are I5 (82/95), L22 (89/95), L36 (92/95), L138 (84/95), and I145 (89/95), the rest (I8, L17, A21, L33, I141, V144, and L148) were below 70% of conservation.

### 3.3.12 Gene expression and conserved residues

We tested whether there was any correlation between the transcript expression evidence we obtained and the presence of the critical residues described above (Tables 3-1 to 3-3). In general, genes that are not expressed may accumulate mutations, including mutations in residues critical to the normal function of the protein. We found that when we analyzed the 33 critical residues in the LuPMEIs as a group, the conservation of these residues among the expressed genes (73.2%) was significantly higher than in the genes

for which transcript expression was not detected (63.9%) (Fisher's Exact Test,  $p < 0.05$ ).

Taking together the 11 critical residues studied in the LuPMEs, we found that in the expressed genes, the level of conservation (86.8%) was significantly higher than the conservation observed in the non-expressed genes (75.3%) (Fisher's Exact Test,  $p < 0.05$ ).

However when we individually analyzed the residues we found that in the LuPMEs only three out of the 11 residues (Q109, F156, and R221) showed significantly higher conservation in the expressed genes when compared to the genes without evidence of transcript expression (Fisher's Exact Test,  $p < 0.05$ ). In the remaining eight there was no significant difference (Fisher's Exact Test,  $p > 0.05$ ). The critical residue that showed the greatest change in conservation in relation to transcript expression was Q109, which was found in 83.1% of expressed LuPMEs, but only 57.1% of LuPMEs with no evidence of transcript expression. In the same way, we individually analyzed the LuPMEIs conserved residues, and found that only 5 of the 33 residues (L36, V144, C18, R27, and D83 ) have a significant higher conservation in expressed genes respect to non-expressed genes (Fisher's Exact Test,  $p < 0.05$ ). In the remaining 28 residues there was no significant difference (Fisher's Exact Test,  $p > 0.05$ ). The two greatest differences were observed in two of the residues contributing to the bundle-hairpin interface: V144 was conserved in 71.1% of the expressed LuPMEs, but only 33.3% of non expressed genes, and R27 from 59.0% to 16.7%. Interestingly, the change in conservation was very different between the two pairs of cysteines of the PMEIs, which generate the disulfide bridges, C74 and C114 conservation changed both from 98.8% in expressed genes, to 100% in both, in non-expressed genes. Conversely, the conservation at positions C9 and C18 was reduced drastically, although not significantly in C9, from 92.8% and 96.4% in expressed genes, to 75.0% and 66.7%, respectively in non expressed genes. This might indicate that there

is more evolutionary pressure on residues C74 and C114, which indicates that they might be more important for the structure of the protein.

The residue at position K31 (respect to PME1\_ACTDE [47]) or P28 (respect to PME1\_ARATH [46]) has been reported to affect the orientation of the N-terminal of the PMEIs. We found 13 different amino acids at this position. There were three prevalent residues in this position, they were P (19/95), A (18/95), and I (17/95) residues. Hothorn and collaborators [46], found that when they mutated P28 to Ala in PME1\_ARATH, the inhibitory activity of the protein diminished. So the 18 PMEIs with alanine could have a decline in activity.

### **3.4 Conclusion**

PMEs, regulated in part by PMEIs, modify cell and tissue properties by demethylesterification of homogalacturonan within cell walls and the middle lamella. Flax phloem fibers elongate intrusively by penetrating the middle lamella of the cells below and above them. This process requires the loosening of the middle lamella, the strengthening of the cell wall of the growing fiber, and then the creation of a new contact interface with the cell. In these processes, PMEs might be involved. Here we described 105 putative PMEs and 95 putative PMEIs in the flax genome, of which 77 and 83 genes, respectively, had evidence of transcript expression. The expression data obtained from the different tissues and organs allowed us to define a list of candidate genes that could play a role in fiber development.

## **4 Chapter 4: Characterization of pectinmethylesterases (PME) and pectinmethylesterase inhibitors (PMEI ) enriched during flax fiber development in (*Linum usitatissimum*)**

### **4.1 Introduction**

The demethylesterification of the cell wall plays a major role in the development of the bast fibers in flax. The flax genome contains 105 putative flax pectin methylesterases (LuPMEs) and 95 putative pectin methylesterase inhibitors (LuPMEIs), the majority of which (77 LuPMEs and 83 LuPMEIs) have been demonstrated to be transcribed in at least one of the following tissues and developmental stages: floral buds, flowers, green capsules, early cortical peels, early fibers, late fibers, shoot apices, xylem, roots, leaf, senescent leaves [117]. Having thus defined the LuPME and LuPMEI families, we can now more precisely characterize the expression and functions of these genes in the context of flax bast fiber development.

The heterologous expression of PMEIs has been successfully achieved in *Escherichia coli* and *Pichia pastoris*. The mature proteins (without the signal peptide) of the Arabidopsis PMEIs AtPMEI-1 and AtPMEI-2 were both expressed in *E. coli* strain Rosetta-gami™ (DE3) [118] and in *P. pastoris* strain X-33 [119] producing, in both cases, functional inhibitors. Also the complete and mature proteins of BoPMEI1, a PMEI from *Brassica oleracea* were effectively expressed in *E. coli* strain ER2566. On the other hand, the expression of PMEs has been more challenging and does not always produce functional proteins. The complete proteins of the type-2 PMEs QUARTET1 [55] and

AtPME31 [120] were successfully expressed in *E. coli* strains BL21(DE3) and JM101, respectively. However, the mature portion (removing signal peptide and pro-region) of a type-1 PME (At1g11580) was expressed in *E. coli* strain M15 but it was not functional, even though the purified native protein from Arabidopsis was functional [121]. One explanation for these results is that post-translational modifications, such as glycosylation, may be necessary for the correct activity of the protein as was reported for PMEs and PMEIs from kiwi fruit (*Actinidia chinensis*) [44, 122] and for PMEs from mandarin orange (*Citrus sp.*) [123].

Based on the preliminary identification of candidate fiber-related genes (Table 3-4), we selected 30 LuPMEs and LuPMEIs to assay by qRT-PCR in nine whole stem segments, and five stem peel segments representing various stages of fiber development. Three of the genes (*LuPME67*, *LuPME79*, and *LuPMEI45*) that showed peak transcript expression during fiber elongation were expressed in *E. coli*. Of these recombinant proteins, LuPMEI45 showed detectable activity *in vitro*. Our analysis allowed definition of a set of candidate PMEs and PMEIs with roles in fiber development during elongation, and during secondary cell wall deposition and maturation, and also a set of genes that could have important roles in xylem development.

## **4.2 Materials and Methods**

### **4.2.1 Plant material**

Plants were grown and tissues were harvested as described in Section 2.2.1. Segments (1 cm long) of whole stems and stem peels were harvested at approximately five weeks after germination.

#### 4.2.2 RNA extraction

Fifteen 1 cm-fragments per tissue were used for the RNA extraction. The RNA was extracted using Trizol extraction coupled with the RNeasy Plant Mini Kit (QIAGEN). Tissue was ground, and 2 ml of Trizol (Sigma) were added, followed by incubation at 60°C for 5 min with vortexing. The supernatant was transferred to a new tube by centrifugation at 12000 rcf at 4°C for 15 min. 0.2 volumes of chloroform were added, mixed, and centrifuged at 12000 rcf at 4°C for 20 min. The supernatant was obtained, and from here the extraction was coupled with the RNeasy Plant Mini Kit. 0.25 volumes of solution RLT plus 0.5 volumes of cold ethanol were added. These were applied to RNeasy columns, and the kit manufacturer's instructions were followed.

The RNA was tested for DNA contamination using a set of primers that flank an intron, producing differential product sizes for gDNA amplicons and for cDNA amplicons (Fw: 5'- TGCATATGCTCAGACCGACT-3', Rv: 5'- TGGTGTAGATTTTCGGAAGAGAC-3). The RNA quality and concentration were assessed using the Agilent 2100 BioAnalyzer (Agilent Technologies, Inc.).

#### 4.2.3 cDNA synthesis and quantitative real time PCR

1 µg of RNA was used to synthesize cDNA using RevertAid H Minus Reverse Transcriptase (Thermo Scientific) and oligo(dT)<sub>18</sub> primers following the manufacturer's protocol.

The Applied Biosystems 7500 Fast Real-Time PCR System was used to conduct quantitative real time PCR (qRT-PCR) on the stem peel tissues, in 96 well-plates. For the whole stem tissues, we used the Applied Biosystems 7900 HT Fast Real-Time PCR

System, in 384 well-plates. Three biological replicates and three technical replicates were used per sample. The cDNA was diluted 1:40. The 10  $\mu$ L sample mix consisted of 2.5  $\mu$ L of diluted cDNA, 0.4  $\mu$ M of each primer, and 1X MBSU buffer Tris (pH 8.3), containing KCl, MgCl<sub>2</sub>, Glycerol, Tween 20, DMSO, dNTPs, ROX as a normalizing dye, SYBR Green (Molecular Probes) as the detection dye, and an antibody-inhibited Taq polymerase. Primers used are listed in the appendix in Table A3-1

#### 4.2.4 Gene clustering based on expression

The STEM (Short Time-series Expression Miner) software package [124], was used to cluster the genes according to their transcript expression patterns. The negative of the dCT values were used as input in STEM, which was run using the “normalize data” option, so the values of the first tissue were transformed to 0. We also used a minimum correlation of 0.8, with a maximum of 9 model profiles for whole stem samples and 5 model profiles for stem peel samples, and also the minimum absolute expression change was adjusted to 2, so those genes in which there was less than a 4-fold difference between the highest and the lowest expression value were not used to generate the clustering.

#### 4.2.5 Heterologous expression

The coding regions of the genes, excluding the signal peptide of LuPMEI45 and LuPME79 (LuPME67 does not possess a predicted signal peptide), were used for heterologous expression. These were synthesized (Bio Basic Inc.) with codon optimization for *E. coli* (Appendix Additional file A4-1) and were introduced into pET22b(+) (Novagen, Madison, WI, USA) via the restriction sites XhoI and NcoI. This plasmid was then transformed into *E. coli* Rosetta-Gami B(DE3)pLysS (Novagen,

Madison, WI, USA). The empty pET22b(+) vector without inserts was used as a negative control in the various assays.

A single colony was grown overnight at 37°C in 2XYT medium plus chloramphenicol (34 µg/ml), tetracycline (12.5 µg/ml), kanamycin (15 µg/ml) and ampicillin (50 µg/ml). 1 mL was then transferred into 1 L of medium, and grown at 37°C until OD<sub>600nm</sub> 0.5-0.6, which was cooled on ice. IPTG at a final concentration of 1 mM was added, followed by growth for 18 hours at 20°C. Cells were pelleted at 4°C at 8000 rpm for 20 min. All the steps from here on were done at 4°C unless otherwise indicated. The pellet was then mixed with 5% v/v of the original volume of 300mM NaCl Tris HCl (the pH was at least one unit away from the predicted pI of the protein). This solution was left for at least 4 hours at -20°C, and was then sonicated at 55% for 30 seconds five times, with the intermediate tip of a Sonic Dismembrator model 300 (Fisher), with at least 1 min on ice between pulses. It was then centrifuged at 15000 rpm for 30 min at 4°C. The supernatant was incubated with 2% v/v of Ni-NTA agarose (QIAGEN) and rocked overnight prior to purification.

The His-tagged proteins were purified using a Poly-Prep chromatography column (0.8x4 cm) which was prepared by adding 2 volumes of 50 mM Tris-HCl and 300 mM NaCl at the selected pH. The protein extract was then added, and it was washed with two volumes of 50 mM Tris-HCl, 1.5 M NaCl, then with 50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, and then with 50 mM Tris-HCl, 300 mM NaCl, 40 mM imidazole. The protein was eluted with 5 ml of 50mM Tris HCl, 1 M NaCl and 250 mM imidazole, containing one cOmplete ULTRA protease inhibitor (Roche) tablet per 10 ml. Five 1 ml-

fractions were obtained, which were dialyzed against 50 mM Tris-HCl, 300 mM NaCl using a Amicon Ultra 3K centrifugal filter unit (Millipore).

#### 4.2.6 LC MS/MS

The proteins observed with the expected size in the Coomassie-stained polyacrylamide gel were confirmed by in-gel tryptic digestion and identification by LC MS/MS analysis in the Institute for Biomolecular Design (University of Alberta).

#### 4.2.7 PME activity

As described in section 2.2.2.

#### 4.2.8 PMEI inhibitory activity

The ability of recombinant PMEI to inhibit native PME activity in proteins extracted from flax stems was tested as performed by Raiola and collaborators [119]. For this purpose, PME activity was assayed as described in section 2.2.2. For inhibition assays, 10  $\mu$ L of flax cell wall proteins (396  $\mu$ g/mL) were mixed with 10  $\mu$ L of heterologous LuPMEI45 dialyzed solution (146  $\mu$ g/mL) and incubated for 30 min at room temperature, and then the mixture was added to a well in the assay plate (20  $\mu$ l per well).

### **4.3 Results**

#### 4.3.1 Selection of candidate genes

To identify genes that affect the development and extractability of flax fibers, we selected 21 LuPMEs and 9 LuPMEIs for further characterization (Table 4-1). The selection of these genes was based on two previous studies: (i) previously published Fluidigm qRT-PCR expression data (Table 3-4) that showed the selected genes to be

enriched in fiber-bearing tissues, and (ii) oligonucleotide microarray data that showed transcripts of the selected genes to be enriched in at least one of the points of the stem [125].

#### 4.3.2 Transcript expression profiling of selected LuPMEs and LuPMEIs

The expression profiling depicted in section 3.3.4 provided a general idea of the expression of the genes in the stem at early and late stages of development of the fibers, however, it did not provide sufficient resolution of the different stages of development of the fibers. Consequently, to determine the expression of genes of interest at key developmental stages along the stem, we measured the transcript expression of 21 LuPMEs and 9 LuPMEIs in 14 tissues described in section 2.3.1, Figure 2-2. The tissues sampled were whole stems at nine different positions along the developing flax stem, plus outer (fiber-bearing) stem peels at the five most basal positions. The expression was assessed from three biological replicates.

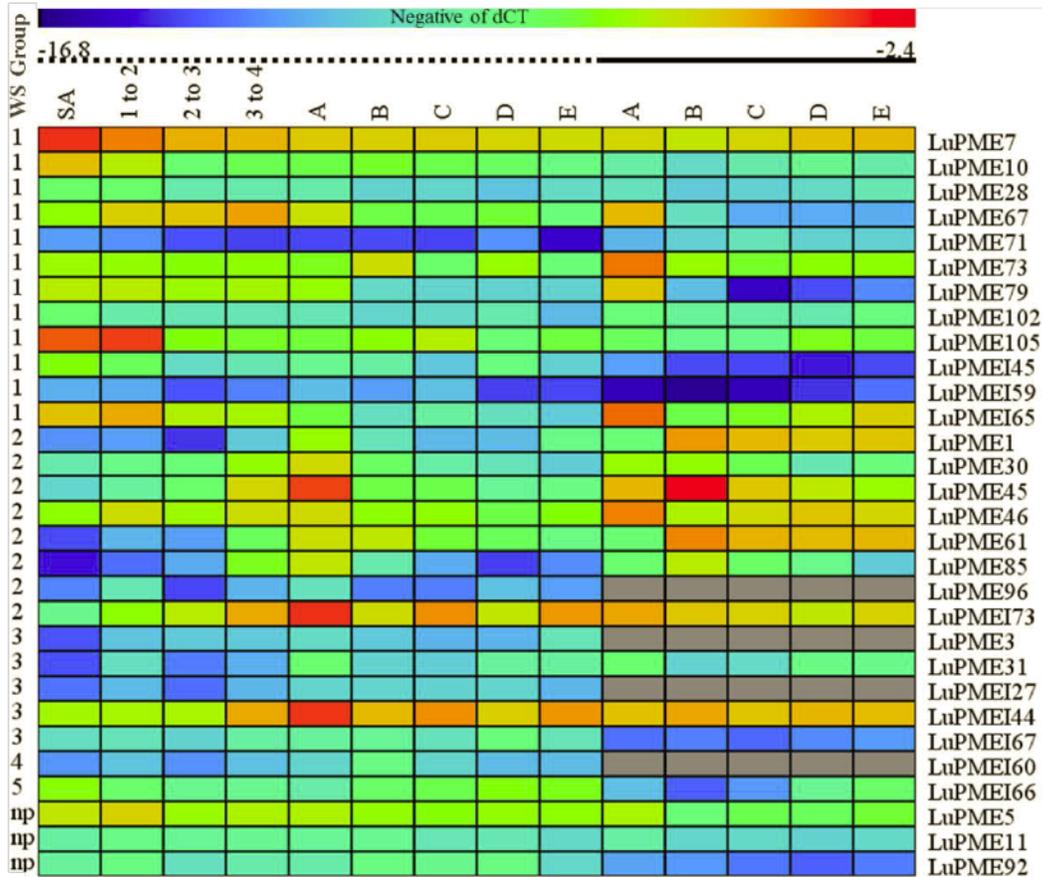
**Table 4 - 1 Selected LuPMEs and LuPMEIs for expression profiling at different fiber developmental stages.**

<b>Code assigned</b>	<b>Genome version 1.0 code</b>
LuPME1	Lus10025510
LuPME3	Lus10039314
LuPME5	Lus10013344
LuPME7	Lus10003933
LuPME10	Lus10004720
LuPME11	Lus10005587
LuPME28	Lus10011760
LuPME30	Lus10001466_Lus10001467_DPL
LuPME31	Lus10013416
LuPME45	Lus10017665
LuPME46	Lus10018103
LuPME61	Lus10026729
LuPME67	Lus10027737
LuPME71	Lus10029866_a_DPL
LuPME73	Lus10029867_g25305
LuPME79	Lus10031470
LuPME85	Lus10033621
LuPME92	Lus10037458_Lus10037459(g6229)_DPL
LuPME96	Lus10038918
LuPME102	Lus10041699
LuPME105	Lus10043035
LuPMEI27	Lus10016318
LuPMEI44	Lus10024595
LuPMEI45	Lus10024596
LuPMEI59	Lus10029877
LuPMEI60	Lus10030292
LuPMEI65	Lus10031138
LuPMEI66	Lus10031197
LuPMEI67	Lus10031483
LuPMEI73	Lus10032232

We tested three genes for their suitability as endogenous controls in the qRT-PCR assays. These three genes (*GAPDH*, *ETIF1*, *ETIF5A*) were selected for evaluation based on the results from Huis and collaborators [96]. We used BestKeeper software [126] to evaluate the expression stability of these genes in the tissues used in this study. *ETIF1* had the least overall variation with a standard deviation of 0.67, followed by *ETIF5A* (0.74) and *GAPDH* (0.75). The best correlation between BestKeeper index and candidate reference gene was for *ETIF5A* (0.995), followed by *GAPDH* (0.992), and then *ETIF1* (0.984). All three genes were therefore considered suitable as endogenous references for the qRT-PCR experiments described here, and the geometric mean of their Ct value was used to calculate the delta-C<sub>T</sub>.

We measured relative transcript abundance of 21 LuPMEs and 9 LuPMEIs in fourteen stem segments and stem peels (Figure 4-1). Transcripts of four genes (*LuPME3*, *LuPME96*, *LuPMEI27*, and *LuPMEI60*) could not be reliably detected in stem peel and these data are therefore not included in the results presented here. We measured the fold change among tissues in every gene between the highest and the lowest dCT values (Table 4-2), we found 10 genes in the whole stem and 6 genes in the stem peel that have at least a 20-fold difference between the highest and the lowest tissue. Among them the three highest fold changes in the whole stem tissues was observed in *LuPME85* (419-fold higher at A with respect to SA), *LuPME61* (307-fold higher at A with respect to SA), and *LuPME1* (191-fold higher at A with respect to 2-3). Among outer stem peels, the three genes with the greatest difference in transcript abundance were *LuPME79* (1085-fold higher at point A respect to C), *LuPME67* (153-fold higher at point A with respect to C), and *LuPMEI66* (37 times higher at point E with respect to B).

We also compared the expression of the genes in the same position in the whole stem and the stem peel (Table 4-3). Five genes (*LuPME67*, *LuPME79*, *LuPME92*, *LuPMEI45*, and *LuPMEI66*) showed an expression at least 20 times higher in at least one of the whole stem tissues in comparison to the corresponding position in the stem peel; all of these observations of differential expression were made in tissues below the snap point (B to E). Meanwhile, transcripts of four genes (*LuPME1*, *LuPME45*, *LuPME85*, and *LuPMEI65*) were at least 20 times more abundant in stem peel tissue than in whole stem tissue. In all the cases the 20-fold change in expression was observed in tissues below the snap point.



**Figure 4 - 1 Transcript expression of PMEs and PMEIs of interest from whole stem and stem peel tissues.**

Dotted line on top of whole stem tissues. Solid line on top of stem peel tissues. dCT was obtained by subtracting the geometric mean of the three endogenous controls used to the Ct value of the genes studied for every biological replicate. Here we show the average of the three biological replicates. The negative of the dCT value is shown in the graph, so a higher value represents higher transcript abundance.

**Table 4 - 2 Maximum fold change of expression between two tissues of selected PME and PMEIs.**

Gene	Highest fold change in whole stem				Highest fold change in stem peel			
	Fold change	CI	Max at	Min at	Fold change	CI	Max at	Min at
LuPME1	191.4	97.8 - 374.6	A	2to3	20.3	6.2 - 66.3	B	A
LuPME3	13.4	4.8 - 36.9	E	SA				
LuPME5	3.4	2 - 5.6	1to2	C	4.9	3.5 - 6.6	A	B
LuPME7	6.5	5.1 - 8.3	SA	E	2.1	2 - 2.3	E	B
LuPME10	15.7	9.4 - 26.1	SA	E	2.7	1.4 - 5.1	D	B
LuPME11	3	1.8 - 5	1to2	E	2.3	1.7 - 3.3	A	D
LuPME28	7.8	3.9 - 15.5	SA	D	2.3	1.6 - 3.4	E	B
LuPME30	45.1	33.4 - 61	A	E	9.2	6.3 - 13.4	A	D
LuPME31	40.2	3.1 - 527.3	A	SA	4.7	2.9 - 7.7	A	B
LuPME45	185.9	88 - 392.6	A	SA	19.5	12.9 - 29.4	B	E
LuPME46	4.6	2.9 - 7.3	A	D	5.2	3.6 - 7.4	A	B
LuPME61	307	171.5 - 549.4	A	SA	24.3	13.5 - 43.9	B	A
LuPME67	19.9	17.7 - 22.4	3to4	E	152.6	94.3 - 246.8	A	C
LuPME71	6	1.2 - 30.4	SA	E	3	1.2 - 7.8	C	A
LuPME73	7.9	5.1 - 12.3	B	E	11.8	10.9 - 12.7	A	C
LuPME79	26	7.3 - 92.3	1to2	E	1084.9	707.9 - 1662.6	A	C
LuPME85	419.4	110.4 - 1593.1	A	SA	31.5	22.5 - 44.2	B	E
LuPME92	3.4	1.7 - 6.6	D	E	2.6	2 - 3.4	A	D
LuPME96	15.8	7.8 - 32.3	1to2	2to3				
LuPME102	8.8	5.5 - 14	SA	E	2.2	0.8 - 6.1	A	D
LuPME105	49.3	10.2 - 237.8	1to2	D	3.6	3 - 4.3	D	C
LuPME127	6.2	3.3 - 11.7	B	2to3				
LuPME144	11.8	6.8 - 20.5	A	SA	1.4	1 - 2.1	B	C
LuPME145	17	7.1 - 40.8	SA	C	4.8	2.4 - 9.8	A	D
LuPME159	6.5	1.6 - 26.8	C	D	6.6	3.3 - 13.3	E	B
LuPME160	13.6	1.3 - 137.8	B	2to3				
LuPME165	91.8	39 - 215.7	1to2	E	18.2	7.4 - 44.9	A	B
LuPME166	6.5	3.9 - 11	SA	A	37.4	21.1 - 66.4	E	B
LuPME167	4.4	3.5 - 5.4	D	2to3	2	1.7 - 2.4	E	C
LuPME173	74.7	43.9 - 127	A	SA	2.9	2.2 - 3.9	A	D

The fold-enrichment between the tissue sample with the highest transcript abundance and the lowest transcript abundance was calculated for each gene. This calculation was done separately for whole stem (WS) and stem peel (SP) samples. Fold enrichment is shown in a linear scale and is the mean of 3 measurements from 3 biologically independent samples. The values not shown are genes that were not detected in those tissues. The confidence interval (CI) was calculated by using one standard deviation of the difference of the dCT between the two tissues compared.

**Table 4 - 3 Tissue enrichment expression of selected flax PME and PMEIs in whole stem compared to stem peels.**

Gene/Point	Fold change of genes with higher expression in whole stem									
	Fold change					CI of genes with higher expression in whole stem				
	A	B	C	D	E	A	B	C	D	E
LuPME1	3.8					1.6-9.3				
LuPME3										
LuPME5	1	3.6	1.8	2.6	1.6	0.7-1.5	2.6-4.9	1.2-2.6	1.6-4.3	0.9-2.6
LuPME7	1.3	1.6	1.2			1.1-1.5	0.7-3.7	0.9-1.5		
LuPME10	3.6	8.5	3.7	1.9	1.8	3-4.2	6-12.2	2.1-6.6	0.7-5	1.1-2.8
LuPME11	1.2	2.8	1.4	1.8		0.8-1.8	2.3-3.4	1.1-1.8	1.7-1.9	
LuPME28	1.4	1.3	1.2			0.9-2.2	1-1.5	0.9-1.6		
LuPME30	2.3					1.5-3.4				
LuPME31	1					0.1-9.1				
LuPME45	3.1					1.6-5.9				
LuPME46										
LuPME61	7.6					3.6-16.1				
LuPME67		5.7	17.8	22.9	9.5		5-6.4	13.4-23.7	14.5-36	6.5-14
LuPME71										
LuPME73		2.1		1.2			1.4-3.3		0.6-2.5	
LuPME79		2.1	21.8	8.6	4		1.5-2.9	13.9-34.3	4.2-17.4	1.0-16
LuPME85	5.8					1.8-18.8				
LuPME92	5.6	12.2	16.3	26.5	5.5	3.1-10.2	10.4-14.3	13.2-20.1	14-50	4.2-7.1
LuPME96										
LuPME102				1					0.7-1.5	
LuPME105	1.6	4.2	7.1		1	1.2-2.1	2.8-6.2	4.9-10.3		0.7-1.4
LuPME127										
LuPMEI44	4.2		1.9		1.5	2.8-6.1		1.5-2.4		1-2.4
LuPMEI45	9.1	21.3	6.9	56.4	8.3	5.1-16.1	9.3-49	3.6-13.2	26.5-119.9	3.1-22
LuPMEI59	13.7	12.7	15	1.2		9.6-19.6	5.4-29.8	6.9-32.3	0.3-3.9	
LuPMEI60										
LuPMEI65										
LuPMEI66	3.5	36.9	23.9	4.3	2.1	2.4-5.3	21.3-63.8	19.3-29.4	2.5-7.4	1.5-3
LuPMEI67	14.3	12.9	9.2	17.2	5.1	9.4-21.6	9.8-17	6.8-12.5	16.6-17.8	4.3-6.2
LuPMEI73	3		2.4	1	2	2.2-4.1		1.8-3.1	0.7-1.5	1.2-3.3

The fold-enrichment between the equivalent stem positions in whole stem (WS) and stem peel (SP) samples was calculated for each gene. Fold enrichment is shown in a linear scale and is the mean of 3 measurements from 3 biologically independent samples. The values not shown are genes that were not detected in those tissues. The confidence interval was calculated by using one standard deviation of the difference of the dCT between the two tissues compared.

**Table 4 - 3 Tissue enrichment expression of selected flax PME and PMEIs in whole stem compared to stem peels. Continued**

Gene/Point	Fold Change of genes with higher expression in stem peel									
	Fold change					Confidence Interval				
	A	B	C	D	E	A	B	C	D	E
LuPME1		53.2	121.2	86.3	15.2		20.6-137.2	99.1-148.1	22.1-337.2	9.5-24.4
LuPME3										
LuPME5										
LuPME7				1.4	1.7				1.1-1.7	1.5-2
LuPME10										
LuPME11					1					0.6-1.7
LuPME28				1.8	1.5				1-3.4	0.5-4.8
LuPME30		2.8	3.4	1.2	4.9		1.6-4.7	2.4-4.7	0.4-3.6	3.7-6.6
LuPME31		1	1.4	1.4	1.4		0.4-2.5	1.0-2.0	0.6-3.4	1-1.9
LuPME45		40.5	6.7	9.6	4.7		26.3-62.3	4.9-9.1	5.9-15.8	3-7.5
LuPME46	2.9	1.5	2.5	6.5	3.1	2-4.3	1.2-1.9	1.7-3.4	3.2-13.3	1.9-5
LuPME61		4	7	9.8	18.2		2.9-5.5	4.4-10.9	4.7-20.5	12.2-27.1
LuPME67	1.9					1.3-2.9				
LuPME71	4.9	8.3	15.4	3.3	17.1	1.7-14.1	2-33.9	3.6-66.2	1-11.6	3.5-85
LuPME73	10.8		2.1		3.4	9.3-12.6		1.6-2.7		2.8-4.1
LuPME79	3.2					1.6-6.4				
LuPME85		12.5	12.8	34.9	3.1		2.9-53.6	6.8-24.2	15.5-78.5	1.3-7.2
LuPME92										
LuPME96										
LuPME102	2.4	2.5	2		6.9	0.9-6.7	0.9-7.2	0.9-4.5		4.4-10.8
LuPME105				2.7					1.3-5.7	
LuPME127										
LuPME144		1.3		1.4			0.8-2		0.9-2.2	
LuPME145										
LuPME159					1.7					1.4-2.1
LuPME160										
LuPME165	16.6	5.9	4.5	15.7	56.1	11.2-24.4	2.3-14.7	1.4-14.7	6.7-37.1	25.4-123.7
LuPME166										
LuPME167										
LuPME173		1.4					1.2-1.6			

The fold-enrichment between the equivalent stem positions in whole stem (WS) and stem peel (SP) samples was calculated for each gene. Fold enrichment is shown in a linear scale and is the mean of 3 measurements from 3 biologically independent samples. The values not shown are genes that were not detected in those tissues. The confidence interval was calculated by using one standard deviation of the difference of the dCT between the two tissues compared.

#### 4.3.3 Clustering of transcript expression data

To identify shared patterns of transcript expression among the genes surveyed, we clustered the qRT-PCR results using STEM (Short Term Expression Miner) software [124]. STEM was designed specifically for time-series expression data and is therefore well-suited to clustering the developmental series represented by the stem and peel segments we analyzed.

Three genes for the whole stem tissues were filtered out and not used for the clustering because the difference in expression between the lowest value and the highest value was less than four-fold. Using STEM we identified five broad patterns among the transcript expression data from segments of the whole stem (Figure 4-2). In Group 1, which contained nine LuPMEs and three LuPMEIs, expression was highest in positions undergoing intrusive growth (SA through A), and decreased as the fibers matured (positions B through E). In Group 2, which contained seven LuPMEs and one LuPMEI, we observed an expression peak just above the snap point, at point A. In Group 3, which contained two LuPMEs and three LuPMEIs, expression was highest in positions below the snap point (B through E) which represent secondary cell wall deposition. In Group 4, which includes only one LuPMEI, peak expression occurred at point B. Finally in Group 5, one LuPMEI showed its lowest transcript expression at point A.

We also applied the same clustering method to transcript expression data from the outer stem peels. 11 genes were eliminated from clustering as the difference between the minimum and maximum dCT value was less than 2. Four different patterns were established (Figure 4-3). In Group 1, which contained four LuPMEs and one LuPMEI, a peak in expression was observed at point A (representing intrusive growth). In Group 2,

which contained two LuPMEs and two LuPMEIs, transcript abundance generally increased as the fiber matured. In Group 3, which contained two LuPMEs, peak in expression occurred in position B, which is associated with the onset of secondary cell wall, and expression decreased rapidly below this point. Group 4 contained three LuPMEs and one PMEI and showed an expression minimum at point B, when secondary cell wall deposition started, and then the expression increased basally (points C, D, and E) as the fibers matured.

To assess the statistical significance of the differences between tissues in a given gene, we performed an ANOVA statistical analysis for the expression of the genes in the whole stem and the stem peel tissues, which is depicted in Tables 4-4 and 4-5, respectively.

**Group 1**

Negative of dCT

**Group 2**

Negative of dCT

**Figure 4 - 2 Clusters of transcript expression patterns in segments of whole stem tissues.**

Stems positions (SA,1-2, 2-3, 3-4, A through E) are as defined in Figure 2-1. Transcript expression (y-axis) is the normalized negative dCT with point SA transformed to 0. Clusters are as defined by STEM software, using genes a minimum fold change of 4 between any two tissues.

Group 3

Negative of dCT

Group 4

Negative of dCT

**Figure 4 - 2 Clustering of genes based on expression on whole stem tissues. Continued.**

Group 5

Negative of dCT

Group without pattern

Negative of dCT

**Figure 4 - 2 Clustering of genes based on expression on whole stem tissues. Continued**

## Group 1

Negative of dCT

### Figure 4 - 3 Clusters of transcript expression patterns in segments of stem peels.

Stems positions (A through E) are as defined in Figure 2-1. Transcript expression (y-axis) is the normalized negative dCT with point SA transformed to 0. Clusters are as defined by STEM software, using genes a minimum fold change of 4 between any two tissues.

**Group 2**

**Negative of dCT**

**Figure 4 – 3 Clustering of genes based on expression on stem peel tissues. Continued**

Group 3

Negative of dCT

**Figure 4 – 3 Clustering of genes based on expression on stem peel tissues. Continued**

**Group 4**

**Negative of dCT**

**Figure 4 – 3 Clustering of genes based on expression on stem peel tissues. Continued**

**Group without pattern**

**Negative of dCT**

**Figure 4 - 3 Clustering of genes based on expression on stem peel tissues. Continued**

**Table 4 - 4 Tukey's multiple comparisons test of dCT expression values between whole stem tissues.**

Group	Gene	SA_1-2	SA_2-3	SA_3-4	SA_A	SA_B	SA_C	SA_D	SA_E
1	LuPME7	ns	ns	ns	*	**	**	**	***
1	LuPME10	ns	****	****	***	**	***	****	****
1	LuPME28	ns	ns	ns	ns	**	**	***	*
1	LuPME67	ns	ns	**	ns	ns	ns	ns	ns
1	LuPME71	ns	ns	ns	ns	ns	ns	ns	**
1	LuPME73	ns	ns	ns	ns	ns	ns	ns	*
1	LuPME79	ns	ns	ns	ns	****	****	****	****
1	LuPME102	ns	ns	ns	ns	*	*	ns	****
1	LuPME105	ns	****	****	****	****	***	****	****
1	LuPMEI45	ns	****	***	*	**	****	ns	****
1	LuPMEI59	ns	ns	ns	ns	ns	ns	*	*
1	LuPMEI65	ns	ns	ns	**	****	****	****	****
2	LuPME1	ns	ns	ns	****	**	ns	ns	****
2	LuPME30	ns	ns	***	****	ns	ns	ns	ns
2	LuPME45	ns	*	****	****	***	**	ns	ns
2	LuPME46	ns	ns	ns	ns	ns	ns	ns	ns
2	LuPME61	*	ns	****	****	****	****	****	****
2	LuPME85	ns	**	****	****	****	**	ns	ns
2	LuPME96	***	ns	ns	**	ns	ns	ns	ns
2	LuPMEI73	*	***	****	****	****	****	***	****
3	LuPME3	*	*	*	***	*	ns	ns	***
3	LuPME31	***	ns	ns	****	**	**	****	****
3	LuPMEI27	ns	ns	ns	*	*	ns	*	ns
3	LuPMEI44	ns	ns	ns	***	ns	*	ns	ns
3	LuPMEI67	ns	ns	ns	ns	ns	ns	ns	ns
4	LuPMEI60	ns	ns	ns	ns	ns	ns	ns	ns
5	LuPMEI66	***	****	****	****	***	*	ns	ns
np	LUPME5	ns	ns	ns	ns	ns	*	ns	ns
np	LUPME11	ns	ns	ns	ns	ns	ns	ns	ns
np	LUPME92	ns	ns	ns	ns	ns	ns	ns	*

An ANOVA test was followed by a Tukey's multiple comparisons test using GraphPad Prism version 6.00 for Windows. The asterisks denote the p-value as follows. \* 0.01-0.05; \*\*0.001-0.01, \*\*\*0.0001-0.001; \*\*\*\*<0.0001. . ns: non-significant difference (p>0.05)

**Table 4 - 4 Tukey's multiple comparisons test of dCT expression values between whole stem tissues.  
Continued**

Group	Gene	1-2_2-3	1-2_3-4	1-2_A	1-2_B	1-2_C	1-2_D	1-2_E	2-3_3-4
1	LuPME7	ns	ns	ns	ns	ns	ns	ns	ns
1	LuPME10	**	ns	ns	ns	ns	*	***	ns
1	LuPME28	ns	ns	ns	**	**	***	*	ns
1	LuPME67	ns	ns	ns	**	***	*	****	ns
1	LuPME71	ns	ns	ns	ns	ns	ns	**	ns
1	LuPME73	ns	ns	ns	ns	ns	ns	ns	ns
1	LuPME79	ns	ns	ns	****	****	****	****	ns
1	LuPME102	ns	ns	ns	ns	ns	ns	ns	ns
1	LuPME105	****	****	****	****	****	****	****	ns
1	LuPMEI45	**	ns	ns	ns	****	ns	***	ns
1	LuPMEI59	ns	ns	ns	ns	ns	*	*	ns
1	LuPMEI65	ns	ns	****	****	****	****	****	ns
2	LuPME1	ns	ns	****	*	ns	ns	****	***
2	LuPME30	ns	*	****	ns	ns	ns	ns	ns
2	LuPME45	ns	****	****	ns	ns	ns	ns	***
2	LuPME46	ns	ns	ns	ns	ns	ns	ns	ns
2	LuPME61	ns	****	****	****	****	****	**	****
2	LuPME85	ns	****	****	****	ns	ns	ns	****
2	LuPME96	****	ns	ns	***	***	ns	*	*
2	LuPMEI73	ns	*	****	ns	**	ns	**	ns
3	LuPME3	ns	ns	ns	ns	ns	ns	ns	ns
3	LuPME31	*	ns	ns	ns	ns	ns	ns	ns
3	LuPMEI27	ns	ns	ns	ns	ns	ns	ns	ns
3	LuPMEI44	ns	ns	***	ns	ns	ns	ns	ns
3	LuPMEI67	ns	ns	ns	ns	ns	ns	ns	ns
4	LuPMEI60	ns	ns	ns	ns	ns	ns	ns	ns
5	LuPMEI66	ns	ns	**	ns	ns	***	**	ns
np	LUPME5	**	ns	ns	**	***	**	**	ns
np	LUPME11	ns	ns	ns	ns	ns	ns	**	ns
np	LUPME92	**	ns	ns	ns	ns	ns	***	ns

**Table 4 - 4 Tukey's multiple comparisons test of dCT expression values between whole stem tissues.  
Continued**

Group	Gene	2-3_A	2-3_B	2-3_C	2-3_D	2-3_E	3-4_A	3-4_B	3-4_C	3-4_D
1	LuPME7	ns								
1	LuPME10	ns								
1	LuPME28	ns								
1	LuPME67	ns	***	***	**	****	ns	****	****	****
1	LuPME71	ns								
1	LuPME73	ns								
1	LuPME79	ns	****	****	****	****	ns	****	****	****
1	LuPME102	ns								
1	LuPME105	ns								
1	LuPMEI45	ns								
1	LuPMEI59	*	ns	**	ns	ns	ns	ns	ns	ns
1	LuPMEI65	ns	****	****	****	****	ns	****	****	****
2	LuPME1	****	****	**	**	****	****	ns	ns	ns
2	LuPME30	***	ns	ns	ns	*	ns	ns	***	****
2	LuPME45	****	ns	ns	ns	ns	*	*	*	****
2	LuPME46	ns								
2	LuPME61	****	****	****	****	****	**	*	ns	ns
2	LuPME85	****	*	ns	*	ns	ns	**	****	****
2	LuPME96	****	ns	ns	**	ns	ns	ns	ns	ns
2	LuPMEI73	****	ns							
3	LuPME3	ns								
3	LuPME31	****	ns	ns	***	***	***	ns	ns	*
3	LuPMEI27	*	*	*	*	ns	ns	ns	ns	ns
3	LuPMEI44	***	ns							
3	LuPMEI67	ns								
4	LuPMEI60	ns								
5	LuPMEI66	ns	ns	**	****	****	ns	ns	***	****
np	LUPME5	ns								
np	LUPME11	ns	ns	ns	ns	*	ns	ns	ns	ns
np	LUPME92	ns	**	*	**	ns	ns	ns	ns	ns

**Table 4 - 4 Tukey's multiple comparisons test of dCT expression values between whole stem tissues.  
Continued**

Group	Gene	3-4_E	A_B	A_C	A_D	A_E	B_C	B_D	B_E	C_D	C_E	D_E
1	LuPME7	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
1	LuPME10	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
1	LuPME28	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
1	LuPME67	****	*	*	ns	****	ns	ns	ns	ns	ns	ns
1	LuPME71	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	**
1	LuPME73	ns	ns	ns	ns	ns	***	ns	***	ns	ns	ns
1	LuPME79	****	****	****	****	****	ns	ns	ns	ns	ns	ns
1	LuPME102	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
1	LuPME105	ns	ns	ns	ns	ns	ns	ns	ns	**	ns	ns
1	LuPMEI45	ns	ns	*	ns	ns	ns	ns	ns	**	ns	*
1	LuPMEI59	ns	ns	ns	**	**	ns	ns	ns	***	**	ns
1	LuPMEI65	****	***	ns	***	****	ns	ns	ns	ns	ns	ns
2	LuPME1	*	****	****	****	*	ns	ns	ns	ns	**	**
2	LuPME30	****	**	****	****	****	ns	ns	**	ns	ns	ns
2	LuPME45	****	****	****	****	****	ns	ns	ns	ns	ns	ns
2	LuPME46	ns	ns	ns	*	ns	ns	ns	ns	ns	ns	ns
2	LuPME61	ns	ns	ns	**	****	ns	*	***	ns	ns	ns
2	LuPME85	****	****	****	****	****	*	****	***	*	ns	ns
2	LuPME96	ns	**	***	ns	*	ns	ns	ns	ns	ns	ns
2	LuPMEI73	ns	**	ns	***	ns	ns	ns	ns	ns	ns	ns
3	LuPME3	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3	LuPME31	*	ns	*	ns	ns	ns	ns	ns	ns	ns	ns
3	LuPMEI27	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3	LuPMEI44	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
3	LuPMEI67	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
4	LuPMEI60	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
5	LuPMEI66	****	**	****	****	****	ns	**	**	ns	ns	ns
np	LUPME5	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
np	LUPME11	*	ns	ns	ns	*	ns	ns	**	ns	ns	ns
np	LUPME92	ns	ns	ns	ns	ns	ns	ns	***	ns	**	***

**Table 4 - 5. Tukey's multiple comparisons test of dCT expression values between stem peel tissues.**

Group	Gene	A_B	A_C	A_D	A_E	B_C	B_D	B_E	C_D	C_E	D_E
1	LuPME30	ns	ns	****	***	ns	****	***	***	ns	ns
1	LuPME67	****	****	****	****	**	**	**	ns	ns	ns
1	LuPME73	****	****	****	****	ns	ns	ns	ns	ns	ns
1	LuPME79	****	****	****	****	****	****	*	*	****	ns
1	LuPMEI45	**	**	****	**	ns	ns	ns	ns	ns	ns
1	LuPME1	****	****	****	****	ns	ns	ns	ns	ns	ns
1	LuPME61	****	****	****	****	ns	ns	ns	ns	ns	ns
1	LuPME159	ns	ns	ns	***	ns	*	****	ns	**	ns
1	LuPMEI66	***	ns	***	****	ns	****	****	****	****	ns
3	LuPME45	**	ns	ns	*	**	****	****	ns	ns	ns
3	LuPME85	**	ns	ns	**	*	**	****	ns	**	*
4	LuPME5	**	ns	*	ns	ns	ns	ns	ns	ns	ns
4	LuPME31	**	**	ns	ns	ns	*	*	*	ns	ns
4	LuPME46	**	ns	ns							
4	LuPMEI65	****	****	***	*	ns	ns	***	ns	**	ns
np	LuPME7	ns	ns								
np	LuPME10	ns	ns	ns	ns	ns	*	ns	ns	ns	ns
np	LuPME11	ns	ns	*	ns	ns	ns	ns	ns	ns	ns
np	LuPME28	ns	ns	ns	ns	ns	ns	*	ns	ns	ns
np	LuPME71	ns	**	ns	ns						
np	LuPME92	ns	ns	*	ns	ns	ns	ns	ns	ns	ns
np	LuPME102	ns	ns								
np	LuPME105	ns	ns	ns	ns	ns	***	*	***	*	ns
np	LuPMEI44	ns	ns								
np	LuPMEI67	ns	ns								
np	LuPMEI73	ns	ns	**	ns	ns	ns	ns	ns	ns	ns

An ANOVA test was followed by a Tukey's multiple comparisons test using GraphPad Prism version 6.00 for Windows. The asterisks denote the p-value as follows. \* 0.01-0.05; \*\*0.001-0.01, \*\*\*0.0001-0.001; \*\*\*\*<0.0001. . ns: non-significant difference (p>0.05).

#### 4.3.4 Heterologous expression

To test whether any of the predicted LuPMEs or LuPMEIs had genuine PME or PMEI activity, we selected two LuPMEs (*LuPME67*, *LuPME79*) and one LuPMEI (*LuPMEI45*) for heterologous expression in *E. coli*. These particular LuPMEs were chosen because their transcript abundance peaked in expression during intrusive growth and diminished towards the bottom of the stem. Furthermore, these genes showed the highest fold-enrichment in intrusively growing cells of any LuPME assayed, especially in the stem peel tissues (Table 4-2). Therefore, these LuPMEs were considered strong candidates for having important roles in fiber elongation. *LuPMEI45* was chosen because it was one of the LuPMEIs, together with *LuPMEI65*, that showed a significant enrichment in expression during intrusive growth in the stem peel tissues (Figure 4-3, Table 4-5).

Three different strains of *E. coli* were tested for their suitability in heterologous expression of the selected genes: BL21(DE3), Rosetta(DE3)pLysS, and Rosetta-Gami B(DE3)pLysS (Novagen, Madison, WI, USA). We also evaluated different IPTG inducer concentrations (0.5 and 1 mM), induction times (2 hours, 4 hours, 18 hours), induction temperatures (20°C, 30°C, and 37°C), and purification methods. We found that 1 mM IPTG, and 18 hours of induction at 20°C were the best parameters for induction in all strains (data not shown). *LuPME67* and *LuPME79* proteins were not detected in the supernatant after expression with Rosetta(DE3)pLysS, and BL21(DE3), but they were successfully detected in supernatant from Rosetta-Gami B(DE3)pLysS cells, and their identity was confirmed by in-gel tryptic digestion and identification by LC MS/MS analysis (Figure 4-4, Table 4-6). However neither of these LuPMEs were functional in radial assays (data not shown).

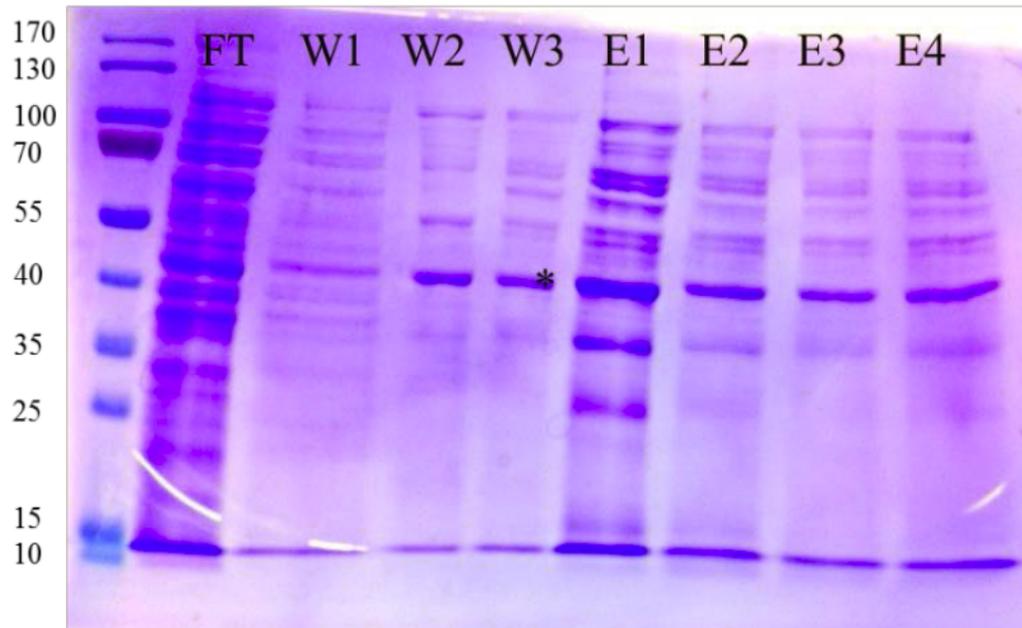
*LuPMEI45* expression was successfully detected in all of the strains, but the concentration was highest in Rosetta-Gami B(DE3)pLysS, so this strain was used in further experiments. The His-tagged heterologous LuPMEI45 protein was partially purified, and its identity was confirmed by LC MS/MS (Table 4-6) analysis and assayed in a radial diffusion assay. The recombinant LuPMEI45 protein was not purified to homogeneity and therefore the extract still contained some residual *E. coli* protein (Figure 4-5). Therefore an empty pET22b(+) vector expressed under the same conditions in Rosetta-Gami B(DE3)pLysS was used as a negative control in subsequent functional assays.

We found that recombinant LuPMEI45 successfully inhibited native PME activity of flax stem protein extracts (Figure 4-6). The purified protein at a concentration of 7310  $\mu\text{g/mL}$  was diluted at 1:12.5, 1:25, 1:50, 1:75, and 1:100. We tested volumes of 10  $\mu\text{L}$  of the different dilutions against proteins extracted from the top of the stem (first 5 cm), middle (11 to 16 cm from apex), and bottom (40 to 45 cm from apex), all at a concentration of 396  $\mu\text{g/mL}$  (10  $\mu\text{L}$  added). We determined that at both pH 6.0 and pH 7.0, a 1:50 dilution, 146  $\mu\text{g/mL}$  of LuPMEI45, was sufficient to reduce native LuPMEs activity by approximately 50%, while a 1:12.5 dilution, 585  $\mu\text{g/mL}$ , was sufficient to achieve a 100% inhibition in all the tissues (Figure 4-7).

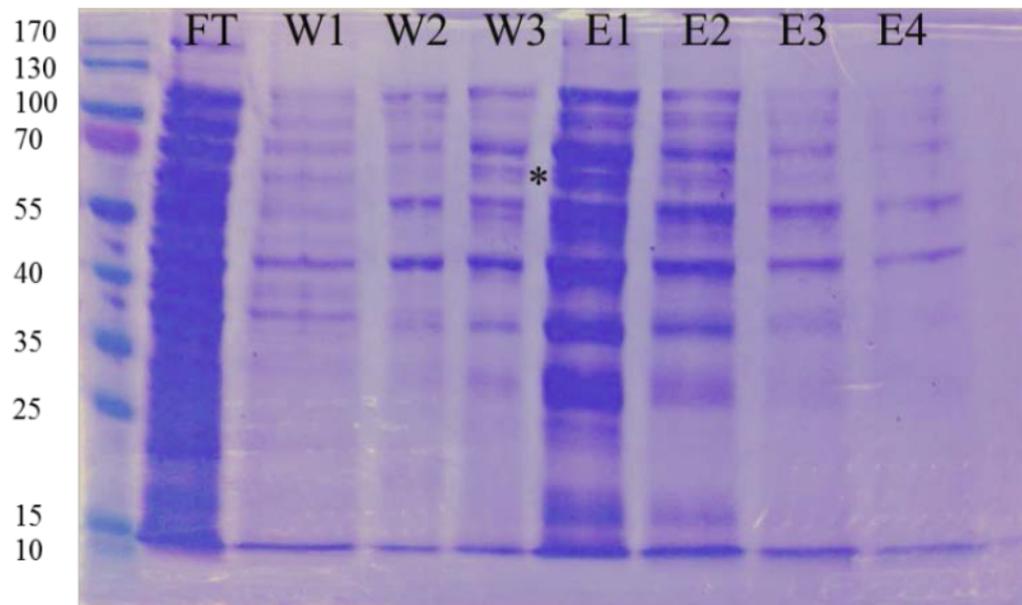
Once we knew the necessary concentration of heterologous LuPMEI45 to inhibit ~50% of the PME activity, we expanded the assessment of the inhibition capacity of LuPMEI45 to cell wall proteins extracted from the nine different points in the whole stem and five different points in the stem peel used in this study (Figure 2-2). It showed significant inhibition ( $p < 0.05$ ) at pH 6.0 at all the points in the whole stem, and all, except

point E, in the stem peel, and at pH 7.0 it inhibited at points SA, 1-2, 2-3, A, B, and E from the whole stem, and at points C, D and E from the stem peel, the activity of the PMEI on the whole stem SA in the stem peel tissues is shown as a reference (Figure 4-8).

#### LuPME67

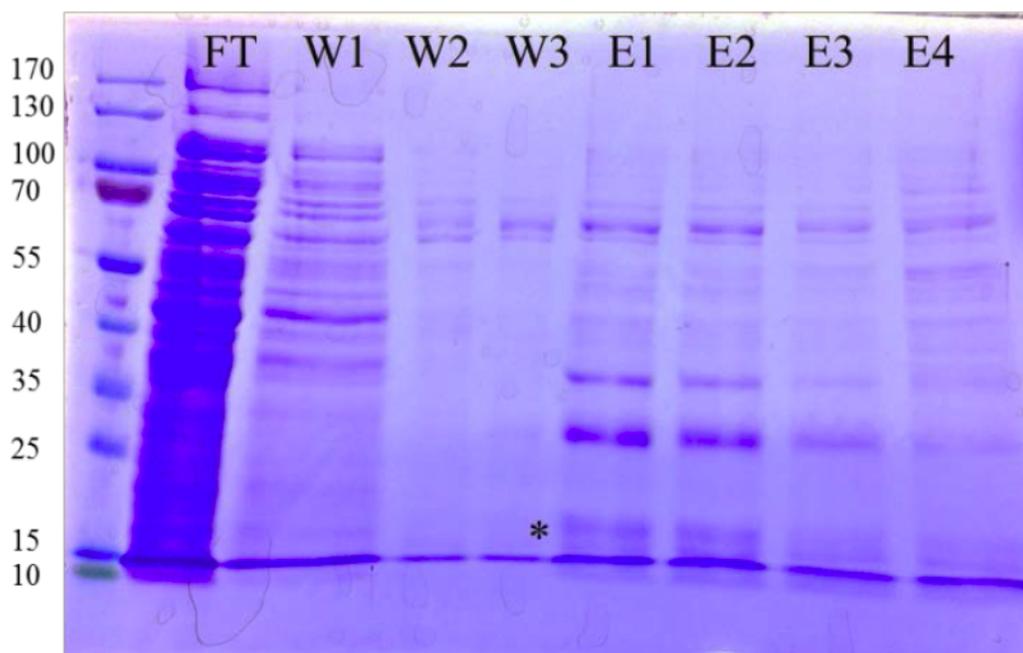


#### LuPME79



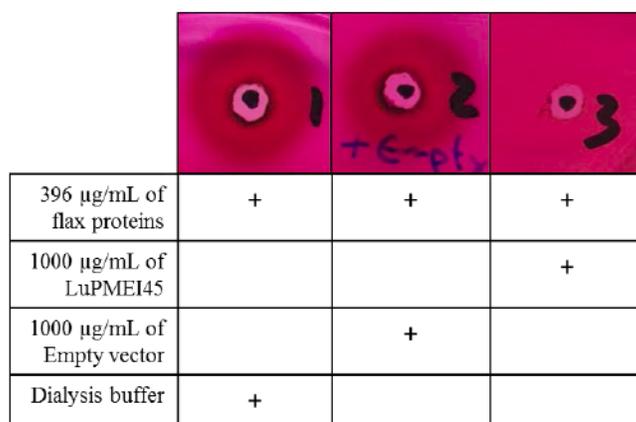
**Figure 4 - 4 Purification of LuPME67 and LuPME79 expressed in *E. coli*.**

\* Proteins successfully identified by LC MS/MS analysis. LuPME67 (~ 49.2 KDa), LuPME79 (~ 62.5 KDa). Left: Protein ladder. FT: Flow through; W: Wash; E: Elution. W1: 50 mM Tris-HCl 1.5 M NaCl. W2: 50 mM Tris-HCl, 300 mM NaCl, 20mM Imidazole. W3: 50 mM Tris-HCl, 300 mM NaCl, 40mM Imidazole. E: 50mM Tris HCl, 1 M NaCl and 250mM Imidazole



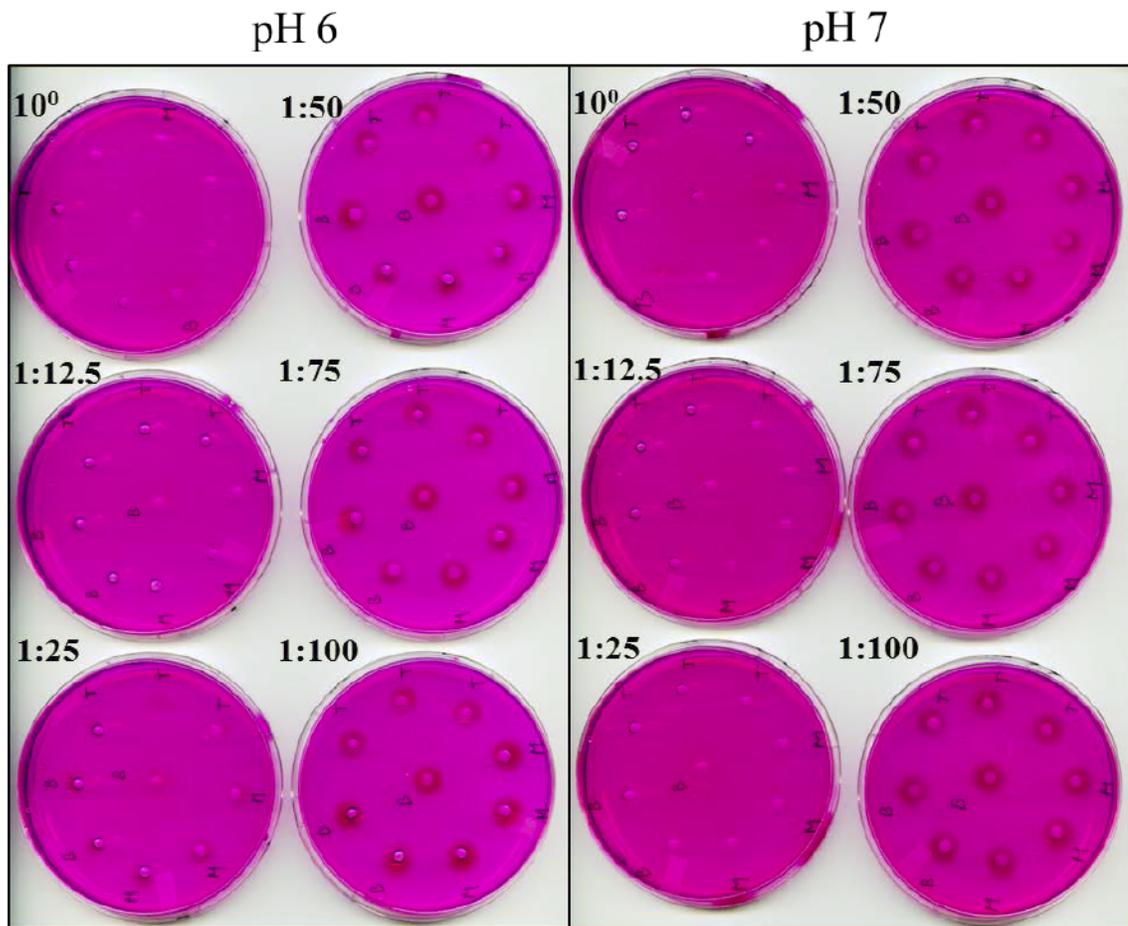
**Figure 4 - 5 Purification of LuPMEI45 expressed in *E. coli*.**

\* LuPMEI45 (~ 20.7 KDa) successfully identified by LC MS/MS analysis. Left: Protein ladder. FT: Flow through; W: Wash; E: Elution. W1: 50 mM Tris-HCl 1.5 M NaCl. W2: 50 mM Tris-HCl, 300 mM NaCl, 20mM Imidazole. W3: 50 mM Tris-HCl, 300 mM NaCl, 40mM Imidazole. E: 50mM Tris HCl, 1 M NaCl and 250mM Imidazole



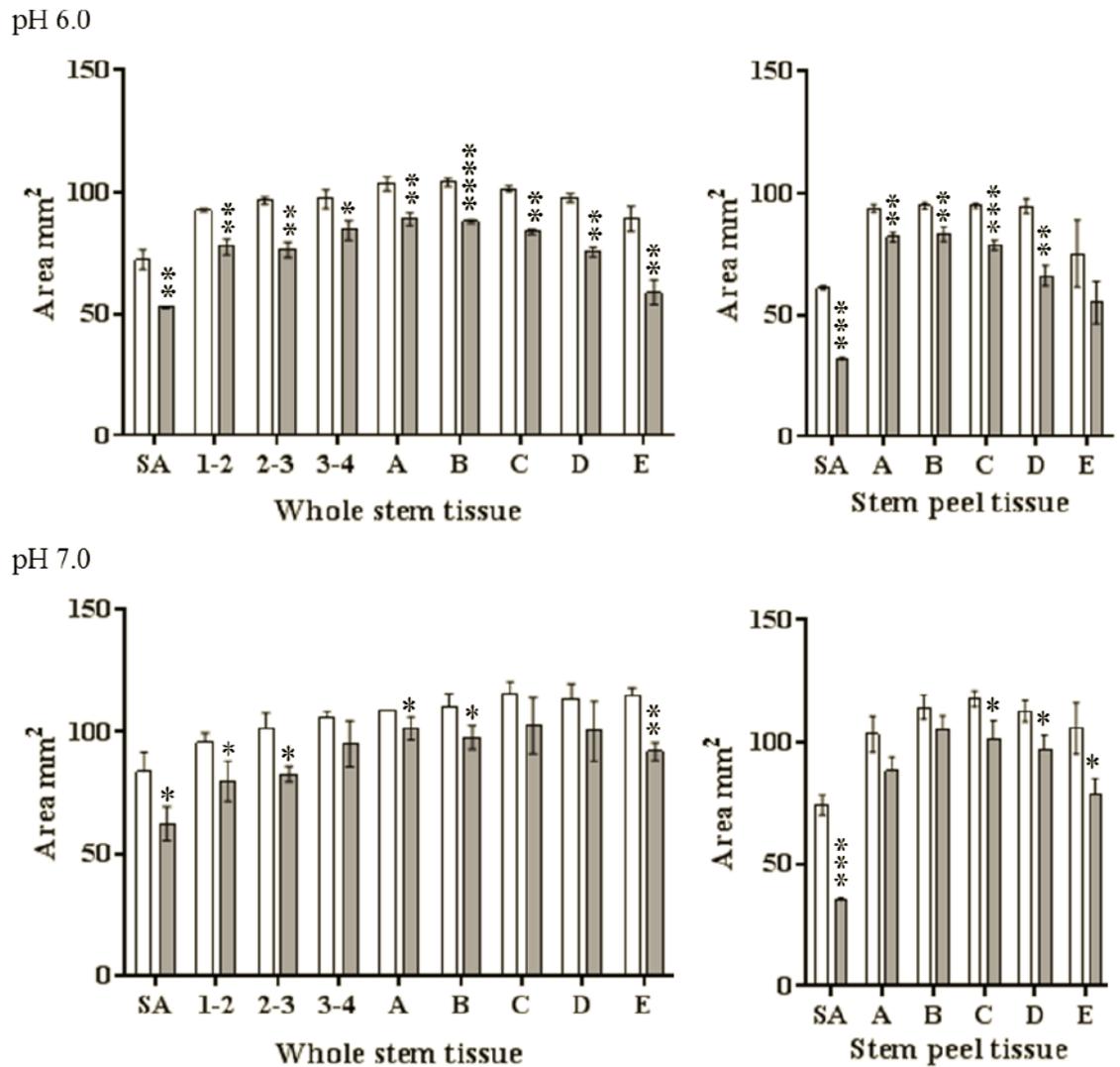
**Figure 4 - 6 Inhibitory capacity of LuPMEI45 expressed in *E. coli*.**

The activity of LuPMEI45 was assessed in a radial assay measured as its capacity to block the activity of flax cell wall proteins extracted from the top 5 cm of a ~5 weeks old plant. Two different controls were used: The buffer used for the dialysis of the protein after purification, and the purified proteins from the empty vector, pET22b(+), expressed in the same system under the same conditions.



**Figure 4 - 7 Radial assay of inhibitory capacity of LuPMEI45.**

Different dilutions of the purified proteins (7310  $\mu\text{g/mL}$ ) were assessed to establish the concentration at which  $\sim 50\%$  of the PME activity (396  $\mu\text{g/mL}$  of flax proteins) was inhibited. The letters in the plates denote the position of the stem where the proteins were extracted: Bottom (B), Medium (M), and Top (T).



**Figure 4 - 8. LuPMEI45 inhibitory activity on flax proteins extracted from whole stem and stem peel tissues at different stages of development of the fiber.**

10  $\mu$ L of proteins extracted from the different tissues (396  $\mu$ g/mL) were mixed with 10  $\mu$ L of LuPMEI45 (146  $\mu$ g/mL) or with 10  $\mu$ L of the buffer in which LuPMEI45 was dialyzed. A t-test was done to determine if the activity of LuPMEI45 significantly reduce the PME activity at the different tissues. The asterisk denotes the p-value as follows. \* 0.01-0.05; \*\*0.001-0.01, \*\*\*0.0001-0.001; \*\*\*\*<0.0001.

**Table 4 - 6 LC MS/MS analysis results of bands marked with an asterisk in figures 4-4 and 4-5.**

**LuPME67**

Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs
Lus10027737	Protein	57.11269951	26.62	1	7	7	18

Sequence	# PSMs	Modifications
LTQEEAASFMDISYINGEQWLQDGR	1	
RLTQEEAASFmDISYINGEQWLQDGR	1	M11(Oxidation)
GcFIQGSIDFIFGNAK	1	C2(Carbamidomethyl)
RLTQEEAASFMDISYINGEQWLQDGR	1	
AVDASPDFGSNTTLILIDSGTYR	2	
RLTQEEAASFMDISYInGEQWLQDGR	8	N17(Deamidated)
QSSGEDTGFSFVNSK	1	
AWGTYSTVVFIR	1	
NSAPAPAPGEVGAQAVAIR	2	

**LuPME79**

Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSM
Lus10031470	Protein	4.23231315	2.86	1	1	1	1

Sequence	# PSMs
SADSDVINSLVAPLLK	1

**LuPMEI45**

Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs
Lus10024596	Protein	1177.93832	57.06	1	9	9	305

Sequence	# PSMs	Modifications
FQVSNVQTWASAAMTcmDTcTDGLVEGEVK	6	C16, C20 (Carbamidomethyl); M17(Oxidation)
FQVSNVQTWASAAMTcmDTcTDGLVEGEVKR	38	M14, M17 (Oxidation); C16, C20 (Carbamidomethyl)
LLASAALYAALAAAK	69	
YPDLCISTLSPQASNITTPK	60	C5(Carbamidomethyl)
FQVSNVQTWASAAMTcmDTcTDGLVEGEVK	4	M14, M17 (Oxidation); C16, C20 (Carbamidomethyl)
FQVSNVQTWASAAMTcMDTcTDGLVEGEVKR	3	C16(Carbamidomethyl); C20(Carbamidomethyl)
AGFLISnALAFVNK	4	N7(Deamidated)
AGFLISNALAFVNK	47	
FQVSNVQTWASAAMTcMDTcTDGLVEGEVKR	58	M14(Oxidation); C16, C20 (Carbamidomethyl)
LLASAALYAALAAAKSTSK	1	
YPDLCISTLSPQASNITTPKLLASAALYAALAAAKSTSK	1	C5(Carbamidomethyl)
SGIVKAGFLISNALAFVNK	1	
FQVSNVqTWASAAMTcmDTcTDGLVEGEVKR	6	Q7(Deamidated); C16, C20 (Carbamidomethyl); M17(Oxidation)
LLASAALYAALAAAKSTSKSIETRPSSWSSR	1	
FQVSNVqTWASAAMTcmDTcTDGLVEGEVKR	6	Q7(Deamidated), M14, M17 (Oxidation); C16, C20 (Carbamidomethyl)

## 4.4 Discussion

To identify PME and PMEIs that were expressed dynamically during fiber development, we calculated the maximum fold difference in transcript abundance for any set of two tissues, for each gene assayed (Table 4-2). This was done separately for both whole stems and stem peels. We also calculated the fold difference in transcript expression at equivalent positions in whole stems and stem peels (Table 4-3). If a gene is important for fiber development, the magnitude of enrichment should be at least as high in stem peels as it is in whole stems. On the other hand, if a pattern was only observed in the whole stem, or if the magnitude of the change was significantly higher in the whole stem than in the stem peel, then that gene could be implicated in xylem development.

### 4.4.1 Genes enriched in fiber containing tissues

#### 4.4.1.1 Genes enriched during fiber elongation

All the genes that had a similar pattern of expression in the whole stem and the stem peel showed enrichment during intrusive growth. As the fold change was of at least the same magnitude in the stem peel compared to the whole stem, this meant that the expression of these genes may be specific to fibers and surrounding tissues. Those genes were LuPME46, LuPME67, LuPME73, LuPME79, LuPMEI45, and LuPMEI65 (Figures 4-2 and 4-3). Here we will discuss the genes that are most likely to have important roles in fiber development, based on the magnitude of their transcript enrichment.

*LuPME67* and *LuPME79* showed the highest change between the lowest and highest dCT values in the stem peel (152 and 1082-fold respectively, Figure 4-2) and a comparatively low change in the whole stem (20 and 26-fold respectively) which was

evidence of fiber-specific enrichment of these genes. *LuPME67* expression drastically diminished ( $p < 0.05$ ) below the snap point, in relation to point A (Figure 4-3, Table 4-4). Based on the whole stem results, it can be concluded that the expression was constant above the snap point ( $p > 0.05$ ), only presenting a difference between points SA to 3-4, where 3-4 was significantly larger ( $p < 0.05$ ) (Table 4-4), which might indicate that as fibers increased in number in a given section [20], the gene expression also increased. *LuPME67* is a type 2 PME, and one of the few LuPMEs with a predicted acidic isoelectric point (pI: 5.63), which implies that its mode of action might be random, leading to cell wall loosening as the pectin becomes a substrate for polygalacturonases and pectate lyases [42, 43]. Consequently, this is a gene that can be implicated in the dissolution of the middle lamella between cells that the fibre is penetrating during intrusive growth.

*LuPME79* showed a drastic decrease in expression below the snap point, its expression was constant above the snap point, and there was not difference ( $p > 0.05$ ) among the stages of development SA to A in whole stem tissue (Table 4-4). *LuPME79* is a type 1 PME, which interestingly does not have a predicted cleavage for separation of the PMEI-like domain from the PME domain [117]. It has a basic isoelectric point (pI predicted 9.04), which indicates *LuPME79* may demethylesterify the homogalacturonan (HG) in a blockwise fashion [42], leading to calcium cross linking between HG domains [35], and ultimately to cell wall stiffening. Based on the results on Chapter 2 and our hypothesis, this LuPME would be involved in strengthening of the cell wall during its elongation, which appears to be regulated by side chains in the HG. It has been shown that *LuPME3* can be secreted to the cell wall without processing the pro-region [53], so *LuPME79* may

likewise be secreted without processing. The persistence of the pro-region (PMEI-like domain) may affect the PME activity, of this protein, so it will be informative to carry out heterologous expression of this gene in the future.

*LuPMEI45* and *LuPMEI65* were the LuPMEIs found to have similar patterns in both the stem peel and in whole stem tissues; they both had significantly higher expression in stem peel tissues ( $p < 0.05$ ) in point A as compared to the tissues below the snap point, meaning that they are genes involved in the regulation of LuPMEs expression in the stem peel above the snap point. *LuPMEI45* was chosen for heterologous expression.

The expression of *LuPME5*, a type-1 PME with a predicted pI of 9.53, did not display major changes among the tissues in the whole stem (less than four-fold change), however, in the stem peel we did observe a higher expression in A than in B, C, and D ( $p < 0.05$ ), although the largest fold change was only 5-fold. This is consistent with the observations of Al-Qsous and collaborators [52], who determined that its highest expression occurs in the elongating parts of the hypocotyl, the apex and the root tip.

#### 4.4.1.2 Genes enriched below the snap point

From the stem peel expression data, we identified two genes that showed increased expression below the snap point (points B to E) with respect to A ( $p < 0.05$ ) and their expression was stable between points B to E ( $p > 0.05$ ). *LuPME1* had a 20-fold change and *LuPME61* had a 24-fold change, between the minimum point (A) and the maximum point (B). The expression pattern of these genes in the stem peel was not similar to their expression in the whole stem, in which the expression, oppositely, diminished from A to B in *LuPME1*, and did not change from A to B in *LuPME61*. This means that the

expression observed in the stem peel for these genes is a specific to this tissue, and indeed expression of *LuPME1* was 53, 121, and 86 times higher in points B, C, and D, respectively, in stem peels as compared to whole stem tissues (Table 4-2). As described above, the mode of action of these genes might be blockwise demethylesterification, so they would aid in the strengthening of the cell wall once the cells stop elongating (below the snap point) [20]. The role suggested for these genes is based on analysis of an orthologous gene from Arabidopsis, AtPME35 [117], which was found to strengthen the inflorescence stem by a blockwise demethylesterification action [77].

#### 4.4.2 Genes enriched in the xylem

We found seven genes that showed a peak in expression in point A of the stem. These included the four genes that showed the highest fold change (between any two stem points) among any of the genes analyzed in the whole stem tissue: *LuPME85* (419-fold), *LuPME61* (306-fold), *LuPME1* (191-fold), and *LuPME45* (186-fold). The other three genes were *LuPME30* (45-fold), *LuPME31* (40-fold), and *LuPME96* (16-fold) (Table 4-2). As this expression pattern was unique for the whole stem, and not observed in the stem peel tissues, we concluded that these genes may play a role in xylem or pith development. The predicted isoelectric point of all of these proteins is basic, so blockwise demethylesterification is expected to occur [42] leading to cell wall rigidification. Furthermore, one PME1, *LuPMEI73*, was observed with this pattern in the whole stem but not in the stem peel; its high expression in point A, respect to the SA (74-fold), leads us to believe that it has an important role in regulating PME activities at this point in the inner tissues, presumably within the xylem.

*LuPME3* (type-1 PME with a predicted pI of 9.8) expression was not detected in the stem peel, while it was detected in low amounts in the whole stem tissues where its expression was significantly lower in SA ( $p < 0.05$ ) with respect to the rest of the tissues (1-2 to E), where the expression was not significantly different ( $p > 0.05$ ). The xylem undergoes differentiation, expansion, and maturation. In the vicinity of the shoot apex, very little vascular tissue maturation is expected to occur and it is only at node 3-5 that thickening starts [127], so if *LuPME3* is involved in the cell wall stiffening of the xylem, it is then expected that its expression is lower in point SA, which we found, and then as more xylem is produced along the stem the maturation of the xylem is a constant process which is observed in the expression of this gene. *LuPME3* was previously found to have detectable expression in the vascular tissue of stems and leaves, and in the root meristem [50], and was found to have similar expression in the whole extension of the hypocotyl and the root in a 10 days old seedling [52]; they did not find lower expression at the top of the seedling, although their detection method (RT-PCR Southern blot) is not as sensitive as qRT-PCR. Based on the phylogenetic analysis (Figure 3-9), it was established that *LuPME3* is one of the most similar genes to *PttPME1* in hybrid aspen [117], a PME that leads to an increase in xylem fiber elongation when it is downregulated, suggesting that *PttPME1* strengthens cellular adhesion, hindering intrusive growth [57]. As the expression of *LuPME3* in flax occurs in the xylem, it is possible that the same situation is occurring in flax.

*LuPME7* and *LuPME92* are the other LuPMEs closely related to *PttPME1*. The expression of *LuPME7* was significantly higher ( $p < 0.05$ ) in the whole stem in point SA respect to A, B, C, D, and E (Table 4-4), while in the stem peel there was not significant

difference between the tissues (Table 4-5). *LuPME92* expression did not show a difference greater than 4-fold between any of the tissues in the whole stem and the stem peel, however the expression was higher in the whole stem tissues (A to E), respect to stem peel, point D was 26 times higher in the whole stem than in stem peel (Table 4-3), suggesting a role in xylem maturation, as the one observed for *PttPME1* [57].

#### 4.4.3 PMEI inhibitory activity

LuPMEI45 was found to effectively inhibit the action of flax LuPMEs along the stem (Figure 4-6). As the expression of *LuPMEI45* is higher during intrusive growth ( $p < 0.05$ ) (Tables 4-4 and 4-5), it could be expected that its inhibitory capacity is higher at the tissues undergoing intrusive growth, however the inhibitory capacity was not significantly different along the stem (data not shown). The preferred target(s) of LuPMEI45 will be important to determine, so its activity can be correlated with the mode of action of a PME.

## 4.5 Conclusion

We were able to characterize in detail the expression of selected LuPMEs and LuPMEIs along the stem, in relation to stages of development of flax fibers. Candidate genes with expression patterns implicating them in specific processes of phloem fibers and xylem development were presented, and a functional heterologous expression of one of them was achieved. The detailed study of these genes by the subcellular localization of the proteins, mutagenesis, silencing and/or overexpression techniques will allow the finding of genes relevant for the improvement of the crop, either by producing longer and

easier to extract fibers or by obtaining plants with shorter fibers avoiding the obstruction of the machinery.

## **5 Chapter 5: A pectate lyase enriched during flax intrusive growth development**

### **5.1 Introduction**

Pectate lyase like (PLL) genes are a family comprising 26 genes in Arabidopsis, 22 in poplar, 14 in rice [128], and 46 in *Brassicca rapa* [129]. They were first reported in plants by Wing and collaborators [130], who found two genes in the pollen of tomato plants with sequence similarity to pectate lyases from phytopathogenic bacteria, and which were hypothesized to aid in the growth of the pollen tube through the style. *GhPEL*, a PLL gene from cotton, was found to have its highest expression during the rapid elongation stage of cotton seed trichomes [62]. When *GhPEL* expression was suppressed, the trichomes were shorter. Because cotton seed trichomes do not penetrate other tissues during their development, it was proposed that *GhPEL* has a role in the remodeling of pectins in trichome walls to allow their rapid expansion. In a study done by Israelsson and collaborators [131] using transgenic aspen (*Populus tremula L. × P. tremuloides* Michx.) over-expressing GA 20-oxidase, which produces longer and more xylem fibers, compared to the wild type, a PLL was the most highly enriched transcript in transgenic plants as compared to the wild-type. PLL genes have also been implicated in other functions such as fruit ripening [132] and in establishing compatible interactions with pathogens, so their presence is detrimental for the plant during infection [133]. PL genes from bacteria and PLL from plants have been previously successfully expressed in *E. coli* [62, 134] and their activity has been determined by measuring the increase in absorbance at 232 nm when PolGalA is treated with the enzyme [135].

Flax phloem fibers elongate by intrusive-diffusive growth [18]. This is believed to require the action of polygalacturonases (PG, EC 3.2.1.15[40]) and pectate lyases (EC 4.2.2.2 [40]) [19]. During elongation, fibers penetrate the compound middle lamella (CML) of neighbouring cells and generate a new CML, in a process that is different from the formation of the CML during cytokinesis [22]. Here we describe a PLL gene from flax, *PLL4*, *Lus10006456*, which was found to have a significantly higher expression during intrusive growth of the fibers.

## **5.2 Materials and Methods**

### **5.2.1 Annotation of PLL in flax**

The pectate lyase domain (PFAM: PF00544) was searched in the predicted flax proteome in a keyword-based ontology search in the Phytozome v9.1 database. The complete protein sequences were used as queries to determine the presence of a signal peptide using SignalP 4.0 [97], transmembrane domains, predicted using TMHMM v.2.0 [98], and to establish the subcellular localization of the protein, using WoLF PSORT and Plant-mPLOC [99, 100].

### **5.2.2 Phylogenetic analysis**

The complete PLL predicted proteins were aligned using MUSCLE [102]. This alignment was used to determine the substitution model that best described the evolutionary process of the proteins. Using ProtTest [103] we found that WAG+I+G+F was the best model. Then the alignment and the substitution model were used to construct a maximum likelihood tree, using GARLI [104] under the CIPRES web interface [105], with 1000 bootstraps and 2 search-replicates.

### 5.2.3 Microarray analysis

The analysis of microarray was done as shown in appendix 7.1.

### 5.2.4 Plant growth and qRT-PCR

Tissues were extracted from stems of 5 week old plants as described in section 2.2.1, RNA was extracted and qRT-PCR was performed as described in sections 4.2.2 and 4.2.3.

The Universal ProbeLibrary Assay Design Center (Roche) was used to design specific primers for the PLL and the PG genes. We evaluated the specificity of primers by BLASTn alignment of the primers against the complete predicted transcriptome and the entire genome assembly. All primer pairs were designed so at least one primer of each pair had three or more mismatches to any off-target gene, near the 3' of the primer. The primers used were as follows: For *PLL4*, forward 5'-TCTTGTCGCCGATGAAGTC-3', reverse 5'-CCGTTTTGCAAGAGAGGAAT-3'; for *PLL19*, forward 5'-CGCCGAGAAGGAATCTCTAC-3', and reverse 5'-AACGCACCATCATTACCTT-3; and for *Lus10034881* 5'-GAGGCTGACATCGTGGAAG-3', reverse 5'-ACATATAGCCACGCCGACAT-3'

### 5.2.5 Heterologous expression

The complete CDS of PLL4 was synthesized and codon optimized (Bio Basic Inc.) for expression in *E. coli* using the methods described in section 4.2.5. (Appendix Additional file 5-1).

### 5.2.6 LC MS/MS

The proteins observed with the expected size in the Coomassie stained polyacrylamide gel were confirmed by in-gel tryptic digestion and identification by LC MS/MS analysis in the Institute for Biomolecular Design (University of Alberta).

### 5.2.7 Pectate lyase assay

The assay was modified from Collmer and collaborators [135] and Wang and collaborators [62], it consisted of 0.3% (w/v) polygalacturonic acid in 0.05M Tris HCl pH 8.4, 2 mM CaCl<sub>2</sub> and 30 µL of the partially purified enzyme or the empty vector (320 µg/ml) in a final volume of 300 µL. The enzyme reaction was incubated for 1 hour at 40°C and the difference in the absorbance at 232 nm, between this time and time 0 was established.

We also assessed the pectate lyase activity by using the antibody PAM1 [136], which recognizes at least 30 contiguous GalA units. 7 µL of a solution containing 0.4% (w/v) of polygalacturonic acid in 0.05M Tris HCl pH 8.4, 1 mM CaCl<sub>2</sub> was mixed with 3.5 µL of the partially purified enzyme or the empty vector (320 µg/ml), incubated for 15 min at 40°C, and 1 µL spots were dispensed in triplicate in a nitrocellulose membrane, and the protocol described in section 2.3.3.3 for LM5 was followed.

## **5.3 Results**

### 5.3.1 Pectate lyase-like proteins of flax

We searched the pectate lyase domain (PFAM: PF00544) in the predicted proteome of flax, using a keyword-based ontology search in the Phytozome v9.1 database. 40 PLL genes were found, with a mean size of predicted proteins of 398 amino acids (Figure 5-1).

The presence of a signal peptide, a transmembrane domain and their predicted subcellular localization were determined using WoLF PSORT and Plant-mPLoc [99, 100] (Table 5-1).

### 5.3.2 Phylogenetic relations of PLL proteins

A phylogenetic tree was constructed to classify the predicted flax PLLs on the basis of amino acid sequence similarity of the complete predicted protein sequences of the putative PLLs in the flax genome. Based on the bootstrap values, and tree topology, three major monophyletic groups of PLLs could be defined.

**Size of protein (AA)**

**Figure 5 - 1 Distribution of size of the predicted PLL proteins in flax.**

**Table 5 - 1. Main features of the predicted PLL proteins in flax.**

Flax genome v1.0 code	PLL	Plant-mPLoc	WoLF PSORT	SP	TM	Length of protein (AA)	Tested in Microarray
Lus10003307	PLL1	cw	cy, cy	-	-	330	NS
Lus10005254	PLL2	cw, ch	v, g, n, pl, ex	+	-	461	NS
Lus10006241	PLL3	cw	ch, ex, v	+	-	387	NS
Lus10006456	PLL4	n	ch, ex, n, cy, m, pl	+	-	407	S
Lus10011257	PLL5	pm, cw	ex, cy, ch, v, er	+	-	407	NA
Lus10011258	PLL6	cw	er, cy, cy, n, m	+	-	405	NS
Lus10011400	PLL7	cw, ch, n	ch, ex, n, cy, m, pl	+	-	406	S
Lus10011758	PLL8	pm, cw, ch, m, n	ch, n, m, pl, ex, v, cy	+	-	421	NS
Lus10011885	PLL9	pm, cw, ch, g, n	ex, v, g, cy	+	+	434	S
Lus10013267	PLL10	cw	cy, n, pe	-	-	231	NA
Lus10013667	PLL11	cw	cy, ex, ch, n	-	-	265	NA
Lus10013668	PLL12	pm, cw	ch, ex, n, cy	-	-	263	NA
Lus10014414	PLL13	cw	n, cy, v, er, cy	-	+	471	NA
Lus10014887	PLL14	pm, cw, ch, m, n	ex, ch, v, er, m	+	-	406	NA
Lus10015359	PLL15	pm, cw	cy, n, m	-	+	275	NA
Lus10018429	PLL16	pm, cw, ch	cy, cy_n, ch, pl, v, n, m	+	-	467	NA
Lus10018430	PLL17	cw	ch, ex, pe, n, cy	+	-	407	NS
Lus10022310	PLL18	pm, cw	ex, er, ch, pl, v	+	-	409	NS
Lus10022817	PLL19	cw, ch	ex, v, g, cy	+	+	435	S
Lus10023542	PLL20	cw	v, ch, ex, g, m	+	-	492	NS
Lus10023623	PLL21	cw, ch	ch, ex, n, er, cy, pl	+	+	439	NS
Lus10023679	PLL22	ch	ch, v, ex	+	+	457	NS
Lus10023917	PLL23	pm, cw	pl, g, v	+	-	487	NS
Lus10024439	PLL24	cw, ch, n	n, cy, m, pl, pe	-	-	382	NS
Lus10025290	PLL25	pm, cw, ch, cy	cy, n, ex, cy	-	-	346	NS
Lus10030313	PLL26	cw	cy, cy, n	-	-	327	NS
Lus10030670	PLL27	cw, ch	v, g, n, pl, ex	+	-	461	NS
Lus10030791	PLL28	cw, ch	cy, ch, m, pl, ex	+	+	437	NS
Lus10031296	PLL29	cw, cy, g	ch, cy	-	-	129	NA
Lus10031867	PLL30	pm, cw	ch, cy, pe, n	-	-	254	NA
Lus10031889	PLL31	ch, g	n, ch, cy, m	-	-	142	NA
Lus10033018	PLL32	cw, ch	pl, g, cy, v	+	-	500	NS
Lus10033037	PLL33	ch, n	ch, pe, ex, cy	+	-	423	NA
Lus10036721	PLL34	pm, cw, ch	ch, ex, cy, v, er	+	-	438	NS
Lus10036946	PLL35	pm, cw, ch, m	ch, ex, m, pl, n	+	-	406	NS
Lus10037207	PLL36	pm, cw, ch	ex, ch, v	+	-	438	NS
Lus10037945	PLL37	cw, ch	pl, v, g	+	-	785	NA
Lus10038157	PLL38	pm, cw, ch	ex, v, er, ch	+	+	447	NS
Lus10040426	PLL39	cw	pl, n, er	-	-	663	NS
Lus10042509	PLL40	pm, cw	cy, cy, n	-	-	205	NS

(SP): Presence of signal peptide; (TM): Presence of transmembrane domain. Subcellular localization: ch: chloroplast; cw: cell wall; cy: cytosol; er: endoplasmic reticulum; ex: extracellular/cell wall; g: Golgi apparatus; m: mitochondria; n: nuclear; pl: plasma membrane; v: vacuolar membrane; pe: peroxisome; ct: cytoskeleton. (NS): not differentially expressed. (S): differentially expressed in at least one tissue. NA not assessed in microarray.

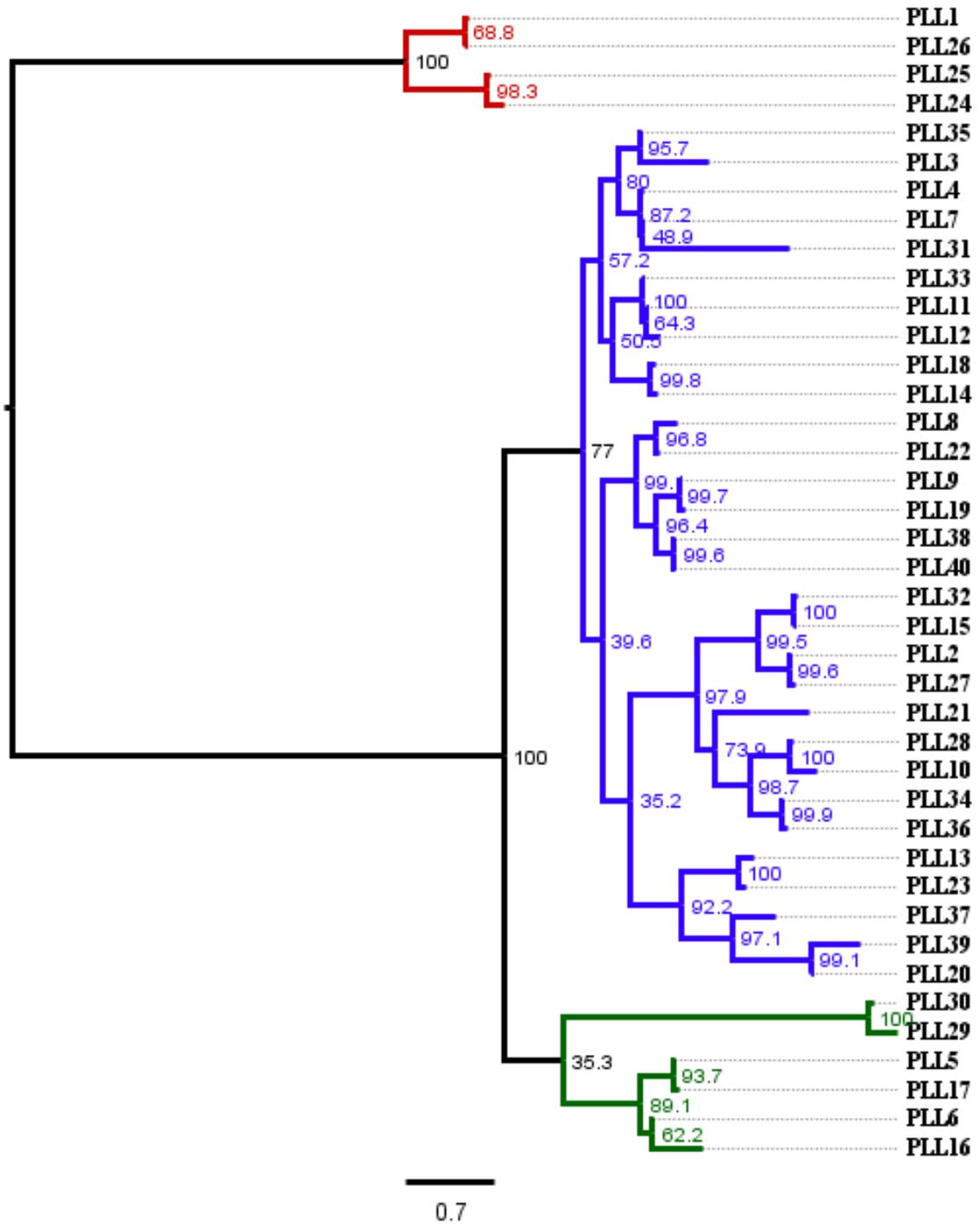


Figure 5 - 2. Maximum likelihood dendrogram of PLL in flax

The three monophyletic clades identified are labeled with three different colors. 1000 bootstraps and 2 search-replicates.

### 5.3.3 Pectin degrading enzymes from microarray data

We searched the results of a microarray analysis of developing flax stems [125] (Appendix 7.1) for PLLs or PG that were differentially expressed in at least one of five positions along the developing stem. Probes for 27 PLL genes were present on the microarray; we found four of these PLLs genes were differentially expressed: *PLL4* and *PLL7* (which were represented by the same probe), *PLL9*, and *PLL19*, all of which had peak expression in the segment 1-2 cm from the shoot apex. Microarray analysis also identified one PG, *Lus10034881*, out of 53 PGs presented in the microarray, which had peak expression in the segment 0-1 cm from the shoot apex. As the expression patterns of the four PLL genes found was very similar, we arbitrarily chose two of the PLL genes (*PLL4* and *PLL9*) and the PG for further analysis by qRT-PCR in the same tissues as were used in the microarray (Figure 5-3). A summary of the ANOVA analysis between the different points is shown in Table 5-2.

Negative of dCT

**Figure 5 - 3 qRT-PCR validation of genes of interest using same tissues as those used by To [125] in the microarray.**

dCT was obtained by subtracting the geometric mean of the three endogenous controls used to the Ct value of the genes studied for every biological replicate. Here we show the average of the three biological replicates. The negative of the dCT value is shown in the graph, so a higher value represents more transcript abundance.

**Table 5 - 2 Tukey's multiple comparisons test for the expression of different genes between the five different tissues used in flax stem microarray.**

	<b>Lus10034881</b>	<b>PLL19</b>	<b>PLL4</b>
1 vs. 2	ns	****	*
1 vs. 3	****	ns	ns
1 vs. 4	****	****	****
1 vs. 5	****	****	****
2 vs. 3	**	****	*
2 vs. 4	****	****	****
2 vs. 5	****	****	****
3 vs. 4	ns	****	****
3 vs. 5	ns	****	****
4 vs. 5	ns	****	****

An ANOVA test was followed by a Tukey's multiple comparisons test using GraphPad Prism version 6.00 for Windows. The asterisks denote the p-value as follows. \* 0.01-0.05; \*\*0.001-0.01, \*\*\*0.0001-0.001; \*\*\*\*<0.0001. ns: no significant difference (p>0.05)

#### 5.3.4 Transcription profiling of a PLL gene in whole stem and stem peel tissues

*PLL4* showed its highest fold change (4.5-fold) between point 3 (intrusive growth) and point 4 (secondary cell wall deposition), so we assessed the expression of this gene in nine different points of the stem (described in Chapter 2, Section 2.2.2). In the stem peel tissues, the expression of *PLL4* was significantly higher ( $p < 0.05$ ) during intrusive growth (point A), than during secondary cell wall deposition (points B through E), with a maximum fold change of 34 between point A and E (Figure 5-4). On the other hand in the whole stem tissues there was no significant difference between tissues undergoing intrusive growth (SA through A) with point B ( $p > 0.05$ ) (Table 5-3), which is in the cell wall thickening stage, but there was difference ( $p < 0.05$ ) with points C and E, finding a 57-fold difference between point B and C, and 30-fold between B and E, which are all points undergoing secondary cell wall deposition of the fibers (Figure 5-4).

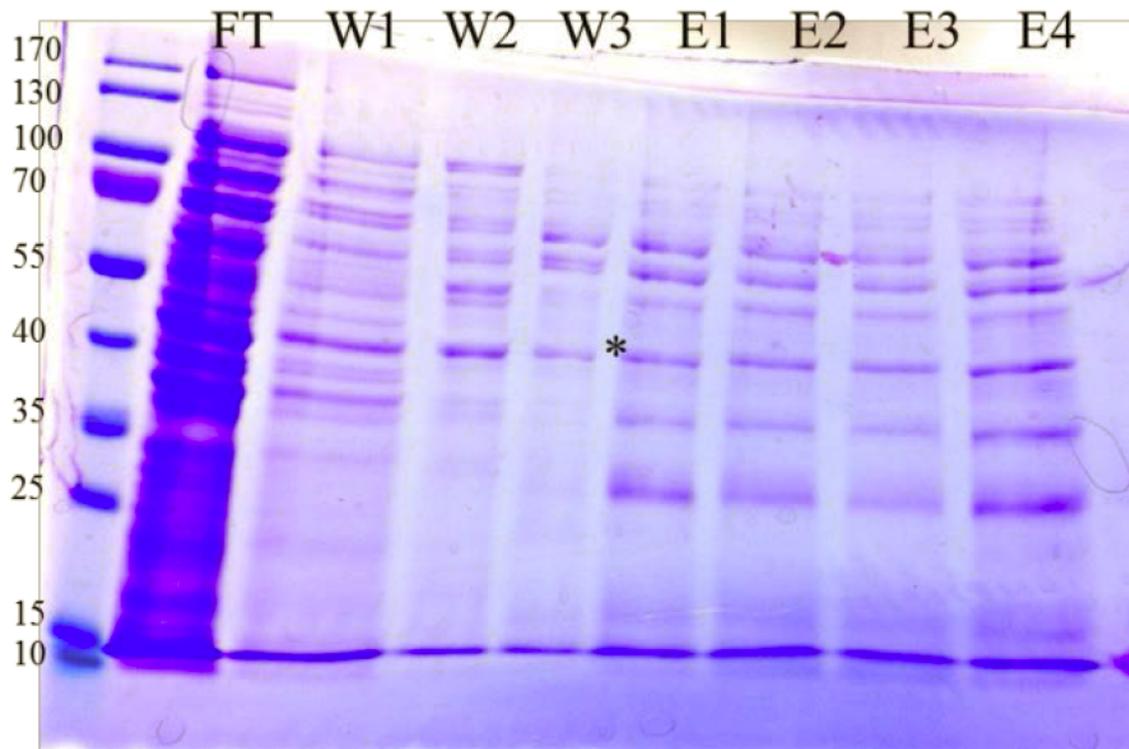
#### 5.3.5 Heterologous expression of PLL4

To establish whether the predicted *PLL4*, that exhibited the highest fold change between the elongation and thickening of the fibers, had authentic pectate lyase activity, we expressed the gene in *E. coli*. The complete CDS of *PLL4* was cloned in the plasmid pET22b(+) (Novagen, Madison, WI, USA) and expressed in *E. coli* strain Rosetta-Gami B(DE3)pLysS. The protein was successfully expressed and partially purified (Figure 5-5), and identified by peptide mass fingerprinting (Table 5-4).

Negative of dCT

**Figure 5 - 4. Expression of PLL4 in whole stem and stem peel tissues.**

The red line indicates expression in the whole stem tissues while the blue line indicates expression in the cortical peel tissues. dCT was obtained by subtracting the geometric mean of the three endogenous controls from the Ct value of the genes studied for every biological replicate. Here we show the average of the three biological replicates. The negative of the dCT value is shown in the graph, so a higher value represents more transcript abundance.



**Figure 5 - 5. Purification of pectate lyase PLL4 expressed in *E. coli*.**

\* PLL4 (~ 48.3 Kda) was successfully identified by LC MS/MS analysis. Left lane: Protein ladder in KDa. FT: Flow through; W: Wash; E: Elution. W1: 50 mM Tris-HCl 1.5 M NaCl. W2: 50 mM Tris-HCl, 300 mM NaCl, 20mM Imidazole. W3: 50 mM Tris-HCl, 300 mM NaCl, 40mM Imidazole. E: 50mM Tris HCl, 1 M NaCl and 250mM Imidazole

**Table 5 - 3 Tukey's multiple comparisons test for the expression of PLL4 in whole stem and stem peel tissues.**

Whole stem tissues	
SA vs. 1-2	ns
SA vs. 2-3	ns
SA vs. 3-4	ns
SA vs. A	ns
SA vs. B	ns
SA vs. C	****
SA vs. D	ns
SA vs. E	****
1-2 vs. 2-3	ns
1-2 vs. 3-4	ns
1-2 vs. A	ns
1-2 vs. B	ns
1-2 vs. C	****
1-2 vs. D	ns
1-2 vs. E	****
2-3 vs. 3-4	ns
2-3 vs. A	ns
2-3 vs. B	ns
2-3 vs. C	****
2-3 vs. D	ns
2-3 vs. E	****
3-4 vs. A	ns
3-4 vs. B	ns
3-4 vs. C	****
3-4 vs. D	ns
3-4 vs. E	****
A vs. B	ns
A vs. C	****
A vs. D	ns
A vs. E	***
B vs. C	****
B vs. D	ns
B vs. E	***
C vs. D	***
C vs. E	ns
D vs. E	**

Stem peel tissues	
A vs. B	****
A vs. C	****
A vs. D	****
A vs. E	****
B vs. C	ns
B vs. D	*
B vs. E	**
C vs. D	ns
C vs. E	ns
D vs. E	ns

An ANOVA test was followed by a Tukey's multiple comparisons test using GraphPad Prism version 6.00 for Windows. The asterisk denote the p-value as follows. \* 0.01-0.05; \*\*0.001-0.01, \*\*\*0.0001-0.001; \*\*\*\*<0.0001. ns: no significant difference (p>0.05)

**Table 5 - 4 LC MS/MS analysis results of band marked with an asterisk in Figure 5-3.**

<b>Accession</b>	<b>Description</b>	<b>Score</b>	<b>Coverage</b>	<b># Proteins</b>	<b># Unique Peptides</b>	<b># Peptides</b>	<b># PSMs</b>
Lus10006456	Protein	31.39829707	12.07	2	4	4	8

<b>Sequence</b>	<b># PSMs</b>	<b>Modifications</b>
SEGDLLNGAFFTR	1	
SEGDLLnGAFFTR	2	N8(Deamidated)
YGVIRDEPLWIVFAR	3	
DKImQVTIAFNHFGGLVQR	1	M4(Oxidation)
DEPLWIVFAR	1	

### 5.3.6 Pectate lyase activity assay

Lionetti and collaborators [80] correlated an increase of PG activity with a decrease in binding of PAM1 antibody (which binds to at least 30 contiguous galacturonic acid units [136]). Based on this, we attempted to use PAM1 to assess the activity of PLL4 on polygalacturonic acid (Sigma-Aldrich). The products of the reaction were assessed by an immunodot assay, by detecting the binding of the PAM1 antibody to PolGalA treated with the heterologous enzyme. However the intensity of the hybridization signal of the antibody was not consistent with the amount of PolGalA in a standard curve (data not shown).

We also assessed the pectate lyase activity by measuring the production of unsaturated products from the cleavage of PolGalA, measured by the change in absorbance at 232 nm. After one hour of incubation we established a pectate lyase activity in PLL4, which was significantly higher ( $p < 0.05$ ) than the basal activity observed in the proteins purified from the expression of the pET22b(+) empty vector growth in the same strain under the same conditions (Figure 5-6).

**Difference in absorbance (232 nm) after 1 hr**

**Figure 5 - 6 Activity of PLL4 measured as the production of unsaturated products from the hydrolysis of PolGalA.**

The assay was repeated three times.

## 5.4 Discussion

As expected based on what has been found in other plants, flax contained a large family of 40 PLL genes. Jiang and collaborators [129] described the distribution of PLL genes in several plant lineages. PLLs are present in mosses and other land plants, but are absent in algae. 22 genes were found in *Physcomitrella patens*, 26 in *Arabidopsis thaliana*, 30 in *Populus trichocarpa*, 12 in *Oryza sativa*, and 46 in *Brassica rapa*.

Interestingly, the four PLL genes and the one PG that were differentially expressed in the microarray showed higher expression during intrusive growth, meaning that they all may have important roles during the elongation of the fibers. Based on the phylogenetic analysis, we could infer that the four PLLs belong to the same monophyletic clade (Figure 5-2); they are two sets of putative paralogous genes: *PLL4* and *PLL7*, and *PLL9* and *PLL19*. We focused our experiments on one of those genes, *PLL4*, as it showed the highest difference in expression between intrusive growing tissues and tissues undergoing fibers thickening.

### 5.4.1 The putative *PLL4* may be important for fiber growth

In the stem peels, we observed significantly higher expression ( $p < 0.05$ ) of *PLL4* during intrusive growth (point A) as compared to all the other tissues undergoing secondary cell wall deposition (Figure 5-4). For example, the fold difference between stages A and B was 14.3. Meanwhile this was not observed in whole stem tissues, in which there was no significant difference between point A and B. Thus, based on a comparison of the results obtained from whole stem and stem peel tissue (Figure 5-4, Table 5-3), we infer that the expression of *PLL4* is significantly higher in the outer, fiber

bearing tissues during the elongation stage of growth, than in these tissues during the thickening stage, when fibers elongation has ceased, in the whole stem tissues, however, there was no significant difference between some of the tissues undergoing fiber elongation respect to those undergoing fibers thickening (Table 5-3).

#### 5.4.2 PLL4 could aid in the intrusive growth of fibers

As the middle lamella of neighbouring cells needs to be hydrolyzed to allow the intrusive growth of fibers[19], it is thought that the higher expression observed during the intrusive growth stage of *PLL4* is an indication that this gene is involved in the hydrolysis of the middle lamella of the cells being penetrated, allowing the extension of the fiber. It is also possible that it maintains the fibers separated from surrounding cells while they are elongating, avoiding the adhesion between the cells through a newly generated middle lamella, until intrusive growth has ceased.

The pectate lyase activity of PLL4 was successfully confirmed by the enzymatic assay (Figure 5-6). Although the difference between the PLL4 expressing clone and the empty vector controls was significant, the overall activity of PLL4 was low, respect to what was observed in the control. This may be due to the low level of purification obtained, which diluted PLL4 with other proteins. It is necessary to improve the purification of the protein by changing the concentrations of imidazole and the number of washes, in order to have a higher enzymatic activity.

## 5.5 Conclusion

In conclusion we detected four PLL genes and one PG gene with differential expression along the stem. They all showed enrichment in expression during intrusive

growth, which suggests an important role of pectin degrading enzymes during fibers elongation, possibly hydrolyzing pectins in the middle lamella to facilitate fibers intrusive growth between adjacent cells. *PLL4* is an important candidate for gene mutagenesis; however, due to the similarity of expression of the other PLLs it may be necessary to mutate multiple PLLs simultaneously to see a change in phenotype.

## **6 Chapter 6: General Discussion**

In this work we wanted to demonstrate the importance of pectin modifying enzymes on flax fiber development, and identify the associated genes. We hypothesized that the modification of pectins of the fiber and surrounding cells is necessary for elongation and maturation of the fibers as follows: (i) fibers must intrude between neighboring cells, which requires loosening of the middle lamella, which in turn requires demethylesterification by pectin methylesterases (PMEs) and cleavage by polygalacturonases (PGs) and/or pectate lyases (PLs); (ii) when fibers are elongating, their cell wall needs to gain certain rigidity, so the fiber can penetrate between cells and avoid being compressed, so PMEs also likely play a role in this rigidification; and (iii) once the fibers stop elongating a new compound middle lamella is generated, which presumably depends on the activity of PMEs.

### **6.1 The methylesterification state of the cell walls and middle lamella changes during fiber development**

Through analysis of epitopes of four antibodies (Chapter 2), we found that the primary cell wall and middle lamella of the fibers have lower methylesterification than surrounding tissues during intrusive growth and after elongation has ceased, before secondary cell wall is detectable (Figures 2-4 to 2-6). This was correlated with a barely detectable binding of LM20 (binds to highly methylesterified HG) in the fibers (Figures 2-4, 2-5 and 2-7), and a higher presence of galactan in the primary cell wall during

coordinated and intrusive growth, and higher calcium cross-linking during intrusive growth (Figures 2-4, 2-5, 2-9).

It is interesting that the labeling of LM19, which binds to demethylesterified HG, was uniformly high in all of the cells in the first millimeter of the stem (Fig 2-6), when all the cells of the stem are expanding, and the fibers had not initiated intrusive growth. However, in point A (4-5 cm) the LM19 labeling was markedly higher in the fibers. It will be important to analyze this pattern at higher spatial resolution, by observing more sections between the first millimeter and point A, to determine whether the onset of intrusive growth is correlated precisely with an increase in the presence of the LM19 epitope. The start of intrusive growth can be distinguished by the formation of a knee-like shape, and smaller diameter of the cell, at both ends of the fiber [18].

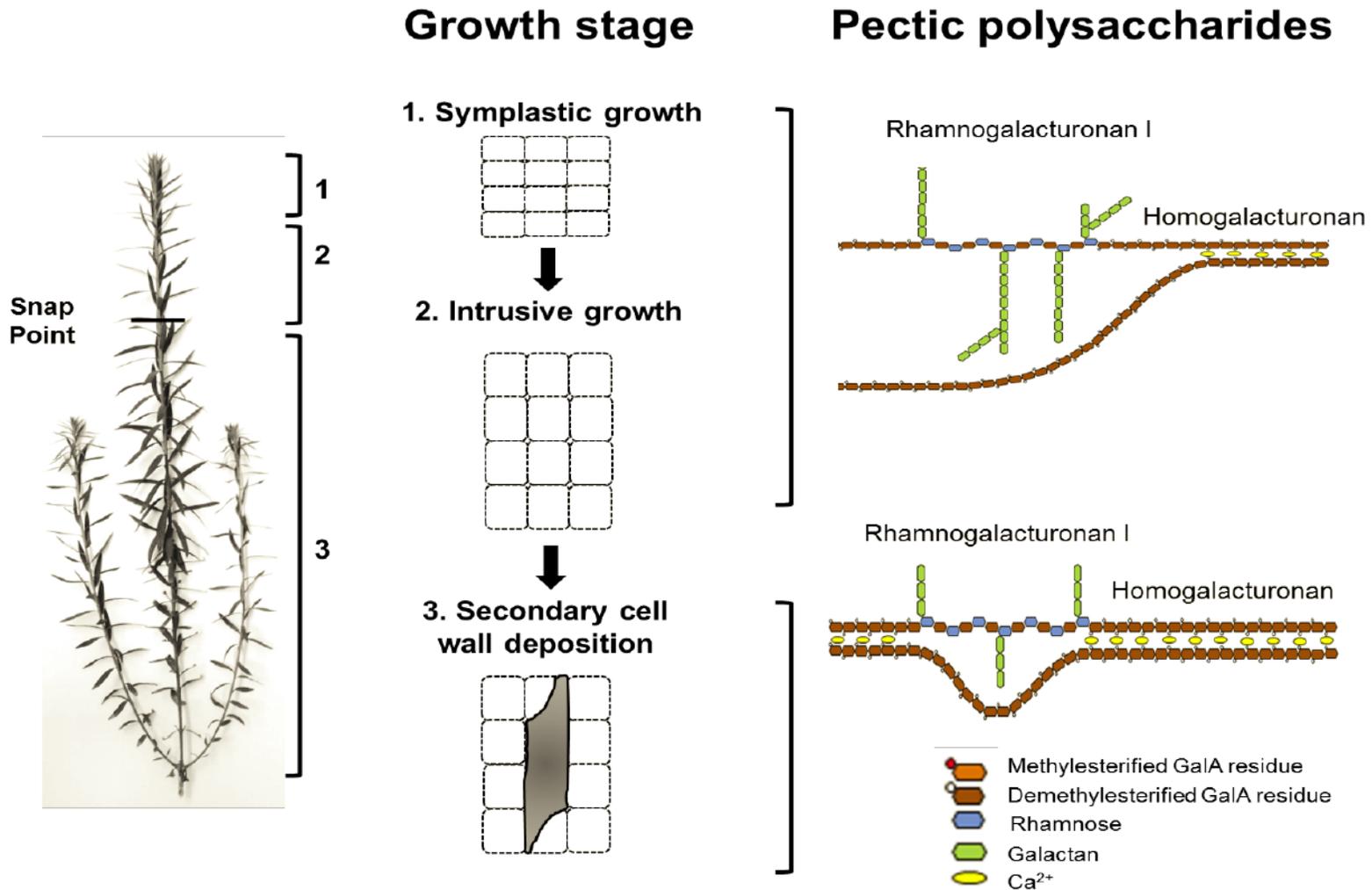
Presumably, the middle lamella between fibers and surrounding tissues reaches its maximum strength only once fibers elongation has ceased. In stage A, we observed that the LM20 epitope (high methylesterification of HG) was absent in the fibers, including the middle lamella binding them (Figs 2-7), however, below the snap-point (stages B-E, Figs 2-7), LM20 could be detected in the fibers, but only in the cell corners of the middle lamella. The possible explanations for why highly methylesterified HG (labeled by LM20) could be detected in the cell corners of the middle lamella, where it was previously undetectable, is that it may have been masked by other epitopes above point B, or, more likely, new HG was deposited at some point between points A and B, before the fibers thicken. This would be correlated with the generation of a new middle lamella, gluing together the fibers with surrounding tissues. It would then be expected that if we studied in detail tissues between stage A and B, we would see a point in which the

binding of LM20 would be high in the whole middle lamella, including the CML, and then these newly synthesized HG would be demethylesterified, leaving some highly methylesterified domains in the cell corners middle lamella.

## **6.2 Galactans side chains of RG-I impede primary cell wall strengthening during elongation**

We found that (1-4)- $\beta$ -galactan is present in higher amounts in the fibers primary cell wall as early as 400  $\mu\text{m}$  from the shoot apex, during coordinate and intrusive growth (Figure 2-9), and then it is degraded in the primary cell wall when elongation ceases (stage B), and a new (1-4)- $\beta$ -galactan is synthesized in the secondary cell wall of the fibers (Stages B to E). Having taken into account that the elongating fibers are largely demethylesterified (Figure 2-6), one possible interpretation for the high abundance of (1-4)- $\beta$ -galactan in the cell wall is that the fiber-enriched galactans acts as spacers to prevent premature formation of calcium cross links between the demethylesterified HG domains associated with elongating fibers. This galactan could then be removed as elongation stops, so that calcium cross-links can be formed and the cell wall can be rigidified. Thus, we hypothesize that the fiber-enriched galactan may play an important role in maintaining plasticity in the cell wall so that fibers can elongate, and removal of this galactan could facilitate rigidification of the primary cell wall by calcium cross linking of demethylesterified HG, once the fiber has stopped growing, so compression from surrounding tissues is avoided, Figure 6-1 illustrates this model. It will be important to establish whether there is a  $\beta$ -galactosidase that is enriched in the transition point between intrusive growth and secondary cell wall degradation, so expression profiling covering

tissues between points A and B will be important to elucidate this. Hobson and Deyholos [137] determined the expression of the  $\beta$ -galactosidases in flax in different tissues and organs in flax; these data will be useful to establish possible candidate genes that can play this role. Mutations of specific  $\beta$ -galactosidases that are active at the transition point could therefore lead to fibers with weaker connections to one another and surrounding tissues. This novel function we have proposed for galactans is somewhat related to what has been proposed by Roach and collaborators [70], wherein galactans prevent premature association and crystallization of cellulose during maturation of the gelatinous-type wall of the fibers.



**Figure 6 - 1 Model showing the effect of galactan side chains on calcium cross linking during fiber elongation.**

The homogalacturonan is largely demethylesterified during the fiber elongation, however long galactan side chains in the RGI impede the formation of calcium cross linking. Once the fibers stop elongating the calcium cross linking increases, presumably due to a reduction in the presence of the galactan side chains.

### **6.3 The PME and PMEIs in flax play diverse roles in the fibers and the plant development in general**

The methylesterification state of the HG in the fibers cell wall and middle lamella seems to be an important factor in the development of the fibers, and as only three flax PMEs had been previously described, we were motivated to characterize the PME and PMEI gene families of flax, and so, create a valuable tool for the definition of their functions in processes such as flower development, xylem development, fiber development, and senescence. 105 putative flax PMEs (LuPMEs) and 95 putative PMEIs (LuPMEIs) were identified within the whole-genome assembly. The proportion of PMEs and PMEIs in the flax genome as a proportion of all predicted proteins was similar to most other dicots (Appendix Figure A3-1). We found experimental evidence for the transcription of 77/105 LuPMEs and 83/95 LuPMEIs, and surveyed the transcript abundance of these in 12 different tissues and stages of development. Six major monophyletic groups of LuPMEs could be defined based on the inferred relationships of flax genes and their presumed orthologs from other species. Three lineage-specific expansions were observed in the LuPMEs (Figure 3-9), the duplications suffered by those genes may indicate roles of the LuPMEs that separate them from the other species analyzed in the phylogenetic tree, it will be important to further investigate them.

Based on their expression profiling and/or homology with previously characterized PMEs/PMEIs in other plants we defined a list of candidate genes that could play a role in fiber development (Table 3-4). However the same approach can be used to identify other

genes involved in processes as important as pollen development in which the activity of PME has been demonstrated [48, 54] to be required for the correct pollen elongation. Flax flowers possess five carpels that are subdivided by a false septum which has up to two ovules [138], so it has the capacity of producing 10 seeds per boll, however, according to the Flax Council of Canada [9], on average it only produces six to eight seeds, which decreases the productivity of the crop. We speculate that LuPMEs could have an effect on the efficiency of the pollen tube elongation and fertilization, which may be another avenue of application of the research presented here.

Certain residues in the tertiary structure of the PMEs and PMEIs are necessary for their correct activity. We searched the LuPMEs and LuPMEIs predicted proteins sequences for conserved residues previously reported [46, 47] to be important for their tertiary structure and function (Figure 3-11). We found a higher conservation on the catalytic residues in the LuPMEs, while in the LuPMEIs, the cysteines forming disulfide bridges between helices  $\alpha 2$  and  $\alpha 3$  were most highly conserved. In reverse genetic approaches that use a fragment of the gene to search for mutations, it is important to establish a region of interest in which, if a silencing mutation is not found, a mutation in a residue important for the activity of the protein can be found. Being able to establish the conserved residues of the LuPMEs and LuPMEIs allowed current studies to screen for mutations in those residues in reverse genetic approaches.

The LuPMEs and LuPMEIs comprise large families with complex patterns of transcript expression and a wide range of physical characteristics. We observed that multiple PMEs and PMEIs were expressed in partially overlapping domains, indicative of

several genes acting redundantly during most processes. The potential for functional redundancy was highlighted also by the phylogenetic analyses.

In order to have a more detailed understanding of the expression of the PME and PMEIs in fiber development, we assessed their expression in nine different stages of fiber development in the stem (Figure 4-1). Among others, two LuPMEs, *LuPME67* and *LuPME79*, showed higher transcript abundance during intrusive growth compared to the bottom of the stem in stem peel tissues, and also had the highest fold change in this specific pattern. These genes could have an important role during the intrusive growth of the fiber, each with a different role as they have predicted acidic (*LuPME67*) and basic isoelectric points (*LuPME79*). Based on our phylogenetic analysis (Figure 3-9) we identified groups of closely related LuPMEs. *LuPME67* is a putative paralog of *LuPME89*, in a clade that contained one PME from poplar (*Populus trichocarpa*), one from *Manihot esculenta* (cassava), one from castor (*Ricinus communis*), and two from Arabidopsis. *LuPME89* was found to be expressed only in reproductive tissues, so it is likely that it doesn't have a redundant activity with *LuPME67*. *LuPME79* is part of a clade composed of four PMEs, and has *LuPME36* as presumed paralog; these both share a common ancestor with *LuPME73* and *LuPME50*, and are all located in a clade with two PMEs from castor, and one from cassava. *LuPME73* expression had a similar pattern in the whole stem and the stem peel (Figure 4- group 2, and Figure 4-3 group 1) but the highest fold difference was only 12, compared to 1085 in *LuPME79*. *LuPME36* expression was not detected in the 12 tissues analyzed (Figure 3-2), and *LuPME50* expression was not assessed individually (only with a common primer with *LuPME73*).

Based on these results, it appears unlikely that the function of *LuPME67* is redundant with its presumptive paralogs. Furthermore, based on the finding of Jiang and collaborators [48] it is not likely that a distant isoform can complement its activity. Thus, if a loss-of-function mutant of *LuPME67* were obtained, it may have a novel phenotype with shorter fibers, resulting from a failure of HG to be randomly demethylesterified, and ultimately inhibiting intrusive growth of fibers between surrounding tissues, this would facilitate the incorporation of the linseed stem into the soil, after harvesting, as the fibers would be shorter and wouldn't affect the machinery [9].

In the case of *LuPME79*, it is possible that *LuPME73* can complement its activity, so if the role of *LuPME79* in flax fiber development were to be studied, it would likely be necessary to mutate both *LuPME73* and *LuPME79*. Two alternative phenotypes could be imagined in a double mutant of these genes: (i) if these genes are required for rigidification of fibers to allow them to intrude between neighbouring cells, mutants would have shorter fibers; (ii) conversely, if the major site of action is in the middle lamella of the surrounding cells, a mutant may have a less rigid middle lamella which would facilitate the intrusive growth of the fibers, and consequently fibers would be longer.

*LuPME61* and *LuPME1* are paralogous genes that had similar patterns of expression. They belong to a clade that includes two Arabidopsis genes, among them *AtPME35*, and one castor PME and one poplar PME. These PMEs appear to be involved in cell wall maturation, because in whole stem tissues they are enriched in point A, where xylem thickening is occurring (Figure 4-2, group 2), while in stem peel tissues they are enriched in point B to E (Figure 4-3 Group 2). *AtPME35* has been found to aid in the strengthening

of the *Arabidopsis* inflorescence stem [77]. Consequently, taking into consideration their expression patterns and marked difference in expression between tissues (based on the fold difference observed), these genes are strong candidates for further studies, so their specific role in fiber and xylem development can be determined. It is interesting that although the estimated divergence of these genes occurred between 2 and 3.8 million years ago, depending on the substitution rate used [108, 109] (Appendix Table A3-3), they maintained very similar levels and patterns of expression (Figure 4-1 to 4-3, Table 4-2).

Another set of paralogous genes are *LuPME45* and *LuPME85*, which diverged approximate 2.3 to 4.2 million years ago. Both of these have similar expression patterns and are among the genes with the highest fold change observed in the whole stem (Table 4-2). They appear to be involved in cell wall maturation. They have maximum expression in point A in the whole stem tissues, and in point B in the stem peel. So, together with *LuPME61* and *LuPME1* they appear to be important targets for understanding the role of the *LuPMEs* during cell wall maturation of xylem cells and fibers. These four genes belong to group C (Figure 3-9) in the phylogenetic tree of the *PMEs*. As their expression is very similar it is possible that their function could be redundant, so in order to analyze their role in flax it may be necessary to generate a mutant plant for the four genes.

When two genes diverge it is expected that their function varies unless higher amounts of the proteins are beneficial for the plant. Consequently, the presence of paralogous genes with similar function is an indication of concerted evolution between the genes [139], so its additive function may be necessary for the fitness of the plant, or in the case of flax for its success as a crop. It is possible that the selection for these genes has been

ligated to the human selection for beneficial phenotypes, meaning that the genotypes in which both genes had the same function were selected.

The expression profile of some paralogous genes was divergent. *LuPMEI44* and *LuPMEI73* are paralogous genes that diverged recently (between 2.72 and 5.04 million years ago, Appendix Table A3-2). Their expression, although not classified in the same cluster by STEM, was similar in the whole stem tissues (enriched in point A) (Figure 4-2), but it was different in the stem peel (Figure 4-3). The silencing or mutants of these inhibitors may lead to an increase in the activity of one or more of the seven LuPMEs that showed a peak in activity in point A, in the whole stem.

As the specific location of the expression of these genes plays a relevant role in the outcome of the phenotype, it will be important to determine the subcellular localization of the proteins, to establish their specific function in fiber development. Promoter fusions and protein fusions are some of the approaches used to detect the localization of the expression of a protein. The use of  $\beta$ -glucuronidase reporters (GUS) has already been successful in flax to determine the expression of a  $\beta$ -galactosidase in the different tissues of the plant at a cellular level [140], and has been used to study PME in Arabidopsis, as in the case of *VANGUARDI* [48]. YFP protein fusions, have been successfully used to study the subcellular localization of the Arabidopsis PME *AtPPMEI* [54], which would allow us to distinguish the expression of the protein in the primary cell wall or the middle lamella.

## 6.4 Pectin degrading enzymes are enriched during fiber elongation

In chapter five we searched for pectin degrading enzymes with differential expression along the stem. We studied in detail one of these genes, a pectate lyase (*PLL4*), which expression was enriched during the intrusive growth of the fibers. As in the case of the LuPMEs and LuPMEIs of interest, it will be important to elucidate the specific parts of the tissues where its expression is highest, in order to establish with certainty if it has a role in fiber development and what kind of role. It could be expected that if this pectate lyase plays a role facilitating fiber elongation, higher amounts of *PLL4* should be found in the middle lamella of the cells surrounding the fibers, which would dismantle the pectin polysaccharides, so the intrusion of the fibers through these cells would be facilitated. Commercial pectate lyases have been used for enzymatic retting of flax fibers, which is an approach to separate the fibers in the plant by using enzymes, such as polygalacturonases and pectate lyases [141, 142], however it is not widely used due to its high cost [143], and dew retting is preferred, although it cannot be implemented in countries with short summers, as Canada, as it relies on endogenous microorganisms on the environment to separate the fibers which need optimal temperature. Being able to use an endogenous pectate lyase from flax, such as *PLL4*, for enzymatic retting, may increase the efficiency of the process, so being able to express higher amounts and a more pure enzyme should be a target in future studies. Furthermore, if the intention is to decrease the length of the fibers [9], a mutant of *PLL4* would be useful, however, based on the phylogenetic tree, it would likely be necessary to mutate or silence its putative paralogous, *PLL7*, and based on the microarray results it would be necessary to mutate also *PLL9* and *PLL19*.

## 6.5 Conclusion

In this study, we established the importance of the methylesterification state of the cell wall and the presence of galactan. We were able to propose a model in which the fiber cell wall exhibits a very low level of methylesterification, but avoided rigidity during its elongation by the presence of long galactan side chains in the RG-I. These long galactan side chains are then degraded once elongation has ceased to allow the strengthening of the fiber cell wall. Also, we characterized the complete families of LuPMEs and LuPMEIs, and we established their importance in fiber development. A subset of genes, including a pectate lyase-like gene, with expression enrichments during different points of development have also been identified, and their possible roles during fiber development were discussed. These genes may provide a basis for the improvement of key traits in industrial feedstocks and a better understanding of the physiological roles of PMEs and PMEIs in general.

## 7 Bibliography

1. Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R *et al*: **The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads**. *Plant Journal* 2012, **72**(3):461-473.
2. McDill J, Repplinger M, Simpson BB, Kadereit JW: **The phylogeny of *Linum* and Linaceae subfamily Linoideae, with implications for their systematics, biogeography, and evolution of heterostyly**. *Systematic Botany* 2009, **34**(2):386-405.
3. Kvavadze E, Bar-Yosef O, Belfer-Cohen A, Boaretto E, Jakeli N, Matskevich Z, Meshveliani T: **30,000-year-old wild flax fibers**. *Science* 2009, **325**(5946):1359.
4. Cullis C: **Flax**. In: *Oilseeds*. Edited by Kole C, vol. 2: Springer Berlin Heidelberg; 2007: 275-295.
5. Foulk J, Akin D, Dodd R, Ulven C: **Production of Flax Fibers for Biocomposites**. In: *Cellulose Fibers: Bio- and Nano-Polymer Composites*. Edited by Kalia S, Kaith BS, Kaur I: Springer Berlin Heidelberg; 2011: 61-95.
6. Vaisey-Genser M, Morris D: **Introduction: History of the cultivation and uses of flaxseed**. In: *Flax : the genus linum Medicinal and aromatic plants--industrial profiles*. Edited by Muir AD, Westcott ND. London: Routledge; 2003: xii, 307 p.
7. Deyholos M: **Bast fiber of flax (*Linum usitatissimum* L.): Biological foundations of its ancient and modern uses**. *Israel Journal of Plant Sciences* 2007, **54**(4):273-280.

8. Akin DE, Foulk JA, Dodd RB, McAlister DD, 3rd: **Enzyme-retting of flax and characterization of processed fibers.** *Journal of Biotechnology* 2001, **89**(2-3):193-203.
9. Rowland G: **Growing Flax, Production, Management & Diagnostic Guide.** Winnipeg: Flax Council of Canada and Saskatchewan Flax Development Commission; 2002.
10. Evert RF, Esau K, Esau K: **Esau's Plant anatomy : Meristems, Cells, and Tissues of the Plant Body : Their Structure, Function, and Development,** 3rd edn. Hoboken, N.J.: Wiley-Interscience; 2006.
11. Morvan C, Andeme-Onzighi C, Girault R, Himmelsbach DS, Driouich A, Akin DE: **Building flax fibres: more than one brick in the walls.** *Plant Physiology and Biochemistry* 2003, **41**(11-12):935-944.
12. Gorshkova TA, Wyatt SE, Salnikov VV, Gibeaut DM, Ibragimov MR, Lozovaya VV, Carpita NC: **Cell-wall polysaccharides of developing flax plants.** *Plant Physiology* 1996, **110**(3):721-729.
13. Day A, Ruel K, Neutelings G, Cronier D, David H, Hawkins S, Chabbert B: **Lignification in the flax stem: evidence for an unusual lignin in bast fibers.** *Planta* 2005, **222**(2):234-245.
14. Sharma HSS, Faughey G, Lyons G: **Comparison of physical, chemical, and thermal characteristics of water-, dew-, and enzyme-retted flax fibers.** *Journal of Applied Polymer Science* 1999, **74**(1):139-143.
15. Gorshkova T, Morvan C: **Secondary cell-wall assembly in flax phloem fibres: role of galactans.** *Planta* 2006, **223**(2):149-158.

16. Gorshkova TA, Chemikosova SB, Sal'nikov VV, Pavlencheva NV, Gur'janov OP, Stolle-Smits T, van Dam JEG: **Occurrence of cell-specific galactan is coinciding with bast fiber developmental transition in flax.** *Industrial Crops and Products* 2004, **19**(3):217-224.
17. Gorshkova TA, Ageeva M, Chemikosova S, Salnikov V: **Tissue-specific processes during cell wall formation in flax fiber.** *Plant Biosystems* 2005, **139**(1):88-92.
18. Ageeva MV, Petrovska B, Kieft H, Sal'nikov VV, Snegireva AV, van Dam JE, van Veenendaal WL, Emons AM, Gorshkova TA, van Lammeren AA: **Intrusive growth of flax phloem fibers is of intercalary type.** *Planta* 2005, **222**(4):565-574.
19. Snegireva AV, Ageeva MV, Amenitskii SI, Chernova TE, Ebskamp M, Gorshkova TA: **Intrusive growth of sclerenchyma fibers.** *Russian Journal of Plant Physiology* 2010, **57**(3):342-355.
20. Gorshkova TA, Sal'nikova VV, Chemikosova SB, Ageeva MV, Pavlencheva NV, van Dam JEG: **The snap point: a transition point in *Linum usitatissimum* bast fiber development.** *Industrial Crops and Products* 2003, **18**(3):213-221.
21. Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, Pilate G: **Plant fiber formation: state of the art, recent and expected progress, and open questions.** *Critical Reviews in Plant Sciences* 2012, **31**(3):201-228.
22. Jarvis MC, Briggs SPH, Knox JP: **Intercellular adhesion and cell separation in plants.** *Plant Cell and Environment* 2003, **26**(7):977-989.

23. Albersheim P, Darvill A, Roberts K, Sederoff R, Staehelin A: **Plant cell walls: From Chemistry to Biology**. New York, NY: Garland Science; 2011.
24. Scheller HV, Ulvskov P: **Hemicelluloses**. *Annual Review of Plant Biology* 2010, **61**(1):263-289.
25. Wolf S, Greiner S: **Growth control by cell wall pectins**. *Protoplasma* 2012, **249**:S169-175.
26. Gurjanov OP, Ibragimova NN, Gnezdilov OI, Gorshkova TA: **Polysaccharides, tightly bound to cellulose in cell wall of flax bast fibre: Isolation and identification**. *Carbohydrate Polymers* 2008, **72**(4):719-729.
27. Mikshina P, Chernova T, Chemikosova S, Ibragimova N, Mokshina N, Gorshkova T: **Cellulosic Fibers: Role of Matrix Polysaccharides in Structure and Function**. In: *Cellulose - Fundamental Aspects*. Edited by van de Ven T, Godbout L: InTech; 2013.
28. Mohnen D: **Pectin structure and biosynthesis**. *Current Opinion in Plant Biology* 2008, **11**(3):266-277.
29. Ridley BL, O'Neill MA, Mohnen D: **Pectins: structure, biosynthesis, and oligogalacturonide-related signaling**. *Phytochemistry* 2001, **57**(6):929-967.
30. Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, 3rd, Orlando R, Scheller HV, Mohnen D: **Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(50):20225-20230.

31. Zhang GF, Stachelin LA: **Functional compartmentation of the golgi-apparatus of plant-cells - immunocytochemical analysis of high-pressure frozen-substituted and freeze-substituted sycamore maple suspension-culture cells.** *Plant Physiology* 1992, **99**(3):1070-1083.
32. Goubet F, Council LN, Mohnen D: **Identification and partial characterization of the pectin methyltransferase "homogalacturonan-methyltransferase" from membranes of tobacco cell suspensions.** *Plant Physiology* 1998, **116**(1):337-347.
33. Willats WG, Orfila C, Limberg G, Buchholt HC, van Alebeek GJ, Voragen AG, Marcus SE, Christensen TM, Mikkelsen JD, Murray BS *et al*: **Modulation of the degree and pattern of methyl-esterification of pectic homogalacturonan in plant cell walls. Implications for pectin methyl esterase action, matrix properties, and cell adhesion.** *Journal of Biological Chemistry* 2001, **276**(22):19404-19413.
34. O'Neill MA, Ishii T, Albersheim P, Darvill AG: **Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide.** *Annual Review of Plant Biology* 2004, **55**(1):109-139.
35. Liners F, Letesson JJ, Didembourg C, Van Cutsem P: **Monoclonal antibodies against pectin: recognition of a conformation induced by calcium.** *Plant Physiology* 1989, **91**(4):1419-1424.
36. Caffall KH, Mohnen D: **The structure, function, and biosynthesis of plant cell wall pectic polysaccharides.** *Carbohydrate Research* 2009, **344**(14):1879-1900.

37. Stolle-Smits T, Beekhuizen JG, Kok MT, Pijnenburg M, Recourt K, Derksen J, Voragen AG: **Changes in cell wall polysaccharides of green bean pods during development.** *Plant Physiology* 1999, **121**(2):363-372.
38. Richard L, Qin LX, Goldberg R: **Clustered genes within the genome of *Arabidopsis thaliana* encoding pectin methylesterase-like enzymes.** *Gene* 1996, **170**(2):207-211.
39. Jolie RP, Duvetter T, Van Loey AM, Hendrickx ME: **Pectin methylesterase and its proteinaceous inhibitor: a review.** *Carbohydrate Research* 2010, **345**(18):2583-2595.
40. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics.** *Nucleic Acids Research* 2009, **37**(Database issue):D233-238.
41. Wolf S, Mouille G, Pelloux J: **Homogalacturonan methyl-esterification and plant development.** *Molecular Plant* 2009, **2**(5):851-860.
42. Micheli F: **Pectin methylesterases: cell wall enzymes with important roles in plant physiology.** *Trends in Plant Science* 2001, **6**(9):414-419.
43. Koch JL, Nevins DJ: **Tomato fruit cell-wall 1. Use of purified tomato polygalacturonase and pectinmethylesterase to identify developmental changes in pectins.** *Plant Physiology* 1989, **91**(3):816-822.
44. Balestrieri C, Castaldo D, Giovane A, Quagliuolo L, Servillo L: **A glycoprotein inhibitor of pectin methylesterase in kiwi fruit (*Actinidia chinensis*).** *European Journal of Biochemistry* 1990, **193**(1):183-187.

45. Marquis H, Bucheli P: **Inhibition of tomato pectin methylesterase by partially purified kiwi pectin methylesterase inhibitor protein.** *International Journal of Food Science and Technology* 1994, **29**(2):121-128.
46. Hothorn M, Wolf S, Aloy P, Greiner S, Scheffzek K: **Structural insights into the target specificity of plant invertase and pectin methylesterase inhibitory proteins.** *Plant Cell* 2004, **16**(12):3437-3447.
47. Di Matteo A, Giovane A, Raiola A, Camardella L, Bonivento D, De Lorenzo G, Cervone F, Bellincampi D, Tsernoglou D: **Structural basis for the interaction between pectin methylesterase and a specific inhibitor protein.** *Plant Cell* 2005, **17**(3):849-858.
48. Jiang L, Yang SL, Xie LF, Pua CS, Zhang XQ, Yang WC, Sundaresan V, Ye D: **VANGUARD1 encodes a pectin methylesterase that enhances pollen tube growth in the Arabidopsis style and transmitting tract.** *Plant Cell* 2005, **17**(2):584-596.
49. Gaffe J, Tiznado ME, Handa AK: **Characterization and functional expression of a ubiquitously expressed tomato pectin methylesterase.** *Plant Physiology* 1997, **114**(4):1547-1556.
50. Roger D, Lacoux J, Lamblin F, Gaillet D, Dauchel H, Klein D, Balange AP, David A, Laine E: **Isolation of a flax pectin methylesterase promoter and its expression in transgenic tobacco.** *Plant Science* 2001, **160**(4):713-721.
51. Lacoux J, Klein D, Domon JM, Burel C, Lamblin F, Alexandre F, Sihachakr D, Roger D, Balange AP, David A *et al*: **Antisense transgenesis of *Linum usitatissimum* with a pectin methylesterase cDNA.** *Plant Physiology and Biochemistry* 2003, **41**(3):241-249.

52. Al-Qsous S, Carpentier E, Klein-Eude D, Burel C, Mareck A, Dauchel H, Gomord V, Balange AP: **Identification and isolation of a pectin methylesterase isoform that could be involved in flax cell wall stiffening.** *Planta* 2004, **219**(2):369-378.
53. Mareck A, Lamour R, Schaumann A, Chan P, Driouich A, Pelloux J, Lerouge P: **Analysis of LuPME3, a pectin methylesterase from *Linum usitatissimum*, revealed a variability in PME proteolytic maturation.** *Plant Signaling & Behavior* 2012, **7**(1):59-61.
54. Tian GW, Chen MH, Zaltsman A, Citovsky V: **Pollen-specific pectin methylesterase involved in pollen tube growth.** *Developmental Biology* 2006, **294**(1):83-91.
55. Francis KE, Lam SY, Copenhaver GP: **Separation of Arabidopsis pollen tetrads is regulated by QUARTET1, a pectin methylesterase gene.** *Plant Physiology* 2006, **142**(3):1004-1013.
56. Rhee SY, Osborne E, Poindexter PD, Somerville CR: **Microspore separation in the quartet 3 mutants of Arabidopsis is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation.** *Plant Physiology* 2003, **133**(3):1170-1180.
57. Siedlecka A, Wiklund S, Peronne MA, Micheli F, Lesniewska J, Sethson I, Edlund U, Richard L, Sundberg B, Mellerowicz EJ: **Pectin methyl esterase inhibits intrusive and symplastic cell growth in developing wood cells of *Populus*.** *Plant Physiology* 2008, **146**(2):554-565.
58. Giovane A, Balestrieri C, Quagliuolo L, Castaldo D, Servillo L: **A glycoprotein inhibitor of pectin methylesterase in kiwi fruit. Purification by affinity**

- chromatography and evidence of a ripening-related precursor.** *European Journal of Biochemistry* 1995, **233**(3):926-929.
59. Zhang GY, Feng J, Wu J, Wang XW: **BoPMEI1, a pollen-specific pectin methylesterase inhibitor, has an essential role in pollen tube growth.** *Planta* 2010, **231**(6):1323-1334.
60. Cooke JE, Martin TA, Davis JM: **Short-term physiological and developmental responses to nitrogen availability in hybrid poplar.** *New Phytologist* 2005, **167**(1):41-52.
61. Marin-Rodriguez MC, Orchard J, Seymour GB: **Pectate lyases, cell wall degradation and fruit softening.** *Journal of Experimental Botany* 2002, **53**(377):2115-2119.
62. Wang H, Guo Y, Lv F, Zhu H, Wu S, Jiang Y, Li F, Zhou B, Guo W, Zhang T: **The essential role of GhPEL gene, encoding a pectate lyase, in cell wall loosening by depolymerization of the de-esterified pectin during fiber elongation in cotton.** *Plant Molecular Biology* 2010, **72**(4-5):397-406.
63. Hatfield RD, Ralph J, Grabber JH: **Cell wall cross-linking by ferulates and diferulates in grasses.** *Journal of the Science of Food and Agriculture* 1999, **79**(3):403-407.
64. Verhertbruggen Y, Marcus SE, Haeger A, Ordaz-Ortiz JJ, Knox JP: **An extended set of monoclonal antibodies to pectic homogalacturonan.** *Carbohydrate Research* 2009, **344**(14):1858-1862.
65. Jones L, Seymour GB, Knox JP: **Localization of pectic galactan in tomato cell walls using a monoclonal antibody specific to (1[->]4)-[Beta]-D-Galactan.** *Plant Physiology* 1997, **113**(4):1405-1412.

66. Liners F, Thibault JF, Van Cutsem P: **Influence of the degree of polymerization of oligogalacturonates and of esterification pattern of pectin on their recognition by monoclonal antibodies.** *Plant Physiology* 1992, **99**(3):1099-1104.
67. Esau K: **Vascular differentiation in the vegetative shoot of *Linum*. III. The origin of the bast fibers.** *American Journal of Botany* 1943, **30**(8):579.
68. Gorshkova TA, Gurjanov OP, Mikshina PV, Ibragimova NN, Mokshina NE, Salnikov VV, Ageeva MV, Amenitskii SI, Chernova TE, Chemikosova SB: **Specific type of secondary cell wall formed by plant fibers.** *Russian Journal of Plant Physiology* 2010, **57**(3):328-341.
69. Gorshkova TA, Chemikosova SB, Lozovaya VV, Carpita NC: **Turnover of galactans and other cell wall polysaccharides during development of flax plants.** *Plant Physiology* 1997, **114**(2):723-729.
70. Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, Gorshkova TA: **Development of cellulosic secondary walls in flax fibers requires beta-galactosidase.** *Plant Physiology* 2011, **156**(3):1351-1363.
71. Mikshina PV, Chemikosova SB, Mokshina NE, Ibragimova NN, Gorshkova TA: **Free galactose and galactosidase activity in the course of flax fiber development.** *Russian Journal of Plant Physiology* 2009, **56**(1):58-67.
72. His I, Andeme-Onzighi C, Morvan C, Driouich A: **Microscopic studies on mature flax fibers embedded in LR white: Immunogold localization of cell wall matrix polysaccharides.** *Journal of Histochemistry & Cytochemistry* 2001, **49**(12):1525-1535.

73. Andeme-Onzighi C, Girault R, His I, Morvan C, Driouich A: **Immunocytochemical characterization of early-developing flax fiber cell walls.** *Protoplasma* 2000, **213**(3-4):235-245.
74. Palin R, Geitmann A: **The role of pectin in plant morphogenesis.** *BioSystems* 2012, **109**(3):397-402.
75. Derbyshire P, McCann MC, Roberts K: **Restricted cell elongation in Arabidopsis hypocotyls is associated with a reduced average pectin esterification level.** *BMC Plant Biology* 2007, **7**:31.
76. Hall HC, Cheung J, Ellis BE: **Immunoprofiling reveals unique cell-specific patterns of wall epitopes in the expanding Arabidopsis stem.** *Plant Journal* 2013, **74**(1):134-147.
77. Hongo S, Sato K, Yokoyama R, Nishitani K: **Demethylesterification of the primary wall by PECTIN METHYLESTERASE35 provides mechanical support to the Arabidopsis stem.** *Plant Cell* 2012, **24**(6):2624-2634.
78. Downie B, Dirk LM, Hadfield KA, Wilkins TA, Bennett AB, Bradford KJ: **A gel diffusion assay for quantification of pectin methylesterase activity.** *Analytical Biochemistry* 1998, **264**(2):149-157.
79. Schneider CA, Rasband WS, Eliceiri KW: **NIH Image to ImageJ: 25 years of image analysis.** *Nature Methods* 2012, **9**(7):671-675.
80. Lionetti V, Francocci F, Ferrari S, Volpi C, Bellincampi D, Galletti R, D'Ovidio R, De Lorenzo G, Cervone F: **Engineering the cell wall by reducing de-methylesterified homogalacturonan improves saccharification of plant tissues for bioconversion.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(2):616-621.

81. Dubois M, Gilles K, Hamilton JK, Rebers PA, Smith F: **A colorimetric method for the determination of sugars.** *Nature* 1951, **168**(4265):167.
82. Willats WG, Knox JP: **Immunoprofiling of pectic polysaccharides.** *Analytical Biochemistry* 1999, **268**(1):143-146.
83. Reynolds ES: **Citation Classic - the Use of Lead Citrate at High Ph as an Electron-Opaque Stain in Electron-Microscopy.** *Current Contents/Life Sciences* 1981, **32**:208-212.
84. Slocum RD, Roux SJ: **An improved method for the sub-cellular localization of calcium using a modification of the antimonate precipitation technique.** *Journal of Histochemistry & Cytochemistry* 1982, **30**(7):617-629.
85. Plavcova L, Hacke UG: **Heterogeneous distribution of pectin epitopes and calcium in different pit types of four angiosperm species.** *New Phytologist* 2011, **192**(4):885-897.
86. Jauneau A, Morvan C, Lefebvre F, Demarty M, Ripoll C, Thellier M: **Differential extractability of calcium and pectic substances in different wall regions of epicotyl cells in young flax plants.** *Journal of Histochemistry & Cytochemistry* 1992, **40**(8):1183-1189.
87. Willats WG, Steele-King CG, Marcus SE, Knox JP: **Side chains of pectic polysaccharides are regulated in relation to cell proliferation and cell differentiation.** *Plant Journal* 1999, **20**(6):619-628.
88. Harholt J, Suttangkakul A, Vibe Scheller H: **Biosynthesis of pectin.** *Plant Physiology* 2010, **153**(2):384-395.

89. Morvan O, Quentin M, Jauneau A, Mareck A, Morvan C: **Immunogold localization of pectin methylesterases in the cortical tissues of flax hypocotyl.** *Protoplasma* 1998, **202**(3-4):175-184.
90. Pelloux J, Rusterucci C, Mellerowicz EJ: **New insights into pectin methylesterase structure and function.** *Trends in Plant Science* 2007, **12**(6):267-277.
91. Wolf S, Rausch T, Greiner S: **The N-terminal pro region mediates retention of unprocessed type-I PME in the Golgi apparatus.** *Plant Journal* 2009, **58**(3):361-375.
92. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al*: **The Pfam protein families database.** *Nucleic Acids Research* 2012, **40**(Database issue):D290-301.
93. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M *et al*: **The Arabidopsis information resource (TAIR): improved gene annotation and new tools.** *Nucleic Acids Research* 2012, **40**(Database issue):D1202-1210.
94. Stanke M, Morgenstern B: **AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints.** *Nucleic Acids Research* 2005, **33**(Web Server issue):W465-467.
95. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al*: **Phytozome: a comparative platform for green plant genomics.** *Nucleic Acids Research* 2012, **40**(Database issue):D1178-1186.

96. Huis R, Hawkins S, Neutelings G: **Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.).** *BMC Plant Biology* 2010, **10**:71.
97. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nature Methods* 2011, **8**(10):785-786.
98. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *Journal of Molecular Biology* 2001, **305**(3):567-580.
99. Horton P, Park KJ, Obayashi T, Nakai K: **Protein subcellular localization prediction with WOLF PSORT.** *Proceedings of the 4th Asia-Pacific Bioinformatics Conference* 2006, **3**:39-48.
100. Chou KC, Shen HB: **Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization.** *PLoS One* 2010, **5**(6):e11335.
101. Li GQ, Moriyama EN: **Vector NTI, a balanced all-in-one sequence analysis suite.** *Briefings in Bioinformatics* 2004, **5**(4):378-388.
102. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research* 2004, **32**(5):1792-1797.
103. Darriba D, Taboada GL, Doallo R, Posada D: **ProtTest 3: fast selection of best-fit models of protein evolution.** *Bioinformatics* 2011, **27**(8):1164-1165.
104. Zwickl DJ: **Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion.** Austin, Texas: The University of Texas at Austin; 2006.

105. Miller MA, Pfeiffer W, Schwartz T: **Creating the CIPRES Science Gateway for inference of large phylogenetic trees.** In: *Gateway Computing Environments Workshop (GCE), 2010: 14-14 Nov. 2010 2010*; 2010: 1-8.
106. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Molecular Biology and Evolution* 2011, **28**(10):2731-2739.
107. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences.** *Journal of Molecular Evolution* 1980, **16**(2):111-120.
108. Koch MA, Haubold B, Mitchell-Olds T: **Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae).** *Molecular Biology and Evolution* 2000, **17**(10):1483-1498.
109. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**(5494):1151-1155.
110. Arabidopsis GI: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**(6814):796.
111. Fenart S, Ndong YP, Duarte J, Riviere N, Wilmer J, van Wuytswinkel O, Lucau A, Cariou E, Neutelings G, Gutierrez L *et al*: **Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray.** *BMC Genomics* 2010, **11**:592.

112. Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M *et al*: **Gene expression analysis of flax seed development**. *BMC Plant Biology* 2011, **11**:74.
113. Gaffe J, Morvan C, Jauneau A, Demarty M: **Partial-purification of flax cell-wall pectin methylesterase**. *Phytochemistry* 1992, **31**(3):761-765.
114. Mareck A, Gaffe J, Morvan O, Alexandre C, Morvan C: **Characterization of isoforms of pectin methylesterase of *Linum usitatissimum* using polyclonal antibodies**. *Plant and Cell Physiology* 1995, **36**(3):409-417.
115. Alexandre F, Morvan O, Gaffe J, Mareck A, Jauneau A, Dauchel H, Balange AP, Morvan C: **Pectin methylesterase pattern in flax seedlings during their development**. *Plant Physiology and Biochemistry* 1997, **35**(6):427-436.
116. Louvet R, Cavel E, Gutierrez L, Guenin S, Roger D, Gillet F, Guerineau F, Pelloux J: **Comprehensive expression profiling of the pectin methylesterase gene family during silique development in *Arabidopsis thaliana***. *Planta* 2006, **224**(4):782-791.
117. Pinzon-Latorre D, Deyholos MK: **Characterization and transcript profiling of the pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) gene families in flax (*Linum usitatissimum*)**. *BMC Genomics* 2013, **14**:742.
118. Wolf S, Grsic-Rausch S, Rausch T, Greiner S: **Identification of pollen-expressed pectin methylesterase inhibitors in *Arabidopsis***. *FEBS Letters* 2003, **555**(3):551-555.
119. Raiola A, Camardella L, Giovane A, Mattei B, De Lorenzo G, Cervone F, Bellincampi D: **Two *Arabidopsis thaliana* genes encode functional pectin methylesterase inhibitors**. *FEBS Letters* 2004, **557**(1-3):199-203.

120. Dedeurwaerder S, Menu-Bouaouiche L, Mareck A, Lerouge P, Guerineau F: **Activity of an atypical *Arabidopsis thaliana* pectin methylesterase.** *Planta* 2009, **229**(2):311-321.
121. De-la-Peña C, Badri DV, Vivanco JM: **Novel role for pectin methylesterase in *Arabidopsis*: a new function showing ribosome-inactivating protein (RIP) activity.** *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 2008, **1780**(5):773-783.
122. Giovane A, Quagliuolo L, Castaldo D, Servillo L, Balestrieri C: **Pectin methyl esterase from *Actinidia-chinensis* fruits.** *Phytochemistry* 1990, **29**(9):2821-2823.
123. Rillo L, Castaldo D, Giovane A, Servillo L, Balestrieri C, Quagliuolo L: **Purification and properties of pectin methylesterase from mandarin orange fruit.** *Journal of Agricultural and Food Chemistry* 1992, **40**(4):591-593.
124. Ernst J, Bar-Joseph Z: **STEM: a tool for the analysis of short time series gene expression data.** *BMC Bioinformatics* 2006, **7**:191.
125. To L: **Genetics of Seed Coat and Stem Development in Flax (*Linum usitatissimum* L.).** Edmonton, AB, Canada: University of Alberta; 2013.
126. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP: **Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations.** *Biotechnology Letters* 2004, **26**(6):509-515.
127. Esau K: **Anatomy of Seed Plants**, 2d edn. New York: Wiley; 1977.
128. Palusa SG, Golovkin M, Shin SB, Richardson DN, Reddy AS: **Organ-specific, developmental, hormonal and stress regulation of expression of putative pectate lyase genes in *Arabidopsis*.** *New Phytologist* 2007, **174**(3):537-550.

129. Jiang J, Yao L, Miao Y, Cao J: **Genome-wide characterization of the Pectate Lyase-like (PLL) genes in *Brassica rapa***. *Molecular Genetics and Genomics* 2013, **288**(11):601-614.
130. Wing RA, Yamaguchi J, Larabell SK, Ursin VM, McCormick S: **Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen *Erwinia***. *Plant Molecular Biology* 1990, **14**(1):17-28.
131. Israelsson M, Eriksson ME, Hertzberg M, Aspeborg H, Nilsson P, Moritz T: **Changes in gene expression in the wood-forming tissue of transgenic hybrid aspen with increased secondary growth**. *Plant Molecular Biology* 2003, **52**(4):893-903.
132. Jimenez-Bermudez S, Redondo-Nevado J, Munoz-Blanco J, Caballero JL, Lopez-Aranda JM, Valpuesta V, Pliego-Alfaro F, Quesada MA, Mercado JA: **Manipulation of strawberry fruit softening by antisense expression of a pectate lyase gene**. *Plant Physiology* 2002, **128**(2):751-759.
133. Vogel JP, Raab TK, Schiff C, Somerville SC: **PMR6, a pectate lyase-like gene required for powdery mildew susceptibility in *Arabidopsis***. *Plant Cell* 2002, **14**(9):2095-2106.
134. Soriano M, Diaz P, Pastor FIJ: **Pectate lyase C from *Bacillus subtilis*: a novel endo-cleaving enzyme with activity on highly methylated pectin**. *Microbiology* 2006, **152**(3):617-625.
135. Collmer A, Ried JL, Mount MS: **Assay-methods for pectic enzymes**. *Methods in Enzymology* 1988, **161**:329-335.

136. Manfield IW, Bernal AJ, Moller I, McCartney I, Riess NP, Knox JP, Willats WGT: **Re-engineering of the PAM1 phage display monoclonal antibody to produce a soluble, versatile anti-homogalacturonan scFv.** *Plant Science* 2005, **169**(6):1090-1095.
137. Hobson N, Deyholos MK: **Genomic and expression analysis of the flax (*Linum usitatissimum*) family of glycosyl hydrolase 35 genes.** *BMC Genomics* 2013, **14**:344.
138. Schewe LC, Sawhney VK, Davis AR: **Ontogeny of floral organs in flax (*Linum usitatissimum*; Linaceae).** *American Journal of Botany* 2011, **98**(7):1077-1085.
139. Zhang JZ: **Evolution by gene duplication: an update.** *Trends in Ecology & Evolution* 2003, **18**(6):292-298.
140. Hobson N, Deyholos MK: **LuFLA1PRO and LuBGAL1PRO promote gene expression in the phloem fibres of flax (*Linum usitatissimum*).** *Plant Cell Reports* 2013, **32**(4):517-528.
141. Zhang JQ, Zhang JC: **Effect of refined processing on the physical and chemical properties of hemp bast fibers.** *Textile Research Journal* 2010, **80**(8):744-753.
142. Akin DE, Condon B, Sohn M, Foulk JA, Dodd RB, Rigsby LL: **Optimization for enzyme-retting of flax with pectate lyase.** *Industrial Crops and Products* 2007, **25**(2):136-146.
143. Deyholos MK, Potter S: **Engineering bast fiber feedstocks for use in composite materials.** *Biocatalysis and Agricultural Biotechnology* 2013(0).
144. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(9):5116-5121.

145. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J: **MeV: MultiExperiment Viewer**. In: *Biomedical Informatics for Cancer Research*. Edited by Ochs MF, Casagrande JT, Davuluri RV: Springer US; 2010: 267-277.

## **8 Appendix**

### **8.1 Microarray analysis**

A flax stem microarray [125] analysis was conducted on 3 week old plants. This analysis compared five different 1-cm points in the stem, the first three corresponding to intrusive growth, at 0-1 cm from the shoot apex, 2-3 cm, and 3-4 cm, and the fourth and fifth ones corresponding to secondary cell wall deposition, at 4-5 cm and 8-9 cm.

To define differentially regulated genes, Significance Analysis of Microarrays (SAM) software developed by Tusher and collaborators [144] was used in MultiExperiment Viewer (MeV, [145]) by running a multi-class SAM, to determine the genes that had differential expression in at least one of the five points in the stem that were assessed. We ran a multi-class SAM using delta 0.32, which generated 1147 significant elements (971 unique genes) and a median number of 71 false positives, out of 94928 elements analyzed.

**Table A3 - 1 Primers used for qRT-PCR, indicating TaqMan probe.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PME-1	LuPME1	Lus10025510	127	77	ccatccgatccatccaata	agttgctcagcgtgtcc
PME-2	LuPME2	Lus10000045	230	69	aaggagccaccgattcat	agatatcgctgcctcgaa
PME-1	LuPME3	Lus10039314	57	77	acgggtcgaccactttcaac	cccttgctaggaaacctg
PME-2	LuPME4	Lus10000936	234	132	tcttatcaacagagatgggtgg	ttatggtcaataaaaaacaaatgg
PME-1	LuPME5	Lus10013344	58	19	ggaggagaggatgggtgg	actacgtggcgctcacc
PME-1	LuPME6	Lus10002976	74	164	cacaagattgccatgaaaa	acaatctccaatgccatcc
PME-1	LuPME7	Lus10003933	61	56	gatccaacaaaacacagg	cccaagtacgtcgggaaac
PME-1	LuPME8	Lus10003934	62	83	tgacggattctccatcagatt	tgacgacgccctctatcc
PME-1	LuPME9	Lus10003968	40	25	ggcctcaaggcttaattgc	tgcatggaatataatgggtaaaaa
PME-2	LuPME10	Lus10004720	41	42	ggagaacaacgacgagaatgt	agcgttatgtacggcttcgt
PME-1	LuPME11	Lus10005587	73	39	cacttctctataattacccccaga	ggaaagggaatttggaggact
PME-2	LuPME12	Lus10005678	133	152	cggttcaaatcaacggaat	caaatccgtccccctagcac
PME-1	LuPME13	Lus10006103	30	65	gattcatcttcggcaacg	tgctgtgtttggtcaga
PME-1	LuPME14	Lus10006940	35	163	ccaaggtcactggtaggaaga	atgaaccagggtgccactac
PME-2	LuPME15	Lus10007113_Lus10007114_DPL	237	49	cagggtatttatttgggcaga	ctctcagcgactgcttg
PME-2	LuPME16	Lus10007136	79	135	cagctccgatagagatgtgg	gtgtggcattaatagggtgg
PME-1	LuPME17	Lus10008203	123	83	tgggctctgacatagatgaca	acgtactgagcgcaaatgag
PME-1	LuPME18	Lus10008937	70	70	tcctcatcctccgactc	gagcttcgggtagagactgg
PME-1	LuPME19	Lus10009110	52	114	ttaccgtccgcaactttt	aacaccgacaggctgactc
PME-2	LuPME20	Lus10009287	239	22	caccggtagccggaatc	ccgaaacacaaattaccgaag
PME-2	LuPME21	Lus10009997	65	1	gactggggcaatgctgac	cacggtagcctccagctat
PME-1	LuPME22	Lus10010170_DPL	139	4	ccagaaaggattggacaaa	tatagcccgtcggctcttt
PME-1	LuPME23	Lus10010309	78	12	acaaacgtgcacccttt	ctggttaccacgaccgfat
PME-2	LuPME24	Lus10010470	83	59	gcaatgctaggctccttacc	taccgcccgtttatgat
PME-2	LuPME25	Lus10010912_a_DPL	90	75	tgcatatgctcagaccgact	tggtgtagatttcggagagac

**Table A3 – 1 Primers used for qRT-PCR, indicating TaqMan probe. Continued.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PME-2	LuPME26	Lus10010912_b_DPL	N.A.			
PME-2	LuPME27	Lus10011132	31	18	cagttccattccctctga	gttggggtgaaftgagcag
PME-2	LuPME28	Lus10011760	240	60	gcaggagcaggaaattgagg	tactccctccggcactg
PME-2	LuPME29	Lus10012942	1	32	acggcttatectgttcgag	atactgttctccctgtgc
PME-1	LuPME30	Lus10001466_Lus10001467_DPL	235	143	ggcgtacttcgagaacgtg	ccatcccatccgatga
PME-1	LuPME31	Lus10013416	241	58	ggcagaacggatcctaata	gcggcctgatgtacaat
PME-1	LuPME32	Lus10013720	93	31	tcaccatcaaaaagtactcc	gccatgaaaatfgatgacg
PME-2	LuPME33	Lus10013721	100	12	acaagaactctctgggacga	tgtgcaacgaatgaaaacag
PME-2	LuPME34	Lus10014338	94	71	gcaacaactccggactctaa	gaaattttggcttgggaaca
PME-2	LuPME35	Lus10014733	242	81	cttatcccggtcaagatcca	cacaatggcagtcacgaat
PME-1	LuPME36	Lus10015210	63	101	gtgtcactggcactcttgc	tgcgaaaaagttggaattct
PME-1	LuPME37	Lus10015211	64	142	acggagccaacaagactttg	ccactgcctggctgtgt
PME-1	LuPME38	Lus10015877	137	2	caccgatcataaacgaaac	ccactctccgctgatgtt
PME-2	LuPME39	Lus10016605	24	86	ggactccatcaccgtcaag	aacctctgattgtgaaatgcc
PME-2	LuPME40	Lus10016678	131	63	ttatcacaccgactctcca	gggcaagtgtgtaattgtg
PME-2	LuPME41	Lus10016711	95	20	tgtaacattggctccgcat	aggaaatccagtgatgcac
PME-2	LuPME42	Lus10016915	N.A.			
PME-1	LuPME43	Lus10017118_Lus10017119_DPL	266	14	gcgtcgacagcatgaaac	tcatgtaatccactggggtt
PME-1	LuPME44	Lus10017375_DPL	71	102	gcgacaaagctgcgattt	atgtgtccggaatcgtcatt
PME-1	LuPME45	Lus10017665	42	75	ttcgggtcatgggtgtt	agttcccgccatcaaat
PME-1	LuPME46	Lus10018103	34	146	agttccgaattgggttgg	tctgccgacacgattatcc
PME-1	LuPME47	Lus10018335	132	37	catggaaaggcagcagaaca	gatgacgagtcgggactt
PME-2	LuPME48	Lus10020313	97	10	ttcatctctcttgcggcta	aaattctctcaagggtgagg
PME-1	LuPME49	Lus10020677	98	139	gtaggagggcgaagtgt	ttggttagaccagctggaactt
PME-1	LuPME50	Lus10020678_Lus10020679_Lus10020680_DPL	267	9	ttctctccattcttacttca	cgaaggaggaaatgttgga

**Table A3 – 1 Primers used for qRT-PCR, indicating TaqMan probe. Continued.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PME-1	LuPME51	Lus10020681	136	88	cgaattcagaccgatatgtgaa	caccacatgaccctactca
PME-2	LuPME52	Lus10020888_DPL	3	132	ggfggtgccgtatgacg	cctcttccaccagattgc
PME-1	LuPME53	Lus10020909	4	12	gttcacgccgtcgaagt	aaggaacaccactgtccgtaa
PME-1	LuPME54	Lus10022412	59	91	agggggaggaggaattaagt	ggcggaa gactcagcagt
PME-2	LuPME55	Lus10023560	231	33	tctcagaattgggtccaag	aggcattgaggagttcca
PME-1	LuPME56	Lus10023775	13	77	attactggcggcaacgac	cacgtcggatacccattctc
PME-2	LuPME57	Lus10024049_Lus10024050_DPL	245	88	ttcgtacacttacatggatca	ctacctggcctgaagcacat
PME-2	LuPME58	Lus10000034	51	40	attccgcctctgcaactg	caaatcgccatgaaacc
PME-2	LuPME59	Lus10026047	94	71	gcaacaactccggactctaa	gaaattttggcttgggaaca
PME-2	LuPME60	Lus10026347	33	58	atggctctgatgctctcc	cttgattccgcgtaaatgc
PME-2	LuPME61	Lus10026729	85	64	ggcca gacctaacgattg	atatcacggcacggctga
PME-1	LuPME62	Lus10027202	44	145	tgccctctttgtccagt	cggaaatccgtctgaatcaa
PME-1	LuPME63	Lus10027203	45	35	ggfcaacggggc gatact	cacgtcatgacgtgactctct
PME-1	LuPME64	Lus10027204	46	99	gacgattctgccgtgga g	gctgcttaaccacgittgg
PME-1	LuPME65	Lus10027206	246	1	tgaa gactctgtattacgggaggt	cgggccacttactctat
PME-1	LuPME66	Lus10027655_Lus10027656_DPL	15	148	gccaa gcaatggggata ga	acgcgtattgctactgcttg
PME-2	LuPME67	Lus10027737	69	77	atcgattccgggacctaca	tgccctcca gtaaccaagt
PME-2	LuPME68	Lus10028364	9	53	caaccctcatcctctgac	cgcttggtttactcattggaa
PME-2	LuPME69	Lus10028536	247	38	caggggtccttcgagacata	gcgctcttataccagtgctc
PME-1	LuPME70	Lus10028882	86	149	ccacttcaattctctgtttc	cgggtttcggaa ggtga
PME-1	LuPME71	Lus10029866_a_DPL	50	33	accgtcatcgtgctctca	gacgctcga gaggaaatc
PME-1	LuPME72	Lus10029866_b_DPL	51	40	attccgcctctgcaactg	caaatcgccatgaaacc
PME-1	LuPME73	Lus10029867_g25305	49	18	ccacctcaattctccaca	ctaaacgtcaaccccacagc
PME-1	LuPME74	Lus10029868	98	139	gtagggacggcgaagtgt	ttggtta gaccagctggaactt
PME-1	LuPME75	Lus10031140	128	70	ttccaaggatgcgtaacg	actgggcggatcatgt

**Table A3 – 1 Primers used for qRT-PCR, indicating TaqMan probe. Continued.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PME-1	LuPME76	Lus10031141	248	7	tcaggacgcgacgatt	aatccctgctcactcttatgat
PME-2	LuPME77	Lus10031421_DPL	5	71	agcga gtaggcgacctg	attgtagaaggcacttctgcg
PME-1	LuPME78	Lus10031469	6	160	cggcgttagcacttatcaca	tctcgtccggaaggctctt
PME-1	LuPME79	Lus10031470	7	52	taaggagacttcggctcgtg	aacaaccacctcttctca
PME-1	LuPME80	Lus10031719	249	105	ccacctcaaccataccttgc	ggtccataaacctgactgg
PME-1	LuPME81	Lus10031720	232	12	cgtctgggaa gaa gttgtca	tcctcttccaatggaacc
PME-1	LuPME82	Lus10033399	140	8	ggagaa gtaga gaa gcatg	gtactgccgctccatct
PME-1	LuPME83	Lus10033466	135	152	tctctccatccttctggtg	tgaaatgcctccgtgac
PME-1	LuPME84	Lus10033486	138	108	gaggccgtggaa gga gtat	caccactcccatttca
PME-1	LuPME85	Lus10033621	250	77	ccgagacaggaaattgttgc	caaccacgacgttgcatta
PME-1	LuPME86	Lus10034859	12	8	tgttagagacgtaga gaaaacac	ccatcttcgccaccactat
PME-2	LuPME87	Lus10034893	75	136	tcggatccaaccaagggtg	gctggtcatgactgccttg
PME-2	LuPME88	Lus10034981	125	17	tgga gaa gtaaacataccacaa	gagcagctcctcccattct
PME-2	LuPME89	Lus10035553_DPL	27	92	gggaatgcacgactcattct	tcgcctgagctgtaattgag
PME-2	LuPME90	Lus10036006_Lus10036007(g942)_DPL	95 and 251	89	atcaccacttaaccgatgc	gaaccaagtctatgatgacag
PME-1	LuPME91	Lus10037456_Lus10037457_DPL	126 and 252	16	agctccctccgactgaatc	gctccgagatgctggagtag
PME-1	LuPME92	Lus10037458_Lus10037459(g6229)_DPL	124	70	gaggfgaa gcc gaacgtg	ctccaccgtcgtgtagtcc
PME-1	LuPME93	Lus10037489	29	33	atagc gac gac gactcatca	aacctcctttctcggagt
PME-2	LuPME94	Lus10037774	254	45	gtccaa gca gtagacaa gtagc	gcccatggaacacgtttg
PME-1	LuPME95	Lus10038917	37	145	gcccttgccttctcagtaaat	cttgaaacccgtctgagtc
PME-1	LuPME96	Lus10038918	233	73	caactcagcaaacatatccac	cagaacttccgcttctcgt
PME-1	LuPME97	Lus10038919	39	77	ggatctc gatacactatcattaccag	tattccacgtgccacaatc
PME-2	LuPME98	Lus10000935	84	10	cggcca gtc gatttac ga	gcctctctgagtagctgtg
PME-1	LuPME99	Lus10039927	80	149	cccacaagacgac gacaaa	ttcaccatttctgtagctgg
PME-2	LuPME100	Lus10040446	130	52	ccccttggagcaaggaa	ccagcttgaaaattgtggac

**Table A3 – 1 Primers used for qRT-PCR, indicating TaqMan probe. Continued.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PME-2	LuPME101	Lus10040541	54	150	ggactacaaaaccatacaagaagcta	gcacggaatgtgacttttccc
PME-2	LuPME102	Lus10041699	56	131	acaccgcctccatgattg	accaaacgtccctgtaccaa
PME-2	LuPME103	Lus10041815	89	6	tgatgttattgctgggtgatcc	gtcctgtctcaggggatg
PME-2	LuPME104	Lus10042315	53	36	cgtctctgcaaacacgacag	agccgattcaaaaatattgtcac
PME-2	LuPME105	Lus10043035	255	31	gtaggggagcaggggcta	ggaaaaggcttagcttcatcg
PMEI	LuPMEI1	Lus10000822	158	156	caatccagcgttgaaggag	cgacattgacgacgacctta
PMEI	LuPMEI2	Lus10000861	170	81	aatcgggtcgaacattc	tcgatggtgtggttaacaatg
PMEI	LuPMEI3	Lus10000961	262	92	atattcattcggcgttgac	acaggtgcttcactctctc
PMEI	LuPMEI4	Lus10001448	256	41	aggcaataaaaagggtttg	tccggttgtattcttctca
PMEI	LuPMEI5	Lus10001464	263	145	accctattcccacacaaaa	gggttctcgaacaaatct
PMEI	LuPMEI6	Lus10001658	178	3	gatgattgcaagggtgcaa	cttaataaatgccggttcc
PMEI	LuPMEI7	Lus10001659	175	9	gtgtcacgtcgaatattgcttattg	gtataataatcttccgcaattt
PMEI	LuPMEI8	Lus10001848	142	149	accgacgtgtggacctatc	tcttgaaggcttctcgttt
PMEI	LuPMEI9	Lus10001988	120	152	ttctcctctctcttctct	ttgagggtctcgaattacg
PMEI	LuPMEI10	Lus10002237	201	64	agaacctccagtcgtgcaat	accaacctgctcgaaggf
PMEI	LuPMEI11	Lus10002420	257	140	gccacagatcatgtcgcgat	ttccggttgtatccttctc
PMEI	LuPMEI12	Lus10002738	156	63	aatctgccgccttttc	gtgtcggcgattagagaaac
PMEI	LuPMEI13	Lus10002739	164	25	ctttcccacaacaacaac	ggaggagaggatgaattgagg
PMEI	LuPMEI14	Lus10002933	166	97	tccggagagcttcacag	ggtgccttctctgcaattc
PMEI	LuPMEI15	Lus10003530	264	157	taatgatgctcccaaggag	gcagccacaacagatgtca
PMEI	LuPMEI16	Lus10004327	122	101	cctgttccataagcagaaa	ctgccaaagtcccatttta
PMEI	LuPMEI17	Lus10007598	183	140	aatcggaaaattgcgagag	caccgtctatcaaccgta
PMEI	LuPMEI18	Lus10007873	218	63	aatgttattcttttctcagtgac	tgcaagcttgtccacgta
PMEI	LuPMEI19	Lus10008201	223	81	cctacatcaagagaatgttgagca	tcgaaagcgttttggagagtc
PMEI	LuPMEI20	Lus10008625	116	152	atcatcctctcatcgtgct	tttaggattcgttctgct

**Table A3 – 1 Primers used for qRT-PCR, indicating TaqMan probe. Continued.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PMEI	LuPMEI21	Lus10008892	169	106	ggaaaattgggatttcaatg	attcgttggtctgtccgatt
PMEI	LuPMEI22	Lus10010002	199	41	cccttgacacgtgcgtaga	gctgctattgcgtccaatg
PMEI	LuPMEI23	Lus10011452_DPL	207	78	agctcaaacatagccccatc	ggatc gatgacgacgag
PMEI	LuPMEI24	Lus10011453	203	19	gagggatccttctcgtctc	ccgtgga gcaagtcttgaa
PMEI	LuPMEI25	Lus10015199	225	12	aaatacaagcccctgtgcat	catgtcctcacctccgatt
PMEI	LuPMEI26	Lus10016317	157	69	ccccacctctcacaacct	ggc ggaattgaggagat
PMEI	LuPMEI27	Lus10016318	150	39	gaaaggacggtggtttaca	cttcaactccgggtcgtc
PMEI	LuPMEI28	Lus10016319	165	102	ttgggggtttaaagggttgac	cgataaaaaggaaaatctggaaca
PMEI	LuPMEI29	Lus10017013	153	55	gtggaa gcggtgagtaaa gg	cctcga gcgtcactcattg
PMEI	LuPMEI30	Lus10017040	151	142	gctcctggattgttcgatt	aactgcctgtcgtgatcgt
PMEI	LuPMEI31	Lus10017074	162	149	gga gtaa ggtgggga gfga g	ttaacaaatctgcggccttt
PMEI	LuPMEI32	Lus10017076	155	1	ccgagcgtgtgatggaat	cgtagtcacctccgtca
PMEI	LuPMEI33	Lus10017077	148	66	agagcagcagcaactacacg	ccttagccaa ggtgatgtcc
PMEI	LuPMEI34	Lus10017345	179	63	cgacatcgttcgtactccttt	ggccgccacataacaaac
PMEI	LuPMEI35	Lus10017346	176	60	cttcggcgaatgcaaafta	ftggcagactcaatgtcacg
PMEI	LuPMEI36	Lus10017347	177	153	gcggctcagacgtttgtc	cctcttgacagctccgaac
PMEI	LuPMEI37	Lus10018615	196	52	atgcgacgctgggttfaat	cgcatctcaacaatcttca
PMEI	LuPMEI38	Lus10019498	194	161	caatcaggcttgcttgca g	ggtttgagggaaggatggtg
PMEI	LuPMEI39	Lus10020664	226	9	aaatgctcacgaatactacca	gaggggtgtgtgtgtgtgt
PMEI	LuPMEI40	Lus10022409	172	101	accggcacgcagtaattc	tccattgaa gfga gaata ggaa
PMEI	LuPMEI41	Lus10023114	173	134	gcataacaacgacaccacctt	ggctgctttgtcgcatac
PMEI	LuPMEI42	Lus10023215	168	71	gtgatgcgctc gctagtgt	ggcaggtcgtttgtttcc
PMEI	LuPMEI43	Lus10024593	111	30	fggttccatccctctgata	gatggtgtgtttgggttg
PMEI	LuPMEI44	Lus10024595	219	160	atcatctccgcaaaactcc	ttttgatgaa gtccgca g
PMEI	LuPMEI45	Lus10024596	216	52	ctcttggagttccgggttg	agccggtccacagagtca

**Table A3 – 1 Primers used for qRT-PCR, indicating TaqMan probe. Continued.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PMEI	LuPMEI46	Lus10025021	198	143	tggtgacgataaggcga	cctctcggaacctcactt
PMEI	LuPMEI47	Lus10025781	204	146	ctcgacacttgcaccagaa	ctggtcttcttcgcaatc
PMEI	LuPMEI48	Lus10025782_a_DPL	258	84	agacgatcgatgccctgg	ctcgcttctctcgaat
PMEI	LuPMEI49	Lus10025782_b_DPL	206	84	ccagttcaagaacctcca	ctcgcttctctcgaat
PMEI	LuPMEI50	Lus10025806	212	11	tgtctgattggtgttgagtt	caccctagtgtgatgacagt
PMEI	LuPMEI51	Lus10025807	213	82	atcatcaccatccatgtcca	gtggaagggaatccgttc
PMEI	LuPMEI52	Lus10026435	268	102	gctgtctaaaggagagatc	gtcgaaggcaagtgtccgat
PMEI	LuPMEI53	Lus10027198	114	69	caaaccctctcataaccatca	caggagggtttgatgagtcg
PMEI	LuPMEI54	Lus10027199	115	70	aaatggcatcaacttctcct	tggtggaaggagattatgatggt
PMEI	LuPMEI55	Lus10027556	260	1	ctgtccaccagaaagacct	ctcattgtctccgcacgctc
PMEI	LuPMEI56	Lus10027768	181	25	ttggaagcgaagggttcg	gcaatcgcgtaaaacttcc
PMEI	LuPMEI57	Lus10027947	159	133	gatcaaggcggctcgtgtag	acttgttccagcgtcgttc
PMEI	LuPMEI58	Lus10028910	105	82	ttcaattcccctgttctat	ctgccgacgttctcattttt
PMEI	LuPMEI59	Lus10029877	227	149	acgacgtcatcga gtc gta	tgaacctatctcacaage
PMEI	LuPMEI60	Lus10030292	121	153	aactcgaaccctcgcacac	ttggcattgctttacgatg
PMEI	LuPMEI61	Lus10030370	217	97	cccgcaccagctcaatag	tattgaaaggacgccatgagt
PMEI	LuPMEI62	Lus10030926	269	108	ttaaatatgttgcattggtgttt	cacagaaccaacaatatactcca
PMEI	LuPMEI63	Lus10031132	259	46	gcgtggaacaatgaa ggt	caccacgtctgca ggtt
PMEI	LuPMEI64	Lus10031133	103	42	gca gctctctctgtcttacgc	gacttactccaacgcagtct
PMEI	LuPMEI65	Lus10031138	102	128	cccagttaccaaaatccag	atgacagcggccatgaa g
PMEI	LuPMEI66	Lus10031197	171	39	gtcctgctcaagtcatcg	atgaggcggctatgttgg
PMEI	LuPMEI67	Lus10031483	224	84	gaa gatgga gggccac gtc	gataatctccccgacgctc
PMEI	LuPMEI68	Lus10031711	110	161	aacatgaa ggttc ggt gga	gaaatccggaccgtccat
PMEI	LuPMEI69	Lus10031712	109	22	tcctctcaccatttcttcc	tttggtactggcgggtgtt
PMEI	LuPMEI70	Lus10031713	108	17	gcctgatcagagc gca gct	gcctcagaccttgaacct

**Table A3 – 1 Primers used for qRT-PCR, indicating TaqMan probe. Continued.**

Type	Gene code	Flax genome version 1.0 identifier	primer #	TaqMan probe #	Fw primer	Rv Primer
PMEI	LuPMEI71	Lus10031717	107	70	ccaccaagatccagaccag	gacttgccc gatgacagg
PMEI	LuPMEI72	Lus10032230	117	70	tcagctccgcctacatct	gaccgcctggtcaaacctc
PMEI	LuPMEI73	Lus10032232	220	149	gcgggga gga tagtggf	tcttctgcattccggta
PMEI	LuPMEI74	Lus10032233	215	30	cgaaatc gatc ga gac gag	ccaccgagtcaccaact
PMEI	LuPMEI75	Lus10035527	182	143	cgcaattcaaaactttcc	tgaggcagtgatggaggat
PMEI	LuPMEI76	Lus10037540	202	83	tgatcgtcgtattgtctctggt	ggttcttgaacacgaattgg
PMEI	LuPMEI77	Lus10037541	209	49	ttaaggacatgaatggtgcaa	gcttggaaaacccaaaagc
PMEI	LuPMEI78	Lus10037791	163	144	tcaattcacgaa gaaacaacaac	tcattttggttggaaattgg
PMEI	LuPMEI79	Lus10037792	154	139	cgctactcaactttccgtctt	tgcagggttccgttatgaga
PMEI	LuPMEI80	Lus10037793	147	132	cgctgctc gaaa gacct	gatcggagggatgagatgc
PMEI	LuPMEI81	Lus10037919	118	160	cgctgccttggtaacagt	tcacgtacgaccgaattgat
PMEI	LuPMEI82	Lus10038294	214	164	aagctaaaaggaa ggtggtgaa	gtatatatagtggtcggatcggaga
PMEI	LuPMEI83	Lus10038295	211	88	ccctcaaataccgacca	tccatcgtc gaa gggagat
PMEI	LuPMEI84	Lus10038645	101	101	tcctcaacctccgtctt	ccgaggaaa gagggttagg
PMEI	LuPMEI85	Lus10038737	187	101	caccatgtcttcccactga	agttcaccgccgatgatg
PMEI	LuPMEI86	Lus10038738	188	69	cattcggctaaaccttctg	tftgtggtgcagctctgc
PMEI	LuPMEI87	Lus10038914	112	144	cctctcatctgcaacaatgg	gaggaggagggtgga gattg
PMEI	LuPMEI88	Lus10038915	113	127	tgtgtaaagctga gttatgtgtct	actccaataacctcaactgttt
PMEI	LuPMEI89	Lus10039120	189	10	atcactccaatgctcaaaag	gga gta gtagta gc gggac gag
PMEI	LuPMEI90	Lus10039849	197	134	ctgccac gfggatcaaga	gcc gtc gac gaacttct
PMEI	LuPMEI91	Lus10040119	193	101	cagtactaacgacggctctgc	tgacggc gaacgaa gttatt
PMEI	LuPMEI92	Lus10040145	184	78	ttcgacttgaaccttgtca	tttgggggaa gctattgaa
PMEI	LuPMEI93	Lus10041650	180	4	cgacaaagacatgaaccgta	ggatatccc gggagtgga
PMEI	LuPMEI94	Lus10042193_DPL	106	155	gtaatagcttggc gattctc g	ccagaacagagaa gaga gatcaa
PMEI	LuPMEI95	Lus10043346	195	10	atgccaccaaccaagtg	ggaatta gc gttgtcgtggt

**Table A3 - 2 Genetic distance between possible paralogs in the LuPMEIs gene family**

Gene 1	Gene 2	t = K/2r Millions	
		r=1.50E <sup>-08</sup>	r=8.1E <sup>-09</sup>
LuPMEI3	LuPMEI92	0.84	1.56
LuPMEI14	LuPMEI15	1.09	2.01
LuPMEI20	LuPMEI94	1.1	2.03
LuPMEI13	LuPMEI28	1.36	2.52
LuPMEI48	LuPMEI49	1.57	2.9
LuPMEI4	LuPMEI11	1.62	2.99
LuPMEI63	LuPMEI68	1.68	3.11
LuPMEI56	LuPMEI75	1.72	3.19
LuPMEI50	LuPMEI83	1.75	3.24
LuPMEI45	LuPMEI74	1.75	3.24
LuPMEI9	LuPMEI60	1.76	3.26
LuPMEI53	LuPMEI87	1.85	3.43
LuPMEI16	LuPMEI58	1.9	3.52
LuPMEI81	LuPMEI84	1.91	3.53
LuPMEI86	LuPMEI89	1.93	3.57
LuPMEI24	LuPMEI76	2.06	3.82
LuPMEI54	LuPMEI88	2.24	4.15
LuPMEI64	LuPMEI70	2.25	4.16
LuPMEI25	LuPMEI67	2.37	4.38
LuPMEI32	LuPMEI79	2.44	4.52
LuPMEI1	LuPMEI57	2.62	4.85
LuPMEI44	LuPMEI73	2.72	5.04
LuPMEI5	LuPMEI19	2.82	5.22
LuPMEI31	LuPMEI78	2.83	5.25
LuPMEI65	LuPMEI71	2.95	5.47
LuPMEI38	LuPMEI95	3.45	6.38
LuPMEI43	LuPMEI72	3.61	6.68
LuPMEI12	LuPMEI26	3.63	6.72
LuPMEI39	LuPMEI59	3.78	7
LuPMEI6	LuPMEI34	3.89	7.2
LuPMEI33	LuPMEI80	4.66	8.62
LuPMEI22	LuPMEI46	4.71	8.72
LuPMEI37	LuPMEI90	6.9	12.77
LuPMEI21	LuPMEI42	7.02	13
LuPMEI29	LuPMEI30	10.26	19
LuPMEI18	LuPMEI61	12.34	22.85
LuPMEI35	LuPMEI36	13.43	24.87
LuPMEI41	LuPMEI93	19.86	36.78
LuPMEI8	LuPMEI55	29.95	55.47
	<b>Mean</b>	4.53	8.39
	<b>St. dev.</b>	5.72	10.59

The Kimura 2-parameter model was used to calculate the genetic distance, which was used to calculate the divergence time using  $t = K/2r$ , where  $t$  is time,  $K$  is the genetic distance, and  $r$  is the substitution rate, either  $1.5 \times 10^{-8}$  [108] or  $8.1 \times 10^{-9}$  [109].

**Table A3 - 3. Genetic distance between possible paralogs in the LuPMEs gene family.**

Gene 1	Gene 2	t = K/2r Millions	
		r=1.50E <sup>-08</sup>	r=8.1E <sup>-09</sup>
LuPME41	LuPME90	0.9	1.7
LuPME103	LuPME68	1.7	3.1
LuPME18	LuPME70	1.8	3.3
LuPME66	LuPME99	1.8	3.4
LuPME56	LuPME9	1.9	3.6
LuPME19	LuPME69	2	3.6
LuPME43	LuPME47	2	3.7
LuPME46	LuPME54	2	3.8
LuPME1	LuPME61	2	3.8
LuPME12	LuPME48	2.1	4
LuPME49	LuPME74	2.2	4
LuPME8	LuPME91	2.2	4
LuPME45	LuPME85	2.3	4.2
LuPME15	LuPME39	2.8	5.2
LuPME20	LuPME38	3.1	5.8
LuPME82	LuPME86	3.2	6
LuPME22	LuPME44	3.8	7
LuPME52	LuPME84	4.1	7.5
LuPME23	LuPME31	4.3	8
LuPME67	LuPME89	10.1	18.8
LuPME28	LuPME5	62.2	115.1
LuPME11	LuPME33	8	14.8
LuPME53	LuPME83	2	3.7
LuPME37	LuPME78	2.2	4
LuPME51	LuPME58	4	7.3
LuPME75	LuPME80	2.2	4
LuPME76	LuPME81	2.7	5.1
LuPME63	LuPME96	6.3	11.7
LuPME62	LuPME95	2.9	5.3
LuPME50	LuPME73	1.4	2.6
LuPME36	LuPME79	2.6	4.9
LuPME7	LuPME92	4.3	8
LuPME17	LuPME35	11.1	20.6
LuPME34	LuPME59	1	1.8
LuPME102	LuPME57	2.7	5
LuPME104	LuPME60	5.9	10.9
LuPME100	LuPME55	6.9	12.8
LuPME42	LuPME94	2.4	4.5
LuPME16	LuPME40	3	5.6
LuPME4	LuPME98	4.3	7.9
LuPME26	LuPME77	76.4	141.4
LuPME105	LuPME27	1.9	3.5
LuPME29	LuPME88	4.3	8
	<b>Mean</b>	6.4	11.9
	<b>St. dev.</b>	14.3	26.47

The Kimura 2-parameter model was used to calculate the genetic distance, which was used to calculate the divergence time using  $t = K/2r$ , where t is time, K is the genetic distance, and r is the substitution rate, either  $1.5 \times 10^{-8}$  [108] or  $8.1 \times 10^{-9}$  [109].

Area of halo  $\text{mm}^2$

**Figure A2 - 1 Standard curve of PME activity by radial assay.**

Proteins extracted from the whole stem were used at different concentration in a radial assay to assess the correlation with the area of the halo they produced.

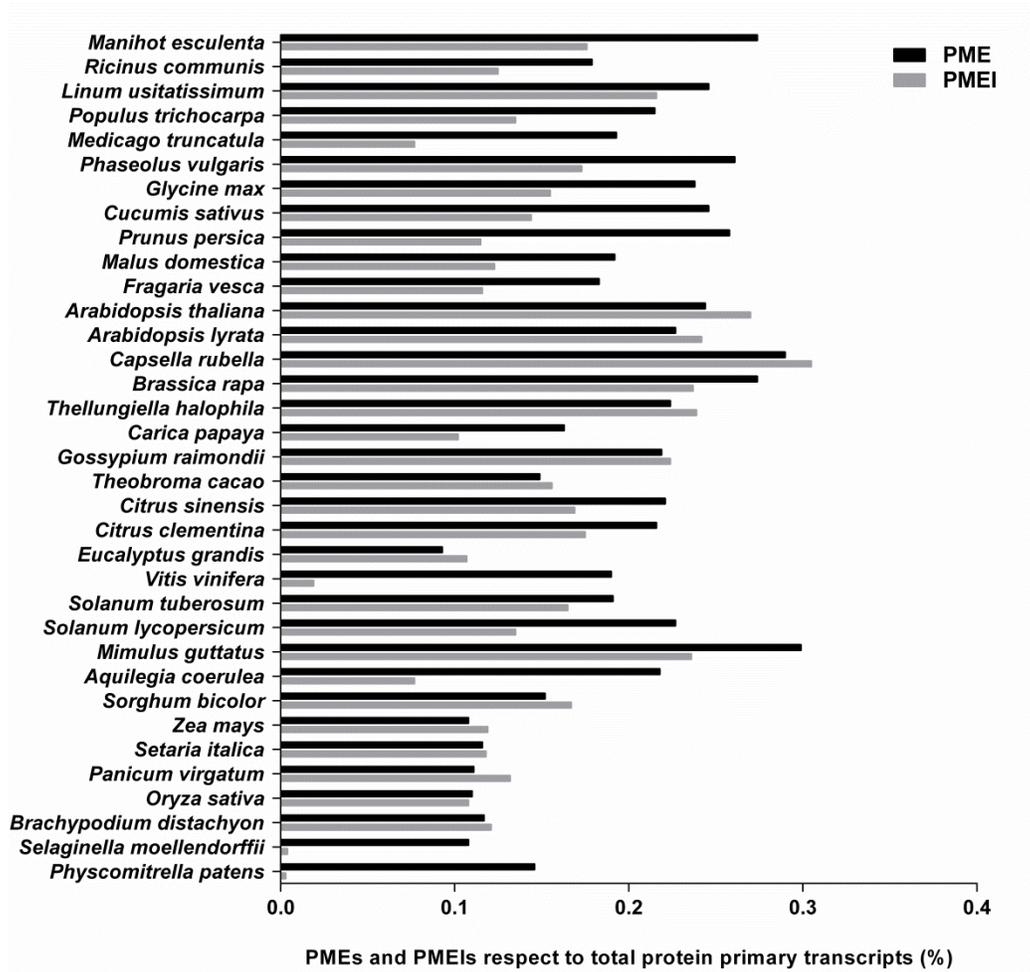


Figure A3 - 1 Percentage of PMEs and PMEIs, respect to the total number of proteins, in Embryophyta plants with available full genomes in Phytozome (version 9.1)

Additional file A4-1

>LuPME79\_no-signal-peptide\_codon-optimized

ATGGTCTGCCAGCGCCGGATAAATTCTGTAACACCACGCCTTTCCCGAGCTTTGCCGTTCCACTCTGCCGT  
ACACTGAAAACCTCCTCATTATGACTATACTCGTAAATCCGTTACACAGAGCCTGTCCAGGCTCGTCAATT  
CCGTCGCCTGATCAATTACCTGCTGAAGCAAGCGTCCAACTGAAGACCTTTCGTACCCTGGTCCCGCTCTG  
CAGGACTGTCAGTTCCTGGCACAGAGCAACTTTGACTCTCTGAGCAACACCATGTCTGCCATCCGCTCCAAA  
GATAAACTGCAGACCTTCCAGGCCTCTGACCTGCAGACTCTGCTGTCTGCTACTCTGACTAACCTGGATACT  
GCCTGGATGGTATTCGTTCTGCGGACTCCGACGTCATTAATAGCCTGGTTGCCCGCTGTGAAAAACGGTAC  
CAAATACTGTAGCGTTTCTCTGGCCTGTTCCGCAAAGGTTGGATTCCGCCACCGAAACGCCGCTCTCCGAAA  
GCTGAATTTTCTCATGAATTCGCGACCCTGGATCTGTCTGACGGCCTGCCGCTGGCTATGTCTGGCCACGACC  
GCTCCTTCTCGAAAGCTTCTCTCGCATGAAAGTTAAAGAAACCAGCGGTTCGCGGGCGGTGGTGGCGGTGTAC  
TGGTAAAGCAAGATGGTGGTTGTTTCTAAGAAAGAAGGCGTCGGCAACTTTACCACTATCAACGAAAGTATTG  
AGTCCGCGCCGAACAGCACCGCCGGTACGGTTACTTCGTTATCTACGTGACCGAGGGCTACTACGAAGAAT  
ACATCTCCATCCGAAAAAGAAAAACACAACTGATGATTATCGGCGACGGCATCGGCAAAAACCACTATTTCCG  
GCAACCGTTCGGTATCGATGGTGGACCACGTTCAATTCCAGCACTTTCGCTGTTGTTGCACAGGGCTTCGT  
GGCGGTTGGTATCACTTTAAAAACACCGCGGGTCCGAAAAACGTCAAGCGGTAGCAGTGCCTAACGGTGC  
GGACATGAGCGCATTCTCAACTGTTCTTCCAGGGTATCAGGATACCCTGTATGTGCACTCTTCCGTCAG  
TTTTATCGCGATTGCGAAGTATACGGCACGGTGGACTTCATCTTTGGCAACGCAGCCGCTGTATTCAGA  
GCCGTATCATGTCCGCTCTGCCGATGGCGCATCAGTTTAAACGCAATTACCGCTCAGAGCCGCTACTGATCCGAA  
TCAGAACACCGGTATCAGCATCCAAAACGCTCTATCAAAAGCTGCTAAAGACCTGGCAGAGTCCAAACGGCAC  
GACCCGTAGTATCTGGGTGCTCCTTGGAAAGCATATTCTCGCACCGTGGTGGTGAATCCTACATCGCGTCT  
TTCTGGATCCGGCAGGTTGGTCCCGTGGTTCGGTAAACGATTCTCTGTCTACTCTGTACTACCGGAAATCA  
ACAATTCTGGTATCGGCGCTCGTACCGGTGGCCGTTGACTGGCCGGGTTCCGCTCTGATTAATGAAACGGA  
AGCGGGCACTTCACCGTTGCTAACTTCAACCAGGGTACGATATGGCTGCCAGCAACCGGTGTCCTCGTCTCGT  
GACTGGTCTGGTACAGCTCGAGTAA

>LuPME67\_codon-optimized

ATGGAGAAAAGTGTGAACCAGACTCCAACCTTCTGGCTGTTGCTTCTCCTCTTTTCATCGCGCTGATTTCTCT  
GCTGTGCCTGATCCGCGCCCGCTGGACCTGCCTTCTAGCTACCTGAGCACCACCTTCCAGATGGCTTCCCTG  
CAATTCGGTTCTTTCTGCCGACTTCTAAAACACCACCACATCACCGTCTGCTGCTCTAAAGTGAATGCGAGG  
AAAACAAGTGGCGTTCCAAGCTGATCTACTGTACCAGGTGAGCCTGGTACTGACTGTAGTACGCGTGGT  
GCGCCAACTTCTCCAGCGTCCAGAAAGCAGTTCGATGCGTCCCGGACTTCGGTAGCAATACCACTCTGATTCT  
GATCGACAGCGGTACCTATCGCGAAAAAGTTGTCGTGAACGCAAAACAAACCAACCTGGTGTGGAAGGCC  
AGAACTATCTGAACACCGCCATCGAATGGAACGACACGGCAACAGCACCGGTGGCACTATCTATCCGCGT  
CCGTTGTGGTGTATGCTCCGAGCTTACCGCATAACAATCAGCTTCAAAAACCTCTGCGCCAGCGCCGGCAC  
CGGGTGAAGTTGGTGCCAAAGCGGTTGCGATTCTGTTGCTGGCGACCAGGCAGCTTTTATCGCTGCGGCTT  
TTATGGCGCTCAGGATACTCTGCACGATGATCATGGTCTGCTACTACTATCGTGGTGTGTTTCATCCAGGGCTCC  
ATCGACTTCATCTTTGGTAAACGCCAAATCTCTGTTTCAATCCTGCACCATCAACTCTATTGCGATCCCGCGA  
CGACCGATTCCGGCGCTACGGGTTCCATTAAGTGCAGGCCCCGCACTTCTGGTGGAGACACCGGCTTTTC  
CTTCGTTAAATCTAAGATTAACGGTACCGGCCAAGTTTGGCTGGTCTGTCATGGGGCACCTACTCCACGGTA  
GTTTTTATCCGTACCTACATGAGCGATGCTGTTAGCCCGGACGGCTGGAATGACTGGCGTGACCCGAGCCGC  
GATCAGACTGTACTGTTTGGTGAATACGATTGTTACGGTCCGGGTGCTAACAACTTCTCTCGGTTTCTTACG  
GCCGTCGCTGACCCAGGAAGAAGCTGCGTCTTTCATGGACATCTCTTACATCAACGGCGAACAGTGGCTGC  
AGGATGGTTCGCACTGTACCGCGCTTCCGGACTACAACGATGGTACGGCTGTGGTAACTACATCTGGGGTA  
ATGGCGACGTGGATCAGGAA

>LuPMEI45\_no-signal-peptide\_codon-optimized

ATGGCCGACACCGACTATATCCAAACTTCTTGGCAGGCGTCCACCCGTTACCCGGATCTGTGTATCTTACCC  
TGTTCCGCGAGGCTTCTAACATTACCACTCCAAACTGCTGGCCTCTGCGGCTGTACGCGCTCTGGCAGC  
GGCAAAAATCCACCTCCAAAAGCATTGAAACCCGCTCTTCTCTTGGAGCTCTCGTCTGCGCGATTGTCGCGAA  
GAGATGAGCGACAGCGTTGACCGTCTGCGTGATTCCCGGAAAGAAATGAAGGGTGAAGTCTGTTCTGCTCCG  
TTCCAGGTAAGCAACGTGCAGACGTGGGCTTCCGCGCAATGACTGTCATGGACACCTGCACGGATGGTCTG  
GTGGAAGGTGAAGTAAAACGCTGGGTCGTTGAACTCTGGTATCGTTAAGGCGGCTTCTGATCAGCAAC  
CGCTGGCTTTTGTAAATAAATACGGCGATGGCCTGGTGAACAG

Additional file A5-1

>PLL4\_Lus10006456\_codon\_optimized

```
ATGGAACTCAACTGCTGTCTCTGCTGCTCCTGCTGACTGTACTGCAGCTGCCGAGCCATATTGCTTCTA
GCCCCGTTCCGGATCCTGACCTGGTTGCGGACGAAGTTAACCGCATTATTACCAACGCTACGAACTCTCGTGG
TAACCTGGGTTTTCTGTCCTGCAAAACTGGTAATCCAATCGATGACTGCTGGCGCTGCGACTCCTCTGGCAC
AAAAACCGTAAACGCCTGGCGGATTGCAGCATCGGCTTCGGTAAAAACGCGATCGGTGGCCGTGACGGTAA
ATTCTACGTTGTGACCGATTCCGGCGATCATCCGTTAATCCGAAACCGGGCACCTGCGCTACGGTGAATT
CGTGATGAGCCGCTGTGGATTGTGTTCGCTCGCGATATGGTGATCAAACGAAAGCAGAGCTGATGATGAAC
AGCTTCAAACCATCGACGCGCCGTGGCGCATCTGTCCACATCGCGGGTGGCCCGTGCATTACCATCCAGTAT
GTTAACAAACATCATCATCCACGGCATCAACGTGCACGACTGTAAAAGCGGTGGCAACCTGAACGTCCGTGAT
AGCCCGAACACTTCGGTTGGCGTACCCGTTCTGATGGCGATGGTATTCCATTTTCGGCGGTCCCACGTCT
GGGTTGACCACGTTTCTGTCCAACCTGCCAGGATGGTCTGATTGACGCAATCCACGGTTCACCGCAATCAC
CATCAGCAACAACATACATGACCCACCATAACAAAGTTATGCTGCTGGGTCATAGCGACGGCTTCGTGCGCGA
CAAAATCATGCAGGTGACTATCGCGTTAATCACTTCGGTGAAGGTCTGGTGCAGCGCATGCCTCGTTGTCGC
CACGGTATTTCACGTTGTTAATAACGACTACCCATTGGCTGATGTACGCTATCGGCGGTAGCGCTTCTC
CGACCATCAACTCTCAGGGCAACCGTTTTCTGGCCCGAACGACAAATTCAACAAAGAAGTAACGAAACGTG
AGGAAGCGGGCGAATCCGAATGGAAGAAATGGAACGGCGTTCTGAAGGCGACCTGCTGCTGAATGGTGCC
TTTTTCACTCGTTCTGGCGCCGGTGCATCTTCTTATGCTAAGGCCTCCTCTCTGGGCGCACGTCCATCTTC
CCTGGTAGGCCCGCTGACTAGCGGTGCGGGTAGCCTGGTATGTAAGAAAGGTTCTCTCTGTA
```