University of Alberta

β-Galactosidases and Fasciclin-like Arabinogalactan Proteins in Flax (*Linum usitatissimum*) Phloem Fibre Development

by

Neil Robert Hobson

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

Department of Biological Sciences

Fall 2013

Edmonton, Alberta

Abstract

The phloem fibres of flax (*Linum usitatissimum*) have specialized, cellulose-rich secondary walls of the gelatinous type (G-type), which give the fibres remarkable strength. G-type walls are also found in tension wood of many trees. β-galactosidases (BGAL) and fasciclin-like arabinogalactan (FLA) genes are expressed during development of G-type walls in flax, but the functions of these genes and their possible interactions are not well understood. The recent assembly of the flax genome, to which I contributed, afforded an opportunity to characterize the BGAL and FLA gene families of flax. Each of these families comprises 43 predicted genes. Comparison of the BGAL family structure between species revealed the expansion and contraction of distinct sub-families, perhaps correlated with specialization of cell wall composition of flax tissues during its evolution and domestication. Phylogenetic analyses of the FLA gene family revealed a sub-family and an amino acid motif unique to the Rosids that is represented by many FLA genes known to be expressed in G-type walls. Transcript expression profiling of the BGAL and FLA families in 12 different flax tissues identified multiple genes from each family that were highly expressed in developing fibres. Transgenic analyses of selected, fibre-enriched genes (*LuBGAL1* and *LuFLA1*) were conducted, first by generating promoter-reporter gene fusions. Within stems, the upstream regions of each of these genes directed reporter gene expression preferentially to developing phloem fibres. Because the RNAi loss-of-funtion phenotype of *LuBGAL1* had been previously shown to reduce fibre strength and crystalinity, I extended this functional analysis by producing and characterizing lines overexpressing *LuBGAL1* transcripts. I was unable to obtain evidence that *LuBGAL1* overexpression affected the composition or mechanical properties of phloem fibres, suggesting that *LuBGAL1* may be necessary but not sufficient for G-type wall development. I also characterized transgenic lines bearing an RNAi construct targeted towards *LuFLA35* (which is closely related to *LuFLA1* and whose transcripts are likewise highly enriched in fibres). *LuFLA35-RNAi* lines were slightly diminished in tensile strength in flax stems, providing the first evidence that FLAs are functionally required for G-type fibre development in flax.

Acknowledgement

Thanks to my family, for all their support. A special thanks goes to Mike, for putting up with my shenanigans, and letting me participate in more studies and publications than I probably warranted. To dear departed Mary, who's moved on to a better place. Or, at least a new job. She was an invaluable help while working as our lab technician, and taught me everything I know about flax tissue culture. Hopefully, cattle research is treating her well. To Melissa, whose research projects I acquired upon her graduation, and with whom we've continued to collaborate. To David, for sharing the burden of experimenting with Fluidigm. And by sharing, I mean doing most of the work while I ran off to China for a summer. Thanks to all my labmates, past and present. I'd also like to thank Dr. Kaushik Ghose and Dr. Bourlaye Fofana from the Crops and Livestock Research Centre in Charlottetown, PE, Canada, for generously aiding in the heterologous expression of LuBGAL1, and validating our results. Lastly, to the double negative, the sentence fragment, and the comma splice. To contractions, contradictions, and colloquialisms. I placed you where I could, and where I couldn't, you were missed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGP, arabinogalactan protein

BA, benzyladenine

BCE, Before Common Era

Ca^{2+}, calcium

CaMV, cauliflower mosaic virus

CIPRES, Cyberinfrastructure for Phylogenetic Research

cDNA, complementary deoxyribonucleic acid

CDS, coding DNA sequence

CTAB, hexadecyl-trimethyl-ammonium bromide

EC, enzyme class

EDTA, ethylenediaminetetraactic acid

EMBOSS, The European Molecular Biology Open Software Suite

EST, expressed sequence tag

FAA, formalin-acetic-alcohol

FTIR, fourier transform infrared spectroscopy

Galp, galactose

GalpA, galactosyl uronic acid

GC-MS, gas chromatography-mass spectrometry

GEO, Gene Expression Omnibus

GH, glycosyl hydrolase

G-layer, gelatinous layer

Gn-layer, galactan enriched layer

GPI, glycophosphatidylinositol

GRP, glycine-rich protein

GUS, beta-glucuronidase

HGA, homogalacturonan

HMM, Hidden Markov Model

HRGP, hydroxyproline-rich glycoprotein

IAA, indole-3-acetic acid

IBA, indole-3-butyric acid

IUBMB, International Union of Biochemistry and Molecular Biology

$K_3Fe(CN)_6$, potassium ferricyanide

$K_4Fe(CN)_6$, potassium ferrocyanide

kDa, kilo dalton

LS, Linsmaier and Skoog

$MgCl_2$, magnesium chloride

mRNA, messenger ribonucleic acid

4-MU, 4-methylumbelliferone

MUG, 4-methylumbelliferyl-beta-D-glucuronide

N_2, nitrogen

NAA, naphthyl-acetic acid

$NaCO_3$, sodium carbonate

$NaHPO_4$, sodium phosphate

NCBI, National Center for Biotechnology Information

nsLTP, non-specific lipid transfer protein

o/n, overnight

ONPG, ortho-nitrophenyl-ß-galactoside

PCR, polymerase chain reaction

PMSF, phenylmethanesulfonylfluoride

PPT, DL-phosphinothricin

PRP, proline-rich protein

qPCR, quantitative polymerase chain reacton

qRT-PCR, quantitative real-time polymerase chain reaction

RG-I, rhamnogalacturonan I

RG-II, rhamnogalacturonan II

Rhap, rhamnose

RNAi, RNA interference

S.E, standard error

TAIR, The Arabidopsis Information Resource

T-DNA, transfer deoxyribonucleic acid

UDP, uridine diphosphate

X-Gal, 5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside

X-Gluc, 5-bromo-4-chloro-3-indolyl-ß-D-glucuronide

XLLG, xyloglucan nonosaccharide

Xylp, xylose

Chapter 1 – Literature Review

Flax

Flax is a eudicot, of the order Malpighiales, and family Linaceae. It is an annual plant, with tall slender stems growing up to 1.2 m in height. The leaves are green and lanceolate, 20-40 mm long and 3 mm broad, and are arranged in an alternate pattern along the stem. The flowers are complete and perfect, borne on panicles and exhibiting sympodial growth. The flowers are actinomorphic and typically 15-25 mm in diameter, with five sepals, five petals (displayed in an open saucer shape), a gynoecium with five carpels, and five stamens. Flower petals are typically blue, however varieties exist with white or pink petals. Stamens are arranged alternate to petals, and their bases widen to form a fused ring at the base of the gynoecium into which the petals are inserted (Schewe et al., 2011). The carpels are divided by false septum and produce up to two ovules. The fruit is a round dry capsule, between 5-9 mm in diameter, and containing up to ten glossy, brown, ovate seeds, which are each 4-7 mm long.

Flax can be cultivated for one of two products. The seed (cultivated from linseed-types), is a nutritious food rich in unsaturated fatty acids. These same oils, α-linolenic acids (C18:3) in particular, are useful in the production of paints, varnishes, and linoleum, and have been reported as beneficial for cardiovascular health and the treatments of certain cancers and auto-immune diseases (Singh et al., 2011). As a food source, linseed has been cultivated as far back as 7,000 BCE, where seeds have been identified amid neolithic occupational debris in Tell Ramad, near Damascus, Syria (van Zeist and Bakker-Heeres, 1975). Alternatively, flax can be grown for the phloem (bast) fibres of the stem. These are long cells rich in crystalline cellulose and having high tensile strength (Mohanty et al., 2000), and are used in the production of textiles such as linen, as well as high quality papers and composite materials. Records of flax cultivation for fibres go back even further than those of seeds, where dyed flax fibres have been identified in 30,000-year-old Paleolithic caves in Georgia (Kvavadze et al., 2009). Unfortunately, no true dual-purpose variety exists, and instead, flax cultivars are developed for one of the two products. In general, oilseed

cultivars tend to be shorter and more branched, producing many seeds, but have lower fibre yield with shorter, low-quality fibres in the stem. Flax fibre cultivars, on the other hand, tend to be taller with fewer branches, produce fewer, smaller seeds, but longer, higher-quality fibres.

As of 2010, linseed production has been led by Canada, China, the United States of America, the Russian Federation, and India. Canada was the greatest producer, yielding 423,000 metric tonnes of seed, or 21.9% of worldwide production (www.faostat3.fao.org). With regards to flax fibres and tow (i.e. coarse broken fibres), worldwide production in 2010 was concentrated in France, China, Belarus, the Russian Federation, and the United Kingdom, with 622,200 metric tonnes harvested in total. Despite its primacy in linseed cultivation, Canada is not known for flax fibre production. Climate plays a large part in this, as retting, a step in fibre extraction whereby stems are exposed to microorganisms to separate the fibres from the surrounding tissues, is untenable in the cold, dry, Canadian prairies. In fact, due to a lack of processing facilities and other factors, the presence of bast fibres in oilseed flax is considered undesirable. The problem of flax straw management is repeatedly cited in surveys of farmers as the greatest hindrance to flax seed production. Therefore, it would be beneficial to understand bast fibre development and develop flax genotypes with fibre properties customized to different industries and applications.

Beyond the obvious economic incentives, flax is an ideal model organism for a number of scientific reasons. With its large, and easily isolated, phloem fibres, flax has become a useful model for dissecting and analysing phloem and fibre development (Gorshkova et al., 2012). These same phloem fibres are also an excellent resource for examinations of cell wall composition and development (Gorshkova et al., 2012). Flax is also of interest due to the peculiarities of its genome. Studies have found that, under certain environmental conditions (variations in soil pH and fertilizer content for instance), varieties of flax will modify its genome, resulting in differences in nuclear DNA content and gene expression between different plant generations (Durrant, 1971; Cullis, 2005). The causative agent of the change has been identified as a 5.7 kb DNA fragment which has been found to insert itself in a targeted manner throughout the genome, resulting in phenotypic changes to plant height, weight, branching, as well as producing hairs on the seed capsule septa (Chen et al., 2005). While the phenotypic changes will be stably inherited in

proceeding generations, the change in DNA content observed in the first generation will return to that observed in the original parent lines (Cullis, 2005).

L. usitatissimum has a number of traits that make it amenable to genetic studies. Flax is treated as a diploid, with n = 15 chromosomes, and a genome size of approximately C = 373 Mb (Wang et al., 2012). As a self-pollinator with a relatively short life cycle, at approximately 100 days per generation (Morvan et al., 2003), it is relatively easy to obtain and propagate mutant and transgenic lines. Flax researchers also have a number of resources at their disposal. At the time of writing, the NCBI-EST database listed 286,852 ESTs, 74.8% of which came from developing seeds (Venglat et al., 2011). The first draft of the genome has been published, a *de novo* assembly of short read sequences, and current analyses have identified 43,384 predicted protein coding loci, to which 93% of published flax ESTs were found to align (Wang et al., 2012). In addition to the sequenced genome, a genetic map using simple sequence repeat markers is available, with high-density SNP maps in development (Cloutier et al., 2012). Flax also has a number of microarray platforms available. A 9600 cDNA microarray was developed from the fibre rich cortical peels of flax stems, and was used to examine the development of fibres in the developing stem as well as in young hypocotyls (Roach and Deyholos, 2007; Roach and Deyholos, 2008). Data from this microarray led to the identification of a number of fibre enriched transcripts, some of which were later characterized as important mediators of phloem fibre development (Roach et al., 2011). A second flax microarray platform was recently described, based on 59,626 ESTs derived from variety of tissues (Fénart et al., 2010), which has been used to begin elucidating the hypolignification observed in flax phloem fibres (Huis et al., 2012). Lastly, flax is amenable to agrobacterium mediated transformation (McHughen, 1989), making it a suitable model for the characterization of individual genes.

Two oilseed cultivars are used in the proceeding chapters: CDC Bethune and Norlin, developed in 1998 and 1982 respectively. The two cultivars differ slightly in growth habit and yield, with Bethune growing slightly taller, with fewer branches and a greater yield of seed. Bethune seeds also have increased oil content, as well as reduced iodine and alanine, in comparison to Norlin. CDC Bethune is also more resistant to powdery mildew, a fungal disease

detrimental to flax cultivation (www.flaxcouncil.ca). Initially, Norlin was used due to earlier successes with plant transformations (Millam et al., 2005), however we later switch to CDC Bethune as it was more prevalent in Canadian agriculture.

Flax Phloem Fibres

Fibres are elongated sclerenchyma cells with thick secondary cell walls and tapered ends, which play a major role in the mechanical properties of plants. While others cell and tissue types, such as tracheids, vascular bundles, and seed trichomes, are occasionally referenced as fibres, they are not true fibres as they lack all of the defining characteristics of fibres listed above (Fahn, 1990).

Fibres are found within many vascular plant species, and can be found in roots, fruits, stems, and leaves. Fibre cells can be uninucleate or mulinucleate, living or dead at maturity, and may or may not have septa dividing the lumen into multiple chambers (Esau, 1965; Fahn, 1990; Courtois-Moreau et al., 2009; Snegireva et al., 2010). While some fibres may be short, such as the 200 μm long fibres in the shoots and root cortex of *Arabidopsis thaliana* (Lev-Yadun, 1997), longer fibres from other species can reach lengths up to 55 cm, such as in *Boehmeria nivea* (Aldava, 2013), and their secondary cell walls can be as thick as 15 μm (Crônier et al., 2005).

In the primary plant body, fibres may be derived from the procambium, examples of which include the phloem fibres of flax and the primary phloem fibres of hemp (Esau, 1965), or they may derive from the ground meristem, such as with the outer interfascicular fibres of Arabidopsis (Little et al., 2002; Altamura et al., 2001). In the secondary plant body, fibres derive from the cambium, developing as bundles in the secondary phloem or in various groupings in the secondary xylem (Esau, 1965; Fahn, 1990).

There exists a limited understanding of the mechanisms underlying the specification of fibre cell fate, be it in phloem or xylem. Currently, the focus has been on understanding the development of the cambium, and the mechanisms underlying the specification of phloem and xylem as a whole (Gorshkova et al., 2012), with the majority of research having elucidated mechanisms of xylem differentiation. In brief, vascular development is regulated strongly by

auxin gradients. As auxin is conducted through the stem from the apex, a feedback loop, involving the transcription factor IAA RESPONSE FACTOR5/MONOPTEROS (MP) and the IAA transporter PIN1, creates an auxin canal through the stem, where cambial tissues accumulating greater levels of IAA develop into xylem (Mellerowicz et al., 2001; Ohashi-Ito and Fukuda, 2010). The initial signal promoting phloem development has not been clearly described; however, a few components of phloem specification have been identified. TRACHEARY ELEMENT DIFFERENTIATION INHIBITORY FACTOR (TDIF), a mobile peptide hormone encoded by *CLE41* and *CLE44*, is produced in developing phloem and, upon transport to the procambium, induces the transcription of WUSCHEL-RELATED HOMEOBOX 4 (WOX4), a master transcription factor that promotes the maintenance of procambial/cambial stem cells (Ji et al., 2010). WOX4 also activates the expression of transcription factors that promote xylem mother cell proliferation (Hirakawa et al., 2010), as well as *ALTERED PHLOEM DEVELOPMENT* (*APL*), a MYB transcription factor, which promotes phloem fate and suppresses xylem fate (Ji et al., 2010). Additional MYB transcription factors from the KANADI family, including KAN1, KAN2, and KAN3, also act to promote phloem fate (Jung and Park, 2007). While many factors downstream of these have been identified (Gorshkova et al., 2012), additional upstream regulatory mechanisms remain to be found.

Upon determination of cell fate, flax phloem fibres undergo two modes of growth. The first is symplastic growth, where the fibres elongate and expand at the same rate as the surrounding cells (Esau, 1965). Throughout this phase, phloem fibres have blunt ends, with diameters of 4-7 μm, and lengths of 70-100 μm (Snegireva et al., 2010). During symplastic growth, which occurs within the top 1.5 mm of the stem, flax phloem fibres can be distinguished as the widest cells in the protophloem (Esau, 1943).

The next stage of fibre elongation is intrusive growth, which can overlap with the symplastic growth of neighboring cells, where fibres continue to elongate despite the cessation of elongation in the surrounding cells (Esau, 1965; Fahn, 1990). This phase of development is limited to several days, and does not overlap with secondary cell wall deposition (Gorshkova et al., 2003; Ageeva et al., 2005; Snegireva et al., 2010). At the beginning of intrusive growth,

starting at 300-500 μm from the shoot apex, fibre ends develop bulges, termed “knees”, caused as the fibre begins to intrude upon the middle lamella of neighboring cells at both apical and basal ends (Ageeva et al., 2005). As intrusive growth continues, the knees eventually smooth away, leaving the cells with long tapered ends as they elongate at a rate of 0.8 mm/hr (Snegireva et al., 2010). During this time, the fibres expand both vertically and horizontally, expanding from 4-10 μm in diameter up to 20-40 μm, as well as from 70-100 μm in length up to 5 cm (Ageeva et al., 2005). The horizontal expansion, continuing after the cessation of growth of sieve elements and companion cells, eventually crushes the other cell types of the protophloem to create a homogenous bundle of phloem fibres (Esau, 1943). Despite initial thoughts that intrusive growth occured by means of tip elongation (Esau, 1965; Fahn, 1990), several studies have shown that flax phloem fibres elongate diffusely, with cell wall extensions occuring along the entire length of the cell (Gorshkova et al., 2003; Ageeva et al., 2005).

Following cell elongation, flax phloem fibres enter a stage of secondary cell wall deposition, where cell walls can thicken enough to eliminate all visual traces of the cell lumen, making them clearly visible in a stem cross-section. The interface between these two stages of development can be clearly observed at the “snap-point”, a point on the stem of increased mechanical strength, conferred by thickened cell walls, that can be detected by the sharp resistance to bending and tearing when pulled (Gorshkova et al., 2003). The cell walls of flax phloem fibres are often described as gelatinous, a cell wall type characterized by hypolignification and a large proportion of crystalline α-cellulose (80-90%) aligned almost parallel to the longitudinal axis of the cell, absent any xylan. The term “gelatinous” comes from observations of these cell walls swelling with water; although they would shrink irreversibly when dry (Esau, 1965; Fahn, 1990). During cell wall deposition, flax phloem fibres display a bipartite cell wall, during the deposition of a Gn-layer and its concurrent maturation into a G-layer. Current models propose that the Gn-layer contains cellulose microfibrils separated by large galactan-rich polysaccharides. The Gn-layer is in turn gradually modified into a crystalline G-layer by the degradation of the galactan-rich polysaccharides and the compression of cellulose (Gorshkova et al., 2006), in large part due to the actions of β-galactosidases (Roach et al., 2011).

Cell Walls

Cell walls are complex structures located outside the cell membrane. Tough and flexible, although occasionally rigid, the cell wall provides structural support to cells, allowing cells to maintain a rigidity imparted by high osmotic pressures. Walls also protect cells from their environment, acting either as filters or barriers to pathogens (Lodish et al., 2000; Underwood, 2012). Cell walls also control the direction of plant growth and modulate the growth rate of the plant cell, according to the relative stiffness of a section of the wall at a given time.

Plant cell wall deposition begins during cell division, when the protoplasts of two sister cells begin secreting a pectin-rich middle lamella into what becomes the cell plate. The middle lamella acts to glue neighboring cells together, thus maintaining the structural integrity of the tissue. The next layer, internal to the middle lamella, is the primary cell wall, which is primarily composed of cellulose microfibrils embedded in a gel-like mix of non-cellulosic polysaccharides and glycoproteins (Popper, 2008). The primary cell wall is continuously synthesized during cell growth to accommodate cell expansion and elongation. Depending upon the cell type, a secondary cell wall may also be deposited, internal to the primary cell wall, and subsequent to the cessation of cell growth. The secondary cell wall is often thicker than the primary, and generally contains lignin, which confers additional rigidity to the cell. In flax phloem fibres, the cell walls typically contain between 60-70% cellulose, 5-15% non-cellulosic polysaccharides, low levels of protein and trace amounts of lignin (Morvan et al., 2003; Day et al., 2005; Deyholos, 2006).

Cellulose

Cellulose is a polymer of (1-4)-β-D-glucopyranosyl residues, chains of which will hydrogen bond to one another and form cellulose microfibrils of 3-5 nm in diameter, several thousand glucose residues in length (Albersheim et al., 2011). On average, a primary cell wall will contain between 20-30% cellulose, whereas secondary cell walls can contain up to 95% cellulose.

Cellulose is synthesized by cellulose synthase (CesA) enzyme complexes, composed of five to six different CesA protein sub-units (three isoforms), which are in turn arranged into

hexameric rosettes embedded in the plasma membrane of the cell (Somerville et al., 2004; Albersheim et al., 2011). The rosette complexes utilize sucrose synthase (SuSy) derived UDP-glucose molecules to form the cellulose chains. Theories postulate that the cellulose chains are initially primed with β-sitosterol-glucoside, and later separated from the cellulose synthase complexes by an endo-β-1,4-glucanase by the name of KORRIGAN (KOR), mutations in which lead to reduced cellulose production in plant cells (Nicol et al., 1998; Albersheim et al., 2011).

Pectin and Hemicellulose

Non-cellulosic polysaccharides in the cell wall include pectic polysaccharides, defined as polysaccharides with high proportions (>20%) of D-galactosyluronic acid, and hemicelluloses, which are defined as polysaccharides capable of hydrogen-bonding to cellulose (Albersheim et al., 2011). Three pectic polysaccharides can be found in the primary cell walls, namely homogalacturonan (HGA), rhamnogalacturonan I (RG-I), and rhamnogalacturonan II (RG-II). By cross-linking to one another, mediated by Ca^{2+} and borate ions, these pectic polysaccharides are responsible for cell adhesion and further control cell wall porosity (Baron-Epel et al., 1988; Albersheim et al., 2011). The first of the pectic polysaccharides, HGA, consists entirely of α-1,4-linked D-galactosyluronic acid residues, whose carboxyl groups (up to 70%) are often methyl esterified. HGA is commonly referred to simply as pectin (Albersheim et al., 2011), and will stiffen as its α-1,4-Gal*p*A backbone cross-link with other strings of α-1,4-Gal*p*A, with the help of Ca^{2+} ions. These Ca^{2+} cross-links are prevented by high levels of methyl-esterification. RG-I, on the other hand, contains a backbone of repeating →4)-α-D-Gal*p*A-(1→2)-α-L-Rha*p*-(1→ disaccharides, where roughly 50% of the rhamnosyl residues have C4 substitutions of L-arabinosyl, D-galactosyl, L-fucosyl, and D-glucosyluronic acid containing side chains. RG-II, similar to HGA, has a backbone of α-1,4-linked D-galactosyluronic acid residues, however it bears C2 and C3 sidechains composed of widespread aldoses and ketoses.

One of the principal hemicelluloses of the primary cell wall is xyloglucan, which can comprise roughly 30% of the primary cell walls of dicotyledons, up to 20% of gymnosperm primary cell walls, and less than 2% of the primary walls of the grass family (Poaceae; Albersheim

et al., 2011). In the grass family of monocots, hemicelluloses of the primary cell wall include arabinoxylans, and can account for 30-40% of their cell walls. Dicotyledonous plants (eudicots and basal angiosperms) and gymnosperms may also contain arabinoxylans, however these account for less that 5% of their cell walls (Scheller and Ulvskov, 2010; Albersheim et al., 2011). Xyloglucans are a highly branched polymer with a β-1,4-linked glucan backbone, where roughly 75% of the glucosyl residues are substituted at C6 with side chains beginning with α-D-xylosyl residues. These α-D-xylosyl residues may in turn be linked at C2 to additional β-D-galactosyl, α-L-fucosyl residues, and/or α-L-arabinosyl residues. Further modifications to these side chains include acetylations to arabinoyl and glucosyl residues of the backbone (Albersheim et al., 2011). In contrast to xyloglucans, with their backbone resembling cellulose, arabinoxylans have a backbone of β-1,4-linked xylosyl residues, with sidechains at C2 and/or C3 positions, on anywhere between 10-90% of the xylose residues. Arabinoxylan sidechains have been found to contain arabinosyl, galactosyl, glucosyluronic acid, and 4-O-methyl glucosyluronic acid residues, in addition to phenolic acids such as ferulic acid and coumaric acid (Albersheim et al., 2011). In secondary cell walls, different forms of hemicelluloses predominate, with compositions differing depending upon the species. Dicotyledonous secondary cell walls can contain between 15-30% glucuronoxylans, which comprise β-1,4-linked xylosyl residues conjugated at C2 (1/10 residues) with 4-O-methyl α-D-glucosyluronic acid, and bearing O-acetyl groups on about 70% of the xylosyl residues at C2 or C3. Glucomannan hemicelluloses may also be present (2-5%), and consist of 1,4-linked glucosyl and mannosyl residues, in ratios between 1:1 and 2:1 (Albersheim et al., 2011). Gymnosperm secondary cell walls differ from dicotyledons, and contain primarily galactoglucomannans and glucuronoarabinoxylans. The galactoglucomannans consist of β-1,4-linked glucosyl and mannosyl residues, in a ratio of 1:4 to 1:3, with side chains attached to the mannosyl C6 consisting of α-D-galactosyl residues, as well as O-acetyl groups on roughly 25% of the mannosyl residues at C2 or C3. Glucuronoarabinoxylans consist of β-1,4-linked xylosyl residues with α-L-arabinosyl residues attached to C3 of some xylosyls, and 4-O-methyl α-D-glucosyluronic acid residues attached to C2 of other xylosyls. In each case, the primary purpose of hemicellulose is as a cross-linker, using hydrogen bonds to form bridges between cellulose

microfibrils, integrating them into the cell wall, and maintaining distance between microfibrils to prevent them from aggregating into larger cables. The resulting cellulose-hemicellulose network thus forms a lattice upon which other cell wall components organize.

In flax phloem fibres, secondary cell wall deposition is noted for the presence of a high molecular weight, galactan-rich pectin deposited in visible layers between layers of cellulose microfibrils by the fusion of Golgi vesicles with the plasma membrane (Gorshkova et al., 2006; Salnikov et al., 2008). This pectic polysaccharide is synthesized as a 2000 kDa RG-I polysaccharide, with side chains of variable length and branching composed primarily of β-1,4-linked galactosyl residues. These undergo extensive degradation to a size of 100-400 kDa after incorporation into the cell wall (Gurjanov et al., 2007; Gurjanov et al., 2008; Mikshina et al., 2012).

Lignin

Primary cell walls do not contain lignin; instead, lignin, a complex phenolic polymer, is deposited in the secondary cell wall after cell elongation has ceased, imparting rigidity and compressive strength to cell walls, impermeability to water, and acting as a barrier to pests and pathogens. In contrast to many biopolymers in the cell wall, lignin is also a disordered structure, composed of diverse linkages between various hydrophobic monolignols derived from phenylalanine in the cytoplasm (Albersheim et al., 2011). In trees, the xylan type secondary cell walls of wood will have lignin comprising between 25-40% of their dry weight. However, in flax, the gelatinous type secondary cell walls of the phloem fibres contain only trace amounts of lignin (Day et al., 2005). While low levels of lignin are a characteristic of gelatinous type secondary cell walls, flax phloem fibre lignin differs in that it has an unusually low levels of syringyl (S) units, and instead has lignin consiting of 12-25% hydoxyphenyl (H) units (Day et al., 2005; Huis et al., 2012). The high H-unit content is also a characteristic of the lignin of xylem fibres, although their H unit content runs slighly lower, at 5% of the lignin monolignols (Huis et al., 2012).

Cell Wall Proteins

In addition to polysaccharides and phenolic compounds, plant cell walls also contain a number of proteins, glycoproteins, and proteoglycans, comprising between 1-10% of the dry weight of a cell wall (Albersheim et al., 2011). These can be classified either as insoluble structural proteins, or soluble proteins. Insoluble structural proteins fall into three major categories, which include hydroxyproline-rich glycoproteins (HRGPs), glycine-rich proteins (GRPs), and proline-rich proteins (PRPs). While these structural proteins are widespread in plants, and are thought to play a role in cell wall architecture, we currently have a limited understanding of their molecular functions. Among the soluble proteins, cell walls contain enzymes, many of which are cell-wall metabolizing enzymes involved in modifying cell wall *in muro* to coordinate the growth, deposition, and incorporation of new polymers; defense proteins, which protect against pathogens as well as abiotic stress; transport proteins; lectins; and arabinogalactan proteins, which are considered soluble HRGPs.

β-Galactosidases

β-D-galactoside galactohydrolases (β-galactosidases; EC 3.2.1.23; BGALs) are defined as enzymes that hydrolyze the terminal non-reducing β-D-galactose residues in β-D-galactosides. Other classes of enzymes are known to hydrolyze bonds involving galactose residues, such as 6-phospho-β-galactosidases (EC 3.2.1.85), arabinogalactan endo-β-1,4-galactanases (EC 3.2.1.89), blood-group-substance endo-1,4-β-galactosidases (EC 3.2.1.102), keratan-sulfate endo-1,4-β-galactosidases (EC 3.2.1.103), galactan 1,3-β-galactosidases (EC 3.2.1.145), galactan endo-1,6-β-galactosidases (EC 3.2.1.164), and galactan endo-β-1,3-galactanases (EC 3.2.1.181), however the nature of the substrate and/or mechanism employed is sufficiently different from EC 3.2.1.23 BGALs as to render these enzyme classes distinct (IUBMB).

BGALs have been known, by one name or another, since the late 19th century, when enzyme preparations were initially found to catalyze the hydrolysis of lactose (Fischer, 1894). Interest in this enzyme, both in industrial applications (Husain, 2010) and pure research (Kuby and Lardy, 1952; Jacob and Monod, 1961), has continued to this day.

Based upon amino acid sequence similarities, BGALs can be grouped into five of 130 glycosyl hydrolase (GH) families, namely families GH1, GH2, GH3, GH35, and GH42 (http://www.cazy.org; Cantarel et al., 2009). Among the five families, GH1, GH2, GH3, and GH42 BGALs have been observed solely in archaea, bacteria, and fungi, whereas the GH35 family of BGALs has been observed in all kingdoms.

In plants, GH35 BGALs have been observed to primarily degrade cell wall polysaccharides, promoting fruit softening (Smith et al., 2002; Lazan et al., 2004), facilitating the organization of cellulose microfibrils in structural fibre cells (Roach et al., 2011; Mokshina et al., 2012), promoting cell elongation in reproductive organs (Sampedro et al., 2012), and facilitating the secretion of seed mucilage (Dean et al., 2007).

Many non-cellulosic cell wall polysaccharides contain galactose. In particular, BGALs have been found to play a major role in the degradation of cell wall xyloglucans (Edwards et al., 1988; Iglesias et al., 2006; de Alcantara et al., 2006; Sampedro et al., 2012), which have been found to increase the extensiblity of the cell wall when cross-linked to cellulose (Whitney et al., 1999; Albersheim et al., 2011). BGALs have also been found to hydrolyze the sidechains of pectic polysaccharides, specifically the galactan rich RGI pectin in developing flax phloem fibres (Roach et al., 2011; Mikshina et al., 2012; Mokshina et al., 2012).

It should be noted that, just as BGALs of EC 3.2.1.23 may have dissimilar sequences and thus are classified in multiple GH families, proteins in a given GH family can have different, or multiple, enzymatic roles, as well as different substrates. Several plant GH35 BGALs have been described as exogalactanases hydrolyzing β-(1,3)- and β-(1,4)-linked galacto-oligosaccharides (Kotake et al., 2005; Ahn et al., 2007; Dean et al., 2007; Gantulga et al., 2008; Tanthanuch et al., 2008), as well as β-(1,6)-linked and β-(1,2)-linked galacto-oligosaccharides (Kotake et al., 2005; Tanthanuch et al., 2008; Gantulga et al., 2009; Sampedro et al., 2012), either solely or in different combinations. One GH35 BGAL, OsBGAL13, has also been described to hydrolyze α-L-arabinosides (Tanthanuch et al., 2008). Interestingly, this variation seems to occur despite all characterized GH35 BGALs containing the same consensus active site of G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY] (Henrissat, 1998).

Fasciclin-like Arabinogalactan Proteins

Arabinogalactan proteins (AGPs) are a family of high molecular weight non-enzymatic proteoglycans (1-10% protein and 90-99% carbohydrate) associated with the plasma membrane, the cell wall, and intercellular spaces in plants (Albersheim et al., 2011; Nguema-Ona et al., 2012). They are thought to play roles in signalling and the regulation of cell differentiation, tissue development, and embryogenesis (Siefert and Roberts, 2007). Wound response is also known to induce the expression of copious amounts of AGPs, which has been harnessed by the food industry in the production of gum arabic, which is a mixture of AGPs isolated from wounded *Acacia senegal* (Albersheim et al., 2011).

In terms of structure, the protein backbone of AGPs is rich in hydroxyproline repeats, making it a member of the HRGPs, soluble due to its extensive glycosylation. Classical AGPs have an 85-170 aa hydroxyproline rich domain, preceded by an N-terminal signal peptide designating the protein for export outside the plasma membrane, and followed by a GPI-anchor signal sequence so that the protein will be bound to the membrane upon transfer to the apoplast (Albersheim et al., 2011). Some classical AGPs also contain short lysine rich domains (Gaspar et al., 2001). Non-classical variants of AGPs include versions with an asparagine rich AGP domain, versions containing one to two fasciclin domains interspersed between one to two AGP domains, as well as versions with AGP domains fused to non-specific lipid transfer protein (nsLTP) domains (Albersheim et al., 2011; Kobayashi et al., 2011). AGPs are noted for extensive O-glycosylation of hydroxyproline residues in the protein backbone, which takes place in the Golgi apparatus (Nguema-Ona et al., 2012), where two types of glycan side-chain are added. The first commonly added sidechain is a short oligo-arabinoside of three to four residues (Kieliszewski and Shpak, 2001). The second commonly added sidechain is a larger polysaccharide, consisting of a β-(1,3)-linked galactan chain with 1,6-linked sidechains of primarily galactose and arabinose, with additions of rhamnose, fucose, and/or glucuronic acid (Tan et al., 2010; Tryfona et al., 2010; Tryfona et al., 2012).

While AGPs have been implicated in a number of developmental processes, notably as a morphogen inducing tracheary element formation in Zinnia cell cultures (Motose et al., 2004), their mode of action has remained shrouded in mystery (Siefert and Roberts, 2007; Albersheim et al., 2011). Only recently has a study indicated that AGPs may act as Ca^{2+} sequestering agents, where their observed roles in signalling and developmental regulation may be due to their role in regulating calcium oscillations in a global calcium signalling system in plants (Lamport and Várnai, 2013).

Fasciclin-like arabinogalactan proteins (FLAs) are a large group of AGPs, comprising ~45% of all AGPs in the Arabidopsis genome (Schultz et al., 2002), and have been divided amidst four sub-classes. In general, group A FLAs contain two AGP domains flanking a single fasciclin domain, whereas the opposite occurs with group B FLAs, which contain a single AGP domain flanked by two fasciclin domains. Group C FLAs generally have two AGP and fasciclin domains, each, while group D contains the remaining FLAs which have very little similarity to each other or the FLAs of groups A, B, and C (Johnson et al., 2003).

The fasciclin domain itself is a ~140 aa region found in extracellular proteins of plants and animals (Zinn et al., 1988; Huber and Sumper, 1994). In animal models, it has been shown to mediate cell adhesion (Elkins et al., 1990; Kim et al., 2002); however, unlike animal systems, where cell adhesion involves the contact of plasma membranes, in plants, cell-cell contact occurs at the level of the cell wall, with the exception of plasmodesmata. Nevertheless, the inhibition of proteins with fasciclin domains has been shown to reduce cell adhesion in *Volvox carteri*, indicating that the function in plants remains the same (Huber and Sumper, 1994).

Current Study

Recent microarray studies have implicated a flax *BGAL* and *FLA* gene in the development of flax phloem fibres (Roach and Deyholos, 2007). Characterization of the flax BGAL, *LuBGAL1*, implicated the gene in the degradation of the galactan rich polysaccharide associated with the Gn-layer of developing flax phloem fibres (Roach et al., 2011). The characterization of the flax FLA gene, *LuFLA35*, started with the development of *LuFLA35*-RNAi

constructs in flax, however, the analysis remained unfinished (Roach, 2009). This study explored the composition, evolutionary relationships, and expression patterns of the entire BGAL and FLA families of flax, then continued to examine the roles of *LuBGAL1, LuFLA1,* and *LuFLA35* in flax development by first examining the expression patterns of *LuBGAL1* and *LuFLA1* in flax using transgenic reporter fusions. We further examined whether the overexpression of *LuBGAL1* could be used to manipulate the quality of flax phloem fibres, before completing the analysis of the *LuFLA35*-RNAi constructs.

References

Ageeva MV, Petrovska B, Kieft H, Sal'nikov VV, Snegireva AV, van Dam JEG, van Veenendaal WLH, Emons AMC, Gorshkova TA, and van Lammeren AAM (2005) Intrusive growth of flax phloem fibers is of intercalary type. Planta **222**: 565-574.

Ahn YO, Zheng M, Bevan DR, Esen A, Shiu S-H, Benson J, Peng HP, Miller JT, Cheng CL, Poulton JE, and Shih MC (2007) Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35. Phytochemistry **68**: 1510–20. doi:10.1016/j.phytochem.2007.03.021

Albersheim P, Darvill A, Roberts K, Sederoff R, and Staehelin A (2011) Plant Cell Walls. Garland Science, Taylor & Francis Group, LLC, New York.

Aldava V (2013) The Structure and Development of the Cell Wall in Plant I. Bast Fibers of *Boehmeria* and *Linum*. American Journal of Botany **14:** 16–24. doi:10.2307/2435518

Altamura MM, Possenti M, Matteucci A, Baima S, Ruberti I, and Morelli G (2001) Development of the vascular system in the inflorescence stem of *Arabidopsis*. New Phytologist **151:** 381–389. doi:10.1046/j.0028-646x.2001.00188.x

Baron-Epel O, Gharyal PK, and Schindler M (1988) Pectins as mediators of wall porosity in soybean cells. Planta **175**: 389-395.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, and Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research **37**: D233-238.

Chen Y, Schneeberger RG, and Cullis CA (2005) A site-specific insertion sequence in flax genotrophs induced by environment. New Phytologist **167**: 171-180. doi: 10.1111/j.1469-8137.2005.01398.x

Cloutier S, Miranda E, Ward K, Radovanovic N, Reimer E, Walichnowski A, Datla R, Rowland G, Duguid S, and Ragupathy R (2012) Simple sequence repeat marker development from bacterial artificial chromosome end sequences and expressed sequence tags of flax (*Linum usitatissimum* L.). Theoretical and Applied Genetics **125**: 685–694. doi:10.1007/s00122-012-1860-4

Courtois-Moreau CL, Pesquet E, Sjödin A, Muñiz L, Bollhöner B, Kaneda M, Samuels L, Jansson S, and Tuominen H (2009) A unique program for cell death in xylem fibers of *Populus* stem. The Plant Journal **58**: 260–274. doi:10.1111/j.1365-313X.2008.03777.x

Crônier D, Monties B, and Chabbert B (2005) Structure and chemical composition of bast fibers isolated from developing hemp stem. Journal of Agricultural and Food chemistry **53**: 8279–8289. doi:10.1021/jf051253k

Cullis CA (2005) Mechanisms and control of rapid genomic changes in flax. Annals of Botany **95**: 201–206. doi:10.1093/aob/mci013

Day A, Ruel K, Neutelings G, Crônier D, David H, Hawkins S, and Chabbert B (2005) Lignification in the flax stem: evidence for an unusual lignin in bast fibers. Planta **222:** 234–245. doi:10.1007/s00425-005-1537-1

Dean GH, Zheng H, Tewari J, Huang J, Young DS, Hwang YT, Western TL, Carpita NC, McCann MC, Mansfield SD, and Haughn GW (2007) The *Arabidopsis MUM2* gene encodes a β-galactosidase required for the production of seed coat mucilage with correct hydration properties. The Plant Cell **19:** 4007–4021. doi:10.1105/tpc.107.050609

de Alcântara PHN, Martim L, Silva CO, Dietrich SMC, and Buckeridge MS (2006) Purification of a β-galactosidase from cotyledons of *Hymenaea courbaril* L. (Leguminosae). Enzyme properties and biological function. Plant Physiology and Biochemistry **44**: 619-627

Deyholos MK (2006) Bast fiber of flax (*Linum usitatissimum* L.): Biological foundations of its ancient and modern uses. Israel Journal of Plant Sciences **54**: 273-280

Durrant A (1971) Induction and growth of flax genotrophs. Heredity **27**: 277-298. doi: 10.1038/hdy.1971.90

Edwards M, Bowman YJL, Dea ICM, and Reid JSG (1988) A β-D-Galactosidase from Nasturtium (*Tropaeolum majus* L.) Cotyledons: Purification, Properties, and Demonstration that Xyloglucan is the Natural Substrate. The Journal of Biological Chemistry **263**: 4333-4337

Elkins T, Hortsch M, Bieber AJ, Snow PM, and Goodman CS (1990) *Drosophila* Fasciclin I Is a Novel Homophilic Adhesion Molecule That along with Fasciclin III Can Mediate Cell Sorting. The Journal of Cell Biology **110**: 1825-1832

Esau K (1943) Vascular Differentiation in the Vegetative Shoot of Linum. III. The Origin of the Bast Fibers. American Journal of Botany **30**: 579-586

Esau K (1965) Plant Anatomy. 2nd Ed. John Wiley & Sons Inc., New York.

Fahn A (1990) Plant Anatomy. 4th Ed. Pergamon Press, Oxford, UK.

Fischer E (1984) Einfluss der Configuration auf die Wirkung der Enzyme. II. Berichte der Deutschen Chemischen Gesellschaft **27**: 3479-3483.

Gantulga D, Turan Y, Bevan DR, and Esen A (2008) The *Arabidopsis* At1g45130 and At3g52840 genes encode β-galactosidases with activity toward cell wall polysaccharides. Phytochemistry **69:** 1661–1670. doi:10.1016/j.phytochem.2008.01.023

Gantulga D, Ahn YO, Zhou C, Battogtokh D, Bevan DR, Winkel BSJ, and Esen A (2009) Comparative characterization of the *Arabidopsis* subfamily a1 β-galactosidases. Phytochemistry **70:** 1999–2009. doi:10.1016/j.phytochem.2009.08.008

Gaspar Y, Johnson KL, McKenna JA, Bacic A, and Schultz CJ (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding function. Plant Molecular Biology **47**: 161-176

Gorshkova TA, Salnikov VV, Chemikosova SB, Ageeva MV, Pavlencheva NV, and van Dam JEG (2003) The snap point: a transient point in *Linum usitatissimum* bast fiber development. Industrial Crops and Products **18**: 213-221.

Gorshkova T, and Morvan C (2006) Secondary cell-wall assembly in flax phloem fibers: role of galactans. Planta **223**: 149-158.

Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, and Pilate G (2012) Plant Fiber Formation: State of the Art, Recent and Expected Progress, and Open Questions. Critical Reviews in Plant Sciences **31**: 201-228. doi: 10.1080/07352689.2011.616096

Gurjanov OP, Gorshkova TA, Kabel, MA, Schols HA, and van Dam JEG (2007) MALDI-TOF MS evidence for the linking of flax bast fibre galactan to rhamnogalacturonan backbone. Carbohydrate Polymers **67**: 86-96.

Gurjanov OP, Ibragimova NN, Gnezdilov OI, and Gorshkova TA (2008) Polysaccharides, tightly bound to cellulose in the cell wall of flax bast fibre: Isolation and identification. Carbohydrate Research **72**: 719-729

Henrissat B (1998) Glycosidase families. Biochemical Society Transactions **26**: 153-156.

Huis R, Morreel K, Fliniaux O, Lucau-Danila A, Fénart S, Grec S, Neutelings G, Chabbert B, Mesnard F, Boerjan W, and Hawkins S (2012) Natural Hypolignification Is Associated with Extensive Oligolignol Accumulation in Flax Stems. Plant Physiology **158**: 1893-1915. doi: 10.1104/pp.111.192328

Huber O, and Sumper M (1994) Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga *Volvox* with homology to *Drosophila* fasciclin I. The EMBO Journal **13**: 4212-4222.

Husain Q (2010) β-galactosidases and their potential applications: a review. Critical Reviews in Biotechnology **30:** 41–62. doi:10.3109/07388550903330497

Iglesias N, Ablenda JA, Rodiño M, Sampedro J, Revilla G, and Zarra I (2006) Apoplastic Glycosidases Active Against Xyloglucan Oligosaccharides of *Arabidopsis thaliana.* Plant Cell Physiology **47**: 55-63.

Jacob F, and Monod J (1961) Genetic Regulatory Mechanisms in the Synthesis of Proteins. Journal of Molecular Biology **3**: 318-356

Ji J, Strable J, Shimizu R, Koenig D, Sinha N, and Scanlon MJ (2010) WOX4 promotes procambial development. Plant Physiology **152**: 1346-1356

Johnson KL, Jones BJ, Bacic A, and Schultz CJ (2003) The Fasciclin-Like Arabinogalactan Proteins of Arabidopsis. A Multigene Family of Putative Cell Adhesion Molecules. Plant Physiology **133**: 1911-1925. doi: 10.1104/pp.103.031237

Jung JH, and Park CM (2007) Vascular Development in Plants: Specification of Xylem and Phloem Tissues. Journal of Plant Biology **50**: 301-305.

Kieliszewski MJ, and Shpak E (2001) Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. Cellular and Molecular Life Sciences **58**: 1386-1398.

Kim JE, Jeong HW, Nam JO, Lee BH, Choi JY, Park RW, Park JY, and Kim IS (2002) Identification of Motifs in the Fasciclin Domains of the Transforming Growth Factor-β-induced Matrix Protein βig-h3 That Interact with the αvβ5 Integrin. Journal of Biological Chemistry **277**: 46159-46165. doi: 10.1074/jbc.M207055200

Kotake T, Dina S, Konishi T, Kaneko S, Igarashi K, Samejima M, Watanabe Y, Kimura K, and Tsumuraya Y (2005) Molecular Cloning of a β-Galactosidase from Radish That Specifically

Hydrolyzes β-(1→3)- and β-(1→6)-Galactosyl Residues of Arabinogalactan Protein. Plant Physiology **138:** 1563–1576. doi:10.1104/pp.105.062562

Kuby SA, and Lardy HA (1953) Purification and Kinetics of β-D-Galacosidase from *Escherichia coli*, Strain K-12. Journal of The American Chemical Society **75**: 890-896.

Kvavadze E, Bar-Yosef O, Belfer-Cohen A, Boaretto E, Jakeli N, Matskevich Z, and Meshveliani T (2009) 30,000-Year-Old Wild Flax Fibers. Science **325**: 1359. doi: 10.1126/science.1175404

Lamport DTA, and Várnai P (2013) Periplasmic arabinogalactan glycoproteins act as a calcium capacitor that regulates plant growth and development. New Phytologist **197**: 58-64. doi: 10.1111/nph.12005

Lazan H, Ng SY, Goh LY, and Ali ZM (2004) Papaya β-galactosidase/galactanase isoforms in differential cell wall hydrolysis and fruit softening during ripening. Plant Physiology and Biochemistry **42**: 847–853. doi:10.1016/j.plaphy.2004.10.007

Lev-Yadun S (1997) Fibres and Fibre-sclereids in Wild-type *Arabidopsis thaliana*. Annals of Botany **80:** 125–129. doi:10.1006/anbo.1997.0419

Little C, MacDonald J, and Olsson O (2002) Involvement of Indole-3-Acetic Acid in Fascicular and Interfascicular Cambial Growth and Interfascicular Extraxylary Fiber Differentiation in *Arabidopsis thaliana* Inflorescence Stems. International Journal of Plant Sciences **163:** 519–529.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, and Darnell J (2000) Molecular Cell Biology. 4th edition. W. H. Freeman, New York

McHughen A (1989) Agrobacterium mediated transfer of chlorsulfuron resistance to commercial flax cultivars. Plant Cell Reports **8**: 445-449.

Mellerowicz EJ, Baucher M, Sundberg B, and Boerjan W (2001) Unravelling cell wall formation in the woody dicot stem. Plant Molecular Biology **47**: 329-274.

Mikshina PV, Gurjanov OP, Mukhitova FK, Petrova AA, Shashkov AS, and Gorshkova TA (2012) Structural details of pectic galactan from the secondary cell walls of flax (*Linum usitatissimum* L.) phloem fibres. Carbohydrate Polymers **87**: 853-861. doi: 10.1016/j.carbpol.2011.08.068

Millam S, Obert B, and Pret'ová A (2005) Plant cell and biotechnology studies in *Linum usitatissimum* – a review. Plant Cell, Tissue and Organ Culture **82**: 93-103. doi: 10.1007/s11240-004-6961-6

Mohanty AK, Misra M, and Hinrichsen G (2000) Biofibres, biodegradable polymers and biocomposites: an overview. Macromolecular Materials and Engineering **276**: 1-24.

Mokshina NE, Ibragimova NN, Salnikov VV, Amenitskii SI, and Gorshkova TA (2012) Galactosidase of plant fibers with gelatinous cell wall: Identification and localization. Russian Journal of Plant Physiology **59**: 246–254. doi:10.1134/S1021443712020082

Morvan C, Andème-Onzighi C, Girault R, Himmelsbach DS, Driouich A, and Akin DE (2003) Building flax fibres: more than one brick in the walls. Plant Physiology and Biochemistry **41**: 935-944.

Motose H, Sgiyama M, and Fukuda H (2004) A proteoglycan mediates inductive interaction during plant vascular development. Nature **429**: 873-878. doi: 10.1038/nature02613

Kobayashi Y, Motose H, Iwamoto K, and Fukuda H (2011) Expression and Genome-Wide Analysis of the Xylogen-Type Gene Family. Plant and Cell Physiology **52**: 1095-1106

Nguema-Ona E, Coimbra S, Vicré-Gibouin M, Mollet JC, and Driouich A (2012) Arabinogalactan proteins in root and pollen-tube cells: distribution and functional aspects. Annals of Botany **110**: 383-404. doi: 10.1093/aob/mcs143

Nicol F, His I, Jauneau A, Vernhettes S, Canut H, and Höfte H (1998) A plasma membrane-bound putative endo-1,4-β-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. The EMBO Journal **17**:5563-5576

Ohashi-Ito K, and Fukuda H (2010) Transcriptional regulation of vascular cell fates. Current Opinions of Plant Biology **13**: 670-676

Popper ZA (2008) Evolution and diversity of green plant cell walls. Current Opinion in Plant Biology **11**: 268-292. doi: 10.1016/j.pbi.2008.02.012

Roach MJ, and Deyholos MK (2007) Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. Molecular Genetics and Genomics **278:** 149–165. doi:10.1007/s00438-007-0241-1

Roach MJ, and Deyholos MK (2008) Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fibre differentiation. Annals of botany **102:** 317–330. doi:10.1093/aob/mcn110

Roach MJ (2009) Examining the Molecular Basis of Bast Fibre Development in Flax (*Linum usitatissimum* L.). Dissertation, University of Alberta.

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, and Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant physiology **156**: 1351–1363. doi:10.1104/pp.111.172676

Salnikov VV, Ageeva MV, and Gorshkova TA (2008) Homofusion of Golgi secretory vesicles in flax phloem fibres during formation of the gelatinous secondary cell wall. Protoplasma **233**: 269-273. doi: 10.1007/s00709-008-0011-x

Sampedro J, Gianzo C, Iglesias N, Guitián E, Revilla G, and Zarra I (2012) AtBGAL10 is the main xyloglucan β-galactosidase in Arabidopsis, and its absence results in unusual xyloglucan subunits and growth defects. Plant physiology **158**: 1146–57. doi:10.1104/pp.111.192195

Scheller HV, and Ulvskov P (2010) Hemicelluloses. Annual Review of Plant Biology **61**: 263-289. doi: 101146/annurev-arplant-042809-112

Schewe LC, Sawhney VK, and Davis AR (2011) Ontogeny of floral organs in flax (*Linum usitatissimum*; Linaceae). American Journal of Botany **98**: 1077-1085. doi: 10.3732/ajb.1000431

Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, and Bacic A (2003) Using Genomic Resources to Guide Research Directions. The Arabinogalactan Proteins Gene Family as a Test Case. Plant Physiology **129**: 1448-1463.

Seifert GJ, and Roberts K (2007) The Biology of Arabinogalactan Proteins. Annual Review of Plant Biology **58**: 137-161. doi: 10.1146/annurev.arplant.58.032806.103801

Singh KK, Mridula D, Rehal J, and Barnwal P (2011) Flaxseed: A Potential Source of Food, Feed and Fiber. Critical Reviews in Food Science and Nutrition **51**: 210-222. doi: 10.1080/10408390903537241

Smith DL, Abbott JA, and Gross KC (2002) Down-Regulation of Tomato β-Galactosidase 4 Results in Decreased Fruit Softening. Plant Physiology **129**: 1755–1762. doi:10.1104/pp.011025

Snegireva AV, Ageeva MV, Amenitskii SI, Chernova TE, Ebskamp M, and Gorshkova TA (2010) Intrusive growth of sclerenchyma fibers. Russian Journal of Plant Physiology **57:** 342–355. doi:10.1134/S1021443710030052

Somerville C, Bauerr S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredex S, Raab T, Vorwerk S, and Youngs H (2004) Towards a Systems Approach to Understanding Plant Cell Walls. Science **306**: 2206-2211. doi: 10.1126/sciencec.1102765

Tan L, Varnai P, Lamport DTA, Yuan C, Xu J, Qiu F, and Kieliszewski MJ (2010) Plant *O*-Hydroxyproline Arabinogalactans Are Composed of Repeating Trigalactosyl Subunits with Short Bifurcated Side Chains. The Journal of Biological Chemistry **285**: 24575-24583.

Tanthanuch W, Chantarangsee M, Maneesan J, and Ketudat-Cairns J (2008) Genomic and expression analysis of glycosyl hydrolase family 35 genes from rice (*Oryza sativa* L.). BMC Plant Biology **8:** 84. doi:10.1186/1471-2229-8-84

Tryfona T, Liang HC, Kotake T, Kaneko S, Marsh J, Ichinose H, Lovegrove A, Tsumuraya Y, Shewry PR, Stephens E, and Dupree P (2010) Carbohydrate structural analysis of wheat flour arabinogalactan protein. Carbohydrate Research **18**: 2648-2656

Tryfona T, Liang HC, Kotake T, Tsumuraya Y, Stephens E, and Dupree P (2012) Structural Characterization of Arabidopsis Leaf Arabinogalactan Polysaccharides. Plant Physiology **160**: 653-666

Underwood W (2012) The Plant Cell Wall: A Dynamic Barrier Against Pathogen Invasion. Frontiers in Plant Science **3**: 85. doi: 10.3389/fpls.2012.00085

van Zeist W, and Bakker-Heeres JAH (1975) Evidence for Linseed Cultivation Before 6000bc. Journal of Archaeological Science **2**: 215-219. doi: 10.1016/0305-4403(75)90059-X

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, and Datla R (2011) Gene expression analysis of flax seed development. BMC Plant Biology **11**:74. doi: 10.1186/1471-2229-11-74

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe A, Wong GK.-S, Wang J, and Deyholos MK. (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. The Plant Journal **72**: 461-473. doi:10.1111/j.1365-313X.2012.05093.x

Whitney SEC, Gothard MGE, Mitchell JT, and Gidley MJ (1999) Roles of Cellulose and Xyloglucan in Determining the Mechanical Properties of Primary Plant Cell Walls. Plant Physiology **121**: 657-664. doi: 10.1104/pp.121.2..657

Zinn K, McAllister L, and Goodman CS (1988) Sequence Analysis and Neuronal Expression of Fasciclin I in Grasshopper and Drosophila. Cell **53**: 577-587

Chapter 2 - LuFLA1$_{PRO}$ and LuBGAL1$_{PRO}$ promote gene expression in the phloem fibres of flax (*Linum usitatissimum*)[*]

Introduction

Flax (*Linum usitatissimum*) is a major temperate field crop that is cultivated for either its seed or for its phloem fibres. Flax phloem (or bast) fibres are derived from the primary meristems of the plant and undergo intrusive growth to increase in length more than 1,000 fold (Fahn, 1990). Elongation is followed by deposition of a gelatinous-type secondary wall, which is rich in crystalline cellulose with a low microfibrillar angle, giving the fibres very high tensile strength (Gorshkova et al., 2012). Because of their length and strength, flax phloem fibres have been used in the production of linen textiles since antiquity (Deyholos, 2007). The unique physical properties of flax phloem fibres also make them attractive feedstock for production of composite materials (Zini and Scandola, 2011). Understanding the development of these unusual fibres, and improving their traits in both linseed-type and fibre-type cultivars of flax has thus become a subject of intense research (Jhala et al., 2009).

Tissue-specific promoters are a pre-requisite to many biotechnological crop improvement approaches. Heterologous gene expression can have a negative impact on plant fitness when gene products are expressed constitutively (Wróbel-Kwiatkowska et al., 2007; Kalinina et al., 2011). Thus, it is useful to limit gene expression to tissues of interest, limiting both the strain on plant resources and the possible build-up of toxic by-products. We therefore sought to identify promoters that would be useful in both basic and applied flax fibre research.

Promoter elements upstream of sucrose synthases (Yang and Russell, 1990; Shi et al., 1994), sucrose-H+ symporters (AtSUC2; Truernit and Sauer, 1995; Zhao et al., 2004; Schneidereit et al., 2008), sucrose binding proteins (GmSBP2; Waclawovsky et al., 2006), galactinol synthases

[*] A version of this chapter has been published as: Hobson N, and Deyholos MK (2013) LuFLA1$_{PRO}$ and LuBGAL1$_{PRO}$ promote gene expression in the phloem fibres of flax (*Linum usitatissimum*). Plant Cell Reports **32**:517-528.

(CmGAS1; Haritatos et al., 2000; Ayre et al., 2003) and lectins (Yoshida et al., 2002), have been shown to direct gene expression in a phloem-specific manner, however, their expression is detected in the parenchymatous cells of the phloem, and not in the sclerenchymatous fibre cells of the phloem. While promoter elements from lignin biosynthesis genes (Chen et al., 2000; Tiimonen et al., 2007) have been observed to drive gene expression in phloem fibres of poplar and birch, expression also extended into the xylem, limiting their effectiveness in cell specific applications.

A flax cDNA microarray identified both a fasciclin-like arabinogalactan protein (*LuFLA35*) and a beta-galactosidase (*LuBGAL1*) as being highly expressed in cortical tissues near the snap point, a region of the stem in which fibres undergo a transition from fibre elongation to secondary cell wall deposition (Gorshkova et al., 2003; Roach and Deyholos, 2007; Roach and Deyholos, 2008). LuBGAL1 proteins were also highly enriched in developing fibres as compared to surrounding tissues (Hotte and Deyholos, 2008). Whereas the function of *LuBGAL1* in phloem fibre development has been previously described (Roach et al., 2011), the function of *LuFLA35*, and FLAs in general, has not been well established. Many transcriptomic studies have implicated FLAs in the development of fibres with gelatinous secondary cell walls (Hobson et al., 2010). In the flax fibres, the development of this thick gelatinous secondary cell wall requires the deposition and later re-modelling of a cell wall layer concentrated with galactan-rich polysaccharides (Gorshkova et al., 2004; Salnikov et al., 2008). The heavy glycosylation displayed by FLAs (Seifert and Roberts, 2007), and their increased transcription in these fibre-rich tissues, suggests that they may play a role in the development of this galactan-rich layer.

LuFLA1 is a gene closely related to *LuFLA35*. Because *LuFLA1* and *LuBGAL1* appeared to be highly enriched in phloem fibres, as compared to other tissues, we characterized their upstream genomic regions ($LuFLA1_{PRO}$ and $LuBGAL1_{PRO}$) in terms of their ability to promote reporter gene expression in the phloem fibres of flax.

Materials and Methods

Fosmid Library Construction and Screening

DNA was extracted from etiolated flax seedlings of variety CDC Bethune. A fosmid library was produced with the CopyControl™ Fosmid Library Production Kit (EPICENTRE Biotechnologies). The library was pooled (roughly 80 clones per well), and screened via PCR for *LuFLA35* and *LuBGAL1*. AGP_FOR2 (5'-CCTTCATCTCCTCTTCTCAGTTTC-3'), AGP_REV2 (5'-GGAGCAACACCTTGTCCAAC-3'), Bgal_FOR1 (*5'-TCAGCATACTGCCTGATTGC* -3'), and Bgal_REV*1 (5'-CCATCCTCGGTGGTTGTATC-3')* primers were designed from EST sequences referenced as probesets 152 and 4738 from Roach and Deyholos (2007). Later analyses revealed that the FLA gene identified was not *LuFLA35* but was the closely related gene *LuFLA1*. These primers were tested on flax genomic DNA and found to amplify fragments of 201 bp or 275 bp, for *LuFLA1* and *LuBGAL1* respectively. Single isolates were identified after successive rounds of re-streaking and PCR. Amplification of the fosmid clones was induced in 25 ml cultures (CopyControl™ Induction solution from Library Kit), which were purified with a Qiagen Plasmid Midi Kit (Cat. No. 12143). BigDye v3.1 sequencing, using existing primers, confirmed fosmid identity. Shotgun sequencing was performed by first nebulizing ~10 µg of fosmid DNA, repairing fragment ends with End-It™ (EPICENTRE Biotechnologies), and size selecting for 1.5-2 kb fragments on a 1% agarose (1X Tris-acetate EDTA) gel. Extracted fragments were ligated into a SmaI digested and shrimp alkaline phosphatase-dephosphorylated pUC19, and transformed into *Escherichia coli* strain DH5α. 288 clones each were submitted to Canada's Michael Smith Genome Sciences Centre for bidirectional sequencing off of the M13 forward and reverse primers. Sequence assembly was performed with Genetool 2.0 (default settings). Up to 39 kb of genomic sequence per clone was obtained, and the fosmid sequences were deposited in NCBI Genbank under accessions JN133301 and HQ902252.

Plasmid Construction

The LuFLA1 and LuBGAL1 fosmid assemblies were submitted to the Augustus v2.1 server (Stanke et al., 2008), using default Arabidopsis settings for gene predictions. A 908 bp fragment upstream of *LuFLA1* ($LuFLA1_{PRO}$) was PCR amplified from fosmid template with primers modified with HindIII and BamHI restriction tags (AGPp_Bam_3 = 5'-AAGGATCCGGAGATATGCGTGCAGCAA-3'; AGPp_Hind_3 = 5'-AAAAGCTTAGTTGCTGCGGGAGTGAG-3'). A second fragment upstream of *LuFLA1*, comprising 1507 bp ($LuFLA1_{PRO-B}$), was also amplified (AGPp_Bam = 5'-AAGGATCCGAGATATGCAGTGCAGCAAAAT-3'; AGPp_Hind = 5'-TTAAGCTTGGGAATGAATACGTCAAGAGGA-3'), as was a 934 bp fragment upstream of *LuBGAL1* (Bgalp_Bam = 5'-AAGGATCCAGAAAGGGGAATTCTTGATGG-3'; Bgalp_Hind = 5'-TTAAGCTTTCTTCATTATTGCATACGTGGTG-3'). $LuFLA1_{PRO}$:*uidA*, $LuFLA1_{PRO-B}$:*uidA*, and $LuBGAL1_{PRO}$:*uidA* fusion vectors were constructed by ligating the restriction products into the pRD420 cloning vector (Datla et al., 1992). Constructs were transformed into *Agrobacterium tumefaciens* GV3101.

Tissue Culture

Plant transformations of flax variety CDC Bethune were conducted using an adaptation of a published protocol (Mlynárová et al., 1994; Wróbel-Kwiatkowska et al., 2007). Flax seedlings were grown for 6 d on half-strength Linsmaier and Skoog (LS) + 1% sucrose + 0.7% phytablend agar plates. Cut hypocotyl segments were inoculated with Agrobacterium, for 2 h with agitation, in 20 mL co-cultivation media containing LS, 3% Sucrose, 1 mg/L benzyladenine (BA), 0.1 mg/L naphthyl-acetic acid (NAA), and 20 mM acetosyringone. Hypocotyls were transferred to co-cultivation media containing LS, 3% sucrose, 1 mg/L BA, 0.1 mg/L NAA, 100 mM acetosyringone, and 0.7% agar for 3 d. Callus formation and transformant selection was achieved by maintaining explants for 2-3 weeks on shoot initiation/selection media containing LS, 3% sucrose, 1 mg/L BA, 0.1 mg/L NAA, 300 mg/L Timentin, 200 mg/L kanamycin, and 0.7% agar. Calli were excised from the infected hypocotyl ends, and maintained on selective shoot

regeneration media containing LS, 2.5% sucrose, 0.02 mg/L BA, 0.001 mg/L NAA, 300 mg/L Timentin, 220 mg/L kanamycin, and 0.7% agar. Calli were transferred to fresh shoot regeneration media every 2 weeks. Developing shoots were excised from the calli and transferred to shoot elongation media containing LS, 1% sucrose, 100 mg/L Timentin, 100 mg/L kanamycin, and 0.7% agar. After 2 weeks, shoots were transferred to rooting media containing half-strength LS, 1% sucrose, 0.2 mg/L indole-3-butyric acid (IBA), 100 mg/L Timentin, 100 mg/L kanamycin, and 0.7% agar. T_0 transgenics which developed roots were transferred to soil and grown to maturity. T_1 plants were screened via PCR for the *nptII* selectable marker, and by GUS staining of stem fragments for *uidA* expression. T_2 plants were again screen via PCR for the *nptII* selectable marker, as were T_3 plants. T_3 plants were used for the final analysis of promoter expression. We confirmed the presence of the LuFLA1$_{PRO}$:*uidA*, LuFLA1$_{PRO-B}$:*uidA*, and LuBGAL1$_{PRO}$:*uidA* transgenes in seven, nine, and six independent transformation events, respectively. Controls included pRD410 transformants, which carried a CaMV35S:*uidA* fusion, and pRD420 transformants, which carried *uidA* with no promoter.

GUS Histochemistry

Tissues were harvested and placed in ice-cold 90% acetone, vacuum infiltrated for 2 minutes, and incubated at -20°C for 30 minutes. Tissues were washed twice in 50 mM $NaHPO_4$ pH 7.2 before infiltrating in GUS staining solution (0.2% triton X-100, 10 mM EDTA, 50 mM $NaHPO_4$ pH 7.2, 4 mM $K_4Fe(CN)_6$, 4 mM $K_3Fe(CN)_6$, 2 mM X-gluc) for 30 minutes, followed by overnight incubation at 37°C . Tissues were subsequently run through an ethanol and fixation series consisting of 30% ethanol for 1 h, FAA (50% ethanol, 5% formaldehyde, 10% glacial acetic acid) o/n, and 70% ethanol for final storage. Tissue samples were hand-sectioned and photographed with an Olympus BX51 microscope, mounted with a Photometrics CoolSnap fx digital camera.

Bioinformatics

$LuFLA1_{PRO}$, $LuBGAL1_{PRO}$, and sequence comprising the entire intergenic region upstream of *LuBGAL1*, denoted $LuBGAL_{Intergenic}$, were submitted to the PlantPan webservice (http://plantpan.mbc.nctu.edu.tw/; Chang et al., 2008) for promoter analysis. Consensus sequences for all identified cis-elements were obtained and used to count their occurrence in three sequences fragments ($LuFLA1_{PRO}$, $LuBGAL1_{PRO}$, and $LuBGAL_{Intergenic}$) and in a repeat-masked copy of the flax genome assembly (version 1.0; Wang et al., 2012). A one-tailed Z-test was employed to determine whether the frequency of a given cis-element in a promoter was significantly different from the frequency observed in the genome. The following formula was used, where Fp was the frequency in the fragment, Fg was the frequency in the genome, and Np was the length of the fragment: $Z = \frac{Fp - Fg}{\sqrt{\frac{Fg*(1-Fg)}{Np}}}$

We performed a MEME analysis (http://meme.nbcr.net/meme/; Bailey and Elkan, 1994) to identify conserved motifs in both $LuFLA1_{PRO}$ and $LuBGAL1_{PRO}$, selecting for 10-base motif lengths occurring one or more times in the sequences. The ten most significant motifs were submitted to TOMTOM (http://meme.nbcr.net/meme/; Gupta et al., 2007) to determine if the motifs had been previously annotated in the JASPAR CORE plants database. Motif matches with a p-value greater than 0.05 were resubmitted to PlantPan for further comparison to motifs in other promoter databases.

MUG Assay

A single representative line was chosen for each construct, and three plants per line were used in the assay. Tissues were harvested and immediately frozen in liquid N_2, then stored at -80°C. MUG assays were performed as previously described (Jefferson et al., 1987) in 96-well microtitre plates. In brief, frozen samples were ground in liquid nitrogen, and resuspended in 500 µl extraction buffer (50 mM $NaPO_4$ pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 0.1% sodium lauryl sarcosine, 10 mM β-mercaptoethanol; Jefferson et al., 1987). Samples were centrifuged at 14,000 g for 10 min, and the supernatants were collected and stored at -80°C. 10 µl of protein

extract was combined with 90 µl extraction buffer and 100 µl 2 mM MUG (4-Methylumbelliferyl b-D-glucuronide dihydrate). Samples were incubated at 37°C, and 10 µl aliquots were removed at 0, 10, 30 and 60 minute timepoints, to be mixed with 200 ul cold 0.2 M $NaCO_3$. 4-MU (4-Methylumbelliferone) fluorescence was measured with a Fluorstar OPTIMA fluorometer, recording emissions at 460 nm, with an excitation wavelength of 355 nm, and concentrations were determined via standard curve. Protein concentrations were determined via Bradford assay (Bradford, 1976). A single factor analysis of variance (ANOVA) was employed to determine if the values observed in a given tissue differed significantly between constructs. Student's t-test was performed between constructs to determine which specific samples differed from one another.

Results

To identify DNA fragments sufficient to direct gene expression in flax phloem fibres, we used promoter fusions to *uidA* to analyze the genomic regions upstream of two genes (*LuFLA1*, *LuBGAL1*) whose transcripts were known to be enriched in fibre-bearing tissues (Roach and Deyholos, 2007; Roach and Deyholos, 2008). While this approach may not as accurately define transcript expression patterns of *LuFLA1* and *LuBGAL1* as would *in situ* hybridizations, due to the potential presence of regulatory elements further upstream or downstream of the selected sequences, it would allow us to better associate observed expression patterns with specific sequences.

LuFLA1

Sequencing of the *LuFLA1*-containing fosmid revealed the presence of three predicted FLA genes. A predicted FLA adjacent to *LuFLA1*, but encoded on the opposite strand (Figure 1-1), was designated *LuFLA2*. The third, further downstream of *LuFLA2*, was designated *LuFLA3*. LuFLA1 and LuFLA2 shared 58.3% amino acid sequence identity according to local alignment, and their start codons were separated by 913 bp. We cloned 908 bp of this intergenic region for further promoter analysis, and referred to this fragment as $LuFLA1_{PRO}$.

We scanned the sequence of $LuFLA1_{PRO}$ for known regulatory motifs, using PlantPan (Chang et al., 2008). The frequency of occurrence for each motif was compared to its frequency in the repeat-masked whole genome assembly of flax (Wang et al., 2012). Of the identified motifs (Table S1), eight were found to be significantly enriched (P-value ≤ 0.05) in the *LuFLA1* promoter, in comparison to the whole genome (Table 2-1). A sucrose response element (SURE1STPAT21; Grierson et al., 1994), and TCA1MOTIF (Goldsbrough et al., 1993) were found to be particularly enriched in comparison to the whole genome, despite occurring only once in the promoter. Other motifs, such as an E-box element (EBOXBNNAPA; Stålberg et al., 1996) implicated in seed expression, and MYC transcription factor binding sites (MYCCONSENSUSAT; Abe et al., 2003) implicated in stress responses, were among the most prevalent motifs in the promoter (ten occurrences each), but were in turn quite prevalent in the genome as a whole, decreasing their significance (Table S1).

When we examined the effects of the $LuFLA1_{PRO}$:*uidA* construct in flax, we observed GUS staining in the phloem fibres of each line, although staining intensity varied between independent lines. No staining was observed in tissues outside of the stem of any line, including developing or mature flowers or fruits (data not shown). When whole, 3-week old plants were exposed to substrate, staining was observed in the stem and hypocotyl (Figure S1), where it was sometimes strongest in the vicinity of a node. Even within the nodes, expression was concentrated in phloem fibres with some staining (possibly due to leakage of the enzyme or product from intensely stained fibres) found in tissues immediately surrounding the fibres (Figure S1b, c). Further analysis of GUS expression was conducted in three different segments along 5-week old stems, corresponding to three different stages of phloem fibre development (Figure 2-2). Near the apex of the stem, where phloem fibres elongate, no GUS staining was observed (Figure 2-2a). Around the snap-point, where phloem fibres undergo a transition from cell elongation to secondary cell wall deposition, GUS staining was observed in a gradient, with weak GUS staining at the top of the segment and more abundant GUS staining at the bottom of the segment (Figure 2-2d). GUS staining was detected in fibres in the snap-point region before these fibres had acquired a distinguishable secondary wall (Figure 2-2b, c). The number of fibres that showed GUS staining

increased towards the bottom of this segment (Figure 2-2e, f). In the segment from the base of the stem, where mature phloem fibres with thick secondary cell walls were observed, staining intensity was uniform along the length of the segment, rather than being distributed in a gradient (Figure 2-2g). As observed elsewhere in the stem, GUS staining was localized to the phloem fibres (Figure 2-2h, i). Staining intensity appeared to be stronger in fibres from the bottom of the stem as compared to fibres from near the snap-point. All LuFLA1$_{PRO}$:*uidA* lines exhibited the same phloem-fibre specific staining patterns, although differences in staining intensity were observed.

We considered the possibility that the promoter of *LuFLA1* extended into the coding region of *LuFLA2*, and that we had not isolated all the cis-elements directing *LuFLA1* expression *in vivo*. To examine this, we created a promoter reporter fusion called LuFLA1$_{PRO-B}$:*uidA* . Compared to LuFLA1$_{PRO}$:*uidA*, LuFLA1$_{PRO-B}$:*uidA* contained an additional 599 bp of upstream sequence extending partway into *LuFLA2*. The expression pattern of LuFLA1$_{PRO-B}$:*uidA* appeared to be indistinguishable from LuFLA1$_{PRO}$:*uidA* (data not shown). Therefore subsequent analyses focused on the shorter construct, LuFLA1$_{PRO}$:*uidA*.

LuBGAL1

Sequencing of the *LuBGAL1*-containing fosmid revealed the presence of a single BGAL gene near the end of the genomic fragment. A promoter analysis tool (Reese, 2001) predicted transcription initiation sites 199 bp and 841 bp from the end of the genomic fragment, upstream of the *LuBGAL1* start codon. We therefore cloned a 934 bp fragment, encompassing both potential transcription initiation sites, for further analysis.

The subsequent cloning of the *LuBGAL1* cDNA (GenBank Accession HQ902251; Roach et al., 2011) and completion of a flax whole genome sequence assembly (Wang et al., 2012), extended the available genomic sequence upstream of *LuBGAL1*, and identified a 2,010 bp intergenic region separating its start codon from the stop codon of a predicted 5'-nucleotidase (Figure 2-1b). Conserved regulatory motifs within either the cloned promoter fragment, or the entire intergenic region, were identified using PlantPan (Chang et al., 2008), and their frequency

within these fragments was compared to their frequency in the repeat-masked whole genome assembly. Of the identified motifs (Table S1), 12 elements were significantly enriched (P-Value ≤ 0.05) in the cloned *LuBGAL1* promoter, in comparison to the whole genome (Table 2-2). A single occurrence of an abscisic acid responsive element (ABREATRD22; Iwasaki et al., 1995) in the promoter was particularly significant. MYB1AT, a drought and abscisic acid responsive element (Abe et al., 2003), was also significantly represented, occurring five times within the promoter. When the intergenic region upstream of *LuBGAL1* was analyzed in its entirety (Table S2), the number of significant motifs increased from 12 to 16 elements. MRNA3ENDTAH3, a cis-element associated with histone genes (Ohtsubo and Iwabuchi, 1994), was the most significantly enriched in comparison to the whole genome, despite occurring only once in the intergenic space. GT1CONSENSUS, a potentially light regulated cis-element (Terzaghi and Cashmore, 1995), was also significantly represented, occurring 25 times within the intergenic region.

When we examined the effects of the LuBGAL1$_{PRO}$:*uidA* construct in flax stems, we observed GUS expression solely in the phloem fibres of the stem, although staining intensities varied between independent transformation events. When we examined GUS expression in whole 3-week old plants, GUS staining was observed in the stem and hypocotyl (Figure S2a). In both tissues, GUS staining was detectable only in the phloem fibres (Figure S2d, e, f, g). No other staining was observed. We then analyzed GUS expression patterns in three segments along 5-week old stems, corresponding to three different stages of phloem fibre development. At the apex of the stem, no GUS staining was observed (Figure 2-3a). Around the snap-point, weak GUS staining was observed at the base of the segment (Figure 2-3b). Transverse sectioning of this segment revealed no observable GUS staining at the top, where fibres had not yet developed a distinguishable secondary cell wall. Lower down the segment, sectioning revealed weak GUS staining in fibres with distinguishable secondary cell walls (Figure 2-3c, d). At the base of the stem, where mature phloem fibres could be observed, GUS staining was distinguishable along the length of the segment (Figure 2-3e), and was again localized to the phloem fibres, appearing with a greater intensity than in less mature tissues (Figure 2-3f, g). This phloem fibre specific

expression pattern was observed in the stems of all six LuBGAL1$_{PRO}$:*uidA* transgenic lines, although differences in staining intensity were observed.

Finally, we examined GUS expression in developing flowers and seed bolls of LuBGAL1$_{PRO}$:*uidA* transgenic lines. Here, we observed GUS staining in the style, filament, and anthers of developing and mature flowers, as well as at the base of the pistil, filaments, and petals (Figure S3a, b, c,). Differences in staining intensity were apparent between independent lines (Figure S3a, b). In only one of the six independent lines, seed bolls exhibited GUS staining along the outer septum (Figure S3d). No staining was observed in bolls of the other five lines.

MUG assay

To compare the strength of the promoters quantitatively, we measured the rate at which protein extracts could hydrolyze 4-methylumbelliferyl-β-D-glucuronide (MUG). Promoter activity was measured in the basal stems of representative 5-week old LuFLA1$_{PRO}$:*uidA* and LuBGAL1$_{PRO}$:*uidA* lines, as well as in positive and negative controls containing, respectively, pRD410, which carried a CaMV35S$_{PRO}$:*uidA* fusion, and pRD420, which carries a promoterless *uidA*. Only the CaMV35S promoter produced detectable GUS activity in all tissues tested (Figure 2-4). GUS activity in the CaMV35S promoter line was comparable between tissues (p-value > 0.05). In both LuFLA1$_{PRO}$ and LuBGAL1$_{PRO}$, for all the tissues tested, activity was detected predominantly in the phloem fibres. Minute beta-glucuronidase activity was detected in the xylem and epidermis of the LuFLA1$_{PRO}$ transgenic flax (p-value < 0.05), as well as the epidermis of the LuBGAL1$_{PRO}$ transgenic flax (p-value < 0.05), however this was at least 100-fold less activity than that observed in the phloem fibres of each line. In fibres, LuFLA1$_{PRO}$ activity in the representative line was at least three fold-greater than found in the CaMV35S line (p-value < 0.05). LuBGAL1$_{PRO}$ and CaMV35S lines showed comparable levels of reporter gene activity in fibres, and the difference in mean activity was not statistically significant. It must be noted, however, that promoter activity is dependent on position effects, and many other variables, not controlled for in this demonstration of relative promoter strength.

Discussion

We characterized the genomic sequences upstream of *LuFLA1* and *LuBGAL1* (Figure 2-1a, b), and demonstrated that both regions contained sufficient regulatory elements to promote gene expression in the phloem fibres of flax, with little or no expression in other tissues. In transgenic flax, $LuFLA1_{PRO}$ initiated reporter gene expression in fibres at the earliest stages of secondary cell wall development, before a secondary wall was even detectable (Figure 2-2b, c). Because expression was generally most intense in mature fibres (Figure 2-2h, i), we infer that $LuFLA1_{PRO}$ promoted expression in fibres throughout secondary wall development and maturation. These results coincide with previous studies of AGP concentrations in fibre-enriched tissues of developing flax stems (Gorshkova et al., 2004; Gorshkova and Morvan, 2006), suggesting $LuFLA1_{PRO}$ accurately reflects the native expression pattern of *LuFLA1*. The enrichment of reporter gene expression nearly exclusively in phloem fibres strongly suggests that LuFLA1 plays a specific role in the development of the gelatinous-type walls of flax fibres. Transcripts of multiple FLA genes have been previously reported to be highly expressed in tissues containing gelatinous-type cell walls, including the phloem fibres of flax and the tension wood of poplar (*Populus trichocarpa*; Hobson et al., 2010). It therefore seems possible that the enrichment of FLA expression we observed would be conserved in other species with gelatinous-type cell walls.

The genomic sequence upstream of *LuFLA1* contained a single sucrose response element (SURE, Table 2-1). SUREs were originally identified in tuber-expressed genes of potato, and induce transcription in response to sucrose (Grierson et al., 1994). The sucrose-responsive function of this motif is conserved in cellulose synthase promoters of poplar, which is a member of the same taxonomic order as flax (Ohmiya et al., 2003). It is therefore possible that *LuFLA1* is likewise induced by sucrose, potentially regulating cell wall deposition in concert with available carbon and energy, possibly as part of a network that includes cellulose synthases. $LuFLA1_{PRO}$ also contained light regulated cis-elements, such as BOXCPSAS1 and SORLIP5AT (Ngai et al., 1997; Hudson and Quail, 2003). Similar to the SURE element, these elements may aid in

regulating expression to coincide with periods of increased available carbohydrates, and active periods of secondary cell wall deposition.

The presence of a second upstream FLA, within 1 kb of *LuFLA1* and in the opposite orientation, may influence the expression of *LuFLA1 in vivo*. However, the expression conferred by LuFLA1$_{PRO-B}$, with an additional 599 bp of sequence extending into *LuFLA2*, appeared indistinguishable from that conferred by LuFLA1$_{PRO}$, suggesting either that *LuFLA2* has a negligible effect on *LuFLA1* expression, or that additional regulatory motifs are further afield. Still, the close proximity of this second anti-parallel FLA gene also raises the possibility that LuFLA1$_{PRO}$ acts bi-directionally. Bi-directional promoters are not uncommon, and have been observed in both plants (Keddie et al., 1994; Wang et al., 2008) and animals (Liu et al., 2011). Neighboring genes, in parallel and divergent orientations, have also been observed to co-express in Arabidopsis (Williams and Bowles, 2004). This raises the possibility that *LuFLA2* might too be expressed primarily in phloem fibres, although the expression pattern of this gene has not yet been characterized.

The second upstream genomic fragment we tested, LuBGAL1$_{PRO}$, promoted reporter gene expression that was first detectable after secondary cell wall deposition was already evident (Figure 2-3e, f). LuBGAL1$_{PRO}$ expression therefore became active later in development than LuFLA1$_{PRO}$ expression. As expression from LuBGAL1$_{PRO}$ generally increased in intensity as fibres matured (Figure 2-3h, i), we inferred that LuBGAL1$_{PRO}$ promoted expression in fibres throughout fibre development and maturation. Phloem fibre development is marked by the presence of two distinct cell wall layers. The inner Gn-layer contains a heavy concentration of large galactan-rich polysaccharides, observed as distinct bands between thin layers of cellulose (Gorshkova and Morvan, 2006; Salnikov et al., 2008), while the outer G-layer is more cellulose rich. As the fibre matures, the galactan-rich polysaccharides of the Gn-layer are degraded by LuBGAL1 until the cell wall is remodelled into a crystalline cellulose rich G-layer (Roach et al., 2011). Our observations thus support the current understanding of the role of the LuBGAL1 as a cell wall remodelling protein, and suggest that a developmental delay is inherent in secondary cell wall maturation, where, likely, a significant Gn-layer is initially established before galactan

degradation begins. LuBGAL1$_{PRO}$ also promoted gene expression in the developing flowers of transgenic flax (Figure S3). Based on our current understanding of the role of LuBGAL1 in cell wall remodelling, this would suggest that galactan rich polysaccharides play an important role in the development of these tissues.

LuBGAL1$_{PRO}$ contained a number of annotated cis-elements, reportedly regulated by abscisic acid (ABREATRD22, Table 2-2; Iwasaki et al., 1995), cold stress (CRTDREHVCBF2; Xue, 2003), and phosphate availability (P1BS; Rubio et al., 2001). Likewise, it contained elements associated with leaf and seed expression (MYB1AT; Abe et al., 2003), seed and endosperm (PROLAMINBOXOSGLUB1; Wu et al., 2000), and developing flowers (CORE; Tan and Irish, 2006). Some of these may explain gene expression in the developing flowers; however, it is currently unclear as to what role these might play in the developmental regulation of gene expression observed during fibre maturation.

LuBGAL1$_{Intergenic}$, which comprised the entire intergenic space upstream of *LuBGAL1*, and contained 1076 bp of intergenic sequence not present in our LuBGAL1$_{PRO}$:*uidA* fusion, was found to contain additional annotated cis-elements. Some of these, such as L1BOXATPDF1, which regulates expression in a cell layer, and cell-type, specific manner (Abe et al., 2001; Wang et al., 2004), and WBOXNTCHN48 (Yamamoto et al., 2004), which regulates elicitor-induced expression, could conceivably influence the native expression of *LuBGAL1*. On the other hand, some of these elements, such as MRNA3ENDTAH3 (Ohtsubo and Iwabuchi, 1994), which regulates 3' end formation in histone mRNAs, may influence the expression of the upstream gene, and are perhaps unrelated to *LuBGAL1* gene expression entirely. Therefore, the identification of previously annotated cis-elements in this sequence should be regarded with caution until further promoter characterization is performed.

While both LuFLA1$_{PRO}$ and LuBGAL1$_{PRO}$ promoted gene expression in the phloem fibres of flax, their sequences shared few annotated motifs. The presence of different light responsive elements (SORLIP1AT, REALPHALGLHCB21; Hudson and Quail, 2003; Degenhardt and Tobin, 1996), sucrose responsive elements (WBOXHVISO1; Sun et al., 2003), and circadian rhythm elements (CIACADIANLELHC; Piechulla et al., 1998), suggest that the timing of gene expression

may be regulated by circadian rhythm and the availability of free carbohydrates, to more efficiently make use of plant resources. Importantly, both DOF-class and MYB-class transcription factor binding sites were quite prevalent in these two sequences. DOF-class transcription factors are unique to the plant kingdom, and some members of this family have been found to be expressed predominantly in the phloem (Skirycz et al., 2006; Guo et al., 2009). The presence of core DOF binding sites, and their importance, in the promoters of phloem-specific genes (Ayre et al., 2003; Freitas et al., 2007; Schneidereit et al., 2008), suggests an equally important role in our phloem fibre enriched genes. Likewise, while the MYB-class of transcription factor has been implicated in a variety of roles (Dubos et al., 2010), its role in regulating secondary cell wall biosynthesis in xylem fibres and vessels marks it as another potentially important regulator of phloem fibre expression (Zhong et al., 2007 ; Zhong et al., 2008 ; Zhou et al., 2009). In support of this view, transcripts for MYB-like genes have been observed to be enriched in cortical peels around the snap-point and lower stem (Roach and Deyholos, 2007), tissues rich in maturing phloem fibres.

Additional motifs may have been present in both promoters, conferring phloem fibre specificity, that were not identified in our PlantPan database search. When we scanned both $LuFLA1_{PRO}$ and $LuBGAL1_{PRO}$ with MEME (data not shown; Bailey and Elkan, 1994), to determine whether novel motifs were present, we found that each motif identified matched a motif previously found using the PlantPan databases. Therefore, novel conserved motifs could not be defined based on the current analyses.

Conclusion

Our findings demonstrated that the $LuFLA1_{PRO}$ and $LuBGAL1_{PRO}$ fragments we extracted from the upstream genomic regions of *LuFLA1* and *LuBGAL1* contained sufficient cis-acting regulatory elements to direct transgene expression in developing and mature phloem fibres of flax. Unlike the phloem-specific promoters described elsewhere (Shi et al., 1994; Haritatos et al., 2000; Yoshida et al., 2002), $LuFLA1_{PRO}$ and $LuBGAL1_{PRO}$ did not drive gene expression in the parenchymatous cells of the phloem, but rather in the sclerenchymatous fibre cells of the

phloem. Of the two fragments, LuFLA1$_{PRO}$ showed the highest activity and highest specificity, and may therefore find the greatest utility in most biotechnological applications. These features of the LuFLA1$_{PRO}$ expression pattern also point to an important role for this protein in development of the secondary wall of fibres. Further characterization is necessary to identify the motifs responsible for cell-specific expression, which may lead to the identification of key regulators of phloem fibre development. Knowledge gained from such studies may in turn lead to new strategies for managing phloem fibre quantity and quality in an important agricultural product.

Figures and Tables

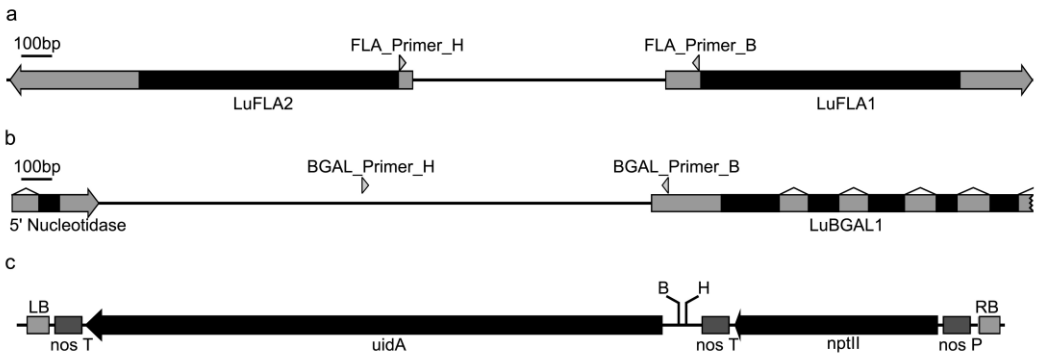

Figure 2-1 LuFLA1 and upstream genomic sequence (a). *LuBGAL1* and upstream genomic sequence (b). T-DNA of pRD420 vector (c). Genomic fragments are to scale (scale bar = 100 bp). T-DNA is not sized to scale. Coding sequences are coloured black. Primers are denoted by arrow heads. BamHI (B) or HindIII (H) restriction sites/tags are listed in vector sequence and primer labels.

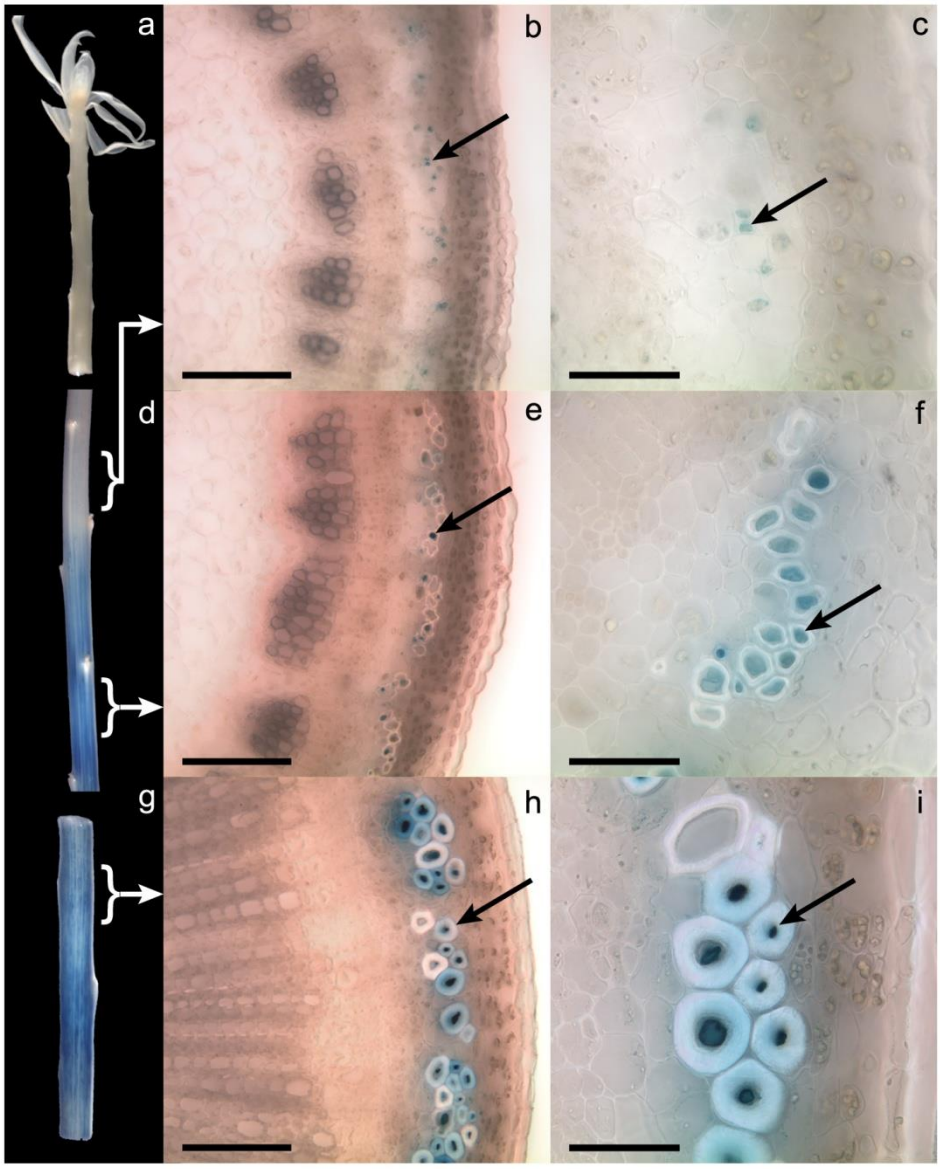

Figure 2-2 GUS staining in 5-week old LuFLA1$_{PRO}$:uidA transgenic flax. Tissues examined were the shoot apex (a), the snap-point (d), and the stem base (g). Transverse sections above the snap-point were examined (b, c), as were sections below the snap-point (e, f), and sections at the stem base (h, i). Black arrows indicate GUS stained bast fibres. Scale bar for b, e, and h is 150 µm. Scale bar for c, f, and i is 50 µm.

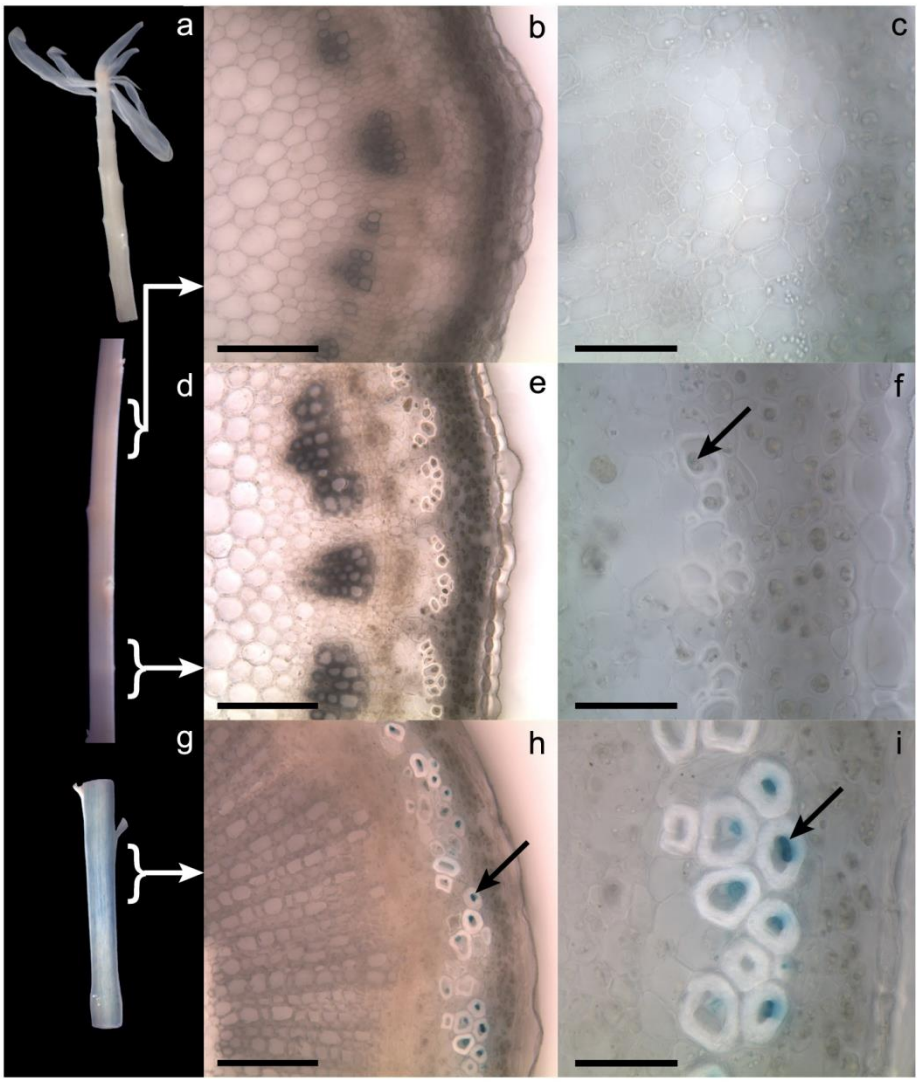

Figure 2-3 GUS staining in 5-week old LuBGAL1$_{PRO}$:uidA transgenic flax. Tissues examined were the shoot apex (a), the snap-point (d), and the stem base (g). Transverse sections above the snap-point were examined (b, c), as were sections below the snap-point (e, f), and sections at the stem base (h, i). Black arrows indicate GUS stained bast fibres. Scale bar for b, e, and h is 150 µm. Scale bar for c, f, and i is 50 µm.

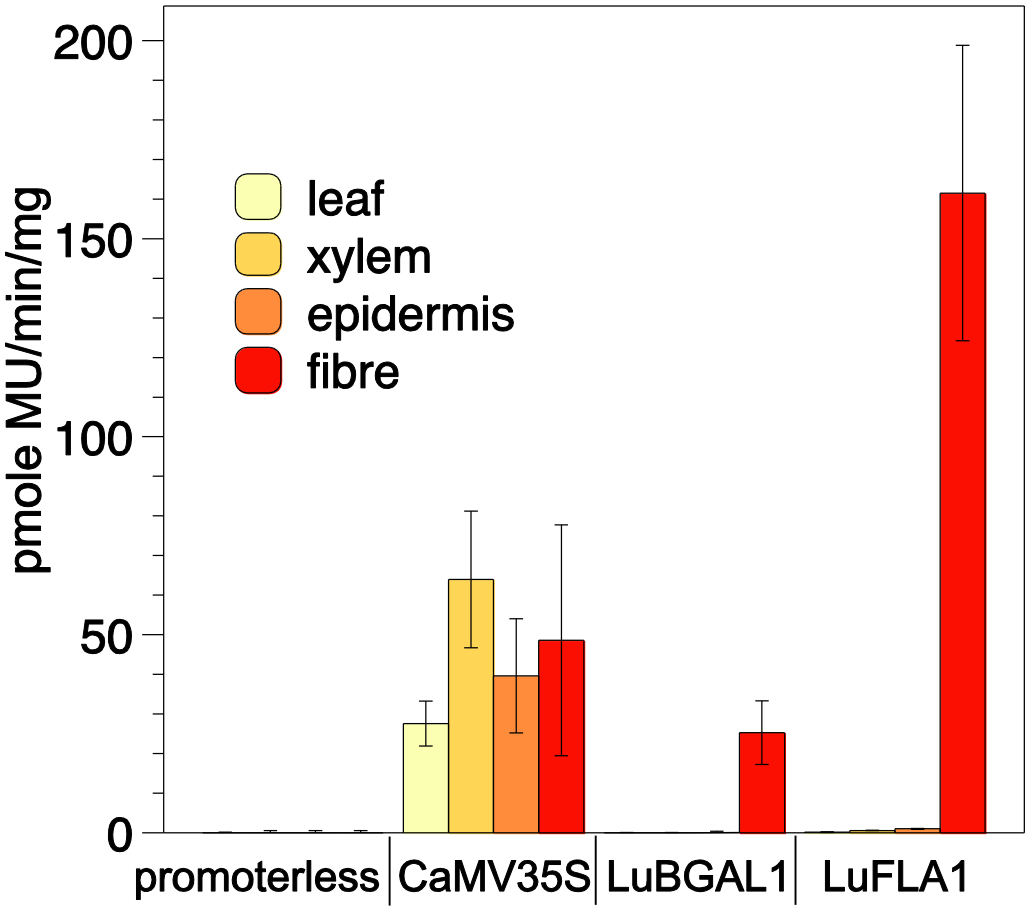

Figure 2-4 Beta-glucuronidase activity in transgenic flax. The hydrolysis of 4-methylumbelliferyl-β-D-glucuronide (MUG) into 4-methylumbelliferone was recorded for protein extracts from leaves, xylem, epidermal tissue, and phloem fibres from transgenic flax bearing promoterless:*uidA*, CAMV35S:*uidA*, LuFLA1$_{PRO}$:*uidA*, or LuBGAL1$_{PRO}$:*uidA*. A single factor analysis of variance (ANOVA) was performed for each tissue, and a significant difference in beta-glucuronidase was found between the transgenic lines (p = value <0.01).

Table 2-1 Cis-elements within LuFLA1$_{PRO}$ (P-Value ≤ 0.05).

Cis-Element	Consensus Sequence	Description	Occurrence	P-value
TCA1MOTIF	TCATCTTCTT	NtTCA-1 binding site (Associated with stress response)	1	4.328E-12
SURE1STPAT21	AATAGAAAA	SURE (Sucrose Response Element)	1	4.796E-07
CEREGLUBOX2PSLEGA	TGAAAACT	PsLEGa cis-element	1	0.0002120
BOXCPSAS1	CTCCCAC	PsAS1 cis-element (light regulated)	1	0.002399
SORLIP5AT	GAGTGAG	SORLIP motif (Sequence Overly Represented in Light Induced Promoters)	1	0.002422
SEBFCONSSTPR10A	YTGTCWC	SEBF motif (Silencing Element Binding Factor)	2	0.007986
CPBCSPOR	TATTAG	CsPOR cis-element (Cytokinin regulated)	2	0.02208
ANAERO3CONSENSUS	TCATCAC	Cis-element of anaerobic genes	1	0.03429

Cis-elements were identified using PlantPan (Chang et al., 2008). Z-Tests were used to determine the probability that the frequency within the DNA fragment was not greater than the frequency within the repeat-masked whole genome shotgun assembly (Wang et al., 2012). IUPAC ambiguity codes are used to describe consensus DNA sequences. Y = C/T, R = A/G, W = A/T, S = G/C, K = T/G, M = C/A, D = A/G/T, V = A/C/G, H = A/C/T, B = C/G/T, N = any base.

Table 2-2 Cis-elements within LuBGAL1$_{PRO}$ (P-Value ≤ 0.05).

Cis-Element	Consensus Sequence	Description	Occurrence	P-value
ABREATRD22	RYACGTGGYR	ABRE (ABA responsive element)	1	0.0000
CRTDREHVCBF2	GTCGAC	HvCBF2 binding site (Temperature regulated)	2	0.0001219
SEF4MOTIFGM7S	RTTTTTR	GmSEF4 binding site	5	0.001058
RUNX1	WWDTGHGGTWW	HsAML-1 binding site	1	0.003241
MYB1AT	WAACCA	MYB-class TF binding site (Abscisic acid and drought regulated)	5	0.004817
P1BS	GNATATNC	AtPHR1 binding site (Phosphate regulated)	2	0.007339
MYB1	MTCCWACC	MYB-class TF binding site	1	0.008506
PROLAMINBOXOSGLUB1	TGCAAAG	Prolamine box (Quantitative regulation)	1	0.01776
id1	TTKYYYYTWBYG	ZmID1 binding site (Flowering time regulation)	1	0.02127
CATATGGMSAUR	CATATG	GmSAUR15 a cis element (Auxin response)	2	0.02936
CAATBOX1	CAAT	CAAT-box	17	0.03144
2SSEEDPROTBANAPA	CAAACAC	Cis-element of storage-protein genes	1	0.04546

Cis-elements were identified using PlantPan (Chang et al., 2008). Z-Tests were used to determine the probability that the frequency within the DNA fragment was not greater than the frequency within the repeat-masked whole genome shotgun assembly (Wang et al., 2012). IUPAC ambiguity codes are used to describe consensus DNA sequences. Y = C/T, R = A/G, W = A/T, S = G/C, K = T/G, M = C/A, D = A/G/T, V = A/C/G, H = A/C/T, B = C/G/T, N = any base.

References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, and Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. The Plant Cell **15**: 63–78. doi:10.1105/tpc.006130

Abe M, Takahashi T, and Komeda Y (2001) Identification of a *cis*-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. The Plant Journal **26**: 487-494. doi:10.1046/j.1365-313x.2001.01047.x

Ayre BG, Blair JE, and Turgeon R (2003) Functional and Phylogenetic Analyses of a Conserved Regulatory Program in the Phloem of Minor Veins. Plant Physiology **133**: 1229–1239. doi:10.1104/pp.103.027714

Bailey TL, and Elkan C (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceeding of the Second International Conference on Intelligent Systems for Molecular Biology, AAAI Press, Menlo Park, California, pp 28–36

Bradford MM (1976) A Rapid and Sensitive Method for the Quantitiation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. Analytical Biochemistry **72**: 248-254. doi:10.1016/0003-2697(76)90527-3

Chang WC, Lee TY, Huang HD, Huang HY, and Pan RL (2008) PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. BMC Genomics **9**: 561. doi:10.1186/1471-2164-9-561

Chen C, Meyermans H, Burggraeve B, De Rycke RM, Inoue K, De Vleesschauwer V, Steenackers M, Van Montagu MC, Engler GJ, and Boerjan WA (2000) Cell-specific and conditional expression of caffeoyl-coenzyme A-3-O-methyltransferase in poplar. Plant Physiology **123**: 853–867. doi:10.1104/pp.123.3.853

Datla RS, Hammerlindl JK, Panchuk B, Pelcher LE, and Keller W (1992) Modified binary plant transformation vectors with the wild-type gene encoding NPTII. Gene **122**: 383-384. doi:10.1016/0378-1119(92)90232-E

Degenhardt J, and Tobin EM (1996) A DNA binding activity for one of two closely defined phytochrome regulatory elements in an Lhcb promoter is more abundant in etiolated than in green plants. The Plant Cell **8**: 31–41. doi:10.1105/tpc.8.1.31

Deyholos M (2007) Phloem fibre of flax (*Linum usitatissimum* L.): Biological foundations of its ancient and modern uses. Israel Journal of Plant Sciences **54**: 273-280. doi:10.1560/IJPS_54_4_273

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, and Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. Trends in Plant Science **15**: 573-581. doi:10.1016/j.tplants.2010.06.005

Fahn A (1990) Plant Anatomy. 4th Ed. Pergamon Press, Oxford, UK.

Freitas RL, Carvalho CM, Fietto LG, Loureiro ME, Almeida AM, and Fontes EPB (2007) Distinct repressing modules on the distal region of the SBP2 promoter contribute to its vascular tissue-specific expression in different vegetative organs. Plant Molecular Biology **65**: 603–14. doi:10.1007/s11103-007-9225-0

Goldsbrough AP, Albrecht H, and Stratford R (1993) Salicylic acid-inducible binding of a tobacco nuclear protein to a 10 bp sequence which is highly conserved amongst stress-inducible genes. The Plant Journal **3**: 563–571. doi:10.1046/j.1365-313X.1993.03040563x

Gorshkova TA, Sal'nikov VV, Chemikosova SB, Ageeva MV, Pavlencheva NV, and van Dam JEG (2003) The snap point: a transition point in *Linum usitatissimum* phloem fibre development. Industrial Crops and Products **18**: 213-221. doi:10.1016/S0926-6690(03)00043-8

Gorshkova TA, Chemikosova SB, Sal'nikov VV, Pavlencheva NV, Gur'janov OP, Stolle-Smits T, and van Dam JEG (2004) Occurrence of cell-specific galactan is coinciding with phloem fiber developmental transition in flax. Industrial Crops and Products **19**: 217-224. doi:10.1016/j.indcrop.2003.10.002

Gorshkova T, and Morvan C (2006) Secondary cell-wall assembly in flax phloem fibres: role of galactans. Planta **223**: 149-158. doi:10.1007/s00425-005-0118-7

Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, and Pilate G (2012) Plant fibre Formation: State of the Art, Recent and Expected Progress, and Open Questions. Critical Reviews in Plant Sciences **31**: 201-228. doi:10.1080/07352689.2011.616096

Grierson C, Du JS, de Torres Zabala M, Beggs K, Smith C, Holdsworth M, and Bevan M (1994) Separate *cis* sequences and *trans* factors direct metabolic and developmental regulation of a potato tuber storage protein gene. The Plant Journal **5**: 815-826. doi:10.1046/j.1365-313X.1994.5060815.x

Guo Y, Qin G, Gu H, and Qu LJ (2009) *Dof5.6/HCA2*, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in *Arabidopsis*. The Plant Cell **21**: 3518–3534. doi:10.1105/tpc.108.064139

Gupta S, Stamatoyannopoulos JA, Bailey TL, and Noble WS (2007) Quantifying similarity between motifs. Genome biology **8**: R24. doi:10.1186/gb-2007-8-2-r24

Haritatos E, Ayre BG, and Turgeon R (2000) Identification of Phloem Involved in Assimilate Loading in Leaves by the Activity of the Galactinol Synthase Promoter. Plant Physiology **123**: 929-937. doi:10.1104/pp.123.3.929

Hobson N, Roach MJ, and Deyholos MK (2010) Gene expression in tension wood and phloem fibres. Russian Journal of Plant Physiology **57**: 321-327. doi:10.1134/S1021443710030039

Hotte NSC, and Deyholos MK (2008) A flax fibre proteome: identification of proteins enriched in phloem fibres. BMC Plant Biology **8**: 52. doi:10.1186/1471-2229-8-52

Hudson ME, and Quail PH (2003) Identification of Promoter Motifs Involved in the Network of Phytochrome A-Regulated Gene Expression by Combined Analysis of Genomic Sequence and Microarray Data. Plant Physiology **133**: 1605-1616. doi:10.1104/pp.103.030437.DNA

Iwasaki T, Yamaguchi-Shinozaki K, and Shinozaki K (1995) Identification of a *cis*-regulatory region of a gene in *Arabidopsis thaliana* whose induction by dehydration is mediated by abscisic acid and requires protein synthesis. Molecular and General Genetics **247**:391–398. doi:10.1007/BF00293139

Jhala AJ, Weselake RJ, and Hall LM (2009) Genetically Engineered Flax: Potential Benefits, Risks, Regulations, and Mitigation of Transgene Movement. Crop Science **49**: 1943-1954. doi:10.2135/cropsci2009.05.0251

Jefferson RA, Kavanagh TA, and Bevan MW (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. The EMBO Journal **6**: 3901-3907

Kalinina O, Zeller SL, and Schmid B (2011) Competitive Performance of Transgenic Wheat Resistant to Powdery Mildew. PLoS ONE **6**: e28091. doi:10.1371/journal.pone.0028091

Keddie JS, Tsiantis M, Piffanelli P, Cella R, Hatzopoulos P, and Murphy DJ (1994) A seed-specific *Brassica napus* oleosin promoter interacts with a G-box-specific protein and may be bi-directional. Plant Molecular Biology **24**: 327–340. doi:10.1007/BF00020171

Kleines M, Elster RC, Rodrigo MJ, Blervacq AS, Salamini F, and Bartels D (1999) Isolation and expression analysis of two stress-responsive sucrose-synthase genes from the resurrection plant *Craterostigma plantagineum* (Hochst.). Planta **209**:13-24

Liu, B, Chen J, and Shen B (2011) Genome-wide analysis of the transcription factor binding preference of human bi-directional promoters and functional annotation of related gene pairs. BMC Systems Biology **5**:S2. doi:10.1186/1752-0509-5-S1-S2

Mlynárová L, Bauer M, Nap JP, and Pretová A (1994) High efficiency Agrobacterium-mediated gene transfer to flax. Plant Cell Reports **13**:282-285. doi:10.1007/BF00233320

Ngai N, Tsai FY, and Coruzzi G (1997) Light-induced transcriptional repression of the pea AS1 gene: identification of cis-elements and transfactors. The Plant Journal **12**:1021-1034. doi:10.1046/j.1365-313X.1997.12051021.x

Ohmiya Y, Nakai T, Park YW, Aoyama T, Oka A, Sakai F, and Hayashi T (2003) The role of PopCel1 and PopCel2 in poplar leaf growth and cellulose biosynthesis. The Plant Journal **33**:1087-1097. doi:10.1046/j.1365-313X.2003.01695.x

Ohtsubo N, and Iwabuchi M (1994) The conserved 3'-flanking sequence, AATGGAAATGG, of the wheat histone H3 gene is necessary for the accurate 3'-end formation of mRNA. Nucleic Acids Research **22**:1052–1058. doi:10.1093/nar/22.6.1052

Piechulla B, Merforth N, and Rudolf B (1998) Identification of tomato *Lhc* promoter regions necessary for circadian expression. Plant Molecular Biology **38**:655-662. doi:10.1023/A:1006094015513

Reese MG (2001) Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Computers and Chemistry **26**:51-56. doi:10.1016/S0097-8485(01)00099-7

Roach MJ, and Deyholos MK (2007) Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. Molecular Genetics and Genomics **278**:149-165. doi:10.1007/s00438-007-0241-1

Roach MJ, and Deyholos MK (2008) Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fibre differentiation. Annals of Botany **102**:317–330. doi:10.1093/aob/mcn110

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, and Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant Physiology **156**:1351-1363. doi:10.1104/pp.111.172676

Rubio V, Linhares F, Solano R, Martín AC, Iglesias J, Leyva A, and Paz-Ares J (2001) A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. Genes & Development 15:2122–2133. doi:10.1101/gad.204401

Salnikov VV, Ageeva MV, and Gorshkova TA (2008) Homofusion of Golgi secretory vesicles in flax phloem fibers during formation of the gelatinous secondary cell wall. Protoplasma **233**:269-273. doi:10.1007/s00709-008-0011-x

Seifert GJ, and Roberts K (2007) The biology of arabinogalactan proteins. Annual Review of Plant Biology **58**:137-161. doi:10.1146/annurev.arplant.58.032806.103801

Shi Y, Wang MB, Powell KS, Damme EV, Hilder VA, Gatehouse AMR, Boulter D, Gatehouse JA (1994) Use of the rice sucrose synthase-1 promoter to direct phloem-specific expression of β-glucuronidase and snowdrop lectin genes in transgenic tobacco plants. Journal of Experimental Botany **45**:623–631. doi:10.1093/jxb/45.5.623

Schneidereit A, Imlau A, and Sauer N (2008) Conserved *cis*-regulatory elements for DNA-binding-with-one-finger and homeo-domain-leucine-zipper transcription factors regulate companion cell-specific expression of the *Arabidopsis thaliana SUCROSE TRANSPORTER 2* gene. Planta **228**:651-662. doi:10.1007/s00425-008-0767-4

Skirycz A, Reichelt M, Burow M, Birkemeyer C, Rolcik J, Kopka J, Zanor MI, Gershenzon J, Strnad M, Szopa J, Mueller-Roeber B, and Witt I (2006) DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in Arabidopsis. The Plant Journal **47**:10–24. doi:10.1111/j.1365-313X.2006.02767.x

Stålberg K, Ellerstöm M, Ezcurra I, Ablov S, and Rask L (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic Brassica napus seeds. Planta **199**:515–519. doi:10.1007/BF00195181

Stanke M, Diekhans M, Baertsch R, and Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics **24**:637-644. doi:10.1093/bioinformatics/btn013

Sun C, Palmqvist S, Olsson H, Borén M, Ahlandsberg S, and Jansson C (2003) A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the *iso1* promoter. The Plant Cell **15**:2076-2092. doi:10.1105/tpc.014597

Tan QK, and Irish VF (2006) The Arabidopsis Zinc Finger-Homeodomain Genes Encode Proteins with Unique Biochemical Properties That Are Coordinately Expressed during Floral Development. Plant Physiology **140**:1095–1108. doi:10.1104/pp.105.070565

Terzaghi WB, and Cashmore AR (1995) LIGHT-REGULATED TRANSCRIPTION. Annual Reviews of Plant Physiology and Plant Molecular Biology **46**:445–474. doi:10.1146/annurev.pp.46.060195.002305

Tiimonen H, Häggman H, Tsai CJ, Chiang V, and Aronen T (2007) The seasonal activity and the effect of mechanical bending and wounding on the PtCOMT promoter in *Betula pendula* Roth. Plant Cell Reports **26**:1205–1214. doi:10.1007/s00299-007-0331-x

Truernit E, and Sauer N (1995) The promoter of the *Arabidopsis thaliana* SUC2 sucrose-H+ symporter gene directs expression of beta-glucuronidase to the phloem: evidence for phloem loading and unloading by SUC2. Planta **196**:564–570. doi:10.1007/BF00203657

Waclawovsky AJ, Freitas RL, Rocha CS, Contim LAS, and Fontes EPB (2006) Combinatorial regulation modules on *GmSBP2* promoter: a distal cis-regulatory domain confines the *SBP2* promoter activity to the vascular tissue in vegetative organs. Biochimica et Biophysica Acta **1759**:89–98. doi:10.1016/j.bbaexp.2006.02.002

Wang C, Ding D, Yan R, Yu X, Li W, and Li M (2008) A Novel Bi-directional Promoter Cloned from Melon and Its Activity in Cucumber and Tobacco. Journal of Plant Biology **51**:108–115. doi:10.1007/BF03030719

Wang S, Wang JW, Yu N, Li, CH, Luo B, Gou JY, Wang LJ, and Chen XY (2004) Control of Plant Trichome Development by a Cotton Fiber MYB Gene. The Plant Cell **16**:2323-2334. doi:10.1105/tpc.104.024844

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe A, Wong GK.-S, Wang J, and Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. The Plant Journal **72**: 461-473. doi:10.1111/j.1365-313X.2012.05093.x

Williams EJB, and Bowles DJ (2004) Coexpression of Neighboring Genes in the Genome of *Arabidopsis thaliana*. Genome Research **14**:1060–1067. doi:10.1101/gr.2131104.1060

Wróbel-Kwiatkowska M, Starzycki M, Zebrowski J, Oszmiański J, and Szopa J (2007) Lignin deficiency in transgenic flax resulted in plants with improved mechanical properties. Journal of Biotechnology **128**:919-934. doi:10.1016/j.jbiotec.2006.12.030

Wróbel-Kwiatkowska M, Zebrowski J, Starzycki M, Oszmiański J, and Szopa J (2007) Engineering of PHB synthesis causes improved elastic properties of flax fibres. Biotechnology Progress **23**:269-277. doi:10.1021/bp0601948

Wu CY, Washida H, Onodera Y, Harada K, and Takaiwa F (2000) Quantitative nature of the Prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal *cis*-element requirements for endosperm-specific gene expression. The Plant Journal **23**:415–421. doi:10.1046/j.1365-313x.2000.00797.x

Xue GP (2003) The DNA-binding activity of an AP2 transcriptional activator HvCBF2 involved in regulation of low-temperature responsive genes in barley is modulated by temperature. The Plant Journal **33**:373–383. doi:10.1046/j.1365-313X.2003.01630.x

Yamamoto S, Nakano T, Suzuki K, Shinshi H (2004) Elicitor-induced activiation of transcription viaWbox-related *cis*-acting elements from a basic chitinase gene by WRKY transcription factors in tobacco. Biochimica et Biophysica Acta **1679**:279–287. doi:10.1016/j.bbaexp.2004.07.005

Yang NS, and Russell D (1990) Maize sucrose synthase-1 promoter directs phloem cell-specific expression of *Gus* gene in transgenic tobacco plants. PNAS **87**:4144–4148. doi:10.1073/pnas.87.11.4144

Yoshida K, Mohri T, Nishiguchi M, and Tazaki K (2002) *Robinia pseudoacacia* inner-bark lectin promoter expresses GUS also predominantly in phloem of transgenic tobacco. Journal of Plant Physiology **159**:757–764. doi:10.1078/0176-1617-0814

Zini E, and Scandola M (2011) Green Composites: An Overview. Polymer Composites **32**:1905-1915. doi:10.1002/pc

Zhao Y, Liu Q, and Davis RE (2004) Transgene expression in strawberries driven by a heterologous phloem-specific promoter. Plant Cell Reports **23**:224–230. doi:10.1007/s00299-004-0812-0

Zhong R, Richardson EA, and Ye ZH (2007) The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis.* The Plant Cell **19**:2776-92. doi:10.1105/tpc.107.053678

Zhong R, Lee C, Zhou J, McCarthy RL, and Ye ZH (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. The Plant Cell **20**:2763-2782. doi:10.1105/tpc.108.061325

Zhou J, Lee C, Zhong R, and Ye ZH (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. The Plant Cell **21**:248-266. doi:10.1105/tpc.108.063321

Chapter 3 - Whereupon we examine the glycosyl hydrolase 35 family of flax[*]

Introduction

In 1894, an enzyme preparation was found to catalyze lactose hydrolysis (Fischer, 1894), initiating the study of proteins we have come to know as β-D-galactoside galactohydrolases (β-galactosidases). In the proceeding decades, a β-galactosidase protein was purified from *Escherichia coli* for kinetic studies (Kuby and Lardy, 1953), and *LacZ*, a bacterial gene coding for a β-galactosidase, was characterized during a seminal examination of the *lac* operon and transcriptional regulation (Jacob and Monod, 1961).

β-Galactosidases (EC 3.2.1.23) hydrolyze the terminal non-reducing β-D-galactose residues in β-D-galactosides, such as lactose, proteoglycans, glycolipids, oligosaccharides, and polysaccharides (IUBMB; www.enzyme-database.org). Other classes of enzymes are known to hydrolyze bonds involving galactose residues (EC 3.2.1.85; EC 3.2.1.89; EC 3.2.1.102; EC 3.2.1.103; EC 3.2.1.145; EC 3.2.1.164; EC 3.2.1.18), however, the nature of the substrate and/or reaction mechanism of these enzymes is sufficiently different from EC 3.2.1.23 β-galactosidases as to render these enzyme classes distinct (IUBMB; www.enzyme-database.org)

Distributed across kingdoms, β-galactosidases are represented in bacteria, fungi, plants and animals. Based on sequence and structural similarity, EC 3.2.1.23 β-galactosidases can be placed in five of the current 131 glycosyl hydrolase (GH) families: GH1, GH2, GH3, GH35, and GH42 (Cantarel et al., 2009). Plant β-galactosidases have been found only in GH35; β-galactosidases from the other four families have been observed solely in bacteria and archaea. Henceforth, we will use the term BGAL to refer to any GH35 β-galactosidase-like gene.

[*] A version of this chapter has been published as: Hobson N, and Deyholos MK (2013) Genomic and expression analysis of the flax (*Linum usitatissimum*) family of glycosyl hydrolase 35 genes. BMC **14**: 344.

In plants, BGALs have been found to play a role in: the degradation of cell wall polysaccharides; promoting fruit softening in tomatoes and papaya (Smith et al., 2002; Lazan et al., 2004); organization of cellulose microfibrils in flax phloem fibres (Roach et al., 2011; Mokshina et al., 2012); promoting cell elongation in arabidopsis (Sampedro et al., 2012); and facilitating the secretion of seed mucilage, also in arabidopsis (Dean et al., 2007).

The BGALs of flax (*Linum usitatissimum*) have not been well studied. To date, only a single flax β-galactosidase (*LuBGAL1*) has been characterized, which has an important role in the development of cell walls of phloem fibres (Roach et al., 2011). The recent publication of a draft flax genome sequence (Wang et al., 2012) now allows researchers to study industrially relevant gene families in their entirety, such as the previously reported analysis of the UDP glycosyltransferase 1 family (Barvkar et al., 2012). We describe here a detailed analysis of the primary structure, evolutionary history, and transcript expression patterns of 43 putative β-galactosidases in flax.

Materials and Methods

Gene Discovery

The 43,384 predicted proteins of the flax genome (Wang et al., 2012), available at Phytozome (version 8.0; www.phytozome.net), were first queried via BLASTP for sequences similar to the 17 known arabidopsis BGALs (AtBGALs 1-17; TAIR v.10; Lamesh et al., 2011). The default settings of BLAST package 2.2.25+ were used. Sequence matches were filtered for e-values $\leq 1^{-10}$, and then assessed via Hidden Markov Model (HMM) with HMMER3 (http://hmmer.janelia.org), using the Pfam-A family database (version 25.0; Punta et al., 2012), for genes encoding a glycosyl hydrolase 35 domain (GH35). Comparisons of gene family size were performed with a one-tailed Z-test of proportions.

Phylogenetics

Predicted protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Physcomitrella patens*, *Populus trichocarpa*, *Ricinus communis*, and *Zea mays* were obtained from Phytozome (version 8.0; www.phytozome.net; Tuskan et al., 2006; Ouyang et al., 2007; Rensing et al., 2008; Swarbreck et al., 2008; Chan et al., 2010). Sequences were assessed via Hidden Markov Model (HMM) with HMMER3 (http://hmmer.janelia.org), using the Pfam-A family database (version 25.0; Punta et al., 2012), for genes putatively encoding a GH35 domain. Retrieved sequences were labelled as BGALs (Table S3), using published BGAL names (e.g. AtBGAL1) wherever possible (Ahn et al., 2007; Tanthanuch et al., 2008). Amino acid sequences were aligned using the default parameters of Muscle 3.7 (Edgar, 2004), with a human beta-galactosidase (GLB1), obtained from NCBI genbank (NP_000395), as outgroup. ProtTest 3.2, with default parameters, was used to determine the best-fit model of amino acid substitution for a maximum likelihood analysis of the sequence alignment (Darriba et al., 2011). All model tests agreed upon the use of the WAG model of amino acid substitution (Whelan and Goldman, 2001), with gamma-distributed rate variations, which was used to performed a maximum likelihood analysis with bootstrap using GARLI v2.0 (www.molecularevolution.org; Zwickl, 2006; Bazinet and Cummings, 2011). The consensus tree of 1000 bootstraps was obtained using CONSENSE (Phylip 3.66) at the CIPRES Science Gateway (Miller et al., 2010).

EST Identification

Genomic sequence of putative flax BGALs, including 1 kb upstream and downstream of their respective start and stop codons, were used as queries in a BLASTN search against the *Linum usitatissimum* NCBI-nr and NCBI-EST datasets (accessed August, 2012), as well as transcript assembly POZS (www.onekp.com), comprising a *de novo* assembly of Illumina sequenced transcripts from three flax stem fragments. All sequence matches were downloaded and aligned to the predicted LuBGAL CDSs using the RNA-Seq analysis tool of CLC Genomics Workbench 5.5. Only sequences aligning to CDSs with 95% identity, along 90% of their length, were recorded.

Microarray Analyses

Flax microarray datasets GSE21868 (Fénart et al., 2010) and GSE29345 (Huis et al., 2012) were obtained from NCBI GEO. Experiment GSE21868 examined expression in a range of tissues and organs: roots (R); leaves (L); outer stem tissues at either the vegetative stage (SOV) or green capsule stage (SOGC); inner stem tissues at either vegetative stage (SIV) or green capsule stage (SIGC); and seeds 10-15 days after flowering (DAF; E1), 20-30 DAF (E2), and 40-50 DAF (E3; Fénart et al., 2010). Experiment GSE29345 focused on the development of stem tissues by comparing: internal (i.e. xylem enriched) stem tissues of either the whole stem (WSI), upper stem (USI), middle stem (MSI), or lower stem (LSI); and external (i.e. phloem and cortex enriched) stem tissues of the whole stem (WSE), upper stem (USE), middle stem (MSE), and lower stem (LSE; Huis et al., 2012). The flax unigenes used in microarray construction (http://tinyurl.com/c888l9d) were aligned to the predicted *LuBGAL* CDSs, using the RNA-Seq function of the CLC Genomics Workbench 5.5, and were classified as matches if at least 90% of their sequence length aligned, with at least 95% sequence identity between the transcript and CDS. Microarray data corresponding to the flax BGALs were then extracted. Robust Multichip Average (RMA)-normalized signal intensities ($\log_2$) were averaged between biological and technical replicates. Heat maps of expression levels were then created with MeV v4.8 (www.tm4.org/mev).

A Combimatrix microarray dataset examining five stages of flax stem development was produced in our laboratory (manuscript in preparation). The array profiled 1 cm stem fragments from the shoot apex (T1), sections of the snap-point corresponding to various stages of fibre development (T2-4), and lower stem with phloem fibres exhibiting a greater degree of secondary cell wall deposition (T5). Probes, 33-40 nt in length, corresponding to predicted *LuBGALs* from an earlier draft of the flax genome (unpublished) were aligned to the current *LuBGAL* CDS predictions (version 1.0; Wang et al., 2012) using the RNA-Seq function of CLC Genomic Workbench 5.5. Only probes with 100% identity to existing *LuBGAL* CDSs were analyzed. Gene signal intensities were normalized as fractions of mean array signal intensity. The $\log_2$ normalized *LuBGAL* intensities, averaged between four biological replicates, were then used to create heat maps of expression levels with MeV v4.8 (www.tm4.org/mev).

Expression Analysis of LuBGALs

Tissue samples from *Linum usitatissimum* (CDC Bethune) were frozen in liquid nitrogen, and stored at -80°C prior to use. Frozen samples were ground in liquid nitrogen, whereupon we followed the CTAB/Acid Phenol/Silica Membrane Method (Johnson et al., 2012) to extract the RNA. DNA was removed using on-column RNase-Free DNase (Qiagen), and/or with the TURBO DNA-Free kit (Invitrogen). cDNA was prepared with RevertAid H Minus Reverse Transcriptase (Fermentas) and oligo$(dT)_{18}$ primer. qPCR primer pairs and hydrolysis probes (Table S4) were designed with the Universal Probe Library Assay Design Center (http://tinyurl.com/7u6s5bh). A 14 cycle pre-amplification of the target sequences was performed with a TaqMan PreAmp Master Mix (ABI) and 5 ng of cDNA, which was subsequently diluted 1:5. Assay master mixes of 3.2 µl 2X Assay Loading Reagent (Fluidigm PN 85000736), 2 µl primer mix (13.3 µM primer and 3.3 µM hydrolysis probe) and 1.3 µl water was prepared, of which 5 µl was loaded into the assay wells of a primed Fluidigm 96*96 well chip. Sample master mixes of 3.63 µl Taqman Universal PCR Master Mix - no AmpErase UNG (PN 4324018), 0.36 µl 20X GE Sample Loading Reagent (Fluidigm PN 85000735), and 2.5 ul diluted pre-amped cDNA were prepared, of which 5 µl was loaded into the sample wells of the primed Fluidigm 96*96 well plate. The Fluidigm chip was run through the following thermal cycles: 95ºC – 10 min, 40X cycles of 95ºC – 15 sec and 60ºC – 1 min. ΔC_T values were calculated based on the geometric mean of reference genes *ETIF1* (eukaryotic translation initiation factor 1), *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase), and *ETIF5A* (eukaryotic translation initiation factor 5A; Schmittgen and Livak, 2008; Huis et al., 2010). We compared expression in 12 different tissues: roots (R); leaves (L); senescing leaves (SL); stem apex (SA); cortical peels from vegetative stage stems (ECP) or green capsule stage stems (LCP); phloem fibres from vegetative stage stems (EF) or green capsule stage stems (LF); xylem from vegetative stage stems (X); budding flowers (FB); open flowers (F); and seed bolls from the green capsule stage (B). A heat map of relative expression values ($\log_2$), averaging technical (two for F, FB, L, and SL; three for all other samples) and biological (three, each of

which is a pooled sample from multiple plants) replicates, was then prepared with MeV v4.8 (www.tm4.org/mev).

Results

Gene discovery and *in silico* analyses

A combination of BLASTP searches and PFAM analyses resulted in the identification of 43 putative flax β-galactosidases (BGALs), on 34 separate scaffolds of the *de novo* flax genome assembly (Table 3-1; Wang et al., 2012). Using the same approach for gene discovery, we compared the size of the flax BGAL families in 23 representative plant genomes obtained through Phytozome (version 8.0; www.phytozome.net). We found that, relative to the number of protein coding loci in the genomes, flax had the second largest BGAL family, comprising 0.0989% of the total gene coding loci (Figure 3-1), significantly larger than the average BGAL family size (p-value < 0.01). In comparison, amongst the 23 species examined, the BGAL gene family represented an average of 0.0596% of the protein coding loci, or roughly 22 BGAL family members per species. The best-characterized examples include the BGAL families of *Arabidopsis thaliana* and *Oryza sativa*, for which 17 and 15 BGALs have been respectively described (Ahn et al., 2007; Darriba et al., 2011). Even other members of the Malpighiales, such as *Populus trichocarpa* and *Ricinus communis*, contained half the number of BGALs as flax, at 23 and 21 members respectively (Table S3).

To determine which of the predicted *LuBGAL* genes were expressed, we used BLASTN to align the *LuBGAL* CDS sequences with the NCBI-nr and NCBI-EST databases (accessed August 2012), and with *de novo* transcriptome assemblies of developing flax stems (www.onekp.com). At the time of writing, the NCBI-EST database contained 286,852 sequences from *Linum usitatissimum*, 74.8% of which were obtained from flax seeds at various stages of development (Venglat et al., 2011). Ninety-three transcript sequences were identified, which aligned unambiguously to 21 of the *LuBGAL* CDSs (Table 3-1), providing evidence for expression of approximately half the predicted LuBGAL family members. However, because only a limited

number of tissues and conditions were represented by the EST sequences queried, it is likely that additional *LuBGALs* may also be expressed.

As described above, the predicted LuBGALs were defined by the presence of a GH35 domain, which was identified by alignment to PFAM HMM profiles. With one exception, in all of these proteins the GH35 domain was located near the N-terminus, beginning within the first 30-70 amino acids (Table 3-2). The one exception, LuBGAL24, contained a GH35 domain that started at position 568 of the peptide sequence, and was further distinguished by the presence of three N-terminal copper oxidase domains preceding the GH35 domain. The predicted LuBGALs were also searched for the presence of a GH35 active site (Henrissat, 1998), which contains the consensus sequence G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY]. Two of the 43 predicted LuBGALs (LuBGAL35 and LuBGAL43) lacked the consensus active site entirely (Figure S4). Another nine LuBGALs contained major deviations from the consensus active site; these either lacked the catalytic glutamate residues, as in LuBGAL26, or contained a series of insertions and substitutions in the active sites, as in LuBGALs 14, 20, 21, 22, 23, 24, 25, and 36. We note, however, that these deviations were not supported by ESTs. In addition to the GH35 domain, plant BGALs have occasionally been found to contain a putative galactose-binding lectin domain at the C-terminal end of the peptide sequence (Ozeki et al., 1991; Trainotti et al., 2001; Ahn et al., 2007; Tanthanuch et al., 2008). This cysteine rich domain has been proposed to increase the catalytic efficiency of BGAL proteins (Ahn et al., 2007), and was found in only 22 of the 43 LuBGALs (Table 3-1), distributed roughly evenly amongst the different BGAL sub-families.

Unlike the described BGALs of rice (Tanthanuch et al., 2008) and arabidopsis (Ahn et al., 2007), which are ~700-900 aa in length, the length of predicted flax BGALs was more variable in size (Table 3-2). Four putative flax BGALs (LuBGALs 14, 18, 25, and 43) were under 300 aa in length, while another two, LuBGALs 22 and 24, were greater than 1300 aa, with LuBGAL24 containing three copper oxidase domains at the N-terminus. Of these six atypically sized BGALs, only *LuBGAL22* and *LuBGAL24* are represented among ESTs or transcript assemblies (Table 3-1). In addition to the arabidopsis and rice BGAL genes previously described (Ahn et al., 2007; Tanthanuch et al., 2008), we also identified an additional putative BGAL in each of these species,

which we designated AtBGAL18 and OsBGAL16, respectively. AtBGAL18 was previously identified (Ahn et al., 2007), but was not named. Both of these predicted proteins were less than 500 aa in length, and both lacked a consensus GH35 active site.

To determine the predicted subcellular localization patterns of the predicted LuBGALs, we analyzed the protein sequences for possible signal peptides, using SignalP 4.0 (Table 3-2; Petersen et al., 2011). We found that 32 of the 43 LuBGAL sequences contained a predicted signal peptide, generally located within the first 19-35 amino acids. The other 11 LuBGAL sequences, ranging in size from 229 to 869 aa, did not contain a signal peptide. We further employed WolfPSORT and Plant-mPLOC (Horton et al., 2007; Chou and Shen, 2010), and obtained a range of predicted subcellular destinations. In the case of Plant-mPLOC, proteins were predominantly predicted to localize to the cell wall, in some cases despite the lack of N-terminal signal peptide. Only eight LuBGALs were given alternative localization predictions, ranging from the cell membrane (LuBGALs 14, 24, 41, and 43), to the cytoplasm (LuBGALs 41, and 42) and chloroplast (LuBGALs 12, 14, 18, and 26). WolfPSORT was more variable in its predictions, with upwards of seven different predictions per putative LuBGAL. Predictions for the transport to the chloroplast and vacuoles were the most common, followed by the endoplasmic reticulum, extracellular space, and the cytoplasm. Surprisingly, a few LuBGALs were even predicted to most likely be localized to the nucleus (LuBGALs 25, 33, and 35). Experimental characterization will be required to validate these predictions.

Phylogenetic analyses

To classify LuBGALs based on sequence similarity, we performed a phylogenetic analysis using deduced amino acid sequences of the predicted BGAL coding genes from the genome assemblies of *L. usitatissimum*, *P. trichocarpa*, *R. communis*, *Physcomitrella patens*, *O. sativa*, *Zea mays*, and *A. thaliana* (Figure 3-2; Table S3). The rice, arabidopsis, and physcomitrella BGAL families were included because they had been studied previously and form the basis of the plant BGAL sub-family nomenclature (Tanthanuch et al., 2008; Gantulga et al., 2009). The poplar and castor bean BGAL families were included as they are members of the order

Malpighiales, and are relatives of flax for which whole genome sequence is available. Flax BGALs were represented in all of the BGAL sub-families, with the exception of sub-family A3, which was a bryophyte-specific cluster. In the majority of sub-families, the BGALs of flax outnumbered the BGALs of other plant species. Two exceptions to this were observed. First, flax was found to have significantly smaller representation in sub-family B (p-value < 0.01), compared to other species, with only LuBGAL43 present. By comparison, *P. trichocarpa* and *R. communis*, sequenced relatives in the same taxonomic order as flax, had five and seven BGALs, respectively, in sub-family B. Second, sub-family A2 also had a single flax representative, although, (in contrast to sub-family B) all other vascular plants in sub-family A2 were also represented by a single member. As with other vascular plants, sub-family A1 contained the largest number of LuBGAL genes, with 14 representatives, including LuBGAL1, which has been described as an important contributor to flax phloem fibre maturation (Roach et al., 2011).

Transcript expression in public microarray datasets

We examined transcript expression patterns of the LuBGAL family using publicly available oligonucleotide microarray data, beginning with two experiments on a Nimblegen 25-mer oligonucleotide array (NCBI GEO experiment accessions GSE21868 and GSE29345; Fénart et al., 2010; Huis et al., 2012). Probes for these microarrays were designed from ESTs, and not the whole genome. Based on alignments where >90% EST length match the LuBGAL CDSs at >95% sequence identity, these microarrays contain probes for four different LuBGAL genes (*LuBGAL3*, *LuBGAL5*, *LuBGAL6*, and *LuBGAL22*). A heat map of expression values from these microarrays (Figure 3-3 a, b) showed that *LuBGAL3* expression was enriched in the stem during vegetative growth (Figure 3-3a), with its highest expression in the phloem rich outer stem tissues of the upper stem (Figure 3-3b). *LuBGAL22* was also enriched in select tissues, and during a narrow developmental timeframe, with its greatest expression occurring in the seeds 10-15 days after flowering. Within the stem, *LuBGAL22* appeared to be more enriched in the outer stem tissues of the lower stem (Figure 3-3b). On the other hand, while *LuBGAL5* expression was not specific to any one tissue (Figure 3-3a), within the stem of vegetatively growing flax, its

expression appeared enriched in the inner stem, especially in the upper stem, around the snap-point (Gorshkova et al., 2003) where resistance to mechanical bending is first detectable, although expression was also quite high in the inner tissues of the lower stem. *LuBGAL6* did not appear to be particularly enriched in any tissue.

We further examined microarray data from a recent Combimatrix oligonucleotide array analysis of flax stem development conducted in our laboratory (manuscript in preparation). Probes for this microarray were designed from a preliminary, unpublished draft of the flax genome. After alignment to the published flax genome assembly (version 1.0; Wang et al., 2012), 27 probes aligned to 15 distinct *LuBGAL* CDS sequences, with multiple probes corresponding to individual genes for added replication. A heat map of expression values (Figure 3-4) showed that a number of genes were enriched at specific developmental stages. *LuBGAL20* was clearly enriched at the shoot apex, with decreasing expression as the stem matured. *LuBGAL9* appeared enriched just above the snap-point, with expression slightly lower just below the snap-point and further down the stem, and at its lowest at the apex. *LuBGAL34* was also enriched at the snap-point, however unlike *LuBGAL9*, its expression was enriched at the lower end of this region. *LuBGAL1* and *LuBGAL2* were the last set of genes to show enrichment at a developmental stage, with their greatest expression occurring in the more mature stem tissue. While whole stem tissues were used in this assay, our previous analysis of the *LuBGAL1* promoter region provides strong evidence that the expression of this gene is specific to the phloem fibres of the stem (Hobson and Deyholos, 2013).

qRT-PCR analysis of LuBGAL expression

Because the available microarray data sets provided transcript expression profiles for only 17 of the 43 predicted LuBGALs, we performed qRT-PCR in a Fluidigm 96*96 array, to obtain additional information about where and when members of the LuBGAL family are transcribed. With the exception of *LuBGAL20* primers, which may have amplified both *LuBGAL20* and *LuBGAL21*, primers used in the qRT-PCR analysis were verified as being gene specific following a series of BLASTN searches against the scaffolds and CDSs of the flax

genome assembly. We were able to detect gene expression for 42 of the 43 LuBGAL genes in at least one of the tissues sampled (Figure 3-5). We could not detect expression for *LuBGAL4* in any of the tissues tested, despite identifying 34 matching ESTs in numerous databases (Table 3-2). This may be a false negative due to the primers; primer design options for the gene were constrained by high sequence identity to other members of the gene family and so were targeted to a putative 3'UTR of *LuBGAL4*. Maturing fibres (EF) had the greatest diversity of LuBGAL family gene expression, with 40/43 genes detected, followed by xylem, with 31/43 genes detected.

Comparing gene expression across tissues, many LuBGALs showed their highest transcript expression in tissues associated with thick secondary cell walls, i.e. the phloem fibres and xylem of vegetative stage flax stems. *LuBGAL7* expression was detected only in the early phloem fibres, whereas *LuBGALs 27, 28,* and *38* were detected in either early phloem fibres and xylem, or in early phloem fibres and budding flowers. Among the more widely expressed genes, *LuBGALs 9, 15, 16, 18, 21,* and *39* were found to be the most highly expressed LuBGALs, with clear expression peaks in the phloem fibres of green-capsule stage flax, as well as in the roots and seed bolls. Lastly, our results confirmed that *LuBGAL1,* whose upstream genomic region was found to drive expression almost exclusively in phloem fibres (Hobson and Deyholos, 2013), showed greater gene expression in the phloem fibres of vegetatively growing flax, in comparison to the other tested tissues.

Discussion

An emerging role for β-galactosidases shows them to be important facilitators of cell wall metabolism in plants. Here, we identified 43 putative BGALs from flax, which were distributed throughout each of the previously defined BGAL sub-families of vascular plants. The relatively large number of genes in the LuBGAL family, and the abundance of LuBGALs compared to BGALs of other species in each of the sub-families (Figure 3-2), is consistent with the recent genome duplication in the flax lineage (Wang et al., 2012). Thus, most LuBGALs exist in pairs and likely share similar functions. Nevertheless, certain variations in the organization of the

LuBGAL proteins suggest a degree of sub-functionalization and selection unique to the species, especially with regards to the reduction in the number of LuBGALs in sub-family B (Figure 3-2).

Aside from being the sole flax representative in sub-family B, LuBGAL43 was also the shortest predicted protein in the LuBGAL family at only 107 amino acids (Table 3-2), compared to the average 700-800 amino acids, and entirely lacked a GH35 active site (Figure S4). While *AtBGAL1*8 and three *RcBGALs* in sub-family B likewise lacked a canonical GH35 active site, other sub-family B BGALs from these (and other) species had the canonical catalytic residues. Currently, no study has yet explored the biochemical function of sub-family B BGALs. Expression data have revealed that *AtBGAL7* and *AtBGAL15*, arabidopsis members of subfamily B, are expressed in flowers and pollen (Ahn et al., 2007; Hrubá et al., 2005), whereas *AtBGAL18* is expressed in seedlings and roots (Hrubá et al., 2005). Similar to *AtBGAL7* and *AtBGAL15*, *OsBGALs 5*, *12*, *14*, and *15*, the rice representatives of sub-family B, have also shown enrichment in reproductive tissues, which led to the hypothesis that the ancestors to sub-family B developed a reproductive-tissue specific role antecedent to the divergence of monocots and dicots (Tanthanuch et al., 2008). We may further speculate that the cell wall development in flax reproductive tissues has a reduced requirement for sub-family B LuBGALs with classical GH35 active sites, as compared to vegetative tissues. Alternatively, a role for BGALs in the development of flax reproductive tissues may yet remain, but may be provided by members of different sub-families, although no individual *LuBGAL* showed enriched expression in these tissues. To better explore these possibilities, it will be important to explore the biochemical and physiological roles of sub-family B in other plant species, including testing their substrate specificity, to determine why sub-family B is not maintained in flax as in other species.

Analyses of the arabidopsis and rice BGAL families had identified 17 and 15 members respectively (Ahn et al., 2007; Tanthanuch et al., 2008; Gantulga et al., 2009). Our own analysis of these genomes added an additional member to each species family, both of which were under 500 amino acids in length, and both of which lacked the putative GH35 active site (Henrissat, 1998). In flax, we identified two LuBGALs, LuBGAL35 and LuBGAL43, which lacked this active site entirely, and another nine, LuBGALs 14, 20, 21, 22, 23, 24, 25, 26, and 36, which

contained either partial active sites, insertions within the active sites, or a series of substitutions in key amino acids (Figure S4). In Arabidopsis, BGAL activity has been characterized in AtBGAL1, AtBGAL2, AtBGAL3, AtBGAL4, AtBGAL5, AtBGAL6, AtBGAL10, and AtBGAL12 (Ahn et al., 2007; Dean et al., 2007; Gantulga et al., 2008; Gantulga et al., 2009; Sampedro et al., 2012), all of which contain consensus GH35 active sites. The radish RsBGAL1, characterized as a BGAL hydrolyzing β-(1→3)- and β-(1→6)-galactosyl residues, also contains the consensus GH35 active site (Kotake et al., 2005), as does a recently characterized chickpea BGAL (Kishore and Kayastha, 2012), and a number of other cloned BGALs (Smith et al., 1998; Smith and Gross, 2000). In fact, all biochemically verified plant BGALs reported to date contain the consensus GH35 active site. Therefore, the absent, partial, and altered GH35 active sites in predicted LuBGAL proteins may indicate a shift in substrate specificity and/or enzyme kinetics, if not a complete lack of enzymatic activity.

LuBGALs 20-25 make up the entirety of sub-family A5 in flax, which, in additional to being composed entirely of LuBGALs with non-conserved GH35 active sites, is also of interest due to the manner in which the sub-family has expanded in comparison to related species (p-value < 0.01). Rice, arabidopsis, poplar, and castor each contain a single member in sub-family A5, whereas flax contained six members. *Arabidopsis lyrata*, *Medicago truncatula*, *Vitis vinifera*, *Aquilegia coerulea*, *Cucumis sativus*, *Prunus persica*, *Mimulus guttatus*, *Brachypodium dystachion*, *Setaria italica*, *Sorghum bicolor*, *Zea mays*, *Nasturtium microphyllum*, *Solanum lycopersicum*, and *Pyrus communis* have also been described as containing a single sub-family A5 representative (Sampedro et al., 2012). Exceptions occur in *Citrus sinensis*, *Citrus clementina*, *Glycine max*, and *Eucalyptus grandis*, where two members of sub-family A5 were recorded (Sampedro et al., 2012). With regards to the changes in the putative GH35 active sites, the mutations shared between protein pairs, such as those observed in LuBGAL22 and LuBGAL24, as well as those observed in LuBGAL20 and LuBGAL21, would suggest that the divergence in sequence from sub-family A5 orthologs occurred early in the evolution of this branch of the Linaceae, at least predating the last genome duplication (Wang et al., 2012). In addition to the changes in the GH35 active site, LuBGAL22, LuBGAL24, and LuBGAL25 are also of uncommon

size. LuBGAL22 and LuBGAL24 are over 1300 aa in length, and, in the case of LuBGAL24, containing additional N-terminal copper oxidase domains, possibly the result of a gene fusion. In contrast, LuBGAL25 appears truncated, coding for a protein 297 aa in length. AtBGAL10, the sole arabidopsis member of sub-family A5, has been described as the main xyloglucan β-galactosidase of arabidopsis, where T-DNA insertions in *AtBGAL10* have led to a 90% decrease in BGAL activity against XLLG substrates, where G refers to an unsubstituted glucose residue of the xyloglucan backbone, X refers to a glucose substituted with α-D-Xylp sidechain, and L refers to a glucose residue substituted with β-D-Galp-(1→2)- α-D-Xylp sidechain (Sampedro et al., 2012). Expression of *AtBGAL10* was observed to be quite strong in developing flowers, the columella cells and elongation zone of the roots, as well as the in the developing vasculature, trichomes, and guard cells of the leaves, all of which are areas of intense cell wall remodelling for cell division and expansion (Sampedro et al., 2012). *LuBGAL21*, too, was strongly expressed in roots, and developing seed bolls. *LuBGAL22* was observed to be expressed strongly in seeds early in development (Figure 3-3a), while *LuBGAL20* appeared to be strongly expressed in the shoot apex (Figure 3-4), all of which might indicate a role in cell division. The remainder of the sub-family A5 *LuBGAL*s were primarily expressed in vegetatively growing phloem fibres (Figure 3-5), which exhibit secondary cell wall deposition as opposed to cell division or elongation.

BGAL sub-family A1 is the best studied of all the BGALs, having been described as encoding exogalactanases, generally hydrolyzing β-(1,3)- and β-(1,4)-linked galacto-oligosaccharides of the cell wall (Ahn et al., 2007; Gantulga et al., 2008), and, in the case of AtBGAL12, additionally hydrolyzing β-(1,6)-galacto-oligosaccharides (Gantulga et al., 2009). In flax, LuBGAL1 has previously been posited to play an important role in the degradation of high molecular weight poly-galactans in the secondary cell walls of phloem fibres. When silenced, the reduction in LuBGAL1 activity (and possible reduction in LuBGAL2 activity) leads to retention of these pectic galactans, which apparently results in reduced crystallization of cellulose, thus reducing the structural integrity of flax stems (Roach et al., 2012). Further characterization of the *LuBGAL1* promoter region supports high specificity of expression in phloem fibres (Hobson and Deyholos, 2013), which our expression analyses reported here have again confirmed (Figures 3-4

and 3-5). It appears likely that other LuBGALs in sub-family A1 share similar functions as LuBGAL1, based on conservation of their coding sequences and similarity of their expression patterns. Sequences sharing the greatest similarity to LuBGAL1 exhibited a very similar pattern of expression: *LuBGALs 2, 3, 7, 6, and 5*, which comprised the same branch of sub-family A1 as LuBGAL1, consistently showed greater expression in tissues rich in secondary cell walls, be it phloem fibres or xylem (Figure 3-5). The sole exception was *LuBGAL4*, for which no expression has been detected in either microarray or qRT-PCR. In some cases, such as *LuBGAL5*, expression was also strong in developing seeds (Figure 3-3a), however this overlap with reproductive tissues has been likewise observed in *LuBGAL1* (Hobson and Deyholos, 2013). Perhaps unsurprisingly, *LuBGAL2*, the most similar paralog of *LuBGAL1*, appears to follow the same expression pattern as it relates to developmental stages in the stem (Figure 3-4), being expressed just below the snap-point, where the secondary cell walls of phloem fibres begin to exhibit the shift from a galactan rich Gn-layer to a more cellulose rich G-layer (Gorshkova et al., 2003). The other major group within sub-family A1 (*LuBGALs 8*, *9*, *12*, *13*, *14*, *15*, and *16*) appear more varied in expression. While some members, such as *LuBGALs 8*, *11*, and *14* are particularly enriched in fibres and xylem, others, such as *LuBGALs 9*, *15*, and *16*, are more strongly expressed throughout the plant, with greater expression in roots (Figure 3-5). We note that these genes do also show expression in stem tissues, however, expression appears restricted to different developmental stages (Figure 3-4). In the case of *LuBGAL9*, expression was observed to occur above the snap-point, which, in the case of phloem fibres, is where cells are still undergoing cell elongation (Gorshkova et al., 2003). All told, the general expression pattern of this branch of sub-family A1 suggests that their function has diverged further from *LuBGAL1* than its immediate sisters.

BGAL sub-family C2 is also a well-characterized group of BGALs. Mutations in *AtBGAL6* (*MUM2*) inhibit the secretion of pectinaceous seed mucilage during hydration (Dean et al., 2007). The LuBGALs with the most sequence similarity to *AtBGAL6* were *LuBGALs 34-37*, and their expression was detected in seed capsules, with the exception of *LuBGAL36.* Greater characterization will be required to determine whether these genes play a similar role in seed coat development.

The remainder of the flax BGALs were observed to express themselves in a variety of tissues, with over half observed to be most strongly expressed in the phloem fibres of vegetatively growing flax stems, relative to the other examined tissues (Figure 3-5). The maturation of flax phloem fibres involves the deposition and later degradation of a large galactan-rich polysaccharide (Gorshkova et al., 2006), which is likely one of the main substrates of these BGAL proteins.

It should be noted that slight differences in expression patterns were observed when comparing genes across Nimblegen, Combimatrix, and Fluidigm platforms. We attribute this to differences in binding efficiencies between cDNA and probes of the microarrays, and cDNA, primers, and hydrolysis probes of the qPCR analyses. Additionally, each platform utilized a different cultivar of flax, grown under dissimilar environmental conditions. Therefore, we attempted to focus not on minor differences in expression between tissues, but rather on the larger differences.

Conclusion

Forty-three putative BGAL genes were identified in the genome of *Linum usitatissimum*. Clustered into eight distinct sub-families, the flax BGAL family was observed to be large in comparison to other sequenced species, with distinct differences in family composition not observed in related species of the order Malpighiales, including a reduction in gene representation in sub-family B, an increased representation in sub-family A5, and many alterations to the typically consensus GH35 active site in a large number of LuBGALs. Using a combination of EST, microarray, and qRT-PCR data, we were able to detect the expression of each member of the LuBGAL family. Almost every *LuBGAL* was expressed in the fibres, the majority of which were predominantly expressed in fibres, compared to other tissues. This suggests that the expansion of the LuBGAL family played an important role in the development of this species as a fibre crop. Further characterization will be necessary to better elucidate their precise function in flax development.

Tables and Figures

Table 3-1 Summary of glycosyl hydrolase 35 encoding gene homologues

Gene Name	Genomic Contig	Gene ID	mRNA[a]	ESTs[a]	Scaffold Gap (bp)
LuBGAL1	scaffold1486	Lus10008974.g	1	3	N
LuBGAL2	scaffold540	Lus10028848.g		4	N
LuBGAL3	scaffold328	Lus10006009.g		16	N
LuBGAL4	scaffold156	Lus10040557.g		5	N
LuBGAL5	scaffold504	Lus10000701.g		0	N
LuBGAL6	scaffold630	Lus10015625.g		8	N
LuBGAL7	scaffold196	Lus10037644.g		6	N
LuBGAL8	scaffold1252	Lus10000803.g		0	N
LuBGAL9	scaffold16	Lus10024292.g		0	N
LuBGAL10	scaffold204	Lus10006733.g		1	N
LuBGAL11	scaffold1376	Lus10011237.g		0	N
LuBGAL12	scaffold275	Lus10014278.g		4	Y (494)
LuBGAL13	scaffold319	Lus10025980.g		4	N
LuBGAL14	scaffold3	Lus10020968.g		0	N
LuBGAL15	scaffold413	Lus10028348.g		4	N
LuBGAL16	scaffold272	Lus10041798.g		7	N
LuBGAL17	C8385757	Lus10000271.g		0	N
LuBGAL18	scaffold76	Lus10036109.g		0	N
LuBGAL19	scaffold915	Lus10016655.g		1	N
LuBGAL20	scaffold1120	Lus10003343.g		0	N
LuBGAL21	scaffold59	Lus10022645.g		3	N
LuBGAL22	scaffold305	Lus10025108.g		3	Y (8602)
LuBGAL23	scaffold305	Lus10025110.g		0	N
LuBGAL24	scaffold177	Lus10023977.g		6	N
LuBGAL25	scaffold177	Lus10023974.g		0	N
LuBGAL26	scaffold1982	Lus10005070.g		0	N
LuBGAL27	scaffold1143	Lus10027843.g		0	N
LuBGAL28	scaffold1247	Lus10014126.g		0	N
LuBGAL29	scaffold1982	Lus10005071.g		0	N
LuBGAL30	scaffold1143	Lus10027844.g		0	N
LuBGAL31	scaffold1247	Lus10014125.g		1	N
LuBGAL32	scaffold1491	Lus10019784.g		1	N
LuBGAL33	scaffold388	Lus10008259.g		0	Y (101+104+975)
LuBGAL34	scaffold711	Lus10020875.g		7	N
LuBGAL35	scaffold711	Lus10020877.g		1	N
LuBGAL36	scaffold701	Lus10033500.g		0	N
LuBGAL37	scaffold701	Lus10033502.g		0	N
LuBGAL38	scaffold112	Lus10018138.g		0	Y (16)
LuBGAL39	scaffold346	Lus10028538.g		0	N
LuBGAL40	scaffold488	Lus10033427.g		0	N
LuBGAL41	scaffold630	Lus10015616.g		6	N
LuBGAL42	scaffold196	Lus10037634.g		1	N
LuBGAL43	scaffold25	Lus10043422.g		0	N

[a]The number of mRNA and ESTs identified from the NCBI Genbank database and transcriptome assembly POZS (www.onekp.com).

Table 3-2 Summary of predicted glycosyl hydrolase 35 protein homologues.

BGAL Sub-family	LuBGAL	AA	MW[a] (kDA)	pI[a]	Signal Peptide[b] (Cleavage Site)	Pfam Domain[c]			Possible Destinations (WolfPSORT)[d]	Possible Destination (Plant-mPLOC)[e]
						GH35	Lectin	Copper oxidase		
D	41	761	84.697	9.07	No	Y	N	N	cl, v, n, cy, m, pm	pm, cy
	42	701	78.278	8.06	No	Y	N	N	cl, n, er, cy	cw, cy
C1	32	816	91.547	9.03	No	Y	Y	N	cy, px, m, , n	cw
	31	756	84.239	9.07	No	Y	Y	N	cy, n, px, v	cw
	29	843	94.393	8.38	Yes (34-35)	Y	Y	N	cl, ex, v, n	cw
	30	833	93.226	7.42	Yes (24-25)	Y	Y	N	cl, ex, v, n	cw
	28	828	93.86	8.92	Yes (24-25)	Y	Y	N	v, cl, er, g, m, p	cw
	27	788	89.565	9.69	Yes (22-23)	Y	N	N	v, ex, er, g, cl	cw
	26	752	85.192	9.7	Yes (25-26)	Y	N	N	v, g, cl, ex, er	cw
C2	40	821	92.792	8.68	Yes (19-20)	Y	Y	N	er, pm, n, m, ex	cw
	38	810	91.236	9.06	Yes (24-25)	Y	N	N	er, v, g, cl, n, cy, pm	cw
	39	871	98.135	8.94	Yes (23-24)	Y	Y	N	v, er, g, cl, n, cy, pm	cw
	33	829	91.265	5.96	Yes (30-31)	Y	Y	N	n, er, pm, cl, cy	cw
	37	718	80.437	5.58	Yes (22-23)	Y	N	N	v, ex, er, g, cl	cw
	34	961	108.198	5.48	Yes (23-24)	Y	N	N	v, g, er	cw
	35	647	71.944	8.88	No	Y	N	N	n, cl, cy	cw
	36	706	79.027	8.79	No	Y	N	N	v, er, g, cl, n	cw
A1	9	727	81.545	8.69	Yes (26-27)	Y	N	N	cl, ex, n, v, er, g	cw
	8	683	76.432	8.72	Yes (25-26)	Y	N	N	cl, ex, er, pm, m, cy, v	cw
	13	849	94.313	6.62	Yes (29-30)	Y	Y	N	v, cy, pm, cl, n, ex	cw
	14	229	25.653	8.58	No	Y	N	N	cl, n, cy	pm, cl
	12	650	72.077	7.12	Yes (28-29)	Y	Y	N	v, er, ex, g, cl, cy	cw, pm

	16	849	94.704	7.37	Yes (30-31)	Y	Y	N	er, pm, cy, cl, n, m, p	cw
	15	802	89.416	6.65	Yes (30-31)	Y	Y	N	er, pm, n, cl, cy, m, px	cw
	5	844	93.587	6.79	Yes (29-30)	Y	Y	N	cl, ex	cw
	6	869	95.928	9.2	No	Y	Y	N	cl, v, g, n, pm	cw
	7	851	94.066	9.13	Yes (24-25)	Y	Y	N	cl, ex	cw
	4	717	80.14	9.16	Yes (23-24)	Y	N	N	cl, n	cw
	3	723	80.594	8.95	Yes (23-24)	Y	N	N	cl, ex	cw
	1	731	80.978	6.74	Yes (29-30)	Y	N	N	cl, ex	cw
	2	740	81.923	6.59	Yes (29-30)	Y	N	N	cl, ex	cw
A4	11	897	100.599	6.38	Yes (24-25)	Y	N	N	pm, g	cw
	10	854	94.48	5.31	Yes (24-25)	Y	Y	N	v, pm, er, g, cl	cw
	18	297	32.849	7.62	No	Y	N	N	m, cy, n, cl, pm, v, er	cl
	17	836	91.017	8.14	No	Y	Y	N	cy, v, n, m, pm, cl	cw
B	43	107	11.805	7.57	Yes (31-32)	Y	N	N	ex, v, cl, cy, m, er	pm
A5	22	1460	162.474	5.41	Yes (19-20)	Y	Y	N	ex, v, cl, n, pm	cw
	24	1330	147.844	8.24	Yes (23-24)	Y	Y	Y (3)	v, cl, n, pm, m, ex	pm, cw
	21	871	96.999	8.57	Yes (26-27)	Y	Y	N	er, n, pm, g, cy	cw
	20	874	97.552	8.75	Yes (26-27)	Y	Y	N	ex, v, er, g, cl, n, cy	cw
	23	718	80.588	5.3	Yes (19-20)	Y	N	N	ex, cl, v, cy	cw
	25	261	29.969	8.26	No	Y	N	N	n, cy, cl	cl
A2	19	880	98.216	6.52	Yes (27-28)	Y	Y	N	cl, v, g, pm	cw

[a]Predictions made with CLC Genomics Workbench 5.5.
[b]SignalP 4.0 prediction (Petersen et al., 2011).
[c]Pfam domains and locations were identified with CLC Genomics Workbench 5.5.
[d]WolfPSORT prediction (Horton et al., 2007), in order of decreasing likelihood.
[e]Plant-mPLOC prediction (Chou and Shen, 2010)
Protein Destinations: cl (chloroplast), cy (cytosol), cs (cytoskeleton) cw (cell wall), er (endoplasmic reticulum), ex (extracellular), g (Golgi apparatus), l (lysosome), m (mitochondria), n (nuclear), px (peroxisome), pm (plasma membrane), v (vacuolar membrane)

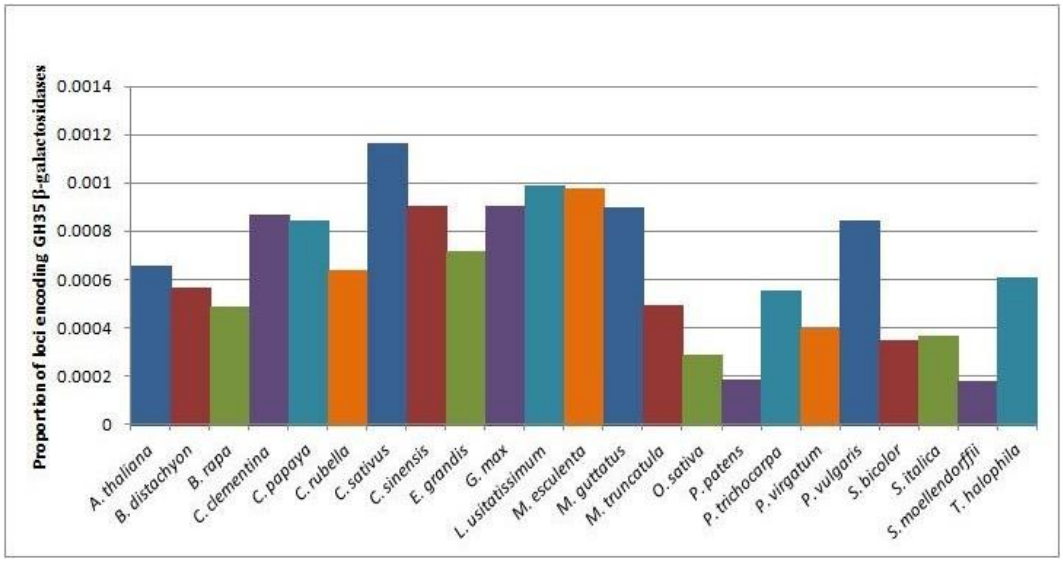

Figure 3-1 Relative quantity of BGAL genes in the genomes of various plant species. Predicted proteomes for *Arabidopsis thaliana*, *Brachypodium distachyon*, *Brassica rapa*, *Citrus clementina*, *Carica papaya*, *Capsella rubella*, *Cucumis sativus*, *Citrus sinensis*, *Eucalyptus grandis*, *Glycine max*, *Linum usitatissimum*, *Manihot esculenta*, *Mimulus guttatus*, *Medicago truncatula*, *Oriza sativa*, *Physcomitrella patens*, *Populus trichocarpa*, *Panicum virgatum*, *Phaseolus vulgaris*, *Sorghum bicolor*, *Setaria italica*, *Selaginella moellendorffii*, and *Thellungiella halophila* were obtained from Phytozome (version 8.0; www.phytozome.net). Sequences were assessed via Hidden Markov Model (HMM) with HMMER3 (https://hmmer.janelia.org), using the Pfam-A family database (version 25.0; Punta et al., 2012), for genes putatively encoding a glycosyl hydrolase 35 domain. The number of putative BGAL genes was compared to the total number of protein coding loci published for each species at Phytozome (version 8.0; www.phytozome.net).

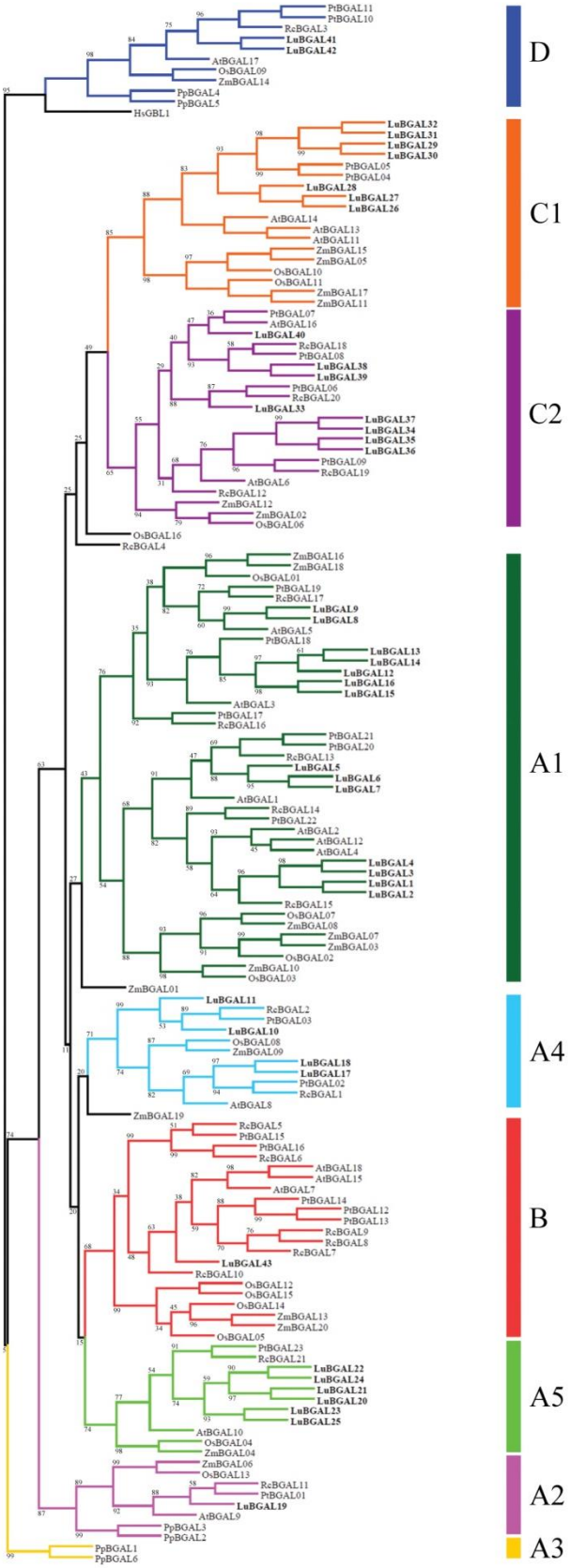

Figure 3-2 Phylogenetic relationship among the glycosyl hydrolase 35 proteins of flax and other species. Deduced amino acid sequences were aligned with MUSCLE (Edgar, 2004). The tree was created with GARLI (Zwickl, 2006), using the maximum likelihood method, following the WAG model of amino acid substitutions (Whelan and Goldman, 2001). A consensus tree of 1000 bootstrap replicates was produced for which percent reproducibility under 100 is shown. The flax sequences are named LuBGAL, and numbered according to Tables 3-1 and 3-2. *Arabidopsis thaliana* sequences are indicated as AtBGAL, and numbered according to existing designations (Ahn et al., 2007). *Oryza sativa* sequences are indicated as OsBGAL, and numbered according to existing designations (Tanthanuch et al., 2008). *Physcomitrella patens* sequences are indicated as PpBGAL, *Populus trichocarpa* sequences are indicated as PtBGAL, and *Ricinus communis* sequences are indicated as RcBGAL. Genomic loci corresponding to these sequences are presented in Table S3. A human beta-galactosidase (GLB1; NP_000395) was used to establish the outgroup.

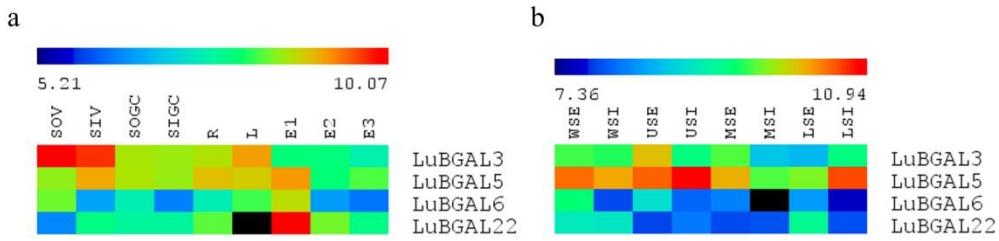

Figure 3-3 Transcript abundance of flax BGAL genes in various tissues, from previously published microarray data sets (Nimblegen platform). RMA-normalized, average log2 signal values of flax BGALs in various tissues were used to produce a heat map. **a**: roots (R); leaves (L); outer stem tissues at either the vegetative stage (SOV) or green capsule stage (SOGC); inner stem tissues at either vegetative stage (SIV) or green capsule stage (SIGC); and seeds 10-15 days after flowering (DAF; E1), 20-30 DAF (E2), and 40-50 DAF (E3; Fénart et al., 2010). **b**: internal stem tissues of either the whole stem (WSI), upper stem (USI), middle stem (MSI), or lower stem (LSI); and external (i.e. phloem and cortex enriched) stem tissues of the whole stem (WSE), upper stem (USE), middle stem (MSE), and lower stem (LSE; Huis et al., 2012).

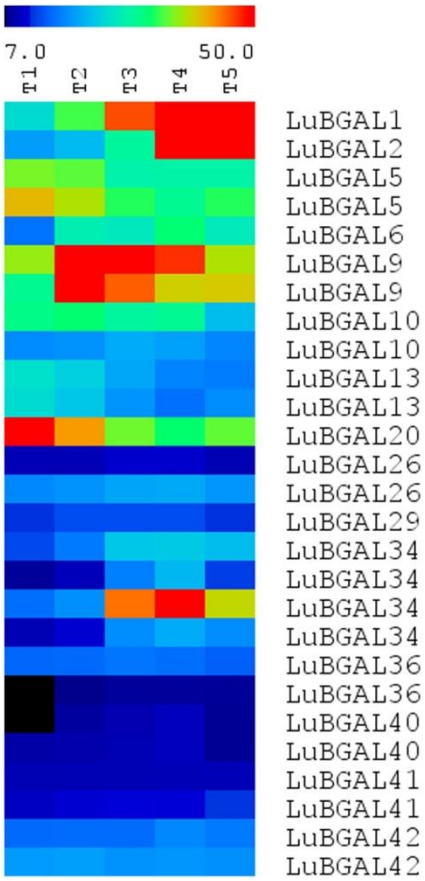

Figure 3-4 Transcript abundance of flax BGAL genes throughout the stem, from unpublished microarray data set (Combimatrix platform). Signal intensities were normalized as fractions of mean signal strength. The log2 signal values of the various flax BGALs were used to produce a heat map. Microarray data examined the shoot apex (T1), the snap-point through various stages of fibre development (T2, T2, and T4), and a lower portion of the stem (T5).

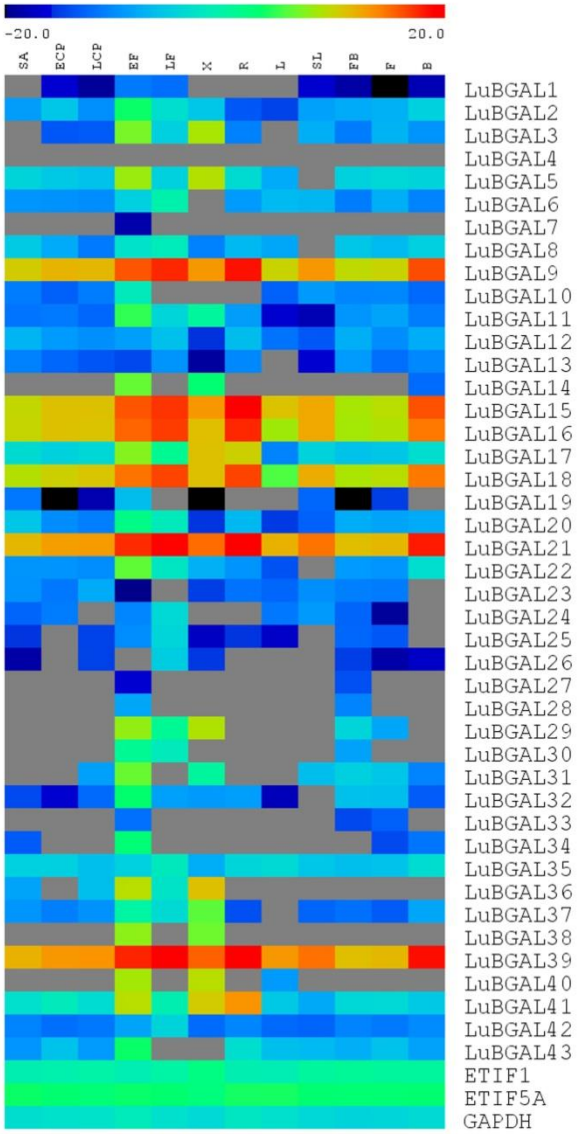

Figure 3-5 Transcript abundance of flax BGAL genes in various tissues, by qRT-PCR (Fluidigm platform). Expression levels (log_2), relative to the reference genes *ETIF1* (eukaryotic translation initiation factor 1), *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase), and *ETIF5A* (eukaryotic translation initiation factor 5A), were used to prepare a heat map, with blue indicating lower expression and red indicating high expression. Gray indicates no detectable expression. Tissue types analysed include: roots (R); leaves (L); senescing leaves (SL); stem apex (SA); cortical peels from vegetative stage stems (ECP) or green capsule stage stems (LCP); phloem fibres from vegetative stage stems (EF) or green capsule stage stems (LF); xylem from vegetative stage stems (X); budding flowers (FB); open flowers (F); and seed bolls from the green capsule stage (B).

References

Ahn YO, Zheng M, Bevan DR, Esen A, Shiu S-H, Benson J, Peng HP, Miller JT, Cheng CL, Poulton JE, and Shih MC (2007) Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35. Phytochemistry **68**: 1510–1520. doi:10.1016/j.phytochem.2007.03.021

Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, and Gupta VS (2012) Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varied expression patterns. BMC Genomics **13**:175.

Bazinet AL and Cummings MP (2011) Computing the Tree of Life - Leveraging the Power of Desktop and Service Grids. Proceedings of the Fifth Workshop on Desktop Grids and Volunteer Computing Systems, Anchorage, May 20.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, and Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research **37**: D233-238.

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, Rabinowicz PD (2010) Draft genome sequence of the oilseed species Ricinus communis. Nature Biotechnology **28**: 951-956.

Chou KC and Shen HB (2010) Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. PLoS ONE **5**: e11335. doi:10.1371/journal.pone.0011335

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics **27**:1164-1165

Dean GH, Zheng H, Tewari J, Huang J, Young DS, Hwang YT, Western TL, Carpita NC, McCann MC, Mansfield SD, and Haughn GW (2007) The *Arabidopsis MUM2* gene encodes a β-galactosidase required for the production of seed coat mucilage with correct hydration properties. The Plant Cell **19:** 4007–4021. doi:10.1105/tpc.107.050609

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**: 1792-1797. doi: 10.1093/nar/gkh340

Fénart S, Ndong Y-PA, Duarte J, Rivière N, Wilmer J, van Wuytswinkel O, Lucau A, Cariou E, Neutelings G, Gutierrez L, Chabbert B, Guillot X, Tavernier R, Hawkins S, and Thomasset B (2010) Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray. BMC Genomics **11**: 592. doi: 10.1186/1471-2164-11-592

Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme. II. Berichte der Deutschen Chemischen Gesellschaft **27**: 3479-3483.

Gantulga D, Turan Y, Bevan DR, and Esen A (2008) The *Arabidopsis* At1g45130 and At3g52840 genes encode β-galactosidases with activity toward cell wall polysaccharides. Phytochemistry **69**: 1661-1670.

Gantulga D, Ahn YO, Zhou C, Battogtokh D, Bevan DR, Winkel BSJ, and Esen A (2009) Comparative characterization of the *Arabidopsis* subfamily a1 β-galactosidases. Phytochemistry **70:** 1999–2009. doi:10.1016/j.phytochem.2009.08.008

Gorshkova TA, Sal'nikova VV, Chemikosova SB, Ageeva MV, Pavlencheva NV, and van Dam JEG (2003) The snap point: a transition point in *Linum usitatissimum* bast fiber development. Industrial Crops and Products **18**: 213-221.

Gorshkova T and Morvan C (2006) Secondary cell-wall assembly in flax phloem fibres: role of galactans. Planta **223**: 149-158.

Henrissat B (1998) Glycosidase families. Biochemical Society Transactions **26**:153-156.

Hobson N and Deyholos MK (2013) LuFLA1$_{PRO}$ and LuBGAL1$_{PRO}$ promote gene expression in the phloem fibres of flax (*Linum usitatissimum*). Plant Cell Reports **32:** 517-528. doi: 10.1007/s00299-013-1383-8

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, and Nakai K (2007) WoLF PSORT: Protein Localization Predictor. Nucleic Acids Research **35**: W585-W587. doi:10.1093/nar/gkm259.

Hrubá P, Honys D, Twell D, Capková V, and Tupý J (2005) Expression of β-galactosidase and β-xylosidase genes during microspore and pollen development. Planta **220:** 931–940. doi:10.1007/s00425-004-1409-0

Huis R, Hawkins S, and Neutelings G (2010) Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). BMC Plant Biology **10**: 71. doi:10.1186/1471-2229-10-71

Huis R, Morreel K, Fliniaux O, Lucau-danila A, Fénart S, Grec S, Neutelings G, Chabbert B, Mesnard F, Boerjan W, and Hawkins S (2012) Natural Hypolignification Is Associated with Extensive Oligolignol Accumulation in Flax Stems. Plant Physiology **158**: 1893–1915. doi:10.1104/pp.111.192328

Jacob F, and Monod J (1961) Genetic Regulatory Mechanisms in the Synthesis of Proteins. Journal of Molecular Biology **3**: 318-356.

Johnson MTJ, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, dePamphilis CW, Edger PP, Goh F, Graham S, Greiner S, Hibberd JM, Jordon-Thaden I, Kutchan TM, Leebens-Mack J, Melkonian M, Miles N, Myburg H, Patterson J, Pires JC, Ralph P, Rolf M, Sage RF, Soltis D, Soltis P, Stevenson D, Stewart CN, Surek B, Thomsen CJM, Villarreal JC, Wu X, Zhang Y, Deyholos MK, and Wong GKS (2012) Evaluating Methods for Isolating Total RNA and Predicting the Success of Sequencing Phylogenetically Diverse Plant Transcriptomes. PLoS ONE **7**: e50226.

Kishore D, and Kayastha AM (2012) A β-galactosidase from chick pea (*Cicer arietinum*) seeds: Its purification, biochemical properties and industrial applications. Food Chemistry **134**: 1113-1122. doi: 10.1016/j.foodchem.2012.03.032

Kotake T, Dina S, Konishi T, Kaneko S, Igarashi K, Samejima M, Watanabe Y, Kimura K, and Tsumuraya Y (2005) Molecular Cloning of a β-Galactosidase from Radish That Specifically Hydrolyzes β-(1→3)- and β-(1→6)-Galactosyl Residues of Arabinogalactan Protein. Plant Physiology **138**: 1563-1576. doi: 10.1104/pp.105.062562

Kuby SA, and Lardy HA (1953) Purification and Kinetics of β-D-Galacosidase from *Escherichia coli*, Strain K-12. Journal of The American Chemical Society **75**: 890-896.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, and Huala E (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research **40**: D1202-1210. doi:10.1093/nar/gkr1090

Lazan H, Ng S-Y, Goh L-Y, and Ali Z M (2004) Papaya beta-galactosidase/galactanase isoforms in differential cell wall hydrolysis and fruit softening during ripening. Plant Physiology and Biochemistry **42**: 847–853. doi:10.1016/j.plaphy.2004.10.007

Miller MA, Pfeiffer W, and Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), New Orleans, November 14. doi: 10.1109/GCE.2010.5676128

Mokshina NE, Ibragimova NN, Salnikov VV, Amenitskii SI, and Gorshkova TA (2012) Galactosidase of plant fibers with gelatinous cell wall: Identification and localization. Russian Journal of Plant Physiology **59**: 246–254. doi:10.1134/S1021443712020082

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Research **35**: D883-887.

Ozeki Y, Matsui T, Suzuki M, and Titani K (1991) Amino Acid Sequence and Molecular Characterization of a D-Galactoside-Specific Lectin Purified from Sea Urchin. Biochemistry **30**: 2391–2394.

Petersen TN, Brunak S, von Heijne G, and Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods **8**: 785-786. doi: 10.1038/nmeth.1701

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, and Finn RD (2012) The Pfam protein families database. Nucleic Acids Research **40**: D290–301. doi:10.1093/nar/gkr1065

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science **319**: 64-69.

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, and Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant Physiology **156**: 1351–1363. doi:10.1104/pp.111.172676

Sampedro J, Gianzo C, Iglesias N, Guitián E, Revilla G, and Zarra I (2012) AtBGAL10 is the main xyloglucan β-galactosidase in Arabidopsis, and its absence results in unusual xyloglucan subunits and growth defects. Plant Physiology **158**: 1146–57. doi:10.1104/pp.111.192195

Schmittgen TD and Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nature Protocols **3**: 1101–1108.

Smith DL, Starrett DA, and Gross KC (1998) A Gene Coding for Tomato Fruit β-Galactosidase II Is expressed during Fruit Ripening. Plant Physiology **117**:417-423. doi: 10.1104/pp.117.2.417

Smith DL, and Gross KC (2000) A Family of at Least Seven β-Galactosidase Genes Is Expressed during Tomato Fruit Development. Plant Physiology **123**:1173-1184. doi: 10.1104/pp.123.3.1173

Smith DL, Abbott JA, and Gross KC (2002) Down-Regulation of Tomato β-Galactosidase 4 Results in Decreased Fruit Softening. Plant Physiology **129**: 1755–1762. doi:10.1104/pp.011025

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbarugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Research **36:** D1009-1014

Tanthanuch W, Chantarangsee M, Maneesan J, and Ketudat-Cairns J (2008) Genomic and expression analysis of glycosyl hydrolase family 35 genes from rice (*Oryza sativa* L.). BMC Plant Biology **8:** 84. doi:10.1186/1471-2229-8-84

Trainotti L, Spinello R, Piovan A, Spolaore S, and Casadoro G (2001) β-Galactosidases with a lectin-like domain are expressed in strawberry. Journal of Experimental Botany **52**:1635-1645. doi: 10.1093/jexbot/52.361.1635

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**: 1596-1604

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, and Datla R (2011) Gene expression analysis of flax seed development. BMC Plant Biology **11**:74. doi: 10.1186/1471-2229-11-74

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe A, Wong GK.-S, Wang J, and Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. The Plant Journal **72**: 461-473. doi:10.1111/j.1365-313X.2012.05093.x

Whelan S and Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution **18**: 691–699.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis. The University of Texas: USA.

Chapter 4 – Characterization of beta-galactosidase (*LuBGAL1*) overexpressing flax plants

Introduction

Flax (*Linum usitatissimum*) is a multipurpose crop, cultivated either for its seed or the phloem fibres of the stem, depending upon the cultivar grown. The long phloem fibre cells have a high tensile strength, being rich in crystalline cellulose, and are used to produce textiles, such as linens, as well as a variety of high quality papers and composite materials (Mohanty et al., 2000). Unfortunately for countries such as Canada, with cold dry climates and a lack of processing facilities impeding fibre extraction, flax phloem fibres are not valued as highly, and in fact are often considered a detriment to the cultivation of flax seed due to the strength of the stems and the ways in which they can become entangled in harvesting machinery.

During the course of their development, flax phloem fibres enter a stage of secondary cell wall deposition, which starts at a point in the stem called the "snap-point", where thickening cell walls confer additional mechanical strength to the stem, detectable by a sharp resistance of the stem to bending and tearing (Gorshkova et al., 2003). As they mature, the cell walls of flax phloem fibres display two distinct layers. Initially, the secondary cell wall is deposited as a Gn-layer, with cellulose microfibrils separated by large galactan rich pectic polysaccharides, which are themselves injected into the cell wall by the Golgi (Gorshkova et al., 2006; Salnikov et al., 2008). As development continues, this pectic polysaccharide is degraded by β-galactosidases (EC 3.2.1.23), ostensibly LuBGAL1, which leads to the formation of a near homogenous layer of crystalline cellulose oriented almost parallel to the longitudinal axis of the cell, called the G-layer (Gurjanov et al., 2007; Gurjanov et al., 2008; Gorshkova et al., 2010; Roach et al., 2011; Mikshina et al., 2012).

LuBGAL1 is a member of the glycosyl hydrolase 35 (GH35) family of flax, whose transcripts and proteins have been found to be enriched in tissues containing phloem fibres, and whose promoter has been observed to drive expression almost exclusively in the phloem fibres of the stem (Roach and Deyholos, 2007; Hotte and Deyholos, 2008; Hobson and Deyholos, 2013). An earlier study examining the role of LuBGAL1 in the degradation of the galactan-rich

polysaccharides of the flax phloem fibre utilized gene-silencing to reduce its activity in developing cell walls. This prevented the degradation of the pectic polysaccharide, and resulted in phloem fibres with swollen cell walls that were permeated with galactan rich polysaccharides (Roach et al., 2011). The lack of galactan degradation resulted in the reduction in the levels of crystalline cellulose, as well as a weakening of the flax stem, traits which may be desirable in flax cultivated solely for seeds.

The purpose of the current study was to directly assess the impact of increased LuBGAL1 activity on the cell wall architecture of flax phloem fibres. As the process of galactan deposition and degradation has been posited to influence the alignment of cellulose microfibrils alongside one another, aiding in the crystallization of cellulose (Gorshkova et al., 2010; Roach et al., 2011), we hypothesized that a decrease in galactan deposition, occurring early in development and in advance of the nominal expression of *LuBGAL1* (Hobson and Deyholos., 2013), would impede the alignment of cellulose microfibrils and crystallization, thus weakening the stem without producing a carbon-sink in the form of undigested galactans. This study also allowed us the opportunity to attempt to purify LuBGAL1 and determine its substrate specificity and enzymatic activity.

Materials and Methods

Plasmid Construction

A fosmid containing the genomic copy of LuBGAL1 (Genbank ID: HQ902252) was submitted to the Augustus v2.1 server (Stanke et al., 2008). Primers flanking the predicted CDS of *LuBGAL1*, containing an artificial ribosomal binding site and the attB sites required for Gateway cloning (Invitrogen Gateway Technology with Clonase II), were designed (BgalattB1rbs = 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACCATGTTGAGCCACAGCAGCCTGG-3'; BgalattB2 = 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTTTTAGGTAAGCACCCCCTGT-3'), and used to amplify the cDNA before Gateway cloning into pEarleyGate100 by way of pDonR201(Earley et al., 2006). Constructs were transformed into *Agrobacterium tumefaciens* GV3101.

Tissue Culture

Plant transformations of flax variety CDC Bethune were conducted using an adaptation of a published protocol (Mlynárová et al. 1994; Wróbel-Kwiatkowska et al. 2007). Flax seedlings were grown for 6 d on half-strength Linsmaier and Skoog (LS) + 1% sucrose + 0.7% phytablend agar plates. Cut hypocotyl segments were inoculated with Agrobacterium, for 2 h with agitation, in 20 mL co-cultivation media containing LS, 3% Sucrose, 1 mg/L benzyladenine (BA), 0.1 mg/L naphthyl-acetic acid (NAA), and 20 mM acetosyringone. Hypocotyls were transferred to co-cultivation media containing LS, 3% sucrose, 1 mg/L BA, 0.1 mg/L NAA, 100 mM acetosyringone, and 0.7% agar for 3 d. Callus formation and transformant selection was achieved by maintaining explants for 2-3 weeks on shoot initiation/selection media containing LS, 3% Sucrose, 1 mg/L BA, 0.1 mg/L NAA, 300 mg/L Timentin, 7.5 mg/L DL-Phosphinothricin (PPT), and 0.7% agar. Calli were excised from the infected hypocotyl ends, and maintained on selective shoot regeneration media containing LS, 2.5% Sucrose, 0.02 mg/L BA, 0.001 mg/L NAA, 300 mg/L Timentin, 7.5 mg/ml PPT, and 0.7% agar. Calli were transferred to fresh shoot regeneration media every 2 weeks. Developing shoots were excised from the calli and transferred to shoot elongation media containing LS, 1% Sucrose, 100 mg/L Timentin, 7.5 mg/L PPT, and 0.7% agar. After 2 weeks, shoots were transferred to rooting media containing half-strength LS, 1% Sucrose, 0.2 mg/L indole-3-butyric acid (IBA), 100 mg/L Timentin, 7.5 mg/L PPT, and 0.7% agar. T_0 transgenics that developed roots were transferred to soil and grown to maturity. T_1 plants were screened via PCR for the *bar* selectable marker. T_2 plants were again screen via PCR for the *bar* selectable marker, as were T_3 plants. We confirmed the presence of the CaMV35S:*LuBGAL1* construct in three independent transformation events.

qRT-PCR

Tissue samples from *Linum usitatissimum* (CDC Bethune) were frozen in liquid nitrogen, and stored at -80°C prior to use. Frozen samples were ground in liquid nitrogen, whereupon 1.5 ml Trizol was added per mg of tissue. Samples were heated at 60°C for 5 min, and centrifuged at 12,000 g for 10 min at 4°C. Supernatant was collected and mixed with 1/5V chloroform, mixed,

allowed to settle for 3 min, and centrifuged at 12,000 g for 20 min at 4°C. The aqueous phase as mixed with 0.25V RLT buffer (Qiagen RNeasy mini kit), and then applied to the RNeasy Mini spin column. An on-column DNase digestion was performed as per kit instructions, with continued washes and elutions as per kit instructions. cDNA was prepared with RevertAid H Minus Reverse Transcriptase (Fermentas) and oligo$(dT)_{18}$ primer. Real-time PCR was performed with an Applied Biosystems 7500 Fast Real-Time PCR System. cDNA derived from 1 µg was diluted to 1/16, 2.5 µl of which was added to a reaction volume of 10 µl containing 0.4 µM of forward and reverse primers (LuBgal1_qPCR_F = 5' – AGAACGAGTACGGGCCGATA – 3'; LuBgal1_qPCR_R = 5' – CCATCTGAGCTGCCCACTGT - 3'), 0.2 mM dNTPs, 0.25X SYBR Green, 1X ROX, and 0.075 U Platinum Taq. Threshold cycles (C_T) were determined by the 7500 Fast Software. $\Delta\Delta C_T$ values were calculated in reference to ETIF1 (eukaryotic translation initiation factor 1; Huis et al., 2010) and wildtype gene expression.

BGAL Activity

Frozen tissue samples were ground in liquid nitrogen, and then 1 ml 50 mM $NaPO_4$ pH 7.2. Phenylmethanesulfonylfluoride (PMSF) was added to the supernatant for a final concentration of 1 mM. 10 µl of protein extract was added to 90 µl ONPG buffer, consisting of 50 mM sodium citrate pH 7.0, 1 mM $MgCl_2$ 5 mM β-mercaptoethanol, and 1 mg/ml ortho-Nitrophenyl-β-galactoside (ONPG), then incubated at 37ºC for 30 minutes. Reactions were stopped by adding 167 µl 1M Na_2CO_3. Hydrolysis was detected by measuring the absorbance of p-nitrophenol at 420 nm. Protein concentrations were determined via Qubit flurometer (Invitrogen). A single factor analysis of variance (ANOVA) was employed to determine if the values observed differed significantly between constructs.

Tensile Strength

Measurements were performed on mature, dry stems, using an Instron 5565. 15 cm stem fragments, located 13 cm above the cotyledon, were used. Instron grips were set 75 mm apart.

Tension was applied at 30 mm/s. Young's modulus (E) was calculated with the following formula, where F is the maximum force (*N*) applied to stems before breakage, A is the area of the stem (mm), ΔL is the change in length due to stretching (mm), and L is the original distance between the two grips (75 mm): $E = \frac{tensilestress}{tensilestrain} = \frac{\frac{F}{A}}{\frac{\Delta L}{L}}$

GC-MS

Cortical peels were harvested from 5-week old flax, starting at the snap-point and extending 16 cm down. Samples were lyophilized, and ground for 10 min at 25 Hz with two 5 mm stainless steel ball bearing and a Retsch mixer mill (MM301). Cell wall isolation and monosaccharide derivitization were performed as described elsewhere (Foster et al., 2010). 1 µl samples were injected into an Agilent 5975 GC/MS with a HP-5MS column (30 m x 0.25 mm x 0.25 µm), running the following program: 2 min at 100°C, 10°C/min until 200°C, 5 min at 200°C, 10°C/min until 250°C, and a hold at 250°C for 10 min. Compounds were identified by comparison to the retention times of individual monosaccharide standards.

Heterologous cloning

LuBGAL1 was first cloned into pPICZA for intracellular expression in *Pichia pastoris* (Invitrogen EasySelect Pichia Expression Kit). The coding sequence, minus the stop codon, was amplified with restriction tagged primers pPICZ_luBgal_EcoRI_F (5'-TCCTATCAGAATTCGTAATGTTGAGCTGAGCCACAGCAGCC-3') and pPICZ_luBgal_NotI_R (5'-TCCTATCAGCGGCCGCACATTCCCGCCCGGCAAAGTTCTTTTC-3'), digested with EcoRI and NotI, and ligated into EcoRI and NotI digested pPICZA. The construct, and parent vector pPICZA, were electroporated into *P. pastoris* GS115, and pure clonal isolates were obtained according to kit instructions. Transformants were maintained on MDH plates (1.34% yeast nitrogen base without amino acids, 4 X 10^{-5}% biotin, 4 X 10^{-3}% histidine, 2% dextrose, and 2% agar). For induction, cultures were grown in MDH liquid medium before induction was

performed in MMH liquid medium (1.34% yeast nitrogen base without amino acids, 4 X 10^{-5}% biotin, 4 X 10^{-3}% histidine, 1% methanol). For functional screens, induction was performed on MMH plates (MMH liquid medium with 2% agar).

Functional screens were performed as described elsewhere (Dean et al., 2007), by transferring yeast colonies onto nitrocellulose membranes, and immersing the membrane in liquid nitrogen before incubating at 37°C atop a buffer of 50 mM sodium acetate pH 5.0, containing 2 mM 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal).

Protein samples, from cell pellets and culture supernatant, were prepared according to the instructions provided in the EasySelect Pichia Expression Kit (Invitrogen) and the ProBond Purification System (Invitrogen) for protein purifications under native conditions.

LuBGAL1 was next cloned into pPICZαA for extracellular expression in *Pichia pastoris* (Invitrogen EasySelect Pichia Expression Kit). The cDNA was amplified with restriction tagged primers (pPICZaA_bgal_fe = 5'-TCCTATCAGAATTCATGTTGAGCCACAGCAGCCTGG-3'; pPICZaA_bgal_rx = 5'-TCCTATCAGCGGCCGCCATTCCCGGCAAAGTTCTTTTC-3'), digested with EcoRI and XbaI, and ligated into EcoRI and XbaI digested pPICZαA. The construct was electroporated into *P.pastoris* GS115, per kit instructions.

LuBGAL1 was last cloned into pET22b+ for heterologous expression in *Escherichia coli*. The cDNA was amplified with restriction tagged primers (pET22b_Bgal_Feco = 5'-TCCTATCAGAATTCAATGTTGAGCCACAGCAGCCTGG-3'; pET22b_Bgal_Rnot = 5'-TCCTATCAGCGGCCGCCATTCCCGGCAAAGTTCTTTTC-3'), digested with EcoRI and NotI, and ligated into EcoRI and NotI digested pET22b+. The construct was chemically transformed into *E. coli* Rosetta™ (Novagen).

Results

To test the hypothesis that increased *LuBGAL1* expression would disrupt gelatinous fiber development in flax, we produced transgenic flax in which *LuBGAL1* was transcribed from the CaMVS35S promoter. Three independent lines (Br1, Br2, and Br3) were obtained. Transgenic lines were propagated until T_3 generation plants could be obtained for detailed analysis. To

quantify the increase in *LuBGAL1* transcript expression, qRT-PCR was performed, comparing *LuBGAL1* transcript abundance in the leaves of transgenic lines to leaves of non-transformed CDC Bethune (Figure 4-1). Only two lines, Br1 and Br2, showed a clear increase in transcript levels, 57 and 44 fold greater, respectively, than that observed in wild-type Bethune. Br3 did not display any significant increase in *LuBGAL1* transcription. Further analyses focused solely on Br1 and Br2.

To determine whether increased *LuBGAL1* transcription in flax tissues contributed to an increase in LuBGAL1 activity, we performed an enzymatic assay that measured the rate of ONPG hydrolysis by crude protein extracts from the cortical peel of the flax stem (Figure 4-2). An analysis of variance (ANOVA) revealed no significant difference was present between the transgenic and non-transgenic flax (p-value = 0.198).

Because LuBGAL1 has been inferred to degrade galactan-rich pectins of the cell walls of fibres (Roach et al. 2011), we wanted to know whether its over-expression would impact the development of cell walls in other tissues, and we therefore examined plant heights to determine whether cell extensibility and/or division had been affected. No change in plant height was observed (Figure 4-3), nor did transgenic lines Br1 and Br2 show any obvious morphological or developmental phenotype.

To determine whether the over-expression of *LuBGAL1* would impact the development of the cell wall of phloem fibres, we first examined the tensile strength of the plant stems (Figure 4-4). No difference between transgenic and wild-type flax stems was observed. We then examined transverse sections of flax stems. In most cases, no obvious phenotype could be seen, and transgenic fibres appeared identical to wild-type fibres (Figure 4-5). However, in a few cases, individual fibres did display a phenotype similar to that observed in the transgenic *LuBGAL1* RNAi lines (Figure 4-6), where the knockdown of *LuBGAL1* resulted in phloem fibres lacking a crystalline cellulose-rich outer G-layer of the cell wall and instead were composed solely of a Gn-layer comprising cellulose microfibrils spaced by thick bands of a galactan-rich pectic polysaccharide (Roach et al., 2011). These RNAi-like fibres appeared sporadically in both Br1 and Br2 lines. Further examination of the stem sections revealed no obvious morphological defects.

While a visual inspection of transgenic flax phloem fibres generally revealed no apparent changes to the cell walls, we chose to examine the monosaccharide content of the cell walls to determine if any change in chemistry occurred. After hydrolyzing and derivitizing the non-cellulosic cell wall polysaccharides of stem cortical peels, we measured the quantity of arabinose, fucose, galactose, glucose, mannose, rhamnose, and xylose (Figure 4-7). While a slight decrease in galactose was observed in Br1, and an increase in galactose was observed in Br2, ANOVA analyses revealed no significant difference between the different lines when comparing the levels of arabinose (p-value = 0.934), fucose (p-value = 0.538), galactose (p-value = 0.068), glucose (p-value = 0.882), mannose (p-value = 0.138), rhamnose (p-value = 0.298), and xylose (p-value = 0.946).

We attempted to purify the LuBGAL1 protein from a heterologous system to determine its substrate specificity, kinetics, and overall function. While we were capable of obtaining transcript expression (data not shown), as well as a faint blue precipitate during the X-Gal functional screen, we were unable to obtain any BGAL activity from protein purification extracts, nor was our protein detectable by mass spectrometry. In fact, every attempt at protein purification, by ourselves and collaborators, in *Pichia pastoris*, *Escherichia coli*, and *Saccharomyces cereviseae*, failed.

Further analyses of the cell walls of the transgenic flax are being performed by collaborators at the University of British Columbia, examining cellulose crystallinity and microfibrillar angle. Results are pending.

Discussion

A characteristic of flax phloem fibre development is the deposition, and later degradation, of a large galactan-rich pectic polysaccharide in the inner Gn-layer of the secondary cell wall. *LuBGAL1* has been described as an important facilitator of this galactan degradation, with studies highlighting its enriched expression in phloem fibres during secondary cell wall deposition (Hobson and Deyholos, 2013), as well as the characterization of transgenic *LuBGAL1* RNAi lines with reduced G-layers and expanded Gn-layers in the phloem fibres (Roach et al., 2011). While it

has been posited that the temporary Gn-layer aids in the organization of cellulose microfibrils (Gorshkova et al., 2010; Roach et al., 2011), aiding in their crystallization and alignment nearly parallel to the longitudinal axis of the fibre cell (Deyholos, 2006), we currently lack experimental evidence supporting this hypothesis.

In an effort to gauge the importance of the Gn-layer on cell wall development in flax phloem fibres, we attempted to increase the expression of *LuBGAL1* and subsequently initiate the degradation of the galactan-rich pectin earlier in secondary cell wall deposition, ideally impeding the formation of a distinct Gn-layer. However, our two transgenic lines overexpressing the *LuBGAL1* transcript did not display any increased BGAL activity (Figure 4-2).

There are a number of reasons why an increase in enzymatic activity may not have been observed despite a >44 fold increase in transcript abundance, including: 1) proteolysis of the expressed protein; 2) barriers to ribosome binding; 3) mRNA instability from co-suppression; 4) mutations in the insert gene introducing stop codons, introducing frameshifts, or substituting amino acids to alter protein folding and/or active sites. We have experimentally excluded 4) by sequencing the *LuBGAL1* insert in the construct. While we recorded a 57 and 44 fold increase in transcript abundance in the leaves of the transgenic lines Br1 and Br2, respectively, we cannot rule out 3) as we have observed random phloem fibres sharing the phenotype of *LuBGAL1*-RNAi phloem fibres (Figure 6). *LuBGAL1* expression has been reported primarily in the phloem fibres of the stem (Hobson and Deyholos, 2013; Hobson and Deyholos, Forthcoming), thus it is possible that co-suppression was unique to the fibres, where endogenous transcripts in select fibres raised expression levels past a co-suppression threshold (Que et al., 1997). Unfortunately, we also cannot exclude factor 2), as we had incorporated an artificial ribosome binding site which may not have been optimum for flax, impeding translation initiated or inducing RNA folding that hindered ribosome movement. Nor can we exclude 1), although we consider this unlikely as the peptide sequence should be identical to the endogenous LuBGAL1, and thus should not be more susceptible to proteolysis than the endogenous protein.

In addition to factors affecting the expression and translation of an introduced gene, we must also consider the appropriateness of the tests employed to measure LuBGAL1 activity.

Previous examinations of BGAL activity in flax stems utilized either ONPG or X-Gal as chromogenic galactoside substrates (Roach and Deyholos 2008). However, these assays, and ours, used crude protein extracts to measure BGAL activity, which likely was produced by any number of the 43 different LuBGALs expressed in the flax stem (Hobson and Deyholos, in press). ONPG and X-GAL are both galactosides and are normally used as analogs of lactose in detecting BGAL activity. However, ONPG and X-Gal may in fact be unsuitable substrates for LuBGAL1. Plant BGALs show a wide range of substrate specificity (Kotake et al., 2005; Ahn et al., 2007; Dean et al., 2007; Gantulga et al., 2008; Tantanuch et al., 2008; Sampedro et al., 2012), therefore future analyses of BGAL activity should look at a wider range of galactoside substrates.

We attempted to obtain purified LuBGAL1 protein from heterologous systems, however no BGAL activity was observed in purified protein preparations, nor were we able to detect any peptides matching LuBGAL1 when we performed peptide fingerprinting on proteins matching the predicted size of LuBGAL1 (80.9 kDA; Table 3-2). In general, the inability to purify a heterologously expressed protein can be attributed to a number of factors. 1) proteolytic degradation of the expressed protein; 2) a lack of gene expression; 3) mutations in the coding region altering the reading frame or introducing stop codons; 4) codon bias impeding protein translation; 5) mRNA instability and increased nucleolytic degradation (Raué, 1994); 6) incorrect localization, aggregating in an insoluble fraction or exporting to the medium. We have experimentally excluded 2) and 5) by amplifying the *LuBGAL1* cDNA from the heterologous system (data not shown), which indicated both that it is expressed and undigested. We have also excluded 3) by sequencing the *LuBGAL1* insert in the expression constructs, which indicated that no stop codons had been introduced, nor had the reading frame shifted, and have excluded 6) by examining insoluble protein fractions and media for the presence of a 6X His-tagged LuBGAL1. As *LuBGAL1* was cloned directly from flax, we cannot rule out 4) codon bias. We also cannot rule out 1) proteolysis of a translated LuBGAL1, especially as we had focused on *Pichia pastoris* as a heterologous system, whose methanol induction system has been noted to induce the expression of proteases (Sinha et al., 2004). Future attempts at expressing *LuBGAL1* should thus

look at using protease deficient host systems (Ganatra et al., 2011), and perhaps examine further codon optimization in an effort to improve the likelihood of complete translation.

We did not observe any consistent morphological or mechanical changes to the plants themselves in response to increased LuBGAL1 transcription (Figures 4-3 and 4-6). On occasion, we did observe phloem fibres with thickened Gn-layers (Figure 4-6), reminiscent of phloem fibres in *LuBGAL1*-RNAi lines (Roach et al., 2011), which are likely the result of co-suppression, a not uncommon phenomenon encountered when overexpressing plant genes (Napoli et al., 1990). Despite this, on average, the levels of galactose in the cell walls amid stem cortical peels were not significantly altered, nor were the levels of the other measured monosaccharides (Figure 4-7). That said, the transgenic lines appeared to experience greater variation in galactose levels, compared to wild-type, with some samples more than doubling the average wild-type levels. This is likely due to the randomly appearing RNAi-like phloem fibres, with a higher incidence in some samples leading to elevated galactose levels.

Setting aside our uncertainty about the effects *LuBGAL1* transcript abundance on LuBGAL1 translation and activity, if we assume that LuBGAL1 activity had increased specifically in fibers then we could make several inferences about its importance in phloem fibre pectic galactan degradation. While the knockdown of *LuBGAL1* was observed to increase the levels of bound galactose (Roach et al., 2011), when *LuBGAL1* was overexpressed, galactan levels did not decrease in the cortical peels of the stem. This would indicate that while LuBGAL1 is necessary for galactan degradation, it is not sufficient to initiate the degradation of the pectic polysaccharide. Structural analyses of the phloem fibre pectic polysaccharide have indicated that the ends of galactan sidechains are capped with arabinose residues (Mikshina et al., 2012). Therefore, if LuBGAL1 is an exo-galactosidase, its activity would be impeded by this cap, and would thus need an exo-arabinosidase to initiate the degradation of the pectin sidechains. Alternatively, LuBGAL1 concentration within the phloem fibres cell wall may already be at saturation, with no space for additional LuBGAL1 proteins to act upon the pectic polysaccharide. This too would result in no perceivable decrease in cell wall bound galactose.

Conclusion

This experiment did not provide evidence that *LuBGAL1* overexpression can alter flax fibre morphology or composition. While transcript levels were increased in transgenic plants, this did not result in a noticeable increase in bulk BGAL activity, nor a concomitant decrease in cell wall galactans. Possibly we obtained sporadic co-suppression of *LuBGAL1* in flax phloem fibres, resulting in phenotypes characteristic of the down-regulation of *LuBGAL1* (Roach et al., 2011). To better understand the role of galactan-rich pectins in the development of the cell wall of phloem fibres, it will undoubtedly be more effective to identify the genes responsible for their production, not their degradation, and focus on analyses of lack of function. To do so, further analysis of the genes and proteins expressed at the initiation of secondary cell wall deposition will be required.

Figures

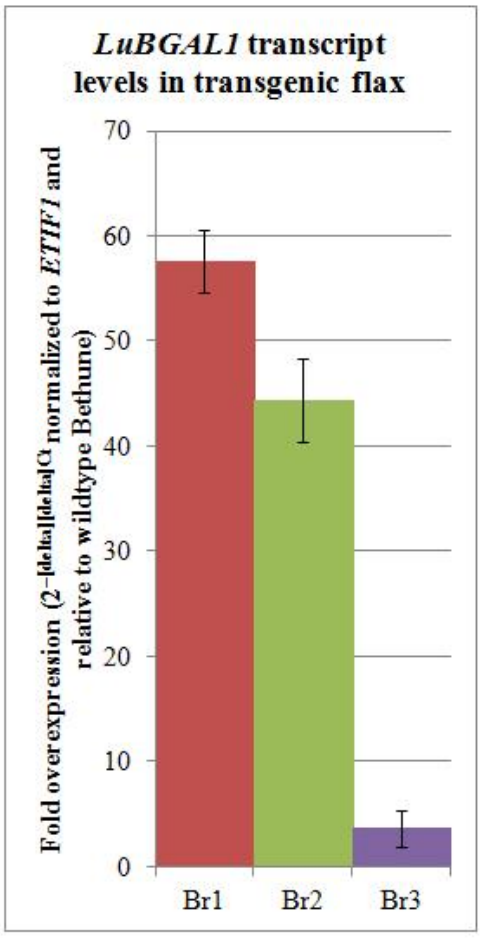

Figure 4-1 Transcript abundance of *LuBGAL1* in transgenic flax. Expression levels, relative to the reference gene *ETIF1* (eukaryotic translation initiation factor 1), were analyzed in the leaf tissue of wild-type and transgenic flax. Depicted values are relative to *LuBGAL1* expression in wild-type flax.

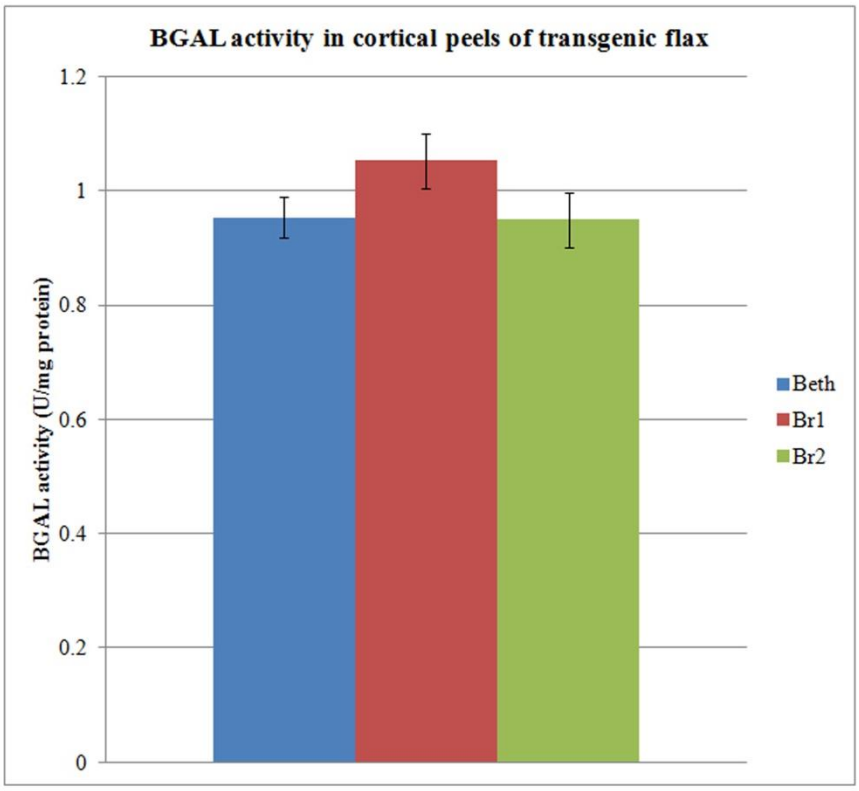

Figure 4-2 β-galactosidase activity in transgenic flax. The hydrolysis of ortho-Nitrophenyl-β-galactoside (ONPG) into galactose and ortho-nitrophenol was recorded for protein extracts from the stems of wild-type and transgenic flax. A single factor analysis of variance (ANOVA) was performed, and a significant difference in BGAL activity was found between the different lines (p-value < 0.01).

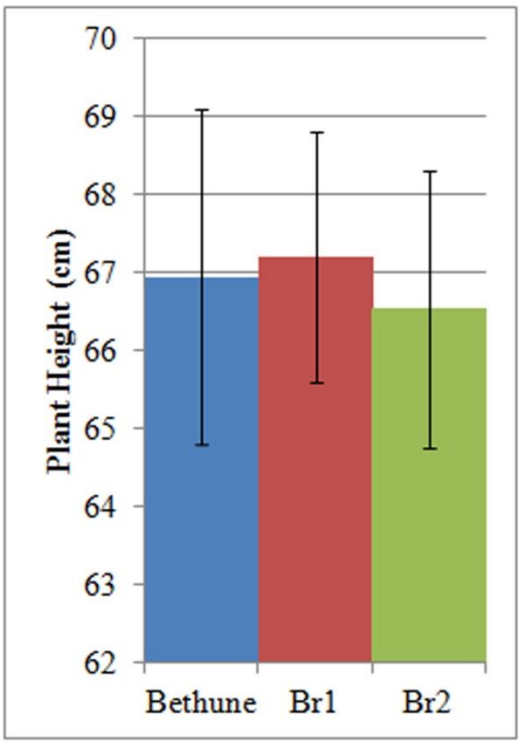

Figure 4-3 Height of transgenic flax. Measurements were performed on senescent flax, measuring from the cotyledons to the lowermost bracts. A single factor analysis of variance (ANOVA) was performed, and no significant difference in stem strength was observed between transgenic and non-transgenic flax (p-value > 0.05).

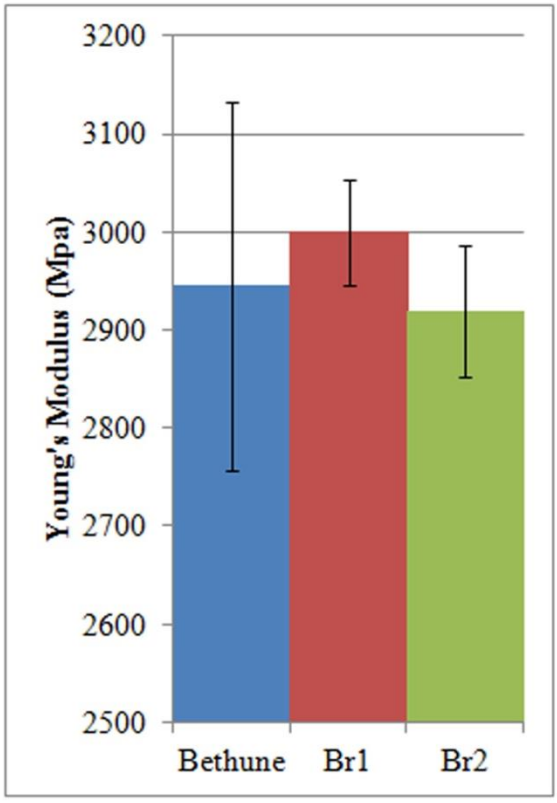

Figure 4-4 Tensile strength of transgenic flax stems. 15cm stem sections were torn apart with an Instron 5565. Young's modulus was calculated for every measured stem. A single factor analysis of variance (ANOVA) was performed, and no significant difference in stem strength was observed between transgenic and non-transgenic flax (p-value > 0.05).

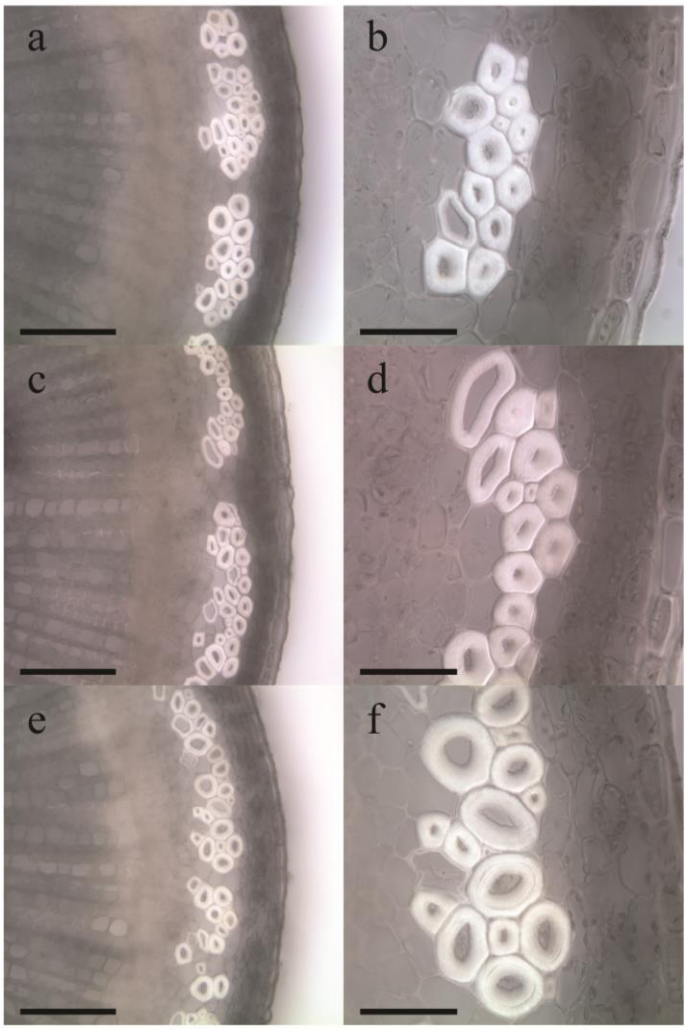

Figure 4-5 Flax phloem fibre bundles of trangenic flax. Transverse sections 24 cm below the snap-point of 5-week old flax were examined in wildtype plants (**a,b**), Br1 transgenics (**c,d**), and Br2 trangenics (**e,f**). *Scale bar* for **a**, **c**, and **e** is 150 μm. *Scale bar* for **b**, **d**, and **f** is 50 μm.

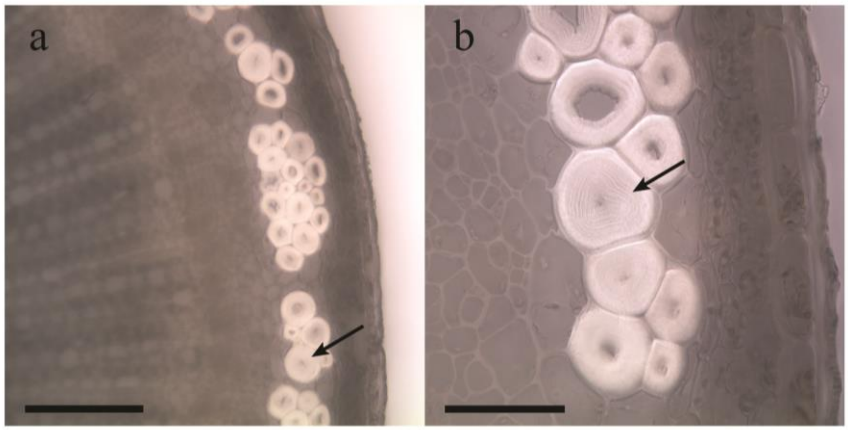

Figure 4-6 LuBGAL1-RNAi-like flax phloem fibres. Transverse sections 24 cm below the snap-point of 5-week old flax were examined. Images are from a member of the Br1 line of transgenics, but are representative of fibres randomly found in both Br1 and Br2 flax stems. Black arrows indicate the RNAi-like fibres. *Scale bar* for **a** is 150 µm. *Scale bar* for **b** is 50 µm.

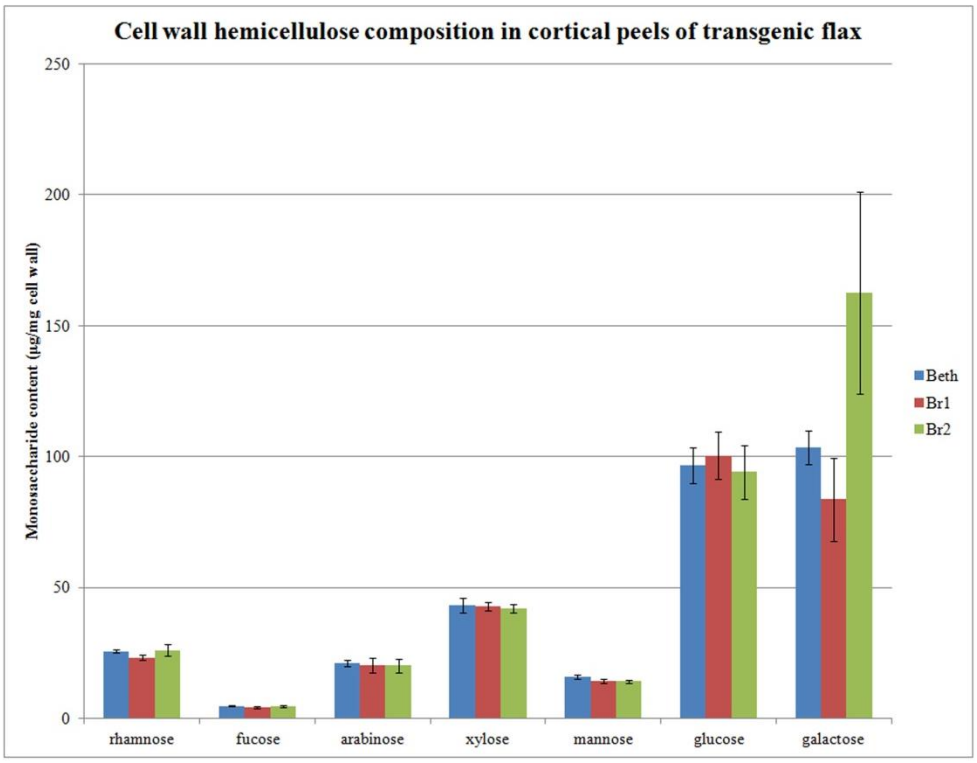

Figure 4-7 Monosaccharide composition of cell wall hemicelluloses in the cortical peels of flax stems. Non-cellulosic cell wall polysaccharides, obtained from cortical peels of 16 cm stem fragments just below the snap-point of 5-week old plants, were hydrolyzed and alditol acetate-derivitized. The derivitized neutral sugars were then analyzed by GC/MS. Inositol was added prior to hydrolysis as an internal control. The experiment was repeated twice, using three biological replicates each. The data is presented as mean monosaccharide concentration ± S.E (µg/mg of cell wall).

References

Ahn YO, Zheng M, Bevan DR, Esen A, Shiu S-H, Benson J, Peng HP, Miller JT, Cheng CL, Poulton JE, and Shih MC (2007) Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35. Phytochemistry **68**: 1510–1520. doi:10.1016/j.phytochem.2007.03.021

Dean GH, Zheng H, Tewari J, Huang J, Young DS, Hwang YT, Western TL, Carpita NC, McCann MC, Mansfield SD, and Haughn GW (2007) The *Arabidopsis MUM2* gene encodes a β-galactosidase required for the production of seed coat mucilage with correct hydration properties. The Plant Cell **19:** 4007–4021. doi:10.1105/tpc.107.050609

Deyholos MK (2006) Bast fiber of flax (*Linum usitatissimum* L.): Biological foundations of its ancient and modern uses. Israel Journal of Plant Sciences **54**: 273-280

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, and Pikaard CS (2006). Gateway-compatible vectors for plant functional genomics and proteomics. The Plant Journal **45**: 616–29.

Foster CE, Martin TM, and Pauly M (2010) Comprehensive Compositional Analysis of Plant Cell Walls (Lignocellulosic biomass) Part II: Carbohydrates. Journal of Visualized Experiments **37**: e1837. Doi: 10.3791/1837

Ganatra MB, Vainauskas S, Hong JM, Talyor TE, Denson J-PM, Esposito D, Read JD, Schmeisser H, Zoon KC, Hartley JL, and Taron CH (2011) A set of aspartyl protease-deficient strains for improved expression of heterologous proteins in *Kuyveromyces lactis*. REMS Yeast Research **11**: 168-178. doi: 10.1111/j.1567-1364.2010.00703.x

Gantulga D, Turan Y, Bevan DR, and Esen A (2008) The *Arabidopsis* At1g45130 and At3g52840 genes encode β-galactosidases with activity toward cell wall polysaccharides. Phytochemistry **69:** 1661–1670. doi:10.1016/j.phytochem.2008.01.023

Gorshkova TA, Salnikov VV, Chemikosova SB, Ageeva MV, Pavlencheva NV, and van Dam JEG (2003) The snap point: a transient point in *Linum usitatissimum* bast fibre development. Industrial Crops and Products **18**: 213-221.

Gorshkova T, and Morvan C (2006) Secondary cell-wall assembly in flax phloem fibers: role of galactans. Planta **223**: 149-158.

Gorshkova TA, Gurjanov OP, Mikshina PV, Ibragimova NN, Mokshina NE, Salnikov VV, AgeevaMV, Amenitskii SI, Chernova TE, Chemikosova SB (2010) Specific type of secondary cell wall formed by plant fibers. Russian Journal of Plant Physiology **57**: 346–360

Gurjanov OP, Gorshkova TA, Kabel, MA, Schols HA, and van Dam JEG (2007) MALDI-TOF MS evidence for the linking of flax bast fibre galactan to rhamnogalacturonan backbone. Carbohydrate Polymers **67**: 86-96.

Gurjanov OP, Ibragimova NN, Gnezdilov OI, and Gorshkova TA (2008) Polysaccharides, tightly bound to cellulose in the cell wall of flax bast fibre: Isolation and identification. Carbohydrate Research **72**: 719-729

Hobson N and Deyholos MK (2013) LuFLA1$_{PRO}$ and LuBGAL1$_{PRO}$ promote gene expression in the phloem fibres of flax (*Linum usitatissimum*). Plant Cell Reports **32:** 517-528. doi: 10.1007/s00299-013-1383-8

Hotte NSC, and Deyholos MK (2008) A flax fibre proteome: identification of proteins enriched in bast fibres. BMC Plant Biology **8**: 52. doi: 10.1186.1471-2229-8-52

Huis R, Hawkins S, and Neutelings G (2010) Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). BMC Plant Biology **10**: 71. doi:10.1186/1471-2229-10-71

Kotake T, Dina S, Konishi T, Kaneko S, Igarashi K, Samejima M, Watanabe Y,Kimura K, and Tsumuraya Y (2005) Molecular Cloning of a β-Galactosidase from Radish That Specifically Hydrolyzes β-(1→3)- and β-(1→6)-Galactosyl Residues of Arabinogalactan Protein. Plant Physiology **138**: 1563–1576. doi:10.1104/pp.105.062562

Mikshina PV, Gurjanov OP, Mukhitova FK, Petrova AA, Shashkov AS, and Gorshkova TA (2012) Structural details of pectic galactan from the secondary cell walls of flax (*Linum usitatissimum* L.) phloem fibres. Carbohydrate Polymers **87**: 853-861. doi: 10.1016/j.carbpol.2011.08.068

Mohanty AK, Misra M, and Hinrichsen G (2000) Biofibres, biodegradable polymers and biocomposites: an overview. Macromolecular Materials and Engineering **276**: 1-24.

Napoli C, Lemieux C, and Jorgensen R (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes *in trans*. The Plant Cell **2**: 279-289

Que Q, Wang HY, English JJ, and Jorgensen RA (1997) The Frequency and Degree of Cosuppression by Sense Chalcone Synthase Transgenes Are Dependent on Transgene Promoter Strength and Are Reduced by Premature Nonsense Codons in the Transgene Coding Sequence. The Plant Cell **9**: 1357-1368.

Raué HA (1994) Metabolic stability of mRNA in yeast – a potential target for modulating productivity? Trends in Biotechnology **12**: 444-449. doi: 10.1016/0167-7799(94)90019-1

Roach MJ, and Deyholos MK (2008) Microarray Analysis of Developing Flax Hypocotyls Identifies Novel Transcripts Correlated with Specific Stages of Phloem Fibre Differentation. Annals of Botany **102**: 317-330. doi: 10.1093/aob/mcn110

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, and Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant Physiology **156**: 1351–1363. doi:10.1104/pp.111.172676

Salnikov VV, Ageeva MV, and Gorshkova TA (2008) Homofusion of Golgi secretory vesicles in flax phloem fibres during formation of the gelatinous secondary cell wall. Protoplasma **233**: 269-273. doi: 10.1007/s00709-008-0011-x

Sampedro J, Gianzo C, Iglesias N, Guitián E, Revilla G, and Zarra I (2012) AtBGAL10 is the main xyloglucan β-galactosidase in Arabidopsis, and its absence results in unusual xyloglucan subunits and growth defects. Plant Physiology **158**: 1146–57. doi:10.1104/pp.111.192195

Sinha J, Plantz BA, Mehmet I, and Meagher MM (2004) Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant ovine interferon-T. Biotechnology and Bioengineering **89**: 102-112. doi: 10.1002/bit.20318

Stanke M, Diekhans M, Baertsch R, and Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics **24**:637-644. doi:10.1093/bioinformatics/btn013

Schmittgen TD and Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nature Protocols **3**: 1101–1108.

Tanthanuch W, Chantarangsee M, Maneesan J, and Ketudat-Cairns J (2008) Genomic and expression analysis of glycosyl hydrolase family 35 genes from rice (*Oryza sativa* L.). BMC Plant Biology **8**: 84. doi:10.1186/1471-2229-8-84

Wohlleben W, Arnold W, Broer I, Hillemann D, Strauch E, and Pühler A (1988) Nucleotide sequence of the phosphinothricin N-acetyltransferase gene from *Streptomyces viridochromogenes* Tü494 and its expression in *Nicotiana tabacum*. Gene **70**: 25–37.

Chapter 5 - Fasciclin-like Arabinogalactan Proteins and Their Role in the Development of the Phloem Fibres of Flax (*Linum usitatissimum*)[*]

Introduction

The phloem fibres of flax (*Linum usitatissimum*) are long sclerenchymatous cells characterized by a low content of lignin and a high content of crystalline cellulose. Their size, the composition of their cell walls, and their high tensile strength, make them valuable in the production of textiles as well as a high quality papers and composite materials (Mohanty et al., 2000).

Flax phloem fibres are derived from the protophloem of the primary plant body, and subsequently elongate to several centimetres in length (Esau, 1965; Fahn, 1990; Ageeva et al., 2005). Following elongation, the phloem fibres of flax undergo secondary cell wall deposition. In the first stage, cellulose microfibrils and fibre-specific, galactan-rich, high-molecular weight pectic polysaccharides are deposited in the cell wall, producing a galactan (Gn)-layer (Gorshkova et al., 2006; Salnikov et al., 2008). As development progresses, the pectic polysaccharides are hydrolyzed by β-galactosidases (EC 3.2.1.23), resulting in replacement of the Gn-layer by a gelatinous (G)-layer, which is rich in crystalline cellulose (Gurjanov et al., 2007; Gurjanov et al., 2008; Gorshkova et al., 2010; Roach et al., 2011; Mikshina et al., 2012).

Fasciclin-like arabinogalactan proteins (FLAs) are a sub-class of the arabinogalactan (AGP) family of proteoglycans, and are characterized by the presence of fasciclin cell adhesion domains (Johnson et al., 2003). Transcripts of some FLA genes are relatively abundant in fibre-bearing phloem tissues of flax and hemp, as well as in tension wood of poplar (Lafarguette et al., 2004; De Pauw et al., 2007; Roach et al., 2007; Roach et al., 2008; Hobson et al., 2010). Both phloem fibres and tension wood are comprised of G-type cell walls (Gorshkova et al., 2012). Other studies have expanded on this theme, linking FLA expression with the deposition of

[*] Portions of this chapter have been published as: Hobson N, Roach MJ, and Deyholos MK (2010) Gene expression in tension wood and bast fibers. Russian Journal of Plant Physiology **57**:321-327

cellulose in arabidopsis (Persson et al., 2005), in eucalyptus tension wood (Qiu et al., 2008), and in the mature wood of radiata pine (Li et al., 2009). Mutations of FLAs have resulted in disruptions of cell wall architecture. A point mutation in *AtFLA4* resulted in thinner cell walls and absence of a middle lamella in roots (Shi et al., 2003), T-DNA mutations in *AtFLA11* (*IRX13*) and *AtFLA12* resulted in cell wall deformations and reduced cellulose content (Persson et al., 2005; MacMillan et al., 2010), while RNAi of *AtFLA3* resulted in reduced cellulose in the cell walls of pollen (Li et al., 2010).

In flax, a FLA (Genbank accession EH791369) was found to be particularly enriched in tissues rich in phloem fibres (Roach and Deyholos, 2007; Roach and Deyholos, 2008). Further analyses of its promoter revealed a propensity for phloem-fibre specific expression (Hobson and Deyholos, 2013). RNAi constructs targeting this gene were previously developed and transformed into flax, although their phenotypes were not described in detail (Roach, 2009). Here, we continue the characterization of the FLA family of flax, including analysis of the effects of RNAi knockdown of a FLA on the development of the secondary cell walls of flax phloem fibres.

Materials and Methods

Gene Discovery

The 43,384 predicted proteins of the flax genome (Wang et al., 2012), available at Phytozome (version 9.1; www.phytozome.net), were screened via Hidden Markov Model (HMM) with HMMER3 (http://hmmer.janelia.org), using the Pfam-A family database (version 25.0; Punta et al., 2012), to identify genes encoding a fasciclin domain. Pairwise sequence alignments were performed with the EMBOSS Water alignment program.

Phylogenetics

Predicted protein sequences from the primary transcripts of *Arabidopsis thaliana*, cabbage (*Brassica rapa*), sweet orange (*Citrus sinensis*), eucalyptus (*Eucalyptus grandis*), cotton (*Gossypium raimondii*), apple (*Malus domestica*), cassava (*Manihot esculenta*), rice (*Oryza*

sativa), poplar (*Populus trichocarpa*), castor bean (*Ricinus communis*), tomato (*Solanum lycopersicum*), spikemoss (*Selaginella moellendorffii*), grape (*Vitis vinifera*), and maize (*Zea mays*) were obtained from Phytozome (version 9.1; www.phytozome.net; Swarbreck et al., 2008; Ouyang et al., 2007; Tuskan et al., 2006; Jaillon et al., 2007; Schnable et al., 2009; Chan et al., 2010; Schmutz et al., 2010; Velasco et al., 2010; Banks et al., 2011; Cheng et al., 2012; Prochnik et al., 2012; Sato et al., 2012). Eucalyptus, poplar, and upland cotton (*Gossypium hirsutum*) FLA sequences were retrieved from GenBank using published accession numbers (Lafarguette et al., 2004; Huang et al., 2008; MacMillan et al., 2010). Retrieved sequences were labelled as FLAs (Table S5), using published FLA names wherever possible (e.g. AtFLA1; Johnson et al., 2003; Lafarguette et al., 2004; Huang et al., 2008; Ma and Zhao, 2010; MacMillan et al., 2010; Jun and Xiaoming, 2012). Amino acid sequences were aligned using the default parameters of Muscle 3.7 (Edgar, 2004). ProtTest 3.2, with default parameters, was used to determine the best-fit model of amino acid substitution for a maximum likelihood analysis of the sequence alignment (Darriba et al., 2011). Using the WAG model of amino acid substitution (Whelan and Goldman, 2001), while employing gamma-distributed rate variations, and empirically determined base frequencies, we performed a maximum likelihood analysis with GARLI (Zwickl, 2006; Bazinet and Cummings, 2011; www.molecularevolution.org). The consensus tree of 1000 bootstraps was obtained using CONSENSE (Phylip 3.66) at the CIPRES Science Gateway (Miller et al., 2010).

EST Identification

Sequences of *Linum usitatissimum* transcripts were obtained from the NCBI-EST database (accessed April, 2013) and from an assembly of Illumina RNA sequencing reads from flax stem segments (accession POZS; www.onekp.com). Sequences were aligned to the WGS predicted *LuFLA* CDSs using the RNA-SEQ analysis tool of CLC Genomics Workbench 5.5. Only sequences aligning to CDSs with 95% similarity, along 90% of their length, were included in further analyses.

Microarray Analyses

Flax microarray datasets GSE21868 (Fénart et al., 2010) and GSE29345 (Huis et al., 2012) were obtained from NCBI GEO. Experiment GSE21868 examined expression in a range of tissues and organs: roots (R); leaves (L); outer stem tissues at either the vegetative stage (SOV) or green capsule stage (SOGC); inner stem tissues at either vegetative stage (SIV) or green capsule stage (SIGC); and seeds 10-15 days after flowering (DAF; E1), 20-30 DAF (E2), and 40-50 DAF (E3; Fénart et al., 2010). Experiment GSE29345 focused on the development of stem tissues by comparing: internal (i.e. xylem enriched) stem tissues of either the whole stem (WSI), upper stem (USI), middle stem (MSI), or lower stem (LSI); and external (i.e. phloem and cortex enriched) stem tissues of the whole stem (WSE), upper stem (USE), middle stem (MSE), and lower stem (LSE; Huis et al., 2012). The flax unigenes used in microarray construction (http://tinyurl.com/c888l9d) were aligned to the predicted *LuBGAL* CDSs, using the RNA-Seq function of the CLC Genomics Workbench 5.5, and were classified as matches if at least 90% of their sequence length aligned, with at least 95% sequence identity between the transcript and CDS. Microarray data corresponding to the flax BGALs were then extracted. Robust Multichip Average (RMA)-normalized signal intensities (log_2) were averaged between biological and technical replicates. Heat maps of expression levels were then created with MeV v4.8 (http://www.tm4.org/mev).

Fluidigm

Tissue samples from *Linum usitatissimum* (CDC Bethune) were frozen in liquid nitrogen, and stored at -80°C prior to use. Frozen samples were ground in liquid nitrogen, whereupon we followed the CTAB/Acid Phenol/Silica Membrane Method to extract the RNA (Johnson et al., 2012). DNA was removed using on-column RNase-Free DNase (Qiagen), and/or with the TURBO DNA-Free kit (Invitrogen). cDNA was prepared with RevertAid H Minus Reverse Transcriptase (Fermentas) and oligo$(dT)_{18}$ primer. qPCR primer pairs and hydrolysis probes (Table S6) were designed with the Universal Probe Library Assay Design Center (http://tinyurl.com/7u6s5bh). A 14 cycle pre-amplification of the target sequences was performed

with a TaqMan PreAmp Master Mix (ABI) and 5 ng of cDNA, which was subsequently diluted 1:5. Assay master mixes of 3.2 µl 2X Assay Loading Reagent (Fluidigm PN 85000736), 2 µl primer mix (13.3 µM primer and 3.3 µM hydrolysis probe) and 1.3 µl water was prepared, of which 5 µl was loaded into the assay wells of a primed Fluidigm 96*96 well chip. Sample master mixes of 3.63 µl Taqman Universal PCR Master Mix - no AmpErase UNG (PN 4324018), 0.36 µl 20X GE Sample Loading Reagent (Fluidigm PN 85000735), and 2.5 µl diluted pre-amped cDNA were prepared, of which 5 µl was loaded into the sample wells of the primed Fluidigm 96*96 well plate. The Fluidigm chip was run through the following thermal cycles: 95ºC – 10 min, 40X cycles of 95ºC – 15 sec and 60ºC – 1 min. ΔC_T values were calculated based on the geometric mean of reference genes *ETIF1* (eukaryotic translation initiation factor 1), *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase), and *ETIF5A* (eukaryotic translation initiation factor 5A; Schmittgen and Livak, 2008; Huis et al., 2010). We compared expression in 12 different tissues: roots (R); leaves (L); senescing leaves (SL); stem apex (SA); cortical peels from vegetative stage stems (ECP) or green capsule stage stems (LCP); phloem fibres from vegetative stage stems (EF) or green capsule stage stems (LF); xylem from vegetative stage stems (X); budding flowers (FB); open flowers (F); and seed bolls from the green capsule stage (B). A heat map of relative expression values (log_2), averaging technical (two for F, FB, L, and SL; three for all other samples) and biological (three, each of which is a pooled sample from multiple plants) replicates, was then prepared with MeV v4.8 (http://www.tm4.org/mev). Hierarchical clustering of genes was performed using Pearson correlation.

qRT-PCR

Tissue samples from *Linum usitatissimum* (CDC Bethune) were frozen in liquid nitrogen, and stored at -80°C prior to use. Frozen samples were ground in liquid nitrogen, whereupon 1.5 ml Trizol was added per mg of tissue. Samples were heated at 60°C for 5 min, and centrifuged at 12,000 g for 10 min at 4°C. Supernatant was collected and mixed with 1/5V chloroform, mixed, allowed to settle for 3 min, and centrifuged at 12,000 g for 20 min at 4°C. The aqueous phase as mixed with 0.25V RLT buffer (Qiagen RNeasy mini kit), and then applied to the RNeasy Mini

spin column. An on-column DNase digestion was performed as per kit instructions, with continued washes and elutions as per kit instructions. cDNA was prepared with RevertAid H Minus Reverse Transcriptase (Fermentas) and oligo$(dT)_{18}$ primer. Real-time PCR was performed with an Applied Biosystems 7500 Fast Real-Time PCR System. cDNA derived from 1 µg was diluted to 1/16, 2.5 µl of which was added to a reaction volume of 10 µl containing 0.4 µM of forward and reverse primers (LuFLA1qpcr_F = 5' – GGACAAATTCATCTGCCTCATC – 3'; LuFLA1qpc_R = 5' – GTTATTGCTGGTAGGGATTTTGTT - 3'), 0.2 mM dNTPs, 0.25X SYBR Green, 1X ROX, and 0.075 U Platinum Taq. Threshold cycles (Ct) were determined by the 7500 Fast Software. $\Delta\Delta C_T$ values were calculated in reference to *ETIF1* (eukaryotic translation initiation factor 1; Huis et al., 2010) and wild-type gene expression.

AGP Quantification

Frozen tissue samples were ground in liquid nitrogen, and then 1 ml 65 mM Tris-Cl pH 6.8, 2% SDS, 2 mM EDTA, and 700 mM β-mercaptoethanol. Samples were boiled for 3 min, and then centrifuged at 14,000 rpm for 10 min, at 4ºC. The supernatant was kept for further analysis.

Protein concentrations were determined via Qubit flurometer (Invitrogen). 10 µg of protein was applied to a 1% agarose gel, buffered with 25 mM Tris-Cl, and 200 mM Glycine, pH 8.3, and containing 10 µg/ml β-glucosyl Yariv reagent. A standard curve of 0.2, 0.4, 0.8, 1.6, 2.4, and 3.2 µg Gum Arabic (protein weight) was used. The gel was run at 100V at 4ºC, for 17 hours. Unbound Yariv reagent was washed from the gel in 4% NaCl, before the gel was dried into a thin film. Lengths of columns of Yariv precipitate, which correspond to the quantity of AGP, were measured and compared to the standard curve.

Tensile Strength

Measurements were performed on mature, dry stems, using an Instron 5565. 15 cm stem fragments, located 13 cm above the cotyledon, were used. Instron grips were set 75 mm apart. Tension was applied at 30 mm/s. Young's modulus (E) was calculated with the following

formula, where F is the maximum force (*N*) applied to stems before breakage, A is the area of the stem (mm), ΔL is the change in length due to stretching (mm), and L is the original distance between the two grips (75 mm): $E = \frac{tensile\ stress}{tensile\ strain} = \frac{\frac{F}{A}}{\frac{\Delta L}{L}}$

GC-MS

Cortical peels were harvested from 5-week old flax, starting at the snap-point and extending 16 cm down. Samples were lyophilized, and ground for 10 min at 25 Hz with two 5 mm stainless steel ball bearing and a Retsch mixer mill (MM301). Cell wall isolation and monosaccharide derivatization were performed as described elsewhere (Foster et al., 2010), with remaining cell wall material conserved for Fourier transform infrared spectroscopy (FTIR). 1 µl samples were injected into an Agilent 5975 GC/MS with a HP-5MS column (30 m x 0.25 mm x 0.25 µm), running the following program: 2 min at 100°C, 10°C/min until 200°C, 5 min at 200°C, 10°C/min until 250°C, and a hold at 250°C for 10 min. Compounds were identified by comparison to the retention times of individual monosaccharide standards.

FTIR

Cell wall material was finely ground with potassium bromide, and pressed into transparent disks. Spectra were collected from a Nicolet Magna-IR Spectrometer 750 with a DTGS detector, recording absorbance between 400 cm^{-1} and 4000 cm^{-1}, at a resolution of 4 cm^{-1}. 32 Scans were co-added using Omnic Software (Thermo Nicolet). Using "The Unscrambler X" (version 10.2; Camo), we reduced the spectra range to 800 cm^{-1} - 1800 cm^{-1}, performed baseline corrections, and area normalized the data. The spectra of three biological replicates were averaged, and differences in spectra were highlighted by subtracting the averaged spectra of the negative sibling (A15) from the averaged spectra of the transgenic lines.

Results

Gene discovery and *in silico* analyses

Fasciclin domains (PFAM: PF02469) were identified in 43 predicted proteins encoded on 33 separate scaffolds of the *de novo* flax genome assembly (Table 5-1; Wang et al., 2012). These included several previously described LuFLA genes. *LuFLA35* was originally identified as a cDNA microarray probe (probeset 152; Genbank accession EH791369) that was highly expressed in fibres. PCR primers based on EH791369 were used to screen a fosmid library, which resulted in identification of *LuFLA1*, *2*, and *3* (Genbank accession JN133301). The upstream promoter region of *LuFLA1* was later found to induce reporter gene expression specifically in the phloem fibres of flax (Hobson and Deyholos, 2013).

The 43 predicted LuFLA genes comprised 0.099% of the total number of predicted flax genes, which was identical to the proportion of predicted β-galactosidase (BGAL) genes in the flax genome (Hobson and Deyholos, in press). However, in comparison to other species (Table S5), the size of the FLA family of flax was not significantly different (p-value = 0.053). This contrasts with the BGAL family of flax, which is large in comparison to most other species (Hobson and Deyholos, in press).

To classify the LuFLAs, we performed a phylogenetic analysis using deduced amino acid sequences of the predicted FLA genes from the genome assemblies of *A. thaliana*, *B. rapa*, *C. sinensis*, *E. grandis*, *G. raimondii*, *M. domestica*, *M. esculenta*, *O. sativa*, *P. trichocarpa*, *R. communis*, *S. lycopersicum*, *S. moellendorffii*, *V. vinifera*, and *Z. mays*, as well as the amino acid sequences corresponding to the FLA cDNAs of *G. hirsutum* (Figure 5-2; Table S5). The arabidopsis, cabbage, and rice FLA families were included because they have been previously characterized, and could therefore be used as a reference for the categorization of FLA sub-families. Poplar, cassava, and castor bean FLA families were included as additional members of the Malpighiales for which whole genome sequences are available. Apple, eucalyptus, grape, and orange FLA families were selected to represent woody plant species, while herbaceous plants were represented by soybean and tomato. Maize and spikemoss were included to increase phylogenetic diversity of the sampling, and represented monocots and lycopodiophytes respectively.

FLAs are typically categorized into one of four groups, based on amino acid sequence similarity (Johnson et al., 2003; Ma and Zhao, 2010; Jun and Xiaoming, 2012), although classification and other conventions are not always consistent between publications. In arabidopsis, group A FLAs have a single fasciclin domain surrounded by AGP regions, and terminate with an attachment site for a glycosylphosphatidylinositol (GPI) anchor (Johnson et al., 2003). Group B FLAs are generally longer, with two fasciclin domains and no GPI attachment signal, whereas group C and D FLAs are of a more variable composition, with 1-2 fasciclin domains, with and without GPI anchors (Johnson et al., 2003). Using the categorization of the AtFLAs as a guide (Johnson et al., 2003), we initially classified the LuFLAs into these same four sub-families. However we noted that the group C AtFLAs could be separated into two clades, which generally reflected the number of fasciclin domains present. We chose to depict this by separating group C into C1 (one fasciclin domain) and C2 (two fasciclin domains; Figure 5-2). Unlike arabidopsis or cabbage (Johnson et al., 2003; Ma and Zhao, 2010), the FLAs of flax were found primarily in group A, which contained 19 of the 43 LuFLA sequences (Table 5-2). Group A was also enriched in FLAs from poplar, cassava, and castor bean, other members of the Malpighiales order, suggestive of a duplication early in the evolution of this order. Group C2 contained the next largest concentration of LuFLAs, with nine members which generally contained two fasciclin domains. The exception to this was LuFLA7, which had a single fasciclin domain. Group C1 was the third largest clade, containing seven LuFLAs, while groups B and D were of equal size, with four members each. While group B was generally composed of FLAs with two fasciclin domains, there was one exception, with LuFLA25 containing only a single fasciclin domain.

As noted above, within group A a large cluster of FLAs from the Malpighiales was identified, comprising 59 sequences (Figure S5). This group contained 11 of the 43 LuFLAs, 15 of the 28 MeFLAs (cassava), 22 of the 44 PtFLAs (poplar), and 11 of the 25 RcFLAs (castor bean). Using MEME (Bailey and Elkan, 1994), we identified a 10 aa motif near the end of the fasciclin domain that distinguished LuFLAs and PtFLAs of this group from LuFLAs and PtFLAs elsewhere in group A, (Figure 5-3). The 10 aa motif was surrounded by sequence that was

conserved among all FLAs. BLASTP of the motif LTNTSLSGTV, against the FLAs extracted from 34 proteomes (www.phytozome.net), returned additional matches from the Malpighiales (flax, poplar, cassava, and castor bean), as well as nine other rosids (papaya, cotton, cocoa, apple, peach, strawberry, cucumber, orange, and cabbage), but no asterids, monocots, or non-seed plants.

Gene expression in public datasets

To determine the expression patterns of LuFLAs, we first retrieved 286,895 flax ESTs from Genbank, of which 74.8% were obtained from flax seeds at various stages of development (Venglat et al., 2011). We also retrieved an Illumina *de novo* transcript assembly from developing flax stems (POZS; www.onekp.com). A total of 80 ESTs and transcript assemblies could be aligned with high stringency to 27 of the 43 LuFLAs predicted from the whole genome assembly (Table 5-1). However, because only a limited number of tissues and conditions were represented by the transcripts queried, it is likely that other predicted LuFLAs may also be expressed under other conditions.

We next examined gene expression patterns of the LuFLA family using publicly available oligonucleotide microarray data from two experiments involving a Nimblegen 25-mer oligonucleotide array (NCBI GEO accessions GSE21868 and GSE29345; Fénart et al., 2010; Huis et al., 2012), developed from a flax EST library. Based on alignments where >90% of the EST length matched the LuFLA CDSs at >95% sequence identity, we identified probes for five different LuFLA genes (*LuFLA11*, *LuFLA14*, *LuFLA15*, *LuFLA24*, and *LuFLA35*). A heat map of expression values from these microarrays (Figure 5-4 a, b) showed that *LuFLA14* expression was enriched in young seeds (10-15 DAF; E1; Figure 5-4 a), with little to moderate amounts of expression in the stem (Figure 5-4 a, b). *LuFLA15* appeared to be strongly expressed in the stems of both vegetatively growing and mature flax, as well as the roots (Figure 5-4 a), with stronger expression in the xylem-rich tissues of the upper stem (Figure 5-4 b) compared to the phloem-fibre bearing tissues. *LuFLA35* showed the strongest expression in the stem, especially the outer phloem fibre-rich tissues of vegetatively growing flax (Figure 5-4 a, b), consistent with earlier reports of *LuFLA35* expression (Roach and Deyholos, 2007).

qRT-PCR analysis of *LuFLA* expression

To further examine the patterns of FLA expression in flax, we performed qRT-PCR in a Fluidigm 96*96 array, examining 12 different tissues: roots (R); leaves (L); senescing leaves (SL); stem apex (SA); cortical peels from vegetative stage stems (ECP) or green capsule stage stems (LCP); phloem fibres from vegetative stage stems (EF) or green capsule stage stems (LF); xylem from vegetative stage stems (X); budding flowers (FB); open flowers (F); and seed bolls from the green capsule stage (B). Primers used in the qRT-PCR analysis (Table S6) were verified as being gene specific following a series of BLASTN alignments against the scaffolds and CDSs of the flax genome assembly (Wang et al., 2012), ensuring that no two primers hybridized within 100 bp of each other, except at their target sequences. Primers were not developed for *LuFLA3* as it had not been annotated at the time the experiment was conducted.

Transcripts were detected for 40 of the 42 genes examined (Figure 5-5), although we could not detect expression for *LuFLA13*, nor *LuFLA25*. Expression of most *LuFLA*s could be detected in a wide range of tissues. However, expression of a subset of FLAs was particularly enriched in phloem fibres, as well as in developing flowers, including *LuFLA1, 2, 7, 10, 14, 16, 19, 30, 32, 33,* and *38*. We note that *LuFLA1* has been previously studied using transgenic promoter-reporter gene fusions, in which expression was detected only in developing phloem fibres (Hobson and Deyholos, 2013). Using Fluidigm qRT-PCR, we here detected strong expression of *LuFLA1* transcripts in both phloem fibres (F) and in developing seed bolls (B), and relatively weaker expression of this gene in several other tissues.

***LuFLA35*-RNAi: Transcript and Protein Abundance**

LuFLA35 was previously shown by microarray analysis to be highly expressed in developing fibres. A corresponding EST (EH791369) was used to design an RNAi-silencing construct to test the effects of loss of function of FLA genes on fibre development (Roach and Deyholos, 2007; Roach and Deyholos, 2008; Roach, 2009). To estimate the number of *FLA*

genes that would have been impacted by the *LuFLA35*-RNAi construct, I performed local pairwise sequence alignments between the RNAi amplicon and each of the predicted *FLA* genes (Table 5-1). Three genes, *LuFLA1*, *10*, and *35*, shared >90% sequence identity with the RNAi amplicon. *LuFLA1*, of the three, exhibited the lowest shared sequence identity at 94.1%. These three genes were all part of a sub-group (Figure S5) within group A that contained only genes from flax and other members of the Malpighiales, including poplar FLAs (*PtFLAs 1-6*) that were reported to be enriched the tension wood (Lafarguette et al., 2004). No other flax FLA was present in this sub-group.

Although *LuFLA35* transcript expression had been previously quantified in the *LuFLA35*-RNAi transgenic lines (Roach, 2009), we wanted to determine the effect of the RNAi construct on the closely related *LuFLA1* gene, as we had previously characterized the expression pattern derived from its promoter (Hobson and Deyholos, 2013). To do this, qRT-PCR was performed, comparing *LuFLA1* transcript abundance in the leaves of transgenic flax and wild-type Norlin (Figure 5-6). Similar to the expression levels previously observed of *LuFLA35* (Roach, 2009), the three lines showed a decrease in *LuFLA1* expression of 0.5-0.4 fold across the different lines, indicating that multiple *FLA* genes were downregulated by the RNAi construct.

To determine whether the knock-down of *FLA* expression led to a decrease in FLA protein, we performed rocket electrophoresis with total protein extracts from the cortical peel of the stem, to measure the concentration of arabinogalactan proteins (AGPs; Figure 5-7). No decrease in total AGP concentration was detected among the transgenic flax. Nevertheless, because the RNAi-targeted genes were only a subset of the 43 FLAs in flax, it is possible that the effects of their silencing were masked by the normal expression of other FLAs in the stem (Figure 5-5). We therefore chose to continue the phenotypic characterization of the *LuFLA35*-RNAi lines.

Physical Characteristics of Cell Walls

To determine whether the suppression of *FLA* transcripts impacted the composition of the phloem fibre cell wall, we measured the tensile strength of the plant stems (Figure 5-8), comparing the three independent *LuFLA35*-RNAi lines (A1, A3, 10) to wild-type Norlin and to previously

characterized *LuBGAL1*-RNAi transgenic lines (Roach et al., 2011). An analysis of variance (ANOVA) indicated that significant differences in tensile strength existed between the different lines (p-value = 0.002). The *LuBGAL1*-RNAi lines were deemed to be significantly weaker than wild-type Norlin, (B4 p-value = 0.01; B6 p-value = 0.002), in agreement with previously reported observations of tensile strength (Roach et al., 2011). While the *LuFLA35*-RNAi lines were not as weak as the *LuBGAL1*-RNAi lines, compared to wild-type flax, line A1 was deemed significantly weaker than Norlin (p-value = 0.04). Neither A3, nor A10, were found to differ significantly from wild-type flax.

We next examined transverse sections of flax stems. In most cases, no obvious phenotype was observed, consistent with previous reports of the *LuFLA35*-RNAi line morphology (Figure 5-9; Figure 5-10 a, b; Roach, 2009). However, in a few lines, a small number of individual phloem fibres displayed an unusual phenotype; appearing as though two individual fibres were present within a single larger fibre (Figure 5-10 c, d). These fibres were observed sporadically within each of the *LuFLA35*-RNAi lines examined, and not in any other genotype; however, this phenotype was also observed only during a single growing cycle, and was not observed in subsequent plantings of the same generation of transgenic flax.

Chemical Composition of Cell Walls

Although the *LuFLA35*-RNAi transgenic phloem fibres were generally indistinguishable from wild-type in terms of their morphology, we considered the possibility that the transgenic lines differed from wild-type in chemical composition. We therefore profiled the monosaccharide constituents of the cell walls to determine whether any changes in chemistry occurred. Upon hydrolyzing and derivatizing the non-cellulosic cell wall polysaccharides of stem cortical peels, we measured the quantity of arabinose, fucose, galactose, glucose, mannose, rhamnose, and xylose (Figure 5-11). While decreases in glucose and galactose content were observed amidst the *LuFLA35*-RNAi lines, compared to wild-type Norlin, ANOVA analyses indicated the differences were insignificant for all monosaccharides (arabinose p-value = 0.98; fucose p-value = 0.42;

galactose p-value = 0.60; glucose p-value = 0.49; mannose p-value = 0.74; rhamnose p-value = 0.87; and xylose p-value = 0.43).

To obtain a more holistic representation of the changes in cell wall chemistry occurring in the transgenic lines, we examined the cell wall material of cortical peels using Fourier-transform infrared spectroscopy (FTIR; Figure 5-12). Some cell wall components can be detected by their absorbance of specific wavelengths of light. Galactans can be distinguished by their absorbance at 893 cm^{-1}, whereas arabinans absorb at a frequency of 895 and 1035 cm^{-1} (Gorgulu et al., 2007). Glucomannans display an absorbance at 1064 cm^{-1}, rhamnogalaturonans at 1073 cm^{-1}, and pectins at 1101 cm^{-1},1250 cm^{-1}, 1651 cm^{-1}, and 1733 cm^{-1}. Cellulose itself absorbs at 1145 cm^{-1}, while proteins in general absorb at 1235 cm^{-1}, 1555 cm^{-1}, and 1651 cm^{-1} (Gorgulu et al., 2007). Some wavelengths are absorbed by multiple cell wall constituents, such as phospholipids, cholesterol, and hemicelluloses at 1733 cm^{-1}, a definitive classification of cell wall changes can be difficult. Nevertheless, by subtracting the IR spectra of a negative sibling from the spectra of the transgenic lines, as well as wild-type Norlin (Figure 5-13), we can more easily identify any differences between spectra. The spectral variation observed in the *LuFLA35*-RNAi lines, such as at 1446 cm^{-1} which corresponds to the C-H bonds in polysaccharides (Gorgulu et al., 2007), was within the variation observed between the negative sibling and wild-type Norlin, even when the average of six biological replicates. Thus, the FTIR analysis failed to provide evidence of differences in composition of the transgenic lines compared to non-transgenic controls.

Further analyses of the cell walls of the transgenic flax are being performed by collaborators at the University of British Columbia, examining cellulose crystallinity and microfibrillar angle. Results are pending.

Discussion

Studies have increasingly related fasciclin-like arabinogalactan proteins to the deposition of cellulose in cell walls (Persson et al., 2005; Qiu et al., 2008; MacMillan et al., 2010; Li et al., 2010). Here, we identified 43 predicted FLAs from flax, which were distributed throughout each of the previously defined FLA sub-families of vascular plants. Unlike the BGAL family of flax,

which displays multiple copies for every BGAL of another related species (Hobson and Deyholos, in press), the LuFLA family appeared of average size, with equally large or larger FLA families found in genomes of other species, including non-woody species (Figure 5-1).

To date, very few FLAs have been characterized experimentally. Aside from expression analyses linking FLA transcript abundance to cells with thick, cellulose-rich cell walls (Lafarguette et al., 2004; De Pauw et al., 2007; Roach et al., 2007; Roach et al., 2008; Qiu et al., 2008; Li et al., 2009; Hobson et al., 2010; Hobson and Deyholos, 2013), mutations and silencing of FLAs has resulted in disruptions of cell wall architecture (Shi et al., 2003; Persson et al., 2005; MacMillan et al., 2010; Li et al., 2010).

Our own expression analysis of LuFLAs was consistent with the prevailing evidence that FLAs are expressed in diverse tissues and cell types. Similar to reports in arabidopsis (Johnson et al., 2003), several LuFLAs were found to be enriched in developing flowers. Similar to expression analyses in poplar and eucalyptus (Lafarguette et al., 2004; Qiu et al., 2008), where many FLAs were enriched in tension wood, many FLAs in flax were found to be enriched in the phloem fibres, an analogue of tension wood. None of the five sub-families of LuFLAs were associated exclusively with any tissue or cell wall type, with LuFLAs of group A (eg. *LuFLA1* and *LuFLA35*), B (eg. *LuFLA16* and *LuFLA32*), C1 (eg. *LuFLA30*), C2 (eg. *LuFLA7*), and D (eg. *LuFLA40* and *LuFLA41*) showing strong expression in the phloem fibres, while *LuFLA1* (A), *LuFLA7* (C2), *LuFLA10* (A), *LuFLA32* (B), *LuFLA33* (A), and *LuFLA40* (D) also expressed strongly in the developing flowers (Figure 5-4 and 5-5). We did note, however, that *LuFLA1* expression deviated from an earlier analysis of its putative promoter (Hobson and Deyholos, 2013), where, instead of displaying transcript expression solely in the phloem fibres as was observed in the promoter analysis, *LuFLA1* transcripts were detected in every tissue studied, with especially strong expression in the developing flowers and seed bolls. As expression was still predominant in the phloem fibres, we believe this data represents a limitation of β-glucuronidase fusions in analysing gene expression, where high transcript abundance, as see in the fibres (Figure 5-5) is translated into detectable protein activity, but lower levels of transcript abundance, such as in the leaves, remain below the threshold of detection by histochemistry. However, we also noted

a high level of *LuFLA1* expression in the flowers, greater than the transcript abundance observed in phloem fibres. This may indicate that additional upstream and downstream regulatory elements, than were examined in the promoter paper, influence *LuFLA1* expression. Alternatively, despite having designed primers specific to *LuFLA1*, where BLASTN searches did not align primer pairs to anything but the target genes, we may be in fact detecting the transcripts of multiple genes, perhaps due to incorrect sequence assembly/annotation.

When we examined the effectiveness of the *LuFLA35*-RNAi construct in gene silencing, we noted a decrease in *LuFLA1* transcripts (Figure 5-6), as well as *LuFLA35* transcripts (Roach, 2009), from whence the construct was derived. As the sequence of *LuFLA10* shared a high degree sequence identity with the RNAi amplicon, between that of *LuFLA1* and *LuFLA35* (Table 5-1), we suspect that it too was silenced by the construct. All three genes were distinguished by a single fasciclin domain, and appear related to the *PtFLAs 1-6*, for which ESTs were only identified in the tension wood of poplar (Lafarguette et al., 2004). *LuFLA35* had been previously noted to be enriched in the phloem fibre rich tissues of flax (Roach et al., 2007; Roach et al., 2008), while the promoter of *LuFLA1* has been found to drive gene expression specifically to the phloem fibres of flax (Hobson et al., 2013). The shared sequence similarity, and the similar expression patterns observed between *LuFLA1* and *LuFLA35*, indicates that *LuFLAs 1*, *10*, and *35* may act redundantly in flax, where the deletion or the down-regulation of any individual member may be masked by unaffected homologues. Their association with PtFLAs 1-6 also suggests that this group of FLAs diverged early in the evolution of the Malpighiales, perhaps earlier, and may be specific to the development of gelatinous type cell walls in this lineage. The presence of a highly conserved motif within the fasiciclin domain (Figure 5-3), and apparently limited to lineages within the Rosids, supports the idea of a specialized function for these genes.

Although transcripts were observed to decrease for *LuFLA1* (Figure 5-6), and *LuFLA35* (Roach, 2009), in the *LuFLA35*-RNAi transgenics, we did not observe a concomitant decrease in protein (Figure 5-7). Possibly, any decrease in protein content was masked by other FLAs as our detection system was not specific enough to quantify only these two proteins. Rather, we made use of a β-glucosyl Yariv reagent, which binds AGPs in general (Yariv et al., 1962; Seifert and

Roberts, 2007), of which FLAs are but a subset. Future characterizations of individual FLAs may thus require the preparation of protein-specific antibodies. In any case, a decrease in tensile strength was observed in the stems of the transgenic flax (Figure 5-8). Though smaller than the decrease in tensile strength observed in *LuBGAL1*-RNAi lines (Figure 5-8; Roach et al., 2010), the results are consistent with previous mutational studies of related FLAs (*AtFLA11*, *AtFLA12*; MacMillan et al., 2010).

In characterizing *AtFLA11* and *AtFLA12*, researchers noted that the reduction in stem strength coincided with reductions in cellulose content (Persson et al., 2005; MacMillan et al., 2010), arabinose and galactose (associated with the glycan moieties of the FLAs; MacMillan et al., 2010), and changes to the microfibrillar angle of the cellulose. In our quantification of the neutral monosaccharides of the cell wall non-cellulosic polysaccharides, we noted slight decreases in the levels of glucose and galactose, although they were deemed statistically insignificant when comparing the overall variation observed in individual lines (Figure 5-11). Nevertheless, the decrease in galactose may provide evidence for a decrease in FLA proteins (specifically their glycan moieties). The decrease in glucose may be explained by a decrease in cellulose, as was observed in the characterization of *AtFLA11/12* mutants (MacMillan et al., 2012). In opposition to this theory, our analysis of the FTIR data (Figure 5-12, 5-13), specifically the absorbance at 1145 cm^{-1}, indicates that the *LuFLA35*-RNAi lines may have increased cellulose, as their absorption at this wavelength exceeds both the negative sibling and wild-type Norlin. Alternatively, the increase in absorption at 1145 cm^{-1} may reflect an increased concentration of cellulose in relation to other cell wall materials, and not an increase in the absolute quantity of cellulose in the transgenic cell walls. Further analyses of the cellulose in the transgenic cell walls are ongoing, and we eagerly await the results.

As a subset of the AGP family of hydroxyproline-rich glycoproteins (HRGPs), FLAs are known to be highly glycosylated. In some cases, AGP ratios of glycan to protein can exceed 9:1 (Seifert and Roberts, 2007). AGPs generally have a distinct pattern of glycosylation consisting of repeating glycan modules (Tan et al., 2004; Seifert and Roberts, 2007); however, the AGP glycan chains can themselves be covalently attached to other cell wall polysaccharides, such as pectin

(Tan et al., 2013). Currently, the ultimate function of AGPs is disputed, where AGPs may act as a signalling molecules (Motose et al., 2004), as hemicellulose cross-linking agents (Tan et al., 2013), as recyclable transporters of cell wall polysaccharides (Burton et al., 2010), or as calcium capacitors in a greater calcium signalling system (Lamport and Várnai, 2013).

The fasciclin domain is known for adhesion to the cell membrane (Zinn et al., 1988; Huber and Sumper, 1994), and plant FLAs have been observed to localize to the cell membrane (Li et al., 2010), suggesting that, whatever the function of the glycan moieties, the LuFLAs do interact with the cell membrane. While the current view is that cell wall polysaccharides are produced in their entirety with the Golgi, alternative theories suggest that hemicelluloses are initially anchored to a lipid or protein, which in turn anchor the cell wall material to the cell membrane as it gets deposited in the extracellular space (Burton et al., 2010). Flax fibres are themselves known for the deposition of thick bands of large pectic polysaccharides during secondary cell wall deposition, and the expression of *LuFLA1* nominally coincides with the beginning of the deposition of this compound (Gorshkova and Morvan, 2006; Hobson and Deyholos, 2013), suggesting that perhaps the secretion of the two compounds are related, with the fasciclin domain acting to anchor the FLA and a bound pectin to the cell membrane to aid in its deposition and dispersal along the cell membrane. However, while a minute decrease in cell wall galactose was observed in our transgenic lines (Figure 5-11), we do not believe it was significant enough to be attributed to a decrease in the deposition of pectin in the cell walls, thus ruling out this particular theory.

We have yet to explain the sporadic occurrence fibres found to house additional internal fibres (Figure 5-10 c,d). Assuming that LuFLAs adhere to the cell membrane, a function of their fasciclin domain (Zinn et al., 1988; Huber and Sumper, 1994; Li et al., 2010), and that their glycan moities connect to the hemicelluloses of the cell wall, either covalently (Tan et al., 2013), or via calcium bridge as is observed in pectins (Albersheim et al., 2011; Lamport and Várnai, 2013), then it is possible that FLAs act as anchors of the cell membrane to the cell wall. The *LuFLA35*-RNAi construct nominally suppresses two *LuFLA*s, *LuFLA1* and *LuFLA35*, but may also suppress additional FLAs such as *LuFLA10*. These sporadic fibres may be representative of episodes of

complete suppression, were no connection is maintained between the cell membranes and their cell walls, essentially producing protoplasts encased in an alien cell wall. While phloem fibres nominally cease cell division early in their development, elongating during symplastic growth while their neighbors undergo cell divisions, they are also multinucleate, having undergone multiple karyokinesis events throughout their development (Snegireva et al., 2010). Protoplasts can be induced to re-enter the cell cycle under the appropriate conditions (Tréhin et al., 1998), and as phloem fibres are already multinucleate, a disengagement from the cell wall may be the necessary cue for resetting the cell cycle. New daughter cells, perhaps no longer experiencing a complete suppression of their FLA genes might thus undergo a new round of cell wall deposition, producing the odd fibres we have observed.

Conclusion

Forty-three putative fasciclin-like arabinogalactan proteins were identified in the genome of *Linum usitatissimum*. Using a combination of microarray and qRT-PCR data, we were able to detect expression for 42 of the 43 genes, the majority of which were expressed in the fibres and developing flowers of flax. The suppression of *LuFLA35* and *LuFLA1* resulted in a decrease in the tensile strength of flax stems in only one of three transgenic lines, as well as a rare sporadic appearance of nested phloem fibres. Further characterization is necessary to better determine their precise function in flax development.

Figures and Tables

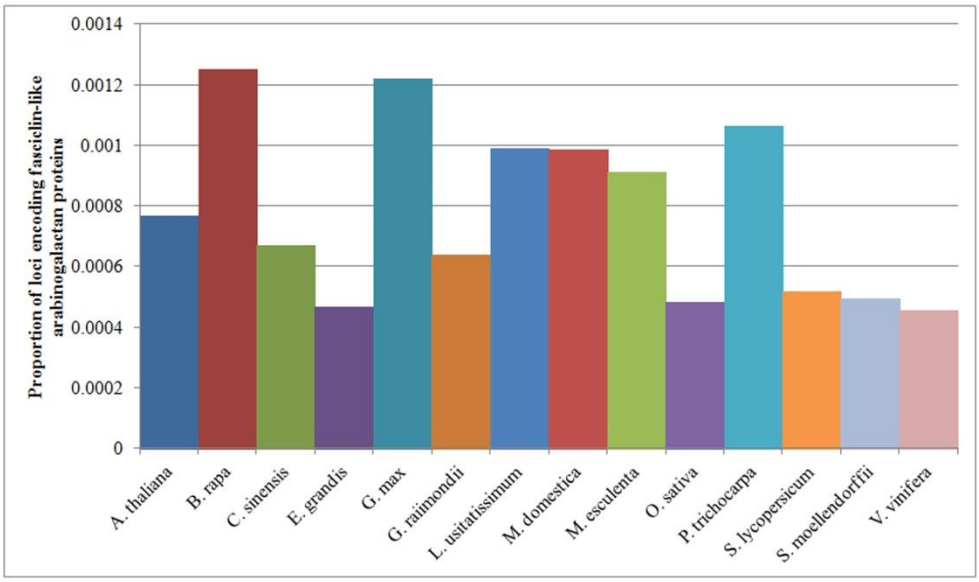

Figure 5-1 Relative quantity of FLA genes in the genomes of selected plant species. Predicted proteomes for *Arabidopsis thaliana*, *Brassica rapa*, *Citrus sinensis*, *Eucalyptus grandis*, *Glycine max*, *Gossypium raimondii*, *Linum usitatissimum*, *Malus domestica*, *Manihot esculenta*, *Oryza sativa*, *Populus trichocarpa*, *Selaginella moellendorffii, Solanum lycopersicum*, and *Vitis vinifera* were obtained from Phytozome (version 9.1; phytozome.net) with only one representative protein per locus. Sequences were assessed via Hidden Markov Model (HMM) with HMMER3 (http://hmmer.janelia.org), using the Pfam-A family database (version 25.0; Punta et al., 2012), for genes putatively encoding a fasciclin domain. The number of putative FLA genes was compared to the total number of protein coding loci published for each species at Phytozome (version 9.1).

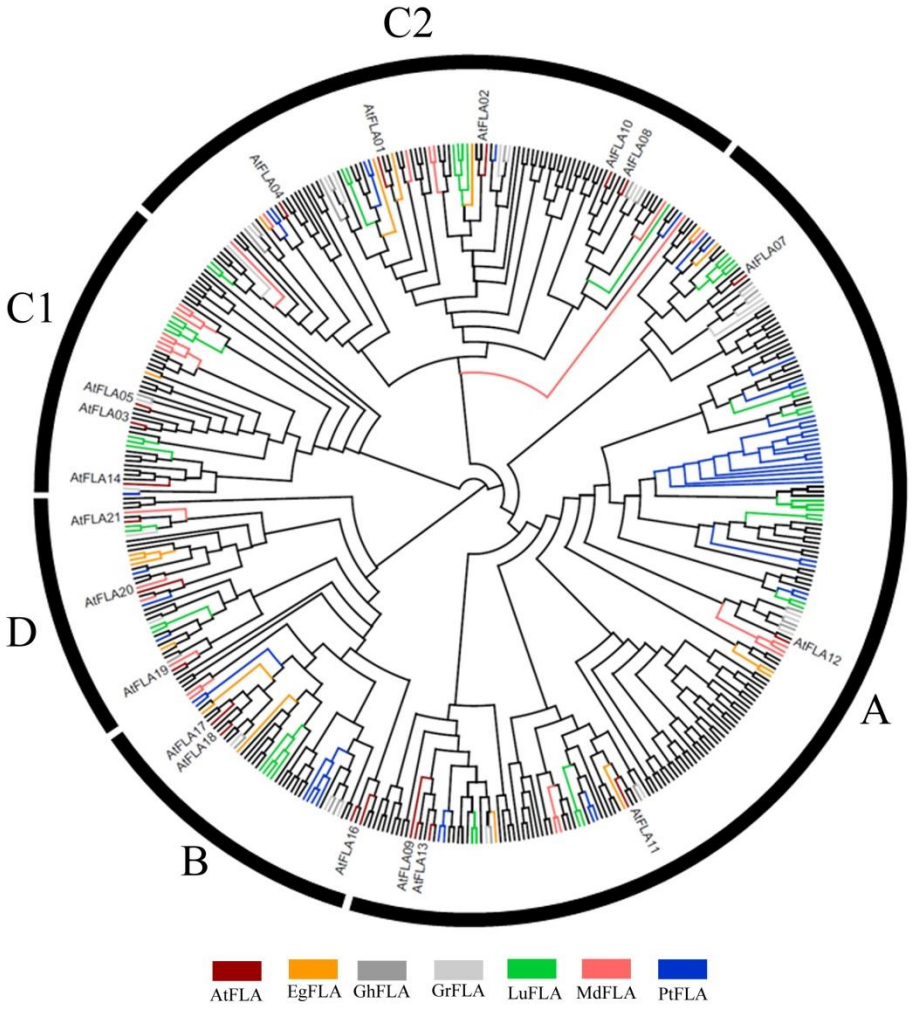

Figure 5-2 Phylogenetic relationship among the fasciclin-like arabinogalactan proteins of flax. Deduced amino acid sequences were aligned with MUSCLE (Edgar, 2004). The tree was created with GARLI (Zwickl, 2006), using the maximum likelihood method, following the WAG model of amino acid substitutions (Whelan and Goldman, 2001), while employing gamma-distributed rate variations and empirically determined base frequencies. A consensus tree of 1000 bootstrap replicates was produced. The flax sequences are named LuFLA, and numbered according to Table 5-1. *Arabidopsis thaliana* sequences are indicated as AtFLA, and numbered according to existing designations (Johnson et al., 2003). *Brassica rapa* sequences are indicated as BrFLA, and numbered according to existing designations (Jun and Xiaoming, 2012). *Oryza sativa* sequences are indicated as OsFLA, and numbered according to existing designations (Ma and Zhao, 2010). *Populus trichocarpa* sequences are indicated as PtFLA, and numbered according to existing designations (Lafarguette et al., 2004). Genomic loci corresponding to these and other sequences are presented in Table S5.

Figure 5-3 Sequence logo of a Malpighiales motif.

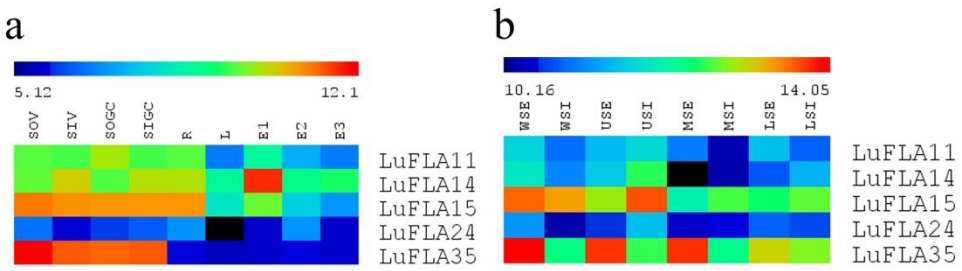

Figure 5-4 Transcript abundance of flax FLA genes in various tissues, from previously published microarray data sets (Nimblegen platform). RMA-normalized, average log2 signal values of flax FLAs in various tissues were used to produce a heat map. **a**: roots (R); leaves (L); outer stem tissues at either the vegetative stage (SOV) or green capsule stage (SOGC); inner stem tissues at either vegetative stage (SIV) or green capsule stage (SIGC); and seeds 10-15 days after flowering (DAF; E1), 20-30 DAF (E2), and 40-50 DAF (E3; Fénart et al., 2010). **b**: internal stem tissues of either the whole stem (WSI), upper stem (USI), middle stem (MSI), or lower stem (LSI); and external (i.e. phloem and cortex enriched) stem tissues of the whole stem (WSE), upper stem (USE), middle stem (MSE), and lower stem (LSE; Huis et al., 2012).

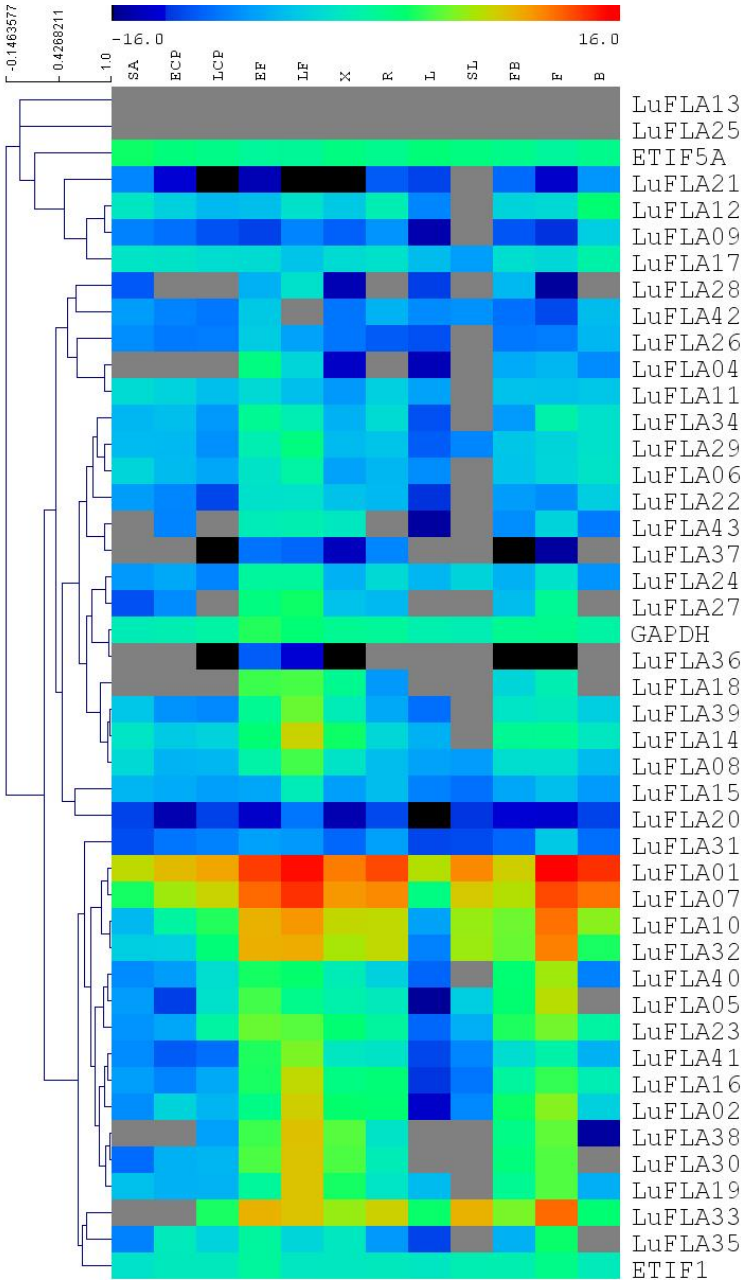

Figure 5-5 Transcript abundance of flax FLA genes in various tissues, by qRT-PCR (Fluidigm platform). Expression levels (log_2), relative to the reference genes *ETIF1* (eukaryotic translation initiation factor 1), *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase), and *ETIF5A* (eukaryotic translation initiation factor 5A), were used to prepare a heat map, with blue indicating lower expression and red indicating high expression. Gray indicates no detectable expression. Hierarchical clustering of the genes, by Pearson correlation, was applied. Tissue types analysed include: roots (R); leaves (L); senescing leaves (SL); stem apex (SA); cortical peels from vegetative stage stems (ECP) or green capsule stage stems (LCP); phloem fibres from vegetative stage stems (EF) or green capsule stage stems (LF); xylem from vegetative stage stems (X); budding flowers (FB); open flowers (F); and seed bolls from the green capsule stage (B).

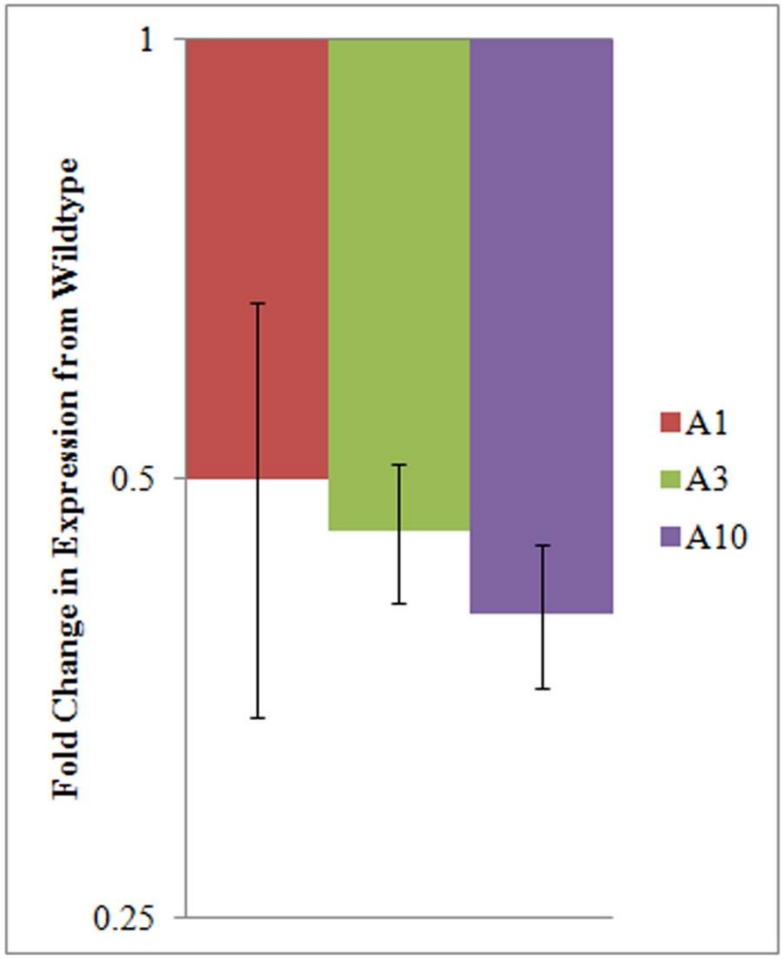

Figure 5-6 ***LuFLA1*** **transcript abundance in transgenic flax**. Expression levels, relative to the reference gene *ETIF1* (eukaryotic translation initiation factor 1), were analyzed in the leaf tissue of wild-type and transgenic flax. Depicted values are relative to *LuFLA1* expression in wild-type flax.

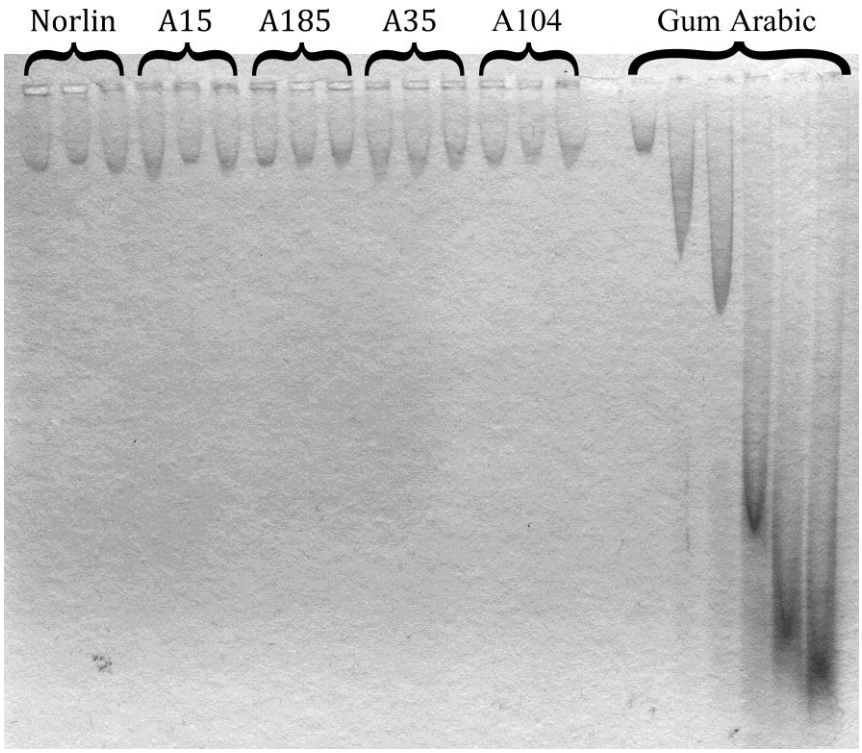

Figure 5-7 AGP abundance in the cortical peels of transgenic flax. Equal amounts of total protein were loaded into the rocket electrophoresis gel. The quantity of AGP can be estimated based on a standard curve of Gum Arabic, present in quantities of 0.2, 0.4, 0.8, 1.6, 2.4, and 3.2 µg (protein weight). No difference in AGP concentration was observed between the different lines (p-value = 0.103).

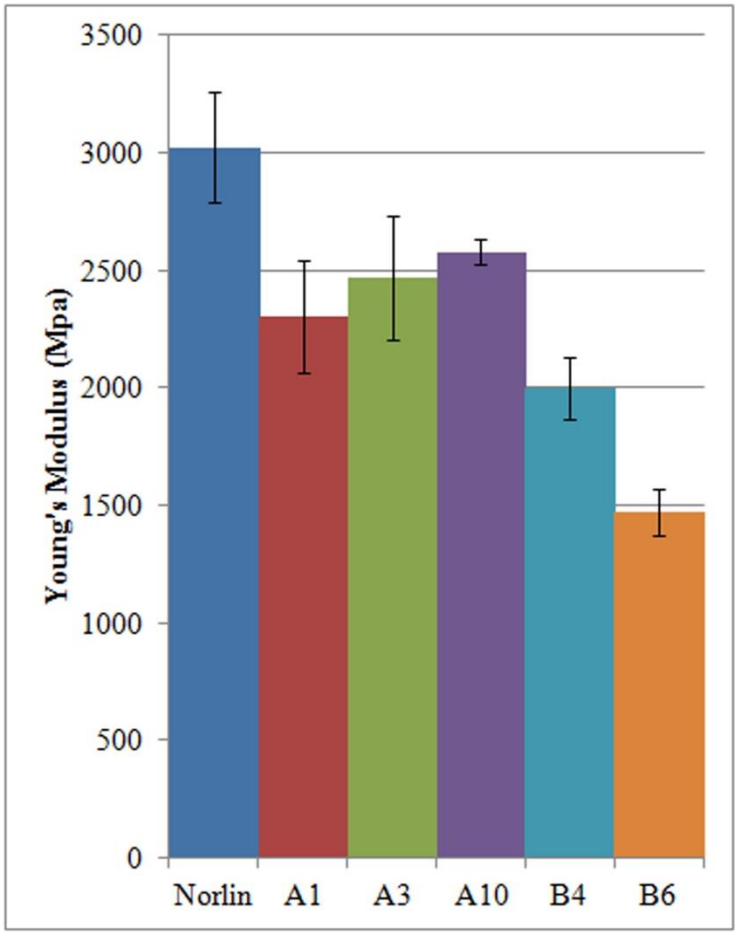

Figure 5-8 Tensile strength of transgenic flax stems. 15 cm stem sections were torn apart with an Instron 5565. Young's modulus was calculated for every measured stem. Transgenic lines A1, A3, and A10 are derived from the LuFLA RNAi construct. Transgenic lines B4 and B6 are derived from *LuBGAL1* RNAi construct (Roach et al., 2011). Only transgenic line A1 was significantly different from wild-type Norlin (p-value = 0.04).

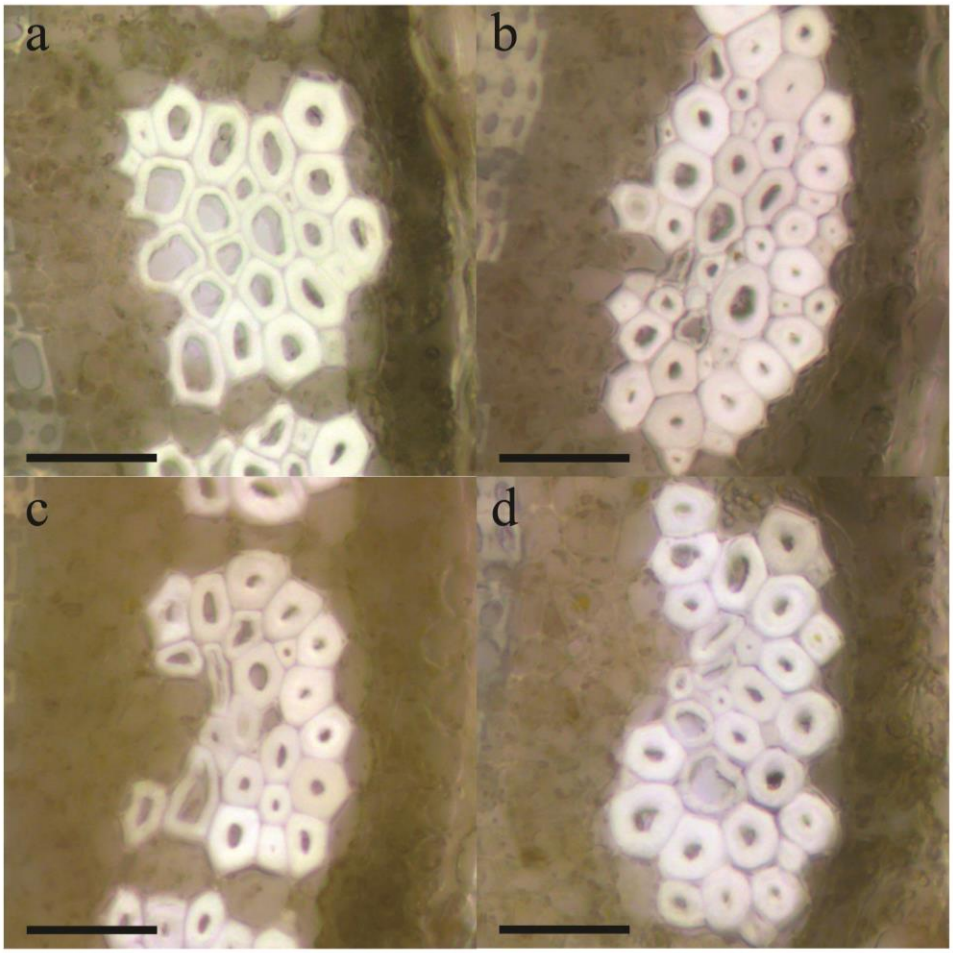

Figure 5-9 Flax phloem fibre bundles of transgenic flax. Transverse sections 16 cm below the first inflorescence of senescent flax were examined in wildtype plants (**a**), A1 transgenics (**b**), A3 transgenics (**c**), and A10 transgenics (**d**). *Scale bar* is 50 μm.

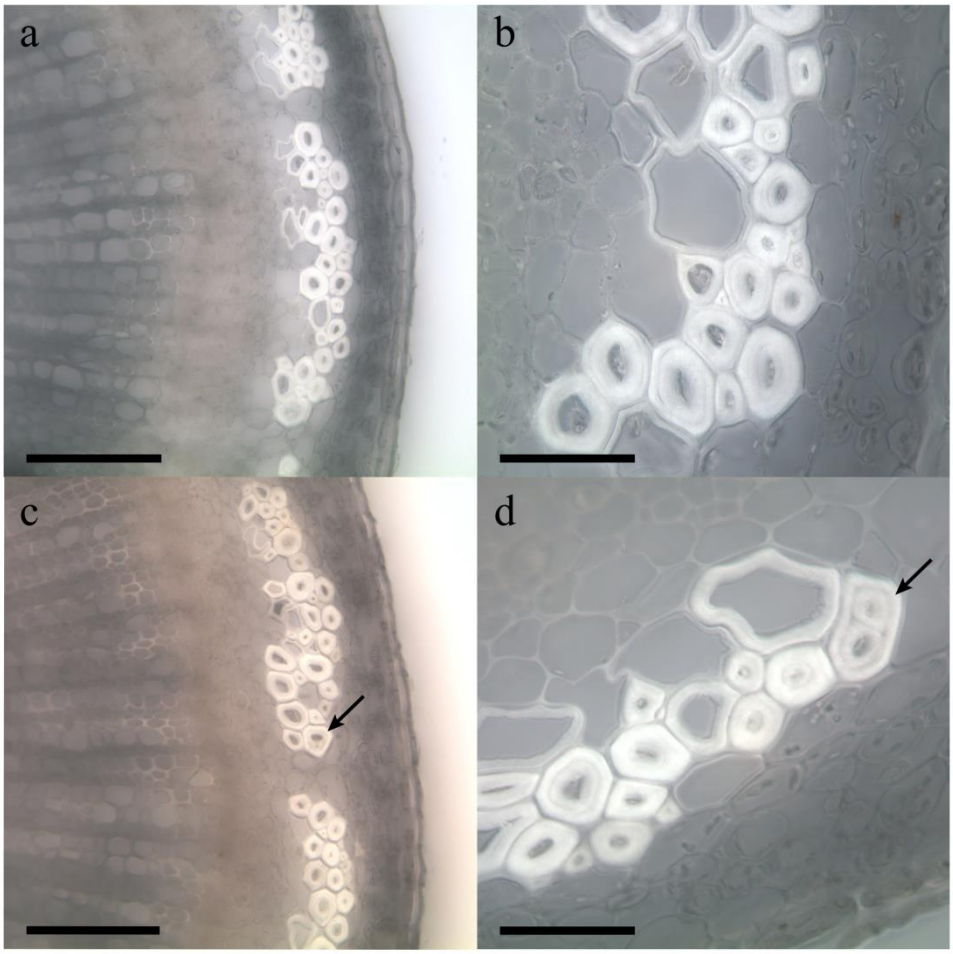

Figure 5-10 Abnormal flax fibres in transgenic flax. Transverse sections 13 cm above the cotyledons of senescent flax were examined in wildtype plants (**a**, **b**), and transgenic flax (**c**, **d**). Black arrows indicate abnormal fibres. Scale bar for **a**, and **c** is 150 μm. *Scale bar* for **b**, and **d** is 50 μm.

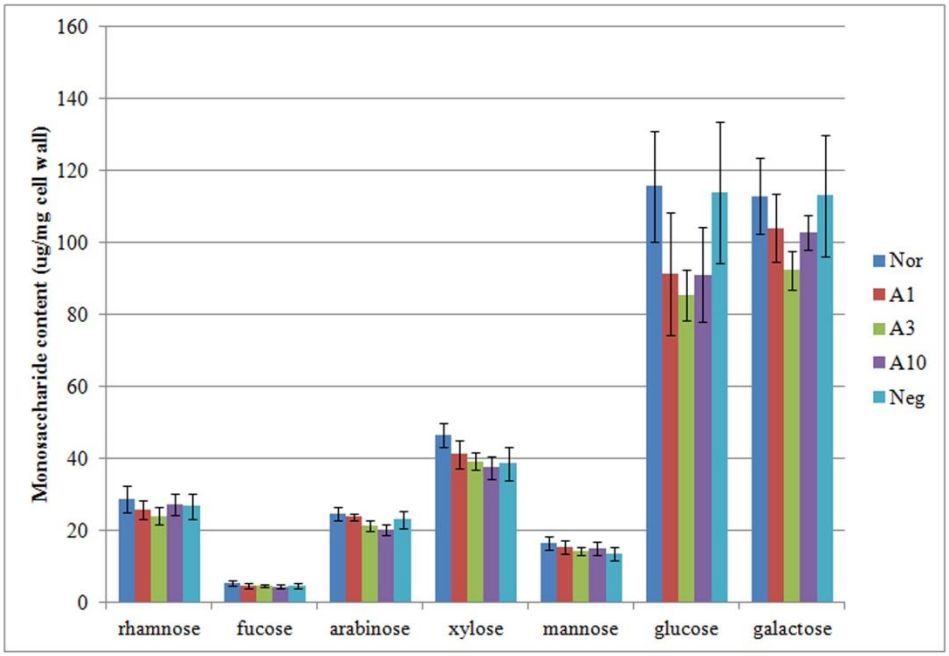

Figure 5-11 Monosaccharide composition of cell wall hemicelluloses in the cortical peels of flax stems. Non-cellulosic cell wall polysaccharides, obtained from cortical peels of 16 cm stem fragments just below the snap-point of 5-week old plants, were hydrolyzed and alditol acetate-derivatized. The derivatized neutral sugars were then analyzed by GC-MS. Inositol was added prior to hydrolysis as an internal control. The experiment was repeated twice, using three biological replicates each. The data is presented as mean monosaccharide concentration ± S.E (μg/mg of cell wall).

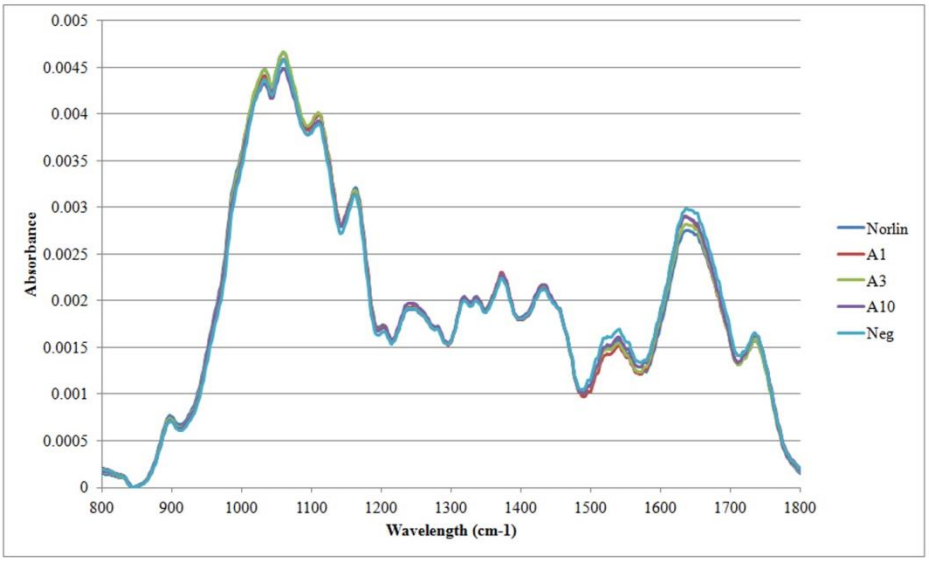

Figure 5-12 FTIR spectra of cortical peel cell walls in transgenic flax. Absorbance values are the average of six biological replicates.

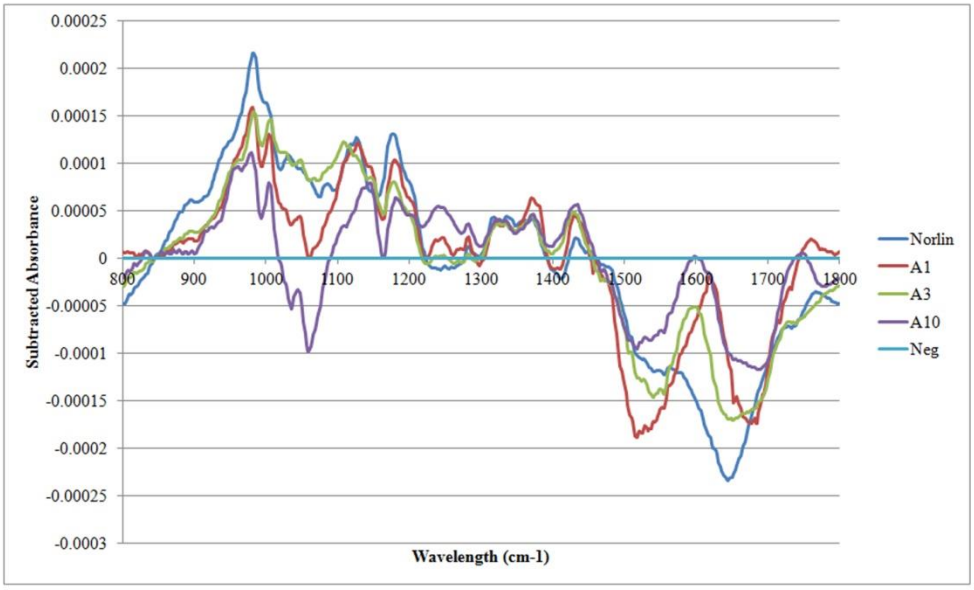

Figure 5-13 Digital subtraction of FTIR spectra of cortical peel cell walls in transgenic flax. Negative sibling absorbance values across the wavelengths were subtracted from the absorbance values of each transgenic line, as well as wild-type Norlin. Absorbance values are the average of six biological replicates.

Gene Name	Genomic Contig	Gene ID	ESTs[a]	%Identity[b]
LuFLA01	scaffold1856	Lus10002984	3	94.4
LuFLA02	scaffold1856	Lus10002985	7	69.4
LuFLA03	scaffold1856	Lus10002986	2	61.5
LuFLA04	scaffold4420	Lus10000415	0	43.6
LuFLA05	scaffold560	Lus10001178	0	45.3
LuFLA06	scaffold561	Lus10001733	0	58.1
LuFLA07	scaffold561	Lus10001752	0	45.1
LuFLA08	scaffold1478	Lus10002273	1	55.9
LuFLA09	scaffold1856	Lus10002978	3	68.1
LuFLA10	scaffold1856	Lus10002979	0	96.1
LuFLA11	scaffold1603	Lus10006391	2	40.5
LuFLA12	scaffold1603	Lus10006399	0	50.9
LuFLA13	scaffold736	Lus10007404	0	61.5
LuFLA14	scaffold267	Lus10009235	12	48.3
LuFLA15	scaffold572	Lus10012344	5	38.3
LuFLA16	scaffold572	Lus10012351	1	34.3
LuFLA17	scaffold62	Lus10013604	6	45.4
LuFLA18	scaffold947	Lus10016194	1	45.5
LuFLA19	scaffold279	Lus10016554	4	45.9
LuFLA20	scaffold915	Lus10016617	0	41.5
LuFLA21	scaffold765	Lus10017696	1	54.0
LuFLA22	scaffold491	Lus10019929	1	62.4
LuFLA23	scaffold454	Lus10020158	1	45.8
LuFLA24	scaffold641	Lus10020296	0	37.7
LuFLA25	scaffold123	Lus10021247	5	41.7
LuFLA26	scaffold465	Lus10022517	0	51.3
LuFLA27	scaffold353	Lus10024089	1	46.6
LuFLA28	scaffold319	Lus10026107	0	35.4
LuFLA29	scaffold617	Lus10026499	1	61.9
LuFLA30	scaffold651	Lus10026957	0	40.6
LuFLA31	scaffold327	Lus10028268	0	40.3
LuFLA32	scaffold360	Lus10029353	2	46.1
LuFLA33	scaffold701	Lus10033651	1	41.9
LuFLA34	scaffold76	Lus10036111	5	69.2
LuFLA35	scaffold76	Lus10036112	3	100.0
LuFLA36	scaffold76	Lus10036113	3	69.9
LuFLA37	scaffold76	Lus10036114	1	66.7
LuFLA38	scaffold76	Lus10036115	1	64.7
LuFLA39	scaffold475	Lus10038004	1	35.1
LuFLA40	scaffold15	Lus10039717	0	47.5
LuFLA41	scaffold86	Lus10040220	0	45.8
LuFLA42	scaffold156	Lus10040821	6	43.6
LuFLA43	scaffold272	Lus10041630	0	46.5

Table 5-1 Summary of fasciclin-like arabinogalactan protein gene homologues
[a]ESTs retrieved from Genbank and assembly POZS of www.onekp.com.
[b]Percent sequence identity to *LuFLA35*-RNAi amplicon according to EMBOSS Water alignment.

Group	ProteinName	AA	MW[a] (kDA)	pI[a]	Signal Peptide[b]	Fasciclin domains[c]	GPI anchor[d]
A	LuFLA08	266	27.69	6.35	Y	1	Y
	LuFLA06	265	27.655	6.54	Y	1	Y
	LuFLA12	265	27.886	5.92	Y	1	Y
	LuFLA16	266	27.92	6.15	Y	1	Y
	LuFLA03	223	24.131	6.64	N	1	Y
	LuFLA37	271	28.587	7.75	Y	1	Y
	LuFLA38	246	25.729	5.21	Y	1	Y
	LuFLA13	304	32.878	10.78	N	1	N
	LuFLA01	262	27.412	8.81	Y	1	Y
	LuFLA10	189	19.732	8.66	Y	1	Y
	LuFLA35	257	26.82	6.74	Y	1	Y
	LuFLA02	262	27.622	9.2	Y	1	Y
	LuFLA36	280	29.63	8.72	Y	1	Y
	LuFLA09	245	25.818	8.72	Y	1	N
	LuFLA34	246	25.72	9.09	Y	1	Y
	LuFLA22	241	25.239	7.47	Y	1	Y
	LuFLA29	241	25.143	6.37	Y	1	Y
	LuFLA21	247	25.487	5.5	Y	1	Y
	LuFLA33	248	25.828	4.75	Y	1	Y
B	LuFLA25	397	43.963	9.61	Y	1	Y
	LuFLA18	383	41.611	6	Y	2	N
	LuFLA32	543	59.185	7.05	Y	2	Y
	LuFLA17	439	48.384	6.63	Y	2	N
C1	LuFLA39	214	22.418	6.58	Y	1	Y
	LuFLA23	261	27.565	8.46	Y	1	Y
	LuFLA30	261	27.466	6.56	Y	1	Y
	LuFLA28	426	43.82	9.64	Y	1	Y
	LuFLA04	198	21.344	10.38	Y	1	Y
	LuFLA43	417	42.886	10.08	Y	1	Y
	LuFLA27	455	46.793	9.78	Y	1	Y
C2	LuFLA19	422	44.943	6.22	Y	2	N
	LuFLA42	434	46.122	6.29	Y	2	N
	LuFLA20	414	43.64	5.44	Y	2	Y
	LuFLA26	409	43.042	5.29	Y	2	Y
	LuFLA15	412	43.567	5.76	Y	2	N
	LuFLA07	342	36.055	5.62	N	1	Y
	LuFLA05	426	45.14	6.18	N	2	Y
	LuFLA11	462	48.629	5.46	Y	2	N
	LuFLA14	416	42.348	5.42	Y	2	Y
D	LuFLA31	196	21.937	6.47	N	1	N
	LuFLA41	130	14.498	4.81	N	1	N
	LuFLA24	411	44.249	10.14	N	2	N
	LuFLA40	381	41.172	9.61	N	2	Y

Table 5-2 Summary of predicted fasciclin-like arabinogalactan protein homologues.

[a]Predictions made with CLC Genomics Workbench 5.5.

[b]SignalP 4.0 prediction (Petersen et al., 2011).

[c]Fasciclin domains and locations were identified at pfam.sanger.ac.uk.

[d]GPI anchor attachment signal predicted with GPI-SOM (Fankhauser and Mäser, 2005).

References

Ageeva MV, Petrovska B, Kieft H, Sal'nikov VV, Snegireva AV, van Dam JEG, van Veenendaal WLH, Emons AMC, Gorshkova TA, and van Lammeren AAM (2005) Intrusive growth of flax phloem fibers is of intercalary type. Planta **222**: 565-574.

Albersheim P, Darvill A, Roberts K, Sederoff R, and Staehelin A (2011) Plant Cell Walls. Garland Science, Taylor & Francis Group, LLC, New York.

Bailey TL, and Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, Menlo Park, August.

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, Ashton NW, Axtell MJ, Barker E, Barker MS, Bennetzen JL, Bonawitz ND, Chapple C, Cheng C, Correa LG, Dacre M, DeBarry J, Dreyer I, Elias M, Engstrom EM, Estelle M, Feng L, Finet C, Floyd SK, Frommer WB, Fujita T, Gramzow L, Gutensohn M, Harholt J, Hattori M, Heyl A, Hirai T, Hiwatashi Y, Ishikawa M, Iwata M, Karol KG, Koehler B, Kolukisaoglu U, Kubo M, Kurata T, Lalonde S, Li K, Li Y, Litt A, Lyons E, Manning G, Maruyama T, Michael TP, Mikami K, Miyazaki S, Morinaga S, Murata T, Mueller-Roeber B, Nelson DR, Obara M, Oguri Y, Olmstead RG, Onodera N, Petersen BL, Pils B, Prigge M, Rensing SA, Riaño-Pachón DM, Roberts AW, Sato Y, Scheller HV, Schulz B, Schulz C, Shakirov EV, Shibagaki N, Shinohara N, Shippen DE, Sørensen I, Sotooka R, Sugimoto N, Sugita M, Sumikawa N, Tanurdzic M, Theissen G, Ulvskov P, Wakazuki S, Weng JK, Willats WW, Wipf D, Wolf PG, Yang L, Zimmer AD, Zhu Q, Mitros T, Hellsten U, Loqué D, Otillar R, Salamov A, Schmutz J, Shapiro H, Lindquist E, Lucas S, Rokhsar D, and Grigoriev IV (2011) The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science **332**:960-963.

Bazinet AL, and Cummings MP (2011) Computing the Tree of Life - Leveraging the Power of Desktop and Service Grids. Proceedings of the Fifth Workshop on Desktop Grids and Volunteer Computing Systems, Anchorage, May 20.

Burton RA, Gidley MJ, and Fincher GB (2010) Heterogeneity in the chemistry, structure and function of plant cell walls. Nature Chemical Biology **6**: 724-732. doi: 10.1038/nchembio.439

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, and Rabinowicz PD (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nature Biotechnology **28**:951-956

Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, and Wang X (2012) BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biology **11**: 136. doi: 10.1186/1471-2229-11-136

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics **27**:1164-1165

De Pauw MA, Vidmar JJ, Collins J, Bennet RA, and Deyholos MK (2007) Microarray analysis of bast fibre producing tissues of *Cannabis sativa* identifies transcripts associated with conserved and specialised processes of secondary wall development. Functional Plant Biology **34**: 737-749. doi: 10.1071/FP07014

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**: 1792-1797. doi: 10.1093/nar/gkh340

Eisenhaber B, Bork P, Yuan Y, Loeffler G, and Eisenhaber F (2000) Automated annotation of GPI anchor sites: case study *C. elegans*. Trends in Biochemical Sciences **25**: 340-341

Esau K (1965) Plant Anatomy. 2nd Ed. John Wiley & Sons Inc., New York.

Fahn A (1990) Plant Anatomy. 4th Ed. Pergamon Press, Oxford, UK.

Fankhauser N, and Mäser P (2005) Identification of GPI anchor attachment signals by a Kohonen self-organizing map. Bioinformatics **21**: 1846-1852. doi: 10.1093/bioinformatics/bti299

Fénart S, Ndong Y-PA, Duarte J, Rivière N, Wilmer J, van Wuytswinkel O, Lucau A, Cariou E, Neutelings G, Gutierrez L, Chabbert B, Guillot X, Tavernier R, Hawkins S, and Thomasset B (2010) Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray. BMC Genomics **11**: 592. doi: 10.1186/1471-2164-11-592

Gorgulu ST, Dogan M, and Severcan F (2007) The Characterization and Differentiation of Higher Plants by Fourier Transform Infrared Spectroscopy. Applied Spectroscopy **61**: 300-308. doi: 10.1366/000370207780220903

Gorshkova T, and Morvan C (2006) Secondary cell-wall assembly in flax phloem fibers: role of galactans. Planta **223**: 149-158.

Gorshkova TA, Gurjanov OP, Mikshina PV, Ibragimova NN, Mokshina NE, Salnikov VV, Ageeva MV, Amenitskii SI, Chernova TE, and Chemikosova SB (2010) Specific Type of Secondary Cell Wall Formed by Plant Fibers. Russian Journal of Plant Physiology **57**: 328-341.

Gurjanov OP, Gorshkova TA, Kabel, MA, Schols HA, and van Dam JEG (2007) MALDI-TOF MS evidence for the linking of flax bast fibre galactan to rhamnogalacturonan backbone. Carbohydrate Polymers **67**: 86-96.

Gurjanov OP, Ibragimova NN, Gnezdilov OI, and Gorshkova TA (2008) Polysaccharides, tightly bound to cellulose in the cell wall of flax bast fibre: Isolation and identification. Carbohydrate Research **72**: 719-729

Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, and Pilate G (2012) Plant Fiber Formation: State of the Art, Recent and Expected Progress, and Open Questions. Critical Reviews in Plant Sciences **31**:201-228. doi: 10.1080/07352689.2011.616096

Hobson N, Roach MJ, and Deyholos MK (2010) Gene Expression in Tension Wood and Bast Fibres. Russian Journal of Plant Physiology **57**: 321-327. doi: 10.1134/S1021443710030039

Hobson N, and Deyholos MK (2013) $LuFLA1_{pro}$ and $LuBGAL1_{pro}$ promote gene expression in the phloem fibres of flax (*Linum usitatissimum*). Plant Cell Reports **32**: 517-528. doi: 10.1007/s00299-013-1383-8

Huang GQ, Xu WL, Gong SY, Li Bing, Wang XL, Xu D, and Li XB (2008) Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. Physiologia Plantarum **134**: 348-359.

Huber O, and Sumper M (1994) Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga *Volvox* with homology to *Drosophila* fasciclin I. The EMBO Journal **13**: 4212-4222.

Huis R, Hawkins S, and Neutelings G (2010) Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). BMC Plant Biology **10**: 71. doi:10.1186/1471-2229-10-71

Huis R, Morreel K, Fliniaux O, Lucau-danila A, Fénart S, Grec S, Neutelings G, Chabbert B, Mesnard F, Boerjan W, and Hawkins S (2012) Natural Hypolignification Is Associated with Extensive Oligolignol Accumulation in Flax Stems. Plant Physiology **158**: 1893–1915. doi:10.1104/pp.111.192328

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Bilault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabin S, Canaguier A, Le Clainch I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè MP, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, and Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **440**: 463-467

Johnson KL, Jones BJ, Bacic A, and Schultz CJ (2003) The Fasciclin-Like Arabinogalactan Proteins of *Arabidopsis*. A Multigene Family of Putative Cell Adhesion Molecules. Plant Physiology **133**: 1911-1925. doi: 10.1104/pp.103.031237

Johnson MTJ, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, dePamphilis CW, Edger PP, Goh F, Graham S, Greiner S, Hibberd JM, Jordon-Thaden I, Kutchan TM, Leebens-Mack J, Melkonian M, Miles N, Myburg H, Patterson J, Pires JC, Ralph P, Rolf M, Sage RF, Soltis D, Soltis P, Stevenson D, Stewart CN, Surek B, Thomsen CJM, Villarreal JC, Wu X, Zhang Y, Deyholos MK, and Wong GKS (2012) Evaluating Methods for Isolating Total RNA and Predicting the Success of Sequencing Phylogenetically Diverse Plant Transcriptomes. PLoS ONE **7**: e50226.

Jun L, and Xiaoming W (2012) Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.). Molecular Biology Reports **39**: 10541-10555. doi: 10.1007/s11033-012-1940-1

Lafarguette F, Leplé, Déjardin A, Laurans F, Costa G, Lesage-Descauses MC, and Pilate G (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. New Phytologist **164**: 107-121. doi: 10.1111/j.1469-8137.2004.01175.x

Lamport DTA, and Várnai P (2013) Periplasmic arabinogalactan glycoproteins act as a calcium capacitor that regulates plant growth and development. New Phytologist **197**: 58-64. doi: 10.1111/nph.12005

Li J, Miao Y, Geng LL, and Zhao J (2010) The fasciclin-like arabinogalactan protein gene, *FLA3*, is involved in microspore development of Arabidopsis. The Plant Journal **64**: 482-497. doi: 10.1111/j.1365-313X.2010.04344.x

Li X, Wu HX, Dillon SK, and Southerton SG (2009) Generation and analysis of expressed sequence tags from six developing xylem libraries in *Pinus radiata* D. Don. BMC Genomics **10**: 41. doi: 10.1186/1471-2164-10-41

Ma H, and Zhao J (2010) Genome-wise identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). Journal of Experimental Botany **61**: 2647-2668. doi: 10.1093/jxb/erq104

MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, and Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and *Eucalyptus*. The Plant Journal **62**: 689-703. doi: 10.1111/j.1365.313X.2010.04181.x

Mikshina PV, Gurjanov OP, Mukhitova FK, Petrova AA, Shashkov AS, and Gorshkova TA (2012) Structural details of pectic galactan from the secondary cell walls of flax (*Linum usitatissimum* L.) phloem fibres. Carbohydrate Polymers **87**: 853-861. doi: 10.1016/j.carbpol.2011.08.068

Miller MA, Pfeiffer W, and Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), New Orleans, Nov. 14. doi: 10.1109/GCE.2010.5676128

Mohanty AK, Misra M, and Hinrichsen G (2000) Biofibres, biodegradable polymers and biocomposites: an overview. Macromolecular Materials and Engineering **276**: 1-24.

Motose H, Sugiyama M, and Fukuda H (2004) A proteoglycan mediates inductive interaction during plant vascular development. Nature **429**: 873-878.

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Research **35**: D883-887.

Persson S, Wei H, Milne J, Page GP, and Somerville CR (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. PNAS **102**: 8633-8638. doi: 10.1073/pnas.0503392102

Petersen TN, Brunak S, von Heijne G, and Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods **8**: 785-786. doi: 10.1038/nmeth.1701

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, and Rounsley S (2012) The Cassava Genome: Current Progress, Future Directions. Tropical Plant Biology **5**:88-94

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, and Finn RD (2012) The Pfam protein families database. Nucleic Acids Research **40**: D290–301. doi:10.1093/nar/gkr1065

Qiu D, Wilson IW, Gan S, Washusen R, Moran GF, and Southerton SG (2008) Gene expression in *Eucalyptus* branch wood with marked variation in cellulose microfibril orientation and lacking G-layers. New Phytologist **179**: 94-103. doi: 10.1111/j.1469-8137.2008.02439.x

Roach MJ, and Deyholos MK (2007) Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. Molecular Genetics and Genomics **278**: 149-165. doi: 10.1007/s00438-007-0241-1

Roach MJ, and Deyholos MK (2008) Microarray Analysis of Developing Flax Hypocotyls Identifies Novel Transcripts Correlated with Specific Stages of Phloem Fibre Differentiation. Annals of Botany **102**: 317-330. doi: 10.1093/aob/mcn110

Roach M (2009) Examining the Molecular Basis of Bast Fibre Development in Flax (*Linum usitatissimum* L.). Ph.D. Thesis. University of Alberta: Canada.

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, and Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant Physiology **156**: 1351–1363. doi:10.1104/pp.111.172676

Salnikov VV, Ageeva MV, and Gorshkova TA (2008) Homofusion of Golgi secretory vesicles in flax phloem fibres during formation of the gelatinous secondary cell wall. Protoplasma **233**: 269-273. doi: 10.1007/s00709-008-0011-x

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, Egholm M, Knight J, Bogden R, Li C, Shuang Y, Xu X, Pan S, Cheng S, Liu X, Ren Y, Wang J, Albiero A, Dal Pero F, Todesco S, Van Eck J, Buels RM, Bombarely A, Gosselin JR, Huang M, Leto JA, Menda N, Strickler S, Mao L, Gao S, Tecle IY, York T, Zheng Y, Vrebalov JT, Lee J, Zhong S, Mueller LA, Stiekema WJ, Ribeca P, Alioto T, Yang W, Huang S, Du Y, Zhang Z, Gao J, Guo Y, Wang X, Li Y, He J, Li C, Cheng Z, Zuo J, Ren J, Zhao J, Yan L, Jiang H, Wang B, Li H, Li Z, Fu F, Chen B, Feng Q, Fan D, Wang Y, Ling H, Xue Y, Ware D, McCombie WR, Lippman ZB, Chia JM, Jiang K, Pasternak S, Gelley L, Kramer M, Anderson LK, Chang SB, Royer SM, Shearer LA, Stack SM, Rose JK, Xu Y, Eannetta N, Matas AJ, McQuinn R, Tanksley SD, Camara F, Guigó R, Rombauts S, Fawcett J, Van de Peer Y, Zamir D, Liang C, Spannagl M, Gundlach H, Bruggmann R, Mayer K, Jia Z, Zhang J, Ye Z, Bishop GJ, Butcher S, Lopez-Cobollo R, Buchan D, Filippis I, Abbott J, Dixit R, Singh M, Singh A, Pal JK, Pandit A, Singh PK, Mahato AK, Gaikwad VD, Sharma RR, Mohapatra T, Singh NK, Causse M, Rothan C, Schiex T, Noirot C, Bellec A, Klopp C, Delalande C, Berges H, Mariette J, Frasse P, Vautrin S, Zouine M, Latché A, Rousseau C, Regad F, Pech JC, Philippot M, Bouzayen M, Pericard P, Osorio S, Fernandez del Carmen A, Monforte A, Granell A, Fernandez-Muñoz R, Conte M, Lichtenstein G, Carrari F, De Bellis G, Fuligni F, Peano C, Grandillo S, Termolino P, Pietrella M, Fantini E, Falcone G, Fiore A, Giuliano G, Lopez L, Facella P, Perotta G, Daddiego L, Bryan G, Orozco M, Pastor X, Torrents D, van Schriek MG, Feron RM, van Oeveren J, de Heer P, daPonte L, Jacobs-Oomen S, Cariaso M, Prins M, van Eijk MJ, Janssen A, van Haaren MJ, Jungeun Kim SH, Kwon SY, Kim S, Koo DH, Lee S, Hur CG, Clouser C, Rico A, Hallab A, Gebhardt C, Klee K, Jöcker A, Warfsmann J, Göbel U, Kawamura S, Yano K, Sherman JD, Fukuoka H, Negoro S, Bhutty S, Chowdhury P, Chattopadhyay D, Datema E, Smit S, Schijlen EG, van de Belt J, van Haarst JC, Peters SA, van Staveren MA, Henkens MH, Mooyman PJ, Hesselink T, van Ham RC, Jiang G, Droege M, Choi D, Kang BC, Kim BD, Park M, Kim S, Yeom SI, Lee YH, Choi YD, Li G, Gao J, Liu Y, Huang S, Fernandez-Pedrosa V, Collado C, Zuñiga S, Wang G, Cade R, Dietrich RA, Rogers J, Knapp S, Fei Z, White RA, Thannhauser TW, Giovannoni JJ, Botella MA, Gilbert L, Gonzalez R, Goicoechea JL, Yu Y, Kudrna D, Collura K, Wissotski M, Wing R, Meyers BC, Gurazada AB, Green PJ, Vyas SM, Solanke AU, Kumar R, Gupta V, Sharma AK, Khurana P, Khurana JP, Tyagi AK, Dalmay T, Mohorianu I, Walts B, Chamala S, Barbazuk WB, Li J, Guo H, Lee TH, Wang Y, Zhang D, Paterson AH, Wang X, Tang H, Barone A, Chiusano ML, Ercolano MR, D'Agostino N, Di Filippo M, Traini A, Sanseverino W, Frusciante L, Seymour GB, Elharam M, Fu Y, Hua A, Kenton S, Lewis J, Lin S, Najar F, Lai H, Qin B, Qu C, Shi R, White D, White J, Xing Y, Yang K, Yi J, Yao Z, Zhou L, Roe BA, Vezzi A, D'Angelo M, Zimbello R, Schiavon R, Caniato E, Rigobello C, Campagna D, Vitulo N, Valle G, Nelson DR, De Paoli E, Szinay D, de Jong HH, Bai Y, Visser RG, Klein R, Beasley H, McLaren K, Nicholson C, Riddle C, and Gianese G (2012) The tomato genome sequence provides insights into flesh fruit evolution. Nature **485**: 635-641.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman

A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, and Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature **463**:178-183

Schmittgen TD and Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nature Protocols **3**: 1101–1108

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, and Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science **326**:1112-1115.

Seifert GJ, and Roberts K (2007) The Biology of Arabinogalactan Proteins. Annual Review of Plant Biology **58**: 137-61. doi: 10.1146/annurev.arplant.58.032806.103801

Shi H, Kim Y, Guo Y, Stevenson B, and Zhu JK (2003) The Arabidopsis *SOS5* Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. The Plant Cell **15**: 19-32. doi: 10.1105/tpc.007872

Snegireva AV, Ageeva MV, Amenitskii SI, Chernova TE, Ebskamp M, and Gorshkova TA (2010) Intrusive Growth of Sclerenchyma Fibers. Russian Journal of Plant Physiology **57**: 342-355. doi: 10.1134/S1021443710030052

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbarugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Research **36:** D1009-1014

Tan L, Qiu F, Lamport DTA, and Kieliszewski MJ (2004) Structure of a Hydroxyproline (Hyp)-Arabinogalactan Polysaccharide from Repetitive Ala-Hyp Expressed in Transgenic *Nicotiana tabacum*. The Journal of Biological Chemistry **279**: 13156-13165. doi: 10.1074/jbc.M3118642000

Tan L, Eberhard S, Pattathil S, Warder C, Glushka J, Yuan C, Hao Z, Zhu X, Avci U, Miller JS, Baldwin D, Pham C, Orlando R, Darvill A, Hahn MG, Kieliszewski MJ, and Mohnen D (2013) An *Arabidopsis* Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein. The Plant Cell **25**: 270-287. doi: 10.1105/tpc.112.107334

Tréhin C, Planchais S, Glab N, Perennes C, Tregear J, and Bergounioux C (1998) Cell cycle regulation by plant growth regulators: involvement of auxin and cytokinin in the re-entry of *Petunia* protoplasts into the cell cycle. Planta **206**: 215-224

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**: 1596-1604

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, and Viola R (2010) The genome of the domesticated apple (*Malus domestica* Borkh.). Nature Genetics **42**:833-839.

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, and Datla R (2011) Gene expression analysis of flax seed development. BMC Plant Biology **11**:74. doi: 10.1186/1471-2229-11-74

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GWS, Wang J, and Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. The Plant Journal **72**: 461-473. doi: 10.1111/j.1365-313X.2012.05093.x

Whelan S, and Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution **18**: 691–699.

Yariv J, Rapport MM, and Graf L (1962) The Interaction of Glycosides and Saccharides with Antibody to the Corresponding Phenylazo Glycosides. Biochemical Journal **85**: 383-388

Zinn K, McAllister L, and Goodman CS (1988) Sequence Analysis and Neuronal Expression of Fasciclin I in Grasshopper and Drosophila. Cell **53**: 577-587

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. thesis. The University of Texas, USA.

Chapter 6 – Concluding Remarks

Development of Genetic Resources for Flax

At the start of my thesis, genetic resources for flax were limited to a tissue specific cDNA library, and a corresponding microarray comprising 9600 cDNAs. In order to obtain genomic material necessary for the cloning and study of genes and promoter elements, I produced a fosmid library comprising 71,522 clones with a genomic insert size ~36 kb; a 7 X coverage of the whole genome. As a result, I was able to screen for the genomic sequence of two flax genes, of which one led to the isolation of nucleotide sequences capable of directing gene expression to the phloem fibres of the flax stem, of potentially great value in future biotechnological endeavors (Hobson and Deyholos, 2013). Since then, additional genetic resources have become available including additional EST libraries encompassing 93% of all predicted flax genes (Venglat et al., 2011; Wang et al., 2012), more comprehensive microarrays (Fénart et al., 2010), libraries of polymorphic markers for use in mapping studies (Cloutier et al., 2012; Kumar et al., 2012), and a *de novo* short-read assembly of the whole genome to which I contributed sequenced fosmids, fosmid end sequences, sequenced BACs, and the genomic DNA sample sequenced for the assembly (Wang et al., 2012). There is therefore now great potential for crop improvement and fundamental discoveries in flax, for example through the combination of genomic information with existing EMS mutant populations, using reverse genetics. However, detailed characterization of target genes in the context of their gene families is an important first step towards these applications.

Gene family analyses

In studying the β-galactosidase family of flax, specifically the genes encoding glycosyl hydrolase 35 domains, several phenomena were observed. First, we noted that, in relation to related species and most sequenced plant genomes, the BGAL family of flax was very large, with gene copies throughout the many BGAL sub-families, suggestive of a recent genome duplication, which has been observed elsewhere (Wang et al., 2012). Duplications aside, two BGAL sub-

families stood out as being of unusual composition. First, the size of sub-family A5 was found to have greatly expanded, in comparison to other plant species. AtBGAL10, the sole arabidopsis member of sub-family A5, is considered the primary xyloglucan BGAL of arabidopsis (Sampedro et al., 2012), therefore, assuming a similar substrate specificity, the expansion of this family in flax would suggest that the usage of xyloglucans in cell wall development has expanded, necessitating additional BGALs for downstream processing. However, not only was sub-family A5 observed to have expanded in flax, but sub-family B was also found to be greatly reduced, with only a single flax representative found, despite multiple representatives in the other examined species. No BGAL of sub-family B has been characterized to date; therefore we cannot yet infer how the composition of flax cell walls may have changed as the sub-family reduced in size. It may be that the expansion observed in sub-family A5 has allowed its members to take up functions normally attributed to sub-family B BGALs. Regardless, the observed shifts in the sizes of glycosyl hydrolase families, such as BGAL sub-families B and A5, offer important targets for further research towards further understanding the evolution of cell walls and gene families in flax.

In studying the fasciclin-like arabinogalactan protein (FLA) family of flax we did not observe an expansion of the gene family, in relation to the FLAs of other species. However, several LuFLAs found to be associated with phloem fibre development (*LuFLA1* and *LuFLA35*), were found to be associated with poplar FLAs enriched in tension wood (Lafarguette et al., 2004), a tissue with a similar cell wall as the phloem fibres of flax. These genes were themselves found in a large clade that appeared specific to the Malpighiales, indicative of a specialization within the lineage, perhaps associated with the development of gelatinous (G) type cell walls. The presence of a unique motif in the fasciclin domain of these genes (Figure 5-3) is indicative of a specialized function, although further study will be required to determine both the role of plant fasciclin domains, and then the effects of the observed differences.

Transgenic analyses

In assessing the impact of over-expressed *LuBGAL1* in flax, we found that an increase in *LuBGAL1* transcript abundance did not lead to any obvious changes to the composition of cell

walls, the appearance of phloem fibres, nor the overall strength of the flax stems. As we also could not detect any increase in BGAL activity, we are unable to conclude that increases in LuBGAL1 activity are meaningless. However, assuming that LuBGAL1 protein expression had increased, and was localized correctly, then it seems likely that the lack of observable phenotype is a consequence of its working in coordination with other enzymes, which were not themselves over-expressed in this study. While LuBGAL1 has shown itself necessary for the degradation of galactan pectic polysaccharides in the cell walls of developing phloem fibres (Roach et al., 2011), the presence of arabinose caps on the ends of the pectic galactan side-chains (Mikshina et al., 2012) would impede LuBGAL1 activity if it acted solely as an exo-galactosidase, rendering it insufficient for pectin degradation in flax phloem fibres.

Transgenic flax bearing a *LuFLA35-RNAi* construct were previously characterized, where the construct was not observed to have any effect on flax development (Roach, 2009). With the availability of new analytical tools, we continued the examination of these lines and observed that the construct decreased the transcript abundance of multiple FLA genes (*LuFLA 1*, *LuFLA35*, and perhaps *LuFLA10*), which in turn was weakening the stems of the transgenic flax, similar to observations in *AtFLA11/AtFLA12* T-DNA insertion lines (MacMillan et al., 2012). Further analyses indicated slight changes in the monosaccharide content of the cell walls, perhaps indicative of a decrease in the polysaccharide side chains of the FLA proteins, and by extension the FLA proteins, although other changes to cell wall composition may be occurring as well. While these changes did not have any noticeable effect on plant morphology, unlike the RNAi analyses of *AtFLA3* in Arabidopsis which reduced fertility and produced shorter siliques (Li et al., 2010a), or the RNAi of *GhAGP4* which suppressed the initiation and elongation of seed trichomes in cotton (Li et al., 2010b), there were sporadic instances of multiple phloem fibres nested within larger phloem fibres, which, to our knowledge, has not been observed in other transgenic studies of FLA function.

Lastly, it will be important to determine how flax BGALs and FLAs interact in the development of the cell wall. BGALs have previously been found to degrade the galactan side-chains of arabinogalactan proteins (AGPs; Kotake et al., 2005). Our analyses of whole gene

family expression patterns, and more specifically the characterization of the promoter elements of *LuBGAL1* and *LuFLA1* (Hobson and Deyholos, 2013), have indicated a large degree of co-expression between the two protein families; however, we currently lack evidence for direct interaction between the two. As future analyses of BGALs progress, especially with regards to the expression and purification of different BGAL isoforms, it will be important to remember that FLAs, and other AGPs, are potential substrates that will need to be tested.

Future Perspectives

GH35 β-galactosidases have been described as important facilitators of cell wall metabolism in plants (Smith et al., 2002; Lazan et al., 2004; Dean et al., 2007; Roach et al., 2011; Sampedro et al., 2012). To better understand the myriad roles they play in cell wall development, especially in the development of G-type cell walls, it will be important to begin characterizing the biochemical function of the BGALs. To do this, we should begin by focussing on the expression and purification of BGALs from the many sub-families, examining the range of glycosidic linkages targeted by the enzymes of the different sub-families. In determining the range of substrates targeted by BGALs, whose expression patterns are becoming better understood, we should gain valuable insight into the chemical composition of the cell walls throughout the range of flax tissues.

A definitive function for FLAs has yet to be proposed; however, evidence suggests that the loss of FLAs results in decreased cellulose deposition in cell walls (Persson et al., 2005; MacMillan et al., 2010). AGP polysaccharide side-chains have been observed to cross-link with hemicelluloses of the cell wall (Tan et al., 2013), and as the fasciclin domain is known to adhere to cell membranes (Zinn et al., 1988; Huber and Sumper, 1994), it seems likely that FLAs act as anchors maintaining a connection between the cell wall and cell membrane, as has been proposed elsewhere (Johnson et al., 2003). The anchor itself may in turn be bound to other cell membrane bound complexes, such as cellulose synthases, where the lack of anchor may prevent the ordered deposition of new cell wall components. An important next step would be to better understand the

localization of FLAs in cell walls, perhaps by developing antibodies specific to FLA protein backbones. The development of such antibodies may also aid in determining if FLAs are part of protein complexes involved in cell wall deposition, by aiding in the extraction of FLA and FLA conjugates from flax tissues.

Conclusion

During the course of this thesis, we have improved our understanding of the size and expression patterns of the BGAL and FLA families of flax, gaining new insight into the evolution of the families and their roles in flax development. Genetic tools were developed that directly aided this research, which will continue to be of use in future studies of flax, including biotechnological modifications of flax phloem fibres. Further work will require a better understanding of the biochemistry of the phloem fibre cell walls, and the development thereof, however our contributions have successfully narrowed the scope of future studies.

References

Cloutier S, Ragupathy R, Miranda E, Radovanovic N, Reimer E, Walichnowski A, Ward K, Rowland G, Duguid S, and Banik M (2012) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theoretical and Applied Genetics **125**: 1783-1795. doi: 10.1007/s00122-012-1953-0

Dean GH, Zheng H, Tewari J, Huang J, Young DS, Hwang YT, Western TL, Carpita NC, McCann MC, Mansfield SD, and Haughn GW (2007) The *Arabidopsis MUM2* gene encodes a β-galactosidase required for the production of seed coat mucilage with correct hydration properties. The Plant Cell **19:** 4007–4021. doi:10.1105/tpc.107.050609

Kumar S, You FM, and Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC Genomics **13**: 684.

Fénart S, Ndong Y-PA, Duarte J, Rivière N, Wilmer J, van Wuytswinkel O, Lucau A, Cariou E, Neutelings G, Gutierrez L, Chabbert B, Guillot X, Tavernier R, Hawkins S, and Thomasset B (2010) Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray. BMC Genomics **11**: 592. doi: 10.1186/1471-2164-11-592

Hobson N and Deyholos MK (2013) $LuFLA1_{PRO}$ and $LuBGAL1_{PRO}$ promote gene expression in the phloem fibres of flax (*Linum usitatissimum*). Plant Cell Reports **32**: 517-528

Huber O, and Sumper M (1994) Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga *Volvox* with homology to *Drosophila* fasciclin I. The EMBO Journal **13**: 4212-4222.

Johnson KL, Jones BJ, Bacic A, and Schultz CJ (2003) The Fasciclin-Like Arabinogalactan Proteins of *Arabidopsis*. A Multigene Family of Putative Cell Adhesion Molecules. Plant Physiology **133**: 1911-1925.

Lafarguette F, Leplé, Déjardin A, Laurans F, Costa G, Lesage-Descauses MC, and Pilate G (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. New Phytologist **164**: 107-121. doi: 10.1111/j.1469-8137.2004.01175.x

Lazan H, Ng S-Y, Goh L-Y, and Ali Z M (2004) Papaya β-galactosidase/galactanase isoforms in differential cell wall hydrolysis and fruit softening during ripening. Plant Physiology and Biochemistry **42**: 847–853. doi:10.1016/j.plaphy.2004.10.007

Li J, Yu M, Geng LL, and Zhao J (2010a) The fasciclin-like arabinogalactan protein gene, *FLA3*, is involved in microspore development of Arabidopsis. The Plant Journal **64**: 482-497. doi: 10.1111/j.1365-313X.2010.04344.x

Li Y, Liu D, Tu L, Zhang X, Wang L Zhu L, Tan J, and Deng F (2010b) Suppression of *GhAGP4* gene expression repressed the initiation and elongation of cotton fiber. Plant Cell Reports **29**: 193-202. doi: 10.1007/s00299-009-0812-1

MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, and Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and *Eucalyptus*. The Plant Journal **62**: 689-703. doi: 10.1111/j.1365.313X.2010.04181.x

Persson S, Wei H, Milne J, Page GP, and Somerville CR (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. PNAS **102**: 8633-8638. doi: 10.1073/pnas.0503392102

Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, and Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β-galactosidase. Plant Physiology **156**: 1351–1363. doi:10.1104/pp.111.172676

Roach M (2009) Examining the Molecular Basis of Bast Fibre Development in Flax (*Linum usitatissimum* L.). Ph.D. Thesis. University of Alberta: Canada.

Sampedro J, Gianzo C, Iglesias N, Guitián E, Revilla G, and Zarra I (2012) AtBGAL10 is the main xyloglucan β-galactosidase in Arabidopsis, and its absence results in unusual xyloglucan subunits and growth defects. Plant Physiology **158**: 1146–57. doi:10.1104/pp.111.192195

Smith DL, Abbott JA, and Gross KC (2002) Down-Regulation of Tomato β-Galactosidase 4 Results in Decreased Fruit Softening. Plant Physiology **129**: 1755–1762. doi:10.1104/pp.011025

Snegireva AV, Ageeva MV, Amenitskii SI, Chernova TE, Ebskamp M, and Gorshkova TA (2010) Intrusive Growth of Sclerenchyma Fibers. Russian Journal of Plant Physiology **57**: 342-355. doi: 10.1134/S1021443710030052

Tan L, Eberhard S, Pattathil S, Warder C, Glushka J, Yuan C, Hao Z, Zhu X, Avci U, Miller JS, Baldwin D, Pham C, Orlando R, Darvill A, Hahn MG, Kieliszewski MJ, and Mohnen D (2013) An *Arabidopsis* Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein. The Plant Cell **25**: 270-287. doi: 10.1105/tpc.112.107334

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, and Datla R (2011) Gene expression analysis of flax seed development. BMC Plant Biology **11**:74. doi: 10.1186/1471-2229-11-74

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe A, Wong GK.-S, Wang J, and Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. The Plant Journal **72**: 461-473. doi:10.1111/j.1365-313X.2012.05093.x

Zinn K, McAllister L, and Goodman CS (1988) Sequence Analysis and Neuronal Expression of Fasciclin I in Grasshopper and Drosophila. Cell **53**: 577-587

Appendix

Table S1 Cis-elements within DNA fragments

Cis-Element	Consensus Sequence	Description	LuFLA1$_{PRO}$	P-value	LuBGAL1$_{PRO}$	P-value	LuBGAL1$_{Intergenic}$	P-value
MRNA3ENDTAH3	AATGGAAATG	Cis-element of histone genes	0	0.5373	0	0.5378	1	9.774E-13
ABREATRD22	RYACGTGGYR	ABRE (ABA responsive element)	0	0.5466	1	0.0000	1	1.375E-08
MYB1AT	WAACCA	MYB-class TF binding site (Abscisic acid and drought regulated)	0	0.8984	5	0.004817	10	0.0003425
PEND	ANTTCTTATK	SOX TF binding site	0	0.5770	0	0.5781	1	0.000763 0
GT1CONSENSUS	GRWAAW	GT-1 binding site (Possibly light regulated)	6	0.5651	6	0.5922	25	0.002007
PROLAMINBOXOSGLUB1	TGCAAAG	Prolamine box (Quantitative regulation)	0	0.6532	1	0.01776	2	0.002365
CATATGGMSAUR	CATATG	GmSAUR15a cis element (Auxin response)	0	0.7720	2	0.02936	4	0.006233
SEF4MOTIFGM7S	RTTTTTR	GmSEF4 binding site	0	0.8772	5	0.001058	7	0.009927
2SSEEDPROTBANAPA	CAAACAC	Cis-element of storage-protein genes	0	0.6764	1	0.04546	2	0.01201

CRTDREHVCBF2	GTCGAC	HvCBF2 binding site (Temperature regulated)	0	0.6826	2	0.0001219	2	0.01690
P	ACCWACCNN	MYB-class TF binding site (Flavonoid biosynthesis regulator)	0	0.6829	1	0.05602	2	0.01715
GT1GMSCAM4	GAAAAA	GT-1-like TF binding site (Pathogen and salt regulation)	1	0.6523	2	0.3475	7	0.01931
MYBPZM	CCWACC	MYB-class TF binding site (Induces expression in floral organs)	1	0.3787	2	0.07631	4	0.03145
WBOXHVISO1	TGACT	W-box (Sucrose responsive element)	1	0.6946	2	0.4093	7	0.04036
L1BOXATPDF1	TAAATGYA	L1-box (SAM L1 tissue specific expression)	0	0.6214	0	0.6231	1	0.04340
WBOXNTCHN48	CTGACY	W-box (Elicitor response)	1	0.2655	0	0.7720	3	0.04934
REALPHALGLHCB21	AACCAA	LgLHCB21 cis-element (Light regulated)	1	0.5381	2	0.2071	5	0.05010
ACGTABOX	TACGTA	A-box (bZIP-class TF binding site)	0	0.7090	0	0.7117	2	0.05231
RUNX1	WWDTGHGGTWW	HsAML-1 binding site	0	0.6267	1	0.003241	1	0.05499
WBOXNTERF3	TGACY	W-box (Possibly wound response)	1	0.8674	3	0.4942	10	0.07722

MYB1	MTCCWACC	MYB-class TF binding site	0	0.6400	1	0.008506	1	0.08993
MYBPLANT	MACCWAMC	MYB-class TF binding site (Induces flower expression)	1	0.1501	1	0.1576	2	0.09595
C1MOTIFZMBZ2	YNAACYA	MYB-class TF binding site	3	0.4183	4	0.2229	9	0.1003
WBOXATNPR1	TTGAC	W-box (Salicylic acid regulated)	0	0.9239	3	0.2682	7	0.1225
P1BS	GNATATNC	AtPHR1 binding site (Phosphate regulated)	0	0.7385	2	0.007339	2	0.1241
CAATBOX1	CAAT	CAAT-box	6	0.9221	17	0.03144	29	0.1244
MYB2CONSENSUSAT	YAACKG	MYB-class TF binding site (Abscisic acid and drought regulated)	2	0.3021	3	0.09429	5	0.1361
id1	TTKYYYYTWBYG	ZmID1 binding site (Flowering time regulation)	0	0.6570	1	0.02127	1	0.1444
ELRECOREPCRP1	TTGACC	W-box (Elicitor response)	0	0.7457	1	0.2057	2	0.1467
TEIL	ATGWAYCT	EIN3-like binding site (Ethylene response)	0	0.6582	0	0.6603	1	0.1486
BOXLCOREDCPAL	ACCWWCC	MYB-class TF binding site	0	0.7545	1	0.2316	2	0.1767
PALBOXAPC	CCGTCC	Box A (Necessary for elicitor/light mediated gene activity)	0	0.6680	0	0.6702	1	0.1837

ACGTATERD1	ACGT	Etiolation induced expression	2	0.7787	6	0.09208	10	0.1874
CANBNNAPA	CNAACAC	BnNAPa cis-element (Seed specificity)	0	0.7576	1	0.2409	2	0.1879
CIACADIANLELHC	CAANNNNATC	LeLHC cis-element (Necessary for circadian rhythm)	1	0.4427	1	0.4539	3	0.2169
HMG-1	SKYBTVHTY	HMG binding site	2	0.8544	6	0.1994	11	0.2727
MYBCOREATCYCB1	AACGG	MYB-class TF binding site	1	0.4911	2	0.1606	3	0.2850
WRKY71OS	TGAC	W-box core motif	6	0.4840	5	0.6687	15	0.2958
DPBFCOREDCDC3	ACACNNG	bZIP-class TF binding site (Abscisic acid response)	2	0.1586	1	0.5114	3	0.2987
SV40COREENHAN	GTGGWWHG	Simian Virus 40 core enhancer	0	0.7061	0	0.7088	1	0.3323
BOXIINTPATPB	ATAGAA	Plastid gene cis-element	1	0.3648	1	0.3756	2	0.3656
QELEMENTZMZM13	AGGTCA	Q-element (Quantitative expression enhancer)	0	0.7150	0	0.7177	1	0.3676
POLLEN1LELAT52	AGAAA	Pollen specific expression	2	0.8823	5	0.4343	11	0.3749
TATABOX2	TATAAAT	TATA-box PsLEGa	0	0.7191	0	0.7219	1	0.3840
BIHD1OS	TGTCA	OsBIHD1 binding site (disease resistance)	4	0.07640	2	0.5122	5	0.3869
ERELEE4	AWTTCAAA	ERE (Ethylene responsive element)	0	0.7251	0	0.7279	1	0.4075

RHERPATEXPA7	KCACGW	RHE motif (Root-Hair-specific cis-Element)	0	0.8119	1	0.4139	2	0.4193
EBOXBNNAPA	CANNTG	BnNAPa E-box (Induces strong seed expression)	10	0.2153	10	0.2427	18	0.4313
MYCCONSENSUSAT	CANNTG	MYC-class TF binding site (Abscisic acid and drought regulated)	10	0.2153	10	0.2427	18	0.4313
LTRE1HVBLT49	CCGAAA	LTRE (Low Temperature Responsive Element)	0	0.7364	0	0.7393	1	0.4512
AtMYC2	CACATG	MYC-class TF binding site (Abscisic acid and drought regulated)	0	0.7409	1	0.1918	1	0.4684
PYRIMIDINEBOXOSRAMY1A	CCTTTT	HvBPBF binding site (Gibberellin response)	1	0.4394	2	0.1172	2	0.4713
INTRONLOWER	TGCAGG	Intron-exon slice site	1	0.1903	0	0.7463	1	0.4777
NODCON2GM	CTCTT	NOD-gene cis-element	1	0.8454	3	0.4351	6	0.4796
OSE2ROOTNODULE	CTCTT	Motif in promoter inducing infected root expression	1	0.8454	3	0.4351	6	0.4796
MYB.ph3	WAACNRWYW	MYB-class TF binding site	0	0.9092	2	0.4510	4	0.4891
ARR10	AGATHHKN	GARP-family TF binding site	4	0.2138	2	0.6795	6	0.4957

DOFCOREZM	AAAG	DOF-class TF binding site	8	0.8399	11	0.5758	25	0.5054
CBFHV	RYCGAC	HvCBF1 &2 binding site (Dehydration response)	0	0.8380	2	0.1583	2	0.5410
S1FBOXSORPS1L21	ATGGTA	SoRPS1 and SoRPL2 cis-element (Negative regulator)	1	0.2447	0	0.7651	1	0.5464
POLASIG1	AATAAA	PolyA signal	0	0.8861	1	0.6572	3	0.5480
MYBCORE	CNGTTR	MYB-class TF binding site	2	0.6954	3	0.4873	6	0.5537
POLASIG2	AATTAAA	PolyA signal	0	0.7640	0	0.7672	1	0.5539
Bellringer	AAATTARW	BELLRINGER binding site	0	0.7702	0	0.7733	1	0.5757
GBF5	ACTCAT	Hypoosmolarity response	0	0.7702	0	0.7734	1	0.5758
PREATPRODH	ACTCAT	PRE (Proline/hypoosmolarity Responsive Element)	0	0.7702	0	0.7734	1	0.5758
TCA1MOTIF	TCATCTTCTT	NtTCA-1 binding site (Associated with stress response)	1	4.328E-12	0	0.5578	0	0.5845
ANAERO2CONSENSUS	AGCAGC	Cis-element of anaerobic genes	0	0.7729	0	0.7761	1	0.5852
GATABOX	GATA	GATA-family TF binding site	8	0.3769	4	0.8936	15	0.5856
RYREPEATBNNAPA	CATGCA	RY repeat (Required for seed-specific expression)	0	0.7733	0	0.7765	1	0.5866

SORLIP1AT	GCCAC	SORLIP motif (Sequence Overly Represented in Light Induced Promoters)	1	0.5330	1	0.5442	2	0.6029
TAAAGSTKST1	TAAAG	DOF-class TF binding site (Guard cell specific expression)	2	0.5177	1	0.7798	4	0.6050
AP1	TTTTTRG	CArG-box (MADS-box TF binding site)	0	0.7788	0	0.7820	1	0.6055
RAV1AAT	CAACA	AtRAV1 binding site	1	0.8791	4	0.3096	6	0.6098
SURE1STPAT21	AATAGAAAA	SURE (Sucrose Response Element)	1	4.796E-07	0	0.5788	0	0.6148
TATABOX5	TTATTT	TATA-box PsGS2	0	0.8976	0	0.9007	3	0.6155
CARGCW8GAT	CWWWWWWWWG	CArG-box (MADS-box AGL15 binding site)	0	0.8569	0	0.8603	2	0.6275
CEREGLUBOX2PSLEGA	TGAAAACT	PsLEGa cis-element	1	0.0002120	0	0.6055	0	0.6527
TBOXATGAPB	ACTTTG	T-Box (Light regulated)	0	0.7954	0	0.7987	1	0.6602
ANAERO1CONSENSUS	AAACAAA	Cis-element of anaerobic genes	0	0.7972	1	0.3658	1	0.6659
HMG-IY	NVWNRRRNRNMRWMRH	HMG binding sites	1	0.7988	3	0.3277	4	0.6721
BOXCPSAS1	CTCCCAC	PsAS1 cis-element (light regulated)	1	0.002399	0	0.6267	0	0.6822
SORLIP5AT	GAGTGAG	SORLIP motif (Sequence Overly Represented in	1	0.002422	0	0.6268	0	0.6824

		Light Induced Promoters)						
POLASIG3	AATAAT	PolyA signal	0	0.8754	1	0.6226	2	0.7078
MYB4	AMCWAMC	MYB-class TF binding site	1	0.6212	1	0.6321	2	0.7199
ARR1AT	NGATT	ARR1 binding site	8	0.7691	9	0.6946	20	0.7303
CGACGOSAMY3	CGACG	Cis-element of OsAMY3 genes	0	0.8223	0	0.8257	1	0.7414
ANAERO3CONSENSUS	TCATCAC	Cis-element of anaerobic genes	1	0.03429	0	0.6730	0	0.7445
HEXMOTIFTAH3H4	ACGTCA	TaHBP-1 binding site	1	0.05842	0	0.6892	0	0.7655
SQUA	MYDWWWWWRGWMAN	MADS-box family TF binding site	1	0.07119	0	0.6963	0	0.7745
RAV1BAT	CACCTG	AtRAV1 binding site	1	0.09189	0	0.7067	0	0.7875
TATABOX4	TATATAA	TATA-box PvPHAS	1	0.1545	0	0.7331	0	0.8193
ARFAT	TGTCTC	ARF binding site	1	0.1760	0	0.7411	0	0.8286
SEBFCONSSTPR10A	YTGTCWC	SEBF motif (Silencing Element Binding Factor)	2	0.007986	0	0.7462	0	0.8344
ABI4	SYGCYYYY	ABI4 binding site (Abscisic acid and sugar response)	0	0.8602	0	0.8636	1	0.8381
IBOX	GATAAG	Cis-element of light responsive genes	1	0.2038	0	0.7511	0	0.8400
E2FCONSENSUS	WTTSSCSS	E2F binding site (Cell cycle regulation)	1	0.2168	0	0.7556	0	0.8451
SORLIP2AT	GGGCC	SORLIP motif (Sequence Overly Represented in	1	0.2593	0	0.7699	0	0.8607

		Light Induced Promoters)						
CPBCSPOR	TATTAG	CsPOR cis-element (Cytokinin regulated)	2	0.02208	0	0.7703	0	0.8612
-10PEHVPSBD	TATTCT	Chloroplast psbD cis-element (Light regulated)	1	0.3416	0	0.7963	0	0.8879
Agamous	CCWWWWNNRGH	AGAMOUS binding site	2	0.05985	0	0.8019	0	0.8933
MYBST1	GGATA	MYB-class TF binding site	0	0.8885	0	0.8917	1	0.8963
NODCON1GM	AAAGAT	NOD-gene cis-element	1	0.4771	0	0.8379	0	0.9259
OSE1ROOTNODULE	AAAGAT	Motif in promoter inducing infected root expression	1	0.4771	0	0.8379	0	0.9259
ASF1MOTIFCAMV	TGACG	Auxin and/or salicylic acid regulated	1	0.4999	0	0.8448	0	0.9316
SURECOREATSULTR11	GAGAC	SURE motif (Sulphur Responsive Element)	1	0.5843	0	0.8704	0	0.9511
Core	ATTA	ZFHD-class TF binding site	12	0.4618	5	0.9791	17	0.9600
INRNTPSADB	YTCANTYY	Chloroplast psdB cis-element (Light regulated)	2	0.4839	0	0.9215	0	0.9810

Cis-elements were identified using PlantPan (Chang *et al.* 2008). Z-Tests were used to determine the probability that the frequency within the DNA fragment was not greater than the frequency within the repeat-masked whole genome shotgun assembly (Wang *et al.*, 2012). IUPAC ambiguity codes are used to describe consensus DNA sequences. Y = C/T, R = A/G, W = A/T, S = G/C, K = T/G, M = C/A, D = A/G/T, V = A/C/G, H = A/C/T, B = C/G/T, N = any base.

Table S2 Cis-elements within *LuBGAL1* upstream intergenic sequence (P-Value ≤ 0.05)

Cis-Element	Consensus Sequence	Description	Occurrence	P-value
MRNA3ENDTAH3	AATGGAAATG	Cis-element of histone genes	1	9.7744E-13
ABREATRD22	RYACGTGGYR	ABRE (ABA responsive element)	1	1.37583E-08
MYB1AT	WAACCA	MYB-class TF binding site (Abscisic acid and drought regulated)	10	0.00034257
PEND	ANTTCTTATK	SOX TF binding site	1	0.000763049
GT1CONSENSUS	GRWAAW	GT-1 binding site (Possibly light regulated)	25	0.002007266
PROLAMINBOXOSGLUB1	TGCAAAG	Prolamine box (Quantitative regulation)	2	0.002365996
CATATGGMSAUR	CATATG	GmSAUR15 a cis element (Auxin response)	4	0.006233978
SEF4MOTIFGM7S	RTTTTTR	GmSEF4 binding site	7	0.009927088
2SSEEDPROTBANAPA	CAAACAC	Cis-element of storage-protein genes	2	0.012017447
CRTDREHVCBF2	GTCGAC	HvCBF2 binding site (Temperature regulated)	2	0.016909906
P	ACCWACCNN	MYB-class TF binding site (Flavonoid biosynthesis regulator)	2	0.01715603
GT1GMSCAM4	GAAAAA	GT-1-like TF binding site (Pathogen and salt regulation)	7	0.019315642
MYBPZM	CCWACC	MYB-class TF binding site (Induces expression in floral organs)	4	0.031456149

WBOXHVISO1	TGACT	W-box (Sucrose responsive element)	7	0.040369938
L1BOXATPDF1	TAAATGYA	L1-box (SAM L1 tissue specific expression)	1	0.043407263
WBOXNTCHN48	CTGACY	W-box (Elicitor response)	3	0.049342073

Cis-elements were identified using PlantPan (Chang *et al.* 2008). Z-Tests were used to determine the probability that the frequency within the DNA fragment was not greater than the frequency within the repeat-masked whole genome shotgun assembly (Wang *et al.*, 2012). IUPAC ambiguity codes are used to describe consensus DNA sequences. Y = C/T, R = A/G, W = A/T, S = G/C, K = T/G, M = C/A, D = A/G/T, V = A/C/G, H = A/C/T, B = C/G/T, N = any base.

Table S3 Genomic loci and accessions of analysed BGALs

Species	Gene	Locus	GenbankAccession
Homo sapiens	*GLB1*		NP_000395
Arabidopsis thaliana	*AtBGAL01*	At3G13750	
	AtBGAL02	At3G52840	
	AtBGAL03	At4G36360	
	AtBGAL04	At5G56870	
	AtBGAL05	At1G45130	
	AtBGAL06	At5G63800	
	AtBGAL07	At5G20710	
	AtBGAL08	At2G28470	
	AtBGAL09	At2G32810	
	AtBGAL10	At5G63810	
	AtBGAL11	At4G35010	
	AtBGAL12	At4G26140	
	AtBGAL13	At2G16730	
	AtBGAL14	At4G38590	
	AtBGAL15	At1G31740	
	AtBGAL16	At1G77410	
	AtBGAL17	At1G72990	
	AtBGAL18	At2G04060	
Physcomitrella patens	*PpBGAL01*	Pp1s24_113V6.1	
	PpBGAL02	Pp1s189_71V6.1	
	PpBGAL03	Pp1s10_144V6.1	
	PpBGAL04	Pp1s49_266V6.1	
	PpBGAL05	Pp1s52_244V6.1	
	PpBGAL06	Pp1s189_65V6.1	
Populus trichocarpa	*PtBGAL01*	Potri.017G057900	
	PtBGAL02	Potri.009G012400	
	PtBGAL03	Potri.011G044300	
	PtBGAL04	Potri.009G134400	
	PtBGAL05	Potri.004G174800	
	PtBGAL06	Potri.013G105100	
	PtBGAL07	Potri.002G080700	
	PtBGAL08	Potri.005G180600	
	PtBGAL09	Potri.005G069200	
	PtBGAL10	Potri.003G040000	
	PtBGAL11	Potri.003G037000	
	PtBGAL12	Potri.T154000	
	PtBGAL13	Potri.001G025700	
	PtBGAL14	Potri.001G025800	

	PtBGAL15	Potri.006G139100	
	PtBGAL16	Potri.018G062800	
	PtBGAL17	Potri.004G159800	
	PtBGAL18	Potri.007G018100	
	PtBGAL19	Potri.005G232600	
	PtBGAL20	Potri.003G038500	
	PtBGAL21	Potri.001G200400	
	PtBGAL22	Potri.006G144500	
	PtBGAL23	Potri.007G099800	
Oryza sativa	*OsBGAL01*	LOC_Os03g06940	
	OsBGAL02	LOC_Os06g37560	
	OsBGAL03	LOC_Os01g39830	
	OsBGAL04	LOC_Os01g65460	
	OsBGAL05	LOC_Os01g34920	
	OsBGAL06	LOC_Os05g35360	
	OsBGAL07	LOC_Os02g12730	
	OsBGAL08	LOC_Os03g15020	
	OsBGAL09	LOC_Os05g46200	
	OsBGAL10	LOC_Os08g43570	
	OsBGAL11	LOC_Os09g36810	
	OsBGAL12	LOC_Os10g18400	
	OsBGAL13	LOC_Os12g24170	
	OsBGAL14	LOC_Os10g19960	
	OsBGAL15	LOC_Os06g42310	
	OsBGAL16	LOC_Os05g35370	
Ricinus communis	*RcBGAL01*	28076.m000428	
	RcBGAL02	29917.m001961	
	RcBGAL03	28694.m000663	
	RcBGAL04	50666.m000014	
	RcBGAL05	29739.m003718	
	RcBGAL06	29739.m003719	
	RcBGAL07	29739.m003720	
	RcBGAL08	29739.m003740	
	RcBGAL09	29739.m003723	
	RcBGAL10	29815.m000493	
	RcBGAL11	29904.m002899	
	RcBGAL12	59437.m00005	
	RcBGAL13	28694.m000675	
	RcBGAL14	29912.m005323	
	RcBGAL15	29912.m005324	
	RcBGAL16	30131.m007094	

	RcBGAL17	30170.m014108	
	RcBGAL18	29648.m002008	
	RcBGAL19	30193.m000721	
	RcBGAL20	30074.m001370	
	RcBGAL21	30193.m000718	
Zea mays	*ZmBGAL01*	AC199908.4	
	ZmBGAL02	AC234152.1	
	ZmBGAL03	GRMZM2G027385	
	ZmBGAL04	GRMZM2G038281	
	ZmBGAL05	GRMZM2G071883	
	ZmBGAL06	GRMZM2G073584	
	ZmBGAL07	GRMZM2G081583	
	ZmBGAL08	GRMZM2G121495	
	ZmBGAL09	GRMZM2G127123	
	ZmBGAL10	GRMZM2G130375	
	ZmBGAL11	GRMZM2G151122	
	ZmBGAL12	GRMZM2G153200	
	ZmBGAL13	GRMZM2G162238	
	ZmBGAL14	GRMZM2G164676	
	ZmBGAL15	GRMZM2G175779	
	ZmBGAL16	GRMZM2G178106	
	ZmBGAL17	GRMZM2G386824	
	ZmBGAL18	GRMZM2G417455	
	ZmBGAL19	GRMZM2G465617	
	ZmBGAL20	GRMZM5G828603	

Genome assemblies for plant species can be obtained from Phytozome (version 8.0; www.phytozome.net).

Table S4 Primers and hydrolysis probes used in qRT-PCR analysis

Gene	Forward Primer	Reverse Primer	Hydrolysis Probe
LuBGAL01	agacccatttctcaatgtcca	cgatgagttttccattgacg	18
LuBGAL02	ctttggaaatggccagtatga	cgttttgcaatcaggcagta	33
LuBGAL03	cgtaaagccataatcgtcaatg	ggagtgcttctggggtagtg	2
LuBGAL04	ccacagtcatgggaatagcc	cttcatgtacaatacccattccac	36
LuBGAL05	atgatggcaggtgggactt	caccettgagaccaatcttgt	29
LuBGAL06	catcccgaagcttcacaataa	tggaggaaatcagaatggaga	5
LuBGAL07	taccacgtccctcgttcg	cggattcctcaaaaacgact	139
LuBGAL08	gaggagaggcgttgaatttg	tcttcagcccactcaacaga	19
LuBGAL09	gctgtggtctcctctgatcc	tgagaacacatgagcctgttg	48
LuBGAL10	cgttgtccagaatgcttgc	gcagccatcgaaagagga	142
LuBGAL11	ttggcctgcctatctctcc	cttcctctgtagtcacagttctcg	46
LuBGAL12	ggtaaatggtccctgcaagtt	aggcgcctttcacgtactc	110
LuBGAL13	taagtggttccccgcaaat	aggcgcctttcaagtactca	110
LuBGAL14	gcctaccaacaggattgtgc	tttcgtagcagattttgtgtcg	137
LuBGAL15	gcaagctgtttggtgtacctg	tccagttttgacagccatgt	26
LuBGAL16	tgaactttacaagtagggccagt	atgaccagtgcccacttacc	29
LuBGAL17	actgaaaattggagcggatg	atgcaaggtcttccacaggt	119
LuBGAL18	ttatatgtaccatggtgggactaact	gtggcgataaagggtccac	12
LuBGAL19	ttgcaagtagacgaccacca	gccaagtaaagcaatgaggtg	153
LuBGAL20	gatggccgatctctcatcat	agtgaatggaagcggaaatg	6
LuBGAL21	cgctagactgggatgcttctt	tcccaccctctccatagaact	39
LuBGAL22	cgatcattgttcaagccatc	gtcgccgcctatctcctc	88
LuBGAL23	accttcagaggtgcccaaa	ggctcatccacaatcaccttt	110
LuBGAL24	gaagtcgtggacggcagtat	gtccgcacgttccacatt	21
LuBGAL25	tcggtggaaaatgaataccaat	catattagcagcccagtaagca	60
LuBGAL26	ctcttcatggcttgctcgat	gaattgagccggagaacaga	34
LuBGAL27	gcttggtacacaaccaagatacaa	tgccctaaacttgccacttc	76
LuBGAL28	aactgttgagttcaggggaaag	tggcagaatgctaacagagg	8
LuBGAL29	ccagccggaatacagtgg	cggcgacaatacgaagaag	153
LuBGAL30	cgttgccgtcttcctcatac	ttattgtcatcgtgagcattgac	131
LuBGAL31	gctaacttcaagggccagaa	aatcaggaaggacgctgatg	69
LuBGAL32	cccctgtctccaagcaagt	atgggacatcgcacgaat	150
LuBGAL33	tttgttgagcgtctctgtagga	taaagccgctcgctcttg	50
LuBGAL34	cggcaaatatatagggaactcg	tgacacagtgagattggagctt	6
LuBGAL35	gaagggaaactgtgcagcat	gcatttcgaaacacaactgc	129
LuBGAL36	ttttcttgctttgctgttgc	cttcatcttcgccttcttgtc	143
LuBGAL37	aacgtcgaggcagcattc	acataacccacgggacttca	44
LuBGAL38	agacgtttaacggagccaac	cttgtcccgttttctttcca	137
LuBGAL39	agctggagttcggagttcac	aggaatggaaaggaaatgagg	143
LuBGAL40	ccctttcattgaagctgagtg	atgtctgggacatcatgcaa	157
LuBGAL41	aacatcccccttgtaactaaatagaat	gcttcaggacacacacctatga	153
LuBGAL42	tggagttgaatccgaaaacc	gtcacgtatagcatgaaaccaaa	123
LuBGAL43	cggtcgagcaatcaccat	gcttcgtgggtaatggactg	101

Oligonucleotide primer sequences and probes for flax BGAL genes were obtained from the Universal Probe Library Assay Design Center (http://tinyurl.com/7u6s5bh).

Table S5 Genomic loci and accessions of additional FLAs

Species	Gene	Locus	Genbank
Arabidopsis thaliana	*AtFLA01*	At5g55730	
	AtFLA02	At4g12730	
	AtFLA03	At2g24450	
	AtFLA04	At3g46550	
	AtFLA05	AT4g31370	
	AtFLA06	At2g20520	
	AtFLA07	At2g04780	
	AtFLA08	At2g45470	
	AtFLA09	At1g03870	
	AtFLA10	At3g60900	
	AtFLA11	At5g03170	
	AtFLA12	At5g60490	
	AtFLA13	At5g44130	
	AtFLA14	At3g12660	
	AtFLA15	At3g52370	
	AtFLA16	At2g35860	
	AtFLA17	At5g06390	
	AtFLA18	At3g11700	
	AtFLA19	At1g15190	
	AtFLA20	At5g40940	
	AtFLA21	At5g06920	
Brassica rapa	*BrFLA01*	Bra000566	
	BrFLA02	Bra001464	
	BrFLA03	Bra002878	
	BrFLA04	Bra003432	
	BrFLA05	Bra004915	
	BrFLA06	Bra005740	
	BrFLA07	Bra005894	
	BrFLA08	Bra005920	
	BrFLA09	Bra006656	
	BrFLA10	Bra007569	
	BrFLA11	Bra007849	
	BrFLA12	Bra009190	
	BrFLA13	Bra010241	
	BrFLA14	Bra013197	
	BrFLA15	Bra015274	
	BrFLA16	Bra017285	
	BrFLA17	Bra019572	
	BrFLA18	Bra023035	
	BrFLA19	Bra025535	
	BrFLA20	Bra026159	
	BrFLA21	Bra027531	
	BrFLA22	Bra028848	
	BrFLA23	Bra028962	
	BrFLA24	Bra029925	
	BrFLA25	Bra032093	
	BrFLA26	Bra033439	
	BrFLA27	Bra033722	
	BrFLA28	Bra034746	
	BrFLA29	Bra034817	
	BrFLA30	Bra035589	
	BrFLA31	Bra038674	

	BrFLA32	Bra038741	
	BrFLA33	Bra039335	
Citrus sinensis	*CsFLA01*	orange1.1g013434m	
	CsFLA02	orange1.1g013841m	
	CsFLA03	orange1.1g015356m	
	CsFLA04	orange1.1g019153m	
	CsFLA05	orange1.1g019462m	
	CsFLA06	orange1.1g024965m	
	CsFLA07	orange1.1g025798m	
	CsFLA08	orange1.1g025959m	
	CsFLA09	orange1.1g026047m	
	CsFLA10	orange1.1g035767m	
	CsFLA11	orange1.1g037512m	
	CsFLA12	orange1.1g039722m	
	CsFLA13	orange1.1g042255m	
	CsFLA14	orange1.1g042775m	
	CsFLA15	orange1.1g043668m	
	CsFLA16	orange1.1g043847m	
	CsFLA17	orange1.1g044183m	
Eucalyptus grandis	*EgFLA01*	Eucgr.B02486	EF534216
	EgFLA02	Eucgr.J00938	EF534217
	EgFLA03	Eucgr.A01158	EF534218
	EgFLA04	Eucgr.A01074	
	EgFLA05	Eucgr.A01871	
	EgFLA06	Eucgr.A02551	
	EgFLA07	Eucgr.B02370	
	EgFLA08	Eucgr.B03801	
	EgFLA09	Eucgr.C02101	
	EgFLA10	Eucgr.F02049	
	EgFLA11	Eucgr.F03689	
	EgFLA12	Eucgr.G02206	
	EgFLA13	Eucgr.G02208	
	EgFLA14	Eucgr.G02210	
	EgFLA15	Eucgr.H00875	
	EgFLA16	Eucgr.K00711	
	EgFLA17	Eucgr.K02662	
Gossypium hirsutum	*GhFLA01*		ABV27472
	GhFLA02		ABV27473
	GhFLA03		ABV27474
	GhFLA04		ABV27475
	GhFLA05		ABV27476
	GhFLA06		ABV27477
	GhFLA07		ABV27478
	GhFLA08		ABV27479
	GhFLA09		ABV27480
	GhFLA10		ABV27481
	GhFLA11		ABV27482
	GhFLA12		ABV27483
	GhFLA13		ABV27484
	GhFLA14		ABV27485
	GhFLA15		ABV27486
	GhFLA16		ABV27487
	GhFLA17		ABV27488

	GhFLA18		ABV27489
	GhFLA19		ABV27490
Glycine max	*GmFLA01*	Glyma02g40950	
	GmFLA02	Glyma02g47790	
	GmFLA03	Glyma02g47880	
	GmFLA04	Glyma03g33720	
	GmFLA05	Glyma03g33730	
	GmFLA06	Glyma03g36260	
	GmFLA07	Glyma04g07890	
	GmFLA08	Glyma04g15200	
	GmFLA09	Glyma05g29430	
	GmFLA10	Glyma05g29440	
	GmFLA11	Glyma06g07940	
	GmFLA12	Glyma06g46530	
	GmFLA13	Glyma08g12580	
	GmFLA14	Glyma08g12590	
	GmFLA15	Glyma08g12600	
	GmFLA16	Glyma08g19605	
	GmFLA17	Glyma08g44210	
	GmFLA18	Glyma09g05311	
	GmFLA19	Glyma09g21240	
	GmFLA20	Glyma09g40420	
	GmFLA21	Glyma10g39110	
	GmFLA22	Glyma10g41200	
	GmFLA23	Glyma11g15960	
	GmFLA24	Glyma11g15990	
	GmFLA25	Glyma11g16000	
	GmFLA26	Glyma11g20720	
	GmFLA27	Glyma11g20760	
	GmFLA28	Glyma11g20770	
	GmFLA29	Glyma11g20780	
	GmFLA30	Glyma11g20791	
	GmFLA31	Glyma11g20800	
	GmFLA32	Glyma11g20810	
	GmFLA33	Glyma11g20820	
	GmFLA34	Glyma12g07370	
	GmFLA35	Glyma12g07400	
	GmFLA36	Glyma12g07411	
	GmFLA37	Glyma12g07420	
	GmFLA38	Glyma12g07430	
	GmFLA39	Glyma12g07440	
	GmFLA40	Glyma12g07450	
	GmFLA41	Glyma12g07460	
	GmFLA42	Glyma12g07490	
	GmFLA43	Glyma12g10240	
	GmFLA44	Glyma12g29670	
	GmFLA45	Glyma12g31691	
	GmFLA46	Glyma12g33530	
	GmFLA47	Glyma13g29790	
	GmFLA48	Glyma13g29800	
	GmFLA49	Glyma13g36930	
	GmFLA50	Glyma13g38730	
	GmFLA51	Glyma13g40210	

	GmFLA52	Glyma13g40220	
	GmFLA53	Glyma13g44921	
	GmFLA54	Glyma14g00720	
	GmFLA55	Glyma14g00830	
	GmFLA56	Glyma14g06700	
	GmFLA57	Glyma14g17200	
	GmFLA58	Glyma15g09240	
	GmFLA59	Glyma15g09250	
	GmFLA60	Glyma15g16650	
	GmFLA61	Glyma17g29710	
	GmFLA62	Glyma18g08530	
	GmFLA63	Glyma18g45420	
	GmFLA64	Glyma19g36470	
	GmFLA65	Glyma19g38910	
	GmFLA66	Glyma20g26070	
Gossypium raimondii	*GrFLA01*	Gorai.001G097900	
	GrFLA02	Gorai.001G110500	
	GrFLA03	Gorai.001G131200	
	GrFLA04	Gorai.001G194400	
	GrFLA05	Gorai.001G219000	
	GrFLA06	Gorai.001G222400	
	GrFLA07	Gorai.002G100100	
	GrFLA08	Gorai.003G132100	
	GrFLA09	Gorai.004G026800	
	GrFLA10	Gorai.004G061700	
	GrFLA11	Gorai.005G208100	
	GrFLA12	Gorai.006G119900	
	GrFLA13	Gorai.006G147500	
	GrFLA14	Gorai.007G092300	
	GrFLA15	Gorai.007G165800	
	GrFLA16	Gorai.007G374100	
	GrFLA17	Gorai.008G114900	
	GrFLA18	Gorai.008G155400	
	GrFLA19	Gorai.010G123500	
	GrFLA20	Gorai.010G169900	
	GrFLA21	Gorai.011G228300	
	GrFLA22	Gorai.013G003200	
	GrFLA23	Gorai.013G152900	
	GrFLA24	Gorai.013G256000	
Malus domestica	*MdFLA01*	MDP0000129025	
	MdFLA02	MDP0000130352	
	MdFLA03	MDP0000147719	
	MdFLA04	MDP0000148467	
	MdFLA05	MDP0000156830	
	MdFLA06	MDP0000171684	
	MdFLA07	MDP0000184487	
	MdFLA08	MDP0000194025	
	MdFLA09	MDP0000223366	
	MdFLA10	MDP0000227011	
	MdFLA11	MDP0000236723	
	MdFLA12	MDP0000257330	
	MdFLA13	MDP0000282305	
	MdFLA14	MDP0000303294	

	MdFLA15	MDP0000322290	
	MdFLA16	MDP0000389735	
	MdFLA17	MDP0000412035	
	MdFLA18	MDP0000437590	
	MdFLA19	MDP0000525641	
	MdFLA20	MDP0000530222	
	MdFLA21	MDP0000533067	
	MdFLA22	MDP0000658332	
	MdFLA23	MDP0000686466	
	MdFLA24	MDP0000803873	
	MdFLA25	MDP0000852699	
	MdFLA26	MDP0000904458	
Manihot esculenta	*MeFLA01*	cassava4.1_007195m	
	MeFLA02	cassava4.1_007317m	
	MeFLA03	cassava4.1_008878m	
	MeFLA04	cassava4.1_008882m	
	MeFLA05	cassava4.1_011291m	
	MeFLA06	cassava4.1_013900m	
	MeFLA07	cassava4.1_014009m	
	MeFLA08	cassava4.1_015156m	
	MeFLA09	cassava4.1_015233m	
	MeFLA10	cassava4.1_017586m	
	MeFLA11	cassava4.1_021867m	
	MeFLA12	cassava4.1_022072m	
	MeFLA13	cassava4.1_022699m	
	MeFLA14	cassava4.1_023463m	
	MeFLA15	cassava4.1_023801m	
	MeFLA16	cassava4.1_025091m	
	MeFLA17	cassava4.1_025211m	
	MeFLA18	cassava4.1_025550m	
	MeFLA19	cassava4.1_027600m	
	MeFLA20	cassava4.1_028966m	
	MeFLA21	cassava4.1_030392m	
	MeFLA22	cassava4.1_030506m	
	MeFLA23	cassava4.1_030640m	
	MeFLA24	cassava4.1_030729m	
	MeFLA25	cassava4.1_031001m	
	MeFLA26	cassava4.1_031631m	
	MeFLA27	cassava4.1_032208m	
	MeFLA28	cassava4.1_033072m	
Oryza sativa	*OsFLA01*	LOC_Os04g48490	
	OsFLA02	LOC_Os03g03600	
	OsFLA03	LOC_Os08g23180	
	OsFLA04	LOC_Os08g38270	
	OsFLA05	LOC_Os08g39270	
	OsFLA06	LOC_Os05g48900	
	OsFLA07	LOC_Os01g47780	
	OsFLA08	LOC_Os01g06580	
	OsFLA09	LOC_Os05g07060	
	OsFLA10	LOC_Os09g30010	
	OsFLA11	LOC_Os09g07350	
	OsFLA12	LOC_Os01g62380	
	OsFLA13	LOC_Os04g39600	

	OsFLA14	LOC_Os04g39590	
	OsFLA15	LOC_Os02g20560	
	OsFLA16	LOC_Os07g06680	
	OsFLA17	LOC_Os03g57490	
	OsFLA18	LOC_Os05g48890	
	OsFLA19	LOC_Os02g20540	
	OsFLA20	LOC_Os02g26320	
	OsFLA21	LOC_Os02g49420	
	OsFLA22	LOC_Os02g26290	
	OsFLA23	LOC_Os06g17460	
	OsFLA24	LOC_Os03g57460	
	OsFLA25	LOC_Os06g44660	
	OsFLA26	LOC_Os05g38500	
	OsFLA27	LOC_Os09g30486	
Populus trichocarpa	*PtFLA01*	Potri.019G121200	AY607753
	PtFLA03	Potri.013G151500	AY607755
	PtFLA04	Potri.013G014200	AY607756
	PtFLA05	Potri.019G123200	AY607757
	PtFLA2/6	Potri.013G151400	AY607754, AY607758
	PtFLA07	Potri.012G015000	AY607759
	PtFLA08	Potri.009G012200	AY607760
	PtFLA09	Potri.004G210600	AY607761
	PtFLA10	Potri.009G012100	AY607762
	PtFLA11	Potri.016G088700	AY607763
	PtFLA12	Potri.014G162900	AY607764
	PtFLA13	Potri.006G129200	AY607765
	PtFLA14	Potri.001G320800	AY607766
	PtFLA15	Potri.015G129400	AY607767
	PtFLA16	Potri.001G037800	
	PtFLA17	Potri.001G367900	
	PtFLA18	Potri.002G223300	
	PtFLA19	Potri.006G174900	
	PtFLA20	Potri.006G200300	
	PtFLA21	Potri.008G012400	
	PtFLA22	Potri.008G127500	
	PtFLA23	Potri.010G192300	
	PtFLA24	Potri.010G244900	
	PtFLA25	Potri.011G093500	
	PtFLA26	Potri.012G006200	
	PtFLA27	Potri.012G127900	
	PtFLA28	Potri.013G120600	
	PtFLA29	Potri.013G151300	
	PtFLA30	Potri.014G071700	
	PtFLA31	Potri.014G168100	
	PtFLA32	Potri.015G013300	
	PtFLA33	Potri.016G066500	
	PtFLA34	Potri.017G111600	
	PtFLA35	Potri.018G097000	
	PtFLA36	Potri.019G008400	
	PtFLA37	Potri.019G093300	
	PtFLA38	Potri.019G120800	
	PtFLA39	Potri.019G120900	
	PtFLA40	Potri.019G121100	

	PtFLA41	Potri.019G121300	
	PtFLA42	Potri.019G122600	
	PtFLA43	Potri.019G122800	
	PtFLA44	Potri.019G123000	
	PtFLA45	Potri.019G123100	
Ricinus communis	*RcFLA01*	27688.m000062	
	RcFLA02	28226.m000841	
	RcFLA03	29333.m001081	
	RcFLA04	29629.m001374	
	RcFLA05	29642.m000283	
	RcFLA06	29643.m000324	
	RcFLA07	29643.m000325	
	RcFLA08	29643.m000327	
	RcFLA09	29643.m000328	
	RcFLA10	29643.m000329	
	RcFLA11	29643.m000330	
	RcFLA12	29643.m000332	
	RcFLA13	29643.m000333	
	RcFLA14	29643.m000334	
	RcFLA15	29643.m000335	
	RcFLA16	29643.m000336	
	RcFLA17	29669.m000803	
	RcFLA18	29854.m001118	
	RcFLA19	29983.m003116	
	RcFLA20	30059.m000469	
	RcFLA21	30073.m002254	
	RcFLA22	30078.m002345	
	RcFLA23	30131.m006914	
	RcFLA24	30174.m008909	
	RcFLA25	32324.m000014	
Solanum lycopersicum	*SlFLA01*	Solyc01g091530	
	SlFLA02	Solyc03g112880	
	SlFLA03	Solyc06g075220	
	SlFLA04	Solyc06g076110	
	SlFLA05	Solyc07g045440	
	SlFLA06	Solyc07g048090	
	SlFLA07	Solyc07g053530	
	SlFLA08	Solyc07g053540	
	SlFLA09	Solyc07g065540	
	SlFLA10	Solyc08g006300	
	SlFLA11	Solyc09g007650	
	SlFLA12	Solyc09g007660	
	SlFLA13	Solyc10g005960	
	SlFLA14	Solyc10g081720	
	SlFLA15	Solyc11g005490	
	SlFLA16	Solyc11g069250	
	SlFLA17	Solyc12g006110	
	SlFLA18	Solyc12g015690	
Selaginella moellendorffii	*SmFLA01*	64323	
	SmFLA02	58104	
	SmFLA03	441324	
	SmFLA04	442211	
	SmFLA05	448915	

	SmFLA06	442972	
	SmFLA07	421625	
	SmFLA08	54466	
	SmFLA09	55113	
	SmFLA10	430281	
	SmFLA11	55277	
Vitis vinifera	*VvFLA01*	GSVIVT01014684001	
	VvFLA02	GSVIVT01020616001	
	VvFLA03	GSVIVT01022545001	
	VvFLA04	GSVIVT01023712001	
	VvFLA05	GSVIVT01025167001	
	VvFLA06	GSVIVT01025660001	
	VvFLA07	GSVIVT01025662001	
	VvFLA08	GSVIVT01025663001	
	VvFLA09	GSVIVT01025665001	
	VvFLA10	GSVIVT01025669001	
	VvFLA11	GSVIVT01025671001	
	VvFLA12	GSVIVT01030085001	
Zea Mays	*ZmFLA01*	AC184130	
	ZmFLA02	AC209784	
	ZmFLA03	AC213621	
	ZmFLA04	AC234156	
	ZmFLA05	GRMZM2G001514	
	ZmFLA06	GRMZM2G003165	
	ZmFLA07	GRMZM2G003752	
	ZmFLA08	GRMZM2G003917	
	ZmFLA09	GRMZM2G021794	
	ZmFLA10	GRMZM2G022931	
	ZmFLA11	GRMZM2G025240	
	ZmFLA12	GRMZM2G027825	
	ZmFLA13	GRMZM2G035933	
	ZmFLA14	GRMZM2G051185	
	ZmFLA15	GRMZM2G065718	
	ZmFLA16	GRMZM2G081017	
	ZmFLA17	GRMZM2G084812	
	ZmFLA18	GRMZM2G109291	
	ZmFLA19	GRMZM2G133053	
	ZmFLA20	GRMZM2G134528	
	ZmFLA21	GRMZM2G144610	
	ZmFLA22	GRMZM2G148534	
	ZmFLA23	GRMZM2G174799	
	ZmFLA24	GRMZM2G177142	
	ZmFLA25	GRMZM2G177242	
	ZmFLA26	GRMZM2G301908	
	ZmFLA27	GRMZM2G329181	
	ZmFLA28	GRMZM2G341698	
	ZmFLA29	GRMZM2G403276	
	ZmFLA30	GRMZM2G421415	

Genome assemblies for plant species can be obtained from Phytozome (version 9.1; www.phytozome.net).

Table S6 Primers and hydrolysis probes used in qRT-PCR analysis

Gene	Forward Primer	Reverse Primer	Hydrolysis Probe
LuFLA01	tcccctattgttcaggtgga	gatagtgctacaagccgcatt	133
LuFLA02	actcccgcagcaactatgaa	gggagactgtgaaagggtga	52
LuFLA04	ttccacatccacacatctcc	gctgatgaggaagcggtaag	101
LuFLA05	tggcgtggataatgatgatg	catccgcagctaccaacttc	78
LuFLA06	gggactgatatccctccaatg	cagcagccggagaaacag	15
LuFLA07	acgggactaagaagcttcacc	aacccgaagtccctggag	22
LuFLA08	ccaaggaatcaccctattcg	gatatggaggcttgttgatgg	13
LuFLA09	caaggataagtccggtggag	gctaaaacagagagcccaacc	49
LuFLA10	gcttgtcaaggtcaacgtca	tcgatagcaattgggttactga	35
LuFLA11	ccaccgtcacaatgtcacac	agaagaaaagggtgaaggtgaa	70
LuFLA12	tcctgatgccgatagtccat	gagcttgctgatgatgatgc	59
LuFLA13	gcaggactgtcaaggtctcc	tacgaaagatgagggcgttc	117
LuFLA14	cgctcatgattttcgcagt	acggagatttcccgaacag	7
LuFLA15	gtttaaagtggataaggttttgcag	tagccgacttagcacccttg	17
LuFLA16	aaatgtgtcatcagaccaactga	ggagtagtagtgcggcaagg	150
LuFLA17	gctgacgacggatcagcta	ctccgggatgatgtggtagt	50
LuFLA18	caattagaatccgccgtctc	cggttggtggagtggttt	7
LuFLA19	gtacttctccttgtccgacctc	cgacccgtcgtttggtag	153
LuFLA20	ttcaaggatgcccccttcc	gctgctagatttgggtgcat	17
LuFLA21	attgggttcggattgctct	aggattcatatgatcaaccaaaataa	50
LuFLA22	tccacatcctcccaactttc	gtcctcagagggttgctcac	154
LuFLA23	gctctccacataaggctcca	aaggagcaatcttcctgtgtct	26
LuFLA24	aacaagccggattctccat	agaagaaggtgtcggggttt	63
LuFLA25	ctcccgtcttcccgattta	gggagctaaggcgtcgtaa	63
LuFLA26	gcctagaataccgccacaat	ctaggaggcgggtaatgttg	63
LuFLA27	ctctgtctcctcgccttcac	gcccaagagggaagttacg	41
LuFLA28	atctcccaggagagccaagt	ccgacgtatcttccgatgag	152
LuFLA29	attcccctgtcgattcttcc	aaaagccaccaaaccaagg	103
LuFLA30	gggacaagctctgcttcatc	tcatcatgaacgcgatgg	38
LuFLA31	agcgatcaaacatccgtaca	tccgaattttcctagcttcttc	8
LuFLA32	gggtgatcatgagggtcaat	atattgcccggactcttcg	144
LuFLA33	acctaaccgcaatccttgaa	acttgggttgtgccaagg	164
LuFLA34	tccaaatgtcctgctctcct	cagagcagcttccgttgag	101
LuFLA35	tgtcaagttccacgtcatcc	cctcggtagcaattgggtta	35
LuFLA36	ttcgtggagttcaatcctca	ttattccaaagcctcccaga	9
LuFLA37	tcatcccctgctcaatcc	gtgggtgtttcagcgtcag	165
LuFLA38	tgaaaatgcaatggcaaact	ctcgtcctgggctagagatg	22
LuFLA39	ttatctccttcaactctctccaatg	gatgttgtgagcggagacg	101
LuFLA40	ctcaatatctacagtcctagccttca	agtgacggtcgagtggagtc	133
LuFLA41	aaactcaaaattcggcgaga	aacaccgtgaaagcaaggtc	31
LuFLA42	tcccatctctattctacctcatca	gaggatggtggcgatgtc	70
LuFLA43	gtcagcgacttgatcactgc	ctagcgccgcctaattgat	63

Oligonucleotide primer sequences and probes for flax FLA genes were obtained from the Universal Probe Library Assay Design Center (http://tinyurl.com/7u6s5bh).

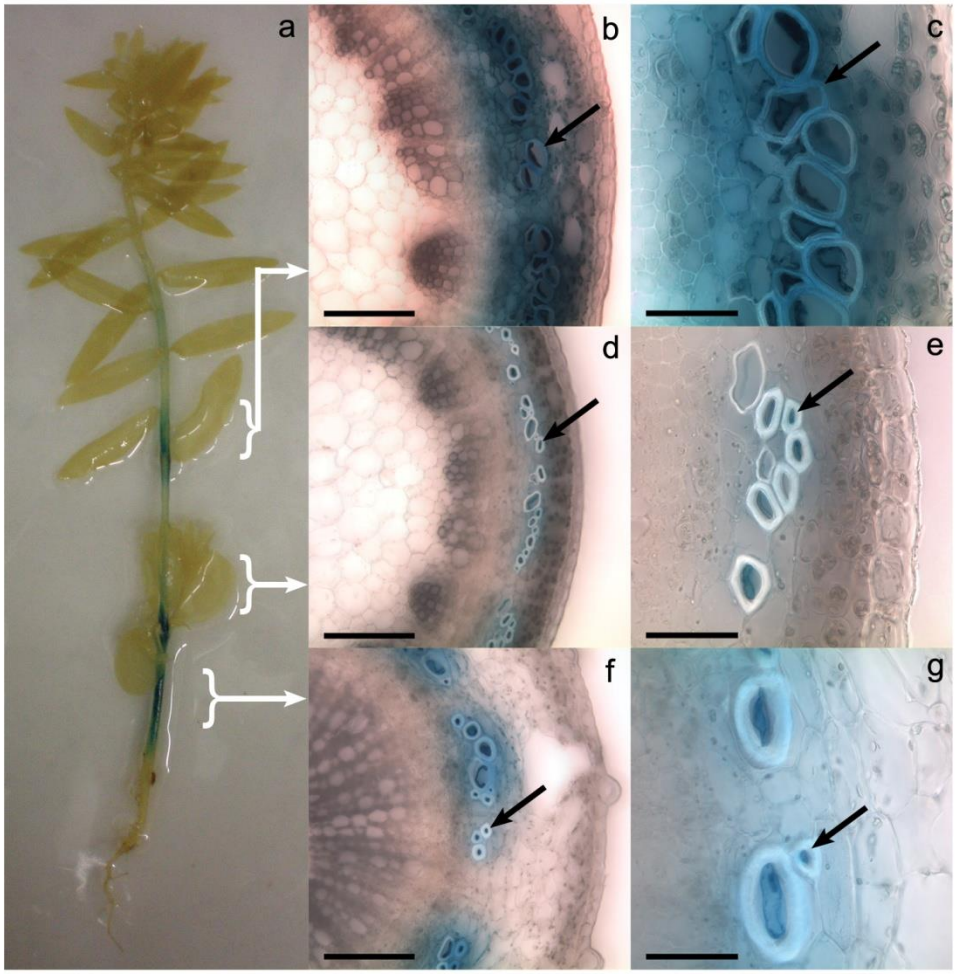

Figure S1 GUS staining in 3-week old LuFLA1$_{PRO}$:*uidA* transgenic flax (a). Transverse sections of a leaf node were examined (b, c), as were sections at the stem base (d, e), and sections at the hypocotyl (f, g). Black arrows indicate GUS stained bast fibres. Scale bar for b, e, and h is 150 µm. Scale bar for c, f, and i is 50 µm.

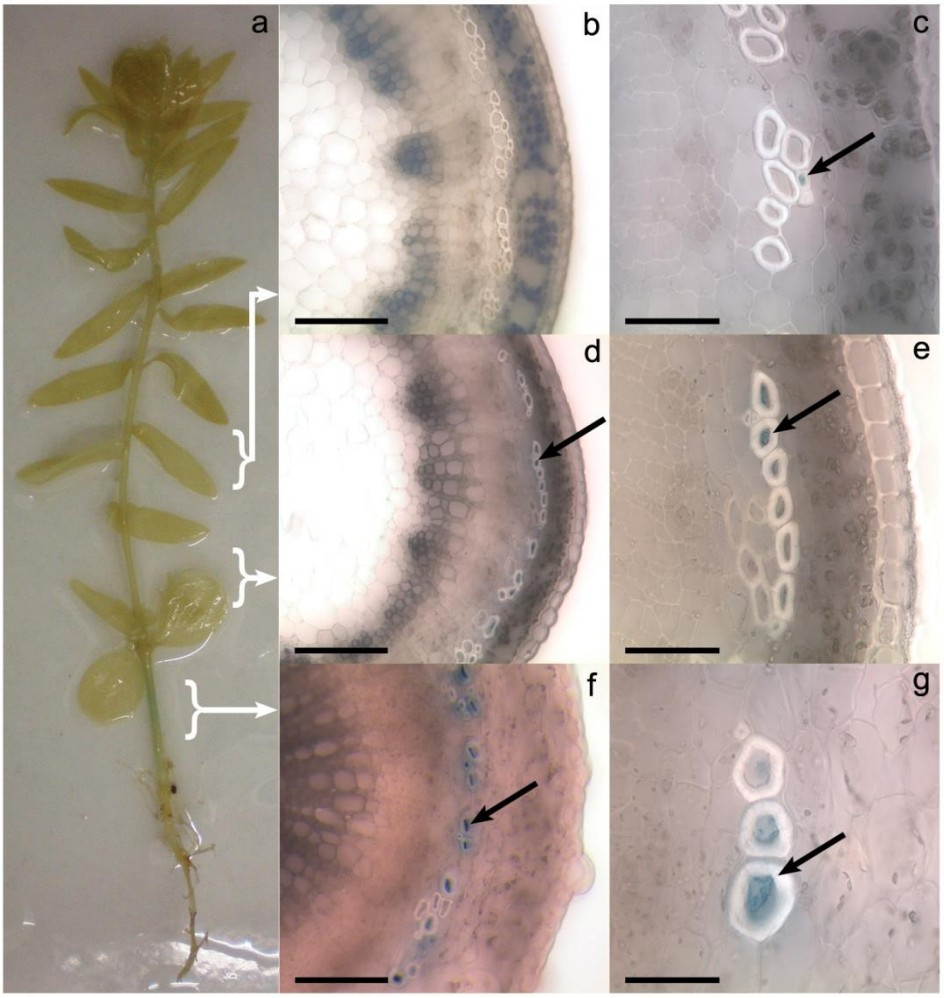

Figure S2 GUS staining in 3-week old LuBGAL1$_{PRO}$:*uidA* transgenic flax (a). Transverse sections of a leaf node were examined (b, c), as were sections at the stem base (d, e), and sections at the hypocotyl (f, g). Black arrows indicate GUS stained bast fibres. Scale bar for b, e, and h is 150 µm. Scale bar for c, f, and i is 50 µm.

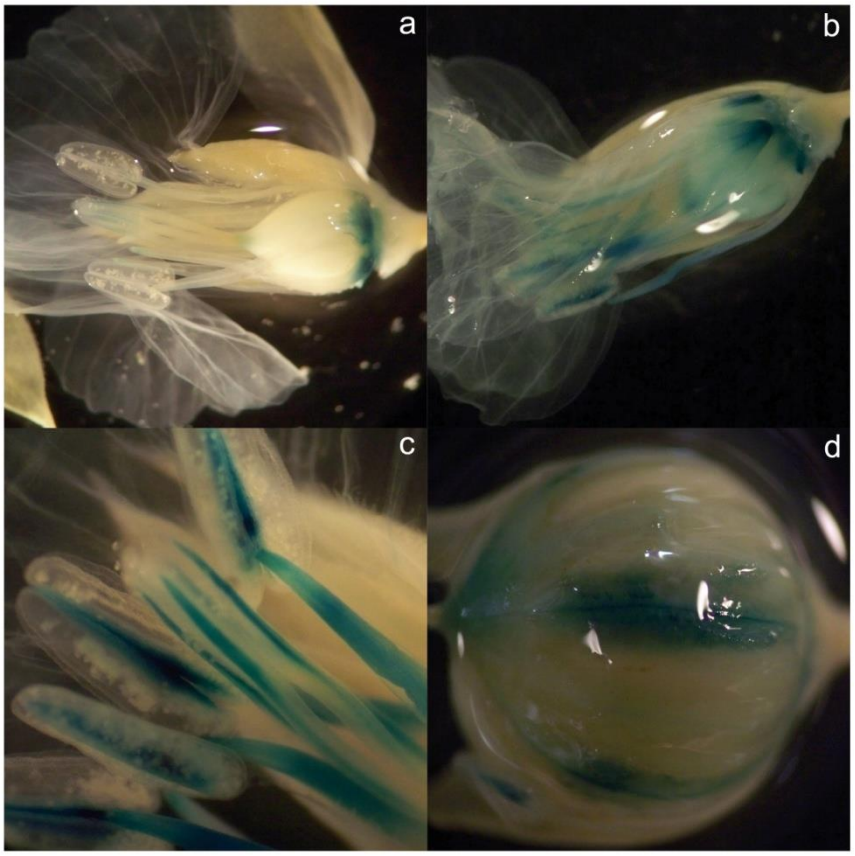

Figure S3 GUS staining in reproductive tissues of LuBGAL1$_{PRO}$:*uidA* transgenic flax.
Developing flowers (a, b), anthers, filaments stigma and style (c), and seed boll (d).

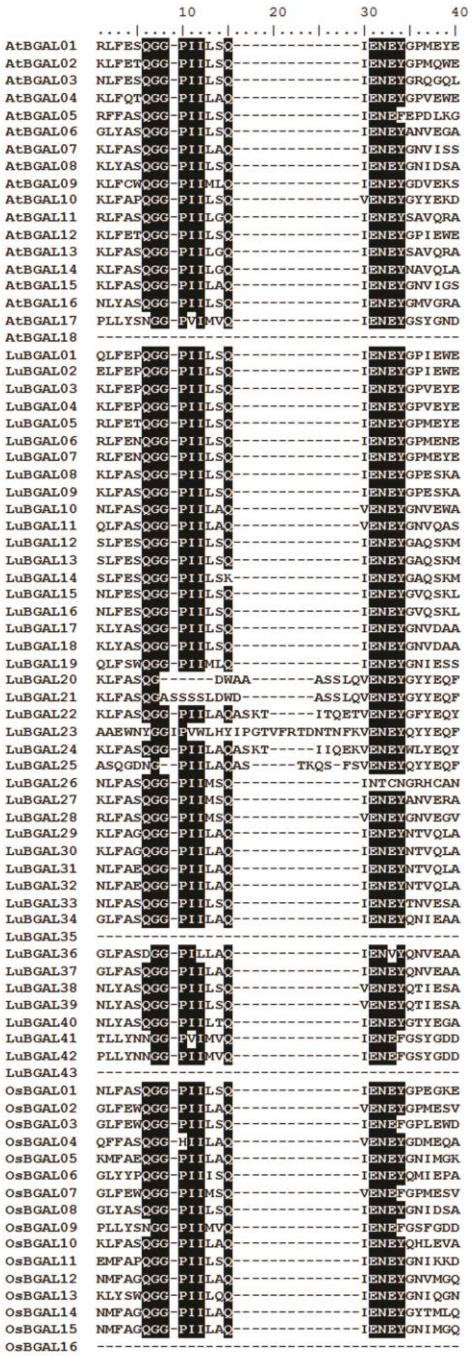

Figure S4 Putative GH35 active site in various plant species. The GH35 active site (Henrissat, 1998), was identified by searching for the consensus sequence G-G-P-[LIVM](2)-x(2)-Q-x-E-N-E-[FY]. Gaps or missing sequence are denoted by dashes '-'. Residues conserved amidst 90% of the sequences are highlighted. The flax sequences are named LuBGAL, and numbered according to Tables 3-1 and 3-2. *Arabidopsis thaliana* sequences are indicated as AtBGAL, and numbered according to existing designations (Ahn et al., 2007). *Oryza sativa* sequences are indicated as OsBGAL, and numbered according to existing designations (Tanthanuch et al., 2008). Genomic loci corresponding to these sequences are presented in Table 3-1 and Table S3.

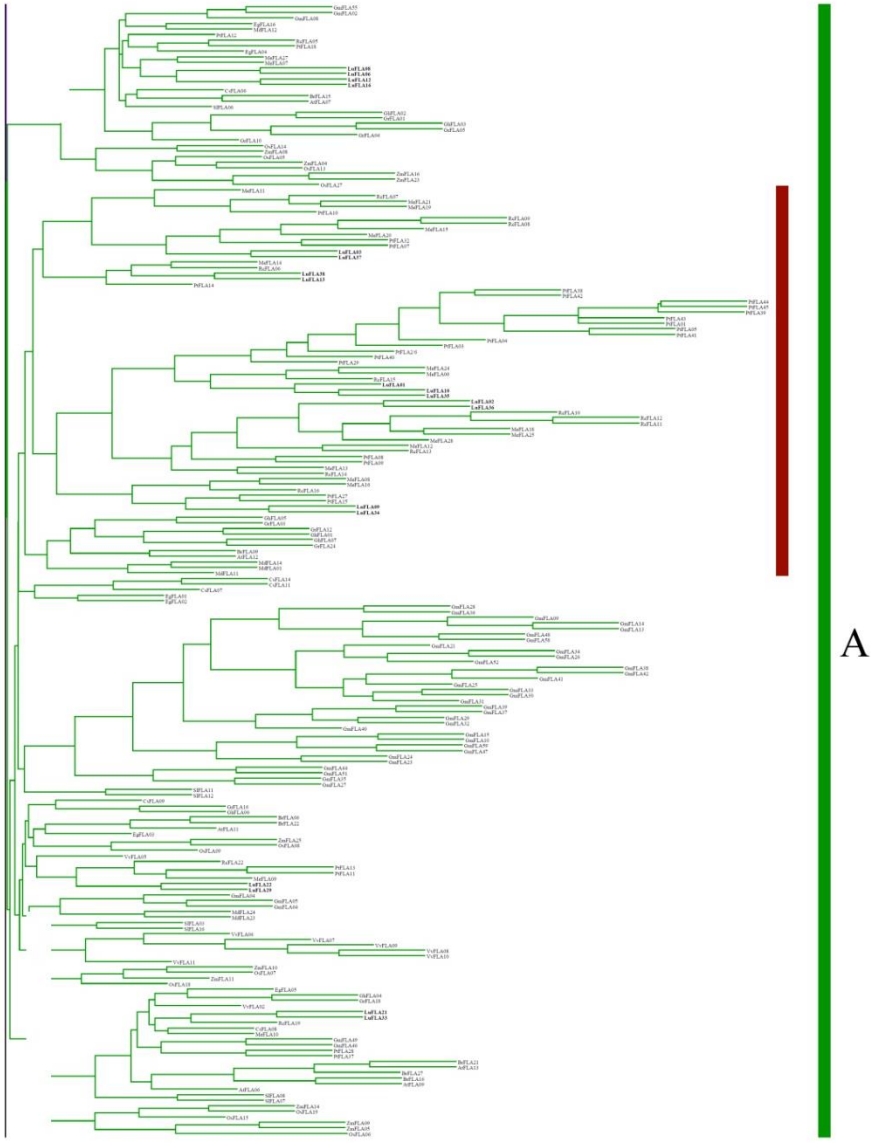

Figure S5: Phylogenetic relationship among the group A fasciclin-like arabinogalactan proteins of flax. Deduced amino acid sequences were aligned with MUSCLE (Edgar, 2004). The tree was created with GARLI (Zwickl, 2006), using the maximum likelihood method, following the WAG model of amino acid substitutions (Whelan and Goldman, 2001), while employing gamma-distributed rate variations and empirically determined base frequencies. A consensus tree of 1000 bootstrap replicates was produced. The flax sequences are named LuFLA, and numbered according to Table 5-1. *Arabidopsis thaliana* sequences are indicated as AtFLA, and numbered according to existing designations (Johnson et al., 2003). *Brassica rapa* sequences are indicated as BrFLA, and numbered according to existing designations (Jun and Xiaoming, 2012). *Oryza sativa* sequences are indicated as OsFLA, and numbered according to existing designations (Ma and Zhao, 2010). *Populus trichocarpa* sequences are indicated as PtFLA, and numbered according to existing designations (Lafarguette et al., 2004). Genomic loci corresponding to these and other sequences are presented in Table S5. The red bar denotes a large clade of Malpighiales.

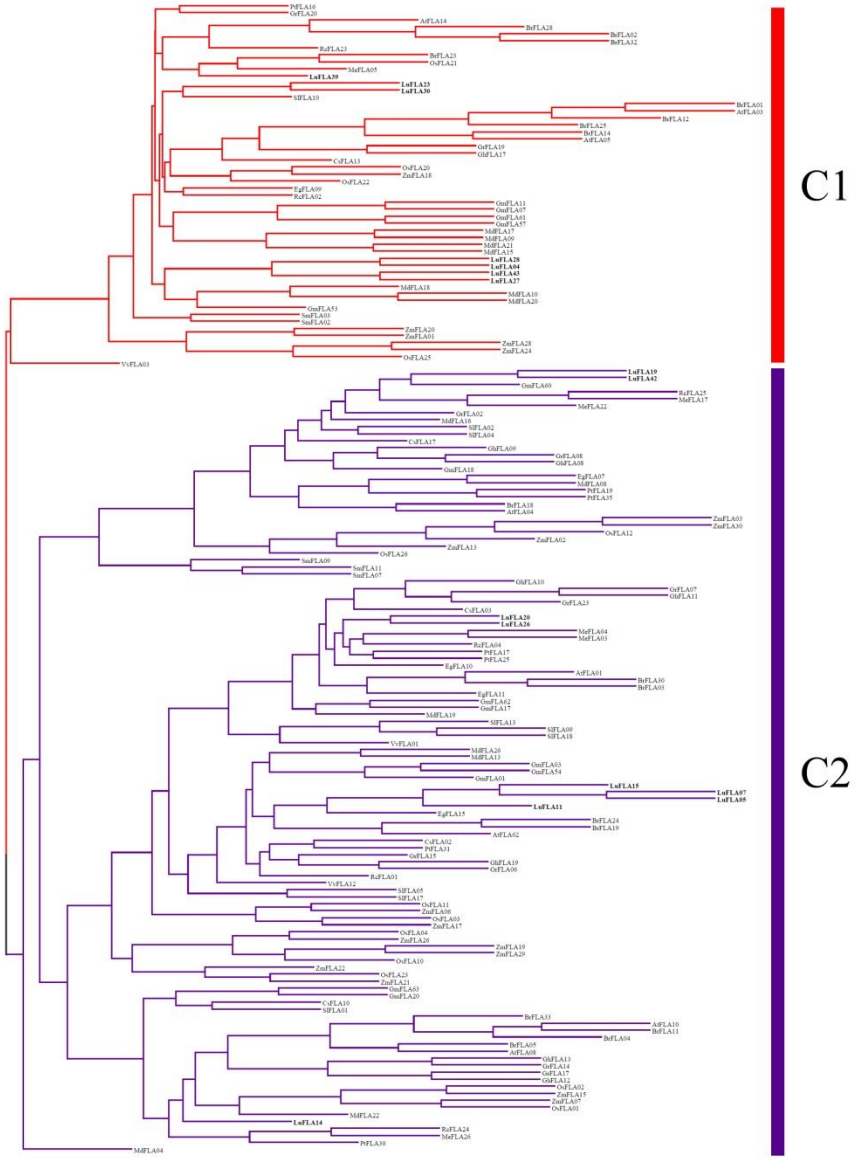

Figure S6: Phylogenetic relationship among the group C1 and C2 fasciclin-like arabinogalactan proteins of flax. Deduced amino acid sequences were aligned with MUSCLE (Edgar, 2004). The tree was created with GARLI (Zwickl, 2006), using the maximum likelihood method, following the WAG model of amino acid substitutions (Whelan and Goldman, 2001), while employing gamma-distributed rate variations and empirically determined base frequencies. A consensus tree of 1000 bootstrap replicates was produced. The flax sequences are named LuFLA, and numbered according to Table 5-1. *Arabidopsis thaliana* sequences are indicated as AtFLA, and numbered according to existing designations (Johnson et al., 2003). *Brassica rapa* sequences are indicated as BrFLA, and numbered according to existing designations (Jun and Xiaoming, 2012). *Oryza sativa* sequences are indicated as OsFLA, and numbered according to existing designations (Ma and Zhao, 2010). *Populus trichocarpa* sequences are indicated as PtFLA, and numbered according to existing designations (Lafarguette et al., 2004). Genomic loci corresponding to these and other sequences are presented in Table S5.

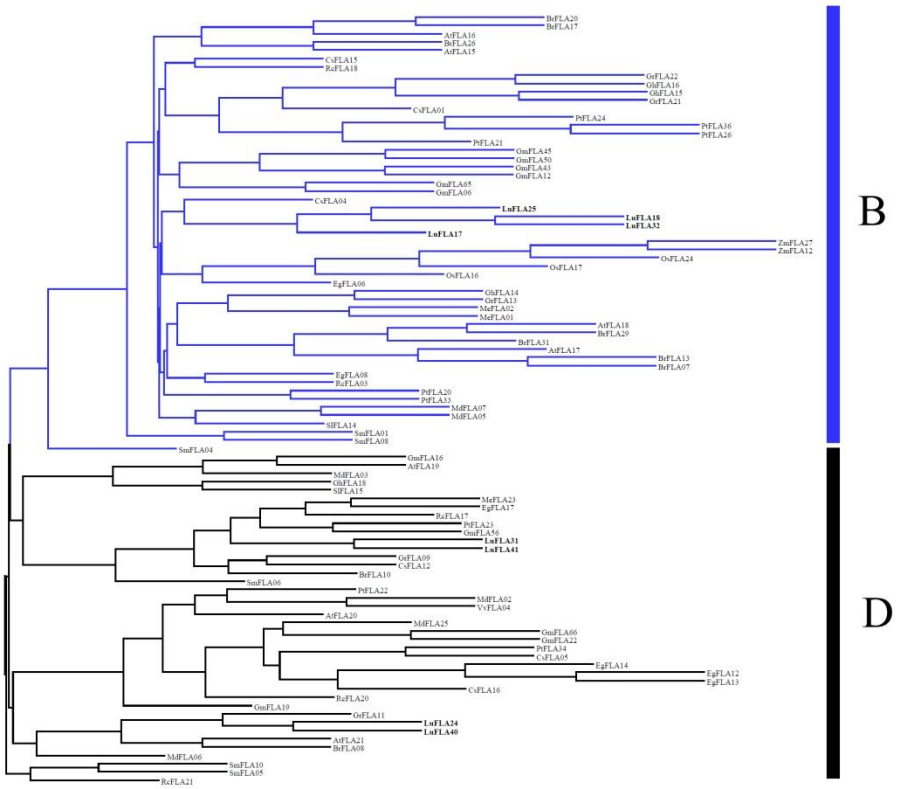

Figure S7: Phylogenetic relationship among the group B and D fasciclin-like arabinogalactan proteins of flax. Deduced amino acid sequences were aligned with MUSCLE (Edgar, 2004). The tree was created with GARLI (Zwickl, 2006), using the maximum likelihood method, following the WAG model of amino acid substitutions (Whelan and Goldman, 2001), while employing gamma-distributed rate variations and empirically determined base frequencies. A consensus tree of 1000 bootstrap replicates was produced. The flax sequences are named LuFLA, and numbered according to Table 5-1. *Arabidopsis thaliana* sequences are indicated as AtFLA, and numbered according to existing designations (Johnson et al., 2003). *Brassica rapa* sequences are indicated as BrFLA, and numbered according to existing designations (Jun and Xiaoming, 2012). *Oryza sativa* sequences are indicated as OsFLA, and numbered according to existing designations (Ma and Zhao, 2010). *Populus trichocarpa* sequences are indicated as PtFLA, and numbered according to existing designations (Lafarguette et al., 2004). Genomic loci corresponding to these and other sequences are presented in Table S5.