



# DYNAMICS OF CELL WALL PROTEINS DURING CELL ELONGATION IN ETIOLATED HYPOCOTYLS OF ARABIDOPSIS THALIANA AS SHOWN BY PROTEOMIC AND TRANSCRIPTOMIC SURVEYS

Muhammad Irshad

## ► To cite this version:

Muhammad Irshad. DYNAMICS OF CELL WALL PROTEINS DURING CELL ELONGATION IN ETIOLATED HYPOCOTYLS OF ARABIDOPSIS THALIANA AS SHOWN BY PROTEOMIC AND TRANSCRIPTOMIC SURVEYS. Vegetal Biology. Université Paul Sabatier - Toulouse III, 2008. English. NNT: . tel-00323217

**HAL Id: tel-00323217**

**<https://theses.hal.science/tel-00323217>**

Submitted on 19 Sep 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université  
de Toulouse



# THÈSE

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ DE TOULOUSE**

Délivré par : **Université Toulouse III-Paul Sabatier**

**UFR Sciences de la Vie et de la Terre (SVT)**

Spécialité : **Biosciences Végétales**

\*\*\*\*\*

présentée et soutenue par

**Muhammad IRSHAD**

le 9 juillet 2008

\*\*\*\*\*

**DYNAMIQUE DES PROTÉINES PARIÉTALES  
AU COURS DE L'ÉLONGATION CELLULAIRE  
DANS DES HYPOCOTYLES ÉTIOLÉS  
D'*ARABIDOPSIS THALIANA* :  
APPROCHES PROTÉOMIQUE ET TRANSCRIPTOMIQUE**

\*\*\*\*\*

Directeurs de thèse : **Dr. Elisabeth JAMET** et **Pr. Dr. Rafael PONT-LEZICA**

\*\*\*\*\*

**Ecole doctorale:** Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénierie

**Unité de recherche:** Surfaces Cellulaires et Signalisation chez les Végétaux, UMR 5546,  
UPS, CNRS, Pôle de Biotechnologies Végétales, 24, chemin de Borde Rouge, BP 42617  
Auzerville, 31326 Castanet-Tolosan, FRANCE

\*\*\*\*\*

## JURY

Dr. Samantha VERNHETTES

Dr. Gilles PILATE

Prof. Dr. Chantal TEULIERES

Dr. Elisabeth JAMET

Chargée de recherche INRA, Versailles

Directeur de recherche INRA, Orléans

Professeur UPS

Directeur de recherche CNRS

Rapporteur

Rapporteur

Président du jury

Directeur de thèse

## MEMBRE INVITÉ

Prof. Dr. Rafael PONT-LEZICA

Professeur UPS

Directeur de thèse





Université  
de Toulouse



## THESIS

for obtaining the degree of

### DOCTORATE FROM UNIVERSITÉ DE TOULOUSE

Delivered by: **Université Toulouse III-Paul Sabatier**

**UFR Sciences de la Vie et de la Terre (SVT)**

Specialization: **Biosciences Végétales**

\*\*\*\*\*

presented and defended by

**Muhammad IRSHAD**

on July 9, 2008

\*\*\*\*\*

### **DYNAMICS OF CELL WALL PROTEINS DURING CELL ELONGATION IN ETIOLATED HYPOCOTYLS OF *ARABIDOPSIS THALIANA* AS SHOWN BY PROTEOMIC AND TRANSCRIPTOMIC SURVEYS**

\*\*\*\*\*

Supervisors: **Dr. Elisabeth JAMET** and **Prof. Dr. Rafael PONT-LEZICA**

\*\*\*\*\*

**Doctoral school:** Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénierie

**Research Lab:** Surfaces Cellulaires et Signalisation chez les Végétaux, UMR 5546,  
UPS, CNRS, Pôle de Biotechnologies Végétales, 24 chemin de Borde Rouge, BP 42617  
Auzeville, 31326 Castanet-Tolosan, FRANCE

\*\*\*\*\*

#### **JURY**

Dr. Samantha VERNHETTES  
Dr. Gilles PILATE  
Prof. Dr. Chantal TEULIERES  
Dr. Elisabeth JAMET

Chargée de recherche INRA, Versailles  
Directeur de recherche INRA, Orléans  
Professeur UPS  
Directeur de recherche CNRS

Reporter  
Reporter  
President of the jury  
Supervisor

#### **INVITED MEMBER**

Prof. Dr. Rafael PONT-LEZICA

Professeur UPS

Supervisor



“If you detect any mistakes of mine, I rely on your superior knowledge to excuse them; for who has ever avoided errors in the wide-extended field of Nature? Who is furnished with sufficient stock of observations? I shall be thankful for your friendly corrections, I have done what I could myself” .

Linnaeus,  
Father of modern Taxonomy

## DEDICATION

I dedicate this humble effort of mine  
to my loving parents, my brother Israr and  
my sweet wife Lubna, without the love, devotion,  
sincerity, guidance and encouragement of whom,  
I would have been lost.

---

## ACKNOWLEDGEMENTS

I would like to thank **Dr. Simon HAWKINS**, **Dr. Gilles PILATE**, **Prof. Dr. Chantal TEULIERES** and **Dr. Samantha VERNHETTES** for evaluating this work.

Sincere thanks are due to **Prof. Dr. Rafael PONT-LEZICA** for accepting me in his group to accomplish this work.

I offer special thanks to my supervisors **Dr. Elisabeth JAMET** and **Prof. Dr. Rafael PONT-LEZICA** for their prudent and companionate guidance, sincere advices, constant encouragement, kind and inspiring behavior, critically reviewing the manuscript and help up to the last moment, without which it would have been impossible to accomplish this work.

I would like to reflect my gratitude to **Dr. Gisèle BORDERIES**, **Miss. Carole PICHEREAUX** and **Dr. Michel ROSSIGNOL** for their help and cooperation during MALDI-TOF analysis.

I am highly grateful to **Dr. Hervé CANUT**, **Mr. David ROUJOL**, **Miss. Hélène SAN CLEMENTE**, **Mrs. Ludivine SOUBIGOU-TACONNAT** and **Dr. Jean-Pierre RENOU** who in addition to my supervisors, contributed to the analyses of proteomic and transcriptomic data or experimental work during this study.

I would like to reflect my heartily gratitude to my colleagues **Cécile**, **Georges**, **Martine**, **Geneviève**, **Tan**, **Philippe** and the whole group of **Jean-Philippe** for their help, encouragement, valuable suggestions, technical help and criticisms during this study.

I must acknowledge the “**services communs**” of UMR 5546 for their sincere help and full cooperation during my work in the lab.

I am highly grateful to **Higher Education Commission of Pakistan**, Islamabad for grant of scholarship, to **SFERE**, France for their wholehearted and sincere cooperation and to **Government of NWFP, Higher Education Department**, Peshawar, for grant of study leave.

Last but not least, I extend my respect, love and thanks to my family members, especially my **parents** and my brother **Israr** without whose affection, whole hearted wishes and help in administrative works, it would have been impossible to complete this work. How can I forget the patience and sacrifice of my wife **Lubna** and my son **Talha** while I have been working during evenings and at week-ends.

---

## CONTENTS

<b>LIST OF ABBREVIATIONS .....</b>	<b>1</b>
------------------------------------	----------

### CHAPTER 1. INTRODUCTION

<b>Chapter summary (French) .....</b>	<b>4</b>
<b>1.1. Cell wall components .....</b>	<b>7</b>
1.1.1. Polysaccharides .....	7
1.1.1.1. Cellulose .....	7
1.1.1.2. Cross-linking glycans .....	8
1.1.1.3. Pectic matrix .....	8
1.1.2. Cell wall proteins .....	9
1.1.2.1. Structural proteins .....	10
1.1.2.2. Arabinogalactan proteins (AGPs) .....	11
1.1.2.3. Other proteins .....	11
1.1.3. Other cell wall components .....	12
1.1.3.1. Lignins .....	12
1.1.3.2. Lipids .....	13
1.1.3.3. Suberin .....	14
1.1.3.4. Inorganic salts .....	14
<b>1.2. Cell wall in relation to cell elongation and growth .....</b>	<b>15</b>
<b>1.3. <i>A. thaliana</i> etiolated hypocotyls as a model for cell elongation .....</b>	<b>17</b>
<b>1.4. Objectives of the work .....</b>	<b>18</b>

### CHAPTER 2. MATERIALS AND METHODS

<b>2.1. Materials .....</b>	<b>19</b>
2.1.1. Plant material .....	19
2.1.2. <i>In vitro</i> culture for obtaining etiolated hypocotyls .....	19
<b>2.2. Methods .....</b>	<b>19</b>
2.2.1. Isolation of cell walls from hypocotyls .....	19
2.2.2. Sequential extraction of proteins from purified cell walls .....	20
2.2.3. Protein fractionation by cation exchange chromatography .....	20
2.2.4. 1D-E (SDS-polyacrylamide gel electrophoresis) and staining procedures .....	21
2.2.4.1. Staining with Coomassie Brilliant Blue (CBB) .....	21

2.2.4.2. Staining with silver nitrate .....	22
2.2.4.3. Staining with Amido Black .....	22
2.2.4.4. Staining with Gelcode® Glycoprotein staining kit .....	22
2.2.5. Transfer of proteins from gel on PVDF membrane .....	23
2.2.6. Protein identification by mass spectrometry .....	23
2.2.7. Extraction of RNAs .....	24
2.2.8. RT-PCR and semi-quantitative PCR .....	25
2.2.9. Microarray analysis .....	26
2.2.10. Statistical analysis of microarray data .....	26
2.2.11. Data Deposition .....	27
2.2.12. Bioinformatic analysis .....	27

### **CHAPTER 3. EVALUATION OF CELL WALL PREPARATIONS FOR PROTEOMICS: A NEW PROCEDURE FOR PURIFYING CELL WALLS FROM ARABIDOPSIS HYPOCOTYLS**

<b>Chapter summary (French) .....</b>	<b>29</b>
<b>Abstract .....</b>	<b>34</b>
<b>Background .....</b>	<b>35</b>
<b>Results and discussion .....</b>	<b>35</b>
Analysis of early methods .....	35
A modified method to prepare plant cell walls .....	39
Sequential salt extraction of proteins from cell walls .....	43
<b>Conclusion .....</b>	<b>43</b>
<b>Methods .....</b>	<b>44</b>
Plant material and isolation of cell walls .....	44
Sequential proteins extraction and identification .....	44
<b>Additional material (List) .....</b>	<b>45</b>
<b>References .....</b>	<b>45</b>

### **CHAPTER 4. A NEW PICTURE OF CELL WALL PROTEIN DYNAMICS IN ELONGATING CELLS OF *ARABIDOPSIS THALIANA*: KNOWN PLAYERS AND NEW COMERS**

<b>Chapter summary (French) .....</b>	<b>47</b>
<b>Summary .....</b>	<b>54</b>



---

<b>Keywords .....</b>	<b>55</b>
<b>Abbreviations .....</b>	<b>55</b>
<b>Introduction .....</b>	<b>56</b>
<b>Material and Methods .....</b>	<b>58</b>
Plant material .....	58
Cell wall purification and protein extraction .....	58
Protein separation by cationic exchange chromatography .....	58
Protein separation by mono-dimensional electrophoresis (1D-E) and identification ..	59
Semi-quantification .....	59
Bioinformatic analyses .....	59
<b>Results .....</b>	<b>61</b>
Establishment of methods for efficient proteomic analysis of hypocotyl cell walls ..	61
Proteins identified in cell wall extracts of Arabidopsis etiolated hypocotyls .....	62
Semi-quantitative comparative analysis of CWPp .....	63
<b>Discussion .....</b>	<b>66</b>
<b>Additional data files (List) .....</b>	<b>70</b>
<b>References .....</b>	<b>71</b>
<b>CHAPTER 5. CELL WALL BIOGENESIS OF <i>ARABIDOPSIS</i> ELONGATING CELLS: TRANSCRIPTOMICS COMPLEMENTS PROTEOMICS</b>	
<b>Chapter summary (French) .....</b>	<b>77</b>
<b>Abstract .....</b>	<b>82</b>
<b>Background .....</b>	<b>83</b>
<b>Results and Discussion .....</b>	<b>84</b>
Level of transcripts of cell wall genes (CWGs) during hypocotyl elongation .....	84
Genes encoding secreted proteins with high or moderate level of transcripts in etiolated hypocotyls .....	86
Are there variations in the level of transcripts between half- and fully-grown hypocotyls? .....	87
Transcriptome vs. proteome .....	89
<b>Conclusions .....</b>	<b>91</b>
<b>Methods .....</b>	<b>92</b>
Plant material .....	92
Total RNA extraction .....	92

---

---

Transcriptome studies .....	93
Statistical analysis of microarray data .....	93
Data deposition .....	94
Semi-quantitative PCR .....	94
Bioinformatic analyses .....	94
<b>References .....</b>	<b>96</b>
<b>Additional data files (List) .....</b>	<b>103</b>
 <b>CHAPTER 6. CONCLUSIONS AND PERSPECTIVES</b>	
French version .....	104
English version .....	108
 <b>LITERATURE CITED .....</b>	 <b>111</b>
 <b>ANNEXES AND SUPPLEMENTARY DATA</b>	
Annexes chapter 2 .....	125
Supplementary data chapter 3 .....	128
Supplementary data chapter 4 .....	144
Supplementary data chapter 5 .....	171

---

## List of abbreviations

1D-E/2D-E: one or two dimensional electrophoresis  
4CL: 4-coumarate CoA ligase  
ACN: acetonitrile  
AG: arabinogalactan  
AGP: arabinogalactan protein  
AGP-PRP: arabinogalactan protein–proline-rich protein  
AGRIKOLA: *Arabidopsis* genomic rna knock-out line analysis  
BSA: bovine serum albumin  
BY2: bright yellow (cv of *Nicotiana tabacum*)  
C3H: p-coumarate 3-hydroxylase  
C4H: cinnamate 4-hydroxylase  
CAD: cinnamyl alcohol dehydrogenase  
CATMA: complete *Arabidopsis* transcriptome microarray  
CBB: Coomassie brilliant blue  
CCoAOMT: caffeoyl-CoA O-methyltransferase  
CCR: hydroxycinnamoyl-CoA reductase  
cDNA: complementary DNA  
CESA: cellulose-synthase  
Col 0: ecotype Columbia 0 of *A. thaliana*  
COMT: caffeic acid/5-hydroxyferulic acid O-methyltransferase  
CSC: cellulose-synthesizing complex  
CSL: cellulose-synthase like  
CWG: cell wall gene  
CWP: cell wall protein  
dNTP: deoxynucleotide triphosphate (dATP, dCTP, dGTP, dTTP)  
DTT: dithiothreitol  
EDTA: ethylenediamine tetra acetic acid  
EGase: endo-1,4- $\beta$ -D-glucanase  
Endo-PG: endopolygalacturonase  
ER: endoplasmic reticulum  
F5H: ferulate 5-hydroxylase  
FLA: fasciclin-like arabinogalactan protein

FPLC: fast protein liquid chromatography  
 FWER: Family Wise Error Rate  
 G: guaiacyl (unit of lignin)  
 GalA: galacturonic acid  
 GAX: glucuronoabinoxylan  
 GFP: green fluorescent protein  
 GH: glycoside hydrolase  
 GPI: glycosylphosphatidylinositol  
 GRP: glycine-rich protein  
 GST: gene-specific sequence tag  
 GT: glycosyl transferase  
 H/PRP: hydroxyproline/proline-rich protein  
 H: p-hydroxyphenyl (unit of lignin)  
 HCT: p-hydroxycinnamoyl-CoA:quinic acid shikimate p-hydroxycinnamoyl-CoA transferase  
 HG: homogalacturonan  
 HGA: homogalacturonic acid  
 HRGP: hydroxyproline-rich protein  
 kDa: kilo Dalton  
 LAE: late-abundant embryogenesis protein  
 LB medium: Luria-Bertani medium  
 LC: liquid chromatography  
 LC-MS/MS: liquid chromatography-tandem mass spectrometry  
 LRR: leucine-rich repeat  
 LRX: leucine-rich repeat extensin  
 LTP: lipid transfer protein  
 MALDI-TOF: matrix-assisted laser desorption ionization-time of flight  
 MS medium: Murashige and Skoog medium  
 MS: mass spectrometry  
 nptI: neomycin phosphotransferase  
 PAGE: polyacrylamide gel electrophoresis  
 PAL: phenylalanine ammonia-lyase  
 PCR: polymerase chain reaction  
 PG: polygalacturonase  
 PGA: polygalacturonic acid

PGIP: poly galacturonase-inhibiting protein  
 PL: pectate lyase  
 PL: polysaccharide lyase  
 PM: plasma membrane  
 PME: pectin methylesterase  
 PRP: proline-rich protein  
 PTM: post-translational modification  
 PVDF: polyvinylidene fluoride  
 QI: semi-quantitative index  
 RG-I: rhamnogalacturonan-I  
 RG-II: rhamnogalacturonan-II  
 ROS: reactive oxygen species  
 rpm: round per min  
 RT-qPCR: reverse transcription-quantitative PCR  
 S: syringyl (unit of lignin)  
 SDS: sodium dodecyl sulphate  
 SPG: secretory pathway gene  
 TE buffer: Tris EDTA-buffer  
 Tris: Trishydroxymethylaminomethane, or 2-amino-2-hydroxymethyl-1,3-propanediol  
 U: unit  
 UHQ water: Ultra High Quality Water  
 XG: xyloglucan  
 XTH: xyloglucan endotransglucosylase-hydrolase

# **CHAPTER 1**

## **INTRODUCTION**

---

**Chapter summary (French)**

L'existence d'une paroi autour des cellules végétales est l'un des caractères qui les différencie des cellules animales. Cette paroi joue des rôles importants au cours du développement des plantes, et au cours de leurs interactions avec l'environnement qu'il s'agisse de stress biotiques ou abiotiques. Elle est également essentielle pour le port dressé des végétaux et permet la circulation de molécules signal entre cellules distantes. La croissance des végétaux fait intervenir deux processus cellulaires, qui sont la division cellulaire et l'augmentation de volume qui peut-être être soit anisotrope (grossissement), soit directionnelle (élongation). Ce travail porte sur la recherche de protéines importantes pour l'élongation des cellules végétales, et plus particulièrement pour les modifications des parois au cours de ce processus.

La première partie de cette introduction décrit les constituants pariétaux principaux : les polysaccharides (cellulose, hémicelluloses, pectines), les protéines, les lignines, les lipides, les cires et les ions. La description des protéines est volontairement succincte puisqu'elle reprend l'état des connaissances avant le début de mes travaux de thèse, *i.e.* avant l'essor des approches protéomiques qui ont permis d'avoir une vue globale des protéines pariétales.

Dans les deux parties suivantes sont présentés successivement les protéines dont le rôle dans l'élongation cellulaire est connu ainsi que le mécanisme particulier de l'élongation cellulaire dans les hypocotyles d'*Arabidopsis thaliana*. En effet, ces hypocotyles présentent la particularité de s'allonger environ 100 fois en un temps très court par simple élongation cellulaire.

Enfin, les objectifs de ma thèse sont développés. Il s'agissait d'avoir une vue globale de la régulation des gènes impliqués dans la biogenèse des parois au cours d'un processus d'élongation cellulaire chez *A. thaliana*. Le modèle expérimental retenu a été celui des hypocotyles étiolés. Les questions posées étaient les suivantes :

- Quelles sont les protéines présentes dans les parois des hypocotyles étiolés à deux stades de leur développement (croissance active vs croissance terminée) ?
- Existe-t-il des différences entre les protéomes pariétaux à ces deux stades développement ?
- Quelles sont les fonctions possibles des protéines identifiées ?

Deux approches complémentaires ont été menées : une approche protéomique impliquant la mise au point de protocoles de purification de parois (chapitre 3) et de séparation des protéines pariétales (chapitre 4) ; et une approche de transcriptomique (chapitre 5). L'ensemble des résultats obtenus a été interprété grâce à la bioinformatique non seulement pour la prédiction de la localisation sub-cellulaire des protéines, mais encore pour celle de domaines fonctionnels.



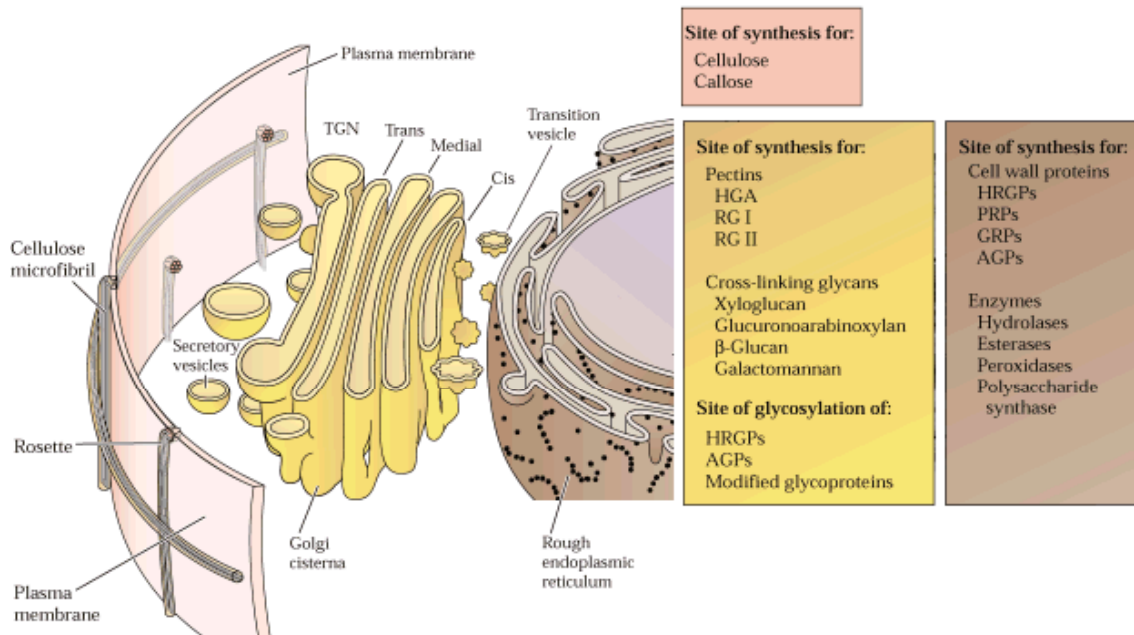
Land plants are among the largest organisms. This achievement is based largely on the growth, and mechanics of their cell wall, a structure that encases cells of plants, algae, fungi and bacteria like an armour. Plant cell wall is remarkable complex and dynamic entity and is one of the most sticking features that differentiate plants from animals and other eukaryotic cells. Cell wall of higher plants is of high economic importance to human especially as raw material of human and animal food, textiles, wood, paper, thickeners, biofuels and other products.

Cell wall was first observed by Robert Hooke in 1665 when he examined thin slices of cork under his microscope. In fact, he found small tiny boxes that he called cells and being dead material, consisted only of cell walls.

Cell wall of higher plants is a complex but organized molecular composite that may comprise many different polysaccharides, lignin, suberin, wax, proteins, aromatic substances, calcium, boron and water. Owing to the diversity of cell shapes and functions, the molecular composition and arrangement of cell wall exhibits a great diversity. Generally the cell wall consists of three parts: middle lamella, primary wall and secondary wall. Middle lamella is the first layer formed between the adjacent cells at the time of cytokinesis. Primary wall is simultaneously laid inner to the middle lamella, and is present in all type of cells and allows changes in cell size and shape. Secondary wall is impregnated on the inner surface of the primary cell wall in some type of cells (*e.g.* conducting cells) after achieving their final shape and size.

The cell wall of higher plants performs a variety of functions during growth and development as well as in plant defense, including maintenance of the osmotic pressure, water movement, rate and direction of cell growth, cell differentiation, intercellular communication and signalling, structural support and morphology of plant, cohesion among the cells of a tissue, storage, protection against pathogens and abiotic stresses. They also contribute to the functional specialization of cell types.

Here a detailed description of the cell wall components, their properties, structural organization and interactions among them, will be given. In addition, the involvement of different cell wall components especially the proteins in growth and development will be elaborated. Finally, we will discuss how cell wall proteins contribute to cell elongation by interacting with other cell wall components which makes the subject of this work.



**Figure 1.1: Biosynthesis of cell wall components** (after Buchanan *et al.*, 2000)  
 Cell wall biogenesis takes place in a series of coordinated steps inside or outside the cell. Synthesis of cellulose microfibrils and callose occurs at the plasma membrane surface. Synthesis and glycosylation of cell wall proteins and wall-modifying enzymes occur at the rough ER while synthesis of non-cellulosic polysaccharides at the Golgi apparatus. From here, the materials destined for cell wall are transported to cell surface in the form of secretory vesicles formed at the *trans*-Golgi network.

## 1.2. Cell wall components

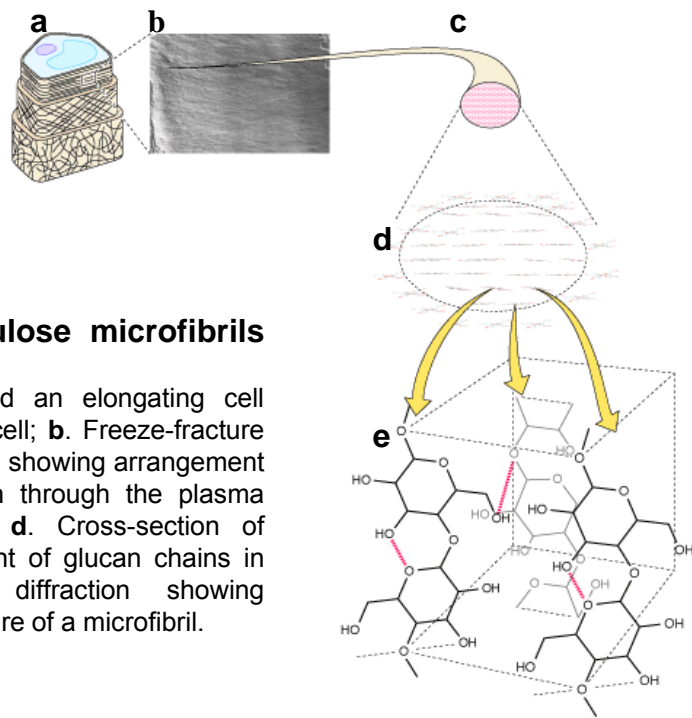
About 10% of the plants genome is devoted to cell wall biogenesis (<http://cellwall.genomics.purdue.edu/>). Cell wall biogenesis has been divided into six distinct stages: substrate generation, polymer synthesis, secretion, assembly, rearrangement during development, and disassembly. These steps are taking place both inside and outside the cell. [Figure 1.1](#) illustrates the cell compartment where synthesis and processing or modifications of major cell wall components occur.

The molecular composition and arrangement of wall polymers differ among species, among tissues of a single species, among individual cells, and even among regions of the wall around a single cell (Buchanan *et al.*, 2000). Following are the major cell wall components.

### 1.1.1. Polysaccharides

#### 1.1.2.1. Cellulose

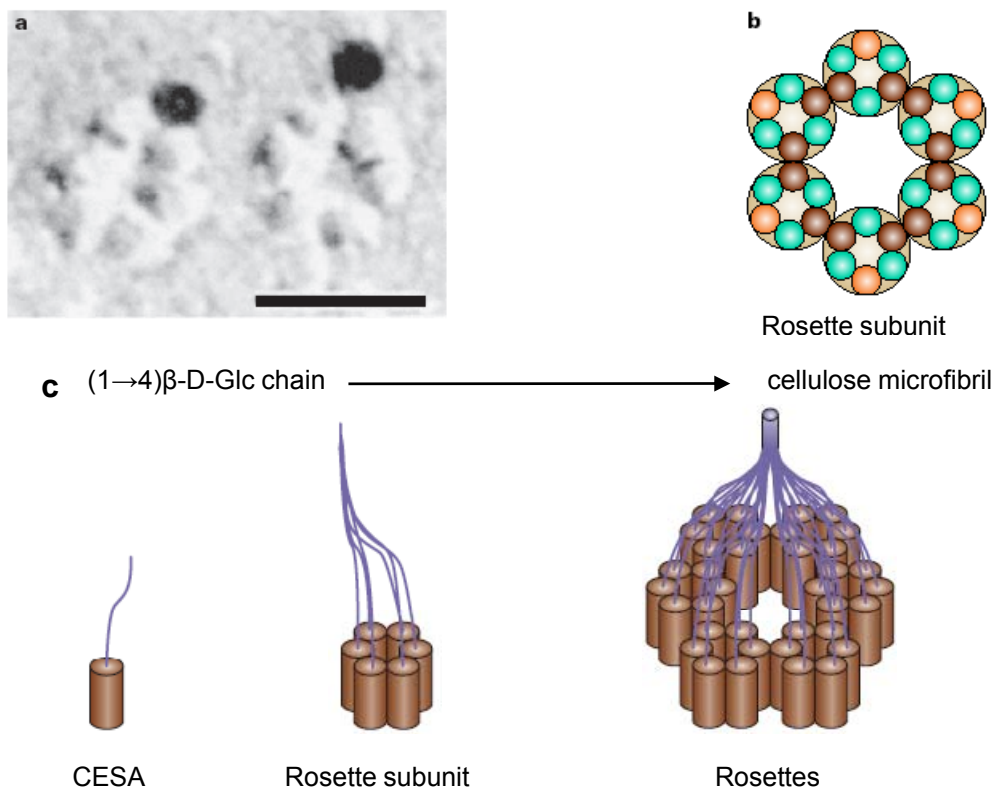
Cellulose is the most abundant plant polysaccharide that accounts for 15-30% of the dry mass of the primary cell wall and even more in the secondary walls. In plants, cellulose occurs in paracrystalline assemblies, *i.e.* microfibrils of about 36 parallel arranged (1→4)β-D-Glc chains hydrogen-bonded to one another along their length ([Figure 1.2](#)). Many microfibrils combine to form a cellulose fiber laid down on the cell surface in several layers distinguished by the different orientation of their fibers (Buchanan *et al.*, 2000). Cellulose synthesis (reviewed in Reiter (Reiter, 2002); and Cosgrove (Cosgrove, 2005) occurs at rosette-like structures, the cellulose-synthesizing complex (CSC) that consists of six hexagonally-arranged subunits (CESA) that are embedded in the plasma membrane ([Figure 1.3](#)). Different CESA genes are assumed to be required to make a functional CSC, and different sets of genes are involved in the formation of the primary and secondary walls. CESA1, CESA3 and CESA6 or CESA6-like (CESA2, CESA9) are required for biosynthesis of the primary wall, whereas CESA4, CESA7 and CESA8 are required to form secondary walls. The most convincing proof of this comes from the study on the CESA1 and CESA3 mutants that result in gametophyte lethality and CESA6 mutant causing cellulose deficiency and growth defects (Persson *et al.*, 2007). In other such studies, the catalytic domain of cotton Cesa was immunolocalized at and near the CSCs (Kimura *et al.*, 1999) and the mutant *atcesA1* (*rsw1*) was shown to cause disassembly of CSC, reduced cellulose accumulation and accumulation of non-crystalline (1→4)β-D-Glc (Arioli *et al.*, 1998). CSCs probably contain other proteins. KORRIGAN is assumed to have some role in polymerization or crystallization of microfibrils



**Figure 1.2: Structure of cellulose microfibrils**

(after Buchanan *et al.*, 2000)

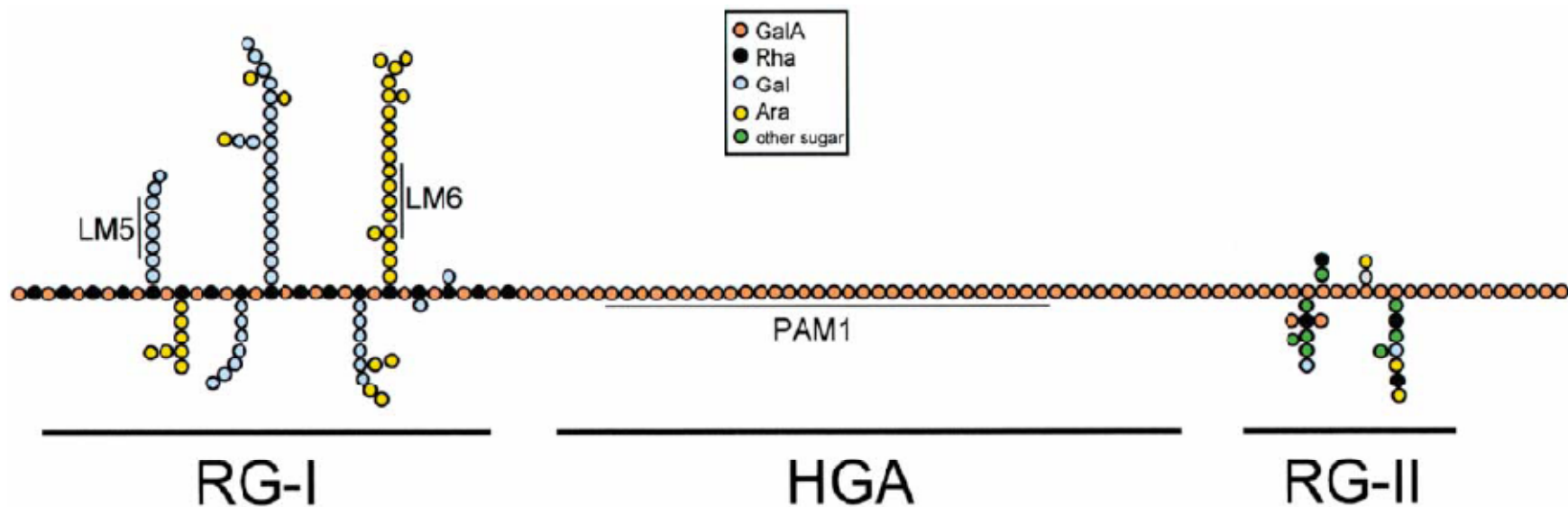
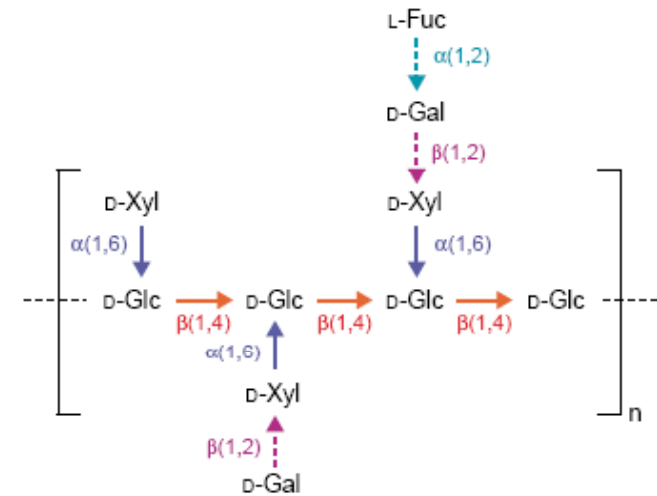
**a.** Helically-winded microfibrils around an elongating cell almost transverse to the long axis of cell; **b.** Freeze-fracture replica of an elongating maize root cells showing arrangement of cellulose microfibrils as impression through the plasma membrane; **c.** A single microfibril; **d.** Cross-section of microfibril showing spacial arrangement of glucan chains in the core of the microfibril; **e.** X-ray diffraction showing arrangement of atoms in the unit structure of a microfibril.



**Figure 1.3: The cellulose-synthesizing machinery of the cell** (after Cosgrove 2005)

**a.** Immunogold labelling of CESA, localized to hexameric cellulose synthesizing complex (CSC) in the plasma membrane. The black circles represent gold nanoparticles that are attached to antibody against CESA. Scale bar, 30 nm; **b.** A model showing how three different CESA proteins (shown in three different colours: orange, brown, green) might be organized into subunits and then into a hexameric CSC; **c.** A model of how CESA complexes synthesize a cellulose microfibril. Each CESA protein can synthesize a single (1→4)β-D-Glc chain. In this model, microfibril is shown to consist of 36 (1→4)β-D-Glc chains, synthesized by a CSC, which is composed of a hexamer of CESA hexamers.

**Figure 1.4: Structure of xyloglucan** (after Reiter 2002)  
XLFG type xyloglucan of Arabidopsis where the solid arrows indicate linkages that are always present, where as dashed arrows denote substitution patterns.



**Figure 1.5: Structure of pectins** (after Willats *et al.*, 2001)

Simplified schematic diagram to indicate some of the features of the three major types of pectins: HGA, RG-I and RG-II. Oligosaccharide epitopes recognized by anti-HGA monoclonal antibody PAM1 and anti-RG-I monoclonal antibodies LM5 and LM6 are also indicated.

or in recycling of sitosterol (Nicol *et al.*, 1998). COBRA is thought to link the complexes to nearby microtubules for guidance along the membrane (Roudier *et al.*, 2005). KOBITO1 is assumed to take part in the cellulose synthesis machinery or to play a role in the coordination between cell elongation and cellulose synthesis (Pagant *et al.*, 2002).

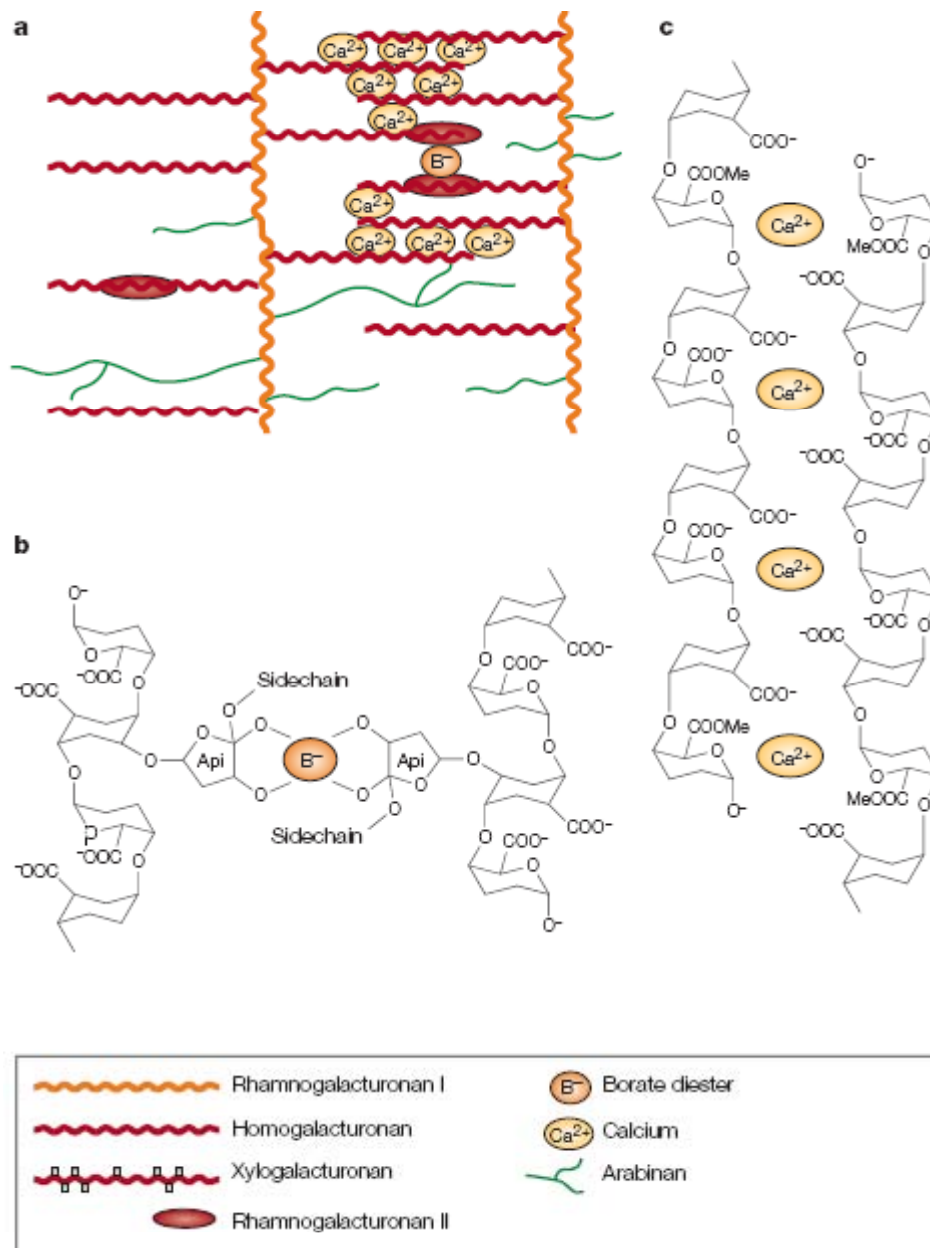
Cellulose is present in cell wall of all types, provides strength and rigidity to cell wall, and prevents the swelling of the cell wall and rupture of the plasma membrane that might occur when osmotic conditions favour water entry into the cell.

#### 1.1.1.2. Cross-linking glycans

The cell wall polysaccharides that can hydrogen-bond to cellulose microfibrils are classified as cross-linking glycans. Most of them are also called hemicelluloses. In type I cell wall, characteristic of dicots, the major cross-linking glycans of primary cell wall are xyloglucans (XGs). XGs typically consist of a (1→4)β-D-Glc backbone carrying (1→6)α-D-Xyl moieties on three consecutive Glc residues. As shown in Figure 1.4, the Xyl residues attached to the second and third Glc residues of the backbone can carry D-Gal in (1→2)β-linkage, and the second of these Gal residues is usually substituted by L-Fuc in (1→2)α-linkage (Reiter, 2002) to make XLFG xyloglucan according to the nomenclature of Fry *et al.* (Fry *et al.*, 1993). The structure and distribution of the side branch (chains) vary in different tissues and species (forming XXFG, XXXG, XXLG, XLLG or XLXG) which seem important for bonding to cellulose. Glucuronoabinoxylans (GAXs) are the major cross-linking polymers of type II cell wall, characteristic of commelinoide monocots and some dicots. The structure of GAXs varies considerably with respect to degree of substitution and position of attachment of α-L-Ara residues (Carpita and Gibeaut, 1993). The other types of cross-linking glycans are “mixed linkage” (1→3, 1→4)β-D-glucan, glucomannans, galactoglucomannans and galactomannans (Carpita and Gibeaut, 1993).

#### 1.1.1.3. Pectic matrix

Pectins (reviewed in Ridley *et al.* (Ridley *et al.*, 2001) and Willats *et al.* (Willats *et al.*, 2001)) are a mixture of heterogenous, branched and highly hydrated polysaccharides (Figure 1.5). They are rich in polygalacturonic acid (PGA) and account for 30% total cell wall mass (Carpita and Gibeaut, 1993). GalA occurs in two major structural features (homo and heteropolymers) that form the backbone of three polysaccharide domains found in almost all pectins: homogalacturonan (HGA), rhamnogalacturonan-I (RG-I), and rhamnogalacturonan-II



**Figure 1.6: Formation of pectin networks by covalent and ionic bonds**  
(after Cosgrove 2005)

**a.** A model of how the different pectins may be covalently linked together to form a macromolecular pectin network. In this model, RG-I serves as the backbone and the other pectin domains are attached as branches; **b.** RG-II chains are cross-linked to form dimers through a borate ester bond; **c** HGA forms stiff gels through  $\text{Ca}^{2+}$ -mediated crosslinking of its carboxyl groups through ionic and coordinate bonds.



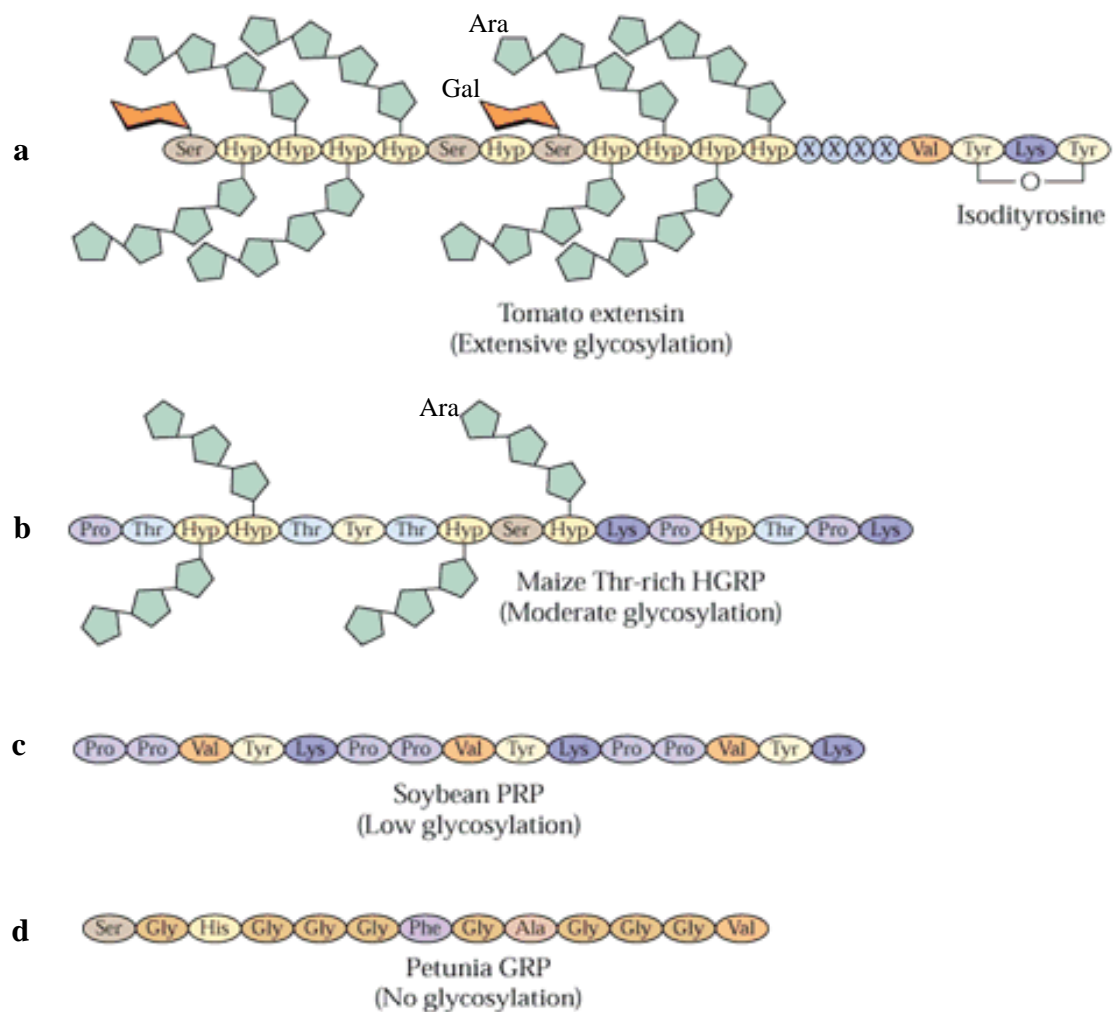
(RG-II). These three types of GalA covalently bond to each other to form a pectic network throughout the primary cell wall matrix and middle lamellae. HGA is a linear, unbranched (unsubstituted) homopolymer containing about 100–200 GalA residues, and is called smooth region of pectins. HGA is an abundant and widespread kind of pectin which is synthesized in the Golgi apparatus and deposited in the cell wall. It has 70–80% GalA residues methylesterified at the C-6 carboxyl. The demethylesterified HGA in the cell wall matrix can be cross-linked by  $\text{Ca}^{2+}$  to form supramolecular assemblies and gels (Figure 1.6). Other modifications and substitutions of HGA are not as widespread as methyl esterification. RG-I (known as the hairy region of pectin) is an acidic pectic domain consisting of as many as 100 repeats of the disaccharide L-Rha-D-GalA. RG-I is abundant and heterogenous and generally glycosidically-attached to HGA domains. Arabinogalactan and arabinan side chains commonly include D-Gal and L-Ara residues. Type I arabinogalactans are (1→4)β-linked D-Gal with non-reducing terminal-Ara (*t*-Ara) substituted at the O-3 of some of the Gal units (Carpita and Gibeaut, 1993). Arabinogalactans of type II with (1→3)β- and (1→6)β-linked-D-Gal residues also occur on pectic backbones. Arabinans can become branched by links through O-2 and O-3. Another major, widespread highly conserved pectic domain is RG-II which is similar to HGA by possessing homopolymer backbone, but is branched like RG-I. It consists of around 9 GalA residues backbone substituted by 4 heteropolymeric side chains each consisting of eleven different sugars including Api, aceric acid and 2-keto-3-deoxy-D-manno-octulosonic acid (kdo). RG-II can dimerize by means of borate ester links through Api residues.

In plants, several enzymes like pectin methylesterases (PMEs) (Micheli, 2001) polygalacturonase (PGs) (Tanaka *et al.*, 2002) and pectate lyases (PLs) (Marin-Rodriguez *et al.*, 2002) modify or degrade pectins in relation with changes in cellular adhesion and plasticity of the cell wall, a prerequisite to elongation and development.

### 1.1.2. Cell wall proteins

Until the last decade cell wall was considered to be made mainly of polysaccharides and structural proteins. But recent studies, especially proteomic studies have reported the presence of many other proteins including enzymes that perform a wide range of functions. The present study has further contributed to this area. Here cell wall proteins (CWPs) will be treated as have been considered before this study: structural proteins, that can form networks and other proteins that contribute to a wide range of functions from growth and development





**Figure 1.7: Cell wall structural proteins, motifs and glycosylation**

(after Buchanan *et al.*, 2000).

**a.** Tomato extensin possessing Ser (Hyp)<sub>4</sub> or related motif with high glycosylation with tetra-Ara as well as Gal at Ser. Isodityrosine linkage is also shown on the Tyr-Lys-Tyr motif, a likely position for this linkage; **b.** Maize extensin like Thr-rich moderately-glycosylated protein; **c.** A Soybean PRP that lacks contiguously hydroxylated Ser, Thr, and Hyp residues (a likely position for glycosylation with Ara) are not heavily glycosylated; **d.** *Petunia* GRP with no glycosylation site.

to response to biotic and abiotic stresses. All CWPs possess a signal peptide and are targeted to secretory pathway (Buchanan *et al.*, 2000).

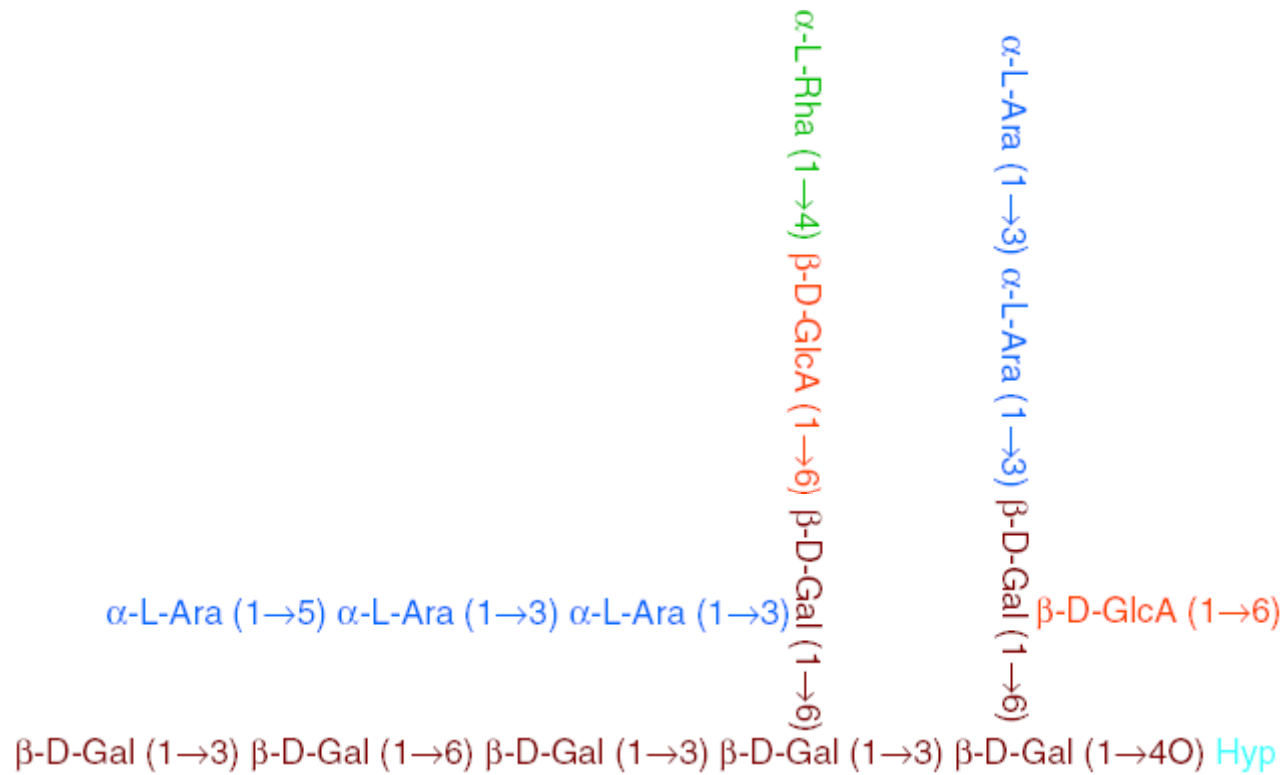
### 1.1.2.1. Structural proteins

Although all types of cell wall are mainly composed of polysaccharides, cellulose-crosslinking glycans network embedded in pectic matrix, another network is formed by cell wall structural proteins. Cell wall structural proteins can be grouped into three principal classes on the basis of their sequence rich in some particular type of amino acids: the extensins, hydroxyproline/proline-rich proteins (H/PRPs) and glycine rich proteins (GRPs). Examples of such proteins are given in [Figure 1.7](#). All these proteins are encoded by large multigene families and are developmentally-regulated. Extensins and H/PRPs are assumed to be cross-linked to make the wall less extendable when the cells achieve their final size. Their relative amount varies among tissues and species.

Extensins are well studied-structural proteins, which are rod-shaped distinguished by their amino acid sequence rich in Pro, Tyr, Lys, Ser, His, Val where Pro may reach up to 50%. They possess the repeating motif Ser-(Pro)<sub>n</sub> ( $n \geq 3$ ) and Tyr-Lys-Tyr sequences that are important for their secondary and tertiary structure making inter- and intra-molecular bonding. They are highly *O*-glycosylated and have basic pIs (Cassab, 1998; Sommer-Knudsen *et al.*, 1998; Buchanan *et al.*, 2000). Normally extensins are *O*-glycosylated with chains of Ara on contiguous Hyp residues and Gal on Ser (Kieliszewski and Lamport, 1994; Kieliszewski, 2001).

H/PRPs represent another large multigene family of structural proteins. Because of their similarity to extensins they are also thought to be rod-shaped (Buchanan *et al.*, 2000). H/PRPs contain the repetitive pentamere motif (Pro-Hyp-Val-Tyr-Lys)<sub>n</sub> or its variants but lack Ser and are *O*-glycosylated (Cassab, 1998).

GRPs (reviewed in Ringli *et al.* (Ringli *et al.*, 2001)) that belong to another major cell wall structural protein family, should be distinguished from intracellular GRPs that are assumed to bind RNAs. Cell wall GRPs have repetitive sequences which may contain more than 60% Gly arranged in short repeating units. The GRP protein sequences often follow the motif (Gly-X)<sub>n</sub> in the Gly-rich region, where X is often Gly, but can also be another amino acid. Ala, Ser, Val, His, Phe, Tyr and Glu are common at the X position. In some cases, the



**Figure 1.8: Structural complexity of an arabinogalactan** (Seifert and Robert 2007)

The first complete structure of an arabinogalactan glycan, derived from synthetic green fluorescent protein (GFP):(Ala-Pro)<sub>51</sub>, expressed in tobacco BY2 cells.

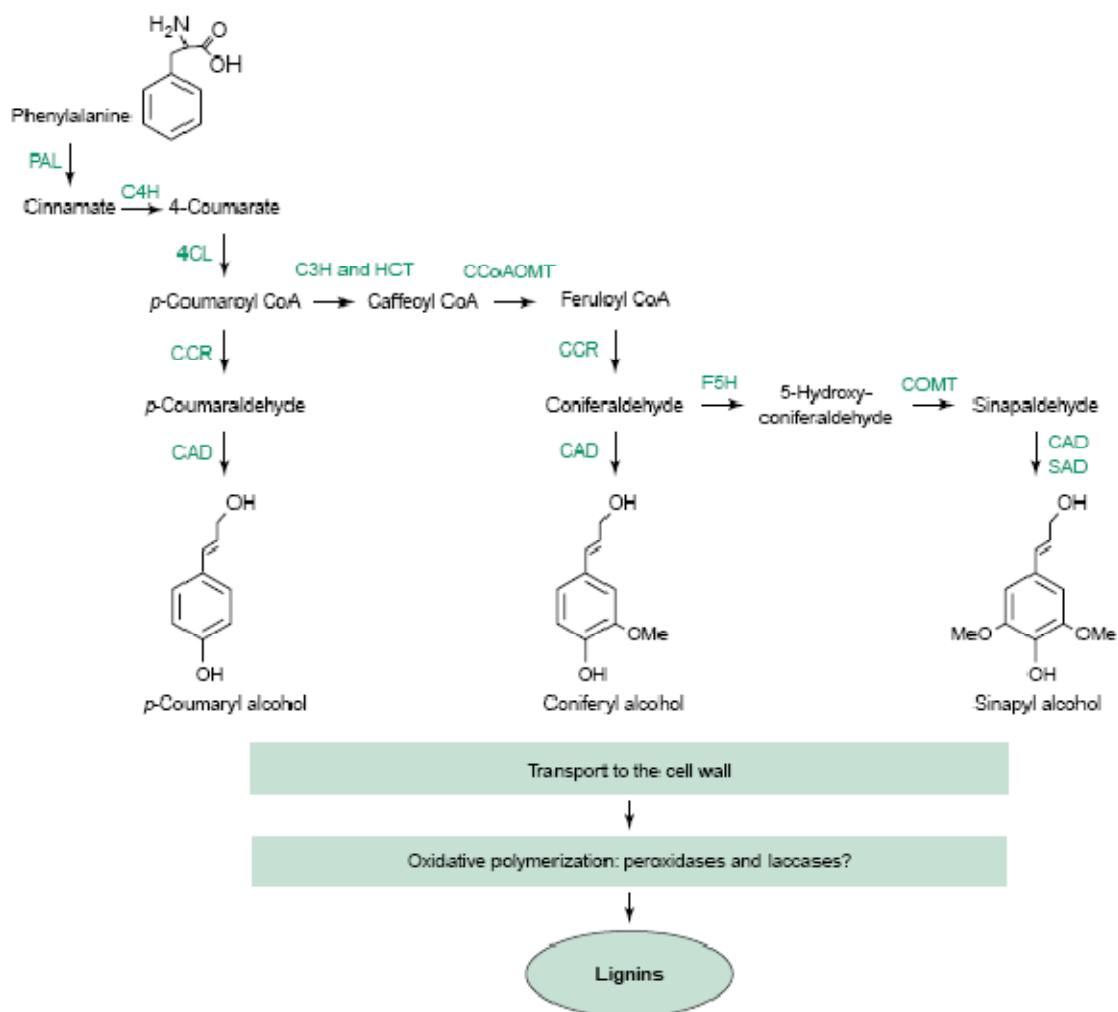
motif varies (Ringli *et al.*, 2001). Beside the (Gly-X)<sub>n</sub> motif, higher-order repetitive sequences that are rarely perfect are sometimes found and were proposed to be important for the formation of the secondary structure of the proteins. GRPs are believed to form plate-like structure rather than rod-shape conformation, on the plasma membrane-cell wall interface. They are thought to crosslink to cell wall polysaccharides are therefore difficult to extract.

### 1.1.2.2. Arabinogalactan proteins (AGPs)

AGPs (reviewed in Seifert and Roberts (Seifert and Roberts, 2007)) consist of a core protein of highly varying length and domain complexity, and one or more arabinogalactan (AG) side chains (Figure 1.8). They often contain a glycosylphosphatidylinositol (GPI) lipid anchor (Borner *et al.*, 2005). The relative ratio of glycan to protein is sometimes higher than 9, but can vary strongly for the same AGP core protein isolated from the same tissue. AGPs are precipitated by treatment with the  $\beta$ -Yariv reagent. In some AGPs, the peptide backbones are 10-13 residue-long and are called AG peptides. Most AGP sequences consist of a single central domain rich in Pro, Ala, Ser, and Thr. Most AGPs are *O*-glycosylated at one or more Hyp residues by AG type II (found in dicots) groups. These consist of (1 $\rightarrow$ 3) and (1 $\rightarrow$ 6) $\beta$ -linked Gal chains connected to each other by (1 $\rightarrow$ 3, 1 $\rightarrow$ 6)-linked branch points, O-3 and O-6 positions substituted with terminal Ara residues. Type I AGs (characteristic of monocots and some dicots) are (1 $\rightarrow$ 4) $\beta$ -linked D-Gal with non-reducing terminal-Ara (*t*-Ara) substituted at the O-3 of some of the Gal units (Carpita and Gibeaut, 1993). AGPs are involved in a variety of functions, *e.g.* embryonic and post-embryonic patterns, pollen tube guidance, growth, secondary wall deposition, abscission and interaction with growth regulators and microbes.

### 1.1.2.3. Other proteins

Recently it has become clear that cell wall is very dynamic and contains many different enzymes and other agents (reviewed in Cosgrove (Cosgrove, 1999) and Fry (Fry, 2004)) conferring it dynamic properties like plasticity and extensibility. Among these proteins were expansins (*e.g.*  $\alpha$ - and  $\beta$ -expansins), hydrolases (*e.g.* cellulases, xylanases, PMEs, PGs), transferases (*e.g.* xyloglucan endotransglycosylase hydrolases (XTHs), lyases (*e.g.* pectate lyases) oxido-reductases (*e.g.* peroxidases and laccases). These non-structural proteins perform a wide range of biological roles. A few examples of these functions are cited below while others, concerning especially the modification of cell wall and regulation of cell elongation are given in section 1.2.



**Figure 1.9: Lignin biosynthesis in dicots** (after Boudet *et al.*, 2003)

The figure shows a simplified view of the most favoured lignin biosynthetic pathway in angiosperms where the implication of laccases in the oxidation of monolignols is still not known.

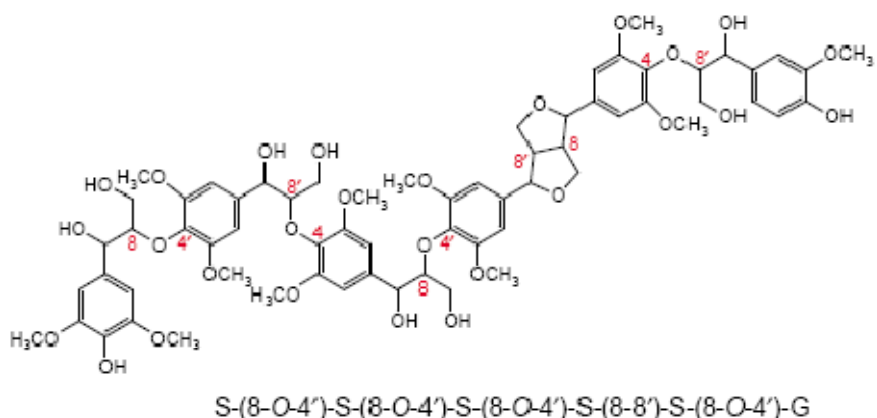
**Abbreviations:** CAD, cinnamyl alcohol dehydrogenase; CCoAOMT, caffeoyl-CoA O-methyltransferase; C3H, p-coumarate 3-hydroxylase; C4H, cinnamate 4-hydroxylase; CCR, hydroxycinnamoyl-CoA reductase; 4CL, 4-coumarate CoA ligase; COMT, caffeic acid/5-hydroxyferulic acid O-methyltransferase also known as AldOMT, 5-hydroxyconiferaldehyde O-methyltransferase; F5H, ferulate 5-hydroxylase also known as Cald5H, coniferaldehyde 5-hydroxylase; HCT, p-hydroxycinnamoyl-CoA:quinic acid p-hydroxycinnamoyl-CoA transferase; PAL, phenylalanine ammonia-lyase; SAD, sinapyl alcohol dehydrogenase.

Inside cells, proteases are involved in all aspects of the plant life cycle (Schaller, 2004), but their role in cell wall is not very clear. Outside cells, some are thought to be involved in cell-cell communication by generating local-signals (Matsubayashi and Sakagami, 2006). The subtilisin-like serine protease *Stomatal Density and Distribution 1 (SDD1)* is thought to control stomata distribution and density in *Too Many Mouths (TMM)* dependent way (Berger and Altmann, 2000; Nadeau and Sack, 2003). The cell wall HRGP, *Root-Shoot-Hypocotyl-Defective (RSH)*, is essential for normal embryo development in *A. thaliana* and the mutant fails to develop normal embryo because mal-positioning of the cell plate at the time of division leading to the formation of abnormal embryo (Hall and Cannon, 2002). Reviewing the role of the cell wall in embryogenesis (Malinowski and Filipecki, 2002), the authors have highlighted that chitinases, XTHs, and peroxidases participate in embryogenesis regulation by involvement in signal transduction, and by influencing cell shape and division plane. AGPs are considered source of signals in a variety of ways (Seifert and Roberts, 2007). They may bind directly to a receptor that activate a signal transduction cascade or may release lipid signals by cleavage of GPI anchor by phospholipase or oligosaccharide as signals by endoglucanases. *Inflorescence Deficient in Abscission (IDA)* gene encoding an AGP playing a role in the abscission where its mutant delayed floral organ abscission and its over-expression produced opposite effect. Pectate lyases play important role in softening ripening fruits (Marin-Rodriguez *et al.*, 2002). Peroxidases that rigidify cell wall by cross-linking of wall compounds are expressed to cope with biotic and abiotic stresses like wound, pathogen interaction, and climatic aggression (Passardi *et al.*, 2004; Passardi *et al.*, 2005). Polygalacturonase-inhibiting proteins (PGIPs) protect plant against fungal attack by inhibiting fungal PGs (De Lorenzo *et al.*, 2001).

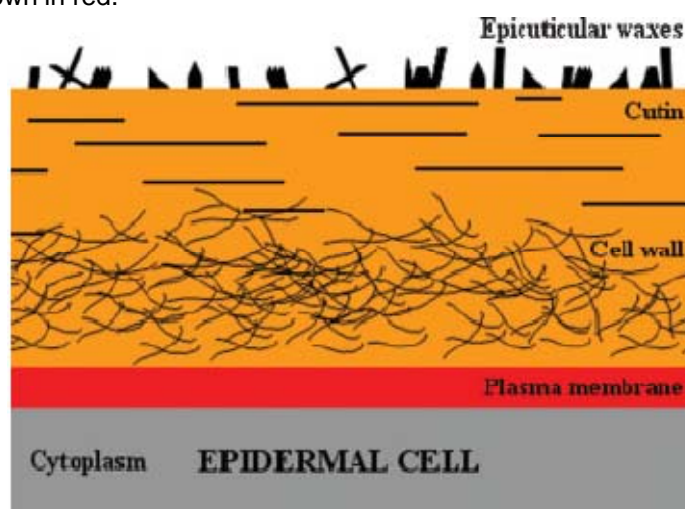
### **1.1.3. Other cell wall components**

#### **1.1.3.1. Lignins**

Lignins (reviewed in Boerjan *et al.* (Boerjan *et al.*, 2003), Boudet *et al.* (Boudet *et al.*, 2003) and Davin and Lewis (Davin and Lewis, 2005)) are the 2<sup>nd</sup> most abundant plant substances in vascular plants. They are usually deposited in the secondary cell wall with few lignins exceptionally occurring in primary cell wall (Buchanan *et al.*, 2000). Lignins are assumed to be obtained by the oxidative polymerization of monolignols in the cell wall by peroxidases and laccases. The first step of lignin biosynthesis is the deamination of phenylalanine (dicots) or tyrosine (monocots) by ammonia lyases, yielding cinnamic or p-coumaric acids, respectively (Figure 1.9). Successive steps of hydroxylation, methylation,



**Figure 1.10: Lignin structure** (after Evtuguin and Amado 2003, in Davin and Lewis 2005)  
An example of putative primary sequence structure of a syringyl lignin-derived hexamer fragment from *Eucalyptus globulus*. The radical-radical coupling linkages between the sinapyl alcohol monomers are shown in red.



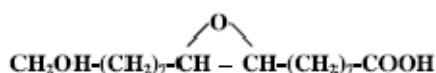
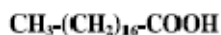
**Figure 1.11: Plant cuticle** (after Heredia-Guerrero *et al.*, 2008)  
Current microscopic model of plant cuticle that shows the epicuticular wax crystals deposited on an amorphous and dense matrix of cutin polymer, containing some intracuticular waxes.

**FAMILY C<sub>16</sub>:**



$$x, y = 6, 7; x+y = 13$$

**FAMILY C<sub>18</sub>:**



$$x, y = 7, 8; x+y = 15$$

**Figure 1.12: Major monomers of plant cutin** (after Heredia-Guerrero *et al.*, 2008)  
Chemical structure of the major monomers present in plant cutin. These monomers are derived from C<sub>16</sub> saturated fatty acid and from C<sub>18</sub> unsaturated fatty acids. Some plants have mainly C<sub>16</sub> family of monomers, whereas others have a mixture of both C<sub>16</sub> and C<sub>18</sub> families of monomers.

formation of hydroxycinnamoyl-CoA thioesters, reduction of hydroxycinnamoyl-CoA thioesters to hydroxycinnamaldehydes, and reduction of hydroxycinnamaldehydes lead to three p-hydroxycinnamyl alcohols (monolignols), p-coumaryl, coniferyl, and sinapyl alcohols which are transported from the cytosol to the apoplast. When reaching the apoplast, monolignols undergo dehydrogenative polymerization via oxidases, forming lignins. They comprise two major components namely guaiacyl (G), derived from coniferyl alcohol, and syringyl (S), derived from sinapyl alcohol (Figure 1.10) and one minor component p-hydroxyphenyl units (H), derived from p-coumaryl alcohol. Although polymers of only three types of monomers, the composition and structure of lignins vary significantly among different plants or within the same plant (Billa *et al.*, 1998).

### 1.1.3.2. Lipids

Lipids constitute the impermeable hydrophobic outer portion of the cell wall called cuticle, a continuous layer that covers aerial parts of leaves, fruits and young non-woody stems. Cuticle weight ranges from 2000  $\mu\text{g}/\text{cm}^2$  (in fruits) to 450-800  $\mu\text{g}/\text{cm}^2$  (in leaves). Forty to 80% of this is constituted by cutin, the extracellular lipids that are interconnected by ester bonds and can be studied after hydrolysing the polyesters (Heredia, 2003). Waxes are soluble extracellular lipids and can be extracted with organic solvents. The structure and composition of the cuticle vary in different plants, plant organs and growth stages but basically comprise a cutin matrix developed on the epidermal plant cell wall with waxes. Waxes are solid, partially crystalline aggregates at room temperature (Schreiber, 2005). They appear either embedded in the matrix (intracuticular) or deposited on its surface (epicuticular) as shown in Figure 1.11 (Riederer and Müller, 2005). Epicuticular wax may exist as a smooth film in some species or as wax crystals in other species (Buschhaus *et al.*, 2007). Chemically, cutin is a polymeric network of polyhydroxylated  $\text{C}_{16}$  and  $\text{C}_{18}$  fatty acids cross-linked by ester bonds (Figure 1.12). Waxes are generally described as mixtures of homologous long-chain aliphatic compounds, like alkanes, alcohols, aldehydes, fatty acids and esters with the addition of varying proportions of cyclic compounds like triterpenoids and hydroxycinnamic acid derivatives (Heredia, 2003).

Cuticle controls non-stomatal water loss, protects plants against ultraviolet radiation. It does not allow the water to stand easily on surface of the plant thus minimizing accumulation of dust, pollen and air pollutants. In addition, surface wax is thought to play important roles in plant defense against bacterial and fungal pathogens (Kunst and Samuels,



2003). Surface wax has also been shown to participate in a variety of plant-insect interactions (Eigenbrode and Espelie, 1995). For example epicuticular lipid extracts and individual lipid component enhance or deter oviposition, movement, and feeding.

#### 1.1.3.3. Suberin

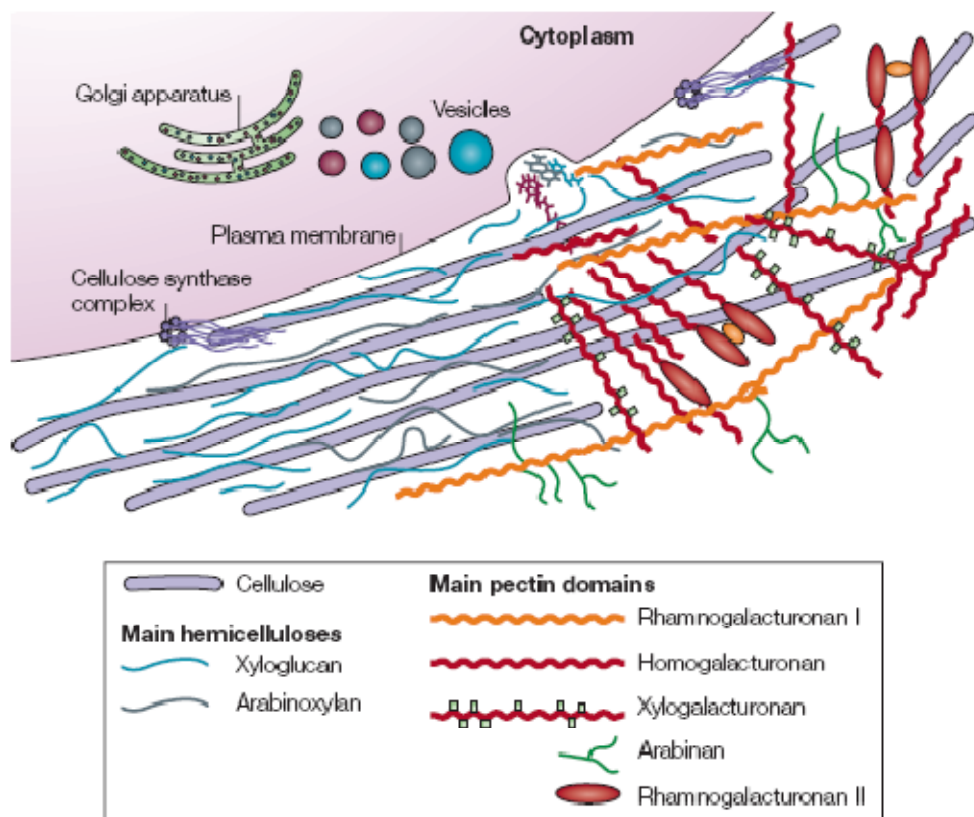
Suberin is a biopolymer making a barrier between plants and the environment in specialized plant tissues (*e.g.* periderm, bark and tuber skin) that protects the internal living tissues from dehydration, injuries, and pathogens (Soler *et al.*, 2007). Suberized cells are also found in the epidermis and hypodermis of roots, the endodermis and the bundle sheath of grasses (Graca and Santos, 2007). Suberin is a complex polyester made of glycerol and long-chain diacids and hydroxyacids (Schreiber *et al.*, 1999; Groh *et al.*, 2002). In suberized cells, suberin represents up to 50% of the mass of the cell wall (Pereira, 1988).

In the outermost tissues of plants, suberized cells play a vital role affording protection against environmental aggressions and pathogens, and controlling temperature and water loss (Schreiber *et al.*, 1999; Groh *et al.*, 2002).

#### 1.1.3.4. Inorganic salts

Calcium and boron are the two major minerals localized in the cell wall. Calcium is also an essential plant nutrient and is required for various structural roles in the cell wall and membranes. Plants take up calcium from soil through roots and calcium is delivered to the shoots via xylem. Calcium enters the plant cells through specific ion channels in their plasma membranes (White and Broadley, 2003). In cell wall,  $\text{Ca}^{2+}$  participates in cross-linking the demethylesterified HGA to form supramolecular assemblies and gels (see Figure 1.6) (Willats *et al.*, 2001) that modify cell wall physical and chemical properties.

Boron is an essential nutrient for vascular plants. In cell wall, it cross-links RG-II to form dimers by borate ester bond between two apioses (see Figure 1.6) (Ridley *et al.*, 2001). Boron deficiency makes the tissues brittle or fragile, while plants grown on high boron levels may have unusually flexible or resilient tissues (Loomis and Durst, 1992). Boron helps establishing an effective legume–*Rhizobium* symbiosis (Bolanos *et al.*, 1994) and is required for the maintenance of nodule cell wall structure (Bonilla *et al.*, 1997).



**Figure 1.13 Structure of the primary cell wall** (after Cosgrove 2005)

Cellulose microfibrils (purple rods) are synthesized by large hexameric complexes in the plasma membrane, whereas hemicelluloses and pectins, which compose the matrix polysaccharides, are synthesized in the Golgi apparatus and are deposited to the wall surface by vesicles. Xyloglucans are tightly bound to microfibrils by hydrogen bonds and form cross-links between them, thus constituting a load-bearing network. For clarity, the hemicellulose–cellulose network is shown on the left part of the cell wall without pectins, which are emphasized on the right part of the figure. In most plant species the main hemicellulose is xyloglucan (blue), while hemicelluloses such as arabinoxylans (grey) and mannans (not shown) are found in lesser amounts. The main pectin polysaccharides include rhamnogalacturonan I and homogalacturonan, with smaller amounts of xylogalacturonan, arabinan, arabinogalactan I (not shown) and rhamnogalacturonan II. Pectin domains are believed to be covalently linked together and to bind to xyloglucan by covalent and non-covalent bonds. Neutral pectin polysaccharides (green) are also able to bind to cellulose surfaces. pH of the cell wall and the action of different enzymes (not shown) like expansins, endo-1,4- $\beta$ -D-glucanases, XTHs, PME and peroxidases on the cell wall components modify the properties of cell wall resulting in wall loosening or rigidification.

---

## 1.2. Cell wall proteins in relation to cell elongation and growth

Growth can be described as an irreversible increase of volume. In plants, growth is the outcome of cell division, enlargement of the new cells and their differentiation into different types of tissues. These processes of growth are accompanied by permanent change in size (usually an increase in length or volume) and an increase in dry mass of the growing part.

Enlargement of cells may be polarized, expanding more in one axis than the other (elongation) or it may be uniform in all directions as in isodiametric cells. Furthermore polarized growth takes place in two ways. In the first case, called tip growth, in which the growth is focused on a single specialized region, the apex of the tip growing cell. This type of growth is found in pollen tube, root hair and fungal hyphae. The other type is called diffuse growth and takes place at any point of a meristematic cell mostly in the cell wall parallel to the axis of elongation. This latter type of elongation generally occurs in the apical meristem of root and shoot or the intercalary meristem. This study is focused exclusively on diffuse polarized cell growth.

Before maturity, plant cells usually enlarge 10- to 1000-fold in volume by the process of vacuolation and irreversible cell wall expansion (Cosgrove, 1997). According to this review, there are two major concepts that account for the molecular basis of expansion. The first concept couples wall expansion to biosynthesis and secretion of wall polymers but did not explain the wall stress relaxation which is essential expansion of the cell *via* water uptake. The second concept considers wall expansion as result of biochemical loosening of the wall which permits turgor-driven extension of the wall polymer network. Though convincing, this latter neglects the need for integration of new polymers into the expanding wall.

To explain cell wall expansion based on rheological properties of the growing wall several potential mechanisms of stress relaxation and expansion can be imagined. Figure 1.13 represents the organization of various cell wall components and the various types of coordinated interactions between them that allow the cell to elongate but maintain cell wall integrity as intact structure. These interactions are described below in slight detail.

The acid-growth theory of cell wall expansion (Hager, 2003), which states that auxin, in an activated form ( $\sim A$ ) activates a  $H^+$ -pumping ATPase at the plasma membrane (PM) which utilizes respiratory energy (ATP) to raise the proton concentration in the

---

apoplast. This triggers cell elongation by activating cell wall enzymes. The transport of protons into the cell wall is compensated by a flow of cations into the cytoplasm, which maintains turgor pressure during cell elongation.

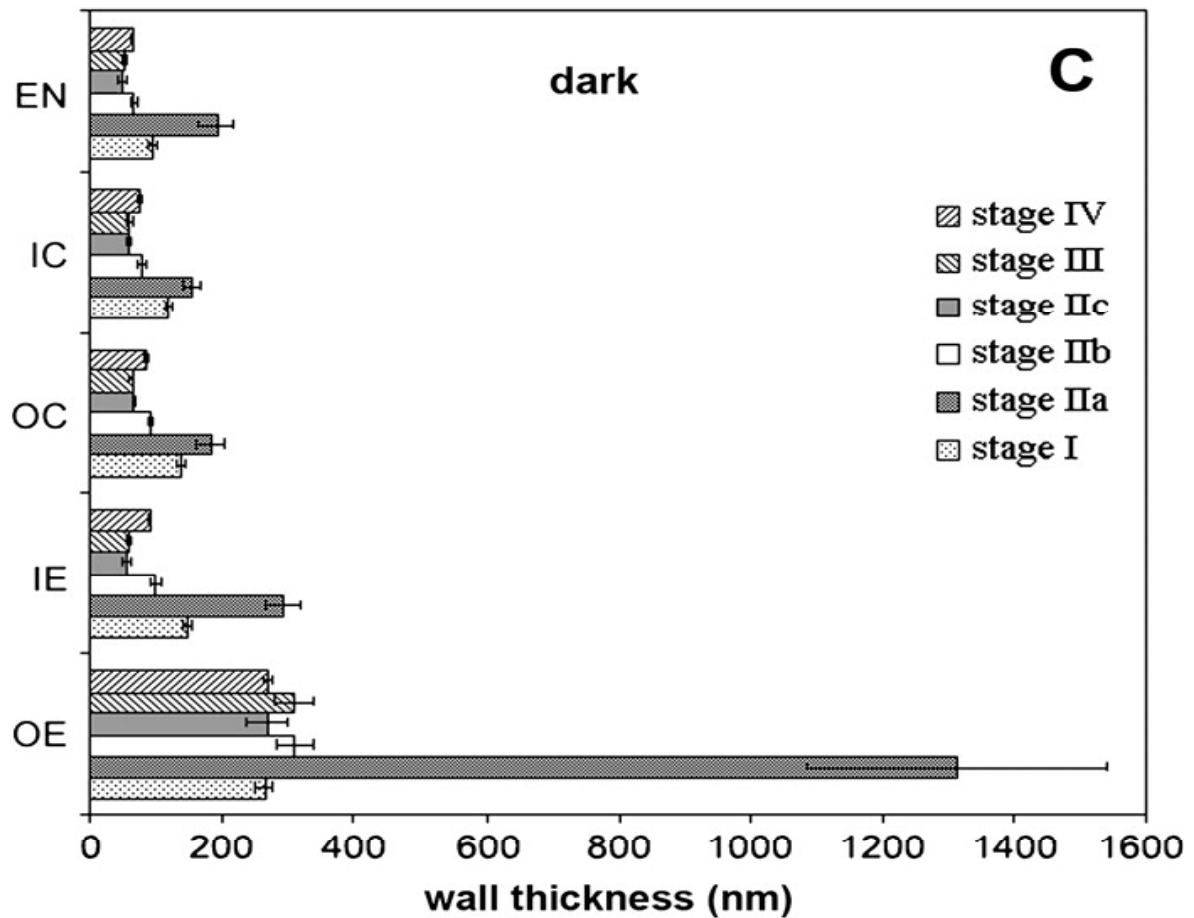
In elongating cells, the cellulose microfibrils are deposited perpendicular to the axis of elongation, forming a spring-like structure (Green, 1962). Such an arrangement reinforces lateral walls on one side and allows directional expansion (elongation) of the cell on the other hand through loosening of cellulose-hemicellulose networks. Expansins that can restore acid-induced creep of denatured walls, also participate this process. In fact, expansins are thought to transiently displace short stretches of hemicelluloses that are bonded to the surface of cellulose microfibril. It makes polymers to creep dragging along other structural components if the wall is in tension (Cosgrove, 1998).

Endo-1,4- $\beta$ -D-glucanases (EGases) are assumed to cleave the  $\beta$ -1,4-glycosidic bonds between glucose of xyloglucans relaxing wall polysaccharides to move apart during cell elongation.

XTHs also perform cell wall modifications required for the process of wall assembly and cell expansion, by either wall loosening or incorporating new xyloglucan chains into extending walls (Chanliaud *et al.*, 2004). They cleave xyloglucans and rejoin the newly generated reducing ends to others by acting on the xyloglucans attached to cellulose microfibrils (Vissenberg *et al.*, 2005).

PMEs can cause wall loosening or wall stiffening according to their mode of action on HGs (Micheli, 2001). Once integrated into the cell wall, they may act randomly (as in fungi) or linearly (as in plants) on HGs. When PMEs act randomly on HGs (at acidic pH), the demethylesterification releases protons that promote the action of endopolygalacturonases (Endo-PGs) and contribute to cell wall loosening. When PMEs act linearly on HGs (at basic pH), PMEs give rise to blocks of free carboxyl groups that could interact with  $\text{Ca}^{2+}$ , so creating a pectate gel. Because the action of Endo-PGs in such a gel is limited, this action pattern of PMEs contributes to cell wall stiffening.

Peroxidases (reviewed in Passardi *et al.* (Passardi *et al.*, 2004)) also play a dual role in this context. By peroxidative cycle, they oxidize various substrates such as tyrosine



**Figure 1.14. Changes in thickness of cell walls of dark-grown hypocotyls**

(after Derbyshire *et al.*, 2007)

Cell wall thicknesses of different tissues [outer epidermis (OE), inner epidermis (IE), outer cortex (OC), inner cortex (IC), and endodermis (EN)] of hypocotyls freeze-fractured at their mid-points at different developmental stages of dark-grown hypocotyls:

**stage I:** at the embryo stage prior to germination

**stage II:** at the onset of germination

**stage III:** at 50% of their final length

**stage IV:** immediately after cessation of growth

residues, monolignols, suberin units and ferulic acids. The radicals produced by the peroxidative cycle, if linked to other polymers can cross-link cell wall polymers like extensins or lignins that block any further wall loosening and hence cell expansion. Inversely, by hydroxylic cycle they produce reactive oxygen species (ROS) like  $\cdot\text{OH}$  that can cleave various polysaccharides non-enzymatically and contribute to wall loosening. Wall loosening is also caused by regulating  $\text{H}_2\text{O}_2$  concentration by both peroxidative and hydroxylic cycles.

Listed above are the proteins potentially involved directly or indirectly in cell wall elongation or its arrest, but still a huge investment is needed in this field to completely understand the process.

### 1.3. *A. thaliana* etiolated hypocotyls as a model for cell elongation

To understand the mechanism of cell elongation, hypocotyls of *A. thaliana* are widely used as model (Gendreau *et al.*, 1997). During hypocotyl elongation, almost the entire cell machinery is devoted to synthesize, export and reorganize cell walls. The cell wall remains dynamic in real sense provided with the required plasticity and elasticity. Etiolation of hypocotyls makes them to elongate more than hypocotyls of light-grown seedlings. In dark-grown seedlings, hypocotyls elongate along a spatially and temporally steep acropetal gradient. This growth takes place mainly by cell elongation (100-fold elongation as compared to embryo cells) and does not involve significant cell divisions. Elongation follows an acropetal gradient and takes place in two distinct, time-separated phases: synthesis and deposition of new cell wall polymers and addition and re-organization of the existing ones (Gendreau *et al.*, 1997; Derbyshire *et al.*, 2007). During the first 3 days after germination, synthesis and deposition of cell wall material are the main processes, producing cells with thick walls. In the following days (until 7 days), hypocotyls grow mainly by extensive polymer disassembly and rearrangement resulting in thinner cell walls (Figure 1.14). Furthermore they respond to growth hormones as normal plant parts do. The model plant *A. thaliana* have the advantage of the existence of mutants with altered hypocotyl growth facilitating gene functional study (Mouille *et al.*, 2003). Owing to these characteristics, etiolated hypocotyls of *A. thaliana* could be considered an ideal model for the study of cell elongation.

---

#### 1.4. Objectives of the project

As discussed above, CWPs contribute to several physiological functions. But these correspond to only a small part of the cell wall genes and role of most of the CWPs is still hidden. This project is aimed at identifying cell wall genes involved in cell elongation or its arrest. Here we have tried to identify the members of the multigene families known to be involved in elongation or elongation arrest in etiolated hypocotyls of *A. thaliana* as well as other proteins that could contribute to these processes.

The questions raised during this work are:

- i) Which are the proteins present in cell wall of elongating hypocotyls and fully-grown hypocotyls that have stopped elongation?
- ii) Do there exist differences between the cell wall proteomes of elongating organs and those that have stopped elongation? What are these differences?
- iii) What are the functions of these proteins?

To achieve these goals, comparative cell wall proteomics was used as primary approach in etiolated hypocotyls of *A. thaliana*: a fast growing stage (5-day-old) and a stage when there is no more elongation (11-day-old) will be compared. Since previous studies showed no defined correlation between the mRNAs and protein concentrations (Moritz and Meyer, 2003), working directly with proteins was preferred here. Proteomics has been defined as a global qualitative analysis of complex protein mixtures, including the post-translationally modified proteins as well as those encoded by alternatively spliced transcripts, and appears as a complementary approach to genomic and transcriptomic studies (Hunter *et al.*, 2002). Comparative proteomics provides an overview of the proteome and detects proteins which are altered between different stages (Moritz and Meyer, 2003) like the two physiological stages in this study. In addition, a comparative transcriptomic study was performed on the same material. Microarrays are a powerful, sensitive, versatile, and easy-to-use genomic tool that can simultaneously determine expression levels for thousands of genes at reasonable cost (Meyers *et al.*, 2004; Trevino *et al.*, 2007). Furthermore, combining the two types of data will give information about the post-transcriptional regulations of genes encoding CWPs. Here a global view of the functions of different genes representing these two physiological stages will be given.

# **CHAPTER 2**

## **MATERIALS AND METHODS**



---

## **2.1. Materials**

### **2.1.1. Plant material**

*Arabidopsis thaliana* ecotype Columbia 0 (Col 0) was used for *in vitro* and *in vivo* cultures for obtaining the hypocotyls, other plant material, and for transformation with the desirable constructs.

### **2.1.2. *In vitro* culture for obtaining etiolated hypocotyls**

One hundred and thirty mg seeds of *A. thaliana* were weighed in an Eppendorf tube, soaked for 2-3 h in 1 mL tap water, sterilized by treating with 4 times diluted Javel (sodium hypochlorite) for 45 min. To remove the sterilizing solution, the seeds were washed 6 times with 1 mL of sterilized ultra high quality (UHQ) water each time. Finally, the seeds were sowed in Magenta boxes containing 50 mL of Murashige and Skoog (MS) medium (Murashige and Skoog, 1962) pH 5.8, supplemented with 2% (w/v) sucrose and 0.8-1.2% (w/v) agar. Synchronization of germination was obtained by 2-4 days chill treatment and 4 h light treatment of the seeds. Seedlings were grown at 23°C in the dark for 5 or 11 days.

## **2.2. Methods**

### **2.2.1. Isolation of cell walls from hypocotyls (after Feiz *et al.* (Feiz *et al.*, 2006))**

For one experiment, hypocotyls were collected from 18 and 36 Magenta boxes for the 11 day-old and 5 day-old samples respectively. Hypocotyls of *A. thaliana* were collected by cutting the seedlings with scissors below the cotyledons and above the crown measuring ~0.7 cm for 5 day-old and ~1.2 for 11 day-old seedlings. The hypocotyls thus obtained were transferred to 5 mM sodium acetate buffer (pH 4.6), 0.4 M sucrose in a large Petri dish kept on ice. To remove the cut cotyledons and testa, hypocotyls were extensively washed on a 2 mm mesh, with 0.8-3.0 L of 5 mM sodium acetate buffer (pH 4.6), 0.4 M sucrose depending on the weight of the hypocotyls. Afterwards, the hypocotyls were transferred to a Moulinex®-type grinder containing 600 mL of the same buffer supplemented with 1 mL of protease inhibitor cocktail (Sigma Aldrich, Saint Louis, USA) and Polyclar® at the rate of 1 g/10 g fresh weight of hypocotyls in order to complex the phenolic compounds (Charmont *et al.*, 2005). The mixture was ground in cold room at full speed for 5 min. Cell walls were separated from soluble cytoplasmic materials by centrifugation for 15 min at 1000 × *g* at 4°C using the Beckman J2-HC centrifuge and a JA 14 rotor. They were further purified by two additional centrifugations in 150 mL per tube of 5 mM sodium acetate buffer (pH 4.6) containing 0.6 M and 1 M sucrose respectively. The cell wall containing pellet was hold at

each step. To ensure the removal of cytosolic proteins and other soluble material, the pellet was extensively washed on nylon net (25  $\mu$ m pore size) with 3-5 L of 5 mM sodium acetate buffer (pH 4.6). Finally it was ground in liquid nitrogen with mortar and pestle in order to obtain very fine powder and to have intimate contact with the protein extraction buffer. The cell wall fraction was lyophilized and stored at -20°C.

### **2.2.2. Sequential extraction of proteins from purified cell walls**

Generally, 0.6-0.9 g of lyophilized cell walls were taken per tube for protein extraction. Proteins were extracted in 4 successive steps by using 10 mL/g of salt solutions in this order: 2 extractions with 6-9 mL 5 mM sodium acetate buffer (pH 4.6), 0.2 M  $\text{CaCl}_2$ , followed by 2 extractions with 6-9 mL 5 mM sodium acetate buffer (pH 4.6), 2 M LiCl. Protease inhibitor cocktail (Sigma Aldrich, Saint Louis, USA), was added at a concentration of 15  $\mu$ L/g cell walls, during the first extraction with each salt. Cell walls were resuspended by vortexing during 5-10 min at room temperature and then centrifuged for 15 min, at 40000  $\times g$  and 4°C. The protein content of each extract was measured using the Bradford method (Bradford, 1976) with the Coomassie<sup>®</sup> protein assay reagent kit (Pierce, Perbio Science, Rockford, USA) taking bovine serum albumine (BSA) as a standard. Proteins were desalted using Econo-Pac<sup>®</sup> 10DG columns (Bio-Rad Laboratories, Inc. Hercules, CA, USA) equilibrated with 0.2 M ammonium formate. Depending on the volume of each extract and of protein concentration, protein solutions could be concentrated by successive centrifugations using the Centriprep<sup>®</sup> centrifugal filter device (YM-10 kDa membrane for volumes greater than 6 mL or 5 kDa for smaller volumes) (Millipore, Billerica, MA, USA) at 4000  $\times g$ . Finally, proteins were lyophilized.

### **2.2.3. Protein fractionation by cation exchange chromatography**

All the lyophilized extracts were combined and redissolved in a total volume of 2 or 3 mL of water. They were again quantified as described above. One mg of proteins was used for chromatographic fractionation on a 1 mL HiTrap<sup>™</sup> SP FF column (Amersham Biosciences, Uppsala, Sweden) equilibrated with 50 mM MES (pH 5.6m adjusted with NaOH) operated with an FPLC<sup>™</sup> System (Amersham Biosciences, Uppsala, Sweden) controlled by FPLCdirector<sup>™</sup> version 1.0 (Amersham Biosciences, Uppsala, Sweden). The protein solution was adjusted to 50 mM MES (pH 5.6) and 20  $\mu$ L protease inhibitor cocktail (Sigma Aldrich, Saint Louis, USA) were added before loading onto the column. Loading onto

the column was accomplished at a flow rate of 0.5 mL/min. A 10 mL unfixed fraction was collected at the same rate. Three mL of first wash with 50 mM MES (pH 5.6) were collected at a flow rate of 1 mL/min. Fixed proteins were eluted by a gradient from 0 to 0.8 M NaCl in 50 mM MES (pH 5.6) and 24 fractions (1 mL each) were collected at a flow rate of 1 mL/min. A modified gradient was applied when enrichment of certain proteins was required. Finally the column was successively washed with 3 mL of 1.2 M and 3 mL of 1.5 M NaCl in 50 mM MES (pH 5.6) at the same flow rate. These washes were also collected as 6 fractions (1 mL/tube). Two  $\mu$ L/mL Protease inhibitor cocktail (Sigma Aldrich, Saint Louis, USA) was added to all the 1 mL collecting tubes. Quantity of proteins in each fraction was measured by the Bradford method (Bradford, 1976). The fractions were combined in groups of 2 or 3 depending on their protein concentration and were desalted as previously described prior to lyophilization.

#### **2.2.4. 1D-E (SDS-polyacrylamide gel electrophoresis) and staining procedures**

Each lyophilized chromatography fraction (group of 2 or 3 fractions) was redissolved in 200  $\mu$ L water and electrophoresis of proteins was performed according to Laemmli (Laemmli, 1970). Samples were loaded on 12  $\times$  15 cm polyacrylamide gel with a concentration of 12.5% and a thickness of 1.5 mm (Annex I). For transfer, 8% polyacrylamide gels of 8  $\times$  6 cm and 1.5 mm thickness were used (Annex II).

Gels were stained with any one of the following staining procedures:

##### **2.2.4.1. Staining with Coomassie Brilliant Blue (CBB)**

CBB staining was performed according to Scheler *et al.* (Scheler *et al.*, 1998). After electrophoresis, the gel was fixed by overnight dipping and gentle shaking in 50% methanol. It was then rinsed 3 times with UHQ water by changing the water after each 30 min. Then, the gel was sensitized by dipping it in sensitising solution (34% methanol, 2% H<sub>3</sub>PO<sub>4</sub> 17% ammonium sulfate) for 1 h. Finally, the gel was put in the staining solution (34% methanol, 2% H<sub>3</sub>PO<sub>4</sub> 17% ammonium sulfate, CBB G (Sigma Aldrich, Saint Louis, USA) 0.66 g/L for 2-3 days. For clearing the background, the gel was rinsed 1-2 times in 2% H<sub>3</sub>PO<sub>4</sub> before scanning the gel. The gel was stored at 4°C in 2% H<sub>3</sub>PO<sub>4</sub>.

---

#### 2.2.4.2. Staining with silver nitrate

For silver nitrate staining, the protocol of Shevchenko *et al.* (Shevchenko *et al.*, 1996) was followed. The gel was first fixed in 45% ethanol, 5% acetic acid, 55% UHQ water for 30 min to one night. Then, the gel was rinsed with 30% ethanol and UHQ water, one after the other (10 min  $\times$  3 each). The gel was sensitized by treating with 0.02% Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>. Then, it was rinsed 2 times for 1 min each, with UHQ water. Lastly, stain was developed on gel with 0.04% formaldehyde, 0.2% Na<sub>2</sub>CO<sub>3</sub>. It was left on a shaker and stopped when enough stained by taking it out and immersing in 1% acetic acid and stored in it at 4°C.

#### 2.2.4.3. Staining with Amido Black

Amido Black staining was used for staining polyacrylamide gel and PVDF transfer membrane (0.2  $\mu$ m, Schleicher & Schuell, Dassel, Germany) after transferring proteins on it. The protocol was provided by Mr. Jacques D'Alayer, Head of "Le Plateau d'Analyse et de Microséquençage des Protéines", Pasteur Institute, Paris.

The gel was fixed in 50% methanol, 10% acetic acid solution for 30 min  $\times$  2. The gel was stained by overnight treatment with 50% methanol, 10% acetic acid and 0.003% (3 mg/100 mL) Amido black (Naphthol Blue Black, Sigma Aldrich, Saint Louis, USA). The gel was rinsed 3 times with UHQ water to remove methanol, acetic acid and excess of colour.

For the PVDF membrane, once the proteins were transferred on it, it was first rinsed with UHQ water and stained in staining solution (40% methanol, 1% acetic acid) for 15-60 s. Water was used for destaining gel background. The membrane was dried in folds of blotting paper. Pieces of 2  $\times$  8 mm were cut for N-terminal sequencing.

#### 2.2.4.4. Staining with Gelcode<sup>®</sup> Glycoprotein staining kit

Pierce kit (Perbio Science, Rockford, USA) was used to detect the glycoproteins present in the crude extract of hypocotyls. After electrophoresis, the gel was completely immersed in 50% methanol for 30 min. Then, the gel was washed twice with 3% acetic acid. The gel was transferred to "Oxidizing Solution" and let to agitate for 15 min. The gel was again washed twice with 3% acetic acid for 5 min. Now, the gel was treated with "Reducing Solution" for 5 min with gentle agitation. Finally, the gel was washed extensively with 3% acetic acid and then with UHQ water. The gel was stored in 3% acetic acid.

### 2.2.5. Transfer of proteins from gel on PVDF membrane

For this, tank transfer method using Mini Trans-Blot<sup>®</sup> Electrophoretic Transfer Cell (Bio-Rad Laboratories, Inc. Hercules, CA, USA), 8% polyacryamide gel, PVDF membrane was followed. PVDF membrane (0.2  $\mu$ m, Schleicher & Schuell, Dassel) measuring 8  $\times$  10 cm, was first immersed for a while in pure methanol, rinsed 2 times for 10 min with UHQ water and finally equilibrated in cold transfer buffer (0.025 M Tris, 0.192 M Glycine, 0.01% SDS). The fiber pads and Whatman 3 mm filter paper were also soaked in cold transfer buffer. The transfer sandwich (also named gel cassette) was prepared inside the transfer buffer contained in a tray. From cathode towards anode, the order was as follows: first fiber pad, 2 folds of Whatman<sup>®</sup> 3 mm filter paper, gel, PVDF membrane, 2 folds of Whatman<sup>®</sup> 3 mm filter paper and fiber pad. The sandwich and an ice cube were inserted into the tank which was completely filled with the transfer buffer. Transfer was carried out in cold room at constant 75 V during 2 h 15 min.

### 2.2.6. Protein identification by mass spectrometry

Protein bands were excised from stained gels with a 2 mm inner diameter Pasteur pipette. Each gel piece was washed twice with 100  $\mu$ L of acetonitrile (ACN)/25 mM ammonium bicarbonate (50/50 v/v), for 10 min. After drying under vacuum for 15 min, gel pieces were rehydrated with 10  $\mu$ L of sequencing grade a trypsin (Promega Corporation, Madison, WI, USA) solution (10  $\mu$ g/mL) in 25 mM ammonium bicarbonate and digested overnight at 37°C. After addition of 1  $\mu$ L ACN, a 5 min sonication was performed. An aliquot of 1  $\mu$ L was deposited on the sample plate of the mass spectrometer with 1  $\mu$ L of the matrix solution (6 g/L of  $\alpha$ -cyano-4-hydroxycyannamic acid in 50% ACN/0.1% trifluoroacetic acid).

All mass spectrometry (MS) analyses were performed on the “Plate-Forme de Protéomique”, Toulouse Midi-Pyrénées, France. Matrix Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometer (Voyager-De<sup>TM</sup> STR, Perseptive Biosystems, USA), was operated in positive reflector mode at the following parameters: accelerating voltage 20 kV; grid voltage 68%; extraction delay time, 200 ns. An external calibration was realized with a mixture of known masses of peptides (monoisotopic MH<sup>+</sup> 904.4681, 1296.6853, 1570.6774, 2093.0867, 2465.1989). Acquisition mass was between 750 and 4000 Da. Mass data were analysed by Data Explorer software (Applied Biosystems/MDS Sciex, Foster City, CA, USA). For each sample, an internal mass calibration was performed

using trypsin autolysis peaks (monoisotopic  $MH^+$  842.5100, 1045.5600 and 2211.1046). Peptide mass fingerprinting data were analyzed by MS-FIT Protein Prospector (<http://prospector.ucsf.edu>). The retained parameters were mass tolerance of 20 ppm, one missed cleavage, NCBI non redundant database (2005.01.06), 6 open reading frames and oxidation of methionine (or in some cases hydroxylation of proline). In some cases peptides were analyzed using, the hybrid linear ion trap mass spectrometer (Q TRAP<sup>®</sup> MS), LC/MS/MS system (Applied Biosystems/MDS Sciex, Foster City, CA, USA) as described in Boudart *et al.* (Boudart *et al.*, 2005). Data were processed by the Analyst software (Applied Biosystems/MDS Sciex, Foster City, CA, USA) and submitted to the search software MASCOT (<http://www.matrixscience.com/home.html>).

### 2.2.7. Extraction of RNAs

RNAs were extracted from *A. thaliana* hypocotyls for comparative transcriptomics. Hypocotyls were cut below cotyledons and above crown with scissors and grinded in liquid nitrogen using a mortar and pestle and stored at -80°C.

RNAs were extracted using “SV total RNA extraction kit” (Promega) following their protocol for “Isolation of Total RNA from Plant Tissue” and “RNA purification by centrifugation (spin)” with minor changes required. The ground tissue was mixed with “RNA Lysis Buffer” (175  $\mu$ L/30 mg). Then added to it “RNA Dilution Buffer” (350  $\mu$ L/30 mg) and mixed by inversion. Now, the mixture was distributed in 1.5 mL Eppendorf tubes (1 tube/30 mg ground tissue). The tubes were centrifuged at 13,000 rpm in a microcentrifuge for 10 min. The cleared lysate solution was transferred to a fresh microcentrifuge tube by pipetting. To this, 200  $\mu$ L of 95% ethanol were added, and mixed. The mixture was then transferred to the “Spin Column Assembly” (SCA) and centrifuged at 13,000 rpm for 1 min. The flow-through was discarded and 600  $\mu$ L of “RNA Wash Solution” were added to the “SCA”. It was again centrifuged at 13,000 rpm for 1 min. Then 50  $\mu$ L of freshly prepared “DNase Incubation Mixture” (40  $\mu$ L Yellow Core Buffer, 5  $\mu$ L 0.09 M  $MnCl_2$  and 5  $\mu$ L of DNase I enzyme per sample) was directly added to the membrane inside the “SCA”. After application of the mixture, “SCA” was incubated for 15 min at 20-25°C. After incubation, 200  $\mu$ L of “DNase Stop Solution” were added to the “SCA”, and centrifuged at 13,000 rpm for 1 min. Now the membrane inside the “SCA” was washed successively with 600  $\mu$ L and 250  $\mu$ L “RNA Wash Solution” by centrifugation at 13,000 rpm for 1 and 2 min respectively. For eluting RNA, the basket was transferred from the “Collection Tube” to the “Elution Tube” and 100  $\mu$ L

“Nuclease-Free Water” were added to the membrane before centrifugation at 13,000 rpm for 1 min. The RNAs were collected in “Elution Tube” and stored at -80°C. Generally 108 µg and 118 µg RNAs were extracted from 800 mg of powder of 5 and 11 days hypocotyls respectively.

### **2.2.8. RT-PCR and semi-quantitative PCR**

RT-PCR was used for comparison of transcript levels of selected genes in 5-day-old hypocotyls to those in 11-day-old hypocotyls of wild type plants. The RNAs extracted from the hypocotyls were used as starting material. Complementary DNA (cDNA) first strand were obtained from mRNA templates through reverse transcription. For this, 1 µL dNTPs (10 mM each), 1 µL oligo (dT)<sub>15</sub> primers (Promega) and 1 µL pork desmin mRNA were added to 1 µg (10 mL) total RNAs (treated with DNase). In order to denature RNAs, the reaction mixture was heated at 65°C for 5 min and immediately put on ice. Then 4 µL of 5× First Strand Buffer (Invitrogen), 2 µL of 0.1 M DTT (Invitrogen) and 1 µL RNasine Plus RNase Inhibitor (Promega) were added to the reaction tube. Then the mixture was incubated at 42°C for 2 min before addition of 1 µL of SuperScript<sup>TM</sup> II reverse transcriptase (Invitrogen). Again the tube was incubated at 42°C for 1 h. At the end, the enzyme was denatured by incubating the tubes at 70°C for 15 min. The cDNAs obtained in this way were kept at -20°C. The details of the reaction are given in Annex III.

For comparing RNA quantity of selected genes in 5- and 11-day-old hypocotyls, qPCR was performed on their respective first strand cDNA using a Roche lightcycler system (Roche Diagnostics, Meylan, France) according to manufacturer’s recommendations and gene specific oligonucleotide primers (Table 2.1). Using the results from quantitative PCR to determine the number of amplification cycles required to be in a linear range for all genes of interest, semi-quantitative PCR was performed to compare relative amounts of the corresponding mRNAs. The amplified fragments were analyzed by electrophoresis in polyacrylamide gels in standardized conditions. In each case, the presence of a fragment of the expected size was checked after staining with ethidium bromide.

Table 2.1 Primers used for RT-qPCR amplifications

Accession number	Primer sequence (5'→3', sense orientation)	Primer sequence (5'→3', antisense orientation)
<i>At1g05570</i>	CAC TTCCTCTGATAACAATCAAAGAC	GC TTACACGCTTGTGCAATG
<i>At1g10550</i>	CAATTCAGTAAGATCGCCATTG	TTTGACACCAACCCAGCTC
<i>At1g28290</i>	AACCATAAGACTCAAACCCTTC	TGAGGGTGGTGGTGATGAG
<i>At1g49240</i>	CACCCGAGAGGAAGTACAGTG	CATACTCTGCCTTAGAGATCCACA
<i>At1g66180</i>	ACCACAAGCTCAACAAATGGT	GGAGGAAGCTTTTACGATGAC
<i>At1g68560</i>	GTCGCCGTTAAATGTTGTTG	CCCATCTACATTGATGAAATCCT
<i>At1g69530</i>	GCATCGCTCAATACAGAGC	GAGTGTCCGTTTATCGTAAACCTT
<i>At2g21140</i>	TCTCATCCTTGATAAAGATGC	TCTACGTAGAAATCAACAAAGC
<i>At2g28790</i>	GTGCCCACTACAACGGAAAA	TGAGCTAGAGAAGCTGGTGGT
<i>At3g07130</i>	GGGATCGGCAACCTGATTA	GCCCATGTCTCGTTCTTCAT
<i>At3g16850</i>	GTAGCCTTCAGAACACCAGA	TCAACGCTTGTGAGATCGAC
<i>At3g43270</i>	TTCCGATCTGCTACATTGCT	GACCGCTTGGTGTTTTTCC
<i>At4g12880</i>	GGCATGAAGCTAGATGTTTTAGTTG	ACATAAAATTGATTTCTTATTGTGCTG
<i>At4g18670</i>	CACCACCTCCAATCTACGAAG	GCCCCTTTTGAGAACATTCTG
<i>At5g05850</i>	TGGTACCACTGGAGGAAAGC	TTGCAACTCCATAGCCACAG
<i>At5g11420</i>	GTCTCTTCTCTTTACTTTGGTCGTC	AGTCGCCGTTTGGTAAACATC
<i>At5g44360</i>	AGGAGCTTATTTGAATTACCGAGA	TCGCATCTTCAAAGCTCGTA
<i>At5g64100</i>	GCAAGACTTCGCTGCTAAAAAC	GCCGTTGAAGTTAACGAACC
<i>desmin<sup>a</sup></i>	CAGCCTCAGTCCTCCAAATCACA	TAGGCCTGAGGTCACAGAGGT

<sup>a</sup> : Pig desmin RNA was used as an internal control for reverse transcription

### 2.2.9. Microarray analysis

Microarray analysis was carried out at the Unité de Recherche en Génomique Végétale (URGV), Evry, France, by Mrs. Ludivine Soubigou-Taconnat and Dr. Jean-Pierre Renou, using the CATMA array (Crowe *et al.*, 2003; Hilson *et al.*, 2004), containing 24,576 gene-specific sequence tags (GSTs) from *Arabidopsis*. RNA samples from 3 independent biological replicates were pooled. For each comparison, one technical replication with fluorochrome reversal was performed for each RNA sample. The reverse transcription of RNAs in the presence of Cy3-dUTP or Cy5-dUTP (Perkin-Elmer-NEN Life Science Products), the hybridization of labeled samples to the slides, and the scanning of the slides were performed as described in Lurin *et al.* (Lurin *et al.*, 2004).

### 2.2.10. Statistical analysis of microarray data

Experiments were designed with the statistics group of the URGV. Statistical analysis was based on 2 dye swaps (*i.e.* 4 arrays, each containing 24,576 GSTs and 384 controls) as described in Lurin *et al.* (Lurin *et al.*, 2004). Controls were used for assessing the quality of the hybridizations, but were not included in the statistical tests or the graphic



representation of the results. For each array, the raw data comprised the logarithm of median feature pixel intensity at wavelengths 635 (red) and 532 nm (green). No background was subtracted. In the following description, log ratio refers to the differential expression between two conditions. It is either  $\log_2$  (red/green) or  $\log_2$  (green/red) according to the experimental design. Array-by-array normalization was performed to remove systematic biases. First, spots that were considered badly formed features were excluded. Then, global intensity-dependent normalization using the LOESS procedure was performed to correct the dye bias. Finally, for each block, the  $\log_2$  ratio median calculated over the values for the entire block was subtracted from each individual  $\log_2$  ratio value to correct print tip effects on each metablock. To determine differentially expressed genes, a paired t test was performed on the  $\log_2$  ratios, assuming that the variance of the  $\log_2$  ratios was the same for all genes. Spots displaying extreme variance (too small or too large) were excluded. The raw p-values were adjusted by the Bonferroni method, which controls the Family Wise Error Rate (FWER). We considered as being differentially expressed the genes with an FWER of 5%. The Bonferroni method (with a type I error equal to 5%) was used in order to keep a strong control of the false positives in a multiple-comparison context (Ge *et al.*, 2003).

### 2.2.11. Data Deposition

Microarray data from this article were deposited at Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>; accession No. E-MEXP-789) and at CATdb (<http://urgv.evry.inra.fr/CATdb/>; Project RS05-11\_Hypocotyls) according to the “Minimum Information About a Microarray Experiment” standards.

### 2.2.12. Bioinformatic analysis

DNA sequences related to each protein or to each gene were collected using data available in the Unigene database (<http://www.ncbi.nlm.nih.gov>). The *Arabidopsis* Genome Initiative (AGI) nomenclature was used (<http://mips.gsf.de/proj/thal/>, <http://www.tigr.org/tdb/tgi/agi/>). Sub-cellular localization and length of signal peptides were predicted using PSORT (<http://psort.nibb.ac.jp>) and TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) (Nielsen *et al.*, 1997; Emanuelsson *et al.*, 2000). Prediction of transmembrane domains was done using Aramemnon (<http://aramemnon.botanik.uni-koeln.de/>) (Schwacke *et al.*, 2003). Molecular mass and pI value were calculated using the aBi program ([http://www.up.univ-mrs.fr/~wabim/d\\_abim/compo-p.html](http://www.up.univ-mrs.fr/~wabim/d_abim/compo-p.html)). Homologies to other proteins were searched using

---

BLAST programs (<http://www.ch.embnet.org/software/BottomBLAST.html>) [Altschul *et al.*, 1990]. Identification of protein families and domains was performed using InterProScan (<http://www.ebi.ac.uk/InterProScan/>) (Quevillon *et al.*, 2005). GHs and CEs were classified according to the CAZy database (<http://www.cazy.org/CAZY/>) (Coutinho and Henrissat, 1999). The GT77 family was annotated according to Egelund *et al.* (Egelund *et al.*, 2004). XTHs and expansins were named according to <http://labs.plantbio.cornell.edu/xth/> and <http://www.bio.psu.edu/expansins/index.htm> respectively. AGPs and FLAs were named according to Schultz *et al.* (Schultz *et al.*, 2002), Johnson *et al.* (Johnson *et al.*, 2003), Van Hengels and Roberts (van Hengel and Roberts, 2003) and Liu and Mehdy (Liu and Mehdy, 2007). Proteins homologous to COBRA, LRXs and Hyp/Pro-rich proteins were annotated according to Roudier *et al.* (Roudier *et al.*, 2002), Baumberger *et al.* (Baumberger *et al.*, 2003), and Fowler *et al.* (Fowler *et al.*, 1999) respectively. The lignin toolbox was proposed by Raes *et al.* (Raes *et al.*, 2003). Peroxidases were named as in the PeroxiBase (<http://peroxidase.isb-sib.ch/index.php>) (Bakalovic *et al.*, 2006). Laccases were annotated as in Pourcel *et al.* (Pourcel *et al.*, 2005) and McCaig *et al.* (McCaig *et al.*, 2005). SKU-like proteins and phytocyanins were described in Jacobs and Roe (Jacobs and Roe, 2005), and Nersissian and Shipp (Nersissian and Shipp, 2002) respectively. Subtilases are listed at <http://csbdb.mpimp-golm.mpg.de/csbdb/dbcawp/psdb.html>.

## **CHAPTER 3**

### **EVALUATION OF CELL WALL PREPARATIONS FOR PROTEOMICS: A NEW PROCEDURE FOR PURIFYING CELL WALLS FROM *ARABIDOPSIS* HYPOCOTYLS**

---

**Chapter summary (French)**

Le premier objectif de ce travail étant d'identifier les protéines présentes dans les hypocotyles en élévation et après l'achèvement de leur croissance, il nous fallait mettre en œuvre une étude de protéomique. En effet, la protéomique est un outil de choix pour une étude globale des protéines. Essentiellement deux types de stratégies, appelées non-destructive ou destructive, ont été mises en œuvre jusqu'à présent, aucune d'elles n'étant capable de révéler l'ensemble des protéines pariétales (Lee *et al.*, 2004 ; Jamet *et al.*, 2008). Dans le premier cas, l'intégrité des cellules est respectée puisqu'il s'agit soit d'analyser des protéines sécrétées dans des milieux de culture, soit de pratiquer des infiltrations dans des espaces intercellulaires avec des solutions salines. Dans le second cas, les tissus sont broyés avant de procéder à la purification des parois et à l'extraction des protéines avec des solutions salines. Cette dernière approche présente un certain nombre de contraintes qui affectent les résultats qualitativement et quantitativement :

- i) les protéines pariétales peuvent être perdues au cours de la purification des parois puisque ce compartiment extracellulaire n'est pas entouré d'une membrane ;
- ii) les protéines extracellulaires peuvent être contaminées par des protéines intracellulaires qui se lient de manière non-spécifique aux polysaccharides des parois au cours de leur purification.

Dans les études protéomiques précédemment publiées, ces problèmes ont rarement fait l'objet d'une grande attention. C'est ainsi qu'un grand nombre de protéines connues ou prédites par bioinformatique pour être intracellulaires (parfois plus de 50%) ont été considérées comme pariétales. Pour pratiquer une étude protéomique des parois végétales, la qualité de la purification des parois apparaît donc essentielle. Avant de commencer la mise au point de notre protocole de purification de parois, nous avons analysé en détail les résultats de

---

plusieurs études publiées précédemment. Nous avons retenu plusieurs points qui nous paraissaient importants et avons essayé d'apporter des solutions :

- i) Perte de protéines au cours de la purification des parois en présence de tampon à forte force ionique : nous avons employé un tampon de faible force ionique (tampon acétate de sodium 5 mM) pendant toute la procédure de purification des parois.
- ii) Contamination des parois par des organelles ou des vésicules membranaires : nous avons employé des centrifugations successives en présence de concentrations croissantes de saccharose (0.4 M, 0.6 M, et 1.0 M). La forte densité des parois permettait ainsi de les recueillir sous forme de culot après chaque centrifugation.
- iii) Contamination des parois par des protéines intracellulaires : nous avons introduit une étape de lavage extensif des parois purifiées avec le tampon à faible force ionique (tampon acétate de sodium 5 mM).
- iv) Protéolyse : des inhibiteurs de protéases ont été ajoutés tout au long de la procédure de purification des parois et des protéines.
- v) Contaminations intracellulaires dues à la présence de détergents dans le tampon d'extraction : aucun détergent n'a été ajouté dans les tampons utilisés.
- vi) Méthodes de vérification de la qualité des préparations de parois : les méthodes traditionnelles se sont révélées inappropriées pour vérifier la présence de contaminants intracellulaires par manque de sensibilité. Nous avons employé systématiquement les outils de la bioinformatique pour prédire l'adressage sub-cellulaire des protéines ainsi que les données de la littérature. Il s'agissait de s'assurer de la présence d'un peptide signal et de l'absence de signal de rétention des protéines dans le reticulum endoplasmique.

---

Le protocole de purification retenu a permis d'obtenir des préparations de parois de bonne qualité, comme en témoigne la grande proportion de protéines dont la localisation sub-cellulaire prédite est extracellulaire (environ 80%). Cette proportion n'a été atteinte dans aucune autre étude publiée précédemment. Les résultats sont détaillés dans l'article qui suit publié dans *Plant Methods*.

As a first objective of the work to identify the proteins that are present in the elongating and fully-elongated hypocotyls, a proteomic study was needed. Proteomics is considered an appropriate tool for large scale studies. Several methods classified as destructive and non-destructive are used (Lee *et al.*, 2004; Jamet *et al.*, 2008) since no single method is sufficient alone to study complete cell wall proteome. In the non-destructive methods living cells are used, and proteins secreted in the culture medium or extracted with washing of cells or vacuum infiltration of tissues with salt solutions are analysed. In other cases, destructive methods are employing the purification of cell wall from the tissue followed by extraction of the proteins with salts. However, the latter approach has several constraints that affect the results qualitatively and quantitatively:

- i) loss of CWPs during cell wall purification due to the lack of a surrounding membrane
- ii) contamination of CWPs by intracellular proteins that non-specifically binds to the cell wall polysaccharides in the cell wall preparation procedure.

In previous cell wall proteomic studies, these problems were not being given due attention and thus resulted in the identification of high number (up to 50% or more) of known or predicted intracellular contaminants. For performing a proteomic study on cell wall, purity of cell wall is of utmost importance. To establish a protocol that could purify a proteomic-grade cell wall, the existing protocols of cell wall preparation, and the results obtained with them were analysed. This analysis brought into our notice some steps that seem to be responsible for these problems. Keeping in mind these points, we have established a new protocol for cell wall preparation. These problems and distinguishing features of our protocol to overcome these problems are listed below:

- (i) Loss of CWPs by using high ionic strength buffers in cell wall preparation: We replaced these buffers by low ionic strength buffer (5 mM acetate buffer) throughout cell wall preparation steps in this new protocol.
- (ii) Contamination of cell wall by cell organelles and other vesicles: To purify organelles-free cell wall, our protocol uses increasing sucrose concentration (0.4 M, 0.6 M and 1.0 M) during three successive centrifugations.
- (iii) Lack of extensive washes of the cell wall to wash away the soluble intracellular contaminants: To eliminate soluble intercellular contaminants, our protocol

---

proposes extensive washing of the cell wall pellet with the low ionic strength buffer (5 mM acetate buffer).

- (iv) Proteolytic degradation of CWPs: For preventing CWP proteolytic degradation, we use protease inhibitors.
- (v) Intracellular contamination caused by the use of detergents in the extraction buffer: Our new method avoids use of detergents in the protein extraction buffer from the purified cell wall that contribute to minimizing the number of intracellular proteins in CWP extract.
- (vi) Use of inappropriate methods to judge the purity of cell wall and to distinguish true CWPs from the false ones: Traditional methods used were not enough sensitive to distinguish between pure and impure cell walls and all the proteins identified were considered as CWPs. We analyse the identified proteins with bioinformatic tools to check for the presence of signal peptide and absence of ER retention signal.

During this study, the above protocol was used and satisfactory results were obtained. CWPs extracted from cell wall purified with this protocol allowed the identification of high number (about 80 %) of predicted CWPs with few intracellular contaminants. None of the previously published studies has reported such high proportion. The detailed results are given in the article published in *Plant Methods*.

The supplementary data files are given as annexes at the end of the report.



## Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis hypocotyls*

Leila Feiz<sup>†1,2</sup>, Muhammad Irshad<sup>†1</sup>, Rafael F Pont-Lezica<sup>1</sup>, Hervé Canut<sup>1</sup> and Elisabeth Jamet<sup>\*1</sup>

Address: <sup>1</sup>Surfaces Cellulaires et Signalisation chez les Végétaux, UMR 5546 CNRS-Université Paul Sabatier-Toulouse III, Pôle de Biotechnologie végétale, 24, Chemin de Borde Rouge, BP 42617 Auzeville, 31326 Castanet-Tolosan, France and <sup>2</sup>Department of Plant Science and Plant Pathology, Agriculture and Biological Sciences Faculty, Montana State University-Bozeman, Bozeman, MT 59717-3150, USA

Email: Leila Feiz - lfeiz@mymail.msu.montana.edu; Muhammad Irshad - muhammad@scsv.ups-tlse.fr; Rafael F Pont-Lezica - lezica@scsv.ups-tlse.fr; Hervé Canut - canut@scsv.ups-tlse.fr; Elisabeth Jamet\* - jamet@scsv.ups-tlse.fr

\* Corresponding author †Equal contributors

Published: 27 May 2006

Received: 27 February 2006

*Plant Methods* 2006, **2**:10 doi:10.1186/1746-4811-2-10

Accepted: 27 May 2006

This article is available from: <http://www.plantmethods.com/content/2/1/10>

© 2006 Feiz et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The ultimate goal of proteomic analysis of a cell compartment should be the exhaustive identification of resident proteins; excluding proteins from other cell compartments. Reaching such a goal closely depends on the reliability of the isolation procedure for the cell compartment of interest. Plant cell walls possess specific difficulties: (i) the lack of a surrounding membrane may result in the loss of cell wall proteins (CWP) during the isolation procedure, (ii) polysaccharide networks of cellulose, hemicelluloses and pectins form potential traps for contaminants such as intracellular proteins. Several reported procedures to isolate cell walls for proteomic analyses led to the isolation of a high proportion (more than 50%) of predicted intracellular proteins. Since isolated cell walls should hold secreted proteins, one can imagine alternative procedures to prepare cell walls containing a lower proportion of contaminant proteins.

**Results:** The rationales of several published procedures to isolate cell walls for proteomics were analyzed, with regard to the bioinformatic-predicted subcellular localization of the identified proteins. Critical steps were revealed: (i) homogenization in low ionic strength acid buffer to retain CWP, (ii) purification through increasing density cushions, (iii) extensive washes with a low ionic strength acid buffer to retain CWP while removing as many cytosolic proteins as possible, and (iv) absence of detergents. A new procedure was developed to prepare cell walls from etiolated hypocotyls of *Arabidopsis thaliana*. After salt extraction, a high proportion of proteins predicted to be secreted was released (73%), belonging to the same functional classes as proteins identified using previously described protocols. Finally, removal of intracellular proteins was obtained using detergents, but their amount represented less than 3% in mass of the total protein extract, based on protein quantification.

**Conclusion:** The new cell wall preparation described in this paper gives the lowest proportion of proteins predicted to be intracellular when compared to available protocols. The application of its principles should lead to a more realistic view of the cell wall proteome, at least for the weakly bound CWP extractable by salts. In addition, it offers a clean cell wall preparation for subsequent extraction of strongly bound CWP.

## Background

Cell walls are natural composite structures, mostly made of high molecular weight polysaccharides, proteins, and lignins, the latter found only in specific cell types. They are dynamic structures contributing to the general morphology of the plant. Cell walls are involved in cell expansion and division, and they are sources of signals for molecular recognition within the same or between different organisms [1-5]. Cell wall proteins (CWP) represent a minor fraction of the wall mass: 5–10% in primary cell walls of dicots, as reported for cell suspension cultures, but accurate determinations in various plant organs are still lacking [6]. Despite their low abundance, CWP contribute, at least in part, to the dynamic of cell walls. CWP can be involved in modification of cell wall components, wall structure, signalling, and interactions with plasma membrane proteins at the cell surface [7].

Proteomics appears to be a suitable method to identify a large number of CWP thus providing information for many genes still lacking a function. Recent publications on cell wall proteomics have shown that more than 50% of the identified proteins were known to be intracellular proteins in higher plants [8,9], green alga [10] and fungi [11]. Different techniques unrelated to proteomics, such as biotinylation of cell surface proteins, or immunoelectron microscopy, also suggested a cell wall location for some glycolytic enzymes, proposing that they are *bona fide* components of the yeast cell wall [11]. However, the reliability of protein profiling for a compartment like the cell wall, strongly depends on the quality of the preparation. Unfortunately, the classical methods to check the purity of a particular fraction are not conclusive for proteomic studies, since the sensibility of the analysis by mass spectrometry is 10 to 1000 times more sensitive than enzymatic or immunological tests using specific markers. Our experience in the field has shown that the most efficient way to evaluate the quality of a cell wall preparation is (i) to identify all the proteins extracted from the cell wall by mass spectrometry, and (ii) to perform extensive bioinformatic analysis to determine if the identified proteins contain a signal peptide, and no retention signals for other cell compartments [12-15]. It is then possible to conclude about the quality of the cell wall preparation by calculating the proportion of predicted secreted proteins to intracellular ones.

The aim of the present study is to present a comparative analysis of different methods previously published to prepare cell walls for proteomic studies. These methods will be evaluated by the proportion of proteins predicted to be secreted after bioinformatic analysis as stated above. A new method is presented, based on classical cell wall preparations, but adapted to the new technologies. The

results indicate that such a method significantly reduces the number of proteins without predicted signal peptide.

## Results and discussion

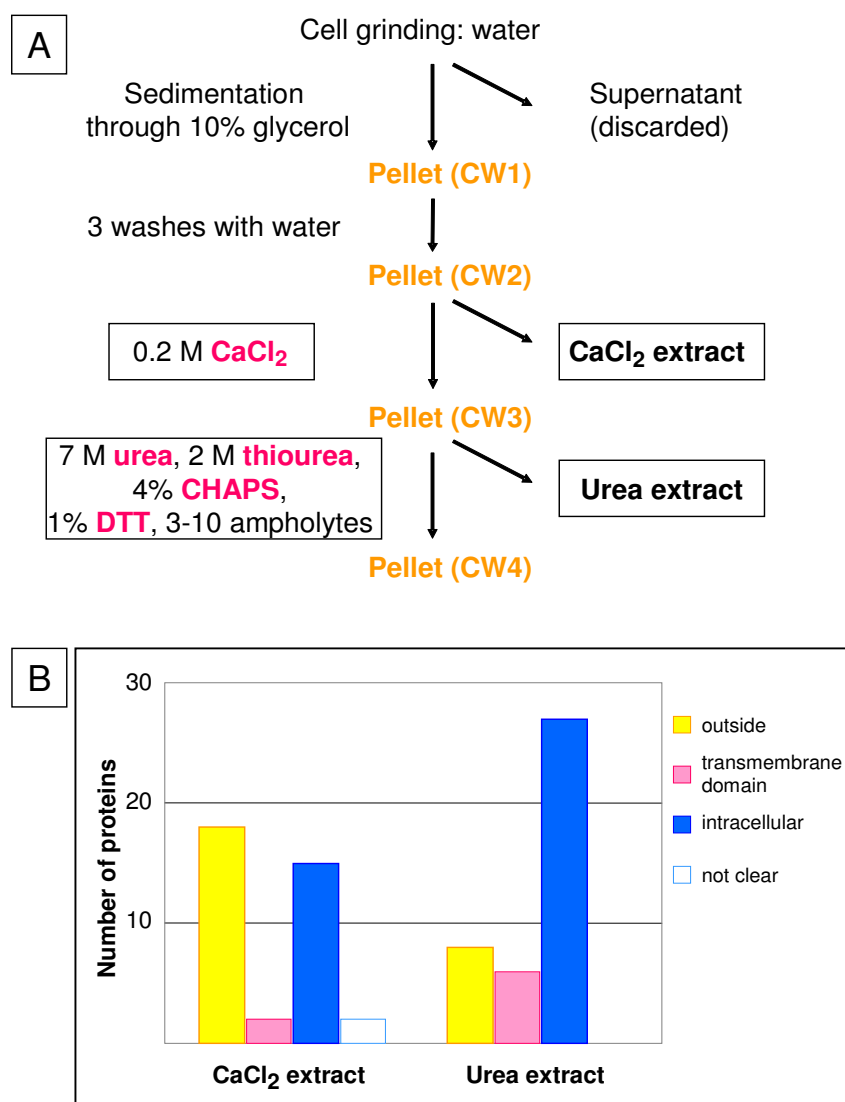
Several strategies have been designed to gain access to CWP. The most labile CWP, *i.e.* those having little or no interactions with cell wall components, can be recovered in culture media of cell suspension cultures [12] or liquid cultured seedlings [14]. Extracellular fluids can be harvested from cell suspension cultures [12,16] or intact organs such as leaves [13]. However, such analysis cannot be done in all cases. It is then necessary to isolate cell walls starting with a drastic mechanical disruption of the material of interest. Consequently, labile CWP may be lost, and intracellular proteins or organelle fragments may contaminate cell wall preparations.

To design a procedure for cell wall isolation and subsequent protein extraction, several general features should be kept in mind. Plant and fungal cell walls are mainly built up with highly dense polysaccharides. This property can be used to purify them through density gradients by centrifugation. The biochemical structure of walls is complex, and CWP can be bound to the matrix by Van der Waals forces, hydrogen bonds, and hydrophobic or ionic attractions. Such interactions can also be modulated by the composition of the isolation medium. Commonly, a low ionic strength is preferred to preserve ionic bonds, but also to dilute the ionic strength of the cell wall itself. An acidic pH is chosen to maintain the interactions between proteins and polysaccharides as *in planta*. Once isolated, cell walls are classically treated by  $\text{CaCl}_2$  buffers to release proteins, and by  $\text{LiCl}$  buffers for extraction of glycoprotein [17,18]. The use of detergents has also been reported to extract proteins strongly embedded in the polysaccharide matrix, like wall associated kinases [19]. Finally, CWP can be covalently bound to cell wall components so that they are resistant to salt-extraction. At present, there is no satisfactory procedure to isolate them. We analyze recent publications using different methods to isolate cell walls from plants or yeast prior to proteomic analysis [8,9,20].

### Analysis of early methods

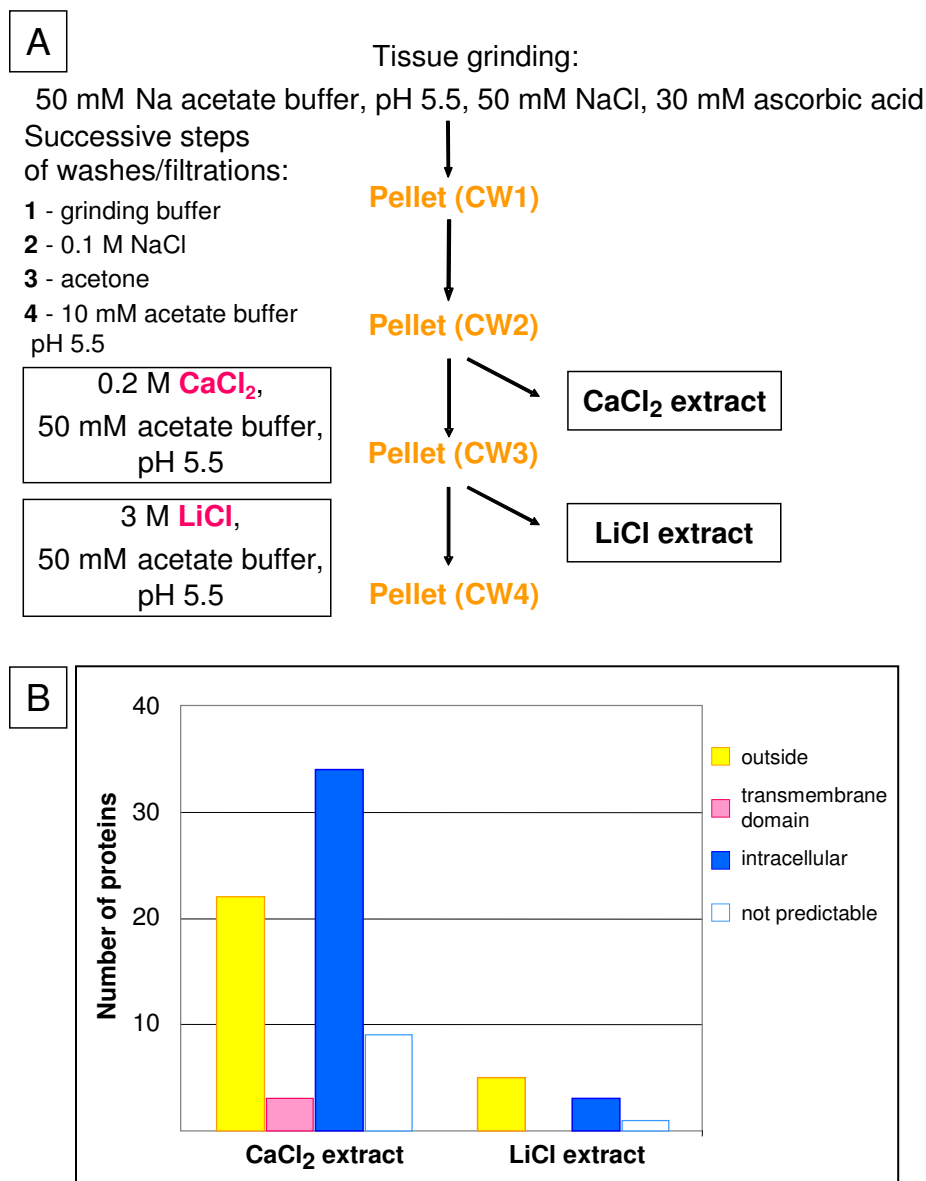
Chivasa et al. [8] used *A. thaliana* cell cultures to purify cell walls. The procedure is schematically represented in Figure 1A. The first step consisted in cell grinding in water. The homogenate was layered onto 10% glycerol and let to sediment for several hours. The cell wall pellet was resuspended in water and washed 3 times by repeated centrifugations. The proteins were sequentially extracted with 0.2 M  $\text{CaCl}_2$ , and urea buffer (7 M urea, 2 M thiourea, 4% CHAPS, 1% DTT, 2% Pharmalytes 3–10). The extracted proteins were separated by 2D-GE and identified by MALDI-TOF mass spectrometry (MS). The identified proteins were analyzed with several bioinformatic programs

## *A. thaliana* cell suspension culture



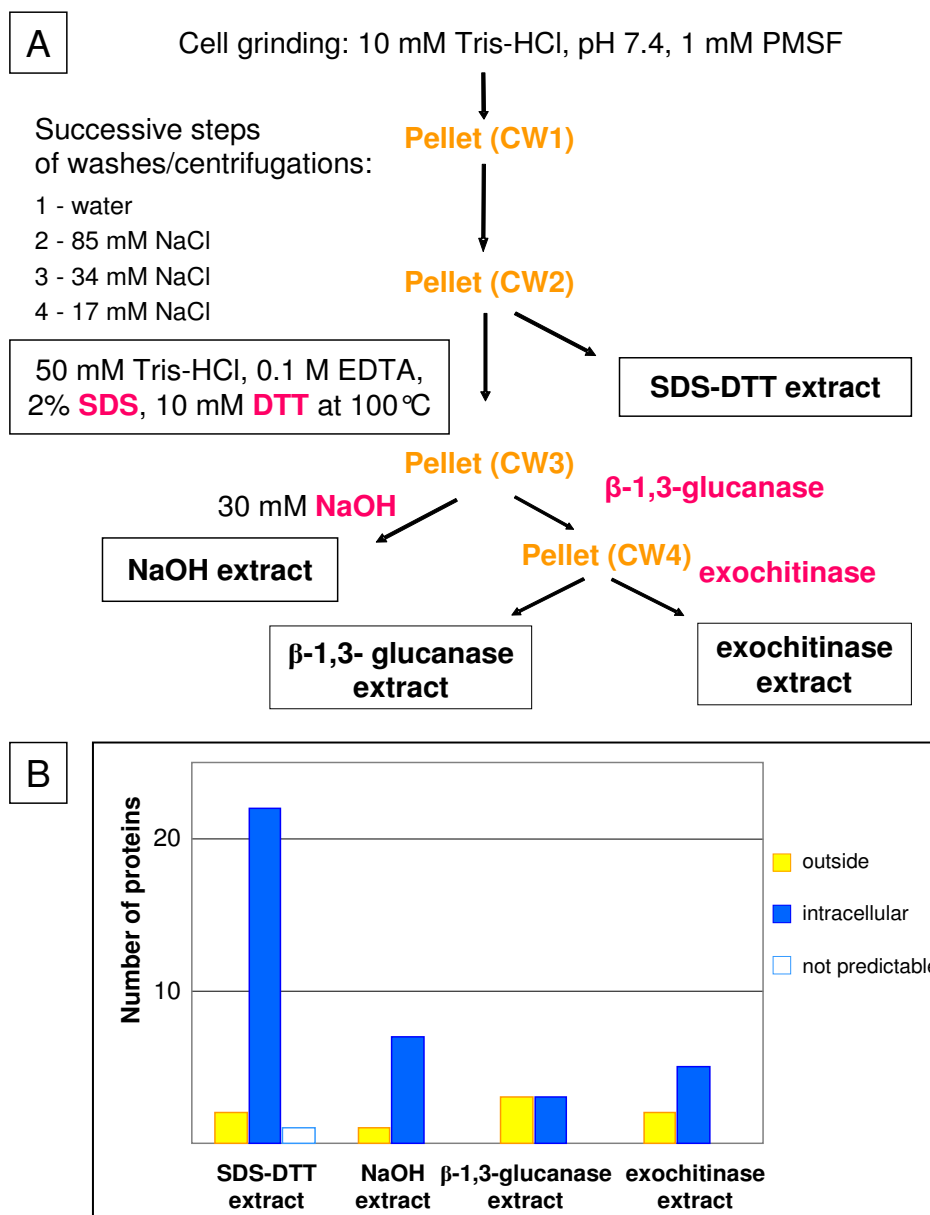
### Figure 1

**Cell wall preparation from *A. thaliana* cell cultures.** **A** – Schematic representation of the purification of cell walls from *A. thaliana* cell suspension cultures, and of the different extracts obtained [8]. **B** – Number of proteins identified in each extract after separation by 2D-GE and MALDI-TOF MS analysis. After bioinformatic analysis, proteins were classified as outside (proteins containing a signal peptide and no other targeting sequences), having at least one trans-membrane domain and intracellular (predicted to be located in any intracellular compartment). Proteins for which predictions by different bioinformatic programs are in conflict are classified as "not clear". Twenty-four different proteins predicted to be secreted were identified in this study (see Additional file 1).

*M. sativa* stems**Figure 2**

**Cell wall preparation from *M. sativa* stems.** **A** – Schematic representation of the purification procedure of cell walls from *M. sativa* stems, and of the different extracts obtained [9]. **B** – Number of proteins identified in each extract after 2D-GE separation and LC-MS/MS analysis. Proteins were classified as indicated in legend to Figure 1. Since the *M. sativa* genome is not fully sequenced, the sequence of the N-terminus of some proteins is not known. They were classified as "not predictable". Twenty-five different proteins predicted to be secreted were identified in this study (see Additional file 2).

## *C. albicans* yeast and hyphae



**Figure 3**

**Preparation of cell walls from *C. albicans* (yeast and hyphae).** **A** – Schematic representation of the purification procedure of cell walls from the dimorphic fungus *C. albicans* (yeast and hyphae), and of the extracts obtained [20]. **B** – Number of proteins identified in each extract after 2D-GE and analysis by MALDI-TOF or MALDI-TOF/TOF MS. Proteins were classified as indicated in legend to Figure 1. Four different proteins predicted to be secreted were identified in this study (see Additional file 3).

and classified as: (i) outside (proteins predicted to be secreted since they contain a signal peptide and no other targeting sequence), (ii) having trans-membrane domain(s), or (iii) intracellular (proteins not fulfilling these criteria). Figure 1B represents the results of both extractions. It appears that the  $\text{CaCl}_2$  extract contains the highest proportion of proteins predicted to be secreted (50%), and that the use of detergents and chaotropic agents brings out mostly intracellular proteins, even if 20% of this fraction corresponds to predicted secreted proteins. Twenty-four different proteins predicted to be secreted were identified with this method.

Stems of *Medicago sativa* (alfalfa) were used for cell wall protein profiling [9]. In this case, a different procedure was used to isolate cell walls (Figure 2A). Tissues were frozen and ground in cold grinding buffer (50 mM Na acetate buffer, pH 5.5, 50 mM NaCl, and 30 mM ascorbic acid) with PVPP. Cell walls were isolated by filtering through a  $47 \mu\text{m}^2$  mesh nylon membrane and washed sequentially with grinding buffer, 0.1 M NaCl, acetone and 10 mM Na acetate pH 5.5. The proteins were sequentially extracted with 0.2 M  $\text{CaCl}_2$ , and 3 M LiCl buffers. The data obtained in this publication was analyzed as mentioned above, and the results are presented in Figure 2B. The proportion of intracellular proteins in the  $\text{CaCl}_2$  extract is quite high (50%). It seems that the first washes do not eliminate such proteins. It is also possible that the wash performed with 0.1 M NaCl eliminates part of the secreted proteins. Twenty-five different proteins predicted to be secreted were identified in this study.

The procedure for isolation of cell walls from the dimorphic fungus *Candida albicans* used by Pitarch et al [20] was based on previous methods designed to isolate proteins covalently linked to the polysaccharide matrix [21,22]. Yeast and hyphae were collected by centrifugation and filtration, washed several times with lysis buffer (10 mM Tris-HCl, pH 7.4, 1 mM PMSF), and mechanically disrupted in lysis buffer (Figure 3A). After centrifugation, the pellet was successively washed with cold water and decreasing concentrations of NaCl (85, 34 and 17 mM) in 1 mM PMSF. Proteins were extracted with boiling SDS-extraction buffer (50 mM Tris-HCl, pH 8.0, 0.1 M EDTA, 2% SDS, 10 mM DTT). The residue was separated in two fractions, one was extracted with alkali (30 mM NaOH), and the other was submitted to sequential digestions by a  $\beta$ -1,3-glucanase followed by an exochitinase to break down the polysaccharide matrix. Each of the four samples were separated by 2D-GE, digested with trypsin and the peptides identified by MALDI-TOF or MALDI-TOF/TOF MS. The proteins identified in this publication were submitted to bioinformatic analysis and the results are represented in Figure 3B. Only four proteins predicted to be

secreted were identified. All the others are predicted to be intracellular proteins (78%).

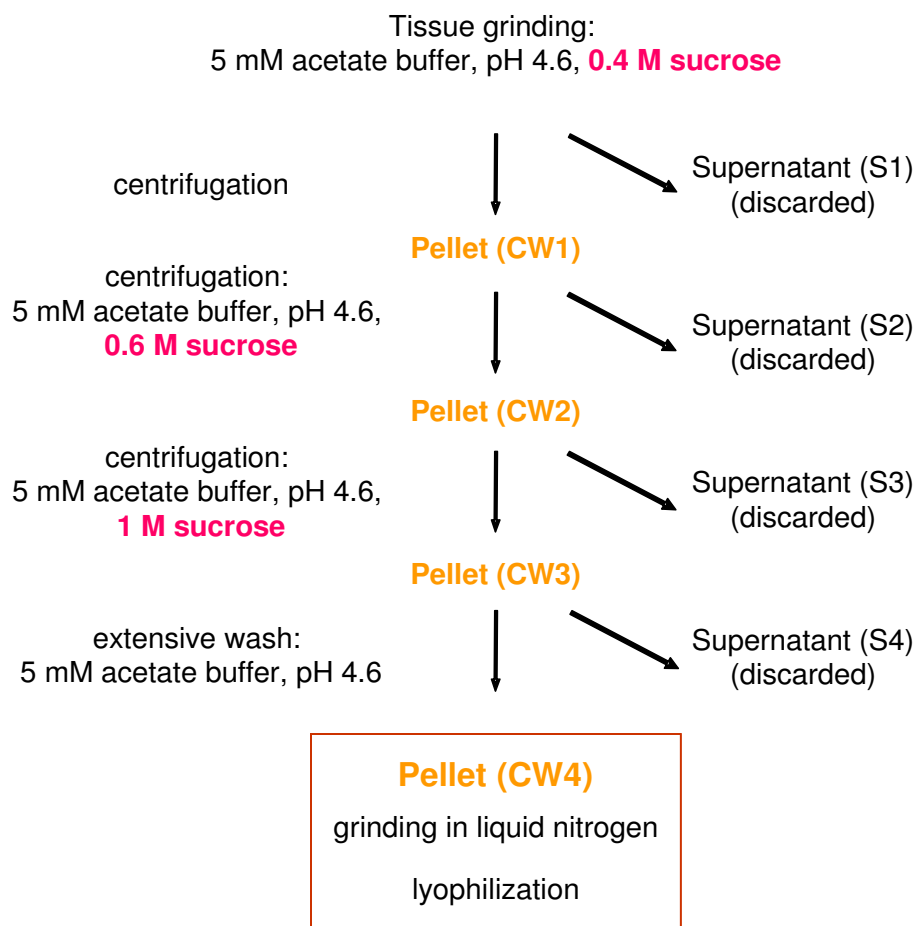
Altogether, this evaluation of three procedures to isolate cell walls from plants or fungi prior to proteomic analyses shows that they all produce a material containing a high proportion of proteins predicted to be intracellular, suggesting they are contaminants. Even if a careful bioinformatic analysis allows the discrimination between secreted and intracellular proteins, the time and effort consumed is not satisfactory.

#### **A modified method to prepare plant cell walls**

From the analysis of the presented methods and other classical cell wall preparations used for the purification of cell wall enzymes [23,24], several points appear to be essential for the purification of cell walls. First, the presence of NaCl at early steps of cell wall preparations of *M. sativa* and *C. albicans* in grinding or washing buffers might induce a release of CWP even at a low concentration [12]. This might indirectly increase the proportion of intracellular proteins sticking non-specifically to cell wall polysaccharides. The use of a low ionic strength buffer for tissue grinding and subsequent washes to purify cell walls appears as an interesting alternative to prevent loss of CWP. Second, the protocol used for *A. thaliana* cell suspension cultures is the only one including a purification of cell walls through a dense medium, *i.e.* sedimentation in 10% glycerol. Proteins predicted to be secreted represented 50% of the identified proteins. It seems that a series of sedimentations/centrifugations in solutions of increasing densities would help in eliminating organelles and other vesicles less dense than cell wall polysaccharides [23,24]. Third, despite repeated washes, the cell wall preparation from *A. thaliana* cell suspension cultures still contained a high proportion of proteins predicted to be intracellular. A way to eliminate soluble contaminants such as intracellular proteins is to perform extensive washes of cell walls with a low ionic strength buffer [24]. Finally, the addition of polyvinyl polypyrrolidone (PVPP) to trap plant phenolic compounds [14] as well as anti-proteases to limit protein degradation during the manipulations [13], improves the quality of protein identification by mass spectrometry. Our procedure was established on the basis of these requirements.

Eleven day-old etiolated hypocotyls were ground in a low ionic strength buffer, 5 mM acetate buffer pH 4.6 in 0.4 M sucrose (Figure 4). PVPP and anti-proteases were added to the homogenate, centrifuged, and the resulting pellet resuspended in 5 mM acetate buffer pH 4.6 with increasing concentrations of sucrose (0.6 M and 1 M) and centrifuged. The residue (CW3) was extensively washed on a nylon membrane (25  $\mu\text{m}$  pore size) with large amounts of 5 mM acetate buffer pH 4.6 (3 L for 16 g fresh material).

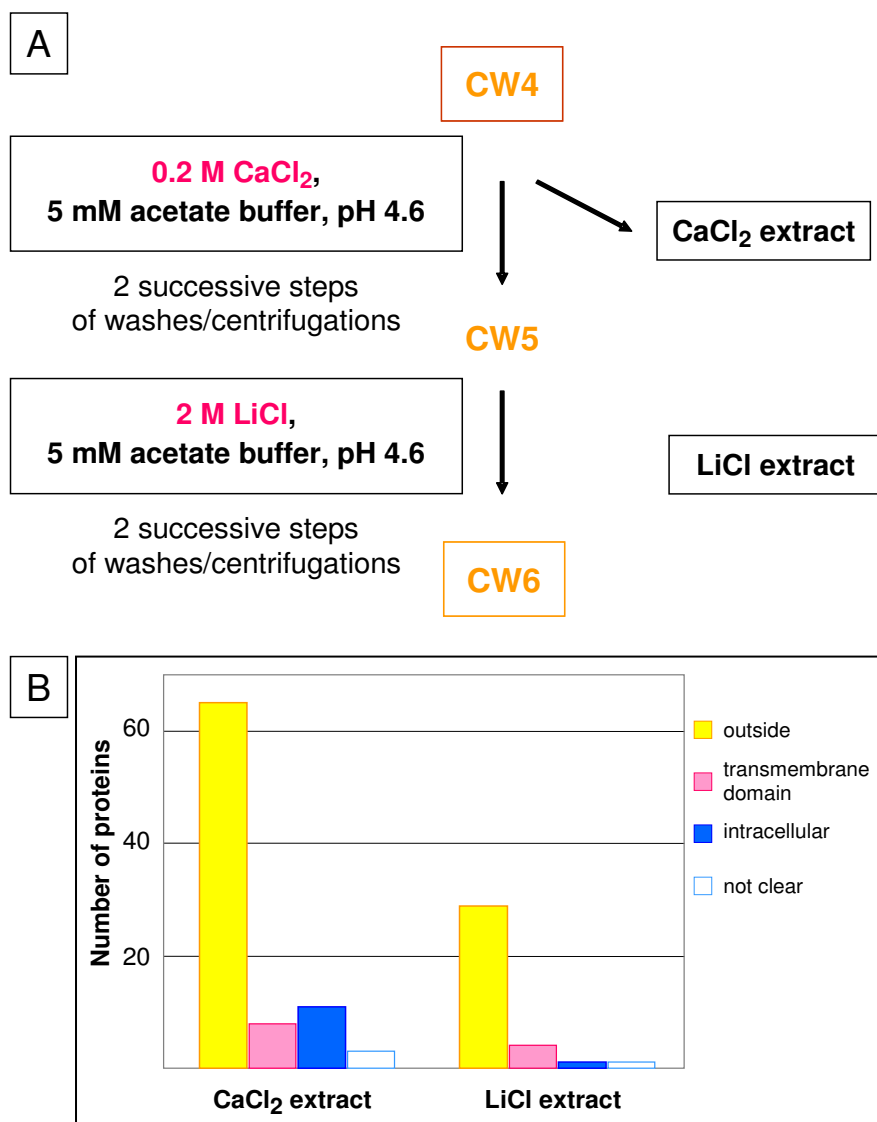
## *A. thaliana* hypocotyls: cell wall preparation



**Figure 4**

**Cell wall preparation from *A. thaliana* hypocotyls.** Schematic representation of the purification procedure of cell walls of *A. thaliana* etiolated hypocotyls. Sixteen g of fresh etiolated hypocotyls were ground. All supernatants were discarded after each centrifugation. The CW3 residue was extensively washed on a nylon net (25  $\mu$ m pore size) with 3 L of 5 mM acetate buffer, pH 4.6. The CW4 pellet was ground in liquid nitrogen in a mortar with a pestle in order to reduce the size of the fragments, and lyophilized, obtaining 1.3 g of dry powder.

## *A. thaliana* hypocotyl cell walls: extraction of proteins with salts

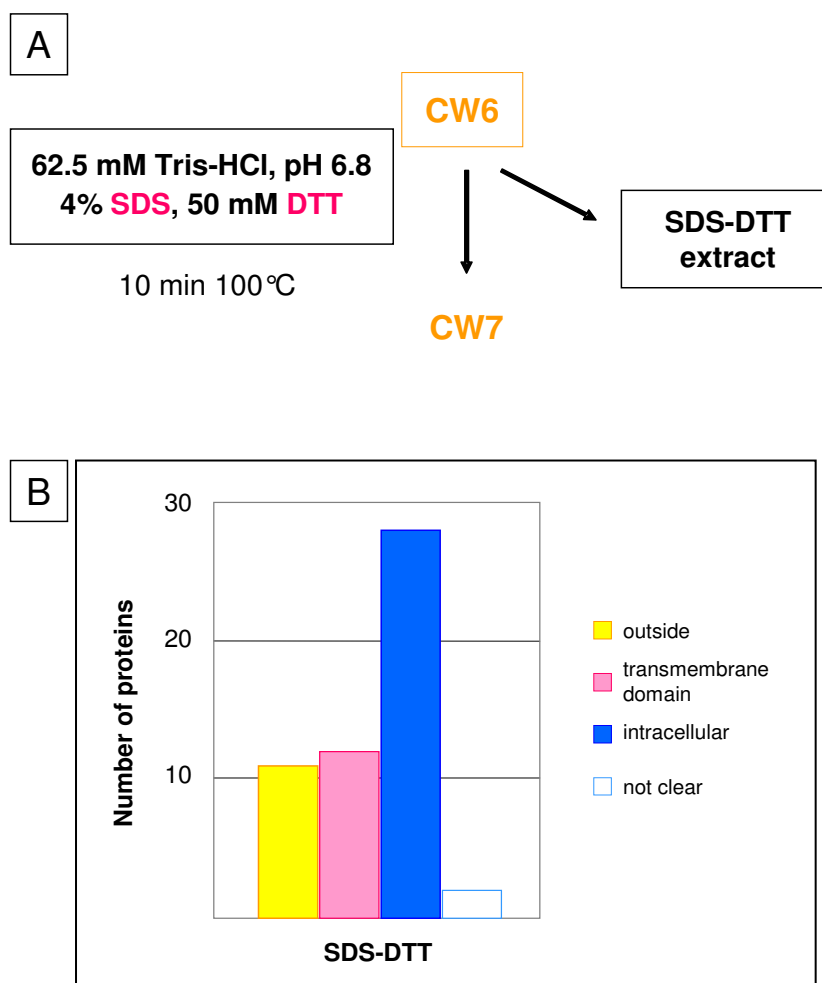


### Figure 5

**Extraction of proteins from *A. thaliana* hypocotyls with salts.** **A** – Half of the CW4 lyophilized powder (0.65 g) was successively suspended in 0.2 M CaCl<sub>2</sub>, 5 mM acetate buffer pH 4.6, and in 2 M LiCl, 5 mM acetate buffer pH 4.6. The CaCl<sub>2</sub> extract contained 400 µg of proteins. The LiCl extract contained 40 µg of proteins. **B** – Number of proteins identified in each extract after 1D-GE separation, and analysis by MALDI-TOF MS or LC-MS/MS. Proteins were classified as indicated in legend to Figure 1. Seventy-three different proteins predicted to be secreted were identified in this study (see Additional file 4).



## *A. thaliana* hypocotyl cell walls: extraction of proteins with boiling SDS and DTT



### Figure 6

**Extraction of proteins from *A. thaliana* hypocotyls cell walls with boiling SDS and DTT.** **A** – The CW6 pellet described in Figure 5 was used for further extraction of proteins using boiling 4% SDS and 50 mM DTT in 62.5 mM Tris-HCl buffer, pH 6.8. The SDS-DTT extract contained 10 µg of proteins. **B** – Number of proteins identified in each extract after ID-GE separation, and analysis by MALDI-TOF MS or LC-MS/MS. Proteins were classified as indicated in legend to Figure 1 (see Additional file 5).

The cell wall fraction was ground in liquid nitrogen in a mortar to reduce the size of the particles, which improves the subsequent extraction of proteins (CW4). In a typical cell wall preparation from *A. thaliana* hypocotyls, 16 g of fresh material resulted in 1.3 g of lyophilized powder.

#### Sequential salt extraction of proteins from cell walls

The lyophilized powder (CW4) was sequentially extracted with salts that can extract proteins ionically bound to the polysaccharide matrix. Calcium chloride has been reported as an efficient salt for the extraction of cell wall proteins [8,9,13], and was used in this procedure (Figure 5A) to identify 65 secreted proteins (see Additional file 4). From 0.65 g of lyophilized powder, 400 µg of proteins were routinely obtained. LiCl was known as potent salt to extract hydroxyproline-rich glycoproteins from *Chlamydomonas reinhardtii* cells, and was successfully used on *M. sativa* cell walls and *A. thaliana* rosettes [9,13,18]. LiCl only extracted 40 µg of proteins, 29 were identified as predicted to be secreted, but only 8 were specifically extracted by this salt (see Additional file 4).

A recent review on 281 CWP identified in proteomic studies by mass spectrometry [7] concluded that more than 60% of them have a basic pI, and around 80% in salt-extractable fractions. This is a serious problem for the separation of CWP by the classical 2D-GE, since it is well known that basic glycoproteins are poorly resolved by this technique [7]. We have then used 1D-GE for the separation, each band stained with Coomassie™ blue was digested by trypsin and further analyzed by MALDI-TOF MS or LC/MS/MS. Each protein was analyzed using several bioinformatic programs as described above. Seventy-three different proteins predicted to be secreted were identified in this study, whereas only 12 proteins predicted to be intracellular and 11 proteins predicted to have trans-membrane domains were found.

The protocol of CWP extraction we used is almost the same as the one employed with *M. sativa* stems [9]. But comparison of Figures 2B and 5B shows big differences in the proportion of proteins predicted to be intracellular or having trans-membrane domains (50% for *M. sativa* vs 27% for *A. thaliana*). Since the main difference between the two protocols is the addition of centrifugations through a dense medium, it shows that this step is critical for the purity of cell walls.

The proteins predicted to be secreted identified in this study belong to the same functional classes as those described in previous cell wall proteomes established from cell wall preparations [8,9,25]. Shortly, these functional classes comprise proteins acting on polysaccharides (e.g. glycoside hydrolases, carbohydrate esterases, expansins), proteases, proteins with interacting domains

(e.g. lectins, leucine-rich repeat proteins, enzyme inhibitors), oxido-reductases (e.g. peroxidases, berberine-bridge enzymes), proteins involved in signaling processes (e.g. arabinogalactan proteins), structural proteins, proteins of unknown function and miscellaneous proteins [7]. The new protocol appears to be more efficient since a large proportion of identified proteins are predicted to be targeted to the compartment of interest.

Since we noticed that the use of detergents described in previous protocols increased the proportion of proteins predicted to be intracellular or having trans-membrane domains [8,20], we wanted to test it on our cell wall preparation. The CW6 pellet (Figure 6) was treated with boiling SDS-DTT buffer (62.5 M Tris-HCl, pH 6.8, 4% SDS, 50 mM DTT). Less than 10 µg of proteins were obtained. As for the other fractions, the proteins were concentrated, separated by 1D-GE, and the Coomassie™ blue stained bands were analyzed by mass spectrometry. Fifty-three proteins were identified in this fraction, among which 11 (20%) were predicted to be secreted, 12 (23%) were predicted to have trans-membrane domains and 30 (57%) were predicted to originate from intracellular compartments including cytoplasm, endoplasmic reticulum, microbodies, and chloroplasts. Comparison of these results with those shown in Figures 1B and 3B, in which detergents such as CHAPS-Urea-DTT or SDS-DTT were also used, confirms that treatment with detergents mainly extract intracellular proteins as well as membrane proteins trapped within the polysaccharide matrix: respectively 66% and 88% of proteins predicted to be intracellular, and 15% of proteins having trans-membrane domain in the case of *A. thaliana* cell walls. Unless looking for specific proteins only extractable in those conditions, this step should be avoided for a large-scale cell wall proteomic study. It can rather be a good method to get rid of contaminants and to have a cleaner preparation for subsequent extraction of strongly bound CWP.

#### Conclusion

The new cell wall preparation procedure followed by salt extraction of proteins described in this paper gives the lowest proportion of proteins predicted to be intracellular when compared with other available protocols and allows the identification of proteins fitting in the same functional classes. Addition of a step including detergent treatment revealed the presence of minor amounts of a few additional proteins predicted to be secreted, but of many proteins predicted to be intracellular. Prediction of the sub-cellular localization of proteins by different bioinformatic programs appeared as an essential tool to evaluate cell wall purification procedures. However, it should not be considered satisfactory to determine the sub-cellular localization of any protein identified by a proteomic analysis. Additional experiments performed *in planta*, such as

immunolocalization or localization of fluorescent protein fusions, are required to confirm it. The application of the principles of cell wall purification described in this paper should lead to a more realistic view of the cell wall proteome, at least for the weakly bound CWP extractable by salts. In addition, it offers a clean cell wall preparation for subsequent extraction of strongly bound CWP.

## Methods

### Plant material and isolation of cell walls

One hundred and fifty mg of *A. thaliana* seeds (ecotype Columbia 0) were sowed per Magenta box containing Murashige and Skoog medium [26] supplemented with 2% w/v sucrose and 1.2% w/v agar. Seedlings were grown at 23°C in the dark for 11 days. For one experiment, hypocotyls from 16 Magenta boxes were collected. One cm long hypocotyls were cut below the cotyledons and above the root, washed with distilled water and transferred into 500 mL of 5 mM acetate buffer, pH 4.6, 0.4 M sucrose and protease inhibitor cocktail (Sigma) 1 mL per 30 g of hypocotyl fresh weight. The mixture was ground in a blender at full speed for 15 min (Figure 4). After adding PVPP (1 g per 10 g fresh weight of hypocotyls), the mixture was incubated in cold room for 30 min while stirring. Cell walls were separated from soluble cytoplasmic fluid by centrifugation of the homogenate for 15 min at 1000 × g and 4°C. The pellet (CW1 in Figure 4) was further purified by two successive centrifugations in 500 mL of 5 mM acetate buffer, pH 4.6, respectively 0.6 M and 1 M sucrose. The residue (CW3) was washed with 3 L of 5 mM acetate buffer, pH 4.6, on a nylon net (25 µm pore size). The resulting cell wall fraction (CW4) was ground in liquid nitrogen in a mortar with a pestle prior to lyophilization. Starting with 16 g fresh weight of hypocotyls, this process resulted in 1.3 g dry powder.

### Sequential proteins extraction and identification

Typically, 0.65 g of lyophilized cell walls was used for one experiment. Proteins were extracted by successive salt solutions in this order: two extractions each time with 6 mL CaCl<sub>2</sub> solution (5 mM acetate buffer, pH 4.6, 0.2 M CaCl<sub>2</sub> and 10 µL protease inhibitor cocktail), followed by two extractions with 6 mL LiCl solution (5 mM acetate buffer, pH 4.6, 2 M LiCl and 10 µL protease inhibitor cocktail). Cell walls were resuspended by vortexing for 5–10 min at room temperature, and then centrifuged for 15 min at 4000 × g and 4°C. Supernatants were desalted using Econo-Pac® 10 DG columns (Bio-Rad) equilibrated with 0.2 formic acid ammonium salt. The extract were lyophilized and resuspended in sample buffer for separation of proteins by 1D-GE, as previously described [12].

The next extraction was carried out by SDS and DTT. The cell wall preparation was treated with 12 mL solution containing 62.5 mM Tris, 4% SDS, 50 mM DTT, pH 6.8

(HCl). The mixture was boiled for 5 min and centrifuged for 15 min at 40000 × g and 4°C. The supernatant was dialyzed against 1 L H<sub>2</sub>O in Spectra/Por® membrane 10 kDa MWCO bags (Spectrum Medical Industries) at room temperature, then concentrated by successive centrifugation using the Centriprep® centrifugal filter devices (YM-10 kDa membrane) (Millipore) at 4000 × g followed by speed vacuum centrifugation.

The protein content of each extract was measured using the Bradford method [27] with the Coomassie™ protein assay reagent kit (Pierce) using bovine serum albumin (BSA) as standard.

Gels were stained with Coomassie™ Brilliant Blue-based method [28]. Colored bands were digested with trypsin and MALDI-TOF MS or LC-MS/MS analyses were performed as previously reported [12,13].

The sequences of the identified proteins were subsequently analyzed with several bioinformatic programs to predict their sub-cellular localization [29–31]. In some cases, predictions were not the same with the three programs. Results are then indicated as "not clear". Data are described in Tables 1–5 (additional data).

### Abbreviations

1D/2D-GE: one or two dimensional gel electrophoresis

CWP: cell wall proteins

CHAPS: 3- [(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid

DTT: dithiothreitol

EDTA: ethylene diamino tetraacetic acid

LC-MS/MS: liquid chromatography-tandem mass spectrometry

MALDI-TOF: matrix-assisted laser desorption ionization-time of flight

MS: mass spectrometry

PMSF: phenylmethylsulfonyl fluoride

PVPP: polyvinyl polypyrrolidone

SDS: sodium dodecyl sulphate

### Competing interests

The author(s) declare that they have no competing interests.

## Authors' contributions

LF and MI carried out cell wall preparations, protein extractions, protein separations and mass spectrometry analyses. HC designed the cell wall preparation procedure and contributed to MALDI-TOF analyses and manuscript writing. EJ conceived data evaluation, performed bioinformatic analyses and was involved in manuscript writing. RPL participated in discussions all along the study, in literature screening, conceived the manuscript, and coordinated its writing. All authors read and approved the final manuscript.

## Additional material

### Additional data file 1

**Table 1** – Bioinformatic analysis of proteins extracted from cell walls of *A. thaliana* cell suspension cultures [8].

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1746-4811-2-10-S1.pdf>]

### Additional data file 2

**Table 2** - Bioinformatic analysis of proteins extracted from cell walls of *M. sativa* stems [9].

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1746-4811-2-10-S2.pdf>]

### Additional data file 3

**Table 3** – Bioinformatic analysis of proteins extracted from cell walls of *C. albicans* [20].

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1746-4811-2-10-S3.pdf>]

### Additional data file 4

**Table 4** - Bioinformatic analysis of proteins extracted from cell walls of *A. thaliana* etiolated hypocotyls with salts.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1746-4811-2-10-S4.pdf>]

### Additional data file 5

**Table 5** - Bioinformatic analysis of proteins extracted from cell walls of *A. thaliana* etiolated hypocotyls with SDS and DTT.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1746-4811-2-10-S5.pdf>]

## Acknowledgements

The authors are grateful to the Université Paul Sabatier (Toulouse, France) and the CNRS (France) for support. M.I. is a fellow from the Higher Education Commission, Islamabad, Pakistan and from the French government on the behalf of SFERE. Mass spectrometry analyses were performed on the Plate-Forme de Spectrométrie de Masse in Toulouse, France.

## References

- Bolwell GP: **Dynamic aspects of the plant extracellular matrix.** *International Review of Cytology* 1993, **146**:261-324.
- Brownlee C: **Role of the extracellular matrix in cell-cell signaling: paracrine paradigms.** *Current Opinion in Cell Biology* 2002, **5**:396-401.
- Carpita N, Gibeaut D: **Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth.** *Plant Journal* 1993, **3**:1-30.
- Pennell R: **Cell walls: structures and signals.** *Current Opinion in Plant Biology* 1998, **1**:504-510.
- Roberts K: **The plant extracellular matrix: in a new expansive mood.** *Current Opinion in Cell Biology* 1994, **6**:688-694.
- Cassab GI, Varner JE: **Cell wall proteins.** *Annual review of Plant Physiology and Plant Molecular Biology* 1988, **39**:321-353.
- Jamet E, Canut H, Boudart G, Pont-Lezica RF: **Cell wall proteins: a new insight through proteomics.** *Trends in Plant Science* 2006, **11**:33-39.
- Chivasa S, Ndimba B, Simon W, Robertson D, Yu X-L, Knox J, Bolwell P, Slabas A: **Proteomic analysis of the *Arabidopsis thaliana* cell wall.** *Electrophoresis* 2002, **23**:1754-1765.
- Watson BS, Lei Z, Dixon RA, Sumner LV: **Proteomics of *Medicago sativa* cell walls.** *Phytochemistry* 2004, **65**:1709-1720.
- Wang SB, Hu Q, Sommerfeld M, Chen F: **Cell wall proteomics of the green alga *Haematococcus pluvialis* (Chlorophyceae).** *Proteomics* 2004, **4**:692-708.
- Pardo M, Ward M, Bains S, Molina M, Blastock W, Gil C, Nombela C: **A proteomic approach for the study of *Saccharomyces cerevisiae* cell wall biogenesis.** *Electrophoresis* 2000, **21**:3396-3410.
- Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerré-Tugayé MT, Boudet A, Pont-Lezica R: **Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: a critical analysis.** *Electrophoresis* 2003, **24**:3421-32.
- Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerré-Tugayé M-T, Pont-Lezica R: **Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics.** *Proteomics* 2005, **5**:212-221.
- Charmont S, Jamet E, Pont-Lezica R, Canut H: **Proteomic analysis of secreted proteins from *Arabidopsis thaliana* seedlings: improved recovery following removal of phenolic compounds.** *Phytochemistry* 2005, **66**:453-461.
- Jamet E: **Bioinformatics as a critical prerequisite to transcriptome and proteome studies.** *Journal of Experimental Botany* 2004, **55**:1977-1979.
- Kwon HK, Yokoyama R, Nishitani K: **A proteomic approach to apoplastic proteins involved in cell wall regeneration in protoplasts of *Arabidopsis* suspension-cultured cells.** *Plant Cell Physiology* 2005, **46**:843-857.
- Melan MA, Cosgrove DJ: **Evidence against the involvement of ionically bound cell wall proteins in pea epicotyl growth.** *Plant Physiology* 1988, **86**:469-74.
- Voigt J, Frank R: **14-3-3 Proteins are constituents of the insoluble glycoprotein framework of the *Chlamydomonas* cell wall.** *The Plant Cell* 2003, **15**:1399-1413.
- He ZE, Fujiki M, Kohorn BD: **A cell wall-associated, receptor-like protein kinase.** *Journal of Biological Chemistry* 1996, **16**:19789-19793.
- Pitarch A, Sanchez M, Nombela C, Gil C: **Sequential fractionation and two-dimensional gel analysis unravel the complexity of the dimorphic fungus *Candida albicans* cell wall proteome.** *Molecular and Cellular Proteomics* 2002, **1**:967-82.
- Mrsa V, Seidi T, Gentzsch M, Tanner VV: **Specific labelling of cell wall proteins by biotinylation, identification of four covalently linked O-mannosylated proteins of *Saccharomyces cerevisiae*.** *Yeast* 1997, **13**:1145-1154.
- Kapteyn JC, Montijn RC, Vink E, de la Cruz J, Llobel A, Douwes JE, Shimoi H, Lipke PN, Klis FM: **Retention of *Saccharomyces cerevisiae* cell wall proteins through a phosphodiester-linked  $\beta$ -1,3- $\beta$ -1,6-glucan heteropolymer.** *Glycobiology* 1996, **6**:337-345.
- Bordenave M, Goldberg R: **Immobilized and free apoplastic pectinmethylesterases in mung bean hypocotyl.** *Plant Physiology* 1994, **106**:1151-1156.

24. Goldberg R: **On possible connections between auxin induced growth and cell wall glucanase activities.** *Plant Science Letters* 1977, **8**:233-242.
25. Bayer EM, Bottrill AR, Walshaw J, Vigouroux M, Naldrett MJ, Thomas CL, Maule AJ: **Arabidopsis cell wall proteome defined using multidimensional protein identification technology.** *Proteomics* 2006, **6**:301-311.
26. Murashige T, Skoog F: **A revised medium for rapid growth and bioassays with tobacco tissue culture.** *Physiologia Plantarum* 1962, **15**:473-497.
27. Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.** *Analytical Biochemistry* 1976, **72**:248-254.
28. Scheler C, Lamer S, Pan Z, Li XP, Salnikow J, Jangblut P: **Peptide mass fingerprint sequences coverage from differentially stained proteins in two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS).** *Electrophoresis* 1998, **19**:918-927.
29. **PSORT** [<http://psort.nibb.ac.jp/form.html>]
30. **TargetP** [<http://www.cbs.dtu.dk/services/TargetP/>]
31. **Aramemnon** [<http://aramemnon.botanik.uni-koeln.de/>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)



## **CHAPTER 4**

### **A NEW PICTURE OF CELL WALL PROTEIN DYNAMICS IN ELONGATING CELLS OF *ARABIDOPSIS THALIANA*: KNOWN PLAYERS AND NEW COMERS**

---

**Chapter summary (French)**

La croissance des cellules végétales est accompagnée par celle de leurs parois qui nécessite d'une part la synthèse de constituants pariétaux, d'autre part leur ajout dans les parois et le remaniement des réseaux existants. Les protéines pariétales sont importantes pour cette dernière phase. Obtenir une image globale des protéines pariétales dans des hypocotyles en cours d'élongation (plantules âgées de 5 jours) et après la fin de leur croissance (plantules âgées de 11 jours) devrait contribuer à la compréhension des mécanismes de réarrangement des constituants pariétaux ainsi que les processus de régulation de la croissance pariétale.

Après avoir surmonté la première difficulté de ce travail consistant à obtenir des préparations de parois de bonne qualité (chapitre 3), je me suis consacré à l'étude protéomique des parois des hypocotyles en élongation comparée à celle des hypocotyles après la fin de leur croissance. Les protéines pariétales ont été extraites avec du  $\text{CaCl}_2$ , puis du  $\text{LiCl}$  à chaque stade de développement. La difficulté suivante est apparue au niveau de la séparation des protéines. En effet, les méthodes existantes ne se sont pas révélées très efficaces dans la mesure où les protéines pariétales sont majoritairement basiques et glycosylées, donc difficiles à séparer avec la méthode traditionnellement utilisée en protéomique, *i.e.* l'électrophorèse bi-dimensionnelle. Par ailleurs, l'électrophorèse mono-dimensionnelle était un outil insuffisamment performant puisque de nombreuses protéines comigraient et devenaient ainsi difficiles à identifier par cartographie peptidique massique (spectrométrie de masse de type MALDI-TOF). Ce problème a été surmonté en effectuant une séparation des protéines en deux étapes. La première étape a consisté en une chromatographie d'échange de cations à haute performance, utilisant le caractère basique des protéines pariétales. La seconde étape a consisté en une électrophorèse mono-dimensionnelle des fractions recueillies le long du gradient d'élution par une solution saline. Ensuite, les protéines colorées au bleu de Coomassie ont été systématiquement identifiées par cartographie peptidique massique et bioinformatique. C'est ainsi (i) que le nombre de protéines identifiées a été doublé, comparé à ce qui avait été identifié après une simple séparation par électrophorèse mono-dimensionnelle, (ii) que la qualité des identifications a été largement améliorée, et (iii) qu'il a été possible d'effectuer une semi-quantification des protéines.

Cette étude a permis d'identifier 137 protéines pariétales (120 à 5 jours, et 101 à 11 jours) parmi lesquelles 51 n'avaient pas encore été identifiées par protéomique. La prédiction

de domaines fonctionnels par bioinformatique a permis de les classer en neuf groupes : des protéines agissant sur les polysaccharides pariétaux, des oxydo-réductases, des protéases, des protéines ayant des domaines d'interaction avec des polysaccharides ou d'autres protéines, des protéines impliquées dans la signalisation, des protéines structurales, des protéines liées au métabolisme des lipides, diverses protéines et des protéines de fonction encore inconnue.

La comparaison des deux stades de développement a montré des différences de profil. Bien que beaucoup de protéines extracellulaires soient communes (84), 36 et 17 protéines ont été identifiées de manière spécifique à 5 et 11 jours respectivement. De plus, la comparaison des niveaux relatifs des 84 protéines communes montre que 61 et 26 protéines sont plus abondantes à 5 et à 11 jours respectivement. L'ensemble de ces résultats montre que les parois hypocotyles en élongation active (5 jours) semblent plus métaboliquement plus actives que celles des hypocotyles dont la croissance est achevée (11 jours).

L'analyse par bioinformatique des fonctions des protéines identifiées dans les hypocotyles au cours de leur élongation ou après la fin de leur croissance a permis d'obtenir les résultats suivants :

- i) Les protéines agissant sur les polysaccharides (XTH, PG, PME et expansines) sont plus nombreuses dans les hypocotyles en cours d'élongation, ce qui est cohérent avec leurs rôles supposés dans le réarrangement des polysaccharides ou l'insertion de nouveaux polymères dans les parois.
- ii) Cependant, ces protéines ainsi que les peroxidases sont trouvées aux deux stades de développement. Ceci indique soit que ces protéines ont des demi-vies importantes, soit qu'elles pourraient jouer des rôles différents. Par exemple, les XTH pourraient intervenir non seulement pour le réarrangement des xyloglucanes au cours de la phase d'élongation cellulaire, mais aussi au cours de la différenciation d'éléments trachéaires plus tardivement. Les peroxidases pourraient d'une part contribuer à la formation de radicaux libres participant à la fragmentation des polysaccharides au cours de l'élongation cellulaire, et d'autre part assurer la formation des réseaux de protéines et de polyphénols après son arrêt.
- iii) La présence d'inhibiteurs d'enzymes telles que les protéases, PME, PG et xyloglucane endoglucanases (XEGIP) montre que la présence d'une enzyme ne signifie pas nécessairement qu'elle est active. L'interaction avec un inhibiteur peut



---

moduler son activité biologique. De même, 14 protéases ont été trouvées : elles pourraient soit assurer la maturation de protéines pour les rendre actives ou au contraire assurer leur dégradation. L'apport d'une approche protéomique est ici essentiel pour envisager ces types de régulation.

- iv) Six protéines liées au métabolisme des lipides ont été identifiées. Leur rôle n'est pas clairement établi. Cependant, on peut imaginer qu'il s'agit de candidats qui pourraient être impliqués dans la formation de la cuticule particulièrement épaisse des hypocotyles ayant poussé à l'obscurité.
- v) Seize protéines de fonction inconnue ont été identifiées. Ces protéines pourraient jouer des rôles importants qui restent à découvrir au cours de la croissance des hypocotyles.

L'ensemble de cette étude a permis de donner une image globale des protéines présentes dans les parois des hypocotyles étiolés d'*Arabidopsis*. Cette étude permet d'identifier des gènes potentiellement importants pour l'élongation cellulaire. Elle permet aussi d'envisager des mécanismes de régulation de l'activité biologique des protéines dans les parois en rapprochant des partenaires potentiels (enzymes et leurs inhibiteurs, protéases). Des études fonctionnelles fines sont encore nécessaires pour préciser le rôle de ces protéines au cours du développement des hypocotyles. Les résultats sont détaillés dans l'article qui suit, actuellement soumis pour publication.

In plants cell, elongation and plant overall growth are accompanied by the synthesis, addition and re-arrangement of cell wall components. CWPs are important for proper assembly and reorganization of cell wall components. A global picture of the CWPs of elongating (5-day-old) and fully elongated (11-day-old) hypocotyls may help to uncover their interactions with other cell wall components and to understand the regulatory mechanisms of cell elongation.

After overcoming the first problem of obtaining good quality cell wall (chapter 3), I focused on comparative proteomic study of cell wall of etiolated hypocotyls during and after elongation. CWPs were extracted with  $\text{CaCl}_2$  and  $\text{LiCl}$  from the cell wall purified from hypocotyls of both stages according to the method described in chapter 3. Good quality resolution of CWPs was a problem:

- i) the CWP extract resolved with one dimensional electrophoresis (1D-E), more than one protein were identified from the same band of CBB-stained polyacrylamide gel because of the complexity of the sample with several proteins of same molecular weight.
- ii) 2D-E is not a suitable resolving tool for CWPs because most of them are glycosylated and have basic PIs.

This problem was overcome by first separating CWP extract by cation exchange chromatography using Fast Protein Liquid Chromatography (FPLC), followed by 1D-E of the salt eluted fractions. The profile of the proteins on CBB-stained polyacrylamide gel and their identification by mass spectrometry and bioinformatics justified the use of FPLC followed by 1D-E as a powerful alternative to 2D-E for CWPs. Use of this technique (i) doubled the number of identified proteins as compared to 1D-E alone, (ii) gave higher quality identification by higher numbers of peptides for identification and repetition of identification in succeeding fractions, and (iii) allowed semi-quantification of proteins.

Overall 137 CWPs (120 and 101 in 5- and 11-day-old respectively) were identified in this study, among which 51 are reported for the first time in a proteomic study. The bioinformatics prediction of functional domains allowed their classification into nine categories: proteins acting on carbohydrates, oxido-reductases, proteins with interaction domains, proteases, structural proteins, proteins involved in signaling, proteins related to lipid metabolism, proteins with miscellaneous functions, and proteins of unknown function.

---

Comparison of the two physiological stages revealed differences regarding presence/absence or abundance of several proteins in one or the other samples. Although most CWP (84) were common to both stages, 36 and 17 were identified only in 5- and 11-day-old hypocotyls respectively. The comparison of relative amount showed 61 and 26 proteins are more abundant at 5- and 11-day-old stages respectively. These results suggest that the cell wall of 5-day-old hypocotyls undergoing active elongation is metabolically more active than at 11-day-old hypocotyls that are fully grown.

The main results obtained from the bioinformatic analysis of the functions of the proteins identified in etiolated hypocotyls during and after elongation are listed below:

- i) Proteins expected to be acting on carbohydrates (XTHs, PGs, PMEs, and expansins) are numerous in elongating hypocotyls where they can be involved in the rearrangement of wall polysaccharides or integration of the new wall polymers.
- ii) However such proteins involved in elongation (XTHs, expansins, PGs PMEs and peroxidases) are found in both stages. This gives clues about long half-lives of these proteins or their involvement in other processes. For example the XTHs could be involved in rearrangement of xyloglucans not only during elongation, but also latter during treachery elements differentiation. The peroxidases produce free radicals that can break cell wall polysaccharides during elongation or cross-link structural proteins and polyphenols after elongation.
- iii) It seems that the proteases found at both stages regulate these proteins by proteolytic degradation.
- iv) Presence of Inhibitors of enzymes like proteases, PME, PGs and xyloglucan endoglucanases (XEGIPs) suggests that the presence of a protein does not mean its active functioning. Thus the interaction of an inhibitor may regulate the biological activity of the enzyme. Fourteen proteases were also found that may be involved in the maturation of proteins or in their degradation. The contribution of proteomic approach is therefore essential to consider such regulations.
- v) The function of proteins related to lipid metabolism is not clear. Six members of this group identified during this study might be involved in the formation of thick cuticle, characteristic of etiolated hypocotyls of *A. thaliana*.
- vi) Sixteen proteins of unknown function were found in both stages that could perform important roles in hypocotyls thus requiring special attention.

This study gives a global picture of proteins present in the cell wall of etiolated hypocotyls of *A. thaliana* and helped to identify genes potentially important for cell elongation. By identifying the proteases and enzyme inhibitors, it pointed out the mechanisms they may regulate the biological activity of proteins in the cell wall. Despite all these, functional characterization of these proteins is needed to know their exact roles during hypocotyle development. The detailed results are given in the following article, recently submitted for publication.

The supplementary data files are given as annexes at the end of the report.

---

**A new picture of cell wall protein dynamics in elongating cells of*****Arabidopsis thaliana*: known players and new comers**

Muhammad IRSHAD, Hervé CANUT, Gisèle BORDERIES, Rafael PONT-LEZICA,

Elisabeth JAMET

UMR 5546 CNRS-Université Toulouse III, Pôle de Biotechnologie Végétale, 24 chemin de  
Borde-Rouge, BP 42617 Auzeville, 31326 Castanet-Tolosan, France

**Running title:** Cell wall proteome of *Arabidopsis* hypocotyls

**Corresponding author:** Elisabeth Jamet

BP 42617, Auzeville, F-31326 Castanet-Tolosan

Phone: 33+ 5 62 19 35 30 - FAX: 33+ 5 32 19 35 02

E-mail: jamet@scsv.ups-tlse.fr

Word count

Total: 6934

Summary: 200

Introduction: 473

Material and methods: 881

Results: 1638

Discussion: 1771

Acknowledgments: 66

References: 1485

Figure legends: 420

Tables: 2

Figures: 4

---

**Summary**

- To figure out the role of cell wall proteins (CWPs) in the addition and rearrangement of plant cell wall components during cell elongation, the identification, quantification, and comparison of the secreted proteins of *Arabidopsis thaliana* half- (active growth) and fully-grown (after growth arrest) etiolated hypocotyls was performed.
- A comparative proteomic study was performed at the two developmental stages of hypocotyls starting from purified cell walls and using a new strategy for CWP separation including high performance cation exchange chromatography and mono-dimensional electrophoresis
- In total, 137 proteins predicted to be secreted were identified, among which 51 were not previously found by proteomic approaches. Expected CWPs known to be involved in cell wall extension such as XTHs, expansins, PGs, PMEAs and peroxidases were identified as well as new CWPs such as proteases, proteins related to lipid metabolism and proteins of unknown function.
- This work highlights a great dynamics of CWPs between the two developmental stages. The presence of proteins well-known to be related to cell wall extension after growth arrest showed that they might play other functions in cell walls. From this global view of cell wall proteins, putative regulatory mechanisms of protein biological activity are discussed.

---

**Keywords:** protein dynamics, cell elongation, *Arabidopsis thaliana*, cell wall, etiolated hypocotyl, proteome.

### Abbreviations

AGP: arabinogalactan protein; CBB: Coomassie brilliant blue; CWP: cell wall protein; 1D-E: mono-dimensional electrophoresis; 2D-E: two-dimensional electrophoresis; FLA: fasciclin arabinogalactan protein; GH: glycoside hydrolase; GRP: glycine-rich protein; LAE: late-abundant embryogenesis protein; LTP: lipid transfer protein; LRR: leucine-rich repeat; LRX: leucine-rich repeat extensin; MALDI-TOF: matrix-assisted laser desorption ionization – time of flight; PG: polygalacturonase; PL: polysaccharide lyase; PME: pectin methylesterase; PTM: post-translational modification; XTH: xyloglucan endotransglucosylase-hydrolase.

---

## Introduction

Plant cell walls are dynamic compartments which composition and structure vary during development and response to environmental stresses. Variability has been observed in developing roots at the level of glycoproteins in carrot (Smallwood *et al.*, 1994), and of polysaccharides in *Arabidopsis thaliana* (Freshour *et al.*, 1996). In both articles, antibodies against specific epitopes showed irregular distribution between cell types, as well as changes during development. To characterize cell wall plasticity, elongation is a particularly appropriate process in which cell walls need to enlarge and to be re-organized. The cell machinery is almost completely dedicated to synthesize, export, and reorganize cell walls that are mainly composed of polysaccharide networks (Cosgrove, 2005).

To gain information on the genes and derived proteins involved in cell wall elongation, it is important to dissect the different phases of this process. *Arabidopsis* hypocotyls are a suitable material, since almost no cell division occurs, and only the cells present in the embryo undergo elongation (Gendreau *et al.*, 1997; Saibo *et al.*, 2003; Refrégier *et al.*, 2004). Another advantage is that in etiolated hypocotyls, synthesis, addition, and reorganization of the cell wall material occur in time-separate phases (Gendreau *et al.*, 1997; Derbyshire *et al.*, 2007). During the first 3 days after germination, synthesis and deposition of cell wall material are the main features, producing cells with thick cell walls. Through the following days, the hypocotyl will grow mainly by extensive polymer disassembly and rearrangement.

Understanding the molecular mechanisms responsible for rearrangement of cell wall polymers during hypocotyl growth requires the identification of proteins present *in muro*. Previous studies identified gene families involved in rearrangement of cell wall components during cell elongation (Nicol and Hofte, 1998; Micheli, 2001; Fry, 2004; Passardi *et al.*, 2004). On the other hand, plant cell wall proteomics has emerged a few years ago providing information on cell wall proteins (CWPs) present in many different types of cells including cell suspension cultures, roots, rosette leaves, and stems (Jamet *et al.*, 2008; Lee and Saravanan, 2004). In most cases, limitations were encountered, *e.g.* presence of intracellular contaminants (Feiz *et al.*, 2006), poor quality of separation of CWPs by two-dimensional electrophoresis (2D-E) (Minic *et al.*, 2007). Based on bioinformatic prediction of functional domains, main families of CWPs were identified (Jamet *et al.*, 2006).



In this paper, we analyzed the cell wall proteome of half- (active elongation) and fully-grown (after growth arrest) etiolated hypocotyls identifying the proteins present at both stages. A new strategy has been established for CWP separation, and semi-quantification. The comparison of these two proteomes revealed significant dynamics of CWPs. The presence of CWPs expected to be involved in polysaccharide rearrangement and modification was confirmed in growing hypocotyls. However, some of them were also present in fully-grown hypocotyls, suggesting other functions. Finally, the presence of unexpected proteins suggested that unknown features of cell elongation should be investigated.

---

## Material and Methods

### *Plant material*

One hundred and thirty mg of *Arabidopsis thaliana* seeds (ecotype Columbia 0) were cultivated in Magenta box as described (Feiz *et al.*, 2006). Magenta boxes were maintained at 4°C in the dark during 48 h and subsequently exposed for 4 h to light to synchronize beginning of germination. Finally, seedlings were grown at 23°C in the dark for 5- or 11-days. Hypocotyls were cut with scissors below the cotyledons and above the roots of seedlings. Typically, 36 and 18 Magenta boxes were required for 5- and 11-day-old seedlings respectively.

### *Cell wall purification and protein extraction*

Cell walls were purified as previously described (Feiz *et al.*, 2006). Proteins were extracted from purified cell walls with two successive extractions with CaCl<sub>2</sub> solution (5 mM sodium acetate buffer, pH 4.6, 0.2 M CaCl<sub>2</sub> and 10 µL protease inhibitor cocktail, Sigma), followed by two extractions with LiCl solution (5 mM sodium acetate buffer, pH 4.6, 2 M LiCl and 10 µL protease inhibitor cocktail). Finally, proteins were desalted and lyophilized.

### *Protein separation by cationic exchange chromatography*

Lyophilized proteins were dissolved in a total volume of 2 mL of water. They were quantified with the Coomassie® protein assay reagent kit (Pierce) taking bovine serum albumin (BSA) as a standard (Bradford, 1976). One mg of proteins was used for chromatographic fractionation on a 1 mL HiTrap™ SP FF column (Amersham Biosciences) equilibrated with 50 mM MES (pH 5.6) operated with an FPLC™ System (Amersham Biosciences) controlled by FPLCdirector™ version 1.0 (Amersham Biosciences). The protein solution was adjusted to 50 mM MES (pH 5.6) and 20 µL protease inhibitor cocktail (Sigma) was added before loading onto the column at a flow rate of 0.5 mL.min<sup>-1</sup>. A 10 mL unfixed fraction was collected at the same rate. Three mL of first wash with 50 mM MES (pH 5.6) were collected at a flow rate of 1 mL.min<sup>-1</sup>. Fixed proteins were eluted by a gradient from 0 to 0.8 M NaCl in 50 mM MES (pH 5.6) and 24 fractions (1 mL each) were collected at a flow rate of 1 mL.min<sup>-1</sup>. Finally the column was successively washed with 3 mL of 1.2 M NaCl and 3 mL of 1.5 M NaCl in 50 mM MES (pH 5.6) at the same flow rate. These washes were also collected as 6 fractions (1 mL per tube). To prevent protein degradation, 2 µL of protease inhibitor cocktail (Sigma) were added to all the 1 mL fractions, 6 µL to the first wash and 20

μL to the unfixed fractions. The fractions were combined by 2 or 3 depending on their protein concentration. Unfixed proteins were concentrated by successive centrifugations using the Centriprep® centrifugal filter device (YM-10 kDa membrane for volumes greater than 6 mL or 5 kDa for smaller volumes, Millipore) at  $4000 \times g$ . All protein fractions were desalted using Econo-Pac® 10DG columns (Bio-Rad) equilibrated with 0.2 M ammonium formate prior to lyophilization

#### *Protein separation by mono-dimensional electrophoresis (1D-E) and identification*

Each lyophilized fraction was dissolved in 200 μL water and electrophoresis of proteins was performed according to Laemmli (Laemmli, 1970). Samples were loaded on  $22 \times 15$  cm SDS-polyacrylamide gel with a concentration of 12.5% of acrylamide and a thickness of 1.5 mm. The staining was carried out by a Coomassie Brilliant Blue (CBB)-based method (Scheler *et al.*, 1998). Colored bands were excised from gels and digested with trypsin. MALDI-TOF MS analyses were performed as previously reported (Borderies *et al.*, 2003; Boudart *et al.*, 2005). MALDI-TOF-TOF MS analysis was performed using a MALDI TOF-TOF Voyager 4700 (AppliedBiosystems/MDS Sciex, USA). N-terminal sequencing of the protein encoded by *At5g14920* was performed at *Plate-Forme d'Analyse et de Microséquençage des Protéines* at Institut Pasteur (Paris, France).

#### *Semi-quantification*

Peptide mass fingerprints were compared to the non-redundant database of *Arabidopsis* of NCBI (<ftp://ftp.ncbi.nih.gov/blast/db>) using ProteinProspector (MS-FIT: <http://prospector.ucsf.edu/cgi-bin/msform.cgi?form=msfitstandard>). A quantification index (QI) was calculated for each protein. It was calculated by adding the percentages of coverage by peptide mass mapping (ratio between the number of amino acids in peptides detected by MS and the total number of amino acids of the protein) in all the bands of the FPLC profile in which the protein was identified.

#### *Bioinformatic analyses*

Sub-cellular localization, length of signal peptides, prediction of transmembrane domains, homologies to other proteins and protein functional domains were predicted as described (Minic *et al.*, 2007). GHs and CEs were classified according to the CAZy database (<http://www.cazy.org/CAZY/>) (Coutinho and Henrissat, 1999). XTHs and expansins were

---

named according to <http://labs.plantbio.cornell.edu/xth/> and <http://www.bio.psu.edu/expansins/index.htm> respectively. AGPs and FLAs were named according to Schultz *et al.* (Schultz *et al.*, 2002), Johnson *et al.* (Johnson *et al.*, 2003) and van Engels and Roberts (van Hengel and Roberts, 2003). Proteins homologous to COBRA, LRXs and Hyp/Pro-rich proteins were annotated according to Roudier *et al.* (Roudier *et al.*, 2002), Baumberger *et al.* (Baumberger *et al.*, 2003), and Fowler *et al.* (Fowler *et al.*, 1999) respectively. Peroxidases were named as in the PeroxiBase (<http://peroxidase.isb-sib.ch/index.php>) (Bakalovic *et al.*, 2006). Laccases were annotated as in Pourcel *et al.* (Pourcel *et al.*, 2005) and McCaig *et al.* (McCaig *et al.*, 2005). SKU-like proteins and phytocyanins were named according to Jacobs and Roe (Jacobs and Roe, 2005), and Nersissian and Shipp (Nersissian and Shipp, 2002) respectively. Protease annotation was done according to the MEROPS database (<http://merops.sanger.ac.uk/>).

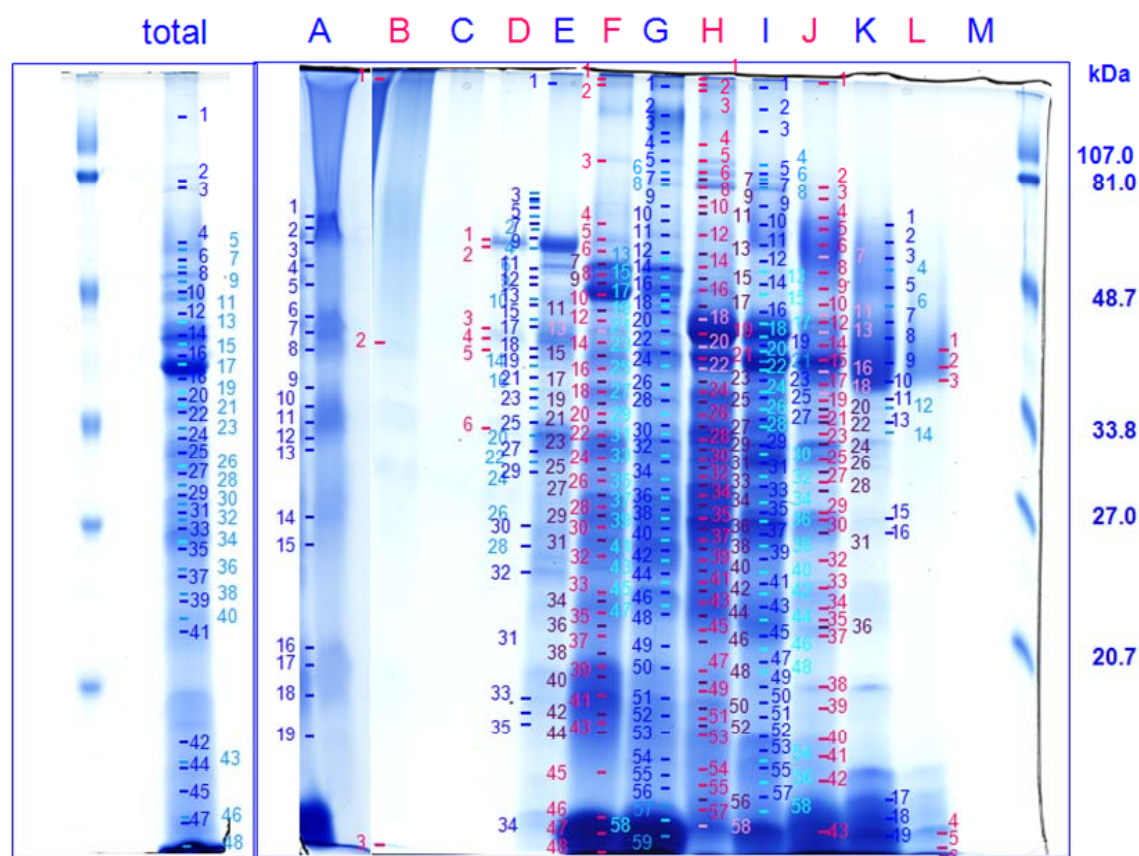
---

## Results

### *Establishment of methods for efficient proteomic analysis of hypocotyl cell walls*

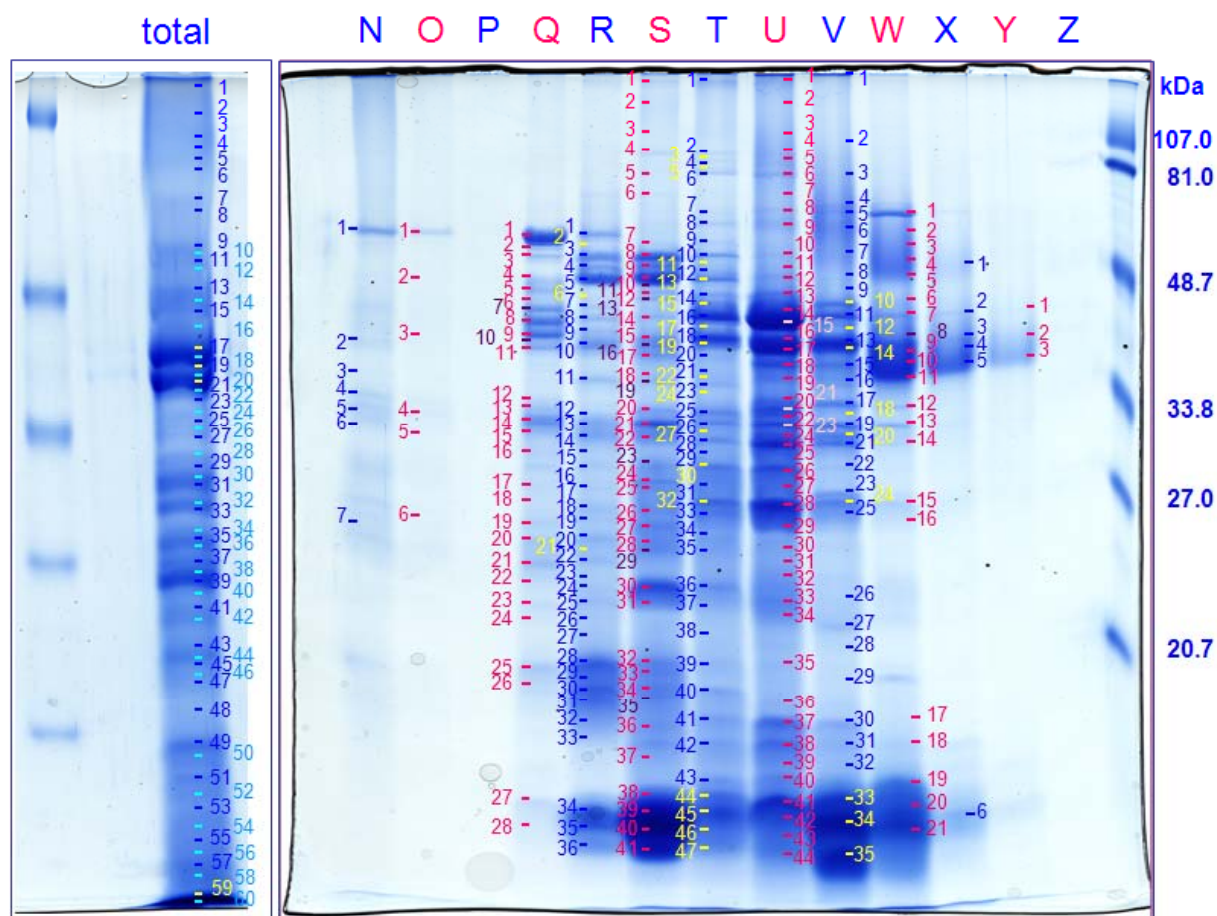
This proteomic study aimed at comparing two developmental stages of *Arabidopsis* hypocotyls and required high amounts of material. It required establishing appropriate culture conditions to get large amounts of homogenous material. They include synchronization of germination, high seedling density on culture medium, and control of culture conditions *in vitro*. Two stages of development were compared: 5-day-old elongating hypocotyls and 11 day-old fully-grown hypocotyls (Gendreau *et al.*, 1997; Derbyshire *et al.*, 2007). First, purified cell walls were prepared avoiding intracellular contamination and loss of proteins (Feiz *et al.*, 2006). To collect high amounts of proteins, two successive extractions were performed with 0.2 M CaCl<sub>2</sub> and 2 M LiCl (Feiz *et al.*, 2006). Both extracts were combined and used for further analysis. Typically, about 1 mg of proteins was obtained from 1 g of dry purified cell walls.

Finally, proteins were separated prior to identification using matrix-assisted laser desorption ionization - time of flight (MALDI-TOF) mass spectrometry (MS) and bioinformatics. Since 2D-E is not appropriate for the separation of CWPs (Jamet *et al.*, 2008), these proteins were separated using mono-dimensional gel electrophoresis (1D-E). About 60 bands could be stained with Coomassie Brilliant Blue (CBB). Fifty-two and 67 proteins could be identified in the extract from 5- and 11-day-old hypocotyls respectively (Figures 1 and 2, and Tables S1 and S2, supplementary material). Because of the limited resolution of 1D-E, many proteins were identified in each band, some of them with a low number of peptides. In order to improve separation and identification of proteins, we introduced an additional step before 1D-E. Since most CWPs are basic proteins (Jamet *et al.*, 2006), a cationic exchange chromatography was performed using a Fast Performance Liquid Chromatography (FPLC) device (Figures S1 and S2, supplementary material). Fractions were collected and combined prior to separation by 1D-E (Figures 1 and 2, S1b and S2b, supplementary material). Now, about 500 bands could be stained with CBB and further analyzed by MALDI-TOF MS. The percentage of successful protein identification from stained bands was about 70%: 141 and 109 proteins were identified in extracts from 5- and 11-day-old hypocotyls respectively. Many of those proteins were identified in several bands, reinforcing their identification. A great improvement of the quality of the analysis was observed: (i) the number of identified



**Figure 1. Analysis of proteins extracted by  $\text{CaCl}_2$  and  $\text{LiCl}$  from purified cell walls of 5-day-old etiolated hypocotyls.**

**Total.** Separation by 1D-E of the total extract of proteins. **Fractions A to M.** Separation by 1D-E of 13 fractions (A to M) obtained after cation exchange chromatography (see Figure S1, supplementary material). Molecular mass markers are on the right. Numbers refer to bands analyzed by MALDI-TOF MS with successful identification (see Table S1, supplementary material).



**Figure 2. Analysis of proteins extracted by CaCl<sub>2</sub> and LiCl from purified cell walls of 11-day-old etiolated hypocotyls.**

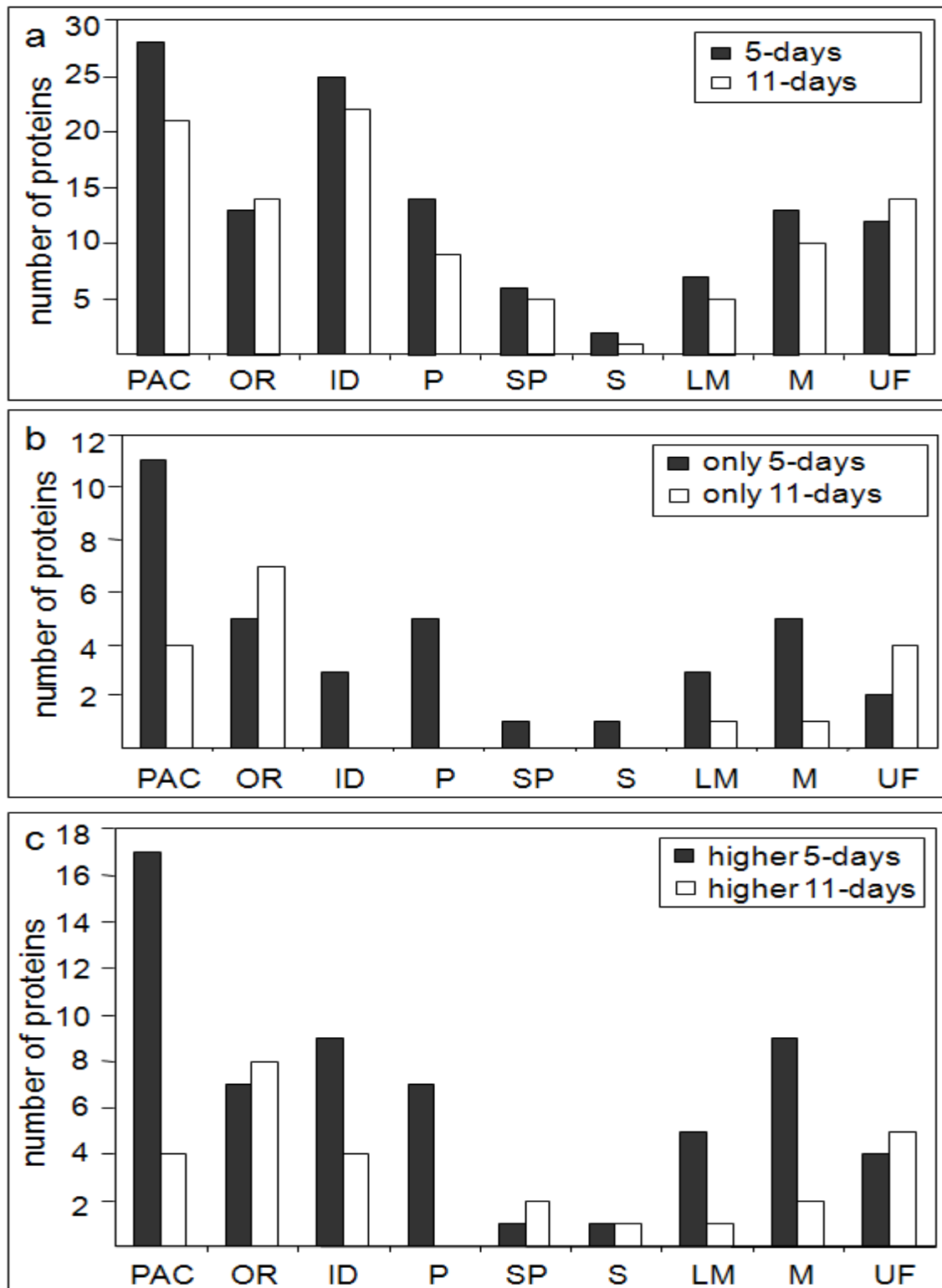
**Total.** Separation by 1D-E of the total extract of proteins. **Fractions N to Z.** Separation by 1D-E of 13 fractions (N to Z) obtained after cation exchange chromatography (see Figure S2, supplementary material). Molecular mass markers are on the right. Numbers refer to bands analyzed by MALDI-TOF MS with successful identification (see Table S2, supplementary material).

**Table 1.** Number of proteins identified in purified cell walls of 5- and 11-day-old etiolated hypocotyls of *Arabidopsis*.

	5-days	11-days	altogether
total number of identified proteins <sup>a</sup>	147	126	173
number of predicted secreted proteins <sup>a</sup>	120	101	137
number of predicted secreted proteins identified only in 5-or 11-day-old etiolated hypocotyls <sup>a</sup>	36	17	
number of predicted intracellular proteins <sup>a</sup>	27	25	36

<sup>a</sup> Proteins are listed in Table S3, supplementary material.





**Figure 3. Sorting of CWPs identified from purified cell walls of 5- and 11-day-old etiolated hypocotyls in functional classes.**

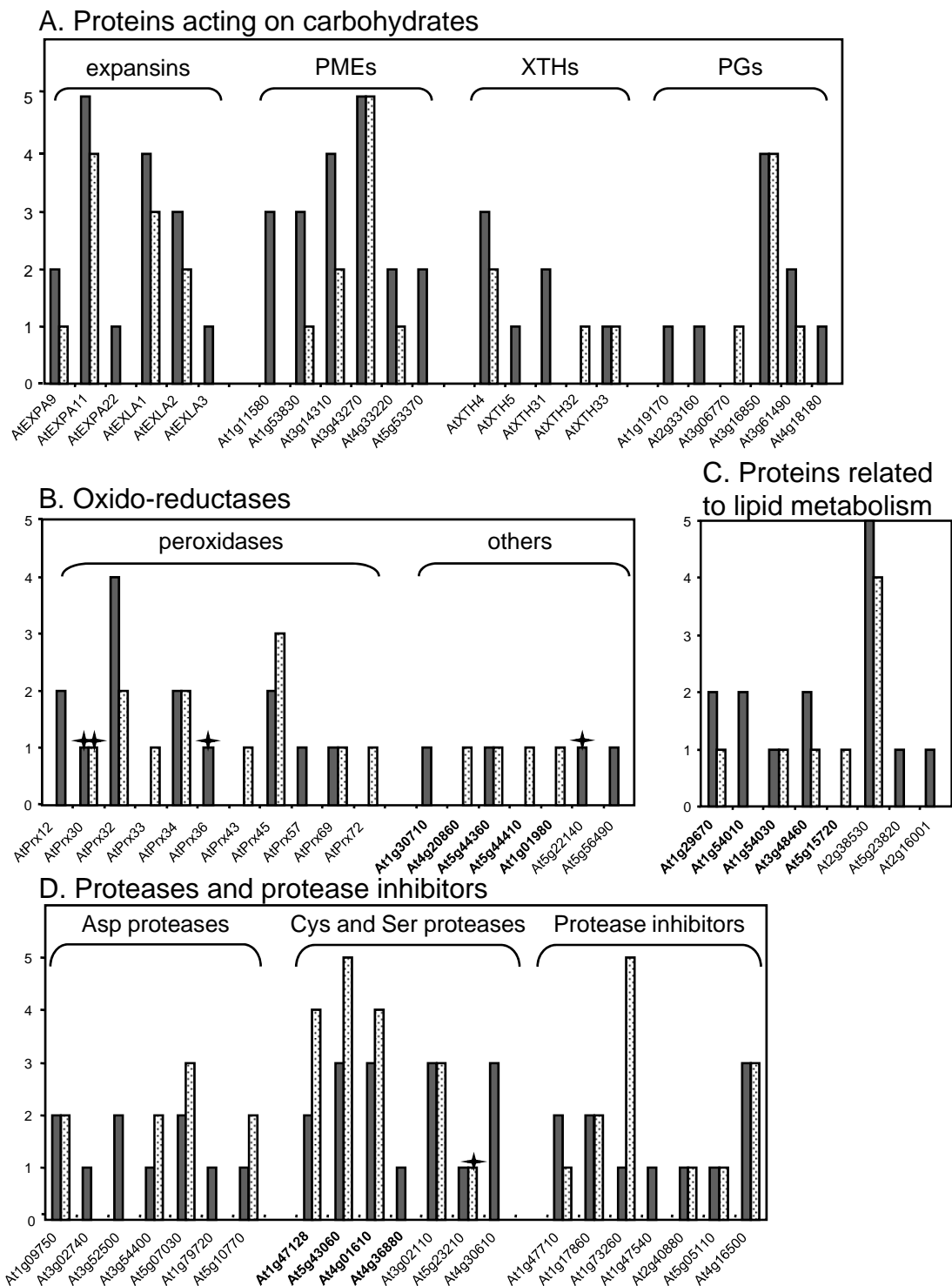
Proteins were classified according to their functional domains as described in Experimental procedures: proteins acting on carbohydrates (PAC), oxido-reductases (OR), proteins with interaction domains (ID), proteases (P), structural proteins (SP), proteins involved in signaling (S), proteins related to lipid metabolism (LM), proteins with miscellaneous functions (M), and proteins of yet unknown function (UF). **a.** Number of proteins found in the nine functional classes in 5- (5-days), and in 11-day-old hypocotyls (11-days). **b.** Number of proteins found only in 5- (only 5-days), or in 11-day-old (only 11-days) hypocotyls. **c.** Number of proteins present in higher amount in 5- (higher 5-days) or in 11-days-old (higher 11 days) hypocotyls.

proteins was doubled; (ii) the quality of the identifications was increased with higher numbers of peptides for identification of most proteins (Figures S3 and S4, supplementary material); (iii) it was possible to make a semi-quantification of proteins to compare the two samples.

*Proteins identified in cell wall extracts of Arabidopsis etiolated hypocotyls*

Combining results from 1D- and 2D-separation, 147 and 126 proteins were identified in 5- and 11-day-old etiolated hypocotyls respectively (Table 1). Bioinformatic prediction of their sub-cellular localization indicated that 120 (82%) and 101 (79%) proteins were secreted in 5- and 11-day-old hypocotyls respectively. On the other hand, 27 and 25 proteins (in 5- and 11-day-old hypocotyls respectively) did not have any predicted signal peptide and were considered as intracellular contaminants. Altogether, 173 proteins were identified in hypocotyls among which 137 (79%) were expected to be secreted, indicating the good quality of cell wall preparations. Those proteins will be called CWP in the following. Although many CWPs (84) were found in both samples, 36 and 17 were identified only in 5- and 11-day-old hypocotyls respectively.

A second bioinformatic analysis of CWPs allowed their classification according to functional domains. For several protein families, nomenclature of experts was used as described in Experimental procedures. The nine functional classes defined by Jamet *et al.* (2008) were employed (Figure 3): proteins acting on carbohydrates, oxido-reductases, proteins with interaction domains, proteases, structural proteins, proteins involved in signaling, proteins related to lipid metabolism, proteins with miscellaneous functions, and proteins of unknown function. Some protein classes were more represented at 5-days than at 11-days, *e.g.* proteins acting on carbohydrates (28 vs 21), proteins with interaction domains (25 vs 22), proteins related to lipid metabolism (7 vs 5), and proteases (14 vs 9) (Figure 3a). In each functional class, some CWPs were only identified at 5- or 11-days (Figure 3b). Differences appeared among proteins acting on carbohydrates (respectively 11 and 4 were found only at 5- or 11-days), among proteases (5 were only found at 5-days), proteins with domains of interactions with proteins or carbohydrates (3 were only found at 5-days), miscellaneous proteins and proteins of unknown function. In particular, the pattern of oxido-reductases appeared to be very different with 5 and 6 proteins being only found at 5- and 11-days respectively. On the contrary, signaling and structural proteins showed minor changes. Due to their specific structural characteristics, they are poorly represented in both proteomes.



**Figure 4. Occurrence of proteins belonging to several families among proteins extracted with CaCl<sub>2</sub> and LiCl from purified cell walls of 5- and 11-day-old etiolated hypocotyls.**

**a.** Families of proteins predicted to encode expansins, PMEs, XTHs and PGs. **b.** Protein families predicted to encode oxido-reductases. Berberine-bridge enzymes are in bold. **c.** Protein families predicted to encode proteins related to lipid metabolism. Proteins having a GDSL Lipase/Acylhydrolase domain are in bold. **d.** Families of proteins predicted to encode proteases or protease inhibitors. Cys proteases are in bold. The number of FPLC fractions in which each protein was identified was counted: black and white bars stand for 5- and 11-day-old etiolated hypocotyls respectively. Stars indicate proteins that have been only identified after separation by 1D-E. Detailed information about the distribution of proteins in total extract and in FPLC fractions are given in Table S3 (supplementary material).

Structural proteins are difficult to extract when they are cross-linked, and to identify because of numerous post-translational modifications (PTMs) (Shpak *et al.*, 2001). Proteins involved in signaling like arabinogalactan proteins (AGPs) may also have many PTMs (Schultz *et al.*, 2000), and proteins with trans-membrane domains are not usually extracted in our working conditions.

Finally, 51 proteins reported in this work were not identified in previous cell wall proteomic studies (Table 2). There are 11 CWPs acting on carbohydrates, 8 oxido-reductases, 5 proteases, 8 carrying interacting domains, 1 possibly involved in signaling, 1 structural protein, 4 proteins related to lipid metabolism, 5 proteins with diverse functions, and 7 without known function.

#### *Semi-quantitative comparative analysis of CWPs*

The comparison between the two physiological stages done in the previous section was based on the single criterion of presence/absence of a protein among proteins identified by MS in our working conditions. A more precise comparison would require quantification of the proteins. However, CBB staining of the gels did not allow such quantification for several reasons. Despite improvement in the separation of proteins by liquid chromatography followed by 1D-E as compared to 1D-E alone, most proteins were found in several FPLC fractions and in several bands of the same FPLC fractions. This was probably due to PTMs, proteolytic maturation or degradation of proteins. Moreover, MALDI-TOF MS analyses are not quantitative due to differences in ionization efficiency of diverse peptides, and to variations related to competitive desorption of peptides at the time of ionization. We propose alternative ways to compare the proteins between both samples. In a first approach, FPLC fractions in which proteins were identified were counted to give a first criterion based on the following rationale: an abundant protein is more difficult to resolve and will be distributed in more fractions than a rare protein. Such a calculation done on members of several gene families gave an overview on the relative abundance of each of them (Figure 4). Two additional criteria were then considered to evaluate the relative amount of each protein at both stages of development: the number of bands per fraction, and the percentage of coverage of the amino acid sequence by peptide mass mapping. A correlation was observed between the abundance of a protein and the number of matching peptides expressed as percentage of coverage (results not shown). The addition of all values for a given protein provided a semi-

**Table 2.** CWP s identified in salt extracts of purified cell walls of 5- and 11-day-old etiolated hypocotyls of *Arabidopsis*.

Accession numbers of genes encoding proteins predicted to have a GPI anchor or trans-membrane domains are in bold or in grey boxes respectively. Accession numbers of genes encoding proteins newly identified in etiolated hypocotyls through cell wall proteomics are underlined. Details of functional annotation are in Tables S1 and S2 (supplementary material).

accession	annotation	5-days	11-days
AGI			
<b>proteins acting on carbohydrates</b>			
At2g06850	GH family 16 (AtXTH4)	✓	✓
At5g13870	GH family 16 (AtXTH5)	✓	
<u>At3g44990</u>	GH family 16 (AtXTH31)	✓	
<u>At2g36870</u>	GH family 16 (AtXTH32)		✓
<u>At1g10550</u>	GH family 16 (AtXTH33)	✓	✓
At4g16260	GH family 17		✓
At4g19810	GH family 18 (chitinase)	✓	
At3g12500	GH family 19 (chitinase)	✓	✓
At3g55260	GH family 20 (beta-hexosaminidase)	✓	
<u>At1g19170</u>	GH family 28 (polygalacturonase)	✓	
At2g33160	GH family 28 (polygalacturonase)	✓	
<u>At3g06770</u>	GH family 28 (polygalacturonase)		✓
At3g16850	GH family 28 (polygalacturonase)	✓	✓
<u>At3g61490</u>	GH family 28 (polygalacturonase)	✓	✓
<u>At4g18180</u>	GH family 28 (polygalacturonase)	✓	
At1g68560	GH family 31 (alpha-xylosidase) (AtXYL1)	✓	✓
At3g13790	GH family 32 (beta-fructofuranosidase)	✓	✓
At5g34940	GH family 79 (endo beta-glucuronidase/heparanase)	✓	✓
At1g11580	CE family 8 (pectin methylesterase)	✓	
<b>At1g53830</b>	CE family 8 (pectin methylesterase) (AtPME2)	✓	✓
<b>At3g14310</b>	CE family 8 (pectin methylesterase) (AtPME3)	✓	✓
At3g43270	CE family 8 (pectin methylesterase)	✓	✓
At4g33220	CE family 8 (pectin methylesterase)	✓	✓
<b>At5g53370</b>	CE family 8 (pectin methylesterase)	✓	
<u>At4g37950</u>	PL family 4 (rhamnogalacturonate lyase)		✓
<b>At1g05570</b>	GT family 48 (callose synthase) (AtCalS1)	✓	✓
At5g02260	alpha-expansin (AtEXPA9)	✓	✓
At1g20190	alpha-expansin (AtEXPA11)	✓	✓
<u>At5g39270</u>	alpha expansin (AtEXPA22)	✓	
At3g45970	expansin-like A (AtEXLA1)	✓	✓
<u>At4g38400</u>	expansin-like A (AtEXLA2)	✓	✓
At3g45960	expansin-like A (AtEXLA3)	✓	
<b>oxido-reductases</b>			
At1g71695	peroxidase (AtPrx12)	✓	✓
At3g21770	peroxidase (AtPrx30)	✓	✓
At3g32980	peroxidase (AtPrx32)	✓	✓
At3g49110	peroxidase (AtPrx33)		✓
At3g49120	peroxidase (AtPrx34)	✓	✓
<u>At3g50990</u>	peroxidase (AtPrx36)	✓	
<u>At5g25980</u>	peroxidase (AtPrx43)		✓

At4g30170	peroxidase (AtPrx45)	✓	✓
At5g17820	peroxidase (AtPrx57)	✓	
At5g64100	peroxidase (AtPrx69)	✓	✓
At5g66390	peroxidase (AtPrx72)		✓
At2g30210	homologous to laccase (AtLAC3)	✓	✓
<u>At1g30710</u>	berberine-bridge enzyme homologue	✓	
<u>At4g20860</u>	berberine-bridge enzyme homologue		✓
<u>At5g44360</u>	berberine-bridge enzyme homologue	✓	✓
<u>At5g44410</u>	berberine-bridge enzyme homologue		✓
<u>At1g01980</u>	berberine-bridge enzyme homologue		✓
At2g02850	plantacyanin ARPN (blue copper binding protein)		✓
At4g12880	early nodulin AtEN20 (blue copper binding protein)	✓	✓
<u>At5g22140</u>	expressed protein (oxido-reductase domain)	✓	
<u>At5g56490</u>	expressed protein (FAD binding domain)	✓	

---

#### proteins with interacting domains

At1g53070	lectin homologue (legume lectin domains)	✓	✓
<u>At1g78820</u>	lectin homologue (curculin-like)	✓	
At1g78830	lectin homologue (curculin-like)	✓	✓
At1g78850	lectin homologue (curculin-like)	✓	✓
At1g78860	lectin homologue (curculin-like)	✓	✓
At5g06860	PGIP1 (LRR domains)	✓	✓
T23B7.10	PGIP1 homologue (LRR protein FLR1)	✓	✓
At5g12940	<i>Phaseolus vulgaris</i> PGIP2 homologue (LRR domains)	✓	✓
At1g33590	expressed protein (LRR domains)	✓	✓
At2g34930	expressed protein (LRR domains)	✓	✓
At3g20820	expressed protein (LRR domains)	✓	✓
<b>At2g17120</b>	expressed protein (LysM domain)	✓	✓
At1g03220	carrot EDGP and tomato XEGIP homologue	✓	✓
At1g03230	carrot EDGP and tomato XEGIP homologue	✓	✓
<u>At5g19110</u>	carrot EDGP and tomato XEGIP homologue	✓	✓
<u>At1g47710</u>	serpin homologue (serine protease inhibitor)	✓	✓
At1g17860	inhibitor family I3 (Kunitz-P family)	✓	✓
At1g73260	inhibitor family I3 (Kunitz-P family)	✓	✓
At1g47540	inhibitor family I18 (mustard trypsin inhibitor-2 family)	✓	✓
<u>At2g40880</u>	inhibitor family I25 (phytostatin)	✓	✓
<u>At5g05110</u>	inhibitor family I25 (phytostatin)	✓	✓
At4g16500	inhibitor family I25 (cystatin family)	✓	✓
<u>At4g25260</u>	invertase/ pectin methylesterase inhibitor homologue	✓	✓
<u>At5g46940</u>	invertase/pectin methylesterase inhibitor homologue	✓	
<u>At5g46960</u>	invertase/pectin methylesterase inhibitor homologue	✓	

---

#### proteases

At1g09750	aspartic protease homologue (pepsin family)	✓	✓
<u>At3g02740</u>	aspartic protease homologue (pepsin family)	✓	
At3g52500	aspartic protease homologue (pepsin family)	✓	
At3g54400	aspartic protease homologue (pepsin family)	✓	✓
At5g07030	aspartic protease homologue (pepsin family)	✓	✓
<u>At1g79720</u>	aspartic protease homologue (CND41 peptidase)	✓	
At5g10770	aspartic protease (CND41 peptidase)	✓	✓
At1g47128	cysteine proteinase homologue (papain family)	✓	✓
At5g43060	cysteine proteinase homologue (papain family)	✓	✓

At4g01610	cysteine proteinase homologue (papain family)	✓	✓
<u>At4g36880</u>	cysteine proteinase homologue (papain family)	✓	
<u>At3g02110</u>	serine carboxypeptidase D (SCPL25)	✓	✓
At5g23210	serine carboxypeptidase (SCPL34)	✓	✓
<u>At4g30610</u>	carboxypeptidase homologue (BRS1 - Brassinosteroid-Insensitive BRI suppressor 1)	✓	
<hr/>			
<b>structural proteins</b>			
At1g28290	proline-rich protein	✓	✓
<u>At5g14920</u>	proline-rich protein	✓	✓
At2g05580	glycine-rich protein	✓	✓
At4g13340	LRR-extensin (AtLRX3)	✓	
At3g24480	LRR-extensin (AtLRX4)	✓	✓
At4g18670	LRR-extensin (AtLRX5)	✓	✓
<hr/>			
<b>signaling</b>			
<u>At4g05200</u>	receptor kinase homologue (RLK, DUF26-1b subfamily)	✓	
<b>At5g55730</b>	fasciclin-like arabinogalactan protein (AtFLA1)	✓	✓
<hr/>			
<b>proteins related to lipid metabolism</b>			
At1g29670	lipase acylhydrolase homologue (GDSL family)	✓	✓
At1g54010	lipase/acylhydrolase homologue (GDSL family)	✓	
<u>At1g54030</u>	lipase/acylhydrolase homologue (GDSL family)	✓	✓
<u>At3g48460</u>	lipase/acylhydrolase homologue (GDSL family)	✓	✓
<u>At5g15720</u>	lipase/acylhydrolase homologue (GDSL family)		✓
At2g38530	non-specific lipid transfer protein type 1 (LTP2)	✓	✓
At5g23820	expressed protein (ML domain - MD-2-related lipid recognition domain)	✓	
<u>At2g16001</u>	expressed protein (lipid recognition domain)	✓	
<hr/>			
<b>miscellaneous functions</b>			
At2g27190	purple acid phosphatase homologue (PAP1)	✓	✓
<u>At3g07130</u>	purple acid phosphatase homologue	✓	✓
At5g34850	purple acid phosphatase homologue	✓	
At4g29270	acid phosphatase homologue	✓	✓
<u>At4g24340</u>	phosphorylase homologue homologue	✓	
<u>At3g02870</u>	myo-inositol monophosphatase homologue	✓	
At5g09440	<i>Nicotiana tabacum</i> phi-I homologue	✓	✓
At5g64260	<i>Nicotiana tabacum</i> phi-I homologue	✓	✓
<u>At5g66590</u>	<i>Nicotiana tabacum</i> pathogenesis-related protein PR1 homologue	✓	✓
At2g28790	<i>Lycopersicon esculentum</i> osmotin homologue	✓	✓
At5g15230	gibberellin-regulated protein (GASA4)	✓	
<b>At4g27110</b>	homologous to COBRA (AtCOBL10)	✓	
At1g09560	germin (subfamily 2, member 1, GLP5)		✓
<hr/>			
<b>unknown function</b>			
<u>At3g56750</u>	expressed protein		✓
<u>At3g22000</u>	expressed protein (DUF26)		✓
<u>At1g26850</u>	expressed protein (DUF248)		✓
At1g80240	expressed protein (DUF642)	✓	✓
At3g08030	expressed protein (DUF642)	✓	✓
At4g32460	expressed protein (DUF642)	✓	✓
At5g11420	expressed protein (DUF642)	✓	✓

At5g25460	expressed protein (DUF642)	✓	✓
<u>At1g78460</u>	expressed protein (SOUL heme binding domain)		✓
<u>At2g04690</u>	expressed protein (homologous to a human brain CREG protein)	✓	✓
At2g15220	expressed protein (Plant Basic Secreted Protein domain)	✓	✓
At2g34700	expressed protein (Ole e1 allergen domain)	✓	✓
At3g20370	expressed protein (MATH domain)	✓	
<u>At2g28490</u>	expressed protein (cupin domain)	✓	✓
At3g22640	expressed protein (cupin domain)	✓	✓
<u>At4g36700</u>	expressed protein (cupin domain)	✓	

---



quantitative index (QI) allowing comparison of the relative amount of each protein between the two samples (Table S4, supplementary material).

Taking into account QIs, 63% of the proteins showed differences between the two developmental stages: 61 proteins (42%) are more abundant at 5-days and 26 (17%) at 11-days. Figure 3c presents the results ordered by functional classes. Differences observed considering only the presence/absence of proteins are enhanced especially for proteins acting on carbohydrates with 17 out of 32 being more abundant at 5- than at 11-days as for proteins having domains of interactions or related to lipid metabolism, and miscellaneous proteins. These results were fully consistent with those described above only taking into account the presence of a protein in FPLC fractions (Figure 3b).

For proteins acting on carbohydrates (Figure 4a), all the identified expansins were found to be more represented at 5- than at 11-days. The situation was similar for PME with the exception of At3g43270 that was found in the same number of FPLC fractions at 5- and 11-days. On the contrary, XTHs and PGs presented a more complex pattern. Three XTHs were more represented at 5- than at 11-days (AtXTH4, AtXTH5, and AtXTH31), AtXTH33 was equally represented at both stages, and AtXTH32 was only found at 11-days. Three PGs were only found at 5-days (At2g33160, At1g19170, and At4g18180), At3g15850 was equally represented at 5- and 11-days, and At3g06770 was only found at 11-days.

The distribution of oxido-reductases between the two stages looked as complicated as that of XTHs (Figure 4b). AtPrx36 and AtPrx57 were only identified at 5-days whereas AtPrx12, AtPrx43, and AtPrx72 were only identified at 11-days. AtPrx32 was more represented at 5-days whereas AtPrx45 was more represented at 11-days. Finally, AtPrx30, AtPrx34, and AtPrx69 were equally represented at both stages. Among other proteins predicted to be oxido-reductases, berberine-bridge enzymes were distributed between the two stages of development.

Most proteins related to lipid metabolism were more represented at 5- than at 11-days (Figure 4c). This is especially the case for proteins containing a GDSL-Lipase/Acylhydrolase domain (At1g29670, At1g54010, and At3g48460).

---

In the case of proteases (Figure 4d), the situation was contrasted depending on the protease family. Three Asp proteases were only found at 5-days (At1g79720, At3g02740, and At3g52500), whereas 3 of them were more abundant at 11- than at 5-days (At3g54400, At5g10770, and At5g07030). At1g09750 was equally represented at the two stages. Cys proteases were more represented at 11- than at 5-days (At1g47128, At5g43060, and At4g01610). On the contrary, a Ser protease was only found at 11-days (At4g30610). Finally, protease inhibitors (Figure 4d) were distributed between the two stages of development with 4 of them equally represented at both stages, 2 of them only found at 5-days, and 1 more represented at 11-days.

---

## Discussion

This work provides a global picture of the cell wall proteome during elongation of etiolated hypocotyls of *Arabidopsis*. It shows the dynamics of CWPs during two phases of hypocotyls development, *i.e.* active elongation and after growth arrest. Expected CWPs known to be involved in cell wall extension such as XTHs, expansins, PGs, PMEs and peroxidases were identified as well as new CWPs such as proteases, proteins predicted to be related to lipid metabolism and proteins of unknown function. In addition, the occurrence of CWPs well-known to be related to cell wall extension after growth arrest showed that such proteins probably play other functions in mature cell walls.

A major advance in plant cell wall proteomics was brought by setting up a new separation method for CWPs. Separation of plant CWPs for proteomic purposes was difficult using 2D-E (Jamet *et al.*, 2008). The window of protein separation is optimal for pIs between 3 and 10 and for molecular masses between 120 and 10 kDa. Since most CWPs are basic glycoproteins, they tend to migrate as a smear on the basic side of 2D-gels (Minic *et al.*, 2007). Alternative methods were proposed that consisted in separation of CWPs into an acidic and a basic fractions by cationic exchange chromatography followed by 2D-E and 1D-E respectively (Borderies *et al.*, 2003; Boudart *et al.*, 2005). The new method includes a first step of separation by cationic exchange chromatography at acidic pH, and a second step of separation by 1D-E. It gives more information on the physico-chemical properties of the proteins, allows comparative semi-quantification between different samples, as well as a better identification through MALDI-TOF MS. In the case of etiolated hypocotyls of *Arabidopsis*, it allowed doubling the number of proteins identified as compared to separation by 1D-E alone. In addition, since many CWPs can now be visualized, this work provided preparative tools for developing biochemical studies on CWPs, *e.g.* for further purification or structural characterization.

Altogether, 137 CWPs were identified in this study among which 51 were not yet identified through cell wall proteomics. This work also presents an overview of the dynamics of CWPs during cell elongation. More CWPs were identified at 5- than at 11-days, suggesting a higher metabolic activity in cell walls during active elongation. Many differences were observed between elongating and full-size hypocotyls. When considering only presence/absence of a CWP, these changes concerned 53 out of the 137 identified CWPs

(38%). When taking into account the proposed semi-quantification method, this percentage even increased to 63% (34 additional proteins). Changes in the same gene family can reflect regulation of gene expression at different stages of development and/or differences in biological activity as discussed below for XTHs, PGs, expansins, PME, and peroxidases. Proteins acting on carbohydrates are more numerous and more abundant in elongating hypocotyls than in full-size hypocotyls. This was expected since rearrangements of cell wall polysaccharides are very important during cell elongation (Derbyshire *et al.*, 2007). More surprising is the fact that proteases are more numerous and in higher amounts at 5- than at 11-days. Nothing is known about their targets in cell walls. Are they contributing to release peptides involved in signaling (Matsubayashi and Sakagami, 2006)? Are they involved in protein maturation (Lee *et al.*, 2004) or in protein degradation? Conversely, two protease inhibitors are much more abundant at 11- than at 5-days. Altogether, it seems that proteolytic activities are more involved when elongation is active than during elongation arrest. Among oxido-reductases, five berberine-bridge enzymes were identified among which three were only present at 11-days. The role of such proteins in cell walls is unknown. For proteins with interacting domains, changes are found in lectins and PME inhibitors that are more abundant at 5-days. Among miscellaneous proteins, the amount of CWPs containing phosphatase domains was found to be higher at 5- than at 11-days. Such proteins were shown to be associated to regeneration of protoplast cell walls (Kaida *et al.*, 2003) and pollen tube growth (Ibrahim *et al.*, 2002), but their precise roles are unknown. A protein homologous to COBRA (AtCOBL10) was only found at 5-days. Although the function of AtCOBL10 is not known, it should be noted that COBRA was shown to participate in orientation of cellulose microfibrils, and dark-grown hypocotyls of *cob-4* mutant have a 95% reduction in length compared to wild-type (Roudier *et al.*, 2005). AtCOBL10 might play such role during elongation of hypocotyl cells.

Many proteins expected to participate in cell wall extension were found such as XTHs, expansins, PGs, PME and peroxidases (Micheli, 2001; Passardi *et al.*, 2004; Cosgrove, 2005; Zhang *et al.*, 2007). But such proteins, *i.e.* same proteins or proteins of the same family were also found after completion of elongation. Several hypotheses can be proposed. Although many proteases were identified at both stages of development suggesting a regulation of CWPs by proteolytic degradation, these proteins can have a long half-life. Some of these proteins probably also participate in differentiation of tracheary elements. XTHs belonging to phylogenetic group 3 such as AtXTH31-33 have been assumed to have

xyloglucan endo-hydrolysis activity (Rose *et al.*, 2002). They might be involved in rearrangement of cell walls of differentiating vessels elements. Such elements can be observed using microscopy (not shown). In the same way, some expansins were found in differentiating tracheary elements (Li *et al.*, 2003). Finally, at least PME and peroxidases were assumed to play roles both during the elongation process and elongation arrest. The enzymatic activity of PMEs might be modulated depending on the pH of the extracellular matrix and on the structure of pectic homogalacturonans. They could have either a local activity favoring the enzymatic activity of endo-PGs thus producing fragments of pectin, or a processing activity leading to de-esterification of stretches of GalA and formation of the so-called egg-boxes that tend to rigidify the pectic network (Micheli, 2001). Moreover, the degree of pectin methyl-esterification was shown to be positively correlated to hypocotyl growth (Derbyshire *et al.*, 2007). The activity of peroxidases is also versatile (Passardi *et al.*, 2004). During the hydroxylic cycle, peroxidases can produce reactive oxygen species that can break cell wall polysaccharides in a non-enzymatic way thus favoring cell wall extension (Fry, 1998; Passardi *et al.*, 2004). On the contrary, during the peroxidative cycle, peroxidases can promote cross-linking of cell wall components such as structural proteins or lignins. In addition, members of most of these protein families were identified in apoplastic fluids of rosette leaves (Boudart *et al.*, 2005). Since leaf cells are surrounded by mature walls, it means that those CWPs could play house-keeping roles.

Proteomics provides information about possible regulatory mechanisms of CWPs. As previously discussed in (Fry, 2004), the presence of a protein does not mean that it has full biological activity. Proteins with putative enzymatic activities are numerous, but inhibitors of these activities are also present. This is the case for proteases (14) and protease inhibitors (7), PMEs (6) and PME inhibitors (3 PMEIs). Some PMEs have a pro-domain consisting in a PMEI. However, such domains are assumed to be cleaved during or just after protein export since they were never found in purified PMEs (Micheli, 2001). In the same way, no peptide matching the PME domains were found during identification of PMEs by peptide mass mapping (data not shown). Other enzyme inhibitors are assumed to be involved in defense reaction such as PG inhibiting proteins (3 PGIPs) and inhibitors of xyloglucan endoglucanases (3 XEGIPs) since some of them were shown to be specifically active against fungal enzymes (Juge, 2006). But this was not proven for all of them. Other regulatory mechanisms include variations in pH of extracellular matrix that occur during growth arrest

---

(Juge, 2006), physical contact between enzymes and their substrates (Fry, 2004) and proteolytic degradation.

Eight proteins predicted to be related to lipid metabolism were identified at both stages of hypocotyl development. At present, little is known about the functions of such proteins in cell walls. Since etiolated hypocotyls have a thicker cuticle than light-grown hypocotyls (Gendreau *et al.*, 1997), presence of proteins involved in cuticle formation is expected. Several genes encoding proteins from the same families have been found to be up-regulated in 35S::AtMYB41 plants having defects in cell expansion and leaf surface permeability (Cominelli *et al.*, 2008). Two mutants affected in genes encoding proteins related to lipid metabolism have been described. *GLIP1* encodes a predicted lipase/acylhydrolase that was shown to have a lipase activity *in vitro* and to disrupt fungal spore integrity at the level of cell wall and/or membrane (Oh *et al.*, 2005). Although none of the proteins of the GDSL family was shown to have an activity towards natural lipids *in vitro*, it cannot be excluded that such proteins are hydrolases acting on cutin or suberin lipids (F. Beisson, personal communication). *BODYGUARD* encodes a protein predicted to belong to an  $\alpha/\beta$ -hydrolase fold superfamily (Kurdyukov *et al.*, 2006). The *bodyguard* mutant shows defects in cuticle formation that could result from incomplete polymerization of carboxyl esters of the cuticle. The function of LTPs is still a matter of debate. They were shown to bind fatty acids and to transfer phospholipids between membranes *in vitro* (Kader, 1997). At2g38530 encoding LTP2 was found to be up-regulated in the epidermis of stems and assumed to contribute to active cuticle formation during stem elongation (Suh *et al.*, 2005). Apart from this role in cuticle formation, many roles were proposed for LTPs including systemic resistance signaling (Maldonado *et al.*, 2002), ability to promote cell wall expansion through binding to a hydrophobic partner in cell walls (Nieuwland *et al.*, 2005), and activation of a PG (Tomassen *et al.*, 2007). CWPs predicted to be related to lipid metabolism identified in this study are candidates for playing roles in cuticle formation.

This proteomic survey provides tools for biochemical studies of CWPs, identifies members of multigene families involved in cell elongation, and gives clues for unraveling the function of many CWPs in etiolated hypocotyls. It also shows a great dynamics of CWPs between the two developmental stages. This is illustrated by changes in protein amount, presence/absence of specific members of multigene families, and presence of many enzymes including proteases and their inhibitors. Interestingly, many CWPs were only found at one

stage of development, either active elongation or after growth arrest. Conversely, different CWP families from the same gene families were found at both stages of development showing stage-specific regulation and suggesting diverse roles in cell walls. A particular attention should be paid to proteins of unknown function, some of which are very abundant. Additional functional studies are now required to shed light on the roles of the identified CWPs during the elongation of hypocotyls in the dark.

### **Acknowledgements**

The authors are grateful to the Université Paul Sabatier and the CNRS (France) for support. They thank Dr C. Albenne and C. Picheraux for MALDI-TOF-TOF analysis and J. d'Alayer at Institut Pasteur. M.I. is a fellow of the Higher Education Commission of Pakistan, Islamabad, and of the French government on the behalf of SFERE. MS analyses were performed on the *Plate-forme Protéomique Toulouse Midi-Pyrénées*, France.

### **Additional data files**

#### **Additional data files1.**

Table S1. Identification of proteins extracted from purified cell walls of 5-day-old Arabidopsis hypocotyls by CaCl<sub>2</sub> and LiCl solutions.

#### **Additional data files2.**

Table S2. Identification of proteins extracted from purified cell walls of 11-day-old Arabidopsis hypocotyls by CaCl<sub>2</sub> and LiCl solutions.

#### **Additional data files3.**

Table S3. Classification of proteins extracted from purified cell walls of 5- and 11-day-old Arabidopsis hypocotyls in FPLC fractions A to Z.

#### **Additional data files4.**

Table S4. Semi-quantification of CWPs identified by MALDI-TOF MS from 5- and 11-day old etiolated Arabidopsis hypocotyls.

---

**References**

**Bakalovic N, Passardi F, Ioannidis V, Cosio C, Penel C, Falquet L, Dunand C. 2006.** PeroxiBase: a class III plant peroxidase database. *Phytochemistry* **67**: 534-539.

**Baumberger N, Doesseger B, Guyot R, Diet A, Parsons RL, Clark MA, Simmons MP, Bedinger P, Goff SA, Ringli C, Keller B. 2003.** Whole-genome comparison of leucine-rich repeat extensins in *Arabidopsis* and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade. *Plant Physiology* **131**: 1313-1326.

**Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerré-Tugayé M-T, Boudet A, Pont-Lezica R. 2003.** Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: A critical analysis. *Electrophoresis* **24**: 3421-3432.

**Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerré-Tugayé M-T, Pont-Lezica R. 2005.** Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics. *Proteomics* **5**: 212-221.

**Bradford MM. 1976.** A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254.

**Cominelli E, Sala T, Calvi D, Gusmaroli G, Tonelli C. 2008.** Overexpression of the *Arabidopsis AtMyb41* gene alters cell expansion and leaf surface permeability. *The Plant Journal* **58**: 53-64.

**Cosgrove DJ. 2005.** Growth of the plant cell wall. *Nature Review. Molecular Cell Biology*: 850-860.

**Coutinho PM, Henrissat B. 1999.** Carbohydrate-active enzymes: an integrated database approach. In: Gilbert HJ, Davies B, Henrissat B, Svensson B, eds. *Carbohydrate-active enzymes: an integrated database approach*. Cambridge: The Royal Society of Chemistry, 3-12.



---

**Derbyshire P, McCann MC, Roberts K. 2007.** Restricted cell elongation in *Arabidopsis* hypocotyls is associated with a reduced average pectin esterification level. *BMC Plant Biology* **7**: 31.

**Derbyshire P, Findlay K, McCann MC, Roberts K. 2007.** Cell elongation in *Arabidopsis* hypocotyls involves dynamic changes in cell wall thickness. *Journal of Experimental Botany* **58**: 2079-2089.

**Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E. 2006.** Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls. *Plant Methods* **2**: 10.

**Fowler TJ, Bernhardt C, Tierney ML. 1999.** Characterization and expression of four proline-rich cell wall protein genes in *Arabidopsis* encoding two distinct subsets of multiple domain proteins. *Plant Physiology* **121**: 1081-92.

**Freshour G, Clay RP, Fuller MS, Albersheim P, Darvill AG, Hahn MG. 1996.** Developmental and tissue-specific structural alterations of the cell-wall polysaccharides of *Arabidopsis thaliana* roots. *Plant Physiology* **110**: 1413-1429.

**Fry SC. 1998.** Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *The Biochemical Journal* **332**: 507-515.

**Fry SC. 2004.** Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. *New Phytologist* **161**: 641-675.

**Gendreau E, Traas J, Desnos T, Grandjean O, Caboche M, Hofte H. 1997.** Cellular basis of hypocotyl growth in *Arabidopsis thaliana*. *Plant Physiology* **114**: 295-305.

**Ibrahim H, Pertl H, Pittertschatscher K, Fadl-Allah E, el-Shahed A, Bentrup FW, Obermeyer G. 2002.** Release of an acid phosphatase activity during lily pollen tube growth involves components of the secretory pathway. *Protoplasma* **219**: 176-183.

---

**Jacobs J, Roe JL. 2005.** *SKS6*, a multicopper oxidase-like gene, participates in cotyledon vascular patterning during *Arabidopsis thaliana* development. *Planta* **222**: 652-666.

**Jamet E, Canut H, Boudart G, Pont-Lezica RF. 2006.** Cell wall proteins: a new insight through proteomics. *Trends in Plant Science* **11**: 33-39.

**Jamet E, Albenne C, Boudart G, Irshad M, Canut H, Pont-Lezica R. 2008.** Recent advances in plant cell wall proteomics. *Proteomics* **8**: 893-908.

**Johnson KL, Jones BJ, Schultz CJ, Bacic A. 2003.** Non enzymic cell wall (glyco)proteins. In: Rose JK, ed, *Non enzymic cell wall (glyco)proteins*. Boca Raton: CRC Press LLC, 111-154.

**Juge N. 2006.** Plant protein inhibitors of cell wall degrading enzymes. *Trends in Plant Science* **11**: 359-367.

**Kader J-C. 1997.** Lipid transfer proteins: a puzzling family of plant proteins. *Trends in Plant Science* **2**: 66-70.

**Kaida R, Sage-Ono K, Kamada H, Okuyama H, Syono K, Kaneko TS. 2003.** Isolation and characterization of four cell wall purple acid phosphatase genes from tobacco cells. *Biochimica et Biophysica Acta* **1625**: 134-140.

**Kurdyukov S, Faust A, Nawrath C, Bär S, Voisin D, Efremova N, Franke R, Shreiber L, Saedler H, Métraux J-P, Yephremov A. 2006.** The epidermis-specific extracellular BODYGUARD controls cuticle development and morphogenesis in *Arabidopsis*. *The Plant Cell* **18**: 321-339.

**Laemmli UK. 1970.** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

**Lee SJ, Saravanan RS, Damasceno CM, Yamane H, Kim BD, Rose JK. 2004.** Digging deeper into the plant cell wall proteome. *Plant Physiology and Biochemistry* **42**: 979-988.

---

**Li Y, Jones L, McQueen-Mason S. 2003.** Expansins and cell growth. *Current Opinion in Cell Biology* **6**: 603-610.

**Maldonado AM, Doerner P, Dixon RA, Lamb CJ, Cameron RK. 2002.** A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature* **26**: 399-403.

**Matsubayashi Y, Sakagami Y. 2006.** Peptide hormones in plants. *Annual Review of Plant Biology* **57**: 649-674.

**McCaig BC, Meagher RB, Dean JF. 2005.** Gene structure and molecular analysis of the laccase-like multicopper oxidase (LMCO) gene family in *Arabidopsis thaliana*. *Planta* **221**: 619-36.

**Micheli F. 2001.** Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends in Plant Science* **6**: 414-419.

**Minic Z, Jamet E, Negroni L, der Garabedian PA, Zivy M, Jouanin L. 2007.** A sub-proteome of *Arabidopsis thaliana* trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. *Journal of Experimental Botany* **58**: 2503-2512.

**Nersissian AM, Shipp EL. 2002.** Blue copper-binding domains. *Advances in Protein Chemistry* **60**: 271-340.

**Nicol F, Hofte H. 1998.** Plant cell expansion: scaling the wall. *Current Opinion in Cell Biology* **1**: 12-17.

**Nieuwland J, Feron R, Huisman BAH, Fasolino A, Hilbers CW, Derksen J, Mariani C. 2005.** Lipid transfer proteins enhance cell wall extension in tobacco. *The Plant Cell* **17**: 2009-2019.

**Oh IS, Park AR, Bae MS, Kwon SJ, Kim YS, Lee JE, Kang NY, Lee S, Cheong H, Park OK. 2005.** Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternaria brassicicola*. *The Plant Cell* **17**: 2832-47.

**Passardi F, Penel C, Dunand C. 2004.** Performing the paradoxical: how plant peroxidases modify the cell wall. *Trends in Plant Science* **9**: 532-540.

**Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I. 2005.** *TRANSPARENT TESTA10* encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat. *The Plant Cell* **17**: 2966-80.

**Refrégier G, Pelletier S, Jaillard D, Hofte H. 2004.** Interaction between wall deposition and cell elongation in dark-grown hypocotyl cells in *Arabidopsis*. *Plant Physiology* **135**: 959-68.

**Rose JK, Braam J, Fry SC, Nishitani K. 2002.** The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. *Plant & Cell Physiology* **43**: 1421-35.

**Roudier F, Schindelman G, DeSalle R, Benfey PN. 2002.** The COBRA family of putative GPI-anchored proteins in *Arabidopsis*: a new fellowship in expansion. *Plant Physiology* **130**: 538-548.

**Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GH, Schindelman G, Song S, Baskin TI, Dupree P, Wasteneys GO, Benfey PN. 2005.** *COBRA*, an *Arabidopsis* extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *The Plant Cell* **17**: 1749-63.

**Saibo NJ, Vriezen WH, Beemster GT, Van Der Straeten D. 2003.** Growth and stomata development of *Arabidopsis* hypocotyls are controlled by gibberellins and modulated by ethylene and auxins. *The Plant Journal* **33**: 989-1000.

**Scheler C, Lamer S, Pan Z, Li XP, Salnikow J, Jangblut P. 1998.** Peptide mass fingerprint sequences coverage from differentially stained proteins in two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* **19**: 918-927.

---

**Schultz CJ, Johnson KL, Currie G, Bacic A. 2000.** The classical arabinogalactan gene family of *Arabidopsis*. *The Plant Cell* **12**: 1751-1767.

**Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YG, Antony Bacic A. 2002.** Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiology* **129**: 1448-1463.

**Shpak E, Barbar E, Leykam JF, Kieliszewski MJ. 2001.** Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. *The Journal of Biological Chemistry* **276**: 11272-8.

**Smallwood M, Beven A, Donovan N, Neill S, Peart J, Roberts K, Knox J. 1994.** Localization of cell wall proteins in relation to the developmental anatomy of the carrot root apex. *The Plant Journal* **5**: 237-246.

**Suh MC, Samuels AL, Jetter R, Kunst L, Pollard M, Ohlrogge J, Beisson F. 2005.** Cuticular lipid composition, surface structure, and gene expression in *Arabidopsis* stem epidermis. *Plant Physiology* **139**: 1649-1665.

**Tomassen MMM, Barrett DM, van der Valk HCPM, Woltering EJ. 2007.** Isolation and characterization of a tomato non-specific lipid transfer protein involved in polygalacturonase-mediated pectin degradation. *Journal of Experimental Botany* **58**: 1151-1160.

**van Hengel AJ, Roberts K. 2003.** AtAGP30, an arabinogalactan-protein in the cell walls of the primary root, plays a role in root regeneration and seed germination. *The Plant Journal* **36**: 256-270.

**Zhang Z, Pierce ML, Mort AJ. 2007.** Changes in homogalacturonans and enzymes degrading them during cotton cotyledon expansion. *Phytochemistry* **68**: 1094-1113.

# **CHAPTER 5**

## **CELL WALL BIOGENESIS OF *ARABIDOPSIS* ELONGATING CELLS: TRANSCRIPTOMICS COMPLEMENTS PROTEOMICS**

---

**Chapter summary (French)**

Pour mieux comprendre la biogenèse des parois, une étude transcriptomique a été réalisée sur le même matériel que l'étude protéomique décrite dans le chapitre 4. Il s'agissait d'une part d'apporter des éléments concernant les niveaux de régulation des gènes codant des protéines impliquées dans la mise en place de la paroi, de rechercher des nouveaux gènes importants pour l'élongation des parois, et de comparer les résultats obtenus à ceux de la protéomique. Les puces CATMA (*Complete Arabidopsis Transcriptome MicroArray*) ont été utilisées dans le cadre d'une collaboration avec l'équipe de Jean-Pierre Renou de l'URGV d'Evry. Les résultats ont été analysés en se focalisant sur des gènes codant des protéines impliquées dans la biogenèse des parois et des gènes codant des protéines pariétales.

Dans un premier temps, le niveau de transcrits d'un ensemble de 846 gènes appelés CWGs (Cell Wall Genes) et annotés par des experts a été analysé. Ces gènes codent pour des protéines qui participent d'une part à la synthèse et au transport des constituants pariétaux, d'autre part à leur assemblage dans les parois ainsi qu'au remodelage des réseaux de polysaccharides, de protéines et de lignine. Environ 55% de ces gènes ont des niveaux de transcrits détectables sur les puces CATMA aux deux stades développement des hypocotyles. Ceci suggère que les processus de synthèse et d'assemblage des constituants pariétaux sont importants à la fois au cours de l'élongation cellulaire et après son arrêt. Les niveaux de transcrits d'autres gènes tels ceux qui codent pour des AGP et des FLA ont également été analysés. Certains d'entre eux ont des niveaux de transcrits très élevés. Cependant leur rôle dans les parois n'est pas encore connu.

Dans un second temps, nous avons recherché des gènes codant pour des protéines secrétées dont les niveaux de transcrits seraient modérés ou élevés. L'idée était d'identifier de nouveaux gènes importants pour les remaniements pariétaux au cours de l'élongation cellulaire. Les 278 gènes identifiés ont été appelés SPGs (*Secretory Pathway Genes*). Ils appartiennent aux mêmes classes fonctionnelles que celles qui ont été définies au cours des études protéomiques. Parmi les SPGs, figurent 55 gènes codant des protéines agissant sur les polysaccharides pariétaux, 6 peroxydases. Après avoir retiré ces gènes déjà identifiés parmi les CWP, ce sont plus de 200 nouveaux gènes qui pourraient jouer des rôles dans l'élongation cellulaire parmi lesquels 70 codent des protéines de fonction inconnue.

Moins de 100 gènes codant des protéines secrétées ont des niveaux de transcrits significativement variables entre les deux stades de développement des hypocotyles (43 ont un niveau de transcrits supérieur à 5 jours, et 38 à 11 jours). Une réduction du niveau de transcrits entre 5 et 11 jours était attendue pour des gènes codant des GT et des GH, mais pas pour des gènes codant des PME dont l'activité de déméthylation des pectines semblait plutôt décrite après la phase d'élongation cellulaire pour favoriser la formation de complexes avec  $\text{Ca}^{2+}$  (« boîtes à œufs »). Par contre, une augmentation du niveau de transcrits de gènes codant des GRP était attendue à 11 jours, puisque les GRP ont été associées à la formation de protoxylème. Les différences entre les résultats attendus et ceux obtenus par ces analyses pourraient être expliquées soit par des nouveaux rôles des gènes concernés, soit par des régulations particulières de leur expression. Il pourrait exister un délai entre la synthèse de certains transcrits et leur traduction en protéine, ou encore, certaines protéines pourraient être particulièrement stables. Enfin, il faudrait prendre en compte l'activité des protéases qui pourraient intervenir dans l'activation de certaines enzymes ou dans leur dégradation.

Enfin, les données de transcriptomique ont été comparées aux données de protéomique. Trois cas ont été relevés : (i) la protéine est identifiée, et le niveau de transcrits du gène correspondant est détectable sur les puces CATMA; (ii) le gène a un niveau de transcrit modéré ou élevé, mais la protéine n'est pas identifiée ; (iii) la protéine est identifiée, mais le niveau de transcrits du gène correspondant n'est pas détectable sur les puces CATMA. Plusieurs explications peuvent être proposées pour le second cas du point de vue de la protéomique. Certaines protéines sont difficiles à extraire des parois, peu abondantes, difficiles à colorer avec le bleu de Coomassie, ou encore difficiles à identifier par spectrométrie de masse du fait de leurs nombreuses modifications post-traductionnelles. Il se pourrait aussi que les protéines aient une demi-vie très courte ou encore, comme évoqué précédemment, qu'il existe un délai entre transcription et traduction d'un ARNm. Dans le troisième cas, les transcrits pourraient avoir une courte demi-vie, et les protéines au contraire être stables. En conclusion, nos travaux ont montré qu'il n'y a pas de cohérence systématique entre les données de transcriptomique et de protéomique. La régulation post-transcriptionnelle semble être importante dans le cas des gènes codant des protéines pariétales dans les hypocotyles étiolés. Les données de transcriptomique et de protéomique semblent donc complémentaires pour étudier la régulation des gènes codant des protéines pariétales. L'ensemble de ces résultats est détaillé dans l'article qui suit actuellement soumis pour publication.



---

To better understand cell wall biogenesis by complementing the results of proteomic study (chapter 4) and to know about the transcriptional regulation during hypocotyl elongation, a transcriptomic study was performed on same material used for proteomic study in chapter 4. For transcript profiling, Complete *Arabidopsis* Transcriptome MicroArray (CATMA) was used at URGV, Evry in collaboration with the research team of Jean-Pierre Renou. This chapter provides an analysis of:

- i) the transcript level of genes that could possibly take part in the biogenesis of cell wall
- ii) the transcript level of the genes for which the proteins were identified in the proteomic study (chapter 4)

In a first selection, the level of a set of 846 genes called cell wall genes (CWGs) and known to be involved in cell wall biogenesis was measured. At both stages of hypocotyl development, about 55% of these genes had detectable transcript level on CATMA. These genes are participating in synthesis, transport, assembly and rearrangement of cell wall polysaccharides (nucleotide-sugar inter-conversion pathway, monolignol biosynthesis, vesicle trafficking, structural proteins, GTs, GHs, XTHs, PME<sub>s</sub>, PL<sub>s</sub>, expansins, peroxidases) suggesting that these two processes are very active during elongation. Levels of transcripts of other genes like those encoding AGPs and ALFs have also been analysed. Some of them have very high levels of transcripts, however, their role in cell wall is not yet known.

In a second step, we looked for genes that encode secretory proteins with moderate or high level of transcripts. The idea was to identify new genes important for assembly and reorganization of cell wall during elongation. The 278 genes identified in this way were called Secretory Pathway Genes (SPGs). They belong to the same functional classes to which belonged the proteins identified in the proteomic study. Among the SPGs, 55 genes encode proteins acting on cell wall polysaccharides and 6 genes encode peroxidases. After removing the genes encoding CWP<sub>s</sub> identified in the proteomic study, more than 200 new genes remained including 70 genes encoding proteins of unknown function that may play roles in cell elongation.

Less than 100 genes encoding secretory proteins had significant differences in their transcript levels between the two stages of hypocotyl development (43 and 38 having higher level in 5- and 11-day-old hypocotyls respectively). Higher level of transcripts in 5-day-old

hypocotyls was expected for genes encoding GTs and GHs, but not for genes encoding PMEs, that seems important after elongation to perform demethylation of pectins and promote the formation of "egg boxes" with  $\text{Ca}^{2+}$ . Conversely, higher level of transcripts of genes encoding GRPs was expected in 11 days, since they have been involved in repair mechanism of elongating protoxylem primary cell wall. The differences between the expected results and those obtained during this work could be explained either by new roles of the concerned genes, or by special regulations of their expression. It could be a delay between the synthesis of mRNAs and their translation, or certain proteins could be particularly stable. Finally, the activity of proteases should also be considered since they may be involved in the activation of certain enzymes or in their degradation.

Comparison of the transcriptomic data to the proteomic data (chapter 4) revealed three different cases: (i) protein identification with detectable level of transcripts of the corresponding gene on CATMA; (ii) high level of transcript but protein was not identified; (iii) Protein identified while the transcripts of the corresponding gene was not detectable on CATMA.

Several explanations can be proposed for the second case from the viewpoint of proteomics; some proteins are difficult to extract from cell wall; others are difficult to stain with Coomassie blue; still others are difficult to identify by mass spectrometry because of their post-translational modifications or absence of trypsin cut sites. It is also possible that these proteins have a very short half-life or as mentioned above, there is a delay between the synthesis mRNA and active protein. In the third case, the transcripts might have a short half-life, and proteins rather are stable.

In brief, this study revealed: (i) that there exists no systematic correlation between the abundance of transcripts and the presence of the corresponding proteins; (ii) that the post-transcriptional regulation of genes encoding CWP in etiolated hypocotyles seems to be important; (iii) the transcriptomic and proteomic data seem complementary to study the regulation of genes encoding CWPs.

The supplementary data files are given as annexes at the end of the report.

---

## **Cell wall biogenesis of Arabidopsis elongating cells: transcriptomics complements proteomics**

Elisabeth Jamet<sup>1§</sup>, David Roujol<sup>1</sup>, H      San-Clemente<sup>1</sup>, Muhammad Irshad<sup>1</sup>,  
Ludivine Soubigou-Taconnat<sup>2</sup>, Jean-Pierre Renou<sup>2</sup>, Rafael Pont-Lezica<sup>1</sup>

<sup>1</sup>UMR 5546 CNRS-Universit   Toulouse III, P     de Biotechnologie V      , 24 chemin de  
Borde Rouge, BP 42617 Auzeville, 31326 Castanet-Tolosan, France

<sup>2</sup>UMR INRA 1165 - CNRS 8114 – UEVE, 2 rue Gaston Cr      , CP 5708, 91057 Evry,  
France

<sup>§</sup>Corresponding author

Email addresses:

EJ: [jamet@scsv.ups-tlse.fr](mailto:jamet@scsv.ups-tlse.fr)

DR: [roujol@scsv.ups-tlse.fr](mailto:roujol@scsv.ups-tlse.fr)

HSC: [sancle@scsv.ups-tlse.fr](mailto:sancle@scsv.ups-tlse.fr)

MI: [muhammad@scsv.ups-tlse.fr](mailto:muhammad@scsv.ups-tlse.fr)

LST: [soubigou@evry.inra.fr](mailto:soubigou@evry.inra.fr)

JPR: [renou@evry.inra.fr](mailto:renou@evry.inra.fr)

RPL: [lezica@scsv.ups-tlse.fr](mailto:lezica@scsv.ups-tlse.fr)

---

## Abstract

### Background

Plant cell growth is a complex process involving cell division and cell elongation. *Arabidopsis thaliana* hypocotyls undergo a 100-fold length increase mainly by cell elongation. Cell enlargement implicates significant changes in the composition and structure of cell walls. In order to understand cell wall biogenesis during cell elongation, mRNA profiling was made on half- (active elongation) and fully-grown (after growth arrest) etiolated hypocotyls of *Arabidopsis*.

### Results

Transcriptomic studies were performed on two sets of genes. The first set of 846 genes called cell wall genes (CWGs) included genes involved in cell wall biogenesis, that is in synthesis and transport of cell wall components and in their assembly and rearrangements in cell walls. A significant proportion of them have detectable levels of transcripts (54.5%), suggesting that these processes are important throughout hypocotyl elongation and after growth arrest. A second set of 3039 genes named secretory pathway genes (SPGs) was studied to search for new genes encoding secreted proteins possibly involved in cell wall expansion. Based on transcript level, about 200 genes were selected. Seventy of them encoded proteins of yet unknown function. Finally, comparison with results of a proteomic study on the same material revealed that 35 out of the 137 identified proteins were products of the genes having high or moderate level of transcripts. It was surprising to find that about 15% of the genes encoding proteins identified by proteomics showed levels of transcripts below background.

### Conclusions

Members of multigenic families known to be involved in cell wall biogenesis were identified during growth of etiolated hypocotyls as well as about 200 genes that might be required for cell wall assembly and/or rearrangement. No clear correlation was found between the abundance of transcripts (transcriptomic data) and the presence of the proteins (cell wall proteomic data) demonstrating (i) the importance of post-transcriptional events for the regulation of genes during cell elongation and (ii) that transcriptomic and proteomic data are complementary.

---

## Background

Plant cell growth occurs mainly by cell elongation, and may or may not be accompanied by cell division. A meristematic cell might enlarge as much as 50000-fold its initial volume. In this process, membrane surface area and amount of cell wall material increase. The primary cell wall plays an essential role since it should allow turgor-driven increase in cell volume by permitting the incorporation of new cell wall material and rearrangement of existing cell wall. Several plant organs including coleoptiles (*poaceae*), internodes (legumes), and hypocotyls (mung bean, sunflower, and *Arabidopsis thaliana*) have been used to study cell elongation [1]. Environmental signals such as light temperature, and hormones, regulate hypocotyl growth [2-5]. *Arabidopsis* seedlings grown in continuous darkness are a material of choice to analyze the cell elongation process. Indeed, cells of hypocotyls undergo a 100-fold length increase compared to embryo cells [6]. Growth occurs mostly by cell expansion, with little cell division [4, 6-8]. Changes in wall thickness during elongation of *Arabidopsis* hypocotyls were investigated using cryo-field-emission scanning electron microscopy [1]. At the germination stage, cell wall thickening occurs and involves high rates of biosynthesis and deposition of cell wall components. During the elongation stage, cell walls undergo remarkable thinning, requiring extensive polymer disassembly and rearrangement.

Many genes are assumed to be involved in cell wall synthesis and rearrangement to support growth of plant cell walls [9]. They encode cellulose synthases (CESAs), cellulose synthases-like (CSLs), endo-glucanases, xyloglucan endotransglucosylase/hydrolases (XTHs) and expansins. They belong to multigenic families, but the members of each family involved in elongation of hypocotyl cells have not been precisely identified. It is also likely that other genes are important for cell elongation.

In this paper, *Arabidopsis* etiolated hypocotyls were compared at two developmental stages, half- and fully-grown, *i.e.* active elongation *vs.* elongation arrest. Transcript profiling was carried out using CATMA (Complete *Arabidopsis* Transcriptome MicroArray) [10]: (i) to look at the level of transcripts of cell wall genes (CWGs) belonging to families known to be involved in cell wall biogenesis; (ii) to identify genes encoding secreted cell wall proteins (SPGs) having high or moderate level of transcripts; (iii) to reveal differential gene expression between half- and fully-grown etiolated hypocotyls; (iv) and to look at the correlation

---

between transcript abundance and protein presence as revealed by a proteomic study performed on the same material (Irshad *et al.*, submitted for publication).

## Results and Discussion

### Level of transcripts of cell wall genes (CWGs) during hypocotyl elongation

Etiolated hypocotyls were compared at two developmental stages. Five-day-old hypocotyls were approximately half the final size. Growth propagated following an acropetal gradient. After 5-days, the bottom cells were fully elongated, whereas the top cells were only starting elongation [8]. Eleven-day-old hypocotyls had reached their maximum size and the cells were more than 100-fold their initial length [6]. CATMA was used for mRNA profiling. Since one of the major modifications during cell elongation is the addition and rearrangement of cell wall components, a selection of genes possibly involved in cell wall biogenesis was made (Additional data file 1). This selection was called “Cell Wall Genes” (CWGs). It was mainly based on the knowledge of gene families known to be involved in biogenesis of cell walls, that means synthesis and transport of cell wall components and their assembly or rearrangement in cell walls (see Methods). It includes genes encoding proteins involved in substrate generation (nucleotide-sugar inter-conversion pathway, monolignol biosynthesis), polysaccharide synthesis (mainly glycosyl transferases), vesicle trafficking, assembly/disassembly of the wall (glycosyl hydrolases, expansins, carbohydrate esterases, carbohydrate lyases), structural proteins, oxido-reductases involved in cross-linking of wall components (mainly peroxidases and laccases). A few gene families encoding cell wall proteins were also included such as arabinogalactan proteins (AGPs), fasciclin AGPs (FLAs), phytocyanins, multicopper oxidases, pectin methylesterase (PME) inhibitors, and subtilases. Only genes annotated by experts were retained (see Methods) and 26 gene families are represented.

Altogether, 1021 genes were selected among which only 846 were analyzed through CATMA for technical reasons. Some genes had no gene-specific tag (GST) on the microarray, others were not considered because of a poor signal of hybridization to the RNA probe, or of inaccurate duplicates. The level of transcripts was expressed as  $\log_2$  of the mean signal intensity. Values of  $\log_2$  below 6.75 were considered as background. Values between 6.75 and 9 corresponded to a low level of transcripts (1 to 4-fold the background level),

**Table 1. Cell Wall Genes (CWGs) with high levels of transcripts in either 5- or 11-day-old hypocotyls.**

CWGs were selected according to several papers and databases as mentioned in Methods. The intensity of the signal is expressed as  $\log_2$ , the ratios between the levels of transcripts at 11-days and 5-days as well as the Bonferroni p-values are indicated. CWGs are classified by gene family.

Functional classes	AGI number	Predicted or known gene function	5-days	11-days	ratio 11-days / 5-days	p-value
<b>Substrate generation</b>						
nucleotide-sugar interconversion pathway	AT5G39320	UGD1	11.43	11.03	-0.40	1.00
	AT1G08200	AXS2	10.34	9.98	-0.36	1.00
lignin toolbox	AT4G39330	Arath;CAD1(cinnamyl alcohol dehydrogenase)	11.03	10.33	-0.70	7.14E-02
	AT2G30490	Arath;C4H (CYP73A5) (trans-cinnamate 4-hydroxylase)	10.84	10.43	-0.41	1.00
	AT1G76790	Arath;COMT-like8 (caffeic acid O-methyltransferase)	10.37	9.48	-0.89	4.45E-05
	AT5G48930	Arath;HCT (hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyl transferase)	10.24	10.28	0.04	1.00
<b>Synthesis of polysaccharides</b>						
glycosyl transferases (GTs)	AT2G22900	GT34	10.09	9.72	-0.36	1.00
	AT2G28110	GT47 (Group E) (FRA8)	11.39	10.73	-0.67	1.86E-01
	AT5G09870	GT2 (AtCesA5)	10.01	9.11	-0.90	3.26E-05
	AT5G05170	GT2 (AtCesA3)	10.36	9.89	-0.47	1.00
<b>Vesicle trafficking</b>						
	AT1G69460	emp24/gp25L/p24 family protein	10.17	10.55	0.37	1.00
<b>Modification of polysaccharides</b>						
glycosyl hydrolases (GHs)	AT1G64390	GH9 (endoglucanases)	10.68	10.19	-0.49	1.00
	AT4G02290	GH9 (endoglucanases)	11.10	10.08	-1.02	1.57E-07
	AT2G06850	GH16 (AtXTH4)	12.98	12.15	-0.83	6.40E-04
	AT4G14130	GH16 (AtXTH15)	12.72	11.98	-0.74	1.62E-02

	AT4G30290	GH16 (AtXTH19)	11.37	11.40	0.03	1.00
	AT4G30270	GH16 (AtXTH24)	10.84	11.64	0.80	1.63E-03
	AT1G32170	GH16 (AtXTH30)	10.06	9.85	-0.21	1.00
	AT5G56870	GH35 (beta galactosidases) (AtBGAL4)	13.18	11.97	-1.22	5.60E-12
	AT1G45130	GH35 (beta galactosidases) (AtBGAL5)	11.21	10.42	-0.78	3.24E-03
carbohydrate esterases (CEs)	AT2G43050	CE8 (pectin methylesterases)	12.56	10.76	-1.80	0.00
	AT3G49220	CE8 (pectin methylesterases)	11.81	10.99	-0.82	8.65E-04
	AT3G14310	CE8 (pectin methylesterases) AtPME3	10.79	9.90	-0.90	4.05E-05
	AT3G05910	CE13 (pectin acylesterases)	10.44	10.14	-0.3	1.00
carbohydrate lyases (CLs)	AT5G48900	PL1 (pectate lyases) (Group A)	10.92	10.23	-0.70	6.69E-02
	AT1G04680	PL1 (pectate lyases) (Group A)	10.73	10.22	-0.52	1.00
	AT1G67750	PL1 (pectate lyases) (Group A)	10.66	10.42	-0.24	1.00
	AT3G07010	PL1 (pectate lyases) (Group A)	10.26	9.56	-0.70	6.16E-02
expansins	AT5G02260	AtEXPA9 (alpha expansin)	12.71	12.44	-0.28	1.00
	AT1G69530	AtEXPA1 (alpha expansin)	10.62	10.86	0.24	1.00
	AT2G37640	AtEXPA3 (alpha expansin)	10.34	10.52	0.18	1.00
	AT4G28250	AtEXPB3 (beta expansin)	9.07	10.38	1.31	0.00
<b>Structural proteins</b>						
Hyp/Pro rich proteins (H/PRP)	AT2G21140	AtPRP2	11.67	11.15	-0.51	1.00
<b>Oxido-reductases</b>						
peroxidases	AT2G37130	AtPrx21	14.28	12.86	-1.42	0.00
	AT4G21960	AtPrx42	13.17	12.40	-0.77	5.29E-03
	AT1G05260	AtPrx03	10.68	10.18	-0.51	1.00
	AT4G33870	AtPrx48	10.52	9.25	-1.27	0.00
	AT1G71695	AtPrx12	10.17	9.54	-0.63	4.86E-01
	AT5G64120	AtPrx71	10.03	9.85	-0.18	1.00
phytochemicals	AT4G12880	AtEN20 (early nodulin)	10.84	11.19	0.36	1.00
	AT2G32300	AtUCC1 (uclacyanin)	10.61	10.60	-0.02	1.00
SKU-like proteins (multi-copper oxidases)	AT1G76160	SKS5	10.33	9.91	-0.43	1.00
<b>Proteins possibly involved in signaling</b>						
arabinogalactan proteins (AGPs)	AT2G14890	AtAGP9	11.71	11.82	0.12	1.00



fasciclin AGPs (FLAs)	AT5G11740	AtAGP15 (AG peptide)	11.03	11.71	0.68	1.10E-01
	AT5G10430	AtAGP4	10.84	10.65	-0.19	1.00
	AT3G13520	AtAGP12 (AG peptide)	10.77	11.51	0.74	1.50E-02
	AT1G28290	AtAGP31	10.97	11.42	0.46	1.00
	AT1G03870	AtFLA9	12.49	12.25	-0.24	1.00
	AT2G45470	AtFLA8	11.96	11.53	-0.43	1.00
	AT4G12730	AtFLA2	10.85	10.44	-0.41	1.00
<hr/>						
<b>Proteases</b>						
subtilases	AT3G14067	AtSBT1.4	11.85	10.71	-1.15	2.80E-10
	AT2G05920	AtSBT1.8	11.54	10.90	-0.64	4.50E-01
	AT5G67360	AtSBT1.7 (ARA12)	11.45	11.12	-0.33	1.00
	AT5G51750	AtSBT1.3	11.04	10.80	-0.24	1.00
	AT5G59090	AtSBT4.12	10.47	9.49	-0.99	7.99E-07

values between 9 and 10 to a moderate level (4 to 8-fold the background level) and values higher than 10 to a high level (more than 8-fold, and up to 64-fold the background level). The level of transcripts of the 846 genes analyzed was below background for 369 genes (45.8%), low for 343 (40.5%), moderate for 53 (6.3%) and high for 55 (6.5%). Most families are represented by genes with low, moderate or high levels of transcripts.

At both stages, of the 55 genes with high levels of transcripts (Table 1), 11 are involved in the synthesis or transport of cell wall components and 27 are involved in modifications of cell wall components. Altogether, 64.6% and 45.5% respectively of the genes related to synthesis or transport of cell wall components and to their modifications in cell walls respectively have detectable level of transcripts. Two separate phases of growth have been described in *Arabidopsis* dark-grown hypocotyls: an early phase of active synthesis of cell wall polysaccharides up to 3-days after beginning of germination, and a late phase of cell expansion [8]. The former phase results in thicker cell walls which later on become thinner as hypocotyls elongate [1]. Our results show that both synthesis and rearrangement of cell wall components are probably required throughout hypocotyl elongation, and even after growth arrest.

Most of these genes were expected to be transcribed during cell elongation. Genes involved in rearrangement of cell wall components encode GHs such as endoglucanases, XTHs, and beta-galactosidases; carbohydrate esterases such as PMEs; pectin acylesterases; PLs; expansins of the alpha- or beta-type; and peroxidases. However, 21 genes encoding PMEs and 6 genes encoding pectin acylesterases have detectable levels of transcripts among which 4 genes have high levels of transcripts (Additional file 1), which seems contradictory with previous results showing that a low level of pectin esterification, associated to abundance of PMEs and acylesterases, restricted cell elongation in *Arabidopsis* hypocotyls [11]. It should be noted that 21 genes encoding PME inhibitors also have detectable levels of transcripts. The interplay between PME and their inhibitors could regulate the activity of PMEs. Nine genes encoding proteins possibly involved in oxido-reduction reactions have high levels of transcripts (peroxidases, phytochemicals, and protein homologous to SKU5). Six peroxidase genes fall in this category, among which *AT2G37130* (AtPrx21) has one of the highest levels of transcripts. The multiple roles of peroxidases during growth have been reviewed [12]. They could either release reactive oxygen species able to cut polysaccharides

---

or restrict growth by cross-linking structural proteins, monolignols, or polysaccharides and ferulate.

Other CWGs are also well-represented such as those encoding arabinogalactan proteins (AGPs), fasciclin AGPs (FLAs) and subtilases (Table 1). Even if AGPs and FLAs have been shown to be associated with wood formation in poplar [13], their role in cell expansion is not very clear at present. Likewise, nothing is known about the role of proteases during cell elongation. Finally, genes of the COBRA gene family have level of transcripts low or below background. However, COBRA was shown to play an important role in microfibril deposition during rapid elongation and in the orientation of cellulose microfibrils [14]. It may mean that such proteins are stable during hypocotyl elongation.

This work gives clues for understanding the function of members of multigene families either during cell elongation or after its arrest. Indeed, general functions have been proposed for most of these gene families, but only scarce information is available for specific members.

### **Genes encoding secreted proteins with high or moderate level of transcripts in etiolated hypocotyls**

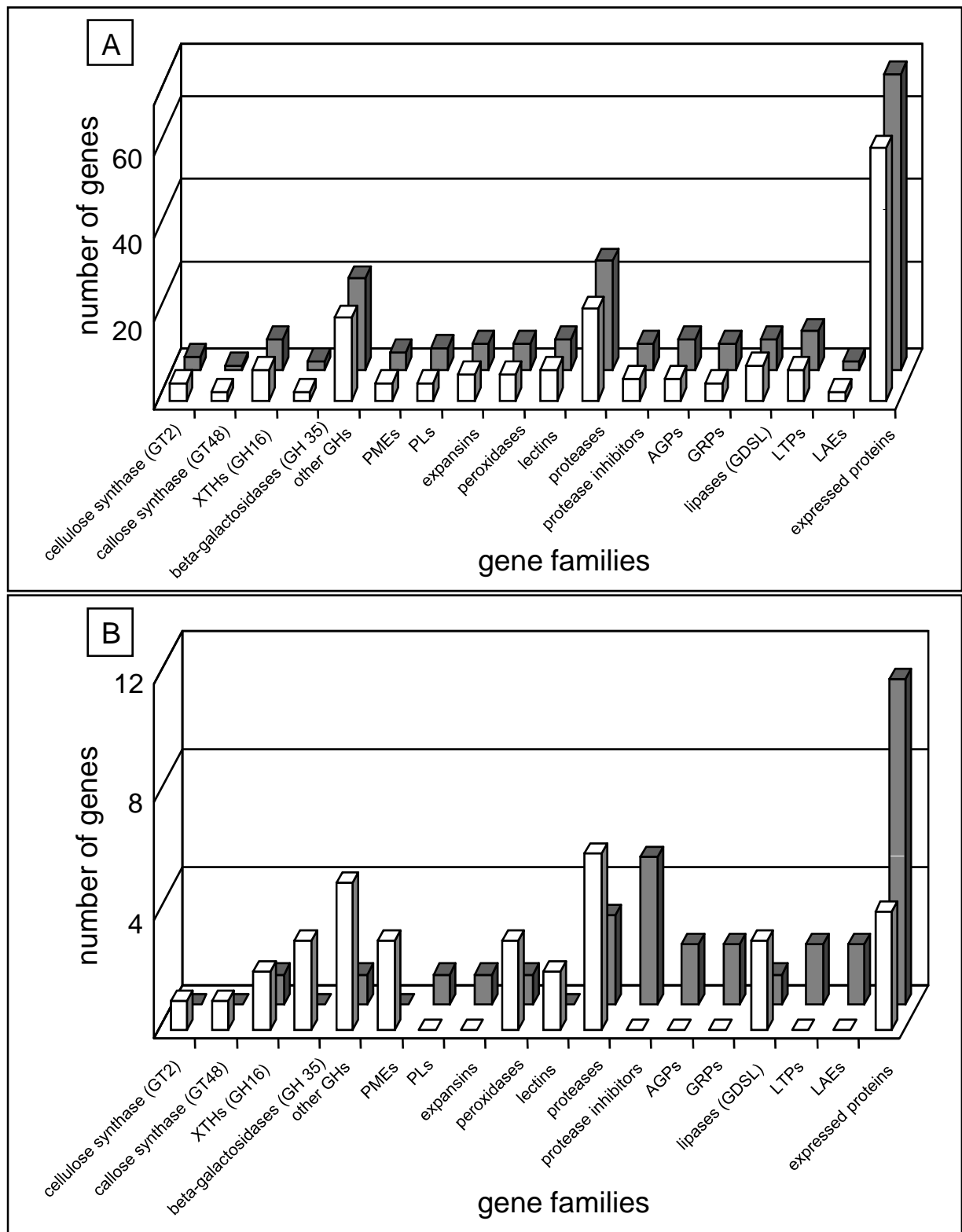
Most of the gene families described above was already known to be involved in cell wall biogenesis. In order to identify new genes that might be involved in cell expansion, a second selection was made based on sub-cellular location of proteins, choosing all the genes encoding proteins with predicted signal peptide thus targeted to the secretion pathway. The 3039 selected genes were ranked by level of transcripts, producing a profile similar to the one obtained with CWGs, namely 1341 genes (44.1 %) above background level, 1363 (44.9%) with a low transcript level, 185 (6.1%) with a moderate level, and 150 (4.9%) with at high level. From this selection, only genes encoding proteins predicted to be located either outside the cell or in the plasma membrane were retained (see Methods). In Additional data file 2, 278 genes named “Secretory Pathway Genes” (SPGs) with moderate or high levels of transcripts are listed and grouped in families according to their predicted functional domains (see Methods). All these protein families were already described in cell wall proteomic studies [15-17]: proteins acting on carbohydrates (55 genes); proteases (25 genes); proteins possibly involved in signaling (23 genes); structural proteins (9 genes); proteins possibly involved in oxido-reduction reactions (20 genes); proteins with interacting domains (20 genes); proteins

related to lipid metabolism (25 genes); miscellaneous proteins (31 genes); proteins of unknown function (70 genes). Main differences lie in the genes encoding proteins possibly involved in signaling since they comprise AGPs, FLAs, and plasma membrane proteins that are difficult to isolate, separate or identify through proteomics [15]. In the same way, the group of proteins of unknown function is very important because 42% of them are predicted to have trans-membrane domains. On the contrary, the group of structural proteins is probably under-represented because of the lack of appropriate GSTs for many of them. Indeed, their repetitive amino acid sequences make the design of specific probes difficult. One should note the abundance of proteases that can be assumed to be essential for protein turnover in tissues undergoing rapid elongation followed by elongation arrest within a short time. They may also be involved in signaling as shown for SDD1 (STOMATAL DENSITY AND DISTRIBUTION 1) and ALE1 (ABNORMAL LEAF EPIDERMIS 1) [18, 19] or in protein maturation as assumed for AtSBT1.7, thus contributing to regulation of cell wall enzyme activities [20]. In addition, there are probably interactions between proteases and protease inhibitors to regulate the proteolytic activities in cell walls. Several of the 70 proteins of yet unknown function have known structural domains (29 genes). Others share domains with other proteins, such as domains of unknown function (DUF), or belong to the so-called uncharacterized protein families (UPF). Many are of particular interest, since they are only present in plants.

Among these 278 SPGs, only the 55 encoding proteins acting on carbohydrates, and the 8 encoding peroxidases or laccases were shown or assumed to contribute to assembly or rearrangement of cell wall components. It means that this study allowed identifying about 200 genes encoding proteins that are candidates to play roles during growth of *Arabidopsis* etiolated hypocotyls. Their functional characterization will be paramount to understand cell wall architecture and assembly during an elongation process.

### **Are there variations in the level of transcripts between half- and fully-grown hypocotyls?**

The comparison of the transcript levels of genes encoding secreted proteins between 5- and 11-day-old hypocotyls showed that 81 genes have a significant differential expression (Additional data file 3). Forty-three and 38 genes have a higher level of transcripts in 5- and 11-day-old hypocotyls respectively. The highest increase (13.8-fold) was found at 11-days for a gene encoding a GRP (AT2G05440). Conversely, the largest decrease (3.5-fold) was



**Figure 1. Overview of SPGs expressed in 5- and 11-day-old hypocotyls.**

**A.** The number of genes belonging to several families of SPGs with moderate and high level of transcripts is shown: white and grey bars are used for 5- and 11-day-old hypocotyls respectively. **B.** Genes of the same families showing significant differences in expression between 5- and 11- day-old hypocotyls are counted: genes with higher level of transcripts in 5-day-old hypocotyls are represented by white bars; genes with higher level of transcripts in 11-day-old hypocotyls are represented by grey bars. AGP: arabinogalactan protein; GRP: glycine-rich protein; GT: glycosyl transferase; GH: glycosyl hydrolase; LAE: late embryogenesis abundant; LTP: lipid transfer protein; PL: polysaccharide lyase; PME: pectin methylesterase; XTH: xyloglucan endotransglucosylase/hydrolase.

observed at 11-days for a gene encoding a putative Asp protease (*AT5G10770*). The number of genes of selected families expressed differentially in both samples is represented in Figure 1. For comparison, the number of genes of the same families having high or moderate levels of transcripts is also represented. All the selected gene families are represented by almost the same number of genes at both developmental stages (Figure 1A). However, there are striking differences when the comparison is done with genes which level of transcripts varies significantly (Figure 1B).

Several genes encoding GTs, GHs, and PME have higher levels of transcripts at 5-days than at 11-days, *i.e.* at a time hypocotyls undergo active elongation. Although GTs and GHs are expected to be expressed during elongation, the role of PMEs appears to be more complex [21]. They have been described as enzymes involved in the arrest of cell elongation acting at the level of esterified pectins to de-esterify methyl residues and to allow binding to  $\text{Ca}^{2+}$  ions, thus promoting the formation of egg boxes. But they could also play roles during elongation in combination with polygalacturonases to favor the fragmentation of pectin chains. The proportion of cells already elongated could also be significant after 5-days of growth. Alternatively, there might be a delay between synthesis of mRNAs, and production of an active protein. This might be the case for PMEs that are usually produced as polyproteins comprising an inhibitor at their N-terminus and an active enzyme at their C-terminus [21].

Genes belonging to other gene families have higher levels of transcripts at 11-days: PLs, expansins, protease inhibitors, AGPs, GRPs, lipid transfer proteins (LTPs), and late-embryogenesis abundant proteins (LEAs). Higher levels of transcripts were not expected for several of them, *e.g.* PLs that are enzymes involved in the deconstruction of pectin backbones [22], expansins that are assumed to be involved in cell expansion [9], and LTPs that have recently been related to cell elongation [23]. On the contrary, GRPs have been shown to be associated to cell walls of xylem and phloem by tissue printing [24]. The great increase in the amount of transcripts of *AT2G05440* is consistent with the development of protoxylem elements that contain GRPs [25]. Concerning AGPs, they are very well represented in the transcriptome of hypocotyls with 12 AGPs having detectable levels of transcripts (Additional data file 1). AGPs are candidates for cell-to-cell communication [26], and were found to be associated to wood formation in poplar [13]. Nothing is known about the role LEAs could play in fully-developed hypocotyls.

**Table 2. Genes with high levels of transcripts in either 5- or 11-day-old hypocotyls for which the encoded proteins were identified in a proteomic study performed on the same plant material.**

Genes are classified by predicted functional domains as described in Methods. The intensity of the signal is expressed as  $\log_2$ , the ratios between the levels of transcripts at 11 and 5 days as well as the Bonferroni p-values are indicated.

Functional class	Gene family	AGI number	Predicted or known gene function	5-days	11-days	ratio 11-days / 5-days	p-value
<b>Proteins acting on carbohydrates</b>	glycoside hydrolase family 16 (xyloglucan endotransglycosidases/hydrolases)	AT2G06850	AtXTH4	12.98	12.15	-0.83	6.40E-04
	glycoside hydrolase family 31	AT1G68560	XYL1 (alpha-xylosidase)	10.60	10.13	-0.47	1.00
	expansin	AT5G02260	AtEXPA9	12.71	12.44	-0.28	1.00
	carbohydrate esterase family 8 (pectin methylesterases)	AT3G14310	AtPME3	10.79	9.9	-0.90	4.05E-05
<b>Proteases</b>	cysteine protease (papain family)	AT4G01610		12.4	12.38	-0.02	1.00
	aspartyl protease	AT1G11910		12.95	12.35	-0.60	1.00
	aspartyl protease	AT5G10770		12.46	9.97	-2.49	0.00E+00
	asparaginyl peptidase	AT4G32940		12.35	10.82	-1.53	0.00E+00
<b>Structural proteins</b>	proline-rich protein (PRP)	AT1G28290		10.97	11.42	0.46	1.00
<b>Proteins involved in oxidation-reduction reactions</b>	peroxidases	AT1G71695	AtPrx12	10.17	9.54	-0.63	4.86E-01
	protein homologous to germin	AT1G09560	GLP5 (subfamily 2, member 1)	10.19	9.79	-0.40	1.00
	phytoecyanin (proteins homologous to blue copper binding proteins)	AT4G12880	plastocyanin	10.84	11.19	0.36	1.00
<b>Proteins with interacting domains</b>							
	protein homologous to lectins	AT1G78850	curculin-like, mannose binding	11.95	11.42	-0.53	1.00

	protein with leucine-rich domains (LRRs)	AT3G20820	expressed protein (LRR domains)	10.49	9.99	-0.50	1.00
	enzyme inhibitor	AT1G73260	serine protease inhibitor	12.00	12.5	0.50	1.00
<hr/>							
<b>Miscellaneous functions</b>		AT2G22170	homologous to <i>Brassica napus</i> dehydration stress-induced protein	10.57	10.54	-0.03	1.00
		AT5G09440	homologous to phosphate-induced proteins (phi)	11.87	10.73	-1.14	3.41E-10
		AT5G15230	gibberellin regulated protein (GASA4)	12.58	13.42	0.84	4.08E-04
		AT2G27190	homologous to purple acid phosphatase	10.67	10.42	-0.25	1.00
<hr/>							
<b>Unknown functions</b>	expressed proteins	AT5G11420	DUF642	11.97	11.68	-0.29	1.00
		AT3G08030	DUF642	12.03	11.69	-0.33	1.00
		AT3G20370	MATH domain	10.31	10.33	0.01	1.00
		AT3G22640	cupin domain	9.52	10.52	1.00	4.06E-07
		AT2G28490	cupin domain	10.09	10.69	0.59	1.00
		AT2G34700	pollen proteins Ole e 1 family	10.14	8.35	-1.80	0.00E+00

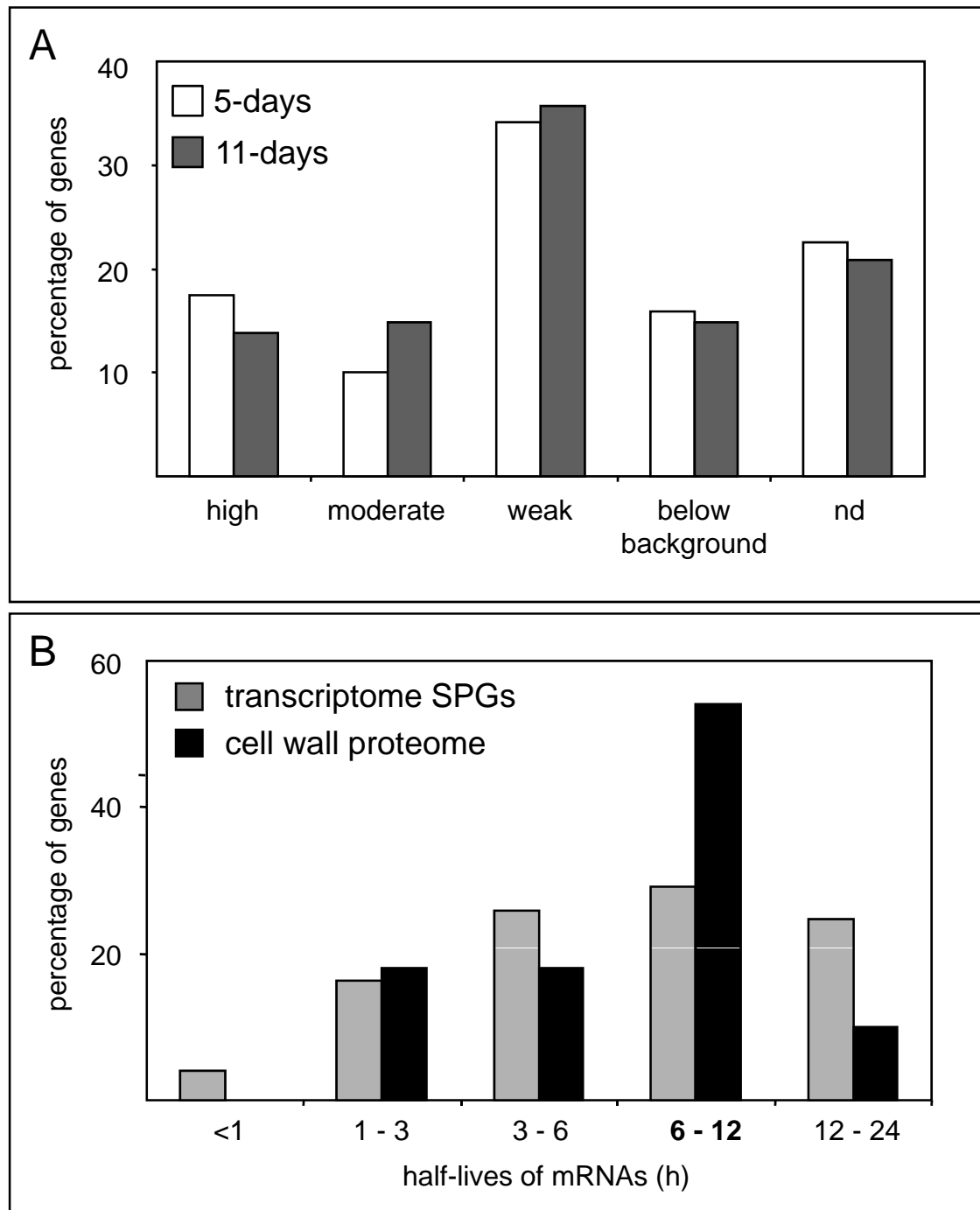


Finally, some gene families have members which level of transcripts is higher either in 5- or in 11-day-old hypocotyls. This is the case of XTHs, peroxidases and proteases. XTHs are expected to play roles in cell elongation by cutting and linking fragments of xyloglucans [27]. However, they have also been identified by proteomics in mature tissues such as rosette leaves [28]. Peroxidases are assumed to cross-link networks of structural proteins and to polymerize monolignols in secondary walls at the end of elongation, thus being involved in the arrest of cell elongation [12]. But they could also be involved in the generation of free radicals able to cut polysaccharides, thus favoring wall extension [12]. The role of proteases in cell walls during elongation has not yet been described. As mentioned above, it can be assumed that they contribute to protein turnover in tissues where many physiological processes occur within a short time. It should be noted that 5 genes encoding protease inhibitors are up-regulated in 11-day-old hypocotyls, suggesting complex regulations of proteolytic activities in cell walls at the end of the arrest of hypocotyls elongation.

These differences in transcript abundance between the two samples should be taken carefully with regard to the possible functions carried out by the proteins, since many other genes from the same families are transcribed in half- and fully-grown hypocotyls, but without significant differences.

### **Transcriptome vs. proteome**

In order to look for the consistency between levels of mRNAs and presence of the corresponding proteins, which are the molecules present in cell walls, a proteomic analysis has been performed on cell walls and the results were compared to those of this transcriptomic analysis. The cell wall proteomes of 5- and 11-days-old hypocotyls were achieved and a total of 137 proteins predicted to be secreted were identified (Irshad *et al.*, submitted for publication). When these 137 proteins were compared to the 278 SPGs with moderate and high levels of transcripts (Additional data file 2), only 35 proteins matched (12.8 %). From the 150 SPGs having high levels of transcripts in etiolated hypocotyls, only 25 (16.6 %) showed the corresponding proteins (Table 2). It was expected that proteomic profiling identified at least the proteins encoded by the highly-transcribed genes. The great inconsistency between the abundance of mRNAs and the presence of the corresponding proteins was surprising, but several reasons may explain this disparity. It is known that CWP extraction and identification can be challenging [15, 16]. Many proteins can remain linked to the polysaccharide matrix, such as the structural proteins [29, 30], or some peroxidases that



**Figure 2. Level of transcripts of genes encoding CWPs identified through proteomics.**

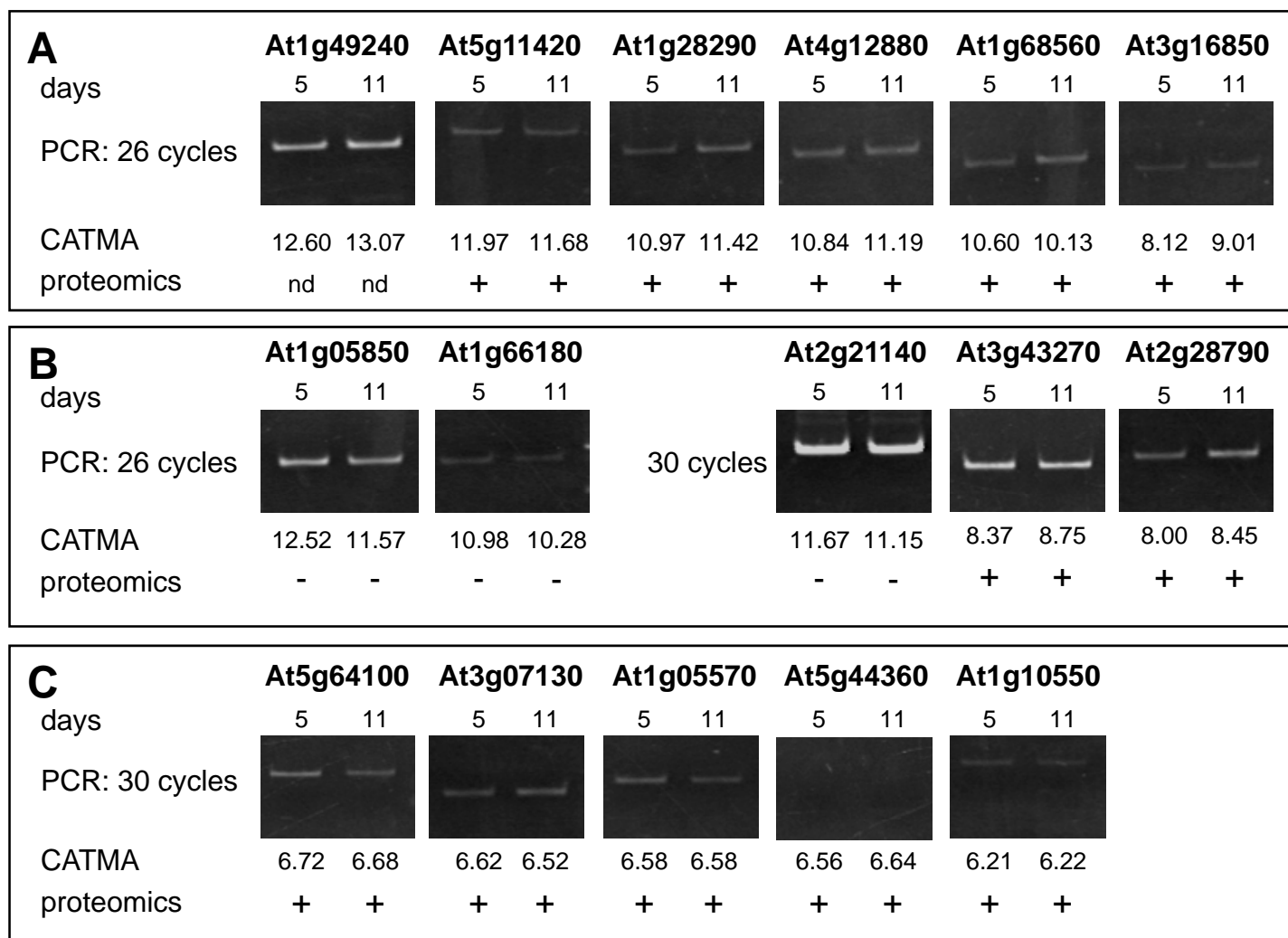
**A.** The levels of transcripts of genes were determined by the CATMA analysis in 5- and 11-day-old hypocotyls (white and dark grey bars respectively). Percentage of genes falling in the three following categories are represented: high transcript level corresponds to  $\log_2$  values of the mean signal intensity higher the 10, moderate to values between 9 and 10, and weak to values between background (6.75) and 9. nd means not determined. **B.** Half-lives of mRNAs (h) corresponding to SPGs having high or moderate levels of transcripts (grey bars) or to proteins identified through cell wall proteomics (black bars). Percentage of genes in each range of half-lives is indicated for each set of gene. Half-lives of mRNAs in cell suspension cultures were from Narsai et al. (2007) [36].

might be strongly bound to pectins [31]. Others are difficult to identify because of their structure, *e.g.* a high level of *O*-glycosylation as for AGPs, which require a special deglycosylation step [32]. Proteins containing few linkages sensitive to trypsin digestion can also escape identification by peptide mass mapping. Finally, low-abundant proteins elude proteomic analyses. For the proteins that were identified without particular problems such as GHs, expansins and proteases, only a few of them correspond to highly-transcribed genes. It indicates that a high level of transcripts is not always correlated with the presence of the protein.

We then looked at the level of transcripts of the 137 genes encoding the proteins identified through proteomics (Figure 2A). The transcript level of 21 genes was not found in the CATMA experiment since some of them have no GST or were eliminated because of poor signals of hybridization to the RNA probe, or of inaccurate duplicates. The levels of transcripts of the 105 remaining genes were surprising, since 34.2 (resp. 35.6%) had low levels of transcripts and 15.8% (resp. 14.9%) had levels of transcripts below background at 5-days (resp. 11-days). However, all the identified proteins are assumed to be the most abundant. This shows that the transcripts could have short half-lives and/or that the proteins could have a low turnover.

To confirm the results obtained with the CATMA analysis, the level of transcripts was estimated by semi-quantitative RT-PCR for several genes corresponding to the three cases described (Figure 3): high or moderate level of transcripts and proteins identified; high level of transcripts and proteins not identified; low or below background level of transcripts and proteins identified. The results obtained for genes having high or moderate level of transcripts are consistent with CATMA results. When the number of PCR cycles was increased to visualize transcripts of some genes having levels of transcripts below background in CATMA analysis (Figure 3C), no band (*At5g44360*) or bands of weak intensity were detected, as compared with *At2g21140* which has a high level of transcripts (Figure 3B). It means that CATMA results are consistent with semi-quantitative RT-PCR results.

Altogether, these results show that there is not a clear correlation between the presence of CWPs as shown by cell wall proteomic analysis and the amount of transcripts of the corresponding genes. The quality of this correlation may depend on genes and/on environmental conditions. For example, the quantification of soluble proteins of yeast at mid-



**Figure 3. RT-PCR analysis of the transcripts of some genes encoding CWPs identified through proteomics.**

**A.** Genes with high or moderate level of transcripts in CATMA analysis and for which the protein was identified by proteomics (+). At1g49240 encoding an actin was used as a control. **B.** Genes with high level of transcripts in CATMA analysis but for which the protein was not identified by proteomics (-); or with low level of transcripts in CATMA analysis for which the protein was identified by proteomics (+). **C.** Genes with transcript level under background in CATMA analysis (lower than 6.75) and for which the proteins were identified by proteomics (+). PCR was run for 26 cycles for genes having high or moderate levels of transcripts in CATMA analysis (A and B), or for 30 cycles for genes having level of transcripts lower than background in CATMA analysis (B and C).

log phase showed that for a given transcript level, protein levels were found to vary by more than 20-fold, whereas for a given protein level, transcript levels were found to vary 30-fold [33]. However, up-regulation of yeast genes in response to glucose or nitrogen limitation was found to be controlled at the transcriptional or post-transcriptional level respectively [34]. In *Arabidopsis* and rice, changes observed in the soluble proteome in response to bacterial challenge were not strictly correlated to changes in transcript levels [35]. These results show that quantitative analysis of transcript levels is not sufficient to infer protein levels. Multilevel analysis must take into account the stability of transcripts, their availability for active translation, as well as the stability of proteins that is certainly essential considering the high number of proteases in cell walls. With regard to transcript stability, data from a recent study aiming at measuring mRNA decay rates in *Arabidopsis* cell suspension cultures [36] were used to look for half-lives of gene transcripts identified through proteomics (Figure 2B). It can be seen that more than half of the proteins (64%) identified by cell wall proteomics correspond to genes having transcripts with rather long half-lives (6-24 h). Conversely, no gene corresponding to proteins identified by cell wall proteomics has transcripts with half-life shorter than 1 h. This distribution differs from that of transcripts of genes having high or moderate level of transcripts since 46% of these genes have half-lives shorter than 6h.

For the particular case of *Arabidopsis* etiolated hypocotyls, only 25 genes both showed a high level of transcripts during cell elongation and after its arrest and encoded proteins identified by cell wall proteomics. They might be considered either as house-keeping genes in this organ or good markers for cell elongation of dark-grown hypocotyls.

## Conclusions

mRNA profiling of the genes potentially involved in cell wall biogenesis (CWGs) in elongating etiolated hypocotyls showed that more than half of them present a detectable level of transcripts. All gene families are expressed. The results suggest that both synthesis and rearrangement of wall components are required throughout hypocotyl elongation. The transcriptomic analysis of genes encoding secreted proteins showed that around 200 new genes might be implicated in this process. Understanding the biochemical and biological functions of these genes might reveal new mechanisms of cell wall expansion or growth arrest, or new functions for the cell wall.

Less than 100 genes encoding secreted proteins had significantly different levels of transcripts between growing and full-elongated hypocotyls. As expected, genes acting on

polysaccharides (GTs or GHs) had higher levels of transcripts at 5-days, whereas others encoding PME or peroxidases were not supposed to have higher level of transcripts during active elongation. Their function during cell elongation should be revisited. On the contrary, several genes encoding LTPs, LEAs, lipases, and protease inhibitors had higher levels of transcripts at 11-days. Their functions remain to be found at the end of the elongation process. Finally, as expected, some genes encoding GRPs were found to have much higher levels of transcripts in fully-grown hypocotyls at a time lignification is an active process. However, since all these genes belong to multigene families, one cannot rule out the fact that a similar function can be shared by several genes.

Finally, a proteomic analysis of the cell walls of half- and fully-grown hypocotyls identified 137 proteins. When the transcript levels of the corresponding genes were analyzed, 15% presented levels of transcripts below CATMA background. On the contrary, only 13% of the genes encoding secreted proteins with high or moderate levels of transcripts corresponded to proteins identified through proteomics. Thus, the comparison between transcript levels and presence of the corresponding proteins suggested that many genes encoding proteins secreted in cell walls are regulated at a post-transcriptional level during the elongation of etiolated hypocotyls. In conclusion, transcriptomic and proteomic data appeared to be complementary to describe the regulation of gene activity.

## Methods

### Plant material

*Arabidopsis thaliana* seedlings (ecotype Columbia 0) were grown in continuous dark on Magenta box on Murashige and Skoog [37] medium supplemented with 2% sucrose. Etiolated hypocotyls were collected after 5- and 11- days of culture.

### Total RNA extraction

Three independent RNA extractions were performed for each sample (5- and 11-day-old hypocotyls). Hypocotyls were cut below cotyledons and above crown with sterile scissors. They were ground in liquid nitrogen in a mortar with a pestle. Extraction of total RNAs was performed using the SV Total RNA Isolation kit according to manufacturer's instructions (Promega France, Charbonnières, France). For each RNA extraction, 750 mg of ground hypocotyls were used. Typically, about 110 µg of total RNAs were obtained.

### Transcriptome studies

Microarray analysis was carried out at the Unité de Recherche en Génomique Végétale (Evry, France), using the CATMA array [10, 38], containing 24,576 GSTs from *Arabidopsis*. RNA samples from three independent biological replicates were pooled. For each comparison, one technical replication with fluorochrome reversal was performed for each RNA sample (*i.e.* four hybridizations per comparison). The reverse transcription of RNA in the presence of Cy3-dUTP or Cy5-dUTP (Perkin-Elmer-NEN Life Science Products), the hybridization of labeled samples to the slides, and the scanning of the slides were performed as described in Lurin *et al.* [39].

### Statistical analysis of microarray data

Experiments were designed with the statistics group of the Unité de Recherche en Génomique Végétale. Statistical analysis was based on two dye swaps (*i.e.* four arrays, each containing 24,576 GSTs and 384 controls) as described [39]. Controls were used for assessing the quality of the hybridizations, but were not included in the statistical tests or the graphic representation of the results. For each array, the raw data comprised the logarithm of median feature pixel intensity at wavelengths 635 (red) and 532 nm (green). To estimate the expression level of each gene, we considered a background value obtained by addition of the average background value to 2 background standard deviation. This background was not subtracted from the data presented in this paper, but was considered for the interpretation of the results. In the following description,  $\log_2$  ratio refers to the differential expression between two conditions. It is either  $\log_2$  (red/green) or  $\log_2$  (green/red) according to the experimental design. Array-by-array normalization was performed to remove systematic biases. First, we excluded spots that were considered badly formed features. Then, we performed global intensity-dependent normalization using the LOESS procedure to correct the dye bias. Finally, for each block, the log ratio median calculated over the values for the entire block was subtracted from each individual log ratio value to correct print tip effects on each metablock. To determine differentially expressed genes, we performed a paired t test on the log ratios, assuming that the variance of the log ratios was the same for all genes. Spots displaying extreme variance (too small or too large) were excluded. The raw p-values were adjusted by the Bonferroni method, which controls the FWER. We considered as being differentially expressed the genes with an FWER of 5%. We use the Bonferroni method (with

a type I error equal to 5%) in order to keep a strong control of the false positives in a multiple-comparison context [40].

### **Data deposition**

Microarray data from this article were deposited at Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>; accession No. E-MEXP-789) and at CATdb (<http://urgv.evry.inra.fr/CATdb/>; Project RS05-11\_Hypocotyls) according to the “Minimum Information About a Microarray Experiment” standards.

### **Semi-quantitative PCR**

cDNA first strand were obtained from total RNAs using 1 µg total RNAs and SuperScript™ II reverse transcriptase (Invitrogen, Carlsbad, San Diego, CA, USA). As a control, the same amount of pig desmin RNA was added in each sample. Quantitative PCR was performed using a Roche lightcycler system (Roche Diagnostics, Meylan, France) according to manufacturer's recommendations. The sequences of oligonucleotide primers used for amplification is provided in Additional data file 5. Using the results from quantitative PCR to determine the number of amplification cycles required to be in a linear range for all genes of interest, semi-quantitative PCR was performed to compare relative amounts of the corresponding mRNAs. The amplified fragments were analyzed by electrophoresis in polyacrylamide gels in standardized conditions. In each case, presence of a fragment of the expected size was checked after staining with ethidium bromide, and the comparison of staining intensities was used for semi-quantification.

### **Bioinformatic analyses**

Sub-cellular localization and length of signal peptides were predicted using PSORT (<http://psort.nibb.ac.jp/>) and TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) [41, 42]. Prediction of transmembrane domains was done with Aramemnon (<http://aramemnon.botanik.uni-koeln.de/>) [43]. Molecular masses and pI values were calculated using the aBi program ([http://www.up.univ-mrs.fr/~wabim/d\\_abim/compo-p.html](http://www.up.univ-mrs.fr/~wabim/d_abim/compo-p.html)). Homologies to other proteins were searched for using BLAST programs (<http://www.ncbi.nlm.nih.gov/BLAST/>) [44]. Identification of protein families and functional domains was performed using MyHits ([http://myhits.isb-sib.ch/cgi-bin/motif\\_scan](http://myhits.isb-sib.ch/cgi-bin/motif_scan)) and InterProScan (<http://www.ebi.ac.uk/InterProScan/>) [45].



---

All the protein families chosen in our CWG list were annotated by experts. GHs and CEs were classified according to the CAZy database (<http://www.cazy.org/CAZY/>) [46] at the Cell Wall Genomics website (<http://cellwall.genomics.purdue.edu/intro/index.html>). The GT77 family was annotated according to Egelund *et al.* [47]. XTHs and expansins were named according to <http://labs.plantbio.cornell.edu/xth/> and <http://www.bio.psu.edu/expansins/index.htm> respectively. AGPs and FLAs were named according to Schultz *et al.* [48], Johnson *et al.* [49], Van Hengels and Roberts [50], and Liu and Mehdy [51]. Proteins homologous to COBRA, LRXs and Hyp/Pro-rich proteins were annotated according to Roudier *et al.* [14], Baumberger *et al.* [52], and Fowler *et al.* [53] respectively. The lignin toolbox was proposed by Raes *et al.* [54]. Peroxidases were named as in the PeroxiBase (<http://peroxidase.isb-sib.ch/index.php>) [55]. Laccases were annotated as in Pourcel *et al.* [56] and McCaig *et al.* [57]. SKU-like proteins and phytocyanins were described in Jacobs and Roe [58], and Nersissian and Shipp [59] respectively. Subtilases are listed at <http://csbdb.mpimp-golm.mpg.de/csbdb/dbcawp/psdb.html>.

---

### List of abbreviations

AGP: arabinogalactan protein; CATMA: Complete *Arabidopsis* Transcriptome MicroArray; CESA: cellulose-synthase; CSL: cellulose-synthase like; CWG: cell wall gene; CWP: cell wall protein; FLA: fasciclin arabinogalactan protein; GH: glycoside hydrolase; GRP: glycine-rich protein; GST: gene-specific tag; GT: glycosyl transferase; LEA: late-embryogenesis abundant protein; LTP: lipid transfer protein; LRR: leucine-rich repeat; LRX: leucine-rich repeat extensin; PL: polysaccharide lyase; PME: pectin methylesterase; SPG: secretory pathway gene; XTH: xyloglucan endotransglucosylase-hydrolase

### Authors' contributions

E.J. conceived the study, participated in its design, coordination, analysis of data, and drafted the manuscript. D.R. and M.I. carried out the culture of plants, RNA extractions and PCR-analysis. HSC was involved in bioinformatic analyses. L.S.-T. and J.-P.R. performed the microarray and statistical analyses of the results. R.P.-L. contributed to the analysis of data and to drafting of the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

The authors thank CNRS (Département des Sciences du Vivant) and Université Paul Sabatier (Toulouse III, France) for financial support. M.I. is a fellow of the Higher Education Commission of Pakistan, Islamabad, and of the French government on the behalf of SFERE. They also thank Dr Hervé Canut for fruitful discussions.

### References

1. Derbyshire P, Findlay K, McCann MC, Roberts K: **Cell elongation in *Arabidopsis* hypocotyls involves dynamic changes in cell wall thickness.** *J Exp Bot* 2007, **58**:2079-2089.
2. Desnos T, Orbovic V, Bellini C, Kronenberger J, Caboche M, Traas J, Hofte H: ***Procuste1* mutants identify two distinct genetic pathways controlling hypocotyl cell elongation, respectively in dark- and light-grown *Arabidopsis* seedlings.** *Development* 1996, **122**:683-693.

3. Collett CE, Harberd NP, Leyser O: **Hormonal interactions in the control of *Arabidopsis* hypocotyl elongation.** *Plant Physiol* 2000, **124**:553-562.
4. Saibo NJ, Vriezen WH, Beemster GT, Van Der Straeten D: **Growth and stomata development of *Arabidopsis* hypocotyls are controlled by gibberellins and modulated by ethylene and auxins.** *Plant J* 2003, **33**:989-1000.
5. Quail PH, Boylan MT, Parks BM, Short TW, Xu Y, Wagner D: **Phytochromes: photosensory perception and signal transduction.** *Science* 1995, **268**:675-680.
6. Gendreau E, Traas J, Desnos T, Grandjean O, Caboche M, Hofte H: **Cellular basis of hypocotyl growth in *Arabidopsis thaliana*.** *Plant Physiol* 1997, **114**:295-305.
7. Raz V, Koornneef M: **Cell division activity during apical hook development.** *Plant Physiol* 2001, **125**:219-226.
8. Refrégier G, Pelletier S, Jaillard D, Hofte H: **Interaction between wall deposition and cell elongation in dark-grown hypocotyl cells in *Arabidopsis*.** *Plant Physiol* 2004, **135**:959-968.
9. Cosgrove DJ: **Growth of the plant cell wall.** *Nat Rev Mol Cell Biol* 2005, **6**:850-861.
10. Crowe ML, Serizet C, Thareau V, Aubourg S, Rouze P, Hilson P, Beynon J, Weisbeek P, van Hummelen P, Reymond P, *et al*: **CATMA: a complete *Arabidopsis* GST database.** *Nucleic Acids Res* 2003, **31**:156-158.
11. Derbyshire P, McCann MC, Roberts K: **Restricted cell elongation in *Arabidopsis* hypocotyls is associated with a reduced average pectin esterification level.** *BMC Plant Biol* 2007, **7**:31.
12. Passardi F, Penel C, Dunand C: **Performing the paradoxical: how plant peroxidases modify the cell wall.** *Trends Plant Sci* 2004, **9**:532-540.
13. Lafarguette F, Leple JC, Dejardin A, Laurans F, Costa G, Lesage-Descauses M-C, Pilate G: **Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood.** *New Phytol* 2004, **164**:107-121.

- 
14. Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GH, Schindelman G, Song S, Baskin TI, Dupree P, Wasteneys GO, Benfey PN: **COBRA, an *Arabidopsis* extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation.** *Plant Cell* 2005, **17**:1749-1763.
  15. Jamet E, Canut H, Boudart G, Pont-Lezica RF: **Cell wall proteins: a new insight through proteomics.** *Trends in Plant Sci* 2006, **11**:33-39.
  16. Boudart G, Minic Z, Albenne C, Canut H, Jamet E, Pont-Lezica R: **Cell wall proteome.** In *Plant Proteomics*. Edited by Samaj S, Thelen J: Springer; 2007: 169-185
  17. Jamet E, Albenne C, Boudart G, Irshad M, Canut H, Pont-Lezica R: **Recent advances in plant cell wall proteomics.** *Proteomics* 2008, **8**:893-908.
  18. Berger D, Altmann T: **A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana*.** *Genes Dev* 2000, **14**:1119-1131.
  19. Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y: **A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants.** *Development* 2001, **128**:4681-4689.
  20. Rautengarten C, Usadel B, Neumetzler L, Hartmann J, Büssis D, Altmann T: **A subtilisin-like serine protease essential for mucilage release from *Arabidopsis* seed coats.** *Plant J* 2008, **in press**.
  21. Micheli F: **Pectin methylesterases: cell wall enzymes with important roles in plant physiology.** *Trends Plant Sci* 2001, **6**:414-419.
  22. Wong D: **Enzymatic deconstruction of backbone structures of the ramified regions in pectins.** *Protein J* 2007.
-

- 
23. Nieuwland J, Feron R, Huisman BA, Fasolino A, Hilbers CW, Derksen J, Mariani C: **Lipid transfer proteins enhance cell wall extension in tobacco.** *Plant Cell* 2005, **17**:2009-2019.
  24. Ye ZH, Song YR, Marcus A, Varner JE: **Comparative localization of three classes of cell wall proteins.** *Plant J* 1991, **1**:175-183.
  25. Ringli C, Keller B, Ryser U: **Glycine-rich proteins as structural components of plant cell walls.** *Cell Mol Life Sci* 2001, **58**:1430-1441.
  26. McCabe PF, Valentine TA, Forsberg LS, Pennell RI: **Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot.** *Plant Cell* 1997, **9**:2225-2241.
  27. Rose JK, Braam J, Fry SC, Nishitani K: **The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature.** *Plant Cell Physiol* 2002, **43**:1421-1435.
  28. Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerré-Tugayé M-T, Pont-Lezica R: **Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics.** *Proteomics* 2005, **5**:212-221.
  29. Brady JD, Sadler IH, Fry SC: **Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: a new potential cross-link.** *Biochem J* 1996, **315**:323-327.
  30. Schnabelrauch LS, Kieliszewski MJ, Upham BL, Alizedeh H, Lamport DTA: **Isolation of pI 4.6 extensin peroxidase from tomato cell suspension cultures and identification of Val-Tyr-Lys as putative intermolecular cross-link site.** *Plant J* 1996, **9**:477-489.
  31. Shah K, Penel C, Gagnon J, Dunand C: **Purification and identification of a Ca<sup>2+</sup>-pectate binding peroxidase from *Arabidopsis* leaves.** *Phytochemistry* 2004, **65**:307-312.

- 
32. Schultz CJ, Ferguson KL, Lahnstein J, Bacic A: **Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*.** *J Biol Chem* 2004, **279**:455103-445511.
  33. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19**:1720-1730.
  34. Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MM, Pronk JT, Slijper M, Heck AJ: **Proteome analysis of yeast response to various nutrient limitations.** *Mol Syst Biol* 2006, **2**:2006.0026.
  35. Jones AM, Thomas V, Truman B, Lilley K, Mansfield J, Grant M: **Specific changes in the *Arabidopsis* proteome in response to bacterial challenge: differentiating basal and R-gene mediated resistance.** *Phytochemistry* 2004, **65**:1805-1816.
  36. Narsai R, Howell K, Millar A, O'Toole N, Small I, Whealan J: **Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*.** *Plant Cell* 2007, **19**:3418-3436.
  37. Murashige T, Skoog F: **A revised medium for rapid growth and bioassays with tobacco tissue culture.** *Physiol Plant* 1962, **15**:473-497.
  38. Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao RP, Bitton F, Caboche M, Cannoot B, *et al*: **Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications.** *Genome Res* 2004, **14**:2176-2189.
  39. Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, *et al*: **Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis.** *Plant Cell* 2004, **16**:2089-2103.
  40. Ge Y, Dudoit S, Speed TP: **Resampling-based multiple testing for microarray data analysis.** *Test* 2003, **12**:1-77.

- 
41. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Prot Eng* 1997, **10**.
  42. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**:1005-1016.
  43. Schwacke R, Schneider A, van der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flugge UI, Kunze R: **ARAMEMNON, a novel database for Arabidopsis integral membrane proteins.** *Plant Physiol* 2003, **131**:16-26.
  44. Altschul SF, Gish W, Miller W, Myers EW, Lipman D: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
  45. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier.** *Nucleic Acids Res* 2005, **33**:W116-120.
  46. Coutinho PM, Henrissat B: **Carbohydrate-active enzymes: an integrated database approach.** In *Recent advances in carbohydrate Bioengineering*. Edited by Gilbert HJ, Davies B, Henrissat B, Svensson B. Cambridge: The Royal Society of Chemistry; 1999: 3-12
  47. Egelund J, Skjot M, Geshi N, Ulvskov P, Petersen BL: **A complementary bioinformatics approach to identify potential plant cell wall glycosyltransferase-encoding genes.** *Plant Physiol* 2004, **136**:2609-2620.
  48. Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YG, Antony Bacic A: **Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case.** *Plant Physiol* 2002, **129**:1448-1463.
  49. Johnson KL, Jones BJ, Schultz CJ, Bacic A: **Non enzymic cell wall (glyco)proteins.** In *The plant cell wall. Volume 8*. Edited by Rose JK. Boca Raton: CRC Press LLC; 2003.[111-154 (Series Editor)]

- 
50. van Hengel AJ, Roberts K: **AtAGP30, an arabinogalactan-protein in the cell walls of the primary root, plays a role in root regeneration and seed germination.** *Plant J* 2003, **36**:256-270.
  51. Liu C, Mehdy M: **A nonclassical arabinogalactan protein gene highly expressed in vascular tissues, AGP31, is transcriptionally repressed by methyl jasmonic acid in Arabidopsis.** *Plant Physiol* 2007, **145**:863-874.
  52. Baumberger N, Doesseger B, Guyot R, Diet A, Parsons RL, Clark MA, Simmons MP, Bedinger P, Goff SA, Ringli C, Keller B: **Whole-genome comparison of leucine-rich repeat extensins in Arabidopsis and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade.** *Plant Physiol* 2003, **131**:1313-1326.
  53. Fowler TJ, Bernhardt C, Tierney ML: **Characterization and expression of four proline-rich cell wall protein genes in Arabidopsis encoding two distinct subsets of multiple domain proteins.** *Plant Physiol* 1999, **121**:1081-1092.
  54. Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W: **Genome-wide characterization of the lignification toolbox in Arabidopsis.** *Plant Physiol* 2003, **133**:1051-1071.
  55. Bakalovic N, Passardi F, Ioannidis V, Cosio C, Penel C, Falquet L, Dunand C: **PeroxiBase: a class III plant peroxidase database.** *Phytochemistry* 2006, **67**:534-539.
  56. Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I: **TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat.** *Plant Cell* 2005, **17**:2966-2980.
  57. McCaig BC, Meagher RB, Dean JF: **Gene structure and molecular analysis of the laccase-like multicopper oxidase (LMCO) gene family in Arabidopsis thaliana.** *Planta* 2005, **221**:619-636.



- 
58. Jacobs J, Roe JL: **SKS6, a multicopper oxidase-like gene, participates in cotyledon vascular patterning during *Arabidopsis thaliana* development.** *Planta* 2005, **222**:652-666.
59. Nersissian AM, Shipp EL: **Blue copper-binding domains.** *Adv Protein Chem* 2002, **60**:271-340.

### **Additional data files**

#### **Additional data file 1.**

Table 1. Cell Wall Genes (CWGs) with detectable levels of transcripts in 5- and 11-day-old hypocotyls.

#### **Additional data file 2.**

Table 2. Secretory Pathway Genes (SPGs) with moderate or high levels of transcripts in 5- and 11-day-old hypocotyls.

#### **Additional data file 3.**

Table 3. Secretory Pathway Genes (SPGs) with modulated levels of transcripts in 5- and 11-day-old hypocotyls.

#### **Additional data file 4.**

Table 4. Proteins extracted and identified by mass spectrometry from purified cell walls of 5- and 11-day-old etiolated hypocotyls of *A. thaliana*.

#### **Additional data file 5.**

Table 5. Nucleotide primers used for PCR amplifications

# **CHAPTER 6**

## **CONCLUSIONS AND PERSPECTIVES**

---

**French version**

La première étape de ce travail a consisté en l'identification des protéines présentes dans les parois d'hypocotyles étiolés d'*Arabidopsis* en élongation ou après la fin de leur élongation en utilisant une approche de protéomique. Pour éviter la perte de protéines au cours de la purification des parois et la contamination par des protéines intracellulaires, nous avons dû établir un nouveau protocole de purification de parois. Les protéines ont été extraites avec des solutions salines avant d'être séparées par électrophorèse 1D. nous avons pu identifier un grand nombre de protéines par cartographie peptidique massique en utilisant la spectrométrie de masse de type MALDI-TOF. Une grande proportion de ces protéines avait un peptide signal prédit par bioinformatique. Une proportion aussi élevée de protéines secrétées n'avait jamais été obtenue dans les études de protéomique de la paroi précédemment publiées. Ces protéines ont été appelées protéines pariétales (CWP). Notre méthode de purification a été récemment utilisée au laboratoire avec succès pour des suspensions cellulaires de tabac (S. Muzammil, M2R, Université Toulouse III, 2007-2008). Ceci montre qu'il doit être possible de l'adapter à différents organes et à d'autres plantes. Des expériences préliminaires indiquent également qu'une fraction de protéines interagissant avec les pectines peut être extraite de telles parois (G. Boudart, résultats non publiés). Des expériences supplémentaires sont maintenant nécessaires pour caractériser ces protéines. Enfin, bien que toutes les protéines appelées CWP aient un peptide signal prédit, leur localisation sub-cellulaire doit encore être vérifiée par immunolocalisation ou par localisation de protéines fusion avec la GFP (*Green Fluorescent Protein*) *in planta*.

Concernant la protéomique des parois végétales, un autre apport majeur de ce travail est le développement d'une nouvelle méthode de séparation des protéines pariétales. Mettant à profit le fait que la plupart des CWPs sont basiques, nous avons utilisé une chromatographie échangeuse de cations à pH acide comme première étape de séparation, mimant ainsi les parois végétales qui sont chargées négativement. La seconde étape de séparation a consisté en une électrophorèse 1D. L'efficacité de cette nouvelle stratégie de séparation 2D a permis d'identifier deux fois plus de protéines que la simple électrophorèse 1D. De plus, les identifications par cartographie peptidique massique ont été de meilleure qualité puisque le pourcentage de recouvrement des séquences des protéines était significativement plus élevé. Enfin, il a été possible de procéder à une semi-quantification des protéines. Cette nouvelle stratégie pourra être employée pour des études biochimiques et structurales futures de

nombreuses protéines pariétales. En effet, il est maintenant possible de les visualiser après seulement deux étapes de séparation.

Cette analyse protéomique a permis l'identification de 137 protéines pariétales à partir des hypocotyles étiolés d'*Arabidopsis*. 51 de ces protéines ont été identifiées pour la première fois par protéomique, ce qui suggère une expression organe-spécifique des gènes correspondants. 84 de ces protéines étaient communes aux deux stades de développement des hypocotyles (élongation active vs élongation terminée) alors que 36 et 17 ont été identifiées seulement à 5 jours (élongation active) et à 10 jours (élongation terminée) respectivement. La semi-quantification des protéines a montré que le niveau de 34 protéines différaient entre les deux stades de développement. D'un point de vue quantitatif, les résultats de cette étude suggèrent que les parois des hypocotyles subissant une élongation sont métaboliquement plus actives que celles des hypocotyles dont l'élongation est terminée.

Parmi les protéines identifiées dans cette étude, celles qui étaient attendues car jouant un rôle soit dans le réarrangement des polysaccharides au cours de l'élongation cellulaire (XTH, PG, expansines), ou encore dans la formation des boîtes à œufs pectiques (PME), des réseaux de protéines structurales et de polyphénols (peroxydases, laccases) en fin d'élongation ont été trouvées. Cependant, certaines d'entre elles ont également été trouvées tout au long du développement des hypocotyles. Ce résultat inattendu suggère l'existence de mécanismes de régulation au niveau de la durée de vie des protéines, ou de l'activité biologique des protéines. Une autre hypothèse est que ces protéines seraient impliquées dans d'autres processus encore à découvrir. Plusieurs protéines liées au métabolisme des lipides ont été identifiées. Aucune fonction ne leur a encore été clairement associée bien qu'il a été suggéré qu'elles pourraient contribuer à la formation de la cuticule particulièrement épaisse des hypocotyles étiolés. De la même manière, les protéines de fonction inconnue sont des bons candidats pour contribuer à l'élongation cellulaire. Des études supplémentaires incluant l'obtention de plantes présentant des niveaux d'expression des gènes correspondants modifiés sont nécessaires pour comprendre la fonction de tous ces gènes.

En plus de l'identification de gènes candidats pour jouer des rôles au cours de l'élongation cellulaire, la protéomique permet d'obtenir une vision globale des protéines d'un compartiment cellulaire, et notamment d'identifier en même les partenaires d'interactions protéine-protéine. Par exemple, des enzymes et leurs inhibiteurs ont été trouvés tels que des

protéases et des inhibiteurs de protéases, des PME et des PME1, des PG et des PGIP. Ceci montre que l'activité biologique d'une protéine peut être régulée par une maturation par une protéase, une interaction avec un inhibiteur, et/ou la protéolyse. L'abondance des protéases et de leurs inhibiteurs dans les parois des hypocotyles doit être soulignée. Des tests fonctionnels devraient permettre de montrer l'existence de ces activités de régulation.

L'étude protéomique des hypocotyles étiolés a été complétée par une étude de transcriptomique réalisée sur le même matériel. Dans un premier temps, des niveaux de transcrits significatifs ont été détectés pour plus de la moitié des gènes supposés être impliqués dans la biogenèse des parois (CWG) avec une proportion importante de gènes codant des protéines impliquées dans la synthèse et le transport des constituants pariétaux. Nos résultats indiquent que la synthèse, le transport et le réarrangement des constituants pariétaux interviendraient tout au long du développement des hypocotyles étiolés. Ce résultat n'était pas attendu dans la mesure où deux phases successives de modification des parois avaient été décrites : une synthèse active de constituants pariétaux ayant pour conséquence une augmentation de l'épaisseur de parois, suivie par une phase de réduction de l'épaisseur des parois pendant l'élongation cellulaire. Dans un second temps, nous avons recherché des gènes codant des protéines sécrétées ayant des niveaux de transcrits modérés ou élevés (SPG). C'est ainsi que nous avons identifié plus de 200 gènes qui pourraient jouer un rôle important pendant le développement des hypocotyles à l'obscurité. Comprendre leurs fonctions pourrait révéler de nouveaux mécanismes importants pour l'élongation cellulaire, l'arrêt de la croissance, ou de nouvelles fonctions pour les parois.

La dernière partie de l'étude a été consacrée à la comparaison entre les données transcriptomique et protéomique. A notre surprise, environ 15% des protéines identifiées par protéomique correspondent à des gènes dont les niveaux de transcrits sont inférieurs au bruit de fond CATMA. Au contraire, nous avons identifié des protéines correspondant à seulement 13% des gènes codant des protéines sécrétées et ayant des niveaux de transcrits modérés ou élevés. Ceci montre que les données de protéomique et de transcriptomique ne sont pas systématiquement cohérentes. Une grande proportion de gènes codant des protéines sécrétées semble être régulée à un niveau post-transcriptionnel dans les hypocotyles étiolés. Ainsi, les niveaux de protéines ne peuvent pas être déduits des niveaux de transcrits. Les deux types de données devraient être utilisées pour mieux décrire la régulation de l'activité des gènes pendant l'élongation des hypocotyles. Des expériences supplémentaires devraient être effectuées pour

---

déterminer la demi-vie des transcrits, leur capacité à être transcrits, et déterminer la demi-vie des protéines. Ceci constitue un champ de recherches complètement nouveau dans lequel peu de données expérimentales existent.

---

**English version**

The first step of this work consisted in the identification of proteins present in cell walls of elongating and fully-elongated hypocotyls of *Arabidopsis* by a proteomic approach using a destructive method. To avoid losing CWPs during the purification of cell walls and contamination by intracellular proteins, we had to establish **a new protocol for purification of cell walls**. Proteins were extracted with salt solutions and separated by 1D-E. We were able to identify a high number of proteins by peptide mass mapping using MALDI-TOF mass spectrometry, and a high proportion of them were predicted to be secreted by bioinformatics. Such a high proportion of secreted proteins was never found before in cell wall proteomic studies. These proteins were called CWPs. This method of purification of cell walls has been recently successfully applied to tobacco cell suspension cultures in the laboratory (S. Muzammil, M2R, Toulouse III University, 2007-2008). This result shows that it should be possible to adapt it to different plant organs and to other plants. Preliminary experiments indicate that a fraction of proteins interacting with pectins can also be released from such purified cell walls (G. Boudart, unpublished results). Additional experiments are now required to characterize these proteins. Finally, although all the so-called CWPs are predicted to have a signal peptide, their sub-cellular localization should be confirmed either by immunolocalization or localization of Green Fluorescent Protein (GFP)-fusions *in planta*.

Another major achievement of this study is the development of **a new separation method for CWPs**. Since most CWPs are basic proteins, cation exchange chromatography with FPLC was used at acidic pH as a first step of separation, thus mimicking cell walls that are negatively charged *in planta*. The second step of separation consisted in 1D-E. The efficiency of this new 2D-separation strategy allowed the identification of twice as much proteins as 1D-E alone. In addition, it permitted a better identification with a higher number of matching peptide by MALDI-TOF MS peptide mass mapping, and semi-quantification of proteins. This method also proved to be a powerful tool for future biochemical and structural studies of CWPs.

**This proteomic analysis** allowed the identification of 137 CWPs from *Arabidopsis* etiolated hypocotyls, thus providing a global image of protein dynamics during cell elongation. Fifty-one CWPs were identified for the first time by cell wall proteomics, supporting the idea of organ-specific expression of CWPs. Eighty-four CWPs were common to both stages of hypocotyl development (active elongation *vs* after elongation arrest) whereas

36 and 17 were only identified at 5-days (active elongation) and 11-days (after elongation arrest) respectively. Semi-quantification showed that the amount of 34 CWPs was different at one or the other developmental stage. Altogether, from a quantitative point of view, the results of this study suggest that cell walls of hypocotyls undergoing elongation are metabolically more active than those of hypocotyls after elongation arrest.

Among the CWPs identified in this study, the proteins expected to play roles in rearrangements of cell wall polysaccharides during cell wall elongation (XTHs, PGs, expansins), in formation of pectin egg boxes (PME) or in cross-linking of proteins or polyphenols (peroxidases, laccases) after the end of elongation were found. However, some of them were also found all along the growth of hypocotyls. This unexpected result suggests the existence of regulatory mechanisms at the level of protein half-lives or protein biological activity. Alternatively, such proteins might be involved in other processes still to be discovered. Several proteins related to lipid metabolism were identified. They have yet no clear function in cell walls although it has been suggested that they could be involved in the synthesis of the thick cuticle that surround etiolated hypocotyls. In the same way, proteins of yet unknown function are good candidates to contribute somehow to cell elongation. Additional analyses including study of plants with modified level of expression of these genes should allow understanding the function of these genes.

Apart from identifying new candidates for performing roles in cell wall elongation, proteomics provides a unique way to get a global view of CWPs, thus putting together proteins able to interact. For example, enzymes and the corresponding inhibitors were found such as proteases and protease inhibitors, PMEs and PMEIs, PG and PGIPs. It suggests that the biological activity of a protein might be regulated by maturation by a protease, interaction with an inhibitor, and/or proteolysis. The great abundance of proteases and of their inhibitors in the cell wall proteome of hypocotyls should be emphasized. Functional tests should allow demonstration of these regulatory activities.

The proteomic study of etiolated hypocotyls was complemented by a **transcriptomic survey** performed with the same material. In a first analysis, significant transcript levels were detected for more than one half of the genes assumed to be involved in cell wall biogenesis (CWGs) with a significant proportion of genes involved in synthesis and transport of cell wall components. Our results indicate that synthesis, transport and



---

rearrangements of cell wall components occur during hypocotyl elongation and growth arrest. This result contrasts with what was previously reported about the existence of two phases of cell wall modification during hypocotyl elongation: active synthesis of cell wall polysaccharides with increase in cell wall thickness, followed by thinning down of cell walls during cell elongation. In a second analysis, we focused on genes encoding secreted proteins (SPGs) and we could identify more than 200 genes having moderate or high level of transcripts. All of them are candidates to play important roles, not yet described, during hypocotyl development. Understanding their function might reveal new mechanisms important for cell wall expansion, growth arrest, or new function for cell walls.

The last part of our study consisted in the **comparison between proteomic and transcriptomic data**. To our surprise, about 15% of the proteins identified by proteomics corresponded to genes having level of transcripts below CATMA background level. Conversely, we identified proteins corresponding to only 13% of the genes encoding secreted proteins and having moderate or high level of transcripts. It means that proteomic and transcriptomic data are not consistent in all cases. A great proportion of genes encoding secreted proteins seems to be regulated at a post-transcriptional level in etiolated hypocotyls. Hence, protein abundance cannot be inferred from transcriptomic data. Both kinds of data should be better used to describe regulation of gene activity during hypocotyl elongation. Additional experiments should be performed to know more about half-lives of transcripts, their ability to be translated, and about half-lives of proteins in cell walls. This is a completely new field of research in which there is no data.

## **LITERATURE CITED**

---

LITERATURE CITED

- [1] Altschul SF, Gish W, Miller W, Myers EW, Lipman D: **Basic local alignment search tool**. *Journal of molecular biology* 1990, **215**:403-410.
- [2] **Aramemnon** [<http://aramemnon.botanik.uni-koeln.de/>]
- [3] Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Hofte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE: **Molecular analysis of cellulose biosynthesis in *Arabidopsis***. *Science* 1998, **279**:717-720.
- [4] Bakalovic N, Passardi F, Ioannidis V, Cosio C, Penel C, Falquet L, Dunand C: **PeroxiBase: a class III plant peroxidase database**. *Phytochemistry* 2006, **67**:534-539.
- [5] Baumberger N, Doesseger B, Guyot R, Diet A, Parsons RL, Clark MA, Simmons MP, Bedinger P, Goff SA, Ringli C, Keller B: **Whole-genome comparison of leucine-rich repeat extensins in *Arabidopsis* and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade**. *Plant Physiology* 2003, **131**:1313-1326.
- [6] Bayer EM, Bottrill AR, Walshaw J, Vigouroux M, Naldrett MJ, Thomas CL, Maule AJ: ***Arabidopsis* cell wall proteome defined using multidimensional protein identification technology**. *Proteomics* 2006, **6**:301-311.
- [7] Berger D, Altmann T: **A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana***. *Genes & development* 2000, **14**:1119-1131.
- [8] Billa E, Koukios EG, Month B: **Investigation of lignins structure in cereal crops by chemical degradation methods**. *Polymer degradation and stability* 1998, **59**:71-75.
- [9] Boerjan W, Ralph J, Baucher M: **Lignin biosynthesis**. *Annual review of plant biology* 2003, **54**:519-546.
- [10] Bolanos L, Esteban E, De Lorenzo C, Fernandez-Pascual M, De Felipe MR, Garate A, Bonilla I: **Essentiality of Boron for Symbiotic Dinitrogen Fixation in Pea (*Pisum sativum*) Rhizobium Nodules**. *Plant Physiology* 1994, **104**:85-90.
- [11] Bolwell GP: **Dynamic aspects of the plant extracellular matrix**. *International Review of Cytology* 1993, **146**:261-324.
- [12] Bonilla I, Mergold-Villasenor C, Campos ME, Sanchez N, Perez H, Lopez L, Castrejon L, Sanchez F, Cassab GI: **The aberrant cell walls of boron-deficient**

- 
- bean root nodules have no covalently bound hydroxyproline-/proline-rich proteins.** *Plant Physiology* 1997, **115**:1329-1340.
- [13] Bordenave M, Goldberg R: **Immobilized and free apoplastic pectinmethylesterases in mung bean hypocotyl.** *Plant Physiology* 1994, **106**:1151-1156.
- [14] Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerré-Tugayé MT, Boudet A, Pont-Lezica R: **Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: a critical analysis.** *Electrophoresis* 2003, **24**:3421-3432.
- [15] Borner GH, Sherrier DJ, Weimar T, Michaelson LV, Hawkins ND, Macaskill A, Napier JA, Beale MH, Lilley KS, Dupree P: **Analysis of detergent-resistant membranes in *Arabidopsis*.** Evidence for plasma membrane lipid rafts. *Plant Physiology* 2005, **137**:104-116.
- [16] Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerré-Tugayé M-T, Pont-Lezica R: **Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics.** *Proteomics* 2005, **5**:212-221.
- [17] Boudart G, Minic Z, Albenne C, Canut H, Jamet E, Pont-Lezica R: **Cell wall proteome.** In *Plant Proteomics*. Edited by Samaj S, Thelen J: Springer; 2007: 169-185
- [18] Boudet AM, Kajita S, Grima-Pettenati J, Goffner D: **Lignins and lignocellulosics: a better control of synthesis for new and improved uses.** *Trends in Plant Science* 2003, **8**:576-581.
- [19] Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.** *Analytical Biochemistry* 1976, **72**:248-254.
- [20] Brady JD, Sadler IH, Fry SC: **Di-isodityrosine, a novel tetrametric derivative of tyrosine in plant cell wall proteins: a new potential cross-link.** *The Biochemical journal* 1996, **315**:323-327.
- [21] Brownlee C: **Role of the extracellular matrix in cell-cell signalling: paracrine paradigms.** *Current Opinion in Cell Biology* 2002, **5**:396-401.
- [22] Buchanan BB, Gruissem W, Jones RL: **Biochemistry and molecular biology of plants.** Rockville: American Society of Plant Physiologists; 2000.
- [23] Buschhaus C, Herz H, Jetter R: **Chemical composition of the epicuticular and intracuticular wax layers on the adaxial side of *Ligustrum vulgare* leaves.** *The New Phytologist* 2007, **176**:311-316.
-

- 
- [24] Carpita N, Gibeaut D: **Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth.** *Plant Journal* 1993, **3**:1-30.
- [25] Cassab GI, Varner JE: **Cell wall proteins.** *Annual review of Plant Physiology and Plant Molecular Biology* 1988, **39**:321-353.
- [26] Cassab GI: **Plant cell wall proteins.** *Annual Review of Plant Physiology and Plant Molecular Biology* 1998, **49**:281-309.
- [27] Chanliaud E, De Silva J, Strongitharm B, Jeronimidis G, Gidley MJ: **Mechanical effects of plant cell wall enzymes on cellulose/xyloglucan composites.** *The Plant Journal* 2004, **38**:27-37.
- [28] Charmont S, Jamet E, Pont-Lezica R, Canut H: **Proteomic analysis of secreted proteins from *Arabidopsis thaliana* seedlings: improved recovery following removal of phenolic compounds.** *Phytochemistry* 2005, **66**:453-461.
- [29] Chivasa S, Ndimba B, Simon W, Robertson D, Yu X-L, Knox J, Bolwell P, Slabas A: **Proteomic analysis of the *Arabidopsis thaliana* cell wall.** *Electrophoresis* 2002, **23**:1754-1765.
- [30] Collett CE, Harberd NP, Leyser O: **Hormonal interactions in the control of *Arabidopsis* hypocotyl elongation.** *Plant Physiology* 2000, **124**:553-562.
- [31] Cominelli E, Sala T, Calvi D, Gusmaroli G, Tonelli C: **Overexpression of the *Arabidopsis* AtMyb41 gene alters cell expansion and leaf surface permeability.** *The Plant Journal* 2008, **58**:53-64.
- [32] Cosgrove DJ: **Cell wall loosening by expansins.** *Plant Physiology* 1998, **118**:333-339.
- [33] Cosgrove DJ: **Enzymes and other agents that enhance cell wall extensibility.** *Annual Review of Plant Physiology and Plant Molecular Biology* 1999, **50**:391-417.
- [34] Cosgrove DJ: **Growth of the plant cell wall.** *Nature Reviews. Molecular Cell Biology Biol* 2005, **6**:850-861.
- [35] Cosgrove DJ: **Relaxation in a high-stress environment: the molecular bases of extensible cell walls and cell enlargement.** *The Plant Cell* 1997, **9**:1031-1041.
- [36] Coutinho PM, Henrissat B: **Carbohydrate-active enzymes: an integrated database approach.** In *Recent Advances in Carbohydrate Bioengineering*. Edited by Gilbert HJ, Davies B, Henrissat B, Svensson B. Cambridge: The Royal Society of Chemistry; 1999: 3-12
-

- 
- [37] Crowe ML, Serizet C, Thareau V, Aubourg S, Rouze P, Hilson P, Beynon J, Weisbeek P, van Hummelen P, Reymond P, Paz-Ares J, Nietfeld W, Trick M: **CATMA: a complete *Arabidopsis* GST database.** *Nucleic Acids Research* 2003, **31**:156-158.
- [38] Davin LB, Lewis NG: **Lignin primary structures and dirigent sites.** *Current Opinion in Biotechnology* 2005, **16**:407-415.
- [39] De Lorenzo G, D'Ovidio R, Cervone F: **The role of polygalacturonase-inhibiting proteins (PGIPs) in defense against pathogenic fungi.** *Annual Review of Phytopathology* 2001, **39**:313-335.
- [40] Derbyshire P, Findlay K, McCann MC, Roberts K: **Cell elongation in *Arabidopsis* hypocotyls involves dynamic changes in cell wall thickness.** *Journal of Experimental Botany* 2007, **58**:2079-2089.
- [41] Derbyshire P, McCann MC, Roberts K: **Restricted cell elongation in *Arabidopsis* hypocotyls is associated with a reduced average pectin esterification level.** *BMC Plant Biology* 2007, **7**:31.
- [42] Desnos T, Orbovic V, Bellini C, Kronenberger J, Caboche M, Traas J, Hofte H: **Procuste1 mutants identify two distinct genetic pathways controlling hypocotyl cell elongation, respectively in dark- and light-grown *Arabidopsis* seedlings.** *Development* 1996, **122**:683-693.
- [43] Egelund J, Skjot M, Geshi N, Ulvskov P, Petersen BL: **A complementary bioinformatics approach to identify potential plant cell wall glycosyltransferase-encoding genes.** *Plant Physiology* 2004, **136**:2609-2620.
- [44] Eigenbrode SD, Espelie KE: **Effects of plant epicuticular lipids on insects herbivores.** *Annual Review of Entomology* 1995, **40**:171-194.
- [45] Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *Journal of molecular biology* 2000, **300**:1005-1016.
- [46] Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E: **Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls.** *Plant Methods* 2006, **2**:10.
- [47] Fowler TJ, Bernhardt C, Tierney ML: **Characterization and expression of four proline-rich cell wall protein genes in *Arabidopsis* encoding two distinct subsets of multiple domain proteins.** *Plant Physiology* 1999, **121**:1081-1092.
- [48] Freshour G, Clay RP, Fuller MS, Albersheim P, Darvill AG, Hahn MG: **Developmental and tissue-specific structural alterations of the cell-wall**
-

- 
- polysaccharides of *Arabidopsis thaliana* roots.** *Plant Physiology* 1996, **110**:1413-1429.
- [49] Fry SC, York WS, Albersheim P, Darvill A, Hayashi T, Joseleau JP, Kato Y, Lorences EP, MacLachlan GA, McNeil M, Mort AJ, Reid JSG, Seitz HU, Selvendran RR, Voragen AGJ, White AR: **An unambiguous nomenclature for xyloglucan-derived oligosaccharides.** *Physiologia Plantarum* 1993, **89**:1-3.
- [50] Fry SC: **Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals.** *The Biochemical journal* 1998, **332**:507-515.
- [51] Fry SC: **Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells.** *The New Phytologist* 2004, **161**:641-675.
- [52] Ge Y, Dudoit S, Speed TP: **Resampling-based multiple testing for microarray data analysis.** *Test* 2003, **12**:1-77.
- [53] Gendreau E, Traas J, Desnos T, Grandjean O, Caboche M, Hofte H: **Cellular basis of hypocotyl growth in *Arabidopsis thaliana*.** *Plant Physiology* 1997, **114**:295-305.
- [54] Goldberg R: **On possible connections between auxin induced growth and cell wall glucanase activities.** *Plant Science Letters* 1977, **8**:233-242.
- [55] Graca J, Santos S: **Suberin: a biopolyester of plants' skin.** *Macromolecular Bioscience* 2007, **7**:128-135.
- [56] Green PB: **Mechanism for Plant Cellular Morphogenesis.** *Science* 1962, **138**:1404-1405.
- [57] Groh B, Hubner C, Lendzian KJ: **Water and oxygen permeance of phellements isolated from trees: the role of waxes and lenticels.** *Planta* 2002, **215**:794-801.
- [58] Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Molecular and Cellular Biology* 1999, **19**:1720-1730.
- [59] Hager A: **Role of the plasma membrane H<sup>+</sup>-ATPase in auxin-induced elongation growth: historical and new aspects.** *Journal of Plant Research* 2003, **116**:483-505.
- [60] Hall Q, Cannon MC: **The cell wall hydroxyproline-rich glycoprotein RSH is essential for normal embryo development in *Arabidopsis*.** *The Plant Cell* 2002, **14**:1161-1172.
- [61] He ZE, Fujiki M, Kohorn BD: **A cell wall-associated, receptorlike protein kinase.** *Journal of Biological Chemistry* 1996, **16**:19789-19793.
-

- 
- [62] Heredia A: **Biophysical and biochemical characteristics of cutin, a plant barrier biopolymer.** *Biochimica et Biophysica Acta* 2003, **1620**:1-7.
- [63] Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao RP, Bitton F, Caboche M, Cannoot B, Chardakov V, Cognet-Holliger C, Colot V, Crowe M, Darimont C, Durinck S, Eickhoff H, de Longevialle AF, Farmer EE, Grant M, Kuiper MT, Lehrach H, Leon C, Leyva A, Lundeberg J, Lurin C, Moreau Y, Nietfeld W, Paz-Ares J, Reymond P, Rouze P, Sandberg G, Segura MD, Serizet C, Tabrett A, Taconnat L, Thareau V, Van Hummelen P, Vercruysse S, Vuylsteke M, Weingartner M, Weisbeek PJ, Wirta V, Wittink FR, Zabeau M, Small I: **Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications.** *Genome Research* 2004, **14**:2176-2189.
- [64] Hunter TC, Andon NL, Koller A, Yates JR, Haynes PA: **The functional proteomics toolbox: methods and applications.** *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences Sci* 2002, **782**:165-181.
- [65] Ibrahim H, Pertl H, Pittertschatscher K, Fadl-Allah E, el-Shahed A, Bentrup FW, Obermeyer G: **Release of an acid phosphatase activity during lily pollen tube growth involves components of the secretory pathway.** *Protoplasma* 2002, **219**:176-183.
- [66] Jacobs J, Roe JL: **SKS6, a multicopper oxidase-like gene, participates in cotyledon vascular patterning during *Arabidopsis thaliana* development.** *Planta* 2005, **222**:652-666.
- [67] Jamet E, Albenne C, Boudart G, Irshad M, Canut H, Pont-Lezica R: **Recent advances in plant cell wall proteomics.** *Proteomics* 2008, **8**:893-908.
- [68] Jamet E, Canut H, Boudart G, Pont-Lezica RF: **Cell wall proteins: a new insight through proteomics.** *Trends in Plant Science* 2006, **11**:33-39.
- [69] Jamet E: **Bioinformatics as a critical prerequisite to transcriptome and proteome studies.** *Journal of Experimental Botany* 2004, **55**:1977-1979.
- [70] Johnson KL, Jones BJ, Schultz CJ, Bacic A: **Non enzymic cell wall (glyco)proteins.** In *The plant cell wall*. Volume 8. Edited by Rose JK. Boca Raton: CRC Press LLC; 2003: 111-154
- [71] Jones AM, Thomas V, Truman B, Lilley K, Mansfield J, Grant M: **Specific changes in the *Arabidopsis* proteome in response to bacterial challenge: differentiating basal and R-gene mediated resistance.** *Phytochemistry* 2004, **65**:1805-1816.
- [72] Juge N: **Plant protein inhibitors of cell wall degrading enzymes.** *Trends in Plant Science* 2006, **11**:359-367.
-



- 
- [73] Kader J-C: **Lipid transfer proteins: a puzzling family of plant proteins.** *Trends in Plant Science* 1997, **2**:66-70.
- [74] Kaida R, Sage-Ono K, Kamada H, Okuyama H, Syono K, Kaneko TS: **Isolation and characterization of four cell wall purple acid phosphatase genes from tobacco cells.** *Biochimica et Biophysica Acta* 2003, **1625**:134-140.
- [75] Kapteyn JC, Montijn RC, Vink E, de la Cruz J, Llobel A, Douwes JE, Shimoi H, Lipke PN, Klis FM: **Retention of *Saccharomyces cerevisiae* cell wall proteins through a phosphodiester-linked  $\beta$ -1,3-  $\beta$ -1,6-glucan heteropolymer.** *Glycobiology* 1996, **6**:337-345.
- [76] Kieliszewski MJ, Lamport DT: **Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny.** *The Plant Journal* 1994, **5**:157-172.
- [77] Kieliszewski MJ: **The latest hype on Hyp-O-glycosylation codes.** *Phytochemistry* 2001, **57**:319-323.
- [78] Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown RM, Jr.: **Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*.** *The Plant Cell* 1999, **11**:2075-2086.
- [79] Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MM, Pronk JT, Slijper M, Heck AJ: **Proteome analysis of yeast response to various nutrient limitations.** *Molecular Systems Biology* 2006, **2**:2006.0026.
- [80] Kunst L, Samuels AL: **Biosynthesis and secretion of plant cuticular wax.** *Progress in Lipid Research* 2003, **42**:51-80.
- [81] Kurdyukov S, Faust A, Nawrath C, Bär S, Voisin D, Efremova N, Franke R, Shreiber L, Saedler H, Métraux J-P, Yephremov A: **The epidermis-specific extracellular BODYGUARD controls cuticle development and morphogenesis in *Arabidopsis*.** *The Plant Cell* 2006, **18**:321-339.
- [82] Kwon HK, Yokoyama R, Nishitani K: **A proteomic approach to apoplastic proteins involved in cell wall regeneration in protoplasts of *Arabidopsis* suspension-cultured cells.** *Plant & Cell Physiology* 2005, **46**:843-857.
- [83] Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227**:680-685.
- [84] Lafarguette F, Leple JC, Dejardin A, Laurans F, Costa G, Lesage-Descauses M-C, Pilate G: **Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood.** *The New Phytologist* 2004, **164**:107-121.
-

- 
- [85] Lee SJ, Saravanan RS, Damasceno CM, Yamane H, Kim BD, Rose JK: **Digging deeper into the plant cell wall proteome.** *Plant physiology and Biochemistry* 2004, **42**:979-988.
- [86] Li Y, Jones L, McQueen-Mason S: **Expansins and cell growth.** *Current Opinion in Cell Biology* 2003, **6**:603-610.
- [87] Liu C, Mehdy MC: **A nonclassical arabinogalactan protein gene highly expressed in vascular tissues, AGP31, is transcriptionally repressed by methyl jasmonic acid in *Arabidopsis*.** *Plant Physiology* 2007, **145**:863-874.
- [88] Loomis WD, Durst RW: **Chemistry and biology of boron.** *Biofactors* 1992, **3**:229-239.
- [89] Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I: **Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis.** *The Plant cell* 2004, **16**:2089-2103.
- [90] Maldonado AM, Doerner P, Dixon RA, Lamb CJ, Cameron RK: **A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*.** *Nature* 2002, **26**:399-403.
- [91] Malinowski R, Filipecki M: **The role of cell wall in plant embryogenesis.** *Cellular & Molecular Biology Letters* 2002, **7**:1137-1151.
- [92] Marin-Rodriguez MC, Orchard J, Seymour GB: **Pectate lyases, cell wall degradation and fruit softening.** *Journal of Experimental Botany* 2002, **53**:2115-2119.
- [93] Matsubayashi Y, Sakagami Y: **Peptide hormones in plants.** *Annual Review of Plant Biology* 2006, **57**:649-674.
- [94] McCabe PF, Valentine TA, Forsberg LS, Pennell RI: **Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot.** *The Plant Cell* 1997, **9**:2225-2241.
- [95] McCaig BC, Meagher RB, Dean JF: **Gene structure and molecular analysis of the laccase-like multicopper oxidase (LMCO) gene family in *Arabidopsis thaliana*.** *Planta* 2005, **221**:619-636.
- [96] Melan MA, Cosgrove DJ: **Evidence against the involvement of ionically bound cell wall proteins in pea epicotyl growth.** *Plant Physiology* 1988, **86**:469-74.
- [97] Meyers BC, Galbraith DW, Nelson T, Agrawal V: **Methods for transcriptional profiling in plants. Be fruitful and replicate.** *Plant Physiology* 2004, **135**:637-652.
-

- 
- [98] Micheli F: **Pectin methylesterases: cell wall enzymes with important roles in plant physiology.** *Trends in Plant Science* 2001, **6**:414-419.
- [99] Minic Z, Jouanin L: **Plant glycoside hydrolases involved in cell wall polysaccharide degradation.** *Plant physiology and Biochemistry* 2006, **44**:435-449.
- [100] Moritz B, Meyer HE: **Approaches for the quantification of protein concentration ratios.** *Proteomics* 2003, **3**:2208-2220.
- [101] Mouille G, Robin S, Lecomte M, Pagant S, Hofte H: **Classification and identification of *Arabidopsis* cell wall mutants using Fourier-Transform InfraRed (FT-IR) microspectroscopy.** *The Plant Journal* 2003, **35**:393-404.
- [102] Mersa V, Seidi T, Gentzsch M, Tanner W: **Specific labelling of cell wall proteins by biotinylation, identification of four covalently linked O-mannosylated proteins of *Saccharomyces cerevisiae*.** *Yeast* 1997, **13**:1145-1154.
- [103] Murashige T, Skoog F: **A revised medium for rapid growth and bioassays with tobacco tissue culture.** *Physiologia Plantarum* 1962, **15**:473-497.
- [104] Nadeau JA, Sack FD: **Stomatal development: cross talk puts mouths in place.** *Trends in Plant Science* 2003, **8**:294-299.
- [105] Narsai R, Howell K, Millar A, O'Toole N, Small I, Whealan J: **Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*.** *The Plant Cell* 2007, **19**:3418-3436.
- [106] Nersissian AM, Shipp EL: **Blue copper-binding domains.** *Advances in Protein Chemistry* 2002, **60**:271-340.
- [107] Nicol F, His I, Jauneau A, Vernhettes S, Canut H, Hofte H: **A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*.** *The EMBO Journal* 1998, **17**:5563-5576.
- [108] Nicol F, Hofte H: **Plant cell expansion: scaling the wall.** *Current Opinion in Cell Biology* 1998, **1**:12-17.
- [109] Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Engineering* 1997, **10**:1-6.
- [110] Nieuwland J, Feron R, Huisman BAH, Fasolino A, Hilbers CW, Derksen J, Mariani C: **Lipid transfer proteins enhance cell wall extension in tobacco.** *The Plant Cell* 2005, **17**:2009-2019.
-

- 
- [111] Oh IS, Park AR, Bae MS, Kwon SJ, Kim YS, Lee JE, Kang NY, Lee S, Cheong H, Park OK: **Secretome analysis reveals an *Arabidopsis* lipase involved in defense against *Alternaria brassicicola*.** *The Plant Cell* 2005, **17**:2832-2847.
- [112] Pagant S, Bichet A, Sugimoto K, Lerouxel O, Desprez T, McCann M, Lerouge P, Vernhettes S, Höfte H: **KOBITO1 encodes a novel plasma membrane protein necessary for normal synthesis of cellulose during cell expansion in *Arabidopsis*.** *The Plant Cell* 2002, **14**:2001–2013.
- [113] Pardo M, Ward M, Bains S, Molina M, Blastock W, Gil C, Nombela C: **A proteomic approach for the study of *Saccharomyces cerevisiae* cell wall biogenesis.** *Electrophoresis* 2000, **21**:3396-3410.
- [114] Passardi F, Cosio C, Penel C, Dunand C: **Peroxidases have more functions than a Swiss army knife.** *Plant Cell Reports* 2005, **24**:255-265.
- [114] Passardi F, Penel C, Dunand C: **Performing the paradoxical: how plant peroxidases modify the cell wall.** *Trends in Plant Science* 2004, **9**:532-540.
- [115] Pennell R: **Cell walls: structures and signals.** *Current Opinion in Plant Biology* 1998, **1**:504-510.
- [116] Pereira H: **Chemical composition and variability of cork from *Quercus suber* L.** *Wood Science and Technology* 1988, **22**:211–218.
- [117] Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR: **Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:15566-15571.
- [118] Pitarch A, Sanchez M, Nombela C, Gil C: **Sequential fractionation and two-dimensional gel analysis unravel the complexity of the dimorphic fungus *Candida albicans* cell wall proteome.** *Molecular and Cellular Proteomics* 2002, **1**:967-82.
- [119] Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I: **TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat.** *The Plant cell* 2005, **17**:2966-2980.
- [120] **PSORT** [<http://psort.nibb.ac.jp/form.html>]
- [121] Quail PH, Boylan MT, Parks BM, Short TW, Xu Y, Wagner D: **Phytochromes: photosensory perception and signal transduction.** *Science* 1995, **268**:675-680.
-

- 
- [122] Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier.** *Nucleic Acids Research* 2005, **33**:W116-120.
- [123] Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W: **Genome-wide characterization of the lignification toolbox in *Arabidopsis*.** *Plant Physiology* 2003, **133**:1051-1071.
- [124] Rautengarten C, Usadel B, Neumetzler L, Hartmann J, Büssis D, Altmann T: **A subtilisin-like serine protease essential for mucilage release from *Arabidopsis* seed coats.** *The Plant Journal* 2008, **54**:466-480.
- [125] Raz V, Koornneef M: **Cell division activity during apical hook development.** *Plant Physiology* 2001, **125**:219-226.
- [126] Refrégier G, Pelletier S, Jaillard D, Hofte H: **Interaction between wall deposition and cell elongation in dark-grown hypocotyl cells in *Arabidopsis*.** *Plant Physiology* 2004, **135**:959-968.
- [127] Reiter WD: **Biosynthesis and properties of the plant cell wall.** *Current Opinion in Plant Biology* 2002, **5**:536-542.
- [128] Ridley BL, O'Neill MA, Mohnen D: **Pectins: structure, biosynthesis, and oligogalacturonide-related signaling.** *Phytochemistry* 2001, **57**:929-967.
- [129] Riederer M, Müller C (Eds.): **The Biology of Plant Cuticle.** London: Blackwell; 2005.
- [130] Ringli C, Keller B, Ryser U: **Glycine-rich proteins as structural components of plant cell walls.** *Cellular and Molecular Life Sciences* 2001, **58**:1430-1441.
- [131] Roberts K: **The plant extracellular matrix: in a new expansive mood.** *Current Opinion in Cell Biology* 1994, **6**:688-694.
- [132] Rose JK, Braam J, Fry SC, Nishitani K: **The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature.** *Plant & Cell Physiology* 2002, **43**:1421-1435.
- [133] Roudier F, Fernandez AG, Fujita M, Himmelspach R, Borner GH, Schindelman G, Song S, Baskin TI, Dupree P, Wasteneys GO, Benfey PN: **COBRA, an *Arabidopsis* extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation.** *The Plant Cell* 2005, **17**:1749-1763.
- [134] Roudier F, Schindelman G, DeSalle R, Benfey PN: **The COBRA family of putative GPI-anchored proteins in *Arabidopsis*. A new fellowship in expansion.** *Plant Physiology* 2002, **130**:538-548.
-

- 
- [135] Saibo NJ, Vriezen WH, Beemster GT, Van Der Straeten D: **Growth and stomata development of *Arabidopsis* hypocotyls are controlled by gibberellins and modulated by ethylene and auxins.** *The Plant Journal* 2003, **33**:989-1000.
- [136] Schaller A: **A cut above the rest: the regulatory function of plant proteases.** *Planta* 2004, **220**:183-197.
- [137] Scheler C, Lamer S, Pan Z, Li XP, Salnikow J, Jangblut P: **Peptide mass fingerprint sequences coverage from differentially stained proteins in two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS).** *Electrophoresis* 1998, **19**:918-927.
- [138] Schnabelrauch LS, Kieliszewski MJ, Upham BL, Alizedeh H, Lamport DTA: **Isolation of pI 4.6 extensin peroxidase from tomato cell suspension cultures and identification of Val-Tyr-Lys as putative intermolecular cross-link site.** *The Plant Journal* 1996, **9**:477-489.
- [139] Schreiber L, Hartmann K, Skrabs M, Zeier J: **Apoplastic barriers in roots: chemical composition of endodermal and hypodermal cell walls.** *Journal of Experimental Botany* 1999, **50**:1267-1280.
- [140] Schreiber L: **Polar paths of diffusion across plant cuticles: new evidence for an old hypothesis.** *Annals of Botany* 2005, **95**:1069-1073.
- [141] Schultz CJ, Ferguson KL, Lahnstein J, Bacic A: **Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*.** *The Journal of Biological Chemistry* 2004, **279**:45503-45511.
- [142] Schultz CJ, Johnson KL, Currie G, Bacic A: **The classical arabinogalactan gene family of *Arabidopsis*.** *The Plant Cell* 2000, **12**:1751-1767.
- [143] Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YG, Antony Bacic A: **Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case.** *Plant Physiology* 2002, **129**:1448-1463.
- [144] Schwacke R, Schneider A, van der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flugge UI, Kunze R: **ARAMEMNON, a novel database for *Arabidopsis* integral membrane proteins.** *Plant Physiology* 2003, **131**:16-26.
- [145] Seifert GJ, Roberts K: **The biology of arabinogalactan proteins.** *Annual Review of Plant Biology* 2007, **58**:137-161.
- [146] Shah K, Penel C, Gagnon J, Dunand C: **Purification and identification of a Ca<sup>2+</sup>-pectate binding peroxidase from *Arabidopsis* leaves.** *Phytochemistry* 2004, **65**:307-312.
-

- 
- [147] Shevchenko A, Wilm M, Vorm O, Mann M: **Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.** *Analytical Chemistry* 1996, **68**:850-858.
- [148] Shpak E, Barbar E, Leykam JF, Kieliszewski MJ: **Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*.** *The Journal of biological chemistry* 2001, **276**:11272-11278.
- [149] Smallwood M, Beven A, Donovan N, Neill S, Peart J, Roberts K, Knox J: **Localization of cell wall proteins in relation to the developmental anatomy of the carrot root apex.** *The Plant Journal* 1994, **5**:237-246.
- [150] Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M: **A genomic approach to suberin biosynthesis and cork differentiation.** *Plant Physiology* 2007, **144**:419-431.
- [151] Sommer-Knudsen J, Bacic A, Clarke AE: **Hydroxyproline-rich plant glycoproteins.** *Phytochemistry* 1998, **47**:483-497.
- [152] Suh MC, Samuels AL, Jetter R, Kunst L, Pollard M, Ohlrogge J, Beisson F: **Cuticular lipid composition, surface structure, and gene expression in *Arabidopsis* stem epidermis.** *Plant Physiology* 2005, **139**:1649-1665.
- [153] Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y: **A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants.** *Development* 2001, **128**:4681-4689.
- [154] Tanaka R, Ikeda M, Funatsuki K, Yukioka H, Hashimoto Y, Fujimoto S, Takata M, Katoh K, Konno H: **Molecular cloning and cytochemical analysis of exopolygalacturonase from carrot.** *Planta* 2002, **215**:735-744.
- [155] **TargetP** [<http://www.cbs.dtu.dk/services/TargetP/>]
- [156] Tomassen MMM, Barrett DM, van der Valk HCPM, Woltering EJ: **Isolation and characterization of a tomato non-specific lipid transfer protein involved in polygalacturonase-mediated pectin degradation.** *Journal of Experimental Botany* 2007, **58**:1151-1160.
- [157] Trevino V, Falciani F, Barrera-Saldana HA: **DNA microarrays: a powerful genomic tool for biomedical and clinical research.** *Molecular Medicine* 2007, **13**:527-541.
- [158] van Hengel AJ, Roberts K: **AtAGP30, an arabinogalactan-protein in the cell walls of the primary root, plays a role in root regeneration and seed germination.** *The Plant Journal* 2003, **36**:256-270.
-

- 
- [159] Vissenberg K, Fry SC, Pauly M, Hofte H, Verbelen JP: **XTH acts at the microfibril-matrix interface during cell elongation.** *Journal of Experimental Botany* 2005, **56**:673-683.
- [160] Voigt J, Frank R: **14-3-3 Proteins are constituents of the insoluble glycoprotein framework of the *Chlamydomonas* cell wall.** *The Plant Cell* 2003, **15**:1399-1413.
- [161] Wang SB, Hu Q, Sommerfeld M, Chen F: **Cell wall proteomics of the green alga *Haematococcus pluvialis* (Chlorophyceae).** *Proteomics* 2004, **4**:692-708.
- [162] Watson BS, Lei Z, Dixon RA, Sumner LW: **Proteomics of *Medicago sativa* cell walls.** *Phytochemistry* 2004, **65**:1709-1720.
- [163] White PJ, Broadley MR: **Calcium in plants.** *Annals of Botany* 2003, **92**:487-511.
- [164] Willats WG, McCartney L, Mackie W, Knox JP: **Pectin: cell biology and prospects for functional analysis.** *Plant Molecular Biology* 2001, **47**:9-27.
- [165] Wong D: **Enzymatic deconstruction of backbone structures of the ramified regions in pectins.** *The Protein Journal* 2008, **27**:30-42.
- [166] Ye ZH, Song YR, Marcus A, Varner JE: **Comparative localization of three classes of cell wall proteins.** *The Plant Journal* 1991, **1**:175-183.
- [167] Zhang Z, Pierce ML, Mort AJ: **Changes in homogalacturonans and enzymes degrading them during cotton cotyledon expansion.** *Phytochemistry* 2007, **68**:1094-1113.



# **ANNEXES AND SUPPLEMENTARY DATA**

## **ANNEXES CHAPTER 2**

## Annex I. Composition of (12 × 15 cm) polyacrylamide gel for 1D-E of proteins

Solutions and buffers	Nature, %age of acrylamide, and volume (mL)							
	Resolving gel (7.5%)	Resolving gel (8%)	Resolving gel (9%)	Resolving gel (10%)	Resolving gel (11%)	Resolving gel (12%)	Resolving gel (12.5%)	Stacking gel (4%)
Acrylamide 40%	7.280	7.776	8.748	9.720	10.692	11.664	12.150	1.944
Bis-acrylamide 2%	4.000	4.288	4.824	5.360	5.896	6.432	6.700	1.072
Buffer A 4x (Tris-HCl 1.5 M pH 8.8)	10.000	10.000	10.000	10.000	10.000	10.000	10.000	0.000
Buffer B 10x (Tris-HCl 1.25 M pH 6.8)	0	0	0	0	0	0	0	2.000
SDS 10% (filtered)	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.200
UHQ water	18.120	17.336	15.828	14.320	12.812	11.304	10.550	14.720
Ammonium persulfate 10%	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.100
TEMED	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020
Total	40.020	40.020	40.020	40.020	40.020	40.020	40.020	20.056

Annex II: Composition of mini ( $6 \times 8$  cm) polyacrylamide gel for 1D-E of proteins

Solutions and buffers	Nature, %age of acrylamide, and volume (mL)							
	Resolving gel (7.5%)	Resolving gel (8%)	Resolving gel (9%)	Resolving gel (10%)	Resolving gel (11%)	Resolving gel (12%)	Resolving gel (12.5%)	Stacking gel (4%)
Acrylamide 40%	1.456	1.555	1.750	1.944	2.138	2.333	2.430	0.486
Bis-acrylamide 2%	0.800	0,858	0.965	1.072	1.179	1.286	1.340	0.268
Buffer A 4x (Tris-HCl 1.5 M pH 8.8)	2.500	2.500	2.500	2.500	2.500	2.500	2.500	0.000
Buffer B 10x (Tris-HCl 1.25 M pH 6.8)	0	0	0	0	0	0	0	0.500
SDS 10% (filtered)	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.050
UHQ water	3.089	2.932	2.631	2.329	2.027	1.726	1.575	3.680
Ammonium persulfate 10%	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.025
TEMED	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
Total	8.000	8.000	8.000	8.000	8.000	8.000	8.000	5.014

## Annex III. Composition of RT-PCR buffer

Components	Volume ( $\mu\text{L}$ )
Oligo (dT) <sub>15</sub> primer (Promega Corporation, Madison, WI, USA) 500 $\mu\text{g}/\text{mL}$	1
Total RNA (100 $\text{ng}/\mu\text{L}$ )	10
dNTPs (10 mM each)	1
Desmine mRNA (Pork mRNA)	1

After 5 min heating at 65°C and putting again on ice, the following components are added:

5× First Strand Buffer (Invitrogen, Carlsbad, San Diego, CA, USA)	4
DTT, 0.1 M (Invitrogen, Carlsbad, San Diego, CA, USA)	2
RNasine Plus RNase Inhibitor, 40 U/ $\mu\text{L}$ (Promega Corporation, Madison, WI, USA)	1

## **SUPPLEMENTARY DATA CHAPTER 3**

### Additional file 1: Bioinformatic analysis of proteins extracted from cell walls of *A. thaliana* cell suspension cultures [8].

Data are from [8]. Proteins were extracted from *A. thaliana* cell walls of cell suspension cultures as described in Figure 1. Two successive extractions were performed : step 1 using CaCl<sub>2</sub> and step 2 using urea. All proteins sequences were analyzed with three bioinformatic programs to predict their sub-cellular localization. Proteins for which predictions by different bioinformatic programs are in conflict are classified as "not clear".

step 1: CaCl<sub>2</sub> extract

Predicted subcellular localization	Gene ( <i>A. thaliana</i> )	PSORT (a)	TargetP (b)	Predicted signal peptide (c)	Aramemnon (d)
transmembrane domain	At3g30810	Golgi (0.900), plasma membrane (0.790), chloroplast thylakoid membrane (0.762)	other (0.796)		yes
	At4g12420	plasma membrane (0.919)	secretory pathway (0.989)	1-20	GPI anchor
outside	At1g06870	outside (0.638)	secretory pathway (0.850)	1-21	
	At1g21670	outside (0.820)	secretory pathway (0.973)	1-21	
	At1g53070	outside (0.528)	secretory pathway (0.972)	1-17 or 1-23	
	At1g78830	outside (0.820)	secretory pathway (0.962)	1-22	
	At1g78850	outside (0.820)	secretory pathway (0.973)	1-22	
	At2g39700	outside (0.810)	secretory pathway (0.961)	1-20	
	At2g41800	outside (0.370)	secretory pathway (0.369)	1-21	
	At2g44450	outside (0.820)	secretory pathway (0.853)	1-22	
	At2g47050	outside (0.820)	secretory pathway (0.928)	1-20	
	At3g07320	outside (0.790)	secretory pathway (0.482)	1-19	
	At3g08030	outside (0.820)	secretory pathway (0.985)	1-21	
	At3g45960	outside (0.805)	secretory pathway (0.744)	1-20	
	At3g45970	endoplasmic reticulum (0.820)	secretory pathway (0.854)	1-20	
	At3g61820	outside (0.628)	secretory pathway (0.828)	1-26 or 1-25	
	At4g20830	outside (0.820)	secretory pathway (0.540)	1-30	
	At4g25900	outside (0.820)	secretory pathway (0.924)	1-20 or 1-24	
	At4g30270	outside (0.800)	secretory pathway (0.975)	1-21	
	At5g06860	outside (0.733)	secretory pathway (0.964)	1-21	
intracellular	At1g08450	endoplasmic reticulum (0.910) (C-terminal HDEL)	secretory pathway (0.995)	1-28	
	At1g13440	microbody (0.539)	other (0.610)		
	At1g30580	microbody (0.507)	other (0.579)		
	At2g39050	microbody (0.515)	other (0.910)		
	At2g43710	microbody (0.495)	chloroplast (0.977)		
	At3g02630	chloroplast stroma (0.593)	chloroplast (0.779)		
	At3g26060	mitochondry (0.679)	chloroplast (0.904)		
	At3g48380	mitochondry (0.596)	mitochondry (0.308)		
	At3g58480	nucleus (0.760)	other (0.802)		
	At4g17260	cytoplasm (0.450)	other (0.435)		
	At4g37870	nucleus (0.880)	other (0.849)		
	At5g08310	mitochondry (0.789)	mitochondry (0.946)		
	At5g20080	mitochondry (0.859)	mitochondry (0.855)		
	At5g46550	nucleus (0.940)	other (0.504)		
	At5g55990	microbody (0.569)	secretory pathway (0.580), other (0.576)		
not clear	At4g22410	cytoplasm (0.450)	secretory pathway (0.922)	1-26	
	At5g37990	endoplasmic reticulum (0.550)	secretory pathway (0.523), chloroplast (0.307)	1-15	

**step 2: urea extract**

transmembrane domain	At1g61500	plasma membrane (0.460)	secretory pathway (0.963)	1-24	yes
	At1g67880	plasma membrane (0.790)	other (0.955)		yes
	At3g25560	plasma membrane (0.460)	secretory pathway (0.742)	1-32	yes
	At3g57400	plasma membrane (0.830)	other (0.792)		yes
	At5g08390	plasma membrane (0.460)	secretory pathway (0.932)	1-22	yes
	At5g58640	plasma membrane (0.640)	secretory pathway (0.326)	1-17	yes
outside	At1g71695	outside (0.820)	secretory pathway (0.829)	1-22 or 1-31	
	At2g16430	outside (0.820)	secretory pathway (0.963)	1-25	
	At2g44450	outside (0.820)	secretory pathway (0.853)	1-22	
	At3g08030	outside (0.820)	secretory pathway (0.985)	1-21	
	At3g45970	endoplasmic reticulum (0.820)	secretory pathway (0.854)	1-20	
	At3g52500	outside (0.820)	secretory pathway (0.649)	1-18	
	At4g08770	outside (0.633)	secretory pathway (0.955)	1-22	
	At4g24890	outside (0.820)	secretory pathway (0.925)	1-20 or 1-26	
intracellular	At1g24360	mitochondry (0.613)	chloroplast (0.545), mitochondry (0.206)		
	At1g30580	microbody (0.507)	other (0.579)		
	At1g43800	nucleus (0.760)	mitochondry (0.430)		
	At1g43800	nucleus (0.760)	mitochondry (0.430)		
	At1g52960	nucleus (0.960)	chloroplast (0.630)		
	At1g56070	mitochondry (0.360)	other (0.823)		
	At1g69290	chloroplast (0.840)	chloroplast (0.528)		
	At1g72560	cytoplasm (0.450)	other (0.872)		
	At2g17190	cytoplasm (0.650)	other (0.797)		
		endoplasmic reticulum (0.910) (C-terminal KDEL)			
	At2g32920		secretory pathway (0.920)	1-17 or 1-24	
	At2g36530	endoplasmic reticulum (0.600)	other (0.747)		
	At2g43710	microbody (0.495)	chloroplast (0.977)		
	At2g44350	mitochondry (0.883)	mitochondry (0.848)		
	At3g04600	nucleus (0.940)	other (0.954)		
	At3g14440	chloroplast (0.905)	chloroplast (0.719)		
	At3g15730	microbody (0.634)	other (0.614)		
	At3g16857	nucleus (0.700)	other (0.376)		
	At3g57400	plasma membrane (0.830)	other (0.792)		
	At4g17260	cytoplasm (0.450)	chloroplast (0.703)		
	At4g31180	chloroplast (0.890)	chloroplast (0.876)		
	At4g36530	microbody (0.594)	other (0.828)		
	At4g37870	nucleus (0.880)	other (0.849)		
	At5g06450	microbody (0.640)	other (0.954)		
	At5g20080	mitochondry (0.859)	mitochondry (0.855)		
	At5g26710	endoplasmic reticulum (0.550)	chloroplast (0.552)		
		endoplasmic reticulum (0.910) (C-terminal HDEL)			
	At5g28540		secretory pathway (0.993)	1-27	
	At5g41550	chloroplast (0.520)	mitochondry (0.517)		

Colour code:

proteins found at both steps 1 and 2

(a) PSORT : <http://psort.nibb.ac.jp/form.html> [29]

(b) TargetP: <http://www.cbs.dtu.dk/services/TargetP/> [30]

(c) Two sizes are indicated when different signal peptides are predicted by PSORT and TargetP. The first one is predicted with PSORT.

(d) Aramemnon: <http://aramemnon.botanik.uni-koeln.de/> [31]



## Additional file 2: Bioinformatic analysis of proteins extracted from cell walls of *M. sativa* stems [9].

Data are from [9]. Proteins were extracted from *M. sativa* cell walls of stems as described in Figure 2. Two successive extractions were performed : step 1 using  $\text{CaCl}_2$ , and step 2 using  $\text{LiCl}$ . All proteins sequences were analyzed with bioinformatic softwares to predict their sub-cellular localization. When only partial protein sequences or no sequence are available, proteins are classified as "not predictable".

step 1: $\text{CaCl}_2$ extract	Predicted subcellular localization	Gene ( <i>A. thaliana</i> )	Protein	PSORT (a)	TargetP (b)	Predicted signal peptide (c)	Aramemnon (d)
	transmembrane domain		CAC34417	plasma membrane (0.460)	secretory pathway (0.986)	1-20	yes
			P16002	chloroplast (0.659), plasma membrane (0.650)	chloroplast (0.924)		yes
			P26291	plasma membrane (0.790), chloroplast (0.753)	chloroplast (0.797)		yes
	outside	At1g09560	AAB51577	outside (0.609)	secretory pathway (0.960)	1-23	
		At3g62020	AAB51752	plasma membrane (0.685)	secretory pathway (0.970)	1-21	no
			BAA81904	plasma membrane (0.790)	secretory pathway (0.758)	1-39	no
			CAA09607	outside (0.820)	secretory pathway (0.996)	1-25 or 1-23	
			CAA10167	outside (0.790)	secretory pathway (0.963)	1-16 or 1-22	
			CAA10287	outside (0.585)	secretory pathway (0.867)	1-33	
			CAC44501	outside (0.820)	secretory pathway (0.985)	1-24	
			CAD29731	outside (0.714)	secretory pathway (0.982)	1-23 or 1-26	
		At1g49750	NP_175397	vacuole (0.750), outside (0.666)	secretory pathway (0.239)	1-15 or 1-22	
		At3g20820	NP_188718	vacuole (0.861), outside (0.820)	secretory pathway (0.752)	1-19	
		At4g34480	NP_195174	plasma membrane (0.460)	secretory pathway (0.942)	1-22	
		At5g14440	NP_196949	outside (0.786)	secretory pathway (0.941)	1-30	
		At5g51890	NP_200002	plasma membrane (0.685)	secretory pathway (0.991)	1-24	no
		At1g09750	NP_563851	outside (0.456)	secretory pathway (0.899)	1-21 or 1-23	
		At1g20850	NP_564126	vacuole (0.808), outside (0.786)	secretory pathway (0.987)	1-26	
		At4g33490	NP_567922	outside (0.820)	secretory pathway (0.913)	1-21	
			P36907	outside (0.757)	secretory pathway (0.985)	1-23	
			Q01806	plasma membrane (0.685)	secretory protein (0.983)	1-30	no
			Q9S8P4	outside (0.820)	secretory pathway (0.967)	1-20	
			S68805	outside (0.781)	secretory pathway (0.961)	1-24	
			T07086	outside (0.743)	secretory pathway (0.771)	1-19	
			T09642	vacuole (0.761), outside (0.657)	secretory pathway (0.981)	1-30 or 1-24	

intracellular	AAC49358	mitochondry (0.555)	chloroplast (0.628)
	AAG34872	microbody (0.540)	other (0.568)
	AAL77589	mitochondry (0.679)	chloroplast (0.765)
	CAA09177	microbody (0.640)	other (0.884)
	At1g16470	NP_173096	cytoplasm (0.450)
	At1g23740	NP_173786	mitochondry (0.594)
	At1g48590	NP_175292	mitochondry (0.480)
	At1g67280	NP_176896	mitochondry (0.807)
	At2g15570	NP_179159	chloroplast (0.919)
	At2g24940	NP_180066	cytoplasm (0.450)
	At2g32520	NP_180811	microbody (0.570)
	At3g20390	NP_188674	mitochondry (0.644)
	At3g26060	NP_189235	mitochondry (0.679)
	At3g43810	NP_189967	cytoplasm (0.650)
	At3g55330	NP_191093	chloroplast (0.647)
	At4g09010	NP_192640	chloroplast (0.512)
	At5g01650	NP_195785	chloroplast (0.890)
	At1g09310	NP_563841	cytoplasm (0.650)
	At3g17440	NP_566578	endoplasmic reticulum (0.850)
	At3g63190	NP_567141	mitochondry (0.838)
	At5g20080	NP_568391	mitochondry (0.859)
	O65194	chloroplast (0.876)	chloroplast (0.710)
	O65198	chloroplast (0.921)	chloroplast (0.941)
	P10933	endoplasmic reticulum (0.550)	chloroplast (0.907)
	P16048	mitochondry (0.870)	mitochondry (0.645)
	P16059	mitochondry (0.750)	chloroplast (0.307),
	P29450	chloroplast (0.950)	chloroplast (0.977)
	P49231	cytoplasm (0.450)	other (0.818)
	P51615	plasma membrane (0.615)	other (0.941)
	Q02610	cytoplasm (0.450)	other (0.404)
	Q42971	endoplasmic reticulum (0.600)	other (0.608)
	Q43636	cytoplasm (0.450)	other (0.907)
		plasma membrane (0.650),	
	T06363	chloroplast (0.539)	chloroplast (0.505)
	T09286	mitochondry (0.920)	mitochondry (0.766)
not predictable	AAD34458	encoded by a truncated cDNA	
	AAL06644	encoded by a truncated cDNA	
	AAL15646	encoded by a truncated cDNA	
	JC4780	data not available in usual databases	
	JC4781	data not available in usual databases	
	JC4782	data not available in usual databases	
	S22489	encoded by a truncated cDNA	
	T05957	encoded by a truncated cDNA	
	T09165	encoded by a truncated cDNA	

**step 2: LiCl extract**

outside	At1g49750	NP_175397	vacuole (0.850), outside (0.666)	secretory pathway (0.239)	1-15 or 1-22
		P36907	outside (0.757)	secretory pathway (0.985)	1-23
		T07086	outside (0.743)	secretory pathway (0.771)	1-19
		T07171	outside (0.820)	secretory pathway (0.981)	1-23
		T09665	outside (0.820)	secretory pathway (0.922)	1-28
intracellular	At5g41550	AAD56659	mitochondry (0.483)	mitochondry (0.839)	
		NP_198970	chloroplast (0.520)	mitochondry (0.517)	
			endoplasmic reticulum (0.910) (C-term HDEL)		
		T93508		secretory pathway (0.991)	1-20
not predictable		P04353	encoded by a truncated cDNA		

Colour code:

proteins found at both steps 1 and 2



(a) PSORT : <http://psort.nibb.ac.jp/form.html> [29]

(b) TargetP: <http://www.cbs.dtu.dk/services/TargetP/> [30]

(c) Two sizes are indicated when different signal peptides are predicted by PSORT and TargetP. The first one is predicted with PSORT.

(d) Aramemnon: <http://aramemnon.botanik.uni-koeln.de/> [31]

### Additional file 3: Bioinformatic analysis of proteins extracted from cell walls of *C. albicans* [20].

Data are from [20]. Proteins were extracted from *C. albicans* cell walls as described in Figure 3. Four successive extractions were performed : step 1 using SDS-DTT, step 2 using NaOH, step 3 using a  $\beta$ -1,3-glucanase and step 4 using an exochitinase. All proteins sequences were analyzed with bioinformatic programs to predict their sub-cellular localization. Proteins for which predictions by different bioinformatic programs are in conflict are classified as “not clear”.

step 1: SDS-DTT extract	Predicted subcellular localization	Accession (a)	PSORT (b)	TargetP (c)	Predicted peptide (d)	signal
	outside	CA1541	outside (0.370)	secretory pathway (0.946)	1-18	
		Phrp *				
	intracellular	CA0362	cytoplasm (0.650)	other (0.911)		
		CA0915	endoplasmic reticulum (0.910) (C-term HDEL)	secretory pathway (0.977)	1-29 or 1-32	
		CA1015	mitochondry (0.541)	other (0.694), mitochondry (0.271)		
		CA1230	nucleus (0.760)	other (0.843)		
		CA1691	mitochondry (0.360)	other (0.789), mitochondry (0.227)		
		CA1755	endoplasmic reticulum (0.910) (C-term HDEL)	secretory pathway (0.955)	1-23	
		CA2474	microbody (0.300)	other (0.798)		
		CA2857	nucleus (0.760)	other (0.842)		
		CA3081	nucleus (0.980)	other (0.885)		
		CA3208	cytoplasm (0.650)	mitochondry (0.663)		
		CA3483	mitochondry (0.360)	other (0.275), mitochondry (0.299)		
		CA3534	cytoplasm (0.450)	other (0.887)		
		CA3874	cytoplasm (0.450)	other (0.793)		
		CA4671	cytoplasm (0.450)	other (0.834)		
		CA4765	mitochondry (0.853)	mitochondry (0.716)		
		CA4844	endoplasmic reticulum (0.550)	other (0.704)		
		CA4862	mitochondry (0.360)	other (0.934)		
		CA4959	nucleus (0.600)	other (0.923)		
		CA5135	cytoplasm (0.450)	other (0.928)		
		CA5180	cytoplasm (0.450)	other (0.797)		
		CA5892	mitochondry (0.471)	mitochondry (0.637)		
		CA5950	microbody (0.800)	other (0.604)		
	not predictable	CA2810	not present in the database			
step 2: NaOH extract	outside	Hsp150p/Pir2p*				
	intracellular	CA1691	mitochondry (0.360)	other (0.789), mitochondry (0.227)		
		CA2474	microbody (0.300)	other (0.798)		
		CA3483	mitochondry (0.360)	other (0.275), mitochondry (0.299)		
		CA3874	cytoplasm (0.450)	other (0.793)		
		CA4765	mitochondry (0.853)	mitochondry (0.716)		
		CA5180	cytoplasm (0.450)	other (0.797)		
		CA5892	mitochondry (0.471)	mitochondry (0.637)		

step 3: $\beta$ -1,3-glucanase extract	outside	putative $\beta$ -1,3-glucanase* Hsp150p/Pir2p* Phrp *		
	intracellular	CA1691 CA3874 CA5892	mitochondry (0.360) cytoplasm (0.450) mitochondry (0.471)	other (0.789), mitochondry (0.227) other (0.793) mitochondry (0.637)
step 4: exochitinase extract	outside	Hsp150p/Pir2p* Phrp *		
	intracellular	CA1691 CA2474 CA3874 CA5180 CA5892	mitochondry (0.360) microbody (0.300) cytoplasm (0.450) cytoplasm (0.450) mitochondry (0.471)	other (0.789), mitochondry (0.227) other (0.798) other (0.793) other (0.797) mitochondry (0.637)

colour code:

	proteins found in at least 2 fractions
	proteins found in at least 2 fractions

\* proteins that failed to be identified to already annotated *C. albicans* proteins

(a) <http://genolist.pasteur.fr/CandidaDB/>

(b) PSORT : <http://psort.nibb.ac.jp/form.html> [29]

(b) TargetP: <http://www.cbs.dtu.dk/services/TargetP/> [30]

(d) Two sizes are indicated when different signal peptides are predicted by PSORT and TargetP. The first one is predicted with PSORT.

#### Additional file 4: Bioinformatic analysis of proteins extracted from cell walls of *A. thaliana* etiolated hypocotyls with salts.

Proteins were extracted from *A. thaliana* cell walls of etiolated hypocotyls as described in Figure 5. Two successive extractions were performed : step 1 using  $\text{CaCl}_2$ , and step 2 using  $\text{LiCl}$ . All proteins sequences were analyzed with bioinformatic programs to predict their sub-cellular localization. Proteins for which predictions by different bioinformatic programs are in conflict are classified as "not clear".

step 1:  $\text{CaCl}_2$  extract

Predicted subcellular localization	Gene ( <i>A. thaliana</i> )	PSORT (a)	TargetP (b)	Predicted signal peptide (c)	Aramemnon (d)
transmembrane domain	At1g23480	plasma membrane (0.919)	secretory pathway (0.875)	1-29	yes
	At1g54010	plasma membrane (0.685)	secretory pathway (0.991)	1-29	yes
	At1g55850	plasma membrane (0.600)	other (0.787)		yes
	At1g62660	plasma membrane (0.790)	other (0.822)		yes
	At1g79680	outside (0.820)	secretory pathway (0.929)	1-26 or 1-24	yes
	At3g20580	plasma membrane (0.600)	secretory pathway (0.860)	1-35	yes (GPI anchor)
	At3g45970	endoplasmic reticulum (0.820)	secretory pathway (0.854)	1-20	yes
	At4g15320	plasma membrane (0.600)	other (0.879)		yes
outside	At1g02335	plasma membrane (0.685)	secretory pathway (0.933)	1-22	no
	At1g03220	outside (0.820)	secretory pathway (0.912)	1-22	
	At1g03230	outside (0.820)	secretory pathway (0.910)	1-23	
	At1g09560	outside (0.609)	secretory pathway (0.960)	1-23	
	At1g09750	outside (0.456)	secretory pathway (0.899)	1-21 or 1-23	
	At1g11580	outside (0.820)	secretory pathway (0.423)	1-34	
		mitochondrion (0.850), plasma			
	At1g18970	membrane (0.650)	secretory pathway (0.932)	1-27	no
	At1g20190	outside (0.820)	secretory pathway (0.961)	1-20	
	At1g28290	outside (0.681)	secretory pathway (0.483)	1-24	
	At1g29670	outside (0.820)	secretory pathway (0.987)	1-24	
	At1g33590	outside (0.738)	secretory pathway (0.963)	1-24	no
	At1g65590	outside (0.819)	secretory pathway (0.948)	1-24	
	At1g68560	outside (0.820)	secretory pathway (0.982)	1-27	
	At1g73260	outside (0.542)	secretory pathway (0.985)	1-26	
	At1g76160	outside (0.820)	secretory pathway (0.984)	1-23	
	At1g78830	outside (0.820)	secretory pathway (0.962)	1-22	
	At1g78850	outside (0.820)	secretory pathway (0.973)	1-22	
	At1g78860	outside (0.652)	secretory pathway (0.981)	1-22	
	At2g05580	outside (0.820)	secretory pathway (0.993)	1-20	no
	At2g18140	outside (0.820)	secretory pathway (0.695)	1-20 or 1-16	
	At2g28790	outside (0.705)	secretory pathway (0.990)	1-24 or 1-21	
	At2g30210	outside (0.820)	secretory pathway (0.983)	1-25	
	At2g38530	plasma membrane (0.685)	secretory pathway (0.967)	1-23	no
	At3g08030	outside (0.820)	secretory pathway (0.985)	1-21	
	At3g13790	outside (0.494)	secretory pathway (0.969)	1-20 or 1-28	
	At3g14220	outside (0.609)	secretory pathway (0.926)	1-20 or 1-28	
	At3g14310	plasma membrane (0.460)	secretory pathway (0.227)	1-40 or 1-37	no
	At3g16850	outside (0.820)	secretory pathway (0.965)	1-19 or 1-20	
	At3g18080	outside (0.370)	secretory pathway (0.374)	1-23	

	At3g20370	outside (0.633)	secretory pathway (0.965)	1-25	
	At3g20820	outside (0.820)	secretory pathway (0.752)	1-19	
	At3g21770	outside (0.820)	secretory pathway (0.967)	1-27	
	At3g22640	outside (0.820)	secretory pathway (0.996)	1-22	
	At3g24480	outside (0.795)	secretory pathway (0.978)	1-25	
	At3g25700	endoplasmic reticulum (0.820)	secretory pathway (0.425)	1-23	
	At3g43270	outside (0.820)	secretory pathway (0.979)	1-24	
	At3g49120	outside (0.609)	secretory pathway (0.604)	1-28 or 1-29	
	At3g54400	outside (0.595)	secretory pathway (0.409)	1-19	
	At3g55260	outside (0.738)	secretory pathway (0.805)	1-22 or 1-20	
	At4g08950	outside (0.820)	secretory pathway (0.871)	1-21	
	At4g12880	outside (0.820)	secretory pathway (0.952)	1-18 or 1-26	
	At4g13340	outside (0.820)	secretory pathway (0.945)	1-20	
	At4g16500	outside (0.820)	secretory pathway (0.929)	1-22	
	At4g18970	plasma membrane (0.685)	secretory pathway (0.980)	1-22	no
	At4g29270	outside (0.820)	secretory pathway (0.983)	1-26	
	At4g30170	outside (0.695)	secretory pathway (0.980)	1-25	
	At4g33220	outside (0.709)	secretory pathway (0.919)	1-19	
	At5g02260	plasma membrane (0.685)	secretory pathway (0.969)	1-20	no
	At5g06230	plasma membrane (0.790)	secretory pathway (0.744)	1-38	
	At5g06860	outside (0.733)	secretory pathway (0.964)	1-21	
	At5g07030	outside (0.820)	secretory pathway (0.559)	1-19	
	At5g09440	outside (0.820)	secretory pathway (0.812)	1-19 or 1-23	
	At5g10770	plasma membrane (0.760)	secretory pathway (0.276)	1-25	
	At5g11420	outside (0.820)	secretory pathway (0.989)	1-22	
	At5g12940	outside (0.820)	secretory pathway (0.985)	1-29 or 1-26	
	At5g23210	outside (0.820)	secretory pathway (0.992)	1-25	
	At5g25460	outside (0.820)	secretory pathway (0.979)	1-19	
	At5g26260	outside (0.547)	secretory pathway (0.926)	1-22 or 1-23	
	At5g26280	outside (0.528)	secretory pathway (0.963)	1-22 or 1-23	
	At5g34850	outside (0.748)	secretory pathway (0.870)	1-16 or 1-22	
	At5g34940	outside (0.820)	secretory pathway (0.996)	1-21	
	At5g43060	outside (0.685)	secretory pathway (0.991)	1-21	
	At5g44380	plasma membrane (0.811)	secretory pathway (0.439)	1-32	no
	At5g63810	outside (0.819)	secretory pathway (0.887)	1-29	
	At5g64260	outside (0.685)	secretory pathway (0.966)	1-19 or 1-23	
intracellular	At1g67090*	chloroplast (0.923)	chloroplast (0.769)		
	At3g09260	endoplasmic reticulum (0.910) (C-term KDEL)	secretory pathway (0.952)	1-24	
	At3g16880	cytoplasm (0.450)	other (0.625)		
	At4g15440	microbody (0.540)	other (0.891)		
	At4g22165	microbody (0.640)	other (0.493)		
	At4g23670	microbody (0.560)	other (0.847)		
	At4g28520°	outside (0.820)	secretory pathway (0.962)	1-23	
	At5g38410*	outside (0.370)	chloroplast (0.741)		
	At5g38420*	endoplasmic reticulum (0.550)	chloroplast (0.806)		
not clear	At5g38430*	outside (0.370)	chloroplast (0.807)		
	At5g44120°	outside (0.700)	secretory pathway (0.975)	1-24 or 1-23	
not clear	At2g21540	plasma membrane (0.700)	other (0.945)		no
	At3g26380	outside (0.690)	other (0.355), secretory pathway (0.168)	1-32	yes
	At3g32980	mitochondry (0.660)	secretory pathway (0.356)	1-29	

step 2: LiCl extract

transmembrane domain	At1g54010	plasma membrane (0.685)	secretory pathway (0.991)	1-29	yes
	At1g73620	Golgi (0.900), plasma membrane (0.790)	other (0.532)		yes
	At4g16590	plasma membrane (0.600)	other (0.879)		yes
	At5g20630	plasma membrane (0.460)	secretory pathway (0.922)	1-20	yes
outside	At1g03220	outside (0.820)	secretory pathway (0.912)	1-22	
	At1g03230	outside (0.820)	secretory pathway (0.910)	1-23	
	At1g09750	outside (0.456)	secretory pathway (0.899)	1-21 or 1-23	
	At1g17860	outside (0.690)	secretory pathway (0.910)	1-19	
	At1g29670	outside (0.820)	secretory pathway (0.987)	1-24	
	At1g47128	outside (0.771)	secretory pathway (0.993)	1-21	
	At1g68560	outside (0.820)	secretory pathway (0.982)	1-27	
	At1g73260	outside (0.542)	secretory pathway (0.985)	1-26	
	At1g78850	outside (0.820)	secretory pathway (0.973)	1-22	
	At2g05580	outside (0.820)	secretory pathway (0.993)	1-20	
	At2g22170	outside (0.820)	secretory pathway (0.990)	1-21	
	At2g28790	outside (0.705)	secretory pathway (0.994)	1-24 or 1-21	
	At2g34700	outside (0.432)	secretory pathway (0.790)	1-23 or 1-28	
	At3g14310	plasma membrane (0.460)	secretory pathway (0.227)	1-40 or 1-37	no
	At3g16850	outside (0.820)	secretory pathway (0.965)	1-19 or 1-20	
	At3g21770	outside (0.820)	secretory pathway (0.967)	1-27	
	At3g49120	outside (0.609)	secretory pathway (0.604)	1-28 or 1-29	
	At3g54400	outside (0.595)	secretory pathway (0.409)	1-19	
	At4g30170	outside (0.695)	secretory pathway (0.980)	1-25	
	At4g34980	outside (0.820)	secretory pathway (0.980)	1-20	
	At5g06860	outside (0.733)	secretory pathway (0.964)	1-21	
	At5g07030	outside (0.820)	secretory pathway (0.559)	1-19	
	At5g09440	outside (0.820)	secretory pathway (0.812)	1-19 or 1-23	
	At5g11420	outside (0.820)	secretory pathway (0.989)	1-22	
	At5g25460	outside (0.820)	secretory pathway (0.979)	1-19	
	At5g26280	outside (0.528)	secretory pathway (0.963)	1-22 or 1-23	
	At5g51260	outside (0.820)	secretory pathway (0.938)	1-19	
	At5g59090	outside (0.820)	secretory pathway (0.978)	1-24	
	At5g66390	outside (0.820)	secretory pathway (0.973)	1-23	
intracellular	At5g20830	microbody (0.475)	other (0.669)	1-20	
not clear	At3g42160	plasma membrane (0.650)	other (0.925)		no

colour code:

proteins found at both steps 1 and 2  
proteins found at both steps 1 and 2



\* small subunit of RUBISCO

° homolog to storage proteins, most probably vacuolar localization

(a) PSORT : <http://psort.nibb.ac.jp/form.html> [29]

(b) TargetP: <http://www.cbs.dtu.dk/services/TargetP/> [30]

(c) Two sizes are indicated when different signal peptides are predicted by PSORT and TargetP. The first one is predicted with PSORT.

(d) Aramemnon: <http://aramemnon.botanik.uni-koeln.de/> [31]



### Additional file 5: Bioinformatic analysis of proteins extracted from cell walls of *A. thaliana* etiolated hypocotyls with SDS and DTT.

Proteins were extracted from *A. thaliana* cell walls of etiolated hypocotyls as described in Figure 6. One extraction was performed using SDS and DTT. All proteins sequences were analyzed with bioinformatic programs to predict their sub-cellular localization. Proteins for which predictions by different bioinformatic programs are in conflict are classified as “not clear”.

step 3: SDS-DTT

Predicted subcellular localization	Gene ( <i>A. thaliana</i> )	PSORT (a)	TargetP (b)	Predicted signal peptide (c)	Aramemnon (d)
transmembrane domain	At2g25300	chloroplast thylakoid membrane (0.765), plasma membrane (0.650)	chloroplast (0.573)		yes
	At4g33230	chloroplast thylakoid membrane (0.615), plasma membrane (0.600)	other (0.932)		yes
	At4g25080	endoplasmic reticulum (0.600), mitochondrion (0.453), chloroplast (0.376)	chloroplast (0.866)		yes
	At2g24820	mitochondrion (0.791), microbody (0.635), chloroplast (0.500)	chloroplast (0.897)		yes
		mitochondrion (0.861), chloroplast (0.660), plasma membrane (0.650)			
	At4g15820	outside (0.820)	mitochondrion (0.537)	1-29	yes
	At1g72990		mitochondrion (0.957)		yes
	At1g27200	plasma membrane (0.790)	other (0.521)		yes
	At3g05610	plasma membrane (0.790)	mitochondrion (0.537)		yes
	At3g61270	plasma membrane (0.790)	other (0.843)		yes
	At4g15320	plasma membrane (0.600)	other (0.879)		yes
	At4g16590	plasma membrane (0.600)	other (0.879)		yes
	At5g41390	plasma membrane (0.600)	mitochondrion (0.750), other (0.544)		yes
outside	At1g66270	outside (0.820)	secretory pathway (0.976)	1-24	
	At1g71695	outside (0.820)	secretory pathway (0.829)	1-22 or 1-31	
	At2g46570	outside (0.370)	secretory pathway (0.868)	1-20 or 1-29	
	At3g14040	outside (0.820)	secretory pathway (0.984)	1-23 or 1-25	
	At4g11310	outside (0.820)	secretory pathway (0.995)	1-18 or 1-23	
	At4g12880	outside (0.820)	secretory pathway (0.952)	1-18 or 1-26	
	At4g28790	outside (0.820)	secretory pathway (0.726)	1-17 or 1-24	
	At4g33810	outside (0.820)	secretory pathway (0.910)	1-24	
	At5g07030	outside (0.820)	secretory pathway (0.559)	1-19	
	At5g25460	outside (0.820)	secretory pathway (0.979)	1-19	
		mitochondrion (0.850), outside (0.820)			
	At5g44020		secretory pathway (0.845)	1-16 or 1-21	

intracellular	ArthCp029	chloroplast		
	ArthCp030	chloroplast		
	At1g55490	chloroplast (0.864)	chloroplast (0.979)	
	At1g67090*	chloroplast (0.923)	chloroplast (0.769)	
	At2g21170	chloroplast (0.888)	chloroplast (0.950)	
	At3g04350	chloroplast (0.842)	other (0.635), chloroplast (0.451)	
	At3g13470	chloroplast (0.950)	chloroplast (0.924)	
	At3g53460	chloroplast (0.903)	chloroplast (0.903)	
	At4g15040	chloroplast (0.496)	other (0.625), chloroplast (0.289)	
	At4g20360	chloroplast (0.950)	chloroplast (0.975)	
	At5g20720	chloroplast (0.895)	chloroplast (0.901)	
	At5g54190	chloroplast (0.924)	chloroplast (0.957)	
		endoplasmic reticulum (0.910) (C-term KDEL)	secretory pathway (0.952)	1-24
	At3g09260		secretory pathway (0.962)	1-23
	At4g28520°	outside (0.820)	secretory pathway (0.975)	1-24 or 1-23
	At5g44120°	outside (0.700)		
	At1g49240	cytoplasm (0.450)	other (0.891)	
	At3g18780	cytoplasm (0.450)	other (0.883)	
	At4g11850	cytoplasm (0.450)	other (0.806)	
	At5g17920	cytoplasm (0.450)	other (0.713)	
	At4g02520	microbody (0.520)	other (0.556)	
	At2g13890	endoplasmic reticulum (0.850)	other (0.609)	
	At1g13440	microbody (0.539)	other (0.610)	
	At1g60120	microbody (0.704)	other (0.753)	
	At1g78380	microbody (0.640)	other (0.523)	
	At3g02230	microbody (0.603)	other (0.765)	
	At3g04120	microbody (0.472)	other (0.619)	
	At3g60120	microbody (0.640)	other (0.712)	
	At5g58690	microbody (0.499)	other (0.598)	
not clear	At3g49400	plasma membrane (0.650)	other (0.338)	no
	At5g37470	plasma membrane (0.600)	other (0.973)	no

colour code:

proteins also extracted by CaCl<sub>2</sub> or LiCl

proteins also extracted by CaCl<sub>2</sub> or LiCl

proteins also extracted by CaCl<sub>2</sub> or LiCl



\* small subunit of RUBISCO

° homolog to storage proteins, most probably vacuolar localization

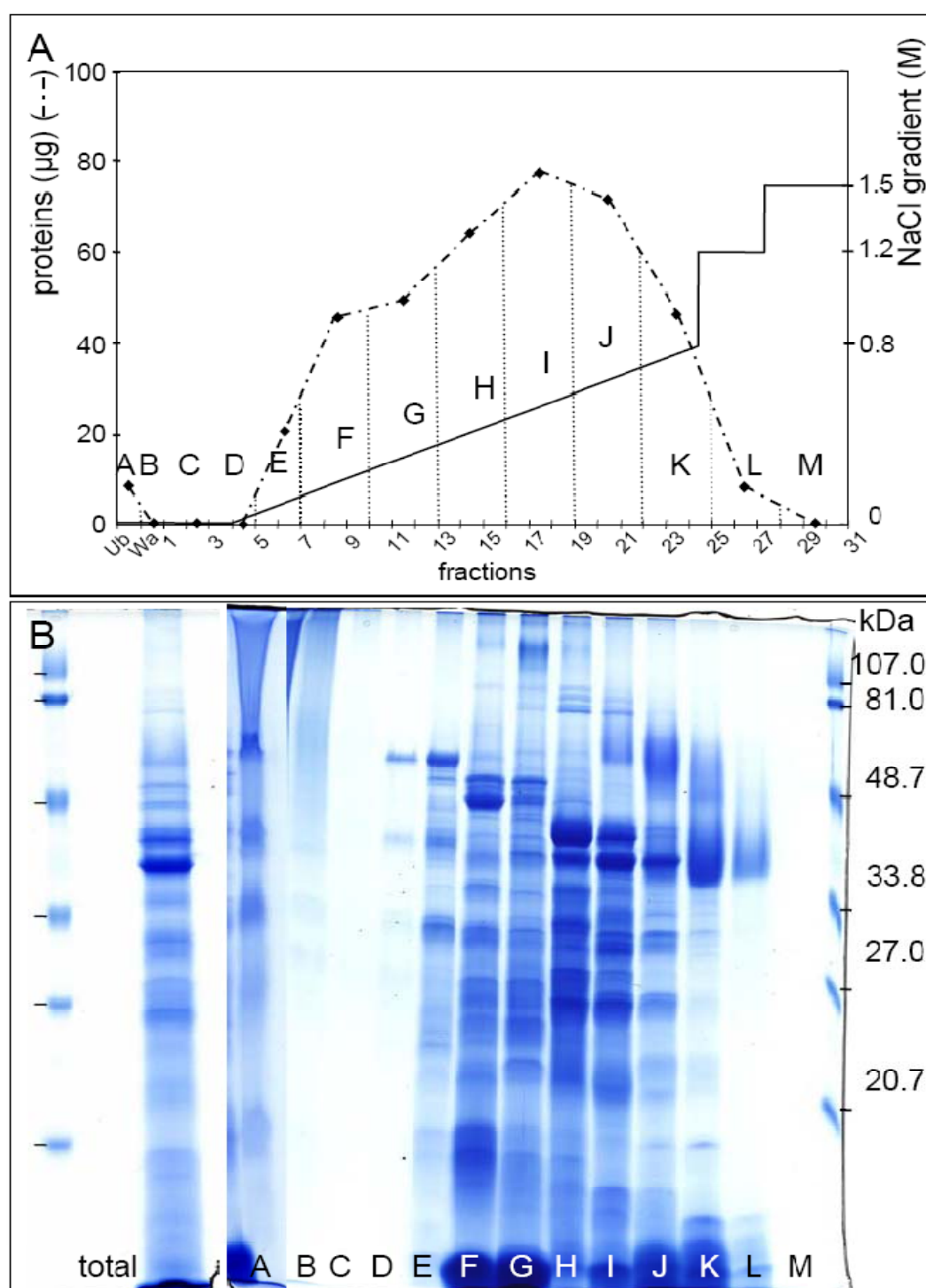
(a) PSORT : <http://psort.nibb.ac.jp/form.html> [29]

(b) TargetP: <http://www.cbs.dtu.dk/services/TargetP/> [30]

(c) Two sizes are indicated when different signal peptides are predicted by PSORT and TargetP. The first one is predicted with PSORT.

(d) Aramemnon: <http://aramemnon.botanik.uni-koeln.de/> [31]

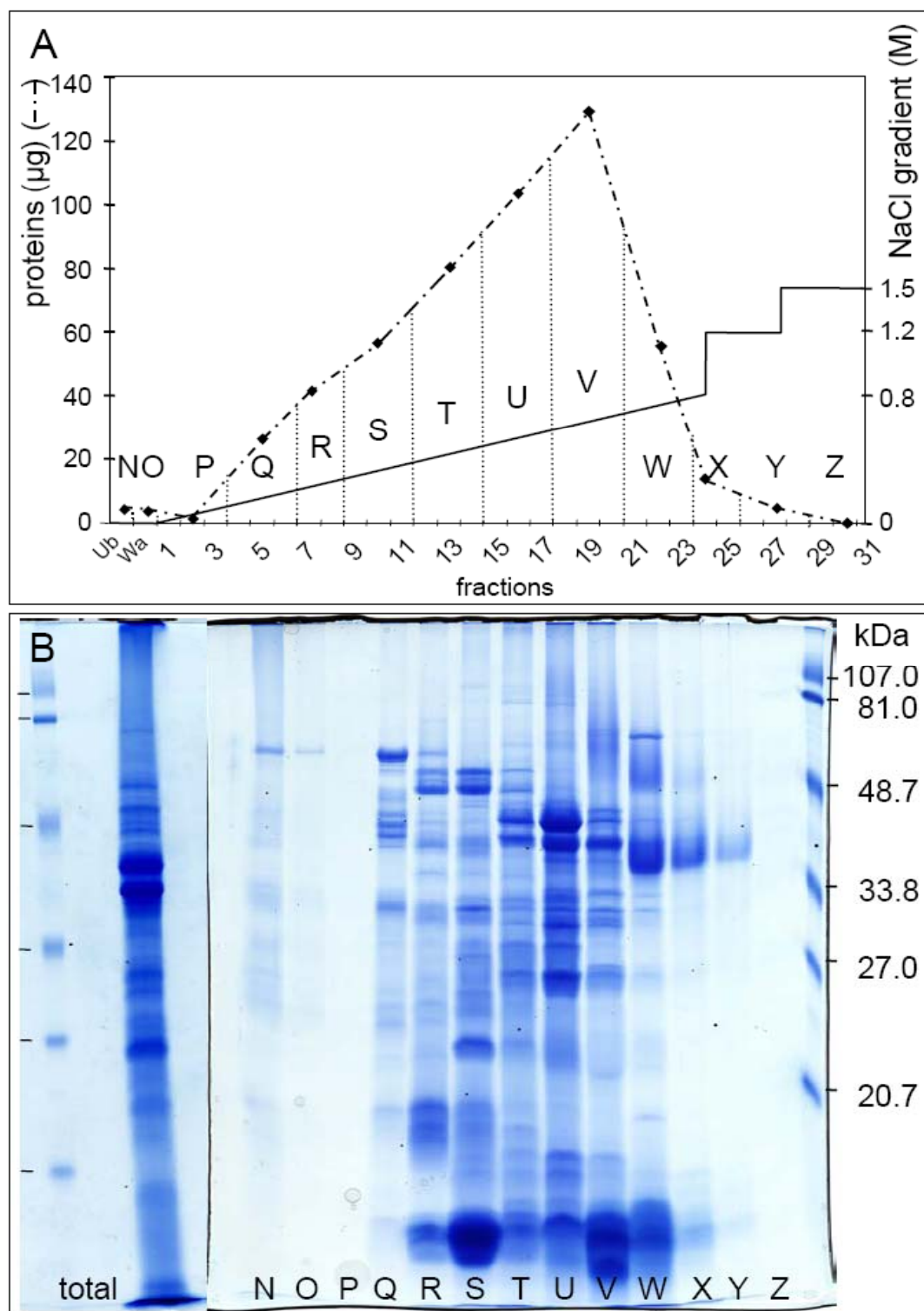
## **SUPPLEMENTARY DATA CHAPTER 4**



**Figure S1, supplementary material. Separation of proteins extracted with  $\text{CaCl}_2$  and  $\text{LiCl}$  from purified cell walls of 5-day-old etiolated hypocotyls by cation exchange chromatography followed by 1D-E.**

**a.** Graph showing amounts of proteins (in  $\mu\text{g}$ ) in each fraction eluted by a NaCl gradient (from 0 to 0.8 M) followed by two steps at 1.2 and 1.5 M NaCl. Dotted vertical lines show the grouping of chromatography fractions whereas the letters (from A to M) correspond to those analyzed by 1D-E.

**b.** 1D-E of total CWP (total) and chromatography fractions (Ub stands for unbound proteins, Wa for proteins eluted with loading buffer). Molecular mass markers are in kDa.



**Figure S2, supplementary material. Separation of proteins extracted with  $\text{CaCl}_2$  and  $\text{LiCl}$  from purified cell walls of 11-day-old etiolated hypocotyls by cation exchange chromatography followed by 1D-E.**

**a.** Graph showing amounts of proteins (in  $\mu\text{g}$ ) in each fraction eluted by a NaCl gradient (from 0 to 0.8 M) followed by two steps at 1.2 and 1.5 M NaCl. Dotted vertical lines show the grouping up of chromatography fractions whereas the letters (from N to Z) correspond to those analyzed by 1D-E.

**b.** 1D-E of total CWP (total) and chromatography fractions (Ub stands for unbound proteins, Wa for proteins eluted with loading buffer). Molecular mass markers are in kDa.

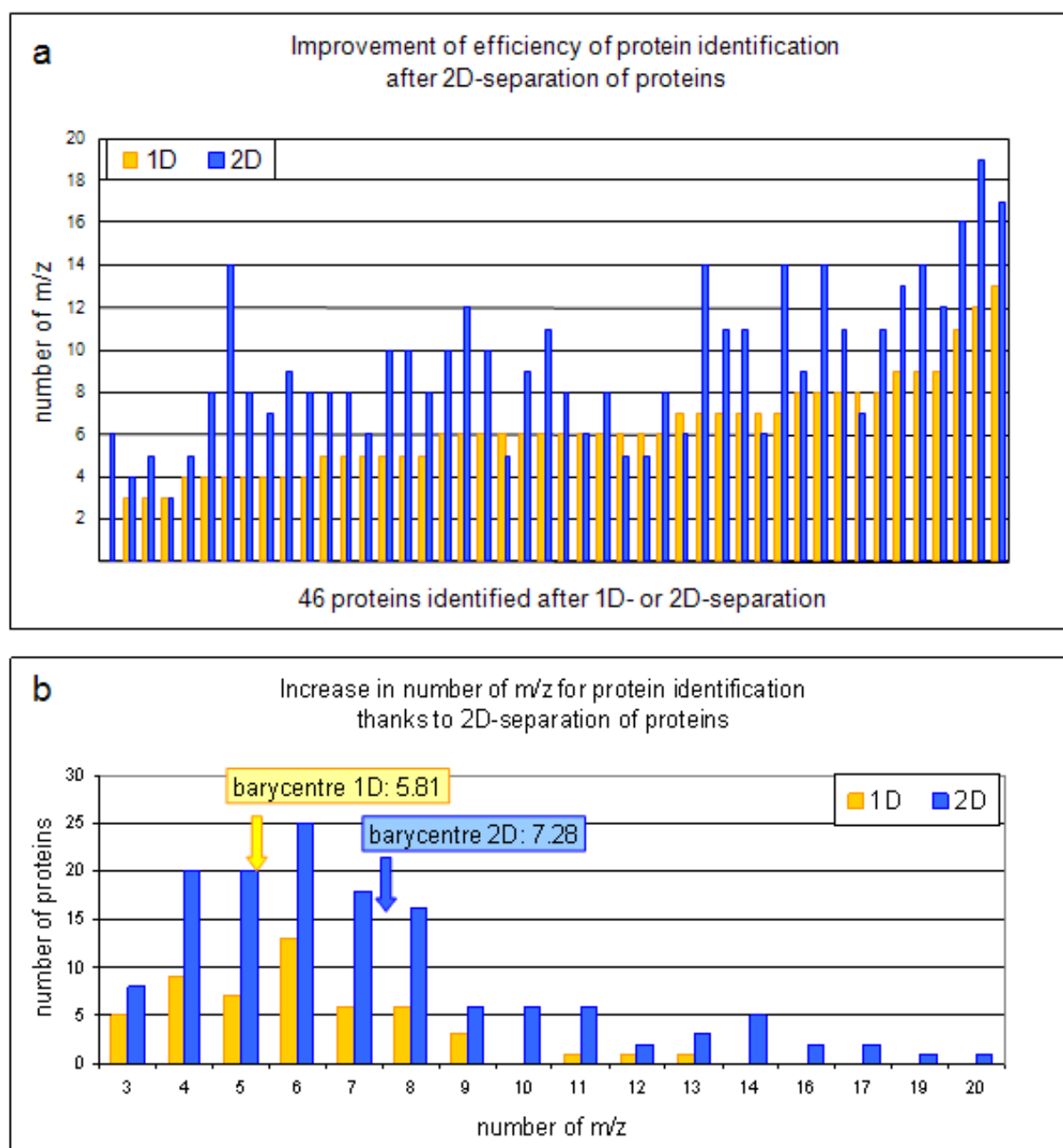
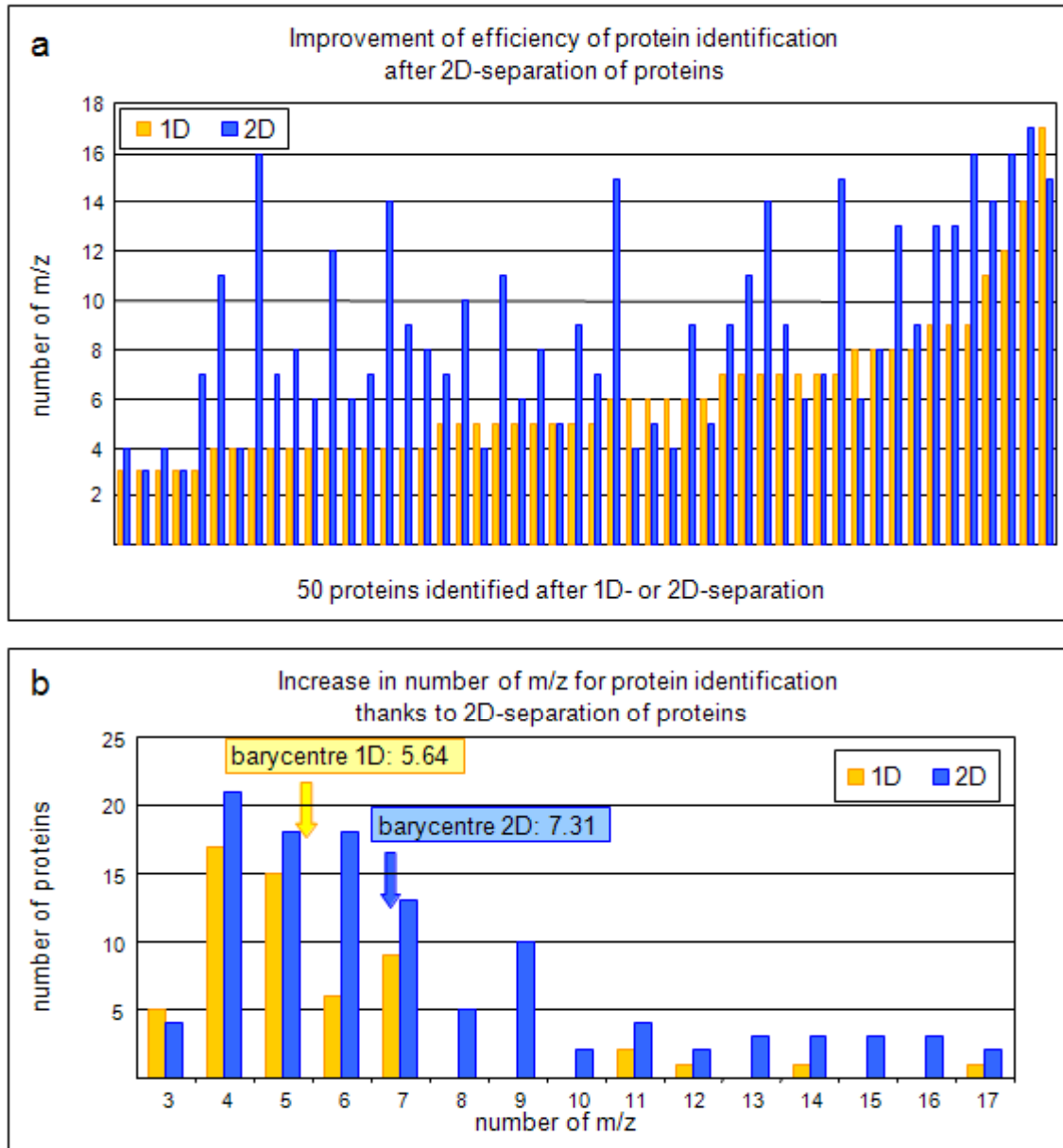


Figure S3, supplementary material. Identification of proteins extracted by  $\text{CaCl}_2$  and  $\text{LiCl}$  from purified cell walls of 5-day-old *Arabidopsis* etiolated hypocotyls by MALDI-TOF MS. **a.** Improvement of efficiency of protein identification after 2D-separation of proteins. Numbers of m/z found for identification of proteins after 1D- (yellow bars) and 2D- (blue bars) separation are shown. Note that only 6 out of 46 proteins are identified with less m/z after 1D- than after 2D-separation. **b.** Increase in number of m/z for protein identification thanks to 2D-separation of proteins. Numbers of proteins identified with 3 to 20 m/z are shown after 1D- (yellow bars) and 2D- (blue bars) separation. Note that the barycentre of this number of m/z used for successful identification is higher after 2D- than after 1D-separation.



**Figure S4, supplementary material. Identification of proteins extracted by CaCl<sub>2</sub> and LiCl from purified cell walls of 11-day-old *Arabidopsis* etiolated hypocotyls by MALDI-TOF MS.**

**a.** Improvement of efficiency of protein identification after 2D-separation of proteins. Numbers of m/z found for identification of proteins after 1D- (yellow bars) and 2D- (blue bars) separation are shown. Note that only 7 out of 50 proteins are identified with less m/z after 1D- than after 2D-separation. **b.** Increase in number of m/z for protein identification thanks to 2D-separation of proteins. Numbers of proteins identified with 3 to 20 m/z are shown after 1D- (yellow bars) and 2D- (blue bars) separation. Note that the barycentre of this number of m/z used for successful identification is higher after 2D- than after 1D-separation.

Table S1, supplementary data

**Table S1, supplementary material.** Identification of proteins extracted from purified cell walls of 5-day-old *Arabidopsis* hypocotyls by CaCl<sub>2</sub> and LiCl solutions.

\* or # proteins that could not be distinguished

Colour code	secretory proteins (Presence of predicted signal peptide was checked using PSORT: <a href="http://psort.ims.u-tokyo.ac.jp/form.html">http://psort.ims.u-tokyo.ac.jp/form.html</a> , and TargetP: <a href="http://www.cbs.dtu.dk/services/TargetP/">http://www.cbs.dtu.dk/services/TargetP/</a> )
	transmembrane domains (Presence of predicted transmembrane domains was checked using Aramemnon: <a href="http://aramemnon.botanik.uni-koeln.de/">http://aramemnon.botanik.uni-koeln.de/</a> )
	GPI anchors (presence or GPI anchors was checked using Aramemnon: <a href="http://aramemnon.botanik.uni-koeln.de/">http://aramemnon.botanik.uni-koeln.de/</a> )
	intracellular proteins
	proteins identified after 1D-separation (1-DE), but not identified after 2D-separation (cationic exchange chromatography followed by 1-DE)

total protein extract	band number	accession AGI	annotation	MALDI-TOF identification % sequence coverage	number of peptides (m/z) <sup>a</sup>	mature protein MM (kDa)	pI	signal peptide	transmembrane domains GPI anchors	functional domains <sup>b</sup>
	1	<b>At5g17820</b>	peroxidase (AtPrx57)	23,0	4	31636,96	11,42	1-22		peroxidase active site (PS00436)
	2, 3	<b>At1g68560</b>	glycoside hydrolase family 31 (alpha-xylosidase) (AtXYL1)	16,4	12	99641,37	6,26	1-27		glycoside hydrolases family 31 active site (PS00129)
	4, 6	<b>At1g28290</b>	proline-rich protein	9,7	4	35913,36	10,77	1-24		histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
	5-7	<b>At3g16850</b>	glycoside hydrolase family 28 (polygalacturonase)	27,9	8	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
	7	<b>At2g27190</b>	homologous to purple acid phosphatase (PAP1)	9,8	4	50962,71	5,75	1-28		metallo-phosphoesterase motif (PS50185), calcineurin-like phosphoesterase (PF00149)
	8, 9, 36-38	<b>At1g78850*</b>	homologous to lectin (curculin-like)	14,5	6	46664,12	8,90	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
	8, 9	<b>At1g78860*</b>	homologous to lectin (curculin-like)	14,4	6	46804,29	6,00	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
	9-11	<b>At1g03220</b>	homologous to carrot EDGP and tomato XEGIP	17,3	6	43394,37	9,99	1-22		
	9	<b>At5g34940</b>	glycoside hydrolase family 79 (endo beta-glucuronidase/heparanase)	10,6	4	57219,88	9,13	1-21		glycoside hydrolase family 79, N-terminal domain (PF03662)
	10, 11	<b>At1g09750</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	16,9	8	45038,78	9,09	1-23		lipase Ser active site (PS00120), Ser-rich region (PS050324), eukaryotic aspartic protease (PF00026)
	10-12	<b>At1g03230</b>	homologous to carrot EDGP and tomato XEGIP	15,9	6	43723,02	10,27	1-23		
	12	<b>At3g54400</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	9,9	4	43313,31	10,26	1-19		eukaryotic aspartyl protease (PF00026)
	13	<b>At3g43270</b>	carbohydrate esterase family 8 (pectin methylesterase)	17,5	9	54929,20	9,55	1-24		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor PME1 (PF04043)
	13-15, 18, 32-34	<b>At5g11420</b>	expressed protein (DUF642)	28,4	11	37344,57	7,70	1-22		domain of unknown function DUF 642 (PF04862)
	13	<b>At5g07030</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	11,8	5	44860,54	10,82	1-19		eukaryotic aspartyl protease (PF00026)
	13, 15	<b>At1g19170</b>	glycoside hydrolase family 28 (polygalacturonase)	11,8	6	49948,02	9,65	1-47		glycosyl hydrolase family 28 (PF00295)
	13	<b>At4g33220</b>	carbohydrate esterase family 8 (pectin methylesterase)	15,6	7	56832,32	6,24	1-19		pectinesterase (PF01095)
	15	<b>At3g49120</b>	peroxidase (AtPrx34)	17,8	6	35695,45	9,41	1-30		peroxidase active site signature (PS00436), peroxidases proximal heme-ligand signature (PS00435)
	15, 23, 40, 43-46	<b>At4g32460</b>	expressed protein (DUF642)	22,7	8	37375,51	9,25	1-21		domain of unknown function DUF 642 (PF04862)
	16, 17	<b>At3g20820</b>	expressed protein (LRR domains)	26,0	8	37850,70	9,90	1-19		typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22-like LRR profile (PS50504)
	16-21, 36-38, 40	<b>At5g25460</b>	expressed protein (DUF642)	35,0	13	37948,52	9,03	1-19		domain of unknown function DUF 642 (PF04862)
	17, 18, 25, 30	<b>At3g08030</b>	expressed protein (DUF642)	23,6	8	36993,02	6,87	1-21		domain of unknown function DUF 642 (PF04862)
	18, 36, 38, 39	<b>At1g70850</b>	homologous to Bet v I allergen family	20,9	6	35569,93	4,96			pathogenesis-related protein Bet v I family (PF00407)
	18-20	<b>At5g12940</b>	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	21,6	7	36784,96	10,45	1-29		plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
	18	<b>At1g78830</b>	homologous to lectin (curculin-like)	11,2	4	48101,66	9,45	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
	20, 21	<b>At1g29670</b>	homologous to lipase/acylhydrolase (GDSL family)	23,1	6	37153,43	9,65	1-24		GDSSL-like Lipase/Acylhydrolase (PF00657)
	20	<b>At5g64100</b>	peroxidase (AtPrx69)	11,2	5	33228,93	10,72	1-23		peroxidases active site signature (PS00436)
	22	<b>At1g53240*</b>	homologous to mitochondrial NAD-dependent malate dehydrogenase	16,7	5	35804,43	9,49			malate dehydrogenase active site signature (PS00068), lactate/malate dehydrogenase, NAD binding domain (PF00056), lactate/malate dehydrogenase, alpha/beta C-terminal domain (PF02866)
	22	<b>At3g15020*</b>	homologous to mitochondrial NAD-dependent malate dehydrogenase	16,7	5	35875,41	9,34			malate dehydrogenase active site signature (PS00068), lactate/malate dehydrogenase, NAD binding domain (PF00056), lactate/malate dehydrogenase, alpha/beta C-terminal domain (PF02866)



Table S1, supplementary data

	23	<b>At3g14310</b> carbohydrate esterase family 8 (pectin methylesterase) (AtPME3)	18,9	9	60050,61	9,59	1-40	transmembrane domains	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095, plant invertase/pectin methylesterase inhibitor (PF04043)
	24, 25	<b>At4g30170</b> peroxidase (AtPrx45)	31,7	9	33052,57	10,43	1-25		haem peroxidase, plant/fungal/bacterial (PS00435, PS00436, PS00873)
	24	<b>At2g06850</b> glycoside hydrolase family 16 (endoxyloglucan transferase) (AtXTH4)	20,3	7	31716,77	9,65	1-24		glycoside hydrolases family 16 (PF00722)
	25	<b>At3g50990</b> peroxidase (AtPrx36)	10,4	3	35273,36	4,57	1-20		peroxidases proximal heme-ligand signature (PS00435), peroxidases active site signature (PS00436), plant heme peroxidase family profile (PS00873), peroxidase (PF00141)
	25-27	<b>At3g45970</b> expansin-like A (AtEXLA1)	18,9	4	26258,12	9,46	1-20		pollen allergen (PF01357), barwin (PF00967)
	26-28	<b>At5g43060</b> homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	13,4	6	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
	26-28	<b>At5g09440</b> homologous to <i>Nicotiana tabacum</i> phi-I	20,5	7	27366,44	10,15	1-23		phosphate-induced protein 1 conserved region (PF04674)
	26	<b>At5g64260</b> homologous to <i>Nicotiana tabacum</i> phi-I	14,1	8	30550,86	9,98	1-19		phosphate-induced protein 1 conserved region (PF04674)
	27	<b>At4g38400</b> expansin-like A (AtEXLA2)	16,6	5	26516,93	9,60	1-20		expansin, cellulose-binding-like domain profile (PS00843), expansin, family-45 endoglucanase-like domain profile (PS00842), rare lipoprotein A (RlpA)-like double-psi beta-barrel (PF03330), pollen allergen (PF01357)
	29	<b>At3g22640</b> expressed protein (cupin domain)	12,3	7	52687,99	6,45	1-22		cupin domain profile (PS00849), cupin (PF00190), cupin domain
	29-31	<b>At2g28790</b> homologous to <i>Lycopersicon esculentum</i> osmotin	12,9	3	24551,60	9,53	1-24		thaumatin family (PF00314)
	31-34	<b>At2g28490</b> expressed protein (cupin domain)	12,3	6	53594,59	5,45	1-19		cupin domain profile (PS00849), cupin (PF00190)
	31	<b>At4g29270</b> homologous to acid phosphatase	18,0	4	25900,84	9,15	1-26		HAD superfamily, subfamily IIIB (acid phosphatase) (PF03767)
	33	<b>At2g40880</b> inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	31,2	3	11955,61	9,71	1-22		cysteine proteases inhibitors signature (PS00287), cystatin domain (PF00031)
	33-35	<b>At1g20190</b> alpha-expansin (AtEXPA11)	19,8	7	24817,74	9,84	1-20		expansin, family-45 endoglucanase-like domain profile (PS00842), expansin, cellulose-binding-like domain profile (PS00843), pollen allergen (PF01357)
	33-37	<b>At4g12880</b> early nodulin (AtEN20) (phytoecyanin)	24,8	5	14273,34	9,64	1-18		plastocyanin-like domain (PF02298) (copper binding proteins)
	36-38	<b>At3g02110</b> homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	14,4	5	51075,46	9,35	1-22		serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
	37	<b>At3g21770</b> peroxidase (AtPrx30)	15,2	4	32901,55	10,54	1-27		peroxidases active site signature (PS00436)
	41	<b>At2g05580</b> glycine-rich protein	21,8	3	21517,58	10,21	1-20		glycine-rich region profile (PS00315)
	43-47	<b>At1g67090</b> RUBISCO small subunit A1	26,1	6	20216,04	9,26			ribulose biphosphate carboxylase, small chain (PF00101)
	42-44	<b>At5g38410*</b> RUBISCO small subunit B3	30,9	6	20284,25	9,40			ribulose biphosphate carboxylase, small chain (PF00101)
	43	<b>At5g38420*</b> RUBISCO small subunit B2	26,0	6	20350,22	9,24			ribulose biphosphate carboxylase, small chain (PF00101)
	48	<b>At2g38530</b> non-specific lipid transfer protein type 1 (LTP2)	25,4	3	9475,92	11,90	1-25		protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00697)
<b>A</b>	1-3, 9, 10, 11	<b>At3g16850</b> glycoside hydrolase family 28 (polygalacturonase)	33,8	14	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
<b>A</b>	4, 5	<b>At1g79720</b> homologous to aspartic protease (CND41 peptidase) (Peptidase family A1, subfamily A1.050, MEROPS)	10,1	5	49324,89	8,88	1-28		eukaryotic and viral aspartyl proteases active site (PS00141), eukaryotic aspartyl protease (PF00026)
<b>A</b>	5-8	<b>At3g32980</b> peroxidase (AtPrx32)	28,7	13	35712,41	6,00	1-29		peroxidases active site signature (PS00436)
<b>A</b>	7, 8	<b>At1g47710</b> homologous to serpin (serine protease inhibitor)	50,6	17	42639,41	4,97		transmembrane domains	serpins signature (PS00284), serpin (serine protease inhibitor) (PF00079)
<b>A</b>	10, 11	<b>At5g22140</b> expressed protein (oxido-reductase domain)	14,2	6	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
<b>A</b>	10-13	<b>At5g43060</b> homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	14,2	5	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
<b>A</b>	11, 12	<b>At1g11840</b> glyoxalase I homologue (lactoylglutathione lyase) (ATGLX1)	27,2	7	31928,38	5,19			glyoxalase I signature 1 (PS00934), glyoxalase I signature 2 (PS00935), glyoxalase/bleomycin resistance protein/dioxygenase superfamily (PF00903)
<b>A</b>	11, 12	<b>At4g01610</b> homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	13,6	5	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
<b>A</b>	13	<b>At2g33160</b> glycoside hydrolase family 28 (polygalacturonase)	8,0	4	72778,41	9,68	1-20		polygalacturonase active site (PS00502), glycoside hydrolases family 28 (PF00295)

Table S1, supplementary data

<b>A</b>	14, 15	<b>At3g11630</b> homologous to 2-Cys peroxiredoxin	24,8	4	29091,87	7,91	1-24	AhpC/TSA family (PF00578) ML domain (PF02221)
<b>A</b>	16-18	<b>At5g23820</b> expressed protein (ML domain - MD-2-related lipid recognition domain)	20,1	4	15240,10	4,17		
<b>A</b>	16, 17	<b>At5g46960</b> homologous to plant invertase/pectin methyltransferase inhibitor (PMEI)	22,4	4	16869,03	4,52	1-20	plant invertase/pectin methyltransferase inhibitor (PF04043)
<b>A</b>	19	<b>At3g20390</b> homologous to endoribonuclease	31,0	5	19815,74	9,57		uncharacterized protein family UPF0076 signature (PS01094), endoribonuclease L-PSP (PF01042)
<b>B</b>	1, 3	<b>At2g38530</b> non-specific lipid transfer protein type 1 (LTP2)	16,1	2	9475,92	11,90	1-25	protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00597)
<b>B</b>	2	<b>At1g47710</b> homologous to serpin (serine protease inhibitor)	17,6	6	42639,41	4,97	transmembrane domains	serpins signature (PS00284), serpin (serine protease inhibitor) (PF00079)
<b>D</b>	1, 2	<b>At3g16850</b> glycoside hydrolase family 28 (polygalacturonase)	29,6	12	46982,42	5,25	1-19	glycoside hydrolase family 28 (PF00295)
<b>D</b>	3, 4	<b>At1g54010</b> homologous to lipase/acylhydrolase (GDSL family)	21,1	5	40021,24	9,05	1-29	GDGL-like Lipase/Acyhydrolase (PF00657)
<b>D</b>	5	<b>At3g32980</b> peroxidase (AtPrx32)	29,0	11	35712,41	6,00	1-29	peroxidases active site signature (PS00436)
<b>D</b>	6	<b>At4g01610</b> homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	15,0	5	36995,63	5,79	1-22	eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
<b>E</b>	1, 5-7, 9	<b>At5g34940</b> glycoside hydrolase family 79 (endo beta-glucuronidase/heparanase)	19,6	8	57219,88	9,13	1-21	glycoside hydrolase family 79, N-terminal domain (PF03662)
<b>E</b>	2-5, 9	<b>At3g16850</b> glycoside hydrolase family 28 (polygalacturonase)	28,5	12	46982,42	5,25	1-19	glycoside hydrolase family 28 (PF00295)
<b>E</b>	5	<b>At3g55260</b> glycoside hydrolase family 20 (beta-hexosaminidase)	19,2	8	58810,86	5,66	1-22	glycoside hydrolase family 20 (PF00295), glycoside hydrolase family 20, catalytic domain (PF00728), glycoside hydrolase family 20, domain 2 (PF02838)
<b>E</b>	8	<b>At5g34850</b> homologous to purple acid phosphatase	22,1	8	53242,23	6,79	1-16	calcineurin-like phosphoesterase (PS00149), phosphoesterase (PF00185)
<b>E</b>	10	<b>At2g17120</b> expressed protein (LysM domain)	17,7	5	35262,82	6,24	1-23	LysM domain (PF01476)
<b>E</b>	11-13, 17	<b>At3g49120</b> peroxidase (AtPrx34)	32,3	8	35695,45	9,41	1-30	peroxidase active site signature (PS00436), peroxidases proximal heme-ligand signature (PS00435)
<b>E</b>	12, 13	<b>At1g54010</b> homologous to lipase/acylhydrolase (GDSL family)	43,0	13	40021,24	9,05	1-29	lipase/acylhydrolase with GDGL-like motif (PS05241), GDGL-like lipase/acylhydrolase (PF00567)
<b>E</b>	12-14	<b>At3g32980</b> peroxidase (AtPrx32)	29,0	11	35712,41	6,00	1-29	peroxidases active site signature (PS00436)
<b>E</b>	14	<b>At1g70850</b> homologous to Bet v I allergen family	37,0	8	35569,93	4,96		pathogenesis-related protein Bet v I family (PF00407)
<b>E</b>	14, 15	<b>At4g19810</b> glycoside hydrolase family 18 (chitinase)	32,2	9	38510,49	9,37	1-24	glycoside hydrolases family 18 (PF00704)
<b>E</b>	14	<b>At1g54030</b> homologous to lipase/acylhydrolase (GDGL family)	11,6	3	46082,76	6,20	transmembrane domains	lipase/acylhydrolase with GDGL-like motif (PS05241), GDGL-like Lipase/Acyhydrolase (PF00657)
<b>E</b>	14	<b>At5g06860</b> PGIP1 (LRR domains)	19,4	5	34324,36	9,56	1-21	LRR (PF00560, PS05052, PS05056)
<b>E</b>	14	<b>At5g39270</b> alpha expansin (AtEXPA22)	11,5	3	27348,10	9,74	1-13	expansin, family-45 endoglucanase-like domain profile (PS050842), expansin, cellulose-binding-like domain profile (PS050843), rare lipoprotein A (RlpA)-like double-psi beta-barrel (PF03330), pollen allergen (PF01357)
<b>E</b>	15, 16	<b>At4g36700</b> expressed protein (cupin domain)	16,4	7	53703,31	5,21	1-25	glutamic acid-rich region profile (PS05313), cupin (PF00190)
<b>E</b>	15-17	<b>At3g20370</b> expressed protein (MATH domain)	24,3	8	40734,84	5,56	1-25	2 MATH domains (PF00917)
<b>E</b>	16, 17	<b>At1g53240</b> homologous to mitochondrial NAD-dependent malate dehydrogenase	16,7	4	35804,43	9,49		malate dehydrogenase active site signature (PS00068), lactate/malate dehydrogenase, NAD binding domain (PF00056), lactate/malate dehydrogenase, alpha/beta C-terminal domain (PF02866)
<b>E</b>	18, 19	<b>At3g12500</b> glycoside hydrolase family 19 (basic endochitinase)	10,2	3	32367,86	6,24	1-20	chitinase family 19 signature 1 (PS00773), signature 2 (PS00774), chitin recognition or binding domain signature (PS00026), chitinase class I (PF00182), chitin recognition protein (PF00187)
<b>E</b>	19	<b>At3g02870</b> homologous to myo-inositol monophosphatase	33,6	9	29121,33	5,03	transmembrane domains	inositol monophosphatase family signature 1 (PS00629), inositol monophosphatase family signature 2 (PS00630), inositol monophosphatase family (PF00459)
<b>E</b>	19-21	<b>At4g01610</b> homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	20,3	7	36995,63	5,79	1-22	eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
<b>E</b>	20-22	<b>At1g11840</b> glyoxalase I homologue (lactoylglutathione lyase) (ATGLX1)	27,2	7	31928,38	5,19		glyoxalase I signature 1 (PS00934), glyoxalase I signature 2 (PS00935), glyoxalase/bleomycin resistance protein/dioxygenase superfamily (PF00903)
<b>E</b>	20, 21	<b>At1g67280</b> glyoxalase I homologue (lactoylglutathione lyase) (ATGLX1)	20,9	6	39167,11	7,73		glyoxalase I signature 1 (PS00934), glyoxalase/bleomycin resistance protein/dioxygenase superfamily (PF00903)
<b>E</b>	21-23	<b>At5g43060</b> homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	15,8	8	36995,63	5,79	1-22	eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)

Table S1, supplementary data

E	21-23	<b>At1g47128</b>	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	11,0	5	48785,27	5,01	1-21		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), cysteine active site (PS00640), asparagine active site (PS00139), Cys-rich region (PS50311), papaine family cysteine protease (PF00112)
E	22	<b>At1g30710</b>	homologous to berberine-bridge enzyme (S)-reticul:n:oxygen oxidoreductase	9,4	5	56821,30	9,63	1-23		berberine and berberine like (PF08031), FAD binding domain (PF01565)
E	24, 25	<b>At3g55440</b>	triose phosphate isomerase	23,2	7	27168,94	5,10			triose phosphate isomerase active site (PS00171)
E	25, 26	<b>At2g28490</b>	expressed protein (cupin domain)	11,0	4	53594,59	5,45	1-19		cupin domain profile (PS50849), cupin (PF00190)
E	27	<b>At2g15220</b>	expressed protein (Plant Basic Secreted Protein domain)	30,2	7	22772,21	9,09	1-21		neutral zinc metalloproteases, zinc-binding region signature (PS00142), plant basic secretory protein (PF04450)
E	28	<b>At1g73260</b>	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	13,0	3	20982,78	5,20	1-26		trypsin and protease inhibitor Kunitz legume (PF00197)
E	28	<b>At5g44360</b>	homologous to berberine-bridge enzyme (S)-reticul:n:oxygen oxidoreductase	8,1	4	57931,72	9,83	1-22		oxygen oxido-reductases covalent FAD binding site (PS00862), FAD binding domain (PF01565), berberine and berberine like (PF08031)
E	29	<b>At5g55730</b>	fasciclin-like arabinogalactan protein (AtFLA1)	10,1	3	42165,78	5,65	1-24	GPI anchor	Beta-Ig-H3/fasciclin domain (PS50213)
E	30, 31	<b>At2g04690</b>	expressed protein (homologous to a human brain CREG protein)	31,9	6	20214,11	8,00	1-28		
E	32	<b>At5g66090</b>	expressed protein	37,1	6	22539,76	6,15			UspA (universal stress protein ) IPR00616
E	33-35	<b>At4g16500</b>	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	33,3	5	10183,74	9,78	1-22		cystatin domain (PF00031)
F	1, 8-10	<b>At1g03220</b>	homologous to carrot EDGP and tomato XEGIP	31,6	10	43394,37	9,99	1-22		
F	2, 4, 46-48	<b>At2g38530</b>	non-specific lipid transfer protein type 1 (LTP2)	33,1	4	9475,92	11,90	1-25		protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00597)
F	3	<b>At2g34930</b>	expressed protein (LRR domains)	22,9	20	97059,60	9,59	1-28		leucine-rich region profile (PS50319), serine-rich region profile (PS50324), typical LRR profile (PS50506), SDS22+-like LRR profile (PS50504), plant-specific LRR profile (PS50502)
F	5	<b>At3g16850</b>	glycoside hydrolase family 28 (polygalacturonase)	24,6	7	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
F	6, 9, 11	<b>At1g78860</b>	homologous to lectin (curculin-like)	21,0	8	46804,29	6,00	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
F	7-9, 11	<b>At1g78850</b>	homologous to lectin (curculin-like)	26,8	11	46664,12	8,90	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
F	8-12, 18-23	<b>At1g09750</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	16,9	9	45038,78	9,09	1-23		lipase Ser active site (PS00120), Ser-rich region (PS050324), eukaryotic aspartic protease (PF00026)
F	11-13	<b>At3g43270</b>	carbohydrate esterase family 8 (pectin methyltransferase)	17,8	14	54929,20	9,55	1-24		pectinesterase (PF01095), plant invertase/pectin methyltransferase inhibitor (PMEI) (PF04043)
F	12-14	<b>At3g49120</b>	peroxidase (AtPrx34)	16,1	6	35695,45	9,41	1-30		peroxidase active site signature (PS00436), peroxidases proximal heme-ligand signature (PS00435)
F	13, 14	<b>At3g48460</b>	homologous to lipase/acylhydrolase (GDLSL family)	22,8	6	39627,85	8,91	1-26		lipase/acylhydrolase with GDLSL-like motif (PS50241), GDLSL-like lipase/acylhydrolase (PF00657)
F	13-17	<b>At1g78830</b>	homologous to lectin (curculin-like)	21,5	8	48101,66	9,45	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
F	14-16	<b>At5g06860</b>	PGIP1 (LRR domains)	27,6	8	34324,36	9,56	1-21		LRR (PF00560, PS50502, PS50506)
F	14-21	<b>At4g36700</b>	expressed protein (cupin domain)	15,2	8	53703,31	5,21	1-25		glutamic acid-rich region profile (PS50313), cupin (PF00190)
F	14	<b>At3g32980</b>	peroxidase (AtPrx32)	15,6	7	35712,41	6,00	1-29		peroxidases active site signature (PS00436)
F	15, 16	<b>At1g70850</b>	homologous to Bet v I allergen family	39,6	9	35569,93	4,96			pathogenesis-related protein Bet v I family (PF00407)
F	15	<b>At1g70830</b>	homologous to Bet v I allergen family	26,0	7	37614,51	5,03			pathogenesis-related protein Bet v I family (PF00407)
F	17-19	<b>At1g29670</b>	homologous to lipase/acylhydrolase (GDLSL family)	38,6	10	37153,43	9,65	1-24		GDLSL-like Lipase/Acylyhydrolase (PF00657)
F	17	<b>At3g01590</b>	homologous to aldose 1-epimerase	19,3	5	34738,64	5,72			aldose 1-epimerase (PF01263)
F	18, 19	<b>At1g53240</b>	homologous to mitochondrial NAD-dependent malate dehydrogenase	24,9	8	35804,43	9,49			malate dehydrogenase active site signature (PS00068), lactate/malate dehydrogenase, NAD binding domain (PF00056), lactate/malate dehydrogenase, alpha/beta C-terminal domain (PF02866)
F	18, 19	<b>At4g18180</b>	glycoside hydrolase family 28 (polygalacturonase)	13,0	4	41118,43	10,42	1-23		polygalacturonase active site (PS00502), glycoside hydrolase family 28 (PF00295)
F	20	<b>At3g14310</b>	carbohydrate esterase family 8 (pectin methyltransferase) (AtPME3)	14,4	7	60050,61	9,59	1-40	transmembrane domains	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methyltransferase inhibitor (PF04043)
F	21	<b>At5g53370</b>	carbohydrate esterase family 8 (pectin methyltransferase)	9,7	5	64240,96	8,80		GPI anchor	plant invertase/pectin methyltransferase inhibitor (PF04043), pectinesterase (PF01095)
F	21	<b>At4g30170</b>	peroxidase (AtPrx45)	14,8	4	33052,57	10,43	1-25		haem peroxidase, plant/fungal/bacterial (PS00435, PS00436, PS50873)

Table S1, supplementary data

F	21, 22	<b>At4g01610</b>	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	15,3	5	36995,63	5,79	1-22	eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
F	22-24	<b>At3g45970</b>	expansin-like A (AtEXLA1)	29,4	8	26258,12	9,46	1-20	pollen allergen (PF01357), barwin (PF00967)
F	22, 23	<b>At1g47128</b>	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	16,2	6	48785,27	5,01	1-21	eukaryotic thiol (cysteine) proteases histidine active site (PS00639), cysteine active site (PS00640), asparagine active site (PS00139), Cys-rich region (PS50311), papaine family cysteine protease (PF00112)
F	22-26	<b>At5g43060</b>	homologous to cysteine proteinase (papain family) (Peptidase family C1, RD21 peptidase, C01.064 MEROPS)	16,0	7	36995,63	5,79	1-22	eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
F	26, 27	<b>At3g55440</b>	triose phosphate isomerase	27,5	6	27168,94	5,10		triose phosphate isomerase active site (PS00171)
F	27-29	<b>At4g29270</b>	homologous to acid phosphatase	29,3	8	25900,84	9,15	1-26	HAD superfamily, subfamily IIIB (acid phosphatase) (PF03767)
F	27-31	<b>At2g28490</b>	expressed protein (cupin domain)	13,7	6	53594,59	5,45	1-19	cupin domain profile (PS50849), cupin (PF00190)
F	27-29	<b>At3g63190</b>	homologous to ribosome recycling factor	25,8	7	30422,14	6,50		ribosome recycling factor (PF01765)
F	27	<b>At4g36880</b>	homologous to cysteine protease (papain family) (Peptidase family C1, Brassicain, C01.021 MEROPS)	16,5	5	39395,74	6,50	1-22	eukaryotic thiol (cysteine) proteases cysteine active site (PS00139), eukaryotic thiol (cysteine) proteases histidine active site (PS00639), eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112)
F	28	<b>At3g57140</b>	homologous to patatin (phospholipase domain)	9,7	6	89933,66	6,31		patatin-like phospholipase (PF01734)
F	28	<b>At4g27110</b>	homologous to COBRA (AICOB10)	8,4	5	71335,23	9,86	1-21	phytochelatin synthetase-like conserved region (PF04833), carbohydrate binding domain (SSF49394)
F	28-32, 35, 36, 38	<b>At4g12880</b>	early nodulin (AtEN20) (phytoecyanin)	24,8	5	14273,34	9,64	1-18	plastocyanin-like domain (PF02298) (copper binding proteins)
F	29	<b>At5g05110</b>	inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	10,3	3	23934,58	8,48	1-24	cysteine proteases inhibitors signature (PS00287), cystatin domain (PF00031)
F	33, 34	<b>At1g17860</b>	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	25,5	7	19718,68	9,76	1-21	soybean trypsin inhibitor (Kunitz) protease inhibitors family signature (PS00283), trypsin and protease inhibitor (PF00197)
F	34-36	<b>At5g01300</b>	homologous to phosphatidylethanolamine-binding protein	50,0	7	71823,40	5,15		phosphatidylethanolamine-binding protein (PF01161),
F	35, 36	<b>At2g04690</b>	expressed protein (homologous to a human brain CREG protein)	28,1	6	20214,11	8,00	1-28	
F	35, 36	<b>At4g25260</b>	homologous to plant invertase/ pectin methyltransferase inhibitor (PMEI)	16,9	4	15737,05	5,50	1-19	plant invertase/pectin methyltransferase inhibitor (PF04043)
F	37, 39-42	<b>At3g15670</b>	homologous to late embryogenesis abundant protein (LEA)	36,4	6	24186,00	9,50		late embryogenesis abundant protein (LEA) (PF02987)
F	39-43	<b>At4g39260</b>	homologous to RNA binding protein (AtGRP8)	41,4	6	16578,63	5,12		glycine-rich region profile (PS50315), eukaryotic RNA Recognition Motif (RRM) profile (PS50102), RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) (PF00076)
F	39-44	<b>At1g09310</b>	expressed protein (DUF538)	38,5	7	19947,00	5,12		domain of unknown function DUF538 (PF04398)
F	40	<b>At4g23680</b>	homologous to Bet v I allergen family	45,0	7	17473,54	5,80		pathogenesis-related protein Bet v I family (PF00407)
F	42-45	<b>At4g01900</b>	P-II nitrogen sensing protein (GLB I)	34,7	7	21275,40	5,12		P-II protein C-terminal region signature (PS00638), nitrogen regulatory protein P-II (PS00543)
F	43-45	<b>At1g67090</b>	RUBISCO small subunit A1	19,4	4	20216,04	9,26		ribulose biphosphate carboxylase, small chain (PF00101)
F	43	<b>At5g38410*</b>	RUBISCO small subunit B3	24,3	4	20284,25	9,40		ribulose biphosphate carboxylase, small chain (PF00101)
F	43	<b>At5g38430*</b>	RUBISCO small subunit B1	19,3	4	20286,18	9,24		ribulose biphosphate carboxylase, small chain (PF00101)
F	48	<b>At1g47540</b>	inhibitor family I18 (mustard trypsin inhibitor-2 family) (family I18 unassigned peptidase inhibitor homologues, MEROPS)	45,9	4	9051,17	4,97	1-18	scorpion toxin-like domain (PF00537)

G	1	<b>At3g15670</b>	homologous to late embryogenesis abundant protein (LEA)	24,0	4	24186,00	9,50		late embryogenesis abundant protein (LEA) (PF02987)
G	2-5, 7, 9-11, 45, 55-59	<b>At2g38530</b>	non-specific lipid transfer protein type 1 (LTP2)	46,6	5	9475,92	11,90	1-25	protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00597)
G	3, 7	<b>At1g05570</b>	glycosyl transferase family 48 (callose synthase) (AtCalS1)	2,8	5	225730,52	9,51		1,3-beta-glucan synthase component (PF02364)
G	5	<b>At2g34930</b>	expressed protein (LRR domains)	10,9	11	97059,60	9,59	1-28	leucine-rich region profile (PS50319), serine-rich region profile (PS50324), typical LRR profile (PS50506), SDS22+like LRR profile (PS50504), plant-specific LRR profile (PS50502)
G	6-8	<b>At1g68560</b>	glycoside hydrolase family 31 (alpha-xylosidase) (AtXYL1)	8,2	7	99641,37	6,26	1-27	glycoside hydrolases family 31 active site (PS00129)
G	10	<b>At3g07130</b>	homologous to purple acid phosphatase	9,6	5	58523,64	5,71	1-17	metallo-phosphoesterase motif (PS50185), calcineurin-like phosphoesterase (PF00149)
G	11, 12	<b>At3g61490</b>	glycoside hydrolase family 28 (polygalacturonase)	10,9	4	49497,06	5,76	1-23	glycoside hydrolase family 28 (PF00295)
G	12, 13	<b>At2g27190</b>	homologous to purple acid phosphatase (PAP1)	19,0	10	50962,71	5,75	1-28	metallo-phosphoesterase motif (PS50185), calcineurin-like phosphoesterase (PF00149)
G	13-15	<b>At1g33590</b>	expressed protein (LRR domains)	40,6	16	48103,60	9,90	1-24	LRR (PF00560), leucine zipper pattern (PS00029)

Table S1, supplementary data

G	13-15	<b>At1g78850</b>	homologous to lectin (curculin-like)	27,0	10	46664,12	8,90	1-22	PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
G	14	<b>At1g78860</b>	homologous to lectin (curculin-like)	20,5	7	46804,29	6,00	1-22	PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
G	14	<b>At5g56490</b>	expressed protein (FAD binding domain)	13,5	7	62107,85	6,49	1-17	FAD binding domain (PF01565)
G	15, 16	<b>At1g03220</b>	homologous to carrot EDGP and tomato XEGIP	23,9	7	43394,37	9,99	1-22	
G	15-17, 19, 21	<b>At1g03230</b>	homologous to carrot EDGP and tomato XEGIP	29,0	10	43723,02	10,27	1-23	
G	17, 18	<b>At1g09750</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	16,9	8	45038,78	9,09	1-23	lipase Ser active site (PS00120), Ser-rich region (PS050324), eukaryotic aspartic protease (PF00026)
G	18, 19, 21	<b>At3g54400</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	16,0	7	43313,31	10,26	1-19	eukaryotic aspartyl protease (PF00026)
G	18	<b>At2g40880</b>	inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	21,6	3	11955,61	9,71	1-22	cysteine proteases inhibitors signature (PS00287), cystatin domain (PF00031)
G	19-22	<b>At3g43270</b>	carbohydrate esterase family 8 (pectin methylesterase)	19,0	14	54929,20	9,55	1-24	pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor PME1) (PF04043)
G	20	<b>At3g02740</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	14,1	4	50452,54	5,38	1-20	eukaryotic and viral aspartyl proteases active site (PS00141), eukaryotic aspartyl protease (PF00026)
G	21, 22	<b>At5g11420</b>	expressed protein (DUF642)	23,2	9	37344,57	7,70	1-22	domain of unknown function DUF 642 (PF04862)
G	22, 23	<b>At3g48460</b>	homologous to lipase/acylhydrolase (GDSL family)	24,9	7	39627,85	8,91	1-26	lipase/acylhydrolase with GDSL-like motif (PS50241), GDSL-like lipase/acylhydrolase (PF00657)
G	23-27, 35, 36	<b>At1g78830</b>	homologous to lectin (curculin-like)	20,0	7	48101,66	9,45	1-22	PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
G	23, 24	<b>At3g20820</b>	expressed protein (LRR domains)	37,5	11	37850,70	9,90	1-19	typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22+like LRR profile (PS50504)
G	24	<b>At5g10770</b>	homologous to aspartic protease (CND41 peptidase) (Peptidase family A1, A01.050, MEROPS)	18,8	7	37068,36	5,48	1-22	eukaryotic aspartyl protease (PF00013)
G	25, 26	<b>At1g78820</b>	homologous to lectin (curculin-like)	15,2	6	48184,38	9,54	1-22	D-mannose binding lectin (PF01453), PAN domain (PF00024)
G	26-29	<b>At1g29670</b>	homologous to lipase/acylhydrolase (GDSL family)	35,3	9	37153,43	9,65	1-24	GDGL-like Lipase/Acylhydrolase (PF00657)
G	27, 29, 30	<b>At4g36700</b>	expressed protein (cupin domain)	16,4	9	53703,31	5,21	1-25	glutamic acid-rich region profile (PS50313), cupin (PF00190)
G	28, 29	<b>At1g53070</b>	homologous to lectin (legume lectin domains)	28,3	8	27805,05	9,15	1-23	legume lectins alpha domain (PF00138), legume lectins beta domain (PF00139)
G	28	<b>T23B7.10</b>	homologous to PGIP1 (LRR protein FLR1)	23,1	4	33134,88	9,62	1-23	typical LRR profile (PS50506), plant specific LRR profile (PS50502), plant disease resistance response protein (PF03018)
G	30-32	<b>At4g30170</b>	peroxidase (AtPrx45)	40,9	12	33052,57	10,43	1-25	haem peroxidase, plant/fungal/bacterial (PS00435, PS00436, PS50873)
G	30, 31	<b>At2g06850</b>	glycoside hydrolase family 16 (endoxyloglucan transferase) (AUXTH4)	17,6	6	31716,77	9,65	1-24	glycoside hydrolases family 16 (PF00722)
G	31-35	<b>At3g45970</b>	expansin-like A (AtEXLA1)	33,6	8	26258,12	9,46	1-20	pollen allergen (PF01357), barwin (PF00967)
G	32-35	<b>At4g38400</b>	expansin-like A (AtEXLA2)	29,1	10	26516,93	9,60	1-20	expansin, cellulose-binding-like domain profile (PS50843), expansin, family-45 endoglucanase-like domain profile (PS50842), rare lipoprotein A (RlpA)-like double-psi beta-barrel (PF03330), pollen allergen (PF01357)
G	34-38, 40, 41	<b>At3g22640</b>	expressed protein (cupin domain)	19,5	10	52687,99	6,45	1-22	cupin domain profile (PS50849), cupin (PF00190), cupin domain
G	34, 35	<b>At5g09440</b>	homologous to <i>Nicotiana tabacum</i> phi-I	17,3	5	27366,44	10,15	1-23	phosphate-induced protein 1 conserved region (PF04674)
G	35-46, 49, 50	<b>At4g12880</b>	early nodulin (AtEN20) (phytoecyanin)	30,5	6	14273,34	9,64	1-18	plastocyanin-like domain (PF02298) (copper binding proteins)
G	37-40	<b>At4g29270</b>	homologous to acid phosphatase	38,7	9	25900,84	9,15	1-26	HAD superfamily, subfamily IIIB (acid phosphatase) (PF03767)
G	38-40	<b>At2g28490</b>	expressed protein (cupin domain)	9,0	4	53594,59	5,45	1-19	cupin domain profile (PS50849), cupin (PF00190)
G	39	<b>At1g20190</b>	alpha-expansin (AtEXPA11)	18,6	5	24817,74	9,84	1-20	expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
G	42-47	<b>At3g02110</b>	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	20,3	8	51075,46	9,35	1-22	serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
G	42, 43, 45	<b>At5g66590</b>	homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	38,9	6	18040,82	9,65	1-19	SCP-like extracellular protein (PF00188) (sperm coating glycoprotein)
G	44, 45	<b>At2g16001</b>	expressed protein (lipid recognition domain)	30,6	4	14715,26	9,61	1-23	E1 protein and Def2/Der2 allergen (ML domain - MD-2-related lipid recognition domain) (PF02221)
G	46-48	<b>At1g17860</b>	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	47,4	10	19718,68	9,76	1-21	soybean trypsin inhibitor (Kunitz) protease inhibitors family signature (PS00283), trypsin and protease inhibitor (PF00197)
G	47, 48	<b>At4g30610</b>	homologous to serine carboxypeptidase (BRS1 - Brassinosteroid-Insensitive BRI suppressor 1) (Peptidase family S10, S10.015 MEROPS)	14,2	6	50571,16	9,15	1-20	serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidases, histidine active site (PS00560), serine peptidase (PF00450)

Table S1, supplementary data

<b>G</b>	51, 52	<b>At1g09310</b>	expressed protein (DUF538)	27,9	5	19947,00	5,12		domain of unknown function DUF538 (PF04398)
<b>G</b>	52, 53, 54	<b>At5g38410*</b>	RUBISCO small subunit B3	27,6	5	20284,25	9,40		ribulose biphosphate carboxylase, small chain (PF00101)
<b>G</b>	52, 53, 57	<b>At5g38430*</b>	RUBISCO small subunit B1	22,7	5	20286,18	9,24		ribulose biphosphate carboxylase, small chain (PF00101)
<b>G</b>	52, 53	<b>At5g38420*</b>	RUBISCO small subunit B2	22,7	5	20350,22	9,24		ribulose biphosphate carboxylase, small chain (PF00101)
<b>G</b>	53-56	<b>At1g67090</b>	RUBISCO small subunit A1	22,8	5	20216,04	9,26		ribulose biphosphate carboxylase, small chain (PF00101)
<b>G</b>	57, 58	<b>At4g16500</b>	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	43,6	6	10183,74	9,78	1-22	cystatin domain (PF00031)
<b>H</b>	1, 3, 4, 10, 12-14, 16-32, 35-42, 48-54, 56-59	<b>At5g11420</b>	expressed protein (DUF642)	32,0	16	37344,57	7,70	1-22	domain of unknown function DUF 642 (PF04862)
<b>H</b>	2	<b>At3g24480</b>	LRR-extensin (AtLRX4)	10,5	6	51682,00	6,39	1-25	plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560), Pro-rich profile (PS50099)
<b>H</b>	3	<b>At4g13340</b>	LRR-extensin (AtLRX3)	5,9	4	79847,66	6,29	1-20	proline-rich region profile (PS50099), leucine-rich repeat (PF00560), extensin-like region (PF04554), leucine-rich repeat, plant specific (PS50502)
<b>H</b>	3	<b>At4g18670</b>	LRR-extensin (AtLRX5)	6,3	4	87040,18	9,02	1-27	histidine-rich region profile (PS50316), plant specific LRR profile (PS50502), serine-rich region profile (PS50324), Leucine Rich Repeat (PF00560)
<b>H</b>	3	<b>At5g53370</b>	carbohydrate esterase family 8 (pectin methylesterase)	8,7	4	64240,96	8,80		GPI anchor plant invertase/pectin methylesterase inhibitor (PF04043), pectinesterase (PF01095)
<b>H</b>	4	<b>At1g28290</b>	proline-rich protein	6,4	3	35913,36	10,77	1-24	histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
<b>H</b>	5-9	<b>At1g68560</b>	glycoside hydrolase family 31 (alpha-xylosidase) (XYL1)	19,3	19	99641,37	6,26	1-27	glycoside hydrolases family 31 active site (PS00129)
<b>H</b>	10	<b>At3g13790</b>	glycoside hydrolase family 32 (beta-fructofuranosidase)	8,9	5	63490,72	9,68	1-25	glycoside hydrolases family 32 active site (PS00609), glycosyl hydrolases family 32 (PF00251)
<b>H</b>	11	<b>At3g07130</b>	homologous to purple acid phosphatase	12,2	6	58523,64	5,71	1-17	metallo-phosphoesterase motif (PS50185), calcineurin-like phosphoesterase (PF00149)
<b>H</b>	12	<b>At2g27190</b>	homologous to purple acid phosphatase (PAP12)	27,9	14	50962,71	5,75	1-28	metallo-phosphoesterase motif (PS50185), calcineurin-like phosphoesterase (PF00149)
<b>H</b>	12	<b>At3g61490</b>	glycoside hydrolase family 28 (polygalacturonase)	12,2	4	49497,06	5,76	1-23	glycoside hydrolase family 28 (PF00295)
<b>H</b>	13	<b>At1g78850</b>	homologous to lectin (curculin-like)	7,7	4	46664,12	8,90	1-22	PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
<b>H</b>	13	<b>At1g33590</b>	expressed protein (LRR domains)	9,0	5	48103,60	9,90	1-24	LRR (PF00560), leucine zipper pattern (PS00029)
<b>H</b>	14-17	<b>At1g03230</b>	homologous to carrot EDGP and tomato XEGIP	33,2	12	43723,02	10,27	1-23	
<b>H</b>	17	<b>At5g19110</b>	homologous to carrot EDGP and tomato XEGIP	23,5	8	41369,63	10,01	1-19	
<b>H</b>	18	<b>At4g05200</b>	homologous to receptor kinase (RLK, DUF26-1b subfamily)	9,2	6	72576,6	6,09	1-25	trans-membrane domains protein kinases ATP-binding region signature (PS00107), Serine/Threonine protein kinases active-site signature (PS00108), protein kinase domain profile (PF50011), domain of unknown function DUF 26 (PF01657), Protein kinase domain (PF00069), protein tyrosine kinase (PF07714)
<b>H</b>	20-23, 32	<b>At3g20820</b>	expressed protein (LRR domains)	37,5	11	37850,70	9,90	1-19	typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22-like LRR profile (PS50504)
<b>H</b>	20	<b>At1g71695</b>	peroxidase (AtPrx12)	26,5	7	37104,36	9,46	1-22	peroxidase (PF00141)
<b>H</b>	21, 22	<b>At5g12940</b>	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	18,6	6	36784,96	10,45	1-29	plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
<b>H</b>	22, 24, 26, 27, 32, 41, 50, 52- 53	<b>At5g25460</b>	expressed protein (DUF642)	18,7	8	37948,52	9,03	1-19	domain of unknown function DUF 642 (PF04862)
<b>H</b>	23-29, 32	<b>At2g06850</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH4)	42,2	14	31716,77	9,65	1-24	glycoside hydrolases family 16 (PF00722)
<b>H</b>	24, 25	<b>At5g64100</b>	peroxidase (AtPrx69)	19,6	8	33228,93	10,72	1-23	peroxidases active site signature (PS00436)
<b>H</b>	25, 26	<b>T23B7.10</b>	homologous to PGIP1 (LRR protein FLR1)	24,0	6	33134,88	9,62	1-23	typical LRR profile (PS50506), plant specific LRR profile (PS50502), plant disease resistance response protein (PF03018)
<b>H</b>	28	<b>At5g13870</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH5)	14,3	6	31752,61	9,53	1-21	xyloglucan endo-transglycosylase (XET) C-terminus (PF06955), glycoside hydrolases family 16 (PF00722)
<b>H</b>	29	<b>At5g64260</b>	homologous to <i>Nicotiana tabacum</i> phi-I	20,7	6	30550,86	9,98	1-19	phosphate-induced protein 1 conserved region (PF04674)
<b>H</b>	29-34	<b>At5g09440</b>	homologous to <i>Nicotiana tabacum</i> phi-I	20,5	6	27366,44	10,15	1-23	phosphate-induced protein 1 conserved region (PF04674)
<b>H</b>	30	<b>At4g38400</b>	expansin-like A (AtEXLA2)	13,2	4	26516,93	9,60	1-20	expansin, cellulose-binding-like domain profile (PS50843), expansin, family-45 endoglucanase-like domain profile (PS50842), rare lipoprotein A (RlpA)-like double-psi beta-barrel (PF03330), pollen allergen (PF01357)



Table S1, supplementary data

H	31-35	<b>At3g22640</b>	expressed protein (cupin domain)	16,0	10	52687,99	6,45	1-22	cupin domain profile (PS50849), cupin (PF00190), cupin domain
H	31	<b>At3g45970</b>	expansin-like A (AtEXLA1)	14,7	4	26258,12	9,46	1-20	pollen allergen (PF01357), barwin (PF00967)
H	33-35	<b>At2g28790</b>	homologous to <i>Lycopersicon esculentum</i> osmotin	16,5	4	24551,60	9,53	1-24	thaumatin family (PF00314)
H	34	<b>At4g29270</b>	homologous to acid phosphatase	17,6	4	25900,84	9,15	1-26	HAD superfamily, subfamily IIIB (acid phosphatase) (PF03767)
H	34-42	<b>At4g12860</b>	early nodulin (AtEN20) (phytoecyanin)	30,5	6	14273,34	9,64	1-18	plastocyanin-like domain (PF02298) (copper binding proteins)
H	35	<b>At2g28490</b>	expressed protein (cupin domain)	9,0	4	53594,59	5,45	1-19	cupin domain profile (PS50849), cupin (PF00190)
H	36-38	<b>At1g20190</b>	alpha-expansin (AtEXPA11)	18,6	6	24817,74	9,84	1-20	expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
H	38-45	<b>At5g66590</b>	homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	38,9	6	18040,82	9,65	1-19	SCP-like extracellular protein (PF00188) (sperm coating glycoprotein)
H	39-40, 47	<b>At3g02110</b>	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	16,1	6	51075,46	9,35	1-22	serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
H	42, 43, 46, 48	<b>At2g34700</b>	expressed protein (Ole e1 allergen domain)	37,1	6	16248,80	10,22	1-23	pollen proteins Ole e 1 family (PF01190)
H	43	<b>At4g30610</b>	homologous to serine carboxypeptidase (BRS1 - Brassinosteroid-Insensitive BRI suppressor 1) (Peptidase family S10, S10.015 MEROPS)	11,8	5	50571,16	9,15	1-20	serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidases, histidine active site (PS00560), serine peptidase (PF00450)
H	51-54, 56	<b>At1g67090</b>	RUBISCO small subunit A1	19,4	4	20216,04	9,26		ribulose biphosphate carboxylase, small chain (PF00101)
H	51-54, 56-58	<b>At5g38410*</b>	RUBISCO small subunit B3	24,3	4	20284,25	9,40		ribulose biphosphate carboxylase, small chain (PF00101)
H	52-54	<b>At5g38430*</b>	RUBISCO small subunit B1	19,3	4	20286,18	9,24		ribulose biphosphate carboxylase, small chain (PF00101)
H	54	<b>At5g38420*</b>	RUBISCO small subunit B2	19,3	4	20350,22	9,24		ribulose biphosphate carboxylase, small chain (PF00101)
H	52, 54, 55	<b>At4g32460</b>	expressed protein (DUF642)	12,6	5	37375,51	9,25	1-21	domain of unknown function DUF 642 (PF04862)
H	57, 58	<b>At4g16500</b>	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	35,0	6	10183,74	9,78	1-22	cystatin domain (PF00031)
H	58	<b>At2g38530</b>	non-specific lipid transfer protein type 1 (LTP2)	31,4	4	9475,92	11,90	1-25	protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00697)
I	1-5, 10, 11, 14, 17-32, 34, 38-58	<b>At5g25460</b>	expressed protein (DUF642)	37,4	17	37948,52	9,03	1-19	domain of unknown function DUF 642 (PF04862)
I	2-5, 8-12	<b>At1g28290</b>	proline-rich protein	9,2	4	35913,36	10,77	1-24	histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
I	6-8	<b>At1g68560</b>	glycoside hydrolase family 31 (alpha-xylosidase) (XYL1)	16,2	13	99641,37	6,26	1-27	glycoside hydrolases family 31 active site (PS00129)
I	12, 14, 16-20, 22-24, 26-32, 34-40, 42, 43, 45, 46, 49, 54, 56-58	<b>At5g11420</b>	expressed protein (DUF642)	35,8	16	37344,57	7,70	1-22	domain of unknown function DUF 642 (PF04862)
I	13, 40	<b>At1g78850</b>	homologous to lectin (curculin-like)	10,7	4	46664,12	8,90	1-22	PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
I	13	<b>At1g78820</b>	homologous to lectin (curculin-like)	11,4	5	48184,38	9,54	1-22	D-mannose binding lectin (PF01453), PAN domain (PF00024)
I	13	<b>At3g52500</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	9,2	4	48996,36	9,49	1-18	eukaryotic and viral aspartyl proteases active site (PS00141), eukaryotic aspartyl protease (PF00026)
I	14, 15	<b>At1g03230</b>	homologous to carrot EDGP and tomato XEGIP	19,6	7	43723,02	10,27	1-23	pectinesterase (PF01095)
I	15	<b>At4g33220</b>	carbohydrate esterase family 8 (pectin methyltransferase)	11,6	11	56832,32	6,24	1-19	eukaryotic aspartyl protease (PF00026)
I	16, 17	<b>At5g19110</b>	homologous to carrot EDGP and tomato XEGIP	23,5	8	41369,63	10,01	1-19	
I	16-18	<b>At5g07030</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	22,3	9	44860,54	10,82	1-19	
I	16, 17	<b>At3g43270</b>	carbohydrate esterase family 8 (pectin methyltransferase)	15,2	8	54929,20	9,55	1-24	pectinesterase (PF01095), plant invertase/pectin methyltransferase inhibitor (PMEI) (PF04043)
I	20, 21, 23	<b>At1g10550</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH33)	15,5	5	32314,04	9,73	1-21	glycoside hydrolases family 16 (PF00722), xyloglucan endo-transglycosylase (XET) C-terminus (PF06955)
I	21-26	<b>At5g12940</b>	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	37,7	14	36784,96	10,45	1-29	plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
I	21-22	<b>At3g20820</b>	expressed protein (LRR domains)	22,5	6	37850,70	9,90	1-19	typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22+like LRR profile (PS50504)
I	23	<b>At1g80240</b>	expressed protein (DUF642)	23,0	7	37817,29	9,57	1-22	domain of unknown function DUF 642 (PF04862)
I	25, 26	<b>At1g53830</b>	carbohydrate esterase family 8 (pectin methyltransferase) (AtPME2)	17,4	9	59942,45	9,67	1-40	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methyltransferase inhibitor (PF04043)
I	26	<b>At1g11580</b>	carbohydrate esterase family 8 (pectin methyltransferase)	16,3	9	58066,93	9,43	1-34	pectinesterase (PF01095), plant invertase/pectin methyltransferase inhibitor (PF04043)
I	26	<b>At4g30610</b>	homologous to serine carboxypeptidase (BRS1 - Brassinosteroid-Insensitive BRI suppressor 1) (Peptidase family S10, S10.015 MEROPS)	9,0	5	50571,16	9,15	1-20	serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidases, histidine active site (PS00560), serine peptidase (PF00450)
I	27	<b>At3g44990</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH31)	16,0	6	31710,24	9,69	1-18	glycoside hydrolase family 16 (PF00722), xyloglucan endo-transglycosylase, C-terminal (PF06955)

Table S1, supplementary data

I	28	<b>At3g14310</b>	carbohydrate esterase family 8 (pectin methylesterase) (AIPME3)	17,9	8	60050,61	9,59	1-40	transmembrane domains	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
I	28-30	<b>At2g06850</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AUXTH4)	39,5	12	31716,77	9,65	1-24		glycoside hydrolases family 16 (PF00722)
I	29-31	<b>At5g64260</b>	homologous to <i>Nicotiana tabacum</i> phi-I	35,7	11	30550,86	9,98	1-19		phosphate-induced protein 1 conserved region (PF04674)
I	29-33	<b>At5g09440</b>	homologous to <i>Nicotiana tabacum</i> phi-I	17,3	5	27366,44	10,15	1-23		phosphate-induced protein 1 conserved region (PF04674)
I	29	<b>At3g45960</b>	expansin-like A (AtEXLA3)	15,2	4	26269,86	9,99	1-20		expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
I	29	<b>At3g45970</b>	expansin-like A (AtEXLA1)	15,1	4	26258,12	9,46	1-20		pollen allergen (PF01357), barwin (PF00967)
I	31	<b>At1g42970</b>	glyceroldehyde 3-phosphate dehydrogenase	14,3	5	47659,41	6,27			glyceraldehyde 3-phosphate dehydrogenase, NAD binding domain (PF00044), glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain (PF02800)
I	32-35	<b>At3g22640</b>	expressed protein (cupin domain)	21,4	11	52687,99	6,45	1-22		cupin domain profile (PS50849), cupin (PF00190), cupin domain
I	33-35	<b>At2g28790</b>	homologous to <i>Lycopersicon esculentum</i> osmotin	12,9	4	24551,60	9,53	1-24		thaumatin family (PF00314)
I	34, 35, 37-40	<b>At4g12880</b>	early nodulin (AtEN20) (phytoecyanin)	24,8	5	14273,34	9,64	1-18		plastocyanin-like domain (PF02298) (copper binding proteins)
I	35-37	<b>At5g02260</b>	alpha-expansin (AtEXPA9)	32,2	7	25579,80	10,19	1-21		expansin, family-45 endoglucanase-like domain profile (PS50842), pollen allergen (PF01357)
I	36, 37	<b>At1g20190</b>	alpha-expansin (AtEXPA11)	18,6	6	24817,74	9,84	1-20		expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
I	36	<b>At2g28490</b>	expressed protein (cupin domain)	12,7	5	53594,59	5,45	1-19		cupin domain profile (PS50849), cupin (PF00190)
I	38-41, 43-46	<b>At5g66590</b>	homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	33,0	5	18040,82	9,65	1-19		SCP-like extracellular protein (PF00188) (sperm coating glycoprotein)
I	39, 42, 47	<b>At5g23210</b>	homologous to serine carboxypeptidase (SCPL34) (Peptidase family S10, S10.005, MEROPS)	11,8	5	53718,78	9,44	1-25		serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
I	39	<b>At3g02110</b>	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	12,3	5	51075,46	9,35	1-22		serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
I	46	<b>At2g34700</b>	expressed protein (Ole e1 allergen domain)	18,3	6	16248,80	10,22	1-23		pollen proteins Ole e 1 family (PF01190)
I	52, 53, 55, 56	<b>At1g67090</b>	RUBISCO small subunit A1	19,0	4	20216,04	9,26			ribulose biphosphate carboxylase, small chain (PF00101)
I	52-55, 57	<b>At5g38410*</b>	RUBISCO small subunit B3	24,3	5	20284,25	9,40			ribulose biphosphate carboxylase, small chain (PF00101)
I	52-55, 57	<b>At5g38430*</b>	RUBISCO small subunit B1	19,3	4	20286,18	9,24			ribulose biphosphate carboxylase, small chain (PF00101)
I	52-55, 57	<b>At5g38420*</b>	RUBISCO small subunit B2	19,3	4	20350,22	9,24			ribulose biphosphate carboxylase, small chain (PF00101)
I	53, 57	<b>At4g32460</b>	expressed protein (DUF642)	11,2	5	37375,51	9,25	1-21		domain of unknown function DUF 642 (PF04862)
I	56	<b>At4g27160</b>	2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	22,0	4	16909,07	7,69	1-17		protease inhibitor/seed storage/LTP family (PF00234)
J	1, 13-27, 30, 32-36, 38-42	<b>At5g25460</b>	expressed protein (DUF642)	35,5	15	37948,52	9,03	1-19		domain of unknown function DUF 642 (PF04862)
J	2-10	<b>At1g28290</b>	proline-rich protein	14,5	5	35913,36	10,77	1-24		histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
J	4, 5	<b>At2g30210</b>	homologous to laccase (AtLAC3)	15,4	8	61172,59	10,11	1-25		multicopper oxidases signature 2 (PS00080), multicopper oxidase (PF00394)
J	9	<b>At3g52500</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	9,2	4	48996,36	9,49	1-18		eukaryotic and viral aspartyl proteases active site (PS00141), eukaryotic aspartyl protease (PF00026)
J	10	<b>At1g03230</b>	homologous to carrot EDGP and tomato XEGIP	13,4	5	43723,02	10,27	1-23		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PMEI) (PF04043)
J	11, 12	<b>At3g43270</b>	carbohydrate esterase family 8 (pectin methylesterase)	20,1	13	54929,20	9,55	1-24		eukaryotic aspartyl protease (PF00026)
J	12, 13	<b>At5g07030</b>	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	23,5	10	44860,54	10,82	1-35		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PMEI) (PF04043)
J	12	<b>At4g33220</b>	carbohydrate esterase family 8 (pectin methylesterase)	14,1	7	56832,32	6,24	1-19		pectinesterase (PF01095)
J	12-14, 18, 19, 21, 25, 39-42	<b>At5g11420</b>	expressed protein (DUF642)	27,6	11	37344,57	7,70	1-22		domain of unknown function DUF 642 (PF04862)
J	16-18	<b>At5g12940</b>	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	31,5	11	36784,96	10,45	1-29		plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
J	19	<b>At1g53830</b>	carbohydrate esterase family 8 (pectin methylesterase) (AIPME2)	12,1	8	59942,45	9,67	1-40	GPI anchor	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
J	20-22	<b>At1g11580</b>	carbohydrate esterase family 8 (pectin methylesterase)	19,7	10	58066,93	9,43	1-34		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
J	21	<b>At3g44990</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AUXTH31)	14,3	4	31710,24	9,69	1-18		glycoside hydrolase family 16 (PF00722), xyloglucan endo-transglycosylase, C-terminal (PF06955)



Table S1, supplementary data

J	21-23	<b>At3g14310</b> carbohydrate esterase family 8 (pectin methylesterase) (AIPME3)	24,8	13	60050,61	9,59	1-40	trans-membrane domains	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095, plant invertase/pectin methylesterase inhibitor (PF04043)
J	23, 25	<b>At5g64260</b> homologous to <i>Nicotiana tabacum</i> phi-I	29,8	9	30550,86	9,98	1-19		phosphate-induced protein 1 conserved region (PF04674)
J	25-27	<b>At5g09440</b> homologous to <i>Nicotiana tabacum</i> phi-I	20,5	6	27366,44	10,15	1-23		expansin, cellulose-binding-like domain profile (PS50843), expansin, family-45 endoglucanase-like domain profile (PS50842), rare lipoprotein A (RlpA)-like double-psi beta-barrel (PF03330), pollen allergen (PF01357)
J	25	<b>At4g38400</b> expansin-like A (AtEXLA2)	14,3	4	26516,93	9,60	1-20		cupin domain profile (PS50849), cupin (PF00190), cupin domain thaumatin family (PF00314)
J	26-28	<b>At3g22640</b> expressed protein (cupin domain)	14,0	5	52687,99	6,45	1-22		expansin, family-45 endoglucanase-like domain profile (PS50842), pollen allergen (PF01357)
J	27, 28	<b>At2g28790</b> homologous to <i>Lycopersicon esculentum</i> osmotin	12,9	3	24551,60	9,53	1-24		plastocyanin-like domain (PF02298) (copper binding proteins)
J	29-31	<b>At5g02260</b> alpha-expansin (AtEXPA9)	24,4	5	25579,80	10,19	1-21		expansin, family-45 endoglucanase-like domain profile (PS50842), pollen allergen (PF01357)
J	29-32	<b>At4g12880</b> early nodulin (AtEN20) (phytoecyanin)	24,8	5	14273,34	9,64	1-18		expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
J	30, 31	<b>At1g20190</b> alpha-expansin (AtEXPA11)	19,8	5	24817,74	9,84	1-20		SCP-like extracellular protein (PF00188) (sperm coating glycoprotein)
J	33, 36, 37	<b>At5g66590</b> homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	28,6	4	18040,82	9,65	1-19		glycine-rich region profile (PS50315)
J	35, 36	<b>At2g05580</b> glycine-rich protein	40,1	6	21517,58	10,21	1-20		plant invertase/pectin methylesterase inhibitor (PF04043)
J	37	homologous to plant invertase/pectin methylesterase inhibitor (PMEI)	19,3	3	16933,85	10,42	1-20		domain of unknown function DUF 642 (PF04862)
J	39, 41, 42	<b>At4g32460</b> expressed protein (DUF642)	17,5	7	37375,51	9,25	1-21		ribulose biphosphate carboxylase, small chain (PF00101)
J	40	<b>At5g38430*</b> RUBISCO small subunit B1	27,6	5	20286,18	9,24			ribulose biphosphate carboxylase, small chain (PF00101)
J	40	<b>At5g38420*</b> RUBISCO small subunit B2	27,6	5	20350,22	9,24			ribulose biphosphate carboxylase, small chain (PF00101)
J	42	<b>At1g67090</b> RUBISCO small subunit A1	19,4	4	20216,04	9,26			protease inhibitor/seed storage/LTP family (PF00234)
J	42, 43	<b>At4g27160</b> 2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	17,7	3	16909,07	7,69	1-17		
K	1	<b>At2g30210</b> homologous to lacase (AILAC3)	12,6	7	61172,59	10,11	1-25		multicopper oxidases signature 2 (PS00080), multicopper oxidase (PF00394)
K	1, 5	<b>At1g28290</b> proline-rich protein	8.9 (32.3)	4 + 3 (Hyp)	35913,36	10,77	1-24		histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
K	6, 7	<b>At3g43270</b> carbohydrate esterase family 8 (pectin methylesterase)	12,5	9	54929,20	9,55	1-24		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PMEI) (PF04043)
K	7	<b>At5g11420</b> expressed protein (DUF642)	11,2	4	37344,57	7,70	1-22		domain of unknown function DUF 642 (PF04862)
K	8-10	<b>At5g14920</b> proline-rich protein	Edman sequencing + MALDI-TOF-TOF MS <sup>c</sup>	3	26753,23	10,98	1-21		gibberellin regulated protein (PF02704)
K	11-13	<b>At1g53830</b> carbohydrate esterase family 8 (pectin methylesterase) (AIPME2)	15,5	11	59942,45	9,67	1-40	GPI anchor	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
K	11-14	<b>At3g14310</b> carbohydrate esterase family 8 (pectin methylesterase) (AIPME3)	16,7	8	60050,61	9,59	1-40	transmembrane domains	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095, plant invertase/pectin methylesterase inhibitor (PF04043)
K	12, 13	<b>At1g11580</b> carbohydrate esterase family 8 (pectin methylesterase)	20,5	10	58066,93	9,43	1-34		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
K	15, 16	<b>At1g20190</b> alpha-expansin (AtEXPA11)	19,8	5	24817,74	9,84	1-20		expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
K	16	<b>At4g12880</b> early nodulin (AtEN20) (phytoecyanin)	24,1	5	14273,34	9,64	1-18		plastocyanin-like domain (PF02298) (copper binding proteins)
K	17, 18	<b>At5g15230</b> gibberellin-regulated protein (GASA4)	31,1	5	9207,55	10,75	1-25		gibberellin regulated protein (PF02704)
K	18, 19	<b>At4g27160</b> 2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	17,1	4	16909,07	7,69	1-17		protease inhibitor/seed storage/LTP family (PF00234)
K	18, 19	<b>At4g27170</b> 2S albumin isoform 4 (protease inhibitor/seed storage/LTP family)	15,7	3	16946,26	7,80	1-21		protease inhibitor/seed storage/LTP family (PF00234)
L	1-3	<b>At5g14920</b> proline-rich protein	Edman sequencing + MALDI-TOF-TOF MS <sup>c</sup>	3	26753,23	10,98	1-21		gibberellin regulated protein (PF02704)
L	3	<b>At4g24340</b> homologous to phosphorylase	29,9	6	33774,65	9,63	1-27		phosphorylase family (PF01048)
L	4, 6	<b>At4g27170</b> 2S albumin isoform 4 (protease inhibitor/seed storage/LTP family)	24,1	4	16946,26	7,80	1-21		protease inhibitor/seed storage/LTP family (PF00234)
L	5	<b>At4g27160</b> 2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	16,5	3	16909,07	7,69	1-17		protease inhibitor/seed storage/LTP family (PF00234)
L	6	<b>At4g27150</b> 2S albumin isoform 2 (protease inhibitor/seed storage/LTP family)	24,7	4	17473,78	6,34	1-17		protease inhibitor/seed storage/LTP family (PF00234)

a: The following criteria were retained for protein identification: at least 4 peptides are required for proteins larger than 15 kDa, only 3 peptides can be used for proteins smaller than 15 kDa. Allowed modifications are 1 miscleavage and Met oxidation. In each series of samples ( i.e., total and A to L), the best score of identification is shown.

b: Functional domains were predicted as described in Material and methods. PF stands for PFAM, PS for PROSITE and IPR for InterPro.

c: N-terminal sequencing, XXTLPS; LC-MS/MS sequencing, CGQHSR.

**Table S2, supplementary material.** Identification of proteins extracted from purified cell walls of 11-day-old *Arabidopsis* hypocotyls by CaCl<sub>2</sub> and LiCl solutions

\* proteins that could not be distinguished

Colour code secretory proteins (Presence of predicted signal peptide was checked using PSORT: <http://psort.ims.u-tokyo.ac.jp/form.html>, and TargetP: <http://www.cbs.dtu.dk/services/TargetP/>)transmembrane domains (Presence of predicted transmembrane domains was checked using Aramemnon: <http://aramemnon.botanik.uni-koeln.de/>)GPI anchors (presence or GPI anchors was checked using Aramemnon: <http://aramemnon.botanik.uni-koeln.de/>)

intracellular proteins

proteins identified after 1D-separation (1-DE), but not identified after 2D-separation (cationic exchange chromatography followed by 1-DE)

	total protein extract	band number	accession AGI	annotation	MALDI-TOF identification % sequence coverage	number of peptides (m/z) <sup>a</sup>	mature protein MM (kDa)	pI	signal peptide	transmembrane domains GPI anchors	functional domains <sup>b</sup>
1		1-16, 18-24, 27, 29, 30, 32-34, 41, 42, 44, 45, 47, 53-56	<b>At5g25460</b>	expressed protein (DUF642)	48,5	17	37948,52	9,03	1-19		domain of unknown function DUF 642 (PF04862)
2		1-23, 26, 28, 37-40, 43, 44, 52-56	<b>At5g11420</b>	expressed protein (DUF642)	34,7	14	37344,57	7,70	1-22		domain of unknown function DUF 642 (PF04862)
3		1, 2	<b>At1g28290</b>	proline-rich protein	47	4	35913,36	10,77	1-24		histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
4		2, 3, 5, 7, 10, 13-15, 24	<b>At1g03230</b>	homologous to carrot EDGP and tomato XEGIP	16,8	6	43723,02	10,27	1-23		
5		3, 7, 9, 10, 12, 14, 15	<b>At1g03220</b>	homologous to carrot EDGP and tomato XEGIP	13,2	4	43394,37	9,99	1-22		
6		5	<b>At2g06850</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH4)	13,2	5	31716,77	9,65	1-24		glycoside hydrolases family 16 (PF00722)
7		5	<b>At5g07030</b>	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	8,6	4	44860,54	10,82	1-19		eukaryotic aspartyl protease (PF00026)
8		9, 10	<b>At3g16850</b>	glycoside hydrolase family 28 (polygalacturonase)	31,9	12	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
9		11	<b>At5g34940</b>	glycoside hydrolase family 79 (endo beta-glucuronidase/heparanase)	9,9	4	57219,88	9,13	1-21		glycoside hydrolase family 79, N-terminal domain (PF03662)
10		11	<b>At1g33590</b>	expressed protein (LRR domains)	7,8	4	48103,60	9,90	1-24		LRR (PF00560), leucine zipper pattern (PS00029)
11		13, 14, 41-43	<b>At1g78850</b>	homologous to lectin (curculin-like)	18,1	7	46664,12	8,90	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
12		13, 41	<b>At1g78860</b>	homologous to lectin (curculin-like)	14,9	6	46804,29	6,00	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
13		13-15	<b>At1g09750</b>	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	12,7	7	45038,78	9,09	1-23		lipase Ser active site (PS00120), Ser-rich region (PS050324), eukaryotic aspartic protease (PF00026)
14		16	<b>At3g43270</b>	carbohydrate esterase family 8 (pectin methylesterase)	6,1	4	54929,20	9,55	1-24		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor PME1 (PF04043)
15		17-19, 21	<b>At3g20820</b>	expressed protein (LRR domains)	27,9	9	37850,70	9,90	1-19		typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22-like LRR profile (PS50504)
16		17, 19, 32	<b>At3g49120</b>	peroxidase (AtPrx34)	8,5	5	35695,45	9,41	1-30		peroxidase active site signature (PS00436), peroxidases proximal heme-ligand signature (PS00435)
17		19-24	<b>At5g12940</b>	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	21,6	7	36784,96	10,45	1-29		plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
18		19	<b>At3g32980</b>	peroxidase (AtPrx32)	15,6	7	35712,41	6,00	1-29		peroxidases active site signature (PS00436)
19		20, 22, 23, 47	<b>At1g70850*</b>	homologous to Bet v I allergen family	35,4	9	35569,93	4,96			pathogenesis-related protein Bet v I family (PF00407)
20		47	<b>At1g70830*</b>	homologous to Bet v I allergen family	17,6	4	37614,51	5,03			pathogenesis-related protein Bet v I family (PF00407)
21		21-23, 30, 31	<b>At3g08030</b>	expressed protein (DUF642)	23,6	7	36993,02	6,87	1-21		domain of unknown function DUF 642 (PF04862)
22		22	<b>At4g37950</b>	polysaccharide lyase family 4 (rhannogalacturonate lyase)	8,8	5	74062,32	9,65	1-30		rhannogalacturonate lyase family (PF06045)
23		23, 24	<b>At1g78830</b>	homologous to lectin (curculin-like)	13,0	8	48101,66	9,45	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
24		23	<b>At1g80240</b>	expressed protein (DUF642)	18,6	5	37817,29	9,57	1-22		domain of unknown function DUF 642 (PF04862)
25		24	<b>T23B7.10</b>	homologous to PGIP1 (LRR protein FLR1)	18,8	4	33134,88	9,62	1-23		typical LRR profile (PS50506), plant specific LRR profile (PS50502), plant disease resistance response protein (PF03018)
26		24	<b>At5g64100</b>	peroxidase (AtPrx69)	18,7	6	33228,93	10,72	1-23		peroxidases active site signature (PS00436)
27		25	<b>ArthCp030</b>	RUBISCO large subunit (RBCL)	21,5	11	52954,98	5,79			ribulose biphosphate carboxylase large chain active site (PS00157), ribulose biphosphate carboxylase large chain, catalytic domain (PF00016), ribulose biphosphate carboxylase large chain, N-terminal domain (PF02788)
28		25, 53-58	<b>At1g67090</b>	RUBISCO small subunit A1	28,9	5	20216,04	9,26			ribulose biphosphate carboxylase, small chain (PF00101)
29		28, 29	<b>At4g30170</b>	peroxidase (AtPrx45)	38,8	11	33052,57	10,43	1-25		haem peroxidase, plant/fungal/bacterial (PS00435, PS00436, PS50873)

Table S2, supplementary data

29-31	At1g47128	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	11	4	48785,27	5,01	1-21	eukaryotic thiol (cysteine) proteases histidine active site (PS00639), cysteine active site (PS00640), asparagine active site (PS00139), Cys-rich region (PS50311), papaine family cysteine protease (PF00112)	
30	30, 31	At5g64260	homologous to <i>Nicotiana tabacum</i> phi-I	27,2	8	30550,86	9,98	1-19	phosphate-induced protein 1 conserved region (PF04674)
31	30, 31	At3g45970	expansin-like A (AtEXLA1)	27,5	7	26258,12	9,46	1-20	pollen allergen (PF01357), barwin (PF00967)
32	30	At4g01610	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	14,2	4	36995,63	5,79	1-22	eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
33	32, 33	At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	13,4	5	36995,63	5,79	1-22	eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
34	32-35	At5g09440	homologous to <i>Nicotiana tabacum</i> phi-I	17,3	5	27366,44	10,15	1-23	phosphate-induced protein 1 conserved region (PF04674)
35	32-41	At4g12880	early nodulin (AtEN20) (phytoecyanin)	30,5	6	14273,34	9,64	1-18	plastocyanin-like domain (PF02298) (copper binding proteins)
36	32, 33	At4g38400	expansin-like A (AtEXLA2)	10,9	3	26516,93	9,60	1-20	expansin, cellulose-binding-like domain profile (PS50843), expansin, family-45 endoglucanase-like domain profile (PS50842), rare lipoprotein A (RlpA)-like double-psi beta-barrel (PF03330), pollen allergen (PF01357)
37	34-36	At2g28790	homologous to <i>Lycopersicon esculentum</i> osmotin	20,5	5	24551,60	9,53	1-24	thaumatin family (PF00314)
38	35-37	At2g28490	expressed protein (cupin domain)	13,7	6	53594,59	5,45	1-19	cupin domain profile (PS50849), cupin (PF00190)
39	35	At3g22640	expressed protein (cupin domain)	9,3	5	52687,99	6,45	1-22	cupin domain profile (PS50849), cupin (PF00190), cupin domain
40	36	At3g56750	expressed protein	9,4	4	42120,52	9,89	1-39	
41	37	At4g29270	homologous to acid phosphatase	12,5	4	25900,84	9,15	1-26	HAD superfamily, subfamily IIIB (acid phosphatase) (PF03767)
42	38, 39	At1g20190	alpha-expansin (AtEXPA11)	25	8	24817,74	9,84	1-20	expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
43	38, 39	At1g09560	germin (subfamily 2, member 1, GLP5)	14,6	3	20472,37	9,71	1-23	germin (PS00725), cupin (PF00190)
44	40	At3g21770	peroxidase (AtPrx30)	15,2	4	32901,55	10,54	1-27	peroxidases active site signature (PS00436)
45	41-43	At3g02110	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	14,6	6	51075,46	9,35	1-22	serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
46	41, 49, 50	At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	23,2	4	20982,78	5,20	1-26	trypsin and protease inhibitor Kunitz legume (PF00197)
47	42	At3g11630	homologous to 2-Cys peroxiredoxin	29,3	7	29091,87	7,91		AhpC/TSA family (PF00578)
48	43-45	At5g66590	homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	33,0	5	18040,82	9,65	1-19	SCP-like extracellular protein (PF00188) (sperm coating glycoprotein)
49	44, 45	At1g17860	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	39,3	9	19718,68	9,76	1-21	soybean trypsin inhibitor (Kunitz) protease inhibitors family signature (PS00283), trypsin and protease inhibitor (PF00197)
50	46	At2g04690	expressed protein (homologous to a human brain CREG protein)	19,5	4	20214,11	8,00	1-28	
51	47	At5g23210	homologous to serine carboxypeptidase D (SCPL34) (Peptidase family S10, S10.005, MEROPS)	8,2	4	53718,78	9,44	1-25	serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
52	48, 49	At4g39260	homologous to RNA binding protein (AtGRP8)	41,4	5	16578,63	5,12		glycine-rich region profile (PS50315), eukaryotic RNA Recognition Motif (RRM) profile (PS50102), RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) (PF00076)
53	49, 50	At4g23680	homologous to Bet v I allergen family	47,0	8	17473,54	5,80		pathogenesis-related protein Bet v I family (PF00407)
54	49	At4g23670	homologous to Bet v I allergen family	43	8	17517,73	5,84		pathogenesis-related protein Bet v I family (PF00407)
55	50-54	At1g09310	expressed protein (DUF538)	40,1	5	19947,00	5,12		domain of unknown function DUF538 (PF04398)
56	52-58	At5g38410*	RUBISCO small subunit B3	27,6	5	20284,25	9,40		ribulose biphosphate carboxylase, small chain (PF00101)
57	52, 53, 55-57	At5g38430*	RUBISCO small subunit B1	22,6	5	20286,18	9,24		ribulose biphosphate carboxylase, small chain (PF00101)
58	52, 53, 55-57	At5g38420*	RUBISCO small subunit B2	22,6	5	20350,22	9,24		ribulose biphosphate carboxylase, small chain (PF00101)
59	53	At3g22000	expressed protein (DUF26)	14,6	4	27296,47	9,97	1-32	domain of unknown function DUF26 (PF01657)
60	53-56, 58, 59	At4g20860	homologous to berberine-bridge enzyme (S)-reticulon:oxygen oxidoreductase	10,4	7	58375,98	9,71	1-16	oxygen oxidoreductases covalent FAD-binding site (PS00862), FAD_binding_4
61	53-57	At4g32460	expressed protein (DUF642)	16,7	7	37375,51	9,25	1-21	FAD binding domain (PF01565), BBE Berberine and berberine like (PF08031)
62	57, 59, 60	At4g27150	2S albumin isoform 2 (protease inhibitor/seed storage/LTP family)	17,6	3	17473,78	6,34	1-17	domain of unknown function DUF 642 (PF04862)
63	59, 60	At4g27160	2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	18,3	3	16909,07	7,69	1-17	protease inhibitor/seed storage/LTP family (PF00234)
64	59	At1g47540	homologous to trypsin inhibitor	12,2	2	9051,17	4,97	1-18	scorpion toxin-like domain (PF00537)
65	59, 60	At2g38530	non-specific lipid transfer protein type 1 (LTP2)	17,8	2	9475,92	11,90	1-25	protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00597)
66	60	At4g27170	2S albumin isoform 4 (protease inhibitor/seed storage/LTP family)	23,7	3	16946,26	7,80	1-21	protease inhibitor/seed storage/LTP family (PF00234)
67	N	At3g16850	glycoside hydrolase family 28 (polygalacturonase)	25,9	8	46982,42	5,25	1-19	glycoside hydrolase family 28 (PF00295)

Table S2, supplementary data

N	2	At1g47710	homologous to serpin (serine protease inhibitor)	20,2	7	42639,41	4,97		transmembrane domains	serpins signature (PS00284), serpin (serine protease inhibitor) (PF00079)
N	3	At4g16260	glycoside hydrolase family 17	12,8	6	35905,19	6,64	1-16		glycoside hydrolases family 17 signature (PS00587), glycoside hydrolases family 17 (PF00332)
N	4, 5	At4g01610	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	14,4	6	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
N	4-6	At1g47128	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	16,7	7	48785,27	5,01	1-21		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), cysteine active site (PS00640), asparagine active site (PS00139), Cys-rich region (PS50311), papaine family cysteine protease (PF00112)
N	4, 5	At3g12500	glycoside hydrolase family 19 (chitinase)	9,8	3	32367,86	6,24	1-20		chitinase family 19 signature 1 (PS00773), signature 2 (PS00774), chitin recognition or binding domain signature (PS00026), chitinase class I (PF00182), chitin recognition protein (PF00187)
N	5, 6	At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	13,6	6	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
N	5, 6	At1g11840	glyoxalase I homologue (lactoylglutathione lyase) (ATGLX1)	17,6	5	31928,38	5,19			glyoxalase I signature 1 (PS00934), glyoxalase I signature 2 (PS00935), glyoxalase/bleomycin resistance protein/dioxygenase superfamily (PF00903)
N	8	At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	19,1	4	20982,78	5,20	1-26		trypsin and protease inhibitor Kunitz legume (PF00197)
O	1	At3g16850	glycoside hydrolase family 28 (polygalacturonase)	25,5	8	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
O	2	At4g16590	glycoside transferase family 2 (cellulose synthase like) (CslA1)	8,8	4	60927,18	9,80	1-27	transmembrane domains	glycoside hydrolase family 2 (PF00535)
O	3	At1g23190	homologous to phosphoglucomutase	12,3	5	63170,37	5,79			phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain I (PF02878), phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain II (PF02879), phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain III (PF02880), phosphoglucomutase/phosphomannomutase, C-terminal domain (PF00408)
O	4	At1g47128	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	16,7	4	48785,27	5,01	1-21		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), cysteine active site (PS00640), asparagine active site (PS00139), Cys-rich region (PS50311), papaine family cysteine protease (PF00112)
O	5	At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	13,6	3	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
O	6	At5g44360	homologous to berberine-bridge enzyme (S)-reticulon: oxygen oxidoreductase	7,9	4	57931,72	9,83	1-22		oxygen oxido-reductases covalent FAD binding site (PS00862), FAD binding domain (PF01565), berberine and berberine like (PF08031)
Q	1-6	At3g16850	glycoside hydrolase family 28 (polygalacturonase)	36,0	16	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
Q	2, 3	At5g34940	glycoside hydrolase family 79 (endo beta-glucuronidase/heparanase)	18,1	9	57219,88	9,13	1-21		glycoside hydrolase family 79, N-terminal domain (PF03662)
Q	4-6	At2g17120	expressed protein (LysM domain, GPI anchor)	12,3	6	35262,82	6,24	1-23	GPI anchor	LysM domain (PF01476)
Q	7, 8, 10	At3g49120	peroxidase (AtPrx34)	30,0	11	35695,45	9,41	1-30		peroxidase active site signature (PS00436), peroxidases proximal heme-ligand signature (PS00435)
Q	6	At3g43270	carbohydrate esterase family 8 (pectin methyltransferase)	13,3	7	54929,20	9,55	1-24		pectinesterase (PF01095), plant invertase/pectin methyltransferase inhibitor PME1 (PF04043)
Q	7	At2g36870	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH32)	16,4	6	31928,02	9,92	1-24		glycoside hydrolase family 16 (PF00722), xyloglucan endo-transglycosidase (XET) C-terminus (PF06955)
Q	8-10	At3g32980	peroxidase (AtPrx32)	31,8	14	35712,41	6,00	1-29		peroxidases active site signature (PS00436)
Q	10, 11	At5g06860	PGIP1 (LRR domains)	21,5	7	34324,36	9,56	1-21		LRR (PF00560, PS05052, PS05056)
Q	10, 11	At1g70850	homologous to Bet v I allergen family	41,1	13	35569,93	4,96			pathogenesis-related protein Bet v I family (PF00407)
Q	10, 11	At1g54030	homologous to lipase/acylhydrolase (GDSL family)	14,5	5	46082,76	6,20		transmembrane domains	lipase/acylhydrolase with GDSL-like motif (PS0241), GDSL-like Lipase/Acylhydrolase (PF00657)
Q	12, 13	At3g12500	glycoside hydrolase family 19 (chitinase)	10,5	4	32367,86	6,24	1-20		chitinase family 19 signature 1 (PS00773), signature 2 (PS00774), chitin recognition or binding domain signature (PS00026), chitinase class I (PF00182), chitin recognition protein (PF00187)
Q	14, 15	At1g47128	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	13,8	6	48785,27	5,01	1-21		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), cysteine active site (PS00640), asparagine active site (PS00139), Cys-rich region (PS50311), papaine family cysteine protease (PF00112)

Table S2, supplementary data

Q	14, 15	At4g01610	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	18,9	6	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
Q	15, 16	At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	10,8	5	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
Q	17	At2g28490	expressed protein (cupin domain)	9,2	5	53594,59	5,45	1-19		cupin domain profile (PS50849), cupin (PF00190)
Q	18-20, 22, 26	At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	23,2	5	20982,78	5,20	1-26		trypsin and protease inhibitor Kunitz legume (PF00197)
Q	19	At2g15220	expressed protein (Plant Basic Secreted Protein domain)	30,1	9	22772,21	9,09	1-21		neutral zinc metalloproteases, zinc-binding region signature (PS00142), plant basic secretory protein (PF04450)
Q	21	At5g55730	fasciclin-like arabinogalactan protein (AtFLA1)	13,9	4	42165,78	5,65	1-24	GPI anchor	Beta-Ig-H3/fasciclin domain (PS50213)
Q	23, 24	At2g04690	expressed protein (homologous to a human brain CREG protein)	30,0	6	20214,11	8,00	1-28		
Q	27, 28	At4g16500	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	60,0	10	10183,74	9,78	1-22		cystatin domain (PF00031)
Q	28	At2g38530	non-specific lipid transfer protein type 1 (LTP2)	30,0	4	9475,92	11,90	1-25		protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00597)
R	1	At3g16850	glycoside hydrolase family 28 (polygalacturonase)	29	11	46982,42	5,25	1-19		glycoside hydrolase family 28 (PF00295)
R	1	At5g25980	peroxidase (AtPrx43)	19,1	4	36806,45	5,45	1-32	transmembrane domains	peroxidases proximal heme-ligand signature (PS00435), plant heme peroxidase family profile (PS50873), peroxidase (PF00141)
R	2	At1g78860	homologous to lectin (curculin-like)	11,7	4	46804,29	6,00	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
R	3	At1g78850	homologous to lectin (curculin-like)	18,8	8	46664,12	8,90	1-22		PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
R	3	At5g34940	glycoside hydrolase family 79 (endo beta-glucuronidase/heparanase)	9,9	4	57219,88	9,13	1-21		glycoside hydrolase family 79, N-terminal domain (PF03662)
R	3	At3g06770	glycoside hydrolase family 28 (polygalacturonase)	8,7	4	46124,62	6,21	1-19		glycoside hydrolase family 28 (PF00295)
R	3	At1g26850	expressed protein (DUF248)	7,1	4	66120,79	6,19	1-32		putative methyltransferase (DUF248) (PF03141)
R	4, 5	At1g03220	homologous to carrot EDGP and tomato XEGIP	19,2	6	43394,37	9,99	1-22		
R	4-6, 11	At1g09750	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	16,9	9	45038,78	9,09	1-23		lipase Ser active site (PS00120), Ser-rich region (PS050324), eukaryotic aspartic protease (PF00026)
R	6	At3g43270	carbohydrate esterase family 8 (pectin methyltransferase)	14,2	8	54929,20	9,55	1-24		pectinesterase (PF01095), plant invertase/pectin methyltransferase inhibitor (PMEI) (PF04043)
R	6, 8, 9	At3g49120	peroxidase (AtPrx34)	27,5	10	35695,45	9,41	1-30		peroxidase active site signature (PS00436), peroxidases proximal heme-ligand signature (PS00435)
R	7, 10	At5g06860	PGIP1 (LRR domains)	29,1	9	34324,36	9,56	1-21		LRR (PF00560, PS50502, PS50506)
R	7	At3g32980	peroxidase (AtPrx32)	21,3	10	35712,41	6,00	1-29		peroxidases active site signature (PS00436)
R	7, 10	At1g70850	homologous to Bet v I allergen family	42,7	6	35569,93	4,96			pathogenesis-related protein Bet v I family (PF00407)
R	8	At3g49110	peroxidase (AtPrx33)	15,5	6	35383,10	7,91	1-33		peroxidase active site signature (PS00436), peroxidases proximal heme-ligand signature (PS00435), plant heme peroxidase family profile (PS50873)
R	12-14, 17	At1g47128	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	13,8	7	48785,27	5,01	1-21		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), cysteine active site (PS00640), asparagine active site (PS00139), Cys-rich region (PS50311), papain family cysteine protease (PF00112)
R	13-16	At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	17,1	9	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases asparagine active site (PS00640), papain family cysteine protease (PF00112), granulin (PF00396)
R	13	At4g01610	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	17,6	6	36995,63	5,79	1-22		eukaryotic thiol (cysteine) proteases histidine active site (PS00639), papain family cysteine protease (PF00112), propeptide_C1 (PF08127)
R	13	At3g45970	expansin-like A (AtEXLA1)	21,9	5	26258,12	9,46	1-20		pollen allergen (PF01357), barwin (PF00967)
R	16	At5g44410	homologous to berberine-bridge enzyme (S)-reticulon: oxygen oxidoreductase	11,8	4	57892,40	9,06	1-29		oxygen oxidoreductases covalent FAD-binding site (PS00862), berberine and berberine like (PF08031), FAD binding domain (PF01565)
R	18, 19, 25	At4g12880	early nodulin (AtEN20) (phytoalexin)	22,7	4	14273,34	9,64	1-18		plastocyanin-like domain (PF02298) (copper binding proteins)
R	19	At2g15220	expressed protein (Plant Basic Secreted Protein domain)	24,9	5	22772,21	9,09	1-21		neutral zinc metalloproteases, zinc-binding region signature (PS00142), plant basic secretory protein (PF04450)
R	19-21, 28, 30	At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	33,0	8	20982,78	5,20	1-26		trypsin and protease inhibitor Kunitz legume (PF00197)
R	21-23	At5g55730	fasciclin-like arabinogalactan protein (AtFLA1)	10,1	3	42165,78	5,65	1-24	GPI anchor	Beta-Ig-H3/fasciclin domain (PS50213)
R	24, 25	At5g01300	homologous to phosphatidylethanolamine-binding protein	39,5	6	71823,40	5,15			phosphatidylethanolamine-binding protein (PF01161),
R	25, 26	At2g04690	expressed protein (homologous to a human brain CREG protein)	33,8	6	20214,11	8,00	1-28		
R	25	At4g25260	homologous to plant invertase/pectin methyltransferase inhibitor (PMEI)	16,9	4	15737,05	5,50	1-19		plant invertase/pectin methyltransferase inhibitor (PF04043)
R	27-33	At1g09310	expressed protein (DUF538)	40,1	5	19947,00	5,12			domain of unknown function DUF538 (PF04398)

Table S2, supplementary data

R				5	16578,63	5,12		glycine-rich region profile (PS50315), eukaryotic RNA Recognition Motif (RRM) profile (PS50102), RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) (PF00076)
	27-29	At4g39260	homologous to RNA binding protein (AtGRP8)	35,5				pathogenesis-related protein Bet v I family (PF00407)
R	28	At4g23670	homologous to Bet v I allergen family	63,6	13	17517,73	5,84	ubiquitin domain signature (PS00299), ubiquitin domain profile (PS50053), ubiquitin family (PF00240)
R	31, 36	At1g31340*	polyubiquitin (AtUBQ7) (*and/or ubiquitin extension protein)	23,0	4	17397,00	5,85	P-II protein C-terminal region signature (PS00638), nitrogen regulatory protein P-II (PS00543)
R	33	At4g01900	PII nitrogen sensing protein (GLB I)	18,4	4	21275,40	5,12	SOUL heme binding protein (PF04832)
R	34	At1g78460	expressed protein (SOUL heme binding domain)	16,9	4	21857,91	9,75	protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00597)
R	34-36	At2g38530	non-specific lipid transfer protein type 1 (LTP2)	40,0	5	9475,92	11,90	
S	1	At3g24480	LRR-extensin (AtLRX4)	9,7	4	51682,00	6,39	1-25
S	1	At4g18670	LRR-extensin (AtLRX5)	5,7	4	87040,18	9,02	1-27
S	2	At1g05570	glycosyl transferase family 48 (callose synthase) (AtCalS1)	3,3	5	225730,52	9,51	
S	3-6, 38-41	At2g38530	non-specific lipid transfer protein type 1 (LTP2)	50,0	6	9475,92	11,90	1-25
S	4	At2g34930	expressed protein (LRR domains)	13,5	12	97059,60	9,59	1-28
S	7	At1g78860	homologous to lectin (curculin-like)	11,7	4	46804,29	6,00	1-22
S	7	At2g27190	homologous to purple acid phosphatase (PAP1)	8,5	3	50962,71	5,75	1-28
S	8, 9	At1g33590	expressed protein (LRR domains)	31,8	12	48103,60	9,90	1-24
S	8, 9	At1g78850	homologous to lectin (curculin-like)	26,5	11	46664,12	8,90	1-22
S	9-11, 13	At1g03220	homologous to carrot EDGP and tomato XEGIP	37,9	11	43394,37	9,99	1-22
S	10, 11, 13	At1g03230	homologous to carrot EDGP and tomato XEGIP	14,7	8	43723,02	10,27	1-23
S	10, 11, 19	At1g09750	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	17,6	9	45038,78	9,09	1-23
S	12, 13	At3g43270	carbohydrate esterase family 8 (pectin methyltransferase)	16,3	12	54929,20	9,55	1-24
S	12, 13	At3g54400	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	13,6	6	43313,31	10,26	1-19
S	14-18, 23	At1g78830	homologous to lectin (curculin-like)	24,0	8	48101,66	9,45	1-22
S	14	At3g48460	homologous to lipase/acylhydrolase (GDGL family)	20,7	6	39627,85	8,91	1-26
S	15-17	At5g06860	PGIP1 (LRR domains)	22,2	6	34324,36	9,56	1-21
S	16, 17	At1g70850	homologous to Bet v I allergen family	24,0	5	35569,93	4,96	
S	16	At4g01610	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	17,6	5	36995,63	5,79	1-22
S	16	At5g10770	homologous to Asp protease (CND41 peptidase) (Peptidase family A1, A01.050, MEROPS)	11,8	3	37068,36	5,48	1-22
S	16	At2g40880	inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	32,0	3	11955,61	9,71	1-22
S	18, 19	At1g29670	homologous to lipase/acylhydrolase (GDGL family)	38,0	11	37153,43	9,65	1-24
S	18	At5g15720	homologous to lipase/acylhydrolase (GDGL family)	13,2	5	48101,66	9,45	1-22
S	19	T23B7.10	homologous to PGIP1 (LRR protein FLR1)	16,9	4	33134,88	9,62	1-23
S	20, 21, 24	At4g30170	peroxidase (AtPrx45)	40,9	12	33052,57	10,43	1-25
S	21-23	At3g45970	expansin-like A (AtEXLA1)	32,4	9	26258,12	9,46	1-20
S	22	At4g38400	expansin-like A (AtEXLA2)	26,0	7	26516,93	9,60	1-20
S	22	At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	16,0	7	36995,63	5,79	1-22
S	24, 25	At4g29270	homologous to acid phosphatase	26,6	7	25900,84	9,15	1-26
S	25-28	At4g12880	early nodulin (AtEN20) (phytoecyanin)	30,5	6	14273,34	9,64	1-18



Table S2, supplementary data

S	28, 29	<b>At3g02110</b>	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	13,0	4	51075,46	9,35	1-22	serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
S	28, 32-35	<b>At1g73260</b>	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	19,1	6	20982,78	5,20	1-26	trypsin and protease inhibitor Kunitz legume (PF00197)
S	30, 31	<b>At1g17860</b>	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	66,8	13	19718,68	9,76	1-21	soybean trypsin inhibitor (Kunitz) protease inhibitors family signature (PS00283), trypsin and protease inhibitor (PF00197)
S	34-36	<b>At1g09310</b>	expressed protein (DUF538)	40,8	7	19947,00	5,12		domain of unknown function DUF538 (PF04398)
S	36	<b>At5g38410</b>	RUBISCO small subunit B3	22,6	5	20284,25	9,40		ribulose biphosphate carboxylase, small chain (PF00101)
S	37	<b>At2g02850</b>	plantacyanin (AIPNC) (phytoeyanin)	20,1	3	11167,60	9,08	1-28	plastocyanin-like domain (PF02298) (copper binding protein)
S	38	<b>At4g16500</b>	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	30,0	4	10183,74	9,78	1-22	cystatin domain (PF00031)
T	1	<b>At3g24480</b>	LRR-extensin (AtLRX4)	8,5	5	51682,00	6,39	1-25	plant specific LRR profile (PS05052), Leucine Rich Repeat (PF00560), Pro-rich profile (PS50099)
T	2-6	<b>At1g68560</b>	glycoside hydrolase family 31 (alpha-xylosidase) (AtXYL1)	22,8	17	99641,37	6,26	1-27	glycoside hydrolases family 31 active site (PS00129)
T	4	<b>At4g18670</b>	LRR-extensin (AtLRX5)	5,0	4	87040,18	9,02	1-27	histidine-rich region profile (PS50316), plant specific LRR profile (PS05052), serine-rich region profile (PS50324), Leucine Rich Repeat (PF00560)
T	7	<b>At3g07130</b>	homologous to purple acid phosphatase	10,9	6	58523,64	5,71	1-17	metallo-phosphoesterase motif (PS50185), calcineurin-like phosphoesterase (PF00149)
T	7	<b>At2g37620#</b>	actin (ACT1)	22,0	7	41797,89	5,12		actins and actin-related proteins signature (PS01132), actins signature 1 (PS00406), actins signature 2 (PS00432), actin (PF00022)
T	7	<b>At3g53750#</b>	actin (ACT3)	22,0	7	41797,89	5,12		actins and actin-related proteins signature (PS01132), actins signature 1 (PS00406), actins signature 2 (PS00432), actin (PF00022)
T	8, 9	<b>At3g61490</b>	glycoside hydrolase family 28 (polygalacturonase)	20,0	7	49497,06	5,76	1-23	glycoside hydrolase family 28 (PF00295)
T	9	<b>At2g27190</b>	homologous to purple acid phosphatase (PAP1)	15,6	7	50962,71	5,75	1-28	metallo-phosphoesterase motif (PS50185), calcineurin-like phosphoesterase (PF00149)
T	10, 11	<b>At1g33590</b>	expressed protein (LRR domains)	40,0	16	48103,60	9,90	1-24	LRR (PF00560), leucine zipper pattern (PS00029)
T	10	<b>At1g78850</b>	homologous to lectin (curculin-like)	16,1	6	46664,12	8,90	1-22	PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
T	10, 12-14, 27	<b>At1g03230</b>	homologous to carrot EDGP and tomato XEGIP	45,6	15	43723,02	10,27	1-23	
T	12	<b>At4g20860</b>	homologous to berberine-bridge enzyme (S)-reticulon:oxygen oxidoreductase	15,8	6	58375,98	9,71	1-16	oxygen oxidoreductases covalent FAD-binding site (PS00862), FAD_binding_4
T	13	<b>At1g03220</b>	homologous to carrot EDGP and tomato XEGIP	12,7	4	43394,37	9,99	1-22	FAD binding domain (PF01565), BBE Berberine and berberine like (PF08031)
T	13, 14	<b>At3g54400</b>	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	20,5	6	43313,31	10,26	1-19	eukaryotic aspartyl protease (PF00026)
T	14, 15	<b>At4g33220</b>	carbohydrate esterase family 8 (pectin methylesterase)	17,5	7	56832,32	6,24	1-19	pectinesterase (PF01095)
T	14-22, 24-27, 29, 38, 39, 41-47	<b>At5g11420</b>	expressed protein (DUF642)	30,0	17	37344,57	7,70	1-22	domain of unknown function DUF 642 (PF04862)
T	17-20, 24, 25, 29-31	<b>At3g20820</b>	expressed protein (LRR domains)	50,0	16	37850,70	9,90	1-19	typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22+like LRR profile (PS50504)
T	17-19	<b>At1g71695</b>	peroxidase (AtPrx12)	23,5	7	37104,36	9,46	1-22	peroxidase (PF00141)
T	19	<b>At5g10770</b>	homologous to Asp protease (CND41 peptidase) (Peptidase family A1, A01.050, MEROPS)	15,6	5	37068,36	5,48	1-22	eukaryotic aspartyl protease (PF00013)
T	19	<b>At1g70850</b>	homologous to Bet v I allergen family	23,4	6	35569,93	4,96		pathogenesis-related protein Bet v I family (PF00407)
T	19	<b>At5g12940</b>	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	20,0	5	36784,96	10,45	1-29	plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
T	20	<b>At1g78830</b>	homologous to lectin (curculin-like)	13,8	5	48101,66	9,45	1-22	PAN domain (PF00024), lectin (probable mannose binding) (PF01453)
T	20, 38, 39, 41, 42	<b>At5g25460</b>	expressed protein (DUF642)	20,0	7	37948,52	9,03	1-19	domain of unknown function DUF 642 (PF04862)
T	22-24, 40	<b>T23B7.10</b>	homologous to PGIP1 (LRR protein FLR1)	34,5	8	33134,88	9,62	1-23	typical LRR profile (PS50506), plant specific LRR profile (PS50502), plant disease resistance response protein (PF03018)
T	22	<b>At5g64100</b>	peroxidase (AtPrx69)	26,0	5	33228,93	10,72	1-23	peroxidases active site signature (PS00436)
T	22	<b>At5g66390</b>	peroxidase (AtPrx72)	18,7	5	35004,65	9,63	1-23	peroxidase (PF00141)
T	23	<b>At1g53070</b>	homologous to lectin (legume lectin domains)	15,8	5	27805,05	9,15	1-23	legume lectins alpha domain (PF00138), legume lectins beta domain (PF00139)
T	24-27	<b>At2g06850</b>	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH4)	30,7	10	31716,77	9,65	1-24	glycoside hydrolases family 16 (PF00722)
T	25, 27	<b>At4g30170</b>	peroxidase (AtPrx45)	40,3	14	33052,57	10,43	1-25	haem peroxidase, plant/fungal/bacterial (PS00435, PS00436, PS00873)
T	26-28	<b>At3g45970</b>	expansin-like A (AtEXLA1)	35,8	7	26258,12	9,46	1-20	pollen allergen (PF01357), barwin (PF00967)

Table S2, supplementary data

T	26-28	At5g64260	homologous to <i>Nicotiana tabacum</i> phi-I	20,0	6	30550,86	9,98	1-19	phosphate-induced protein 1 conserved region (PF04674)
T	27-29	At5g09440	homologous to <i>Nicotiana tabacum</i> phi-I	20,0	8	27366,44	10,15	1-23	phosphate-induced protein 1 conserved region (PF04674)
T	28	At4g38400	expansin-like A (AtEXLA2)	20,0	7	26516,93	9,60	1-20	expansin, cellulose-binding-like domain profile (PS0843), expansin, family-45 endoglucanase-like domain profile (PS0842), rare lipoprotein A (RlpA)-like double-psi beta-barrel (PF03330), pollen allergen (PF01357)
T	30, 31	At2g28790	homologous to <i>Lycopersicon esculentum</i> osmotin	10,0	4	24551,60	9,53	1-24	thaumatin family (PF00314)
T	31	At4g29270	homologous to acid phosphatase	25,4	4	25900,84	9,15	1-26	HAD superfamily, subfamily IIIB (acid phosphatase) (PF03767)
T	31-35, 37, 38	At4g12880	early nodulin (AtEN20) (phytoecyanin)	44,7	8	14273,34	9,64	1-18	plastocyanin-like domain (PF02298) (copper binding proteins)
T	32-33	At1g20190	alpha-expansin (AtEXPA11)	19,8	5	24817,74	9,84	1-20	expansin, family-45 endoglucanase-like domain profile (PS0842), expansin, cellulose-binding-like domain profile (PS0843), pollen allergen (PF01357)
T	32	At1g09560	germin (subfamily 2, member 1, GLP5)	14,6	4	20472,37	9,71	1-23	germin (PS00725), cupin (PF00190)
T	34-37	At5g66590	homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	38,9	5	18040,82	9,65	1-19	SCP-like extracellular protein (PF00188) (sperm coating glycoprotein)
T	34	At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	19,1	3	20982,78	5,20	1-26	trypsin and protease inhibitor Kunitz legume (PF00197)
T	35	At3g02110	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	20,0	4	51075,46	9,35	1-22	serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
T	36, 37	At1g17860	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	25,5	4	19718,68	9,76	1-21	soybean trypsin inhibitor (Kunitz) protease inhibitors family signature (PS00283), trypsin and protease inhibitor (PF00197)
T	44-46	At2g38530	non-specific lipid transfer protein type 1 (LTP2)	40,0	5	9475,92	11,90	1-25	protease inhibitor/seed storage/LTP family (PF00234), plant lipid transfer proteins signature (PS00597)
T	44-46	At4g16500	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	50,0	8	10183,74	9,78	1-22	cystatin domain (PF00031)
T	45	At5g05110	inhibitor family I25 (phytoecyanin) (cystatin family, I25.014, MEROPS)	16,8	4	23934,58	8,48	1-24	cysteine proteases inhibitors signature (PS00287), cystatin domain (PF00031)
U	1-40	At5g11420	expressed protein (DUF642)	30,0	13	37344,57	7,70	1-22	domain of unknown function DUF 642 (PF04862)
U	1-14, 16-40	At5g25460	expressed protein (DUF642)	30,0	12	37948,52	9,03	1-19	domain of unknown function DUF 642 (PF04862)
U	2-7	At1g28290	proline-rich protein	6,4	5	35913,36	10,77	1-24	histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
U	3	At5g59370*	actin (ACT4)	14,1	5	41778,87	5,80		actins and actin-related proteins signature (PS01132), actins signature 1 (PS00406), actins signature 2 (PS00432), actin (PF00022)
U	3	At3g46520*	actin (ACT12)	14,1	5	41794,87	5,20		actins and actin-related proteins signature (PS01132), actins signature 1 (PS00406), actins signature 2 (PS00432), actin (PF00022)
U	3	At2g37620#	actin (ACT1)	13,5	5	41797,89	5,12		actins and actin-related proteins signature (PS01132), actins signature 1 (PS00406), actins signature 2 (PS00432), actin (PF00022)
U	3	At3g53750#	actin (ACT3)	13,5	5	41797,89	5,12		actins and actin-related proteins signature (PS01132), actins signature 1 (PS00406), actins signature 2 (PS00432), actin (PF00022)
U	8	At3g13790	glycoside hydrolase family 32 (beta-fructofuranosidase)	16,4	9	63490,72	9,68	1-25	glycoside hydrolases family 32 active site (PS00609), glycoside hydrolases family 32 (PF00251)
U	11, 12, 20	At1g03230	homologous to carrot EDGP and tomato XEGIP	35,9	11	43723,02	10,27	1-23	
U	11	At1g03220	homologous to carrot EDGP and tomato XEGIP	12,5	4	43394,37	9,99	1-22	
U	11	At4g16260	glycoside hydrolase family 17	12,4	4	35905,19	6,64	1-16	glycoside hydrolases family 17 signature (PS00587), glycoside hydrolases family 17 (PF00332)
U	13, 14	At5g19110	homologous to carrot EDGP and tomato XEGIP	23,9	9	41369,63	10,01	1-19	pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor PME1 (PF04043)
U	13, 14	At3g43270	carbohydrate esterase family 8 (pectin methylesterase)	20,0	8	54929,20	9,55	1-24	eukaryotic aspartyl protease (PF00026)
U	14	At5g07030	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	10,0	6	44860,54	10,82	1-19	typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22-like LRR profile (PS50504)
U	16, 17	At3g20820	expressed protein (LRR domains)	35,9	10	37850,70	9,90	1-19	plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
U	16-19, 21	At5g12940	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	40,0	15	36784,96	10,45	1-29	
U	16	At1g10550	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH33)	19,7	6	32314,04	9,73	1-21	glycoside hydrolases family 16 (PF00722), xyloglucan endo-transglycosidase (XET) C-terminus (PF06955)
U	21, 22, 24	At2g06850	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH4)	22,3	8	31716,77	9,65	1-24	glycoside hydrolases family 16 (PF00722)
U	22	At4g30170	peroxidase (AtPrx45)	24,3	6	33052,57	10,43	1-25	haem peroxidase, plant/fungal/bacterial (PS00435, PS00436, PS00873)



Table S2, supplementary data

U	23-25	At5g64260	homologous to <i>Nicotiana tabacum</i> phi-I	30,0	8	30550,86	9,98	1-19	phosphate-induced protein 1 conserved region (PF04674)
U	25, 26	At5g09440	homologous to <i>Nicotiana tabacum</i> phi-I	20,0	5	27366,44	10,15	1-23	phosphate-induced protein 1 conserved region (PF04674)
U	26	At2g28790	homologous to <i>Lycopersicon esculentum</i> osmotin	20,0	4	24551,60	9,53	1-24	thaumatin family (PF00314)
U	27-29	At1g20190	alpha-expansin (AtEXPA11)	30,2	6	24817,74	9,84	1-20	expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
U	29	At4g12880	early nodulin (AtEN20) (phytoeyanin)	12,1	3	14273,34	9,64	1-18	plastocyanin-like domain (PF02298) (copper binding proteins)
U	30, 31	At3g02110	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	10,0	4	51075,46	9,35	1-22	serine carboxypeptidases, histidine active site (PS00560), serine carboxypeptidases, serine active site (PS00131), serine carboxypeptidase (PF00450)
U	30-35	At5g66590	homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	38,9	7	18040,82	9,65	1-19	SCP-like extracellular protein (PF00188) (sperm coating glycoprotein)
U	32, 34	At1g77810	glycosyl transferase family 31 (galactosyl transferase)	19,6	4	41656,86	6,07	1-26	galactosyl transferase (PF01762)
U	32, 33	At2g34700	expressed protein (Ole e1 allergen domain)	30,3	6	16248,80	10,22	1-23	pollen proteins Ole e 1 family (PF01190)
U	35	At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	17,2	3	20982,78	5,20	1-26	trypsin and protease inhibitor Kunitz legume (PF00197)
U	37, 38, 40	At4g32460	expressed protein (DUF642)	14,0	7	37375,51	9,25	1-21	domain of unknown function DUF 642 (PF04862)
U	39, 41-44	At4g27160	2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	20,0	4	16909,07	7,69	1-17	protease inhibitor/seed storage/LTP family (PF00234)
U	42	At4g27150	2S albumin isoform 2 (protease inhibitor/seed storage/LTP family)	20,0	3	17473,78	6,34	1-17	protease inhibitor/seed storage/LTP family (PF00234)
V	1-4, 6, 7-18, 20-22, 26-32	At5g25460	expressed protein (DUF642)	37,1	15	37948,52	9,03	1-19	domain of unknown function DUF 642 (PF04862)
V	2-7	At1g28290	proline-rich protein	9,2	4	35913,36	10,77	1-24	histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
V	4, 5	At2g30210	homologous to laccase (AtLAC3)	15,4	8	61172,59	10,11	1-25	multicopper oxidases signature 2 (PS00080), multicopper oxidase (PF00394)
V	7-14, 16, 20, 22, 32	At5g11420	expressed protein (DUF642)	26,2	13	37344,57	7,70	1-22	domain of unknown function DUF 642 (PF04862)
V	8, 9	At1g03230	homologous to carrot EDGP and tomato XEGIP	13,6	5	43723,02	10,27	1-23	
V	10, 11	At5g07030	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	26,7	14	44860,54	10,82	1-19	eukaryotic aspartyl protease (PF00026)
V	12	At1g71695	peroxidase (AtPrx12)	16,5	5	37104,36	9,46	1-22	peroxidase (PF00141)
V	13-16	At5g12940	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	34,2	13	36784,96	10,45	1-29	plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
V	13	At3g20820	expressed protein (LRR domains)	19,7	5	37850,70	9,90	1-19	typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22+like LRR profile (PS50504)
V	17, 18	At3g14310	carbohydrate esterase family 8 (pectin methylesterase) (PME3)	17,4	9	60050,61	9,59	1-40	trans-membrane domains pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
V	18	At1g53830	carbohydrate esterase family 8 (pectin methylesterase) (PME2)	7,9	4	64172,61	9,76	1-40	trans-membrane domains pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
V	19-21	At5g64260	homologous to <i>Nicotiana tabacum</i> phi-I	28,9	9	30550,86	9,98	1-19	phosphate-induced protein 1 conserved region (PF04674)
V	20-22	At5g09440	homologous to <i>Nicotiana tabacum</i> phi-I	20,5	6	27366,44	10,15	1-23	phosphate-induced protein 1 conserved region (PF04674)
V	22	At2g28790	homologous to <i>Lycopersicon esculentum</i> osmotin	12,9	3	24551,60	9,53	1-24	thaumatin family (PF00314)
V	23, 24	At5g02260	alpha-expansin (AtEXPA9)	24,0	6	25579,80	10,19	1-21	expansin, family-45 endoglucanase-like domain profile (PS50842), pollen allergen (PF01357)
V	24, 25	At1g20190	alpha-expansin (AtEXPA11)	18,7	6	24817,74	9,84	1-20	expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
V	25	At4g12880	early nodulin (AtEN20) (phytoeyanin)	23,4	9	14273,34	9,64	1-18	plastocyanin-like domain (PF02298) (copper binding proteins)
V	26, 27	At2g05580	glycine-rich protein	40,1	6	21517,58	10,21	1-20	glycine-rich region profile (PS50315)
V	28	At1g01980	homologous to berberine-bridge enzyme (S)-reticulon: oxygen oxidoreductase	12,0	6	58023,25	8,98	1-20	berberine and berberine like (PF08031), FAD binding domain (PF01565)
V	31	At4g32460	expressed protein (DUF642)	14,0	5	37375,51	9,25	1-21	domain of unknown function DUF 642 (PF04862)
V	33-35	At4g27150	2S albumin isoform 2 (protease inhibitor/seed storage/LTP family)	20,6	3	17473,78	6,34	1-17	protease inhibitor/seed storage/LTP family (PF00234)
V	33-35	At4g27160	2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	18,3	3	16909,07	7,69	1-17	protease inhibitor/seed storage/LTP family (PF00234)

Table S2, supplementary data

W	1	At2g30210	homologous to laccase (AtLAC3)	14,2	7	61172,59	10,11	1-25		multicopper oxidases signature 2 (PS00080), multicopper oxidase (PF00394)
W	2, 5	At1g28290	proline-rich protein	9,2	4	35913,36	10,77	1-24		histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
W	6	At3g43270	carbohydrate esterase family 8 (pectin methylesterase)	20,0	10	54929,20	9,55	1-24		pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PMEI) (PF04043)
W	6	At5g07030	homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	20,0	8	44860,54	10,82	1-19		eukaryotic aspartyl protease (PF00026)
W	6, 7, 18	At5g11420	expressed protein (DUF642)	30,0	11	37344,57	7,70	1-22		domain of unknown function DUF 642 (PF04862)
W	7-9, 12, 17, 18	At5g25460	expressed protein (DUF642)	20,0	7	37948,52	9,03	1-19		domain of unknown function DUF 642 (PF04862)
W	8	At3g20820	expressed protein (LRR domains)	10,0	4	37850,70	9,90	1-19		typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22+like LRR profile (PS50504)
W	9, 11	At5g12940	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	20,0	6	36784,96	10,45	1-29		plant specific LRR profile (PS50502), Leucine Rich Repeat (PF00560)
W	10, 11	At5g14920	proline-rich protein	Edman sequencing + MALDI-TOF-TOF	3	26753,23	10,98	1-21		gibberellin regulated protein (PF02704)
W	12	At3g14310	carbohydrate esterase family 8 (pectin methylesterase) (PME3)	19,7	7	60050,61	9,59	1-40	trans-membrane domains	pectinesterase signature 1 (PS00800), pectinesterase signature 2 (PS00503), pectinesterase (PF01095), plant invertase/pectin methylesterase inhibitor (PF04043)
W	13	At5g64260	homologous to <i>Nicotiana tabacum</i> phi-I	11,0	4	30550,86	9,98	1-19		phosphate-induced protein 1 conserved region (PF04674)
W	14	At5g09440	homologous to <i>Nicotiana tabacum</i> phi-I	17,3	5	27366,44	10,15	1-23		phosphate-induced protein 1 conserved region (PF04674)
W	15, 16	At1g20190	alpha-expansin (AtEXPA11)	19,8	5	24817,74	9,84	1-20		expansin, family-45 endoglucanase-like domain profile (PS50842), expansin, cellulose-binding-like domain profile (PS50843), pollen allergen (PF01357)
W	16	At4g12880	early nodulin (AtEN20) (phytoalexin)	33,6	4	14273,34	9,64	1-18		plastocyanin-like domain (PF02298) (copper binding proteins)
W	19, 20	At4g27150*	2S albumin isoform 2 (protease inhibitor/seed storage/LTP family)	13,5	2	17473,78	6,34	1-17		protease inhibitor/seed storage/LTP family (PF00234)
W	19-21	At4g27160*	2S albumin isoform 3 (protease inhibitor/seed storage/LTP family)	14,0	2	16909,07	7,69	1-17		protease inhibitor/seed storage/LTP family (PF00234)
X	1	At1g28290	proline-rich protein	6,4	3	35913,36	10,77	1-24		histidine-rich region profile (PS50316), proline-rich region profile (PS50099), pollen proteins Ole e 1 family (PF01190)
X	2, 3	At5g11420	expressed protein (DUF642)	11,7	5	37344,57	7,70	1-22		domain of unknown function DUF 642 (PF04862)
X	4	At3g20820	expressed protein (LRR domains)	7,7	3	37850,70	9,90	1-19		typical LRR profile (PS50506), plant specific LRR profile (PS50502), SDS22+like LRR profile (PS50504)
X	4	At5g25460	expressed protein (DUF642)	12,5	4	37948,52	9,03	1-19		domain of unknown function DUF 642 (PF04862)
X	5	At5g14920	proline-rich protein	Edman sequencing + MALDI-TOF-TOF	3	26753,23	10,98	1-21		gibberellin regulated protein (PF02704)
X	6	At4g27170	2S albumin isoform 4 (protease inhibitor/seed storage/LTP family)	15,1	3	16946,26	7,80	1-21		protease inhibitor/seed storage/LTP family (PF00234)
Y	1	At5g11420	expressed protein (DUF642)	10,0	5	37344,57	7,70	1-22		domain of unknown function DUF 642 (PF04862)
Y	2	At5g25460	expressed protein (DUF642)	10,0	4	37948,52	9,03	1-19		domain of unknown function DUF 642 (PF04862)
Y	3	At5g14920	proline-rich protein	Edman sequencing + MALDI-TOF-TOF MS <sup>c</sup>	3	26753,23	10,98	1-21		gibberellin regulated protein (PF02704)

a: The following criteria were retained for protein identification: at least 4 peptides are required for proteins larger than 15 kDa, only 3 peptides can be used for proteins smaller than 15 kDa. Allowed modifications are 1 miscleavage and Met oxidation. In each series of samples ( i.e., total and M to Z), the best score of identification is shown.

b: Functional domains were predicted as described in Material and methods. PF stands for PFAM, PS for PROSITE and IPR for InterPro.

c: N-terminal sequencing, XXTLPS; LC-MS/MS sequencing, CGQHSR.

**Table S3, supplementary material.** Classification of proteins extracted from purified cell walls of 5- and 11-day-old *Arabidopsis* hypocotyls in FPLC fractions A to Z.

\* or # proteins that could not be distinguished

Colour code

	secretory proteins (Presence of predicted signal peptide was checked using PSORT: <a href="http://psort.ims.u-tokyo.ac.jp/form.html">http://psort.ims.u-tokyo.ac.jp/form.html</a> , and TargetP: <a href="http://www.cbs.dtu.dk/services/TargetP/">http://www.cbs.dtu.dk/services/TargetP/</a> )
	transmembrane domains (Presence of predicted transmembrane domains was checked using Aramemnon: <a href="http://aramemnon.botanik.uni-koeln.de/">http://aramemnon.botanik.uni-koeln.de/</a> )
	GPI anchors (presence or GPI anchors was checked using Aramemnon: <a href="http://aramemnon.botanik.uni-koeln.de/">http://aramemnon.botanik.uni-koeln.de/</a> )
	intracellular proteins
	proteins identified after 1D-separation (1-DE), but not identified after 2D-separation (cationic exchange chromatography followed by 1-DE)
	proteins only identified in cell walls of 5-day-old etiolated hypocotyls
	proteins only identified in cell walls of 11-day-old etiolated hypocotyls

	accession AGI	annotation	5-day-old hypocotyls													11-day-old hypocotyls														
			total	A	B	C	D	E	F	G	H	I	J	K	L	M	total	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
predicted secreted proteins																														
proteins acting on carbohydrates																														
1	At2g06850	glycoside hydrolase family 16 (endoxyloglucan transferase) (AtXTH4)	1							1	1	1					1							1	1					
2	At5g13870	glycoside hydrolase family 16 (endoxyloglucan transferase) (AtXTH5)									1																			
3	At3g44990	glycoside hydrolase family 16 (endoxyloglucan transferase) (AtXTH31)										1	1																	
4	At2g36870	glycoside hydrolase family 16 (endoxyloglucan transferase) (AtXTH32)																			1									
5	At1g10550	glycoside hydrolase family 16 (endoxyloglucan transferase) (AtXTH33)										1																		
6	At4g16260	glycoside hydrolase family 17																1												
7	At4g19810	glycoside hydrolase family 18 (chitinase)							1																1					
8	At3g12500	glycoside hydrolase family 19 (chitinase)							1																					
9	At3g55260	glycoside hydrolase family 20 (beta-hexosaminidase)							1																					
10	At1g19170	glycoside hydrolase family 28 (polygalacturonase)	1																											
11	At2g33160	glycoside hydrolase family 28 (polygalacturonase)		1																										
12	At3g06770	glycoside hydrolase family 28 (polygalacturonase)																												
13	At3g16850	glycoside hydrolase family 28 (polygalacturonase)	1	1			1	1	1								1	1												
14	At3g61490	glycoside hydrolase family 28 (polygalacturonase)																												
15	At4g18180	glycoside hydrolase family 28 (polygalacturonase)							1																					
16	At1g68560	glycoside hydrolase family 31 (alpha-xylosidase) (AtXYL1)	1								1	1	1																	
17	At3g13790	glycoside hydrolase family 32 (beta-fructofuranosidase)										1																		
18	At5g34940	glycoside hydrolase family 79 (endo beta-glucuronidase/heparanase)	1					1																						
19	At1g11580	carbohydrate esterase family 8 (pectin methylesterase)										1	1	1																
20	At1g53830	carbohydrate esterase family 8 (pectin methylesterase) (AtPME2)										1	1	1																
21	At3g14310	carbohydrate esterase family 8 (pectin methylesterase) (AtPME3)	1						1			1	1	1																
22	At3g43270	carbohydrate esterase family 8 (pectin methylesterase)	1						1	1		1	1	1			1													
23	At4g33220	carbohydrate esterase family 8 (pectin methylesterase)	1									1	1																	
24	At5g53370	carbohydrate esterase family 8 (pectin methylesterase)							1			1																		
25	At4g37950	polysaccharide lyase family 4 (rhamnogalacturonate lyase)															1													
26	At1g05570	glycosyl transferase family 48 (callose synthase) (AtCalS1)									1																			
27	At5g02260	alpha-expansin (AtEXPA9)											1	1	1															
28	At1g20190	alpha-expansin (AtEXPA11)	1								1	1	1	1	1		1													
29	At5g39270	alpha expansin (AtEXP22)							1																					
30	At3g45970	expansin-like A (AtEXLA1)	1							1	1	1	1																	
31	At4g38400	expansin-like A (AtEXLA2)	1								1	1		1																
32	At3g45960	expansin-like A (AtEXLA3)										1																		

164

165

166

**Table S4, supplementary material.** Semi-quantification of CWP identified by MALDI-TOF MS from 5- and 11-day-old etiolated *Arabidopsis* hypocotyls.

A quantification index (QI) is attributed to each protein in each sample ( 5- and 11-day-old etiolated hypocotyls). It was calculated by adding the percentages of coverage by peptide mass mapping in all the bands in which the protein was identified. When the protein was only identified in the total extracts after 1D-E (1D-separation), the values obtained for this identification were retained. When the protein was identified only after cation exchange chromatography and 1D-E of the eluted fractions (2D-separation), the values obtained for this identification were retained. When the protein was identified in both cases, only the values corresponding to the latter were taken into account.

Only differences higher than 2-fold were considered as significant and are indicated in bold red.

CWP CWP identified only in 5-day-old etiolated hypocotyls  
CWP CWP identified only in 11-day-old etiolated hypocotyls  
 nd QI was not calculated since the protein was identified in the same separation conditions in the two samples.

accession AGI	annotation	5 day-old hypocotyls QI	11day-old hypocotyls QI	1D- and 2D- separations	1D- separation only	2D- separation only
<b>proteins acting on carbohydrates</b>						
At2g06850	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH4)	<b>4,59</b>	1,35	1		
At5g13870	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH5)	0,40				1
At3g44990	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH31)	0,41				1
At2g36870	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH32)		0,16			1
At1g10550	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH33)	<b>0,56</b>	0,20			1
At4g16260	glycoside hydrolase family 17		0,32			1
At4g19810	glycoside hydrolase family 18 (chitinase)	<b>0,74</b>	0,18			1
At3g12500	glycoside hydrolase family 19 (chitinase)	0,34	0,41			1
At3g55260	glycoside hydrolase family 20 (beta-hexosaminidase)	0,17				1
At1g19170	glycoside hydrolase family 28 (polygalacturonase)	0,21			1	
At4g18180	glycoside hydrolase family 28 (polygalacturonase)	0,25				1
At3g16850	glycoside hydrolase family 28 (polygalacturonase)	3,77	2,38	1		
At3g06770	glycoside hydrolase family 28 (polygalacturonase)		0,28			1
At3g61490	glycoside hydrolase family 28 (polygalacturonase)	0,39	0,33			1
At2g33160	glycoside hydrolase family 28 (polygalacturonase)	0,08				1
At1g68560	glycoside hydrolase family 31 (alpha-xylosidase) (AtXYL1)	1,62	0,82	1		
At3g13790	glycosyl hydrolase family 32 (beta-fructofuranosidase)	0,09	0,16			1
At5g34940	glycosyl hydrolase family 79 (endo-beta-glucuronidase/heparanase)	<b>0,94</b>	0,11	1		
At1g53830	carbohydrate esterase family 8 (pectin methylesterase) (AtPME2)	<b>0,78</b>	0,10			1
At3g14310	carbohydrate esterase family 8 (pectin methylesterase) (AtPME3)	1,08	0,65	1		
At3g43270	carbohydrate esterase family 8 (pectin methylesterase)	2,24	1,23	1		
At4g33220	carbohydrate esterase family 8 (pectin methylesterase)	0,42	0,77	1		
At1g11580	carbohydrate esterase family 8 (pectin methylesterase)	0,96				1
At5g53370	carbohydrate esterase family 8 (pectin methylesterase)	0,10				1
At4g37950	polysaccharide lyase family 4 (rhamnogalacturonate lyase)		0,90	1		
At1g05570	glycoside transferase family 48 (callose synthase 1) (AtCalS1)	0,04	0,03			1
At5g02260	alpha-expansin (AtEXP9)	<b>1,22</b>	0,43			1
At1g20190	alpha-expansin (AtEXPA11)	2,22	2,59	1		
At5g39270	alpha-expansin (AtEXPA22)	0,11				1
At3g45970	expansin-like A (AtEXLA1)	3,45	2,04	1		
At4g38400	expansin-like A (AtEXLA2)	<b>1,48</b>	0,67	1		

Table S4, supplementary data

At3g45960	expansin-like A (AtEXLA3)	0,14			1	
<b>oxido-reductases</b>						
At1g71695	peroxidase (AtPrx12)	0,27	0,87			1
At3g21770	peroxidase (AtPrx30)	0,15	0,15		1	
At3g32980	peroxidase (AtPrx32)	3,18	0,87			1
At3g49110	peroxidase (AtPrx33)		0,16			1
At3g49120	peroxidase (AtPrx34)	1,53	2,04	1		
At3g50990	peroxidase (AtPrx36)	0,10			1	
At4g25980	peroxidase (AtPrx43)		0,19		1	
At4g30170	peroxidase (AtPrx45)	1,92	2,57	1		
At5g17820	peroxidase (AtPrx57)	0,23			1	
At5g64100	peroxidase (AtPrx69)	0,35	0,26	1		
At5g66390	peroxidase (AtPrx72)		0,19			1
At2g30210	homologous to laccase (AtLAC3)	0,40	0,56			1
At4g20860	homologous to berberine-bridge enzyme (S)-reticul:n:oxyen oxidoreductase		0,16	1		
At5g44360	homologous to berberine-bridge enzyme (S)-reticul:n:oxyen oxidoreductase	0,08	0,16			1
At5g44410	homologous to berberine-bridge enzyme (S)-reticul:n:oxyen oxidoreductase		0,12			1
At1g30710	homologous to berberine-bridge enzyme (S)-reticul:n:oxyen oxidoreductase	0,16				1
At1g09610	homologous to berberine-bridge enzyme (S)-reticul:n:oxyen oxidoreductase		0,12			
At4g12880	early nodulin (AtEN20) (phytoctyanin)	11,84	5,66	1		
At2g02850	plantacyanin (AtPNC) (phytoctyanin)		0,20			1
At5g22140	expressed protein (oxidoreductase domain)	0,28				1
At5g56490	expressed protein (FAD binding domain)	0,15				1
<b>proteases</b>						
At5g07030	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	1,43	0,86	1		
At1g09750	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	1,94	0,97	1		
At3g54400	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	0,59	0,78	1		
At5g10770	homologous to aspartic protease (CND41 peptidase) (Peptidase family A1, A01.050, MEROPS)	0,34	0,42			1
At3g02740	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	0,14				1
At1g79720	homologous to aspartic protease (CND41 peptidase) (Peptidase family A1, subfamily A1.050, MEROPS)	0,20				1
At3g52500	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	0,28				1
At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	1,78	1,65	1		
At1g47128	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	0,73	1,30	1		
At4g01610	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	1,43	1,50	1		
At4g36880	homologous to cysteine proteinase (papain family) (Peptidase family C1, Brassicain, C01.021 MEROPS)	0,16				1
At4g30610	homologous to serine carboxypeptidase (BRS1 - Brassinosteroid-Insensitive BRI suppressor 1) (Peptidase family S10, S10.015 MEROPS)	0,45				1
At3g02110	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	1,69	0,71	1		
At5g23210	homologous to serine carboxypeptidase (SCPL34) (Peptidase family S10, S10.005, MEROPS)		nd	1		



Table S4, supplementary data

proteins with interacting domains						
At1g53070	homologous to lectin (legume lectin domains)	0,71	0,16			1
At1g78820	homologous to lectin (curculin-like)	0,39				1
At1g78830	homologous to lectin (curculin-like)	2,38	1,48	1		
At1g78850	homologous to lectin (curculin-like)	1,93	1,98	1		
At1g78860	homologous to lectin (curculin-like)	0,84	0,24	1		
At5g06860	PGIP1	1,16	1,91			1
T23B7.10	homologous to PGIP1 (LRR protein FLR1)	0,61	1,55	1		
At5g12940	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	3,85	4,44	1		
At1g33590	expressed protein (LRR domains)	1,19	1,11	1		
At2g34930	expressed protein (LRR domains)	0,46	0,24			1
At3g20820	expressed protein (LRR domains)	2,15	5,12	1		
At2g17120	expressed protein (LysM domain)	1,18	0,35			1
At5g19110	homologous to carrot EDGP and to tomato XEGIP	0,84	0,39			1
At1g03220	homologous to carrot EDGP and to tomato XEGIP	2,37	4,02	1		
At1g03230	homologous to carrot EDGP and to tomato XEGIP	1,65	1,55	1		
At1g47710	homolog to serpin (serine protease inhibitor)	1,33	0,20			1
At1g17860	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	1,78	1,81	1		
At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	0,30	5,13	1		
At1g47540	inhibitor family I18 (mustard trypsin inhibitor-2 family) (family I18 unassigned peptidase inhibitor homologues, MEROPS)	nd	nd	1		
At2g40880	inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	0,22	0,31	1		
At5g05110	inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	0,10	0,17			1
At4g16500	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	2,12	5,61			1
At4g25260	homologous to plant invertase/ pectin methylesterase inhibitor (PMEI)	0,34	0,17			1
At5g46940	homologous to plant invertase/pectine methylesterase inhibitor (PMEI)	0,19				1
At5g46960	homologous to plant invertase/pectine methylesterase inhibitor (PMEI)	0,44				1
structural proteins						
At1g28290	proline-rich protein	2,61	2,14	1		
At5g14920	proline-rich protein	0,17	0,29			1
At2g05580	glycine-rich protein	0,74	1,61	1		
At4g13340	LRR-extensin (AtLRX3)	0,08				1
At3g24480	LRR-extensin (AtLRX4)	0,12	0,33			1
At4g18670	LRR-extensin (AtLRX5)	0,07	0,11			1
signaling						
At4g05200	homologous to receptor-kinase (RLK, DUF26-1b subfamily)	0,09				1
At5g55730	fasciclin-like arabinogalactan protein (AtFLA1)	0,19	0,58			1

Table S4, supplementary data

proteins related to lipid metabolism						
At1g29670	homologous to lipase/acylhydrolase (GDSL family)	2,27	0,75	1		
At1g54010	homologous to lipase/acylhydrolase (GDSL family)	1,67				1
At1g54030	homologous to lipase/acylhydrolase (GDSL family)	0,12	0,23		1	
At5g15720	homologous to lipase/acylhydrolase (GDSL family)		0,13			1
At3g48460	homologous to lipase/acylhydrolase (GDSL family)	0,98	0,21			1
At2g38530	non-specific lipid transfer protein type 1 (LTP2)	5,26	9,07	1		
At5g23820	expressed protein (ML domain - MD-2-related lipid recognition domain)	0,80				1
At2g16001	expressed protein (lipid recognition domain)	0,61				1
miscellaneous functions						
At2g28790	homologous to <i>Lycopersicon esculentum</i> osmotin	1,06	0,96	1		
At5g15230	gibberellin-regulated protein (GASA4)	0,47				1
At3g07130	homologous to <i>Glycine max</i> phytase and to purple acid phosphatase	0,25	0,11			1
At2g27190	homologous to purple acid phosphatase (PAP1)	0,60	0,24	1		
At5g34850	homologous to purple acid phosphatase	0,38				1
At4g29270	homologous to acid phosphatase	2,53	0,79	1		
At4g24340	homologous to phosphorylase	0,33				1
At3g02870	homologous to myo-inositol monophosphatase	0,34				1
At5g09440	homologous to <i>N. tabacum</i> phi-I	3,35	2,52	1		
At5g64260	homologous to <i>N. tabacum</i> phi-I	2,37	1,67	1		
At4g27110	homologous to COBRA (AtCOBL10)	0,08				1
At5g66590	homologous to <i>Nicotiana tabacum</i> pathogenesis-related protein PR1	7,86	5,2	1		
At1g09560	germin (subfamily 2, member 1, GLP5)		0,29			1
unknown function						
At3g56750	expressed protein		0,09		1	
At3g22000	expressed protein (DUF26)		0,15		1	
At1g26850	expressed protein (DUF248)		0,07			1
At1g80240	expressed protein (DUF642)	nd	nd	1		
At3g08030	expressed protein (DUF642)	0,73	0,76		1	
At4g32460	expressed protein (DUF642)	1,21	0,50	1		
At5g11420	expressed protein (DUF642)	20,05	24,30	1		
At5g25460	expressed protein (DUF642)	20,12	23,77	1		
At3g20370	expressed protein (MATH domain)	0,57				1
At2g04690	expressed protein (homologous to a human CREG protein)	1,07	1,11	1		
At1g78460	expressed protein (SOUL heme binding domain)		0,17			1
At2g34700	expressed protein (Ole e1 allergen domain)	1,51	1,00			1
At2g15220	expressed protein (Plant Basic Secreted Protein domain)	0,30	0,85			1
At2g28490	expressed protein (cupin domain)	1,23	0,09	1		
At3g22640	expressed protein (cupin domain)	0,12	0,09	1		
At4g36700	expressed protein (cupin domain)	1,96				1

## **SUPPLEMENTARY DATA CHAPTER 5**

## Additional data file 1

**Table 1. Cell Wall Genes (CWGs) with detectable levels of transcripts in 5- and 11-day-old hypocotyls.** CWGs were selected according to several papers and databases as mentioned at the bottom of the Table. The intensity of the signal is expressed as log<sub>2</sub>, the ratio between the levels of transcripts at 11 days and 5 days as well as the p-value are indicated. CWGs are classified by gene family and level of transcripts: below background in black, low level in blue (between background and 9), moderate level in green (between 9 and 10) and high level in red (higher than 10). Significant Bonferroni p-values are highlighted in green (p<5%)

Gene family	AGI number	Predicted or known gene function	CATMA probe	5 day-old hypocotyls	11 day-old hypocotyls	ratio 11 days / 5 days	p-value
prolyl-4-hydroxylases <sup>a</sup>	AT5G18900	AtP4H1	CATMA5A17270	9,67	9,60	-0,07	1,00
	AT2G43080		CATMA2A41480	8,77	8,77	-0,01	1,00
	AT4G33910		CATMA4A35710	8,70	7,96	-0,74	1,85E-02
	AT1G20270		CATMA1A19270	7,54	7,62	0,08	1,00
	AT3G28480		CATMA3A28370	7,23	7,07	-0,16	1,00
	AT4G25600		CATMA4A27290	6,87	7,00	0,13	1,00
nucleotide-sugar interconversion pathway <sup>b</sup>	AT5G39320	UGD1	CATMA5A34920	11,43	11,03	-0,40	1,00
	AT1G08200	AXS2	CATMA1A07190	10,34	9,98	-0,36	1,00
	AT5G15490	UGD3	CATMA5A13740	9,73	9,46	-0,27	1,00
	AT5G28840	GME1	CATMA5A26920	9,53	9,49	-0,04	1,00
	AT1G53500	RHM2	CATMA1A44530	9,11	8,83	-0,28	1,00
	AT1G17890	GER2	CATMA1A16900	9,00	8,83	-0,18	1,00
	AT4G30440	GAE1	CATMA4A32050	8,86	8,69	-0,17	1,00
	AT2G27860	AXS1	CATMA2A26260	8,75	8,68	-0,07	1,00
	AT5G66280	GMD1	CATMA5A61645	8,59	8,17	-0,42	1,00
	AT3G23820	GAE6	CATMA3A23780	8,54	8,51	-0,02	1,00
	AT3G46440	SUD1	CATMA3A39510	8,49	8,36	-0,13	1,00
	AT4G00110	GAE3	CATMA4A00110	7,89	7,81	-0,08	1,00
	AT1G73250	GER1	CATMA1A62530	7,82	8,02	0,20	1,00
	AT1G30620	UXE1 (MUR4)	CATMA1A28670	7,78	7,82	0,03	1,00
	AT2G47650	AUD2	CATMA2A46090	7,75	7,61	-0,13	1,00
	AT1G02000	GAE2	CATMA1A00990	7,70	7,58	-0,12	1,00
	AT1G26570	UGD4	CATMA1A24800	7,58	7,31	-0,27	1,00
	AT2G34850	UXE2	CATMA2A32970	7,51	7,49	-0,02	1,00
	AT2G45310	GAE4	CATMA2A43710	7,46	7,71	0,25	1,00
	AT1G12780	UGE1	CATMA1A11765	7,29	7,78	0,49	1,00
	AT1G64440	UGE4	CATMA1A53740	7,25	7,04	-0,21	1,00
	AT3G62830	AUD	CATMA3A55995	7,21	7,06	-0,15	1,00
	AT1G63000	UER1	CATMA1A52170	7,18	7,08	-0,10	1,00
	AT3G51160	GMD2 (MUR1)	CATMA3A44165	6,91	6,87	-0,05	1,00
glycosyl transferases (GTs)							
GT8 <sup>c</sup>	AT5G18480	Group E	CATMA5A16780	9,66	9,36	-0,30	1,00
	AT1G77130	Group A	CATMA1A66340	9,40	8,95	-0,45	1,00
	AT2G20810	Group D (GAUT10)	CATMA2A19380	9,11	8,64	-0,47	1,00
	AT3G62660	Group C (GATL7)	CATMA3A55830	8,70	8,52	-0,18	1,00
	AT3G02350	Group D (GAUT9)	CATMA3A01330	8,69	8,47	-0,22	1,00
	AT5G15470	Group D (GAUT14)	CATMA5A13720	8,26	8,01	-0,25	1,00
	AT1G18580	Group D (GAUT11)	CATMA1A17620	8,09	7,88	-0,21	1,00
	AT1G19300	Group C (GATL1=Parvus)	CATMA1A18330	8,06	7,93	-0,14	1,00
	AT3G01040	Group D (GAUT13)	CATMA3A00060	8,06	7,82	-0,24	1,00
	AT2G38650	Group D (GAUT7)	CATMA2A36930	8,03	8,02	-0,01	1,00
	AT3G61130	Group D (GAUT1)	CATMA3A54300	7,99	7,87	-0,12	1,00
	AT1G24170	Group C (GATL8)	CATMA1A23060	7,89	7,94	0,05	1,00

Table S1, supplementary data

	AT4G38270	Group D (GAUT3)	CATMA4A39780	7,75	7,92	0,17	1,00
	AT3G18660	Group A (PGS1P1)	CATMA3A18270	7,64	7,18	-0,46	1,00
	AT1G13250	Group C (GATL3)	CATMA1A12260	7,58	7,33	-0,25	1,00
	AT1G70090	Group C (GATL9)	CATMA1A59370	7,38	7,28	-0,10	1,00
	AT5G47780	Group D (GAUT4)	CATMA5A43750	7,37	7,49	0,12	1,00
	AT3G28340	Group C (GATL10)	CATMA3A28190	7,11	6,99	-0,13	1,00
	AT1G02720	Group C (GATL5)	CATMA1A01680	6,91	6,84	-0,07	1,00
	AT3G58790	Group D (GAUT15)	CATMA3A51790	6,78	6,72	-0,06	1,00
	AT2G35710	Group E	CATMA2A33890	6,77	6,80	0,03	1,00
	AT1G56600	Group B (AtGolS2)	CATMA1A47690	6,73	7,26	0,53	1,00
	AT5G30500	Group B	CATMA5A27960	6,68	6,77	0,10	1,00
GT31 <sup>d</sup>	AT4G00300	Group A	CATMA4A00350	9,68	8,84	-0,84	4,22E-04
	AT1G05170	Group E	CATMA1A04030	9,55	9,00	-0,55	1,00
	AT1G77810	Group E	CATMA1A66970	8,26	8,19	-0,07	1,00
	AT3G14960	Group C	CATMA3A14310	8,22	8,38	0,16	1,00
	AT2G32430	Group E	CATMA2A30720	8,19	7,96	-0,23	1,00
	AT3G11420	Group A	CATMA3A10410	7,86	8,03	0,17	1,00
	AT2G26100	Group C	CATMA2A24460	7,68	7,71	0,04	1,00
	AT5G41460	Group A	CATMA5A37060	7,00	6,72	-0,28	1,00
	AT1G33250	Group A	CATMA1A31530	6,95	7,11	0,16	1,00
	AT5G53340	Group D	CATMA5A49250	6,93	6,95	0,02	1,00
	AT1G53290	Group C	CATMA1A44310	6,88	6,86	-0,02	1,00
GT34 <sup>e</sup>	AT2G22900		CATMA2A21380	10,09	9,72	-0,36	1,00
	AT4G02500		CATMA4A02820	9,31	8,99	-0,31	1,00
	AT3G62720	XT1	CATMA3A55880	7,63	7,47	-0,17	1,00
	AT1G74380		CATMA1A63770	7,60	7,37	-0,23	1,00
	AT2G22900		CATMA2B21380	7,35	7,21	-0,14	1,00
	AT1G18690		CATMA1A17720	6,68	6,77	0,09	1,00
GT37 <sup>f</sup>	AT2G03220	FUT1 (MUR2)	CATMA2A02125	6,74	6,79	0,05	1,00
	AT1G14070	FUT7	CATMA1A13020	6,77	6,69	-0,08	1,00
GT47 <sup>g</sup>	AT2G28110	Group E (FRA8)	CATMA2A26510	11,39	10,73	-0,67	1,86E-01
	AT5G22940	Group E	CATMA5A20440	9,40	8,74	-0,66	2,06E-01
	AT5G61840	Group E (AtGUT2)	CATMA5A57440	8,58	8,34	-0,24	1,00
	AT4G16745	Group D	CATMA4A17710	8,23	7,85	-0,38	1,00
	AT2G20370	Group A (MUR3)	CATMA2A18850	8,06	7,78	-0,27	1,00
	AT4G22580	Group A (AtGT19)	CATMA4A24310	8,04	8,05	0,01	1,00
	AT1G27440	Group E (AtGUT1)	CATMA1A25670	7,83	7,82	0,00	1,00
	AT4G38040	Group D	CATMA4A39540	7,68	7,72	0,03	1,00
	AT3G57630	Group E	CATMA3A50640	7,61	7,28	-0,33	1,00
	AT5G62220	Group A (AtGT18)	CATMA5A57810	7,60	7,53	-0,06	1,00
	AT5G33290	Group C	CATMA5A28910	7,57	7,32	-0,26	1,00
	AT2G31990	Group A (AtGT15)	CATMA2A30260	7,56	7,55	-0,01	1,00
	AT1G21480	Group D	CATMA1A20530	7,43	7,38	-0,05	1,00
	AT2G35100	Group B (ARAD1)	CATMA2A33240	7,34	7,17	-0,17	1,00
	AT1G34270	Group B	CATMA1A32570	7,26	7,15	-0,11	1,00
	AT1G67410	Group B	CATMA1A56730	7,12	6,89	-0,23	1,00
	AT5G16890	Group B	CATMA5A15200	7,10	6,95	-0,15	1,00
	AT4G16745	Group D	CATMA4A17700	7,09	7,20	0,11	1,00
	AT1G74680	Group B	CATMA1A64050	7,09	6,87	-0,22	1,00
	AT5G33290	Group C	CATMA5A28900	6,98	6,93	-0,04	1,00
	AT5G44930	Group B	CATMA5A40770	6,91	7,11	0,20	1,00
	AT5G11610	Group D	CATMA5A10380	6,90	6,90	0,00	1,00

Table S1, supplementary data

	AT4G13990	Group A (AtGT14)	CATMA4A14190	6,88	6,83	-0,05	1,00
	AT5G03800	Group C	CATMA5A02990	6,84	6,80	-0,04	1,00
	AT5G19670	Group D	CATMA5A18100	6,80	6,83	0,03	1,00
	AT5G25310	Group C	CATMA5A23020	6,80	6,63	-0,17	1,00
	AT5G03800	Group C	CATMA5A03000	6,79	6,77	-0,02	1,00
GT77 <sup>h</sup>	AT1G19360		CATMA1A18380	9,97	9,52	-0,45	1,00
	AT2G35610		CATMA2A33740	7,60	7,41	-0,20	1,00
	AT2G35610		CATMA2A33750	7,33	7,17	-0,16	1,00
	AT5G44820		CATMA5A40660	7,23	7,01	-0,22	1,00
	AT1G75110	RRA2	CATMA1A64440	7,01	6,90	-0,12	1,00
	AT1G14590		CATMA1A13610	6,94	6,95	0,01	1,00
GT2 (cellulose synthases) <sup>i</sup>	AT5G05170	AtCESA3 (primary wall)	CATMA5A04375	10,36	9,89	-0,47	1,00
	AT5G09870	AtCESA5 (unknown function)	CATMA5A08637	10,01	9,11	-0,90	3,26E-05
	AT4G39350	AtCESA2 (unknown function)	CATMA4A40743	9,80	8,99	-0,82	9,50E-04
	AT4G32410	AtCESA1 (primary wall)	CATMA4A34150	9,80	9,53	-0,27	1,00
	AT2G25540	AtCesA10 (unknown function)	CATMA2A23880	7,63	7,25	-0,38	1,00
	AT5G44030	AtCESA4 (secondary wall)	CATMA5A39800	7,60	7,28	-0,31	1,00
	AT5G17420	AtCesA7 (secondary wall)	CATMA5A15680	6,94	6,87	-0,08	1,00
GT2 (cellulose synthases-like) <sup>i</sup>	AT5G03760	AtCSLA9	CATMA5A02940	8,94	8,47	-0,47	1,00
	AT3G28180	AtCSLC4	CATMA3A06563	8,93	8,66	-0,26	1,00
	AT5G22740	AtCSLA2	CATMA5A20200	8,63	8,45	-0,18	1,00
	AT2G35650	AtCSIA7	CATMA2A33860	8,60	8,21	-0,39	1,00
	AT3G07330	AtCSLC6	CATMA3A28010	8,57	9,03	0,46	1,00
	AT2G33100	AtCsID1	CATMA2A31270	7,32	6,99	-0,33	1,00
	AT1G55850	AtCSLE1	CATMA1A47044	7,19	7,92	0,73	1,98E-02
	AT1G23480	AtCSIA3	CATMA1A22512	7,04	6,99	-0,05	1,00
GT48 (callose synthases) <sup>j</sup>	AT1G05570	AtGSL6 (CAL51)	CATMA1A04500	9,17	8,14	-1,03	8,89E-08
	AT2G31960	AtGSL3	CATMA2A30240	9,15	8,71	-0,44	1,00
	AT5G13000	AtGSL12	CATMA5A11220	8,68	8,64	-0,04	1,00
	AT2G36850	AtGSL8	CATMA2A35110	7,90	7,92	0,02	1,00
	AT3G07160	AtGSL10	CATMA3A06350	7,87	7,77	-0,09	1,00
vesicle trafficking (emp24/gp25L/p24 family protein) <sup>k</sup>	AT1G69460		CATMA1A58750	10,17	10,55	0,37	1,00
	AT1G57620		CATMA1A47920	8,70	9,08	0,37	1,00
	AT3G07680		CATMA3A06940	8,26	8,69	0,43	1,00
	AT1G14010		CATMA1A12960	8,03	8,03	0,00	1,00
	AT1G21900		CATMA1A21000	7,69	7,75	0,06	1,00
	AT1G09580		CATMA1A08430	7,05	7,09	0,03	1,00
glycosyl hydrolases (GHs)							
GH9 (endoglucanases) <sup>l</sup>	AT4G02290		CATMA4A02560	11,10	10,08	-1,02	1,57E-07
	AT1G64390		CATMA1A53700	10,68	10,19	-0,49	1,00
	AT5G49720	AtKOR	CATMA5A45658	9,64	9,94	0,30	1,00
	AT1G70710	AtCEL1	CATMA1A59990	9,32	8,93	-0,39	1,00
	AT1G75680		CATMA1A64980	8,37	8,20	-0,17	1,00
	AT4G24260		CATMA4A25995	7,70	7,88	0,18	1,00
	AT1G71380		CATMA1A60700	7,70	7,68	-0,02	1,00
	AT2G32990		CATMA2A31180	7,35	7,23	-0,12	1,00
	AT4G39000		CATMA4A40450	6,68	6,83	0,15	1,00
GH16 (xyloglucan endotransglycosylases/hydrolases) (XTHs)	AT2G06850	At-XTH4	CATMA2A05540	12,98	12,15	-0,83	6,40E-04
	AT4G14130	At-XTH15	CATMA4A14375	12,72	11,98	-0,74	1,62E-02
	AT4G30290	At-XTH19	CATMA4A31900	11,37	11,40	0,03	1,00

Table S1, supplementary data

	AT4G30270	At-XTH24 (MERI-5)	CATMA4A31885	10,84	11,64	0,80	1,63E-03
	AT1G32170	At-XTH30	CATMA1A30510	10,06	9,85	-0,21	1,00
	AT2G01850	At-XTH27	CATMA2A00905	9,92	9,56	-0,36	1,00
	AT1G11545	At-XTH8	CATMA1A10510	9,18	9,03	-0,15	1,00
	AT1G14720	At-XTH28	CATMA1A13765	8,78	8,92	0,14	1,00
	AT4G37800	At-XTH7	CATMA4A39300	8,70	8,17	-0,52	1,00
	AT2G36870	At-XTH32	CATMA2A35140	8,31	7,86	-0,45	1,00
	AT5G13870	At-XTH5	CATMA5A12105	8,17	8,15	-0,02	1,00
	AT3G23730	At-XTH16	CATMA3A23670	7,79	7,81	0,02	1,00
	AT5G48070	At-XTH20	CATMA5A44050	7,60	7,56	-0,05	1,00
	AT4G25820	At-XTH14	CATMA4A27496	6,87	6,76	-0,11	1,00
	AT2G14620	At-XTH10	CATMA2A13330	6,73	6,78	0,05	1,00
	AT5G65730	At-XTH6	CATMA5A61080	6,69	6,78	0,10	1,00
GH17 <sup>n</sup>	AT5G42100	Group C	CATMA5A37820	9,93	9,88	-0,05	1,00
	AT3G04010	Group B	CATMA3A02980	9,85	8,30	-1,55	0,00E+00
	AT5G58090	Group B	CATMA5A53850	9,33	9,56	0,23	1,00
	AT3G07320	Group E	CATMA3A06560	8,55	8,29	-0,25	1,00
	AT5G42720	Group C	CATMA5A38510	8,54	8,49	-0,05	1,00
	AT4G31140	Group B	CATMA4A32830	8,47	8,22	-0,24	1,00
	AT2G26600	Group C	CATMA2A24920	8,41	8,09	-0,32	1,00
	AT2G01630	Group D	CATMA2A00680	8,40	8,37	-0,03	1,00
	AT5G20330	Group A (AtBG4)	CATMA5A18800	8,26	8,04	-0,22	1,00
	AT1G64760	Group B	CATMA1A54070	8,13	8,09	-0,04	1,00
	AT5G58480	Group B	CATMA5A54240	7,81	7,61	-0,20	1,00
	AT3G55430	Group C	CATMA3A48400	7,78	7,75	-0,03	1,00
	AT2G27500	Group C	CATMA2A25930	7,73	8,01	0,28	1,00
	AT4G29360	Group D	CATMA4A30990	7,64	7,60	-0,04	1,00
	AT5G20870	Group B	CATMA5A19400	7,36	7,20	-0,16	1,00
	AT5G56590	Group D	CATMA5A52400	7,29	7,03	-0,26	1,00
	AT5G18220	Group B	CATMA5A16490	7,25	7,14	-0,11	1,00
	AT1G30080	Group C	CATMA1A28100	7,08	7,02	-0,06	1,00
	AT1G66250	Group D	CATMA1A55520	6,99	6,89	-0,10	1,00
	AT4G17180	Group B	CATMA4A18200	6,92	6,88	-0,04	1,00
	AT5G20560	Group A	CATMA5A19060	6,87	6,87	-0,01	1,00
	AT1G32860	Group C	CATMA1A31160	6,84	7,01	0,17	1,00
	AT3G15800	Group C	CATMA3A15190	6,82	6,73	-0,09	1,00
GH18 (yieldins) <sup>o</sup>	AT5G24090		CATMA5A21620	7,22	7,30	0,09	1,00
	AT4G19730		CATMA4A20960	6,98	6,91	-0,08	1,00
	AT4G19810		CATMA4A21020	6,88	6,89	0,00	1,00
GH28 (polygalacturonases) <sup>p</sup>	AT4G23820	Group A	CATMA4A25640	9,36	8,91	-0,45	1,00
	AT3G06770	Group A	CATMA3A05990	9,01	8,78	-0,23	1,00
	AT5G41870	Group A	CATMA5A37590	8,64	8,35	-0,29	1,00
	AT3G55790	Group A	CATMA3A48760	8,13	8,58	0,45	1,00
	AT3G61490	Group A	CATMA3A54630	8,09	8,55	0,45	1,00
	AT3G62110	Group A	CATMA3A55250	7,93	7,77	-0,16	1,00
	AT1G80170	Group D	CATMA1A69350	7,64	7,64	0,00	1,00
	AT5G49215	Group A	CATMA5A45170	7,36	7,19	-0,16	1,00
	AT3G42950	Group A	CATMA3A35240	7,32	7,30	-0,02	1,00
	AT3G15720	Group B	CATMA3A15120	7,27	7,42	0,15	1,00
	AT3G16850	Group A	CATMA3A16240	7,08	6,95	-0,13	1,00
	AT4G32375	Group B	CATMA4A34110	6,97	6,84	-0,13	1,00

Table S1, supplementary data

	AT5G14650 AT4G23500 AT1G70500	Group C Group A Group D	CATMA5A12860 CATMA4A25330 CATMA1A59770	6,90 6,79 6,77	6,91 6,75 6,75	0,01 -0,04 -0,02	1,00 1,00 1,00
GH35 (beta-galactosidases) <sup>q</sup>	AT5G56870 AT1G45130 AT5G20710 AT5G63810 AT3G52840 AT2G28470 AT5G63800 AT1G72990	AtBGAL4 AtBGAL5 AtBGAL7 AtBGAL10 AtBGAL2 AtBGAL8 AtBGAL6 AtBGAL17	CATMA5A52625 CATMA1A37985 CATMA5A19240 CATMA5A59336 CATMA3A45773 CATMA2A26900 CATMA5A59333 CATMA1A62200	13,18 11,21 8,89 8,51 8,44 8,12 7,59 7,16	11,97 10,42 7,98 8,25 7,71 7,77 7,34 7,03	-1,22 -0,78 -0,92 -0,26 -0,73 -0,35 -0,25 -0,12	5,60E-12 3,24E-03 1,72E-05 1,00 2,32E-02 1,00 1,00 1,00
carbohydrate esterases (CEs)							
CE8 (pectin methylesterases) <sup>r</sup>	AT2G43050 AT3G49220 AT3G14310 AT1G53840 AT3G43270 AT1G53830 AT3G49220 AT4G33220 AT5G09760 AT3G10720 AT3G29090 AT5G64640 AT5G53370 AT5G64640 AT3G59010 AT4G02330 AT1G05310 AT5G47500 AT5G04970 AT3G05610 AT3G47400	AtPME3 AtPME1 AtPME2	CATMA2A41450 CATMA3A42250 CATMA3A13580 CATMA1A44933 CATMA3A35650 CATMA1A44930 CATMA3A42240 CATMA4A34970 CATMA5A08570 CATMA3A09760 CATMA3A29170 CATMA5A60070 CATMA5A49270 CATMA5A60070 CATMA3A52040 CATMA4A02630 CATMA1A04190 CATMA5A43480 CATMA5A04150 CATMA3A04630 CATMA3A40400	12,56 11,81 10,79 9,39 8,75 8,35 8,20 8,08 7,80 7,54 7,50 7,44 7,41 7,35 7,31 7,05 7,03 6,90 6,89 6,83 6,74	10,76 10,99 9,90 9,04 8,37 7,61 8,14 7,75 7,64 7,26 7,78 7,93 7,30 7,17 6,81 7,00 6,81 7,06 6,86 6,91 6,76	-1,80 -0,82 -0,90 -0,35 -0,38 -0,74 -0,06 -0,33 -0,16 -0,28 0,28 0,49 -0,11 -0,18 -0,50 -0,05 -0,22 0,17 -0,03 0,08 0,03	0,00E+00 8,65E-04 4,05E-05 1,00 1,00 1,61E-02 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00
CE13 (pectin acylesterases) <sup>s</sup>	AT3G05910 AT4G19410 AT2G46930 AT3G09410 AT3G62060 AT5G23870		CATMA3A04940 CATMA4A20590 CATMA2A45390 CATMA3A08300 CATMA3A55190 CATMA5A21340	10,44 8,25 7,97 7,25 7,15 6,94	10,14 8,09 8,07 7,31 7,01 6,87	-0,30 -0,17 0,10 0,06 -0,14 -0,06	1,00 1,00 1,00 1,00 1,00 1,00
carbohydrate lyases (CLs)							
PL1 (pectate lyases) <sup>t</sup>	AT5G48900 AT1G04680 AT1G67750 AT3G07010 AT4G24780 AT5G63180 AT3G55250 AT3G09540 AT5G04310 AT3G55140	Group A Group A Group A Group A Group A Group A Group B Group B Group C Group B	CATMA5A44810 CATMA1A03520 CATMA1A57110 CATMA3A06210 CATMA4A26470 CATMA5A58780 CATMA3A48250 CATMA3A08400 CATMA5A03510 CATMA3A48150	10,92 10,73 10,66 10,26 8,40 8,03 7,95 7,76 7,42 7,37	10,23 10,22 10,42 9,56 9,43 7,52 8,14 7,56 7,17 7,27	-0,70 -0,52 -0,24 -0,70 1,03 -0,52 0,18 -0,20 -0,26 -0,11	6,69E-02 1,00 1,00 6,16E-02 1,03E-07 1,00 1,00 1,00 1,00 1,00



Table S1, supplementary data

	AT3G24670	Group A	CATMA3A24620	7,37	7,22	-0,15	1,00
	AT4G13710	Group A	CATMA4A13890	7,29	7,03	-0,27	1,00
	AT5G09280	Group B	CATMA5A08090	7,14	7,37	0,23	1,00
	AT5G15110	Group B	CATMA5A13390	6,89	6,92	0,03	1,00
	AT1G30350	Group B	CATMA1A28360	6,80	6,78	-0,03	1,00
	AT1G14420	Group B	CATMA1A13420	6,77	6,73	-0,04	1,00
	AT3G27400	Group A	CATMA3A27260	6,59	6,94	0,35	1,00
PL4 (rhamnogalacturonan lyases) <sup>u</sup>	AT1G09910		CATMA1A08780	8,04	7,87	-0,17	1,00
	AT1G09890		CATMA1A08750	6,80	6,73	-0,07	1,00
expansins <sup>v</sup>	AT5G02260	AtEXPA9 (alpha expansin)	CATMA5A01340	12,71	12,44	-0,28	1,00
	AT1G69530	AtEXPA1 (alpha expansin)	CATMA1A58850	10,62	10,86	0,24	1,00
	AT2G37640	AtEXPA3 (alpha expansin)	CATMA2A35920	10,34	10,52	0,18	1,00
	AT2G40610	AtEXPA8 (alpha expansin)	CATMA2A38930	9,95	9,85	-0,10	1,00
	AT2G28950	AtEXPA6 (alpha expansin)	CATMA2B27340	9,26	9,80	0,54	1,00
	AT4G28250	AtEXPB3 (beta expansin)	CATMA4A29890	9,07	10,38	1,31	0,00E+00
	AT1G20190	AtEXPA11 (alpha expansin)	CATMA1A19190	8,76	8,72	-0,03	1,00
	AT2G03090	AtEXPA15 (alpha expansin)	CATMA2A02000	8,01	8,56	0,55	1,00
	AT4G38400	AtEXLA2 (expansin like family A)	CATMA4A39920	7,98	7,74	-0,24	1,00
	AT4G38210	AtEXPA20 (alpha expansin)	CATMA4A39720	7,91	8,42	0,51	1,00
	AT2G28950	AtEXPA6 (alpha expansin)	CATMA2A27340	7,88	8,09	0,21	1,00
	AT1G26770	AtEXPA10 (alpha expansin)	CATMA1A24995	7,87	7,97	0,10	1,00
	AT2G20750	AtEXPB1 (beta expansin)	CATMA2A19310	7,47	7,31	-0,16	1,00
	AT3G55500	AtEXPA16 (alpha expansin)	CATMA3A48450	7,43	7,30	-0,13	1,00
	AT5G05290	AtEXPA2 (alpha expansin)	CATMA5A04475	7,27	7,64	0,37	1,00
	AT5G56320	AtEXPA14 (alpha expansin)	CATMA5A52120	7,25	6,99	-0,26	1,00
	AT3G45960	AtEXLA3 (expansin like family A)	CATMA3A38990	6,97	6,91	-0,06	1,00
	AT3G29030	AtEXPA15 (alpha expansin)	CATMA3A29035	6,71	6,80	0,09	1,00
arabinogalactan proteins (AGPs) <sup>w</sup>	AT2G14890	AtAGP9	CATMA2A13650	11,71	11,82	0,12	1,00
	AT5G11740	AtAGP15 (AG peptide)	CATMA5A10540	11,03	11,71	0,68	1,10E-01
	AT5G10430	AtAGP4	CATMA5A09183	10,84	10,65	-0,19	1,00
	AT1G28290	AtAGP31	CATMA1A26460	10,97	11,42	0,46	1,00
	AT3G13520	AtAGP12 (AG peptide)	CATMA3A12660	10,77	11,51	0,74	1,50E-02
	AT5G53250	AtAGP22 (AG peptide)	CATMA5A49170	9,48	9,98	0,50	1,00
	AT3G61640	AtAGP20 (AG peptide)	CATMA3A54770	8,69	9,53	0,85	3,14E-04
	AT2G47930	AtAGP26	CATMA2A46360	8,46	9,10	0,63	5,19E-01
	AT5G18690	AtAGP25	CATMA5A17040	7,90	8,07	0,17	1,00
	AT5G64310	AtAGP1	CATMA5A59763	7,85	8,47	0,62	7,55E-01
	AT4G09030	AtAGP10	CATMA4A09010	7,75	8,34	0,58	1,00
	AT4G37450	AtAGP18 (Lys-rich AGP)	CATMA4A39020	7,68	7,73	0,04	1,00
	AT1G68725	AtAGP19 (Lys-rich AGP)	CATMA1A58110	7,64	7,57	-0,08	1,00
fasciclin AGPs (FLAs) <sup>x</sup>	AT1G03870	AtFLA9	CATMA1A02720	12,49	12,25	-0,24	1,00
	AT2G45470	AtFLA8	CATMA2A43870	11,96	11,53	-0,43	1,00
	AT4G12730	AtFLA2	CATMA4A12870	10,85	10,44	-0,41	1,00
	AT2G04780	AtFLA7	CATMA2A03690	9,27	9,69	0,42	1,00
	AT3G12660	AtFLA14	CATMA3A11630	8,54	8,09	-0,46	1,00
	AT5G55730	AtFLA1	CATMA5A51470	8,18	7,96	-0,22	1,00
	AT3G11700	AtFLA18	CATMA3A10640	8,01	8,28	0,27	1,00
	AT5G44130	AtFLA13	CATMA5A39880	7,76	7,48	-0,28	1,00
	AT5G03170	AtFLA11	CATMA5A02280	7,75	7,66	-0,09	1,00

Table S1, supplementary data

	AT5G60490 AT3G60900 AT2G24450	AtFLA12 AtFLA10 AtFLA3	CATMA5A56200 CATMA3A53870 CATMA2A22780	7,57 7,28 7,19	7,42 7,12 7,12	-0,15 -0,16 -0,07	1,00 1,00 1,00
GPI-anchored peptide (GAPEP) family <sup>y</sup>	AT3G01950 AT5G40970 AT5G40960 AT3G01940 AT5G50660		CATMA3A00940 CATMA5A36640 CATMA5A36630 CATMA3A00930 CATMA5A46550	7,46 7,29 6,96 6,85 6,83	7,94 7,74 8,07 7,15 6,84	0,48 0,44 1,11 0,31 0,01	1,00 1,00 1,58E-09 1,00 1,00
COBRA-like proteins <sup>z</sup>	AT5G15630 AT4G16120 AT3G02210 AT3G29810 AT5G60920 AT5G49270	AtCOBL4 AtCOBL7 AtCOBL1 AtCOBL2 COB COBL9	CATMA5A13880 CATMA4A16865 CATMA3A01200 CATMA3A30250 CATMA5CTRL29 CATMA5A45240	7,65 7,51 7,44 7,41 6,83 6,78	7,23 7,38 7,50 7,53 6,79 6,76	-0,43 -0,12 0,06 0,11 -0,04 -0,02	1,00 1,00 1,00 1,00 1,00 1,00
leucine-rich repeat extensins (LRXs) <sup>aa</sup>	AT3G24480 AT2G15890 AT4G33980 AT1G62440 AT4G13340 AT3G22800 AT4G18670	AtLRX4 AtPEX3 AtPEX4 AtLRX2 AtLRX3 AtLRX6 AtLRX5	CATMA3A24360 CATMA2A14770 CATMA4A35800 CATMA1A51580 CATMA4A13500 CATMA3A22740 CATMA4A19790	9,68 9,07 8,52 7,57 6,89 6,78 6,70	9,56 9,38 8,89 7,44 6,83 6,84 6,79	-0,11 0,31 0,37 -0,12 -0,06 0,06 0,09	1,00 1,00 1,00 1,00 1,00 1,00 1,00
Hyp/Pro-rich proteins (H/PRP) <sup>ab</sup>	AT2G21140 AT4G38770	AtPRP2 AtPRP4	CATMA2A19790 CATMA4A40195	11,67 8,94	11,15 8,78	-0,51 -0,16	1,00 1,00
extensins <sup>ac</sup>	AT1G23720 AT3G28550 AT3G54580		CATMA1A22610 CATMA3A28430 CATMA3A47520	7,01 6,91 7,01	6,92 6,96 6,94	-0,08 0,05 -0,07	1,00 1,00 1,00
lignin toolbox <sup>ad</sup>	AT4G39330 AT2G30490 AT1G76790 AT5G48930 AT2G40890 AT1G20480 AT5G54160 AT5G58490 AT1G15950 AT3G21240 AT3G10340 AT1G72680 AT3G19450 AT3G61990	Arath;CAD1(cinnamyl alcohol dehydrogenase) Arath;C4H (CYP73A5) (trans-cinnamate 4-hydroxylase) Arath;COMT-like8 (caffeic acid O-methyltransferase) Arath;HCT (hydroxycinnamoyl-CoA:shikimate/quinic acid hydroxycinnamoyl transferase) Arath;C3H2 (CYP98A3) (p-coumarate 3-hydroxylase) Arath;4CL-like4 (4-coumarate CoA ligase) Arath;COMT (caffeic acid O-methyltransferase) Arath;CCR-like5 (cinnamoyl-CoA reductase) Arath;CCR1 (cinnamoyl-CoA reductase) Arath;4CL2 (4-coumarate CoA ligase) Arath;PAL4 (Phe ammonia lyase) Arath;CAD9 (cinnamyl alcohol dehydrogenase) Arath;CAD2 (cinnamyl alcohol dehydrogenase) Arath;CCoAOMT3 (caffeoyl-CoA O-methyltransferase)	CATMA4A40735 CATMA2A28746 CATMA1A66010 CATMA5A44840 CATMA2A39220 CATMA1A19480 CATMA5A50073 CATMA5A54250 CATMA1A14998 CATMA3A21086 CATMA3A09340 CATMA1A61900 CATMA3A19045 CATMA3A55120	11,03 10,84 10,37 10,24 9,83 9,57 9,55 8,96 8,74 8,62 8,46 8,40 8,40 8,27	10,33 10,43 9,48 10,28 9,16 9,02 9,02 9,14 8,79 8,40 8,52 8,53 7,96 8,06	-0,70 -0,41 -0,89 0,04 -0,67 -0,55 -0,53 0,18 0,05 -0,22 0,06 0,13 -0,44 -0,21	7,14E-02 1,00 4,45E-05 1,00 1,45E-01 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00 1,00

Table S1, supplementary data

	AT2G02400	Arath;CCR-like2 (cinnamoyl-CoA reductase)	CATMA2A01315	8,09	8,02	-0,08	1,00
	AT5G04230	Arath;PAL3 (Phe ammonia lyase)	CATMA5A03420	8,06	7,83	-0,23	1,00
	AT4G34230	Arath;CAD6 (cinnamyl alcohol dehydrogenase)	CATMA4A36050	7,83	7,53	-0,30	1,00
	AT1G20510	Arath;4CL-like1 (4-coumarate CoA ligase)	CATMA1A19510	7,78	7,40	-0,38	1,00
	AT2G21890	Arath;CAD8 (cinnamyl alcohol dehydrogenase)	CATMA2A20490	7,66	7,41	-0,25	1,00
	AT1G51680	Arath;4CL1 (4-coumarate CoA ligase)	CATMA1A42785	7,64	8,15	0,51	1,00
	AT5G63380	Arath;4CL-like8 (4-coumarate CoA ligase)	CATMA5A58930	7,60	7,42	-0,18	1,00
	AT2G21730	Arath;CAD7 (cinnamyl alcohol dehydrogenase)	CATMA2A20370	7,44	7,31	-0,13	1,00
	AT4G26220	Arath;CCoAOMT7 (caffeoyl-CoA O-methyltransferase)	CATMA4A27740	7,23	7,48	0,24	1,00
	AT5G04330	Arath;F5H2 (CYP84A4)	CATMA5A03540	7,15	6,89	-0,27	1,00
	AT2G37040	Arath;PAL1 (Phe ammonia lyase)	CATMA2A35330	7,01	7,94	0,93	8,18E-06
	AT3G21230	Arath;4CL4 (4-coumarate CoA ligase)	CATMA3A21070	6,99	6,91	-0,07	1,00
	AT1G80820	Arath;CCR2 (cinnamoyl-CoA reductase)	CATMA1A70050	6,90	6,81	-0,09	1,00
	AT4G19010	Arath;4CL-like6 (4-coumarate CoA ligase)	CATMA4A20160	6,89	6,82	-0,07	1,00
	AT5G53810	Arath;COMT-like13 (caffeic acid O-methyltransferase)	CATMA5A49700	6,82	6,72	-0,10	1,00
	AT4G36220	Arath;F5H1 (CYP84A1)	CATMA4A37873	6,68	6,84	0,15	1,00
peroxidases <sup>ae</sup>	AT2G37130	AtPrx21	CATMA2A35430	14,28	12,86	-1,42	0,00
	AT4G21960	AtPrx42	CATMA4A23655	13,17	12,40	-0,77	5,29E-03
	AT1G05260	AtPrx03	CATMA1A04113	10,68	10,18	-0,51	1,00
	AT4G33870	AtPrx48	CATMA5A59590	10,52	9,25	-1,27	0,00
	AT1G71695	AtPrx12	CATMA1A60876	10,17	9,54	-0,63	4,86E-01
	AT5G64120	AtPrx71	CATMA4A35650	10,03	9,85	-0,18	1,00
	AT3G21770	AtPrx30	CATMA3B21635	9,54	9,77	0,23	1,00
	AT3G21770	AtPrx30	CATMA3A21635	9,29	9,52	0,23	1,00
	AT5G51890	AtPrx66	CATMA5A47820	8,92	9,12	0,21	1,00
	AT5G66390	AtPrx72	CATMA5A61725	8,70	8,57	-0,13	1,00
	AT4G37520	AtPrx50	CATMA4A39085	8,28	8,32	0,04	1,00
	AT5G64110	AtPrx70	CATMA5A59580	8,24	8,37	0,13	1,00
	AT5G39580	AtPrx62	CATMA5A35160	7,83	7,95	0,11	1,00
	AT2G38390	AtPrx23	CATMA2A36690	7,80	8,96	1,15	1,96E-10
	AT3G17070	AtPrx29	CATMA3A16470	7,73	7,62	-0,11	1,00
	AT1G44970	AtPrx09	CATMA1A37915	7,72	7,56	-0,16	1,00
	AT2G41480	AtPrx25	CATMA2A39890	7,65	8,09	0,44	1,00
	AT4G30170	AtPrx45	CATMA4A31806	7,52	7,65	0,13	1,00
	AT2G22420	AtPrx17	CATMA2A20940	7,34	7,48	0,14	1,00
	AT4G37530	AtPrx51	CATMA4A39090	7,24	7,29	0,05	1,00
	AT5G05340	AtPrx52	CATMA5A04520	7,04	7,24	0,20	1,00
	AT5G19890	AtPrx59	CATMA5A18330	7,13	7,20	0,08	1,00
	AT5G40150	AtPrx63	CATMA5A35810	7,32	7,19	-0,13	1,00
	AT5G19880	AtPrx58	CATMA5A18320	6,98	6,99	0,01	1,00
	AT2G24800	AtPrx18	CATMA2A23170	7,03	6,95	-0,08	1,00
	AT2G18980	AtPrx16	CATMA2A17560	7,11	6,93	-0,18	1,00
	At4g36430	AtPrx49	CATMA4A38005	6,77	6,90	0,13	1,00
	AT2G18150	AtPrx15	CATMA2A16820	6,63	6,88	0,25	1,00
	AT5G22410	AtPrx60	CATMA5A19865	6,86	6,81	-0,05	1,00
	AT5G17820	AtPrx57	CATMA5A16105	6,99	6,76	-0,23	1,00
	AT3G01190	AtPrx27	CATMA3A00185	6,78	6,73	-0,05	1,00

Table S1, supplementary data

laccases <sup>af</sup>	AT5G60020	AtLAC17	CATMA5A55770	9,53	9,23	-0,30	1,00
	AT5G03260	AtLAC11	CATMA5A02370	9,23	9,39	0,17	1,00
	AT2G38080	AtLAC4	CATMA2A36370	8,91	8,33	-0,58	1,00
	AT2G30210	AtLAC3	CATMA2A28490	8,23	8,14	-0,09	1,00
	AT2G29130	AtLAC2	CATMA2A27560	7,29	7,01	-0,28	1,00
	AT5G05390	AtLAC12	CATMA5A04570	7,01	7,04	0,03	1,00
	AT2G40370	AtLAC5	CATMA2A38650	6,93	6,86	-0,07	1,00
	AT5G01190	AtLAC10	CATMA5A00220	6,88	6,77	-0,12	1,00
	AT5G48100	AtLAC15	CATMA5A44080	6,77	6,68	-0,09	1,00
SKU-like proteins (multi-copper oxidases) <sup>ag</sup>	AT1G76160	SKS5	CATMA1A65380	10,33	9,91	-0,43	1,00
	AT1G41830	SKS6	CATMA1A35650	9,76	9,27	-0,49	1,00
	AT4G12420	SKU5	CATMA4A12530	9,28	8,97	-0,31	1,00
	AT5G48450	SKS3	CATMA5A44430	9,11	8,93	-0,18	1,00
	AT4G22010	SKS4	CATMA4A23690	8,47	8,11	-0,36	1,00
	AT3G13990	SKS11	CATMA3A13220	7,66	7,51	-0,15	1,00
	AT1G41830	SKS6	CATMA1A35660	7,11	6,96	-0,14	1,00
	AT4G28090	SKS10	CATMA4A29740	6,82	6,72	-0,09	1,00
phytycyanins <sup>ah</sup>	AT4G12880	AtEN20 (early nodulin)	CATMA4A13020	10,84	11,19	0,36	1,00
	AT2G32300	AtUCC1 (uclacyanin)	CATMA2A30615	10,61	10,60	-0,02	1,00
	AT5G15350	AtEN22 (early nodulin)	CATMA5A13625	9,07	9,35	0,28	1,00
	AT1G08500	AtEN21 (early nodulin)	CATMA1A07465	8,78	9,08	0,30	1,00
	AT5G53870	AtEN1 (early nodulin)	CATMA5A49760	8,24	8,22	-0,02	1,00
	AT1G72230	AtUCC8 (uclacyanin)	CATMA1A61460	8,10	8,34	0,24	1,00
	AT4G28360	AtEN2 (early nodulin)	CATMA4A30000	8,07	8,52	0,46	1,00
	AT3G20570	AtEN14 (early nodulin)	CATMA3A20300	7,83	8,16	0,33	1,00
	AT3G60280	AtUCC3 (uclacyanin)	CATMA3A53305	7,78	7,81	0,03	1,00
	AT4G31840	AtEN13 (early nodulin)	CATMA4A33490	7,53	8,19	0,66	2,47E-01
	AT2G27035	AtEN17 (early nodulin)	CATMA2A25400	7,33	7,57	0,24	1,00
	AT4G27520	AtEN12 (early nodulin)	CATMA4A29090	7,33	7,49	0,16	1,00
	AT5G26330	AtSTC3 (stellacyanin)	CATMA5A24030	7,23	7,32	0,09	1,00
	AT3G27200	AtUCC6 (uclacyanin)	CATMA3A27010	7,20	7,63	0,43	1,00
	AT5G25090	AtEN6 (early nodulin)	CATMA5A22800	6,98	7,12	0,15	1,00
	AT2G25060	AtEN7 (early nodulin)	CATMA2A23390	7,06	7,40	0,34	1,00
	AT4G33930	AtEN26 (early nodulin)	CATMA4A35740	6,91	6,87	-0,03	1,00
	AT2G02850	AtPNC (plantacyanin)	CATMA2A01770	6,84	6,90	0,06	1,00
	AT5G07475	AtUCC5 (uclacyanin)	CATMA5A06680	6,81	6,84	0,03	1,00
	AT5G20230	AtSTC1 (stellacyanin)	CATMA5A18670	6,79	7,42	0,62	0,67
	AT5G57920	AtEN3 (early nodulin)	CATMA5A53680	6,69	6,82	0,13	1,00
subtilases <sup>ai</sup>	AT3G14067	AtSBT1.4	CATMA3A13290	11,85	10,71	-1,15	2,80E-10
	AT2G05920	AtSBT1.8	CATMA2A04690	11,54	10,90	-0,64	4,50E-01
	AT5G67360	AtSBT1.7 (ARA12)	CATMA5A62810	11,45	11,12	-0,33	1,00
	AT5G51750	AtSBT1.3	CATMA5A47680	11,04	10,80	-0,24	1,00
	AT5G59090	AtSBT4.12	CATMA5A54840	10,47	9,49	-0,99	7,99E-07
	AT3G14240	AtSBT1.5	CATMA3A13520	9,92	9,76	-0,16	1,00
	AT1G20160	AtSBT5.2	CATMA1A19170	8,84	8,90	0,06	1,00
	AT4G34980	AtSBT1.6	CATMA4A36760	8,49	8,32	-0,17	1,00
	AT4G00230	AtSBT4.14 (XSP11)	CATMA4A00245	8,20	8,20	0,00	1,00
	AT2G04160	AtSBT5.3 (AIR3)	CATMA2A03130	7,94	7,37	-0,57	1,00
	AT1G30600	AtSBT2.1	CATMA1A28650	7,35	7,20	-0,15	1,00
	AT5G19660	AtSBT6.1	CATMA5A18090	7,27	7,12	-0,15	1,00
	AT5G44530	AtSBT2.3	CATMA5A40320	7,15	6,88	-0,27	1,00
	AT5G45650	AtSBT5.6	CATMA5A41630	7,03	6,83	-0,20	1,00

Table S1, supplementary data

	AT2G39850	AtSBT4.2	CATMA2A38090	6,73	6,82	0,09	1,00
	AT1G32950	AtSBT3.4	CATMA1A31250	6,79	6,80	0,01	1,00
	AT5G59190	AtSBT4.3	CATMA5A54940	6,76	6,79	0,03	1,00
	AT5G58840	AtSBT4.9	CATMA5A54600	6,78	6,77	0,00	1,00
	AT5G67090	AtSBT1.9	CATMA5A62540	6,81	6,77	-0,04	1,00
	AT5G03620	AtSBT4.15	CATMA5A02800	6,79	6,74	-0,06	1,00
	AT2G19170	AtSBT2.5	CATMA2A17780	6,89	6,72	-0,18	1,00
	AT5G59100	AtSBT4.4	CATMA5A54850	6,77	6,65	-0,11	1,00
a) (PMEIs)	AT2G31430		CATMA2A29670	8,79	8,76	-0,03	1,00
	AT3G12880		CATMA3A11840	8,78	8,51	-0,27	1,00
	AT4G25260		CATMA4A26950	8,67	9,10	0,43	1,00
	AT1G62770		CATMA1A51910	8,22	8,37	0,15	1,00
	AT5G64620		CATMA5A60045	8,20	8,28	0,08	1,00
	AT5G62350		CATMA5A57950	8,09	8,47	0,38	1,00
	AT3G47670		CATMA3A40660	7,72	7,72	0,00	1,00
	AT3G47380		CATMA3A40380	7,53	7,80	0,27	1,00
	AT2G01610		CATMA2A00660	7,52	7,88	0,36	1,00
	AT4G12390		CATMA4A12500	7,44	7,34	-0,10	1,00
	AT1G47960		CATMA1A39055	7,19	7,83	0,65	3,44E-01
	AT3G17220		CATMA3A16630	7,18	7,37	0,19	1,00
	AT3G49330		CATMA3A42360	7,05	7,10	0,05	1,00
	AT5G20740		CATMA5A19270	7,02	6,90	-0,12	1,00
	AT1G70720		CATMA1A60000	6,97	6,91	-0,06	1,00
	AT1G23205		CATMA1A22290	6,95	7,17	0,22	1,00
	AT5G24370		CATMA5A21980	6,94	6,99	0,05	1,00
	AT1G14890		CATMA1A13920	6,94	7,35	0,41	1,00
	AT1G48010		CATMA1A39095	6,83	6,81	-0,02	1,00
	AT3G62820		CATMA3A55990	6,77	6,65	-0,12	1,00
	AT3G55680		CATMA3A48660	6,75	6,80	0,05	1,00

a: <http://cellwall.genomics.purdue.edu/families/4-6-5.html>  
b: <http://cellwall.genomics.purdue.edu/families/1-1.html>  
c: <http://cellwall.genomics.purdue.edu/families/2-3-1.html>  
d: <http://cellwall.genomics.purdue.edu/families/2-3-5.html>  
e: <http://cellwall.genomics.purdue.edu/families/2-3-4.html>  
f: <http://cellwall.genomics.purdue.edu/families/1-1.html>  
g: <http://cellwall.genomics.purdue.edu/families/2-3-2.html>  
h: <http://www.cazy.org/fam/GT77.html>, [47]  
i: <http://cellwall.genomics.purdue.edu/families/2-2.html>  
j: <http://cellwall.genomics.purdue.edu/families/2-4.html>  
k: <http://cellwall.genomics.purdue.edu/families/3-1.html>  
l: <http://cellwall.genomics.purdue.edu/families/4-3-2-1.html>  
m: <http://labs.plantbio.cornell.edu/xth/genes.htm>  
n: <http://cellwall.genomics.purdue.edu/families/4-3-2-2.html>  
o: <http://cellwall.genomics.purdue.edu/families/4-1-2.html>  
p: <http://cellwall.genomics.purdue.edu/families/4-3-3.html>  
q: <http://cellwall.genomics.purdue.edu/families/4-3-1-1.html>  
r: <http://cellwall.genomics.purdue.edu/families/4-5-1.html>

s: <http://cellwall.genomics.purdue.edu/families/4-5-2.html>  
t: <http://cellwall.genomics.purdue.edu/families/4-4-1.html>  
u: <http://cellwall.genomics.purdue.edu/families/4-4-2.html>  
v: <http://www.bio.psu.edu/expansins/arabidopsis.htm#At-EXLB1>  
w: [48-51]  
x: [48]  
y: <http://cellwall.genomics.purdue.edu/families/6-4-10.html>  
z: [14]  
aa: [52]  
ab: [53]  
ac: our own annotation  
ad: [54]  
ae: <http://peroxidase.isb-sib.ch/listing.php?action=view&id=4362>  
af: [56, 57]  
ag: [58]  
ah: [59]  
ai: <http://csbdb.mpiimp-golm.mpg.de/csdbw/psdb/pub/sgenes.html>  
aj: Pr Jérôme Pelloux (Amiens University, France), personal communication

## Additional data file 2

**Table 2. Secretory Pathway Genes (SPGs) with moderate or high levels of transcripts in 5- and 11-day-old hypocotyls.** SPGs were annotated using bioinformatic softwares for sub-cellular localization and functional domains as described in Methods. The intensity of the signal is expressed as log the ratio between the levels of transcripts at 11-days and 5-days as well as the p-value are indicated. CWGs are classified by gene family and level of transcript: low level in blue (between background and 9), moderate level in green (between 9 and 10) and high level in red (higher than 10). Significant Bonferroni p-values are highlighted in green (p<5%).

Genes encoding proteins predicted to be outside the cell are highlighted in yellow, proteins predicted to secreted and to have trans-membrane domains are in pink, and proteins predicted to be secreted and to have a GPI anchor are in dark yellow. Genes encoding proteins identified in the proteomic study (Ishadet *et al.*, submitted for publication) performed with the same plant material are in bold characters.

Functional class	Gene family	AGI number	gene name	CATMA probe	5-days	11-days	ratio 11-days / 5-days	p-value
Proteins acting on carbohydrates	glycoside hydrolase family 1 (beta-glucosidase)	AT3G60130		CATMA3A53140	10.84	9.70	-1.14	3.41E-10
		AT3G18080		CATMA3A17550	10.40	10.53	0.13	1.00
		AT3G21360		CATMA3A21250	9.96	11.47	1.51	0.00E+00
		AT5G26000		CATMA5A23655	9.34	9.08	-0.25	1.00
		AT1G02850		CATMA1A01780	8.98	9.36	0.38	1.00
	glycoside hydrolase family 3	AT1G52400		CATMA1A43455	7.36	9.04	1.68	0.00E+00
		AT5G49360		CATMA5A45350	12.67	11.94	-0.73	2.20E-02
		AT5G10560		CATMA5A09290	10.88	10.42	-0.46	1.00
		AT5G20950		CATMA5A19500	9.46	9.74	0.28	1.00
		AT1G64390		CATMA1A53700	10.68	10.19	-0.49	1.00
	glycoside hydrolase family 9	AT4G02990		CATMA4A02570	11.10	10.08	-1.02	1.57E-07
		AT5G49720	AIKOR	CATMA5A45658	9.64	9.94	0.30	1.00
		AT1G70710	AIKEL1	CATMA1A59990	9.32	8.93	-0.39	1.00
	glycoside hydrolase family 16 (xyloglucan endotransglycosidases/hydrolases)	<b>AT2G06850</b>	<b>AXTH4</b>	CATMA2A05540	12.98	12.15	-0.83	6.40E-04
		AT1G11545	AXTH8	CATMA1A10510	9.18	9.03	-0.15	1.00
		AT4G14130	AXTH15	CATMA4A14375	12.72	11.98	-0.74	1.62E-02
		AT4G30290	AXTH19	CATMA4A31900	11.37	11.40	0.03	1.00
		AT4G30270	AXTH24	CATMA4A31885	10.84	11.64	0.80	1.63E-04
		AT2G01850	AXTH27	CATMA2A00905	9.92	9.56	-0.36	1.00
		AT1G32170	AXTH30	CATMA1A30510	10.06	9.85	-0.21	1.00
		AT5G42100		CATMA5A37820	9.93	9.88	-0.05	1.00
		AT5G58980		CATMA5A33850	9.33	9.56	0.23	1.00
		<b>AT3G04010</b>		<b>CATMA202980</b>	<b>8.30</b>	<b>9.95</b>	<b>-1.55</b>	<b>0.00E+00</b>
	glycoside hydrolase family 19 (chitinase class I)	AT1G05850	ELP: ECTOPIC DEPOSITION OF LIGNIN IN PITH	CATMA1A04870	12.52	11.97	-0.56	1.00
		AT2G43590	glycoside hydrolase family 19	CATMA2B42010	9.36	9.40	0.04	1.00
	glycoside hydrolase family 20	<b>AT5G52950</b>		CATMA1A6320	9.64	9.13	-0.51	1.00
	glycoside hydrolase family 27	AT5G03830		CATMA5A07660	8.97	9.42	0.44	1.00
	glycoside hydrolase family 31	<b>AT1G68560</b>	<b>XYL1 (alpha-xylosidase)</b>	CATMA1A57930	<b>10.6</b>	<b>10.13</b>	<b>-0.47</b>	<b>1.00</b>
	glycoside hydrolase family 35 (beta-galactosidase)	AT5G56870	AI8GAL4	CATMA5A52625	13.18	11.97	-1.22	5.60E-12
		AT1G45130	AI8GAL5	CATMA1A37985	11.21	10.42	-0.78	3.24E-03
	glycoside hydrolase family 38 (alpha-mannosidase)	AT3G26720	glycoside hydrolase family 38 (alpha-mannosidase)	CATMA3A26465	11.33	10.38	-0.95	4.81E-06
	glycoside hydrolase family 47 pectate lyase family 1	AT1G27520		CATMA1A25750	9.15	9.08	-0.07	1.00
		AT1G67750		CATMA1A57110	10.66	10.42	-0.24	1.00
		AT5G48900		CATMA5A44810	10.92	10.23	-0.70	6.69E-02
		AT1G04680		CATMA1A03520	10.73	10.22	-0.52	1.00
		AT3G07010		CATMA3A06210	10.26	9.56	-0.70	6.2E-02
		AT4G24780		CATMA4A26470	8.40	9.43	1.03	1.03E-07
	carbohydrate esterase family 8 (pectin methylsterases)	AT3G49220		CATMA3A42250	11.81	10.99	-0.82	8.65E-04
		AT2G43050		CATMA2A41450	12.56	10.76	-1.80	0.00E+00
		<b>AT3G14310</b>	<b>AI2PME3</b>	CATMA3A13580	10.79	9.90	-0.90	4.05E-05
	carbohydrate esterase family 13 (pectin acylesterases)	AT1G53840		CATMA1A44933	9.49	9.04	-0.45	1.00
		AT3G05910		CATMA3A04940	10.44	10.14	-0.30	1.00
	glycosyl transferase family 2 (cellulose synthases)	<b>AT5G57655</b>	<b>homologous to xylose isomerase</b>	CATMA5A53360	<b>10.55</b>	<b>10.19</b>	<b>-0.36</b>	<b>1.00</b>
		AT5G05170	AICESA3	CATMA5A04375	10.36	9.89	-0.47	1.00
		AT5G08870	AICESA5	CATMA5A08637	10.01	9.11	-0.90	3.26E-05
		AT4G39350	AICESA2	CATMA4A40743	9.80	8.99	-0.82	9.50E-04
		AT4G32410	AICESA1	CATMA4A34150	9.53	9.27	-0.27	1.00
		AT1G05570	AICALS1	CATMA1A04500	9.17	8.14	-1.03	8.89E-08
	expansins	AT2G31960	AI8GL3	CATMA2A30240	9.15	8.71	-0.44	1.00
		AT1G69530	AI8KPA1	CATMA1A15865	10.52	10.86	0.24	1.00
		AT2G37640	AIEXPA3	CATMA2A39820	10.34	10.06	-0.18	1.00
		AT2G28950	AIEXPA6	CATMA2B27340	9.26	9.80	0.54	1.00
		AT2G40610	AIEXPA8	CATMA2A38930	9.95	9.85	-0.10	1.00
		<b>AT5G02260</b>	<b>AIEXPA9</b>	CATMA5A01340	<b>12.71</b>	<b>12.44</b>	<b>-0.28</b>	<b>1.00</b>
		AT4G28250	AIEXPB3	CATMA4A29890	9.07	10.38	1.31	0.00
Proteases	cysteine proteases (papain family)	AT5G06360		CATMA5A56105	12.54	11.90	-0.65	3.43E-01
		AT3G54940		CATMA3A47930	8.35	10.33	1.98	0.00E+00
		AT1G62290		CATMA1A51380	11.77	12.48	0.72	3.29E-02
		<b>AT4G01610</b>		CATMA4A01830	<b>12.40</b>	<b>12.38</b>	<b>-0.02</b>	<b>1.00</b>
		AT1G20850	XCP2 (XYLEM CYSTEINE PEPTIDASE 2)	CATMA1A19916	9.33	9.19	-0.14	1.00
	serine proteases (proteins homologous to serine proteases)	AT5G67360	AI8BT1.7, ARA12	CATMA5A62810	11.45	11.12	-0.33	1.00
		AT5G51750	AI8BT1.3	CATMA5A47680	11.04	10.80	-0.24	1.00
		AT3G14067	AI8BT1.4	CATMA3A13290	11.85	10.71	-1.15	2.80E-10
		AT3G14240	AI8BT1.5	CATMA3A13520	9.92	9.76	-0.16	1.00
		AT2G05920	AI8BT1.8	CATMA2A04690	11.54	10.90	-0.64	4.50E-01
	aspartyl proteases	AT5G59290	AI8BT4.12	CATMA5A54840	10.47	9.49	-0.99	7.99E-07
		AT1G66180		CATMA1A55440	10.98	10.28	-0.69	7.75E-02
		<b>AT1G11910</b>		CATMA1A10930	<b>12.95</b>	<b>12.35</b>	<b>-0.60</b>	<b>1.00</b>
		<b>AT5G10770</b>		CATMA5A09510	<b>12.46</b>	<b>9.97</b>	<b>-2.49</b>	<b>0.00E+00</b>
		AT5G19120		CATMA5A17530	9.29	9.41	0.12	1.00
	peptidases	AT3G51820		CATMA5A57420	9.24	9.52	0.28	1.00
		AT1G78680	carboxypeptidase	CATMA1A67745	10.76	10.53	-0.23	1.00
		AT3G10450	SCPL7 (serine carboxypeptidase)	CATMA3A09450	9.65	9.70	0.05	1.00
		AT5G08260	SCPL35 (serine carboxypeptidase)	CATMA5A07520	8.87	9.21	0.34	1.00
		AT4G32940	asparaginyl peptidase	CATMA4A34690	12.35	10.82	-1.53	0.00E+00
		AT3G01720	metallopeptidase	CATMA3A00730	8.54	9.47	0.93	9.02E-06
		AT4G38220	metallopeptidases M20/M25/M40	CATMA4A39730	9.76	9.04	-0.72	3.02E-02
		AT5G56660	metallopeptidases M20/M25/M40	CATMA5A52456	9.97	9.64	-0.33	1.00
		AT1G62710	C13 family	CATMA1A51830	8.47	9.19	0.72	2.72E-02
		AT5G18740	M28 family	CATMA5A18160	9.55	9.22	-0.33	1.00
Proteins possibly involved in signaling	arabinogalactan proteins (AGPs)	AT5G10430	AAAGP4	CATMA5A09183	10.84	10.65	-0.19	1.00
		AT2G14890	AAAGP9	CATMA3A12660	11.71	11.82	0.12	1.00
		AT3G13520	AAAGP12	CATMA3A12660	10.77	11.51	0.74	1.50E-02
		AT5G11740	AAAGP15	CATMA5A10540	11.03	11.71	0.68	1.10E-01
		AT3G61640	AAAGP20	CATMA3A54770	8.69	9.53	0.85	3.14E-04
	fasciclin-AGPs (FLAs)	AT5G53250	AAAGP22	CATMA5A49170	9.48	9.98	0.50	1.00
		AT2G47930	AAAGP26	CATMA2A46360	8.46	9.10	0.63	5.19E-01
		AT4G12730	AI8LA2	CATMA4A12870	10.85	10.44	-0.41	1.00
		AT2G04780	AI8LA7	CATMA2A03690	9.27	9.69	0.42	1.00
		AT2G45470	AI8LA8	CATMA2A22160	11.96	11.53	-0.43	1.00
	expressed proteins	AT1G03870	AI8LA9	CATMA1A02720	12.49	12.25	-0.24	1.00
		AT2G23810	tetraspanin domain, signal transduction?	CATMA2A22170	10.91	10.60	-0.31	1.00
	proteins homologous to receptor-kinases (RLKs)	AT1G32400	tetraspanin domain, signal transduction?	CATMA1A30780	10.49	10.35	-0.14	1.00
		AT2G37710	AI8RK1/AILecRKs (RLK/Pelle, L-Lectin subfamily)	CATMA2A35995	10.44	9.83	-0.61	9.33E-01
		AT1G53730	AI8SRF (RLK/Pelle, LRR-V subfamily/STRUBELIG family)	CATMA1A44820	10.06	9.38	-0.68	1.34E-01
		AT2G39180	AI8SR2	CATMA2A37410	9.23	8.30	-0.93	8.33E-06
		AT1G08590	CLAVATA1	CATMA1A07510	9.22	8.32	-0.90	3.05E-05
		AT1G69270	AI8RPK1	CATMA1A58565	10.30	9.39	-0.91	2.35E-05
		AT5G16000	RLK/Pelle, LRR-II subfamily	CATMA5A14280	9.23	9.08	-0.15	1.00
		AT1G02710	RLK/Pelle, LRR-Xa subfamily	CATMA1A25440	9.98	9.78	-0.20	1.00
		AT2G39360	RLK/Pelle, CRLK1L-1 subfamily	CATMA2A37610	8.60	9.21	0.61	2.12E-05
		AT1G09970	LRR-XI subfamily	CATMA1A08840	10.79	10.39	-0.40	1.00
		AT2G01820	LRR IX subfamily	CATMA2A00870	10.98	10.37	-0.61	8.98E-01

1	Structural proteins	glycine-rich proteins (GRPs)	AT2G05380 AT2G05520 AT2G05440 AT1G62240 AT2G05540 AT4G28920	AtGRP3	CATMA2A04205 CATMA2A04295 CATMA2A04270 CATMA1A51310 CATMA2A04310 CATMA4A36695	11.88 11.69 7.22 9.33 12.20 8.65	12.04 11.63 11.01 10.43 12.54 9.15	0.06 -0.06 3.79 1.10 0.34 0.52	1.00 1.00 0.00E+00 3.00E-09 1.00 1.00
7		proline-rich proteins (PRPs)	AT1G02890 AT2G21140	AtPRP2	CATMA1A26460 CATMA2A19790	10.97 11.67	11.42 11.15	0.46 -0.51	1.00 1.00
8		leucine-rich-repeat extensin (LRX)	AT3G24480	AtLRX4 (LRR-extensin)	CATMA3A24360	9.68	9.56	-0.11	1.00
1	Proteins possibly involved in oxidoreduction reactions	peroxidases	AT1G05260 AT1G71695 AT2G37130 AT4G21960 AT4G33870 AT5G64120	AtPrx3 AtPrx12 AtPrx21 AtPrx42 AtPrx48 AtPrx71	CATMA1A04113 CATMA1A60876 CATMA2A35430 CATMA4A23655 CATMA4A38650 CATMA5A59590	10.68 10.17 14.28 13.17 10.52 10.03	10.18 9.54 12.86 12.40 9.25 9.85	-0.51 -0.63 -1.42 -0.77 -1.27 -0.18	1.00 4.88E-01 0.00E+00 5.29E-03 0.00E+00 1.00
7		proteins homologous to berberine bridge enzymes	AtSg44380		CATMA5A40150	8.49	9.53	-1.45	0.00E+00
8		proteins homologous to multicopper oxidases	AT5G21105 AT1G76160 AT1G41830	SKS5 (homologous to SKU5) SKS6 (homologous to SKU5)	CATMA5A19660 CATMA1A65380 CATMA1A35650	9.96 10.33 9.76	9.61 9.91 9.27	-0.35 -0.43 -0.49	1.00 1.00 1.00
11		proteins homologous to germin	AT5G03460 AT5G03690 AT1G09560	GLP AGER3 GLP5 (subfamily 2, member 1)	CATMA5A02640 CATMA5A19130 CATMA1A08413	10.31 11.47 10.19	10.89 9.17 9.79	0.57 0.17 -0.40	1.00 1.00 1.00
14		laccases	AT5G03260 AT5G60020 AT3G03330	AtLAC11 AtLAC17	CATMA5A02370 CATMA5A55770 CATMA3A02250	9.23 9.53 10.65	9.39 9.23 10.73	0.17 -0.30 0.08	1.00 1.00 1.00
17		expressed protein (oxido-reductase domain)							
17		phytocyanins (proteins homologous to blue copper binding proteins)	AT2G32300 AT4G12880 AT1G03650 AT5G15350	AtUCC1 (uclacyanin) plastocyanin AtPC1 (plastocyanin domain) AtEN22 (plastocyanin domain)	CATMA2A30615 CATMA4A13020 CATMA1A07465 CATMA5A13625	10.61 10.84 8.78 9.07	10.60 11.19 9.68 9.35	-0.02 0.36 -0.18 0.28	1.00 1.00 1.00 1.00
1	Proteins with interacting domains	proteins homologous to lectins	AT1G78850 AT1G78830 AT1G13356 AT3G54080 AT5G01090 AT1G69295 AT2G17120	curculin-like, mannose binding curculin-like, mannose-binding legume lectin legume lectin expressed protein (XB domain, carbohydrate binding?) expressed protein (peptidoglycan-binding LyxM domain)	CATMA1A67950 CATMA1A67930 CATMA3A14725 CATMA3A47030 CATMA5A00120 CATMA1A58590 CATMA2A15860	11.95 9.82 10.81 9.50 9.37 11.06 9.22	11.42 9.22 9.86 9.76 9.15 11.34 9.03	-0.53 -0.60 -0.75 0.26 -0.22 0.27 -0.19	1.00 1.00 3.38E-04 1.00 1.00 1.00 1.00
7		proteins with leucine-rich repeat (LRR) domains	AT5G12940 AT3G20820 AT1G28340	expressed protein (LRR domains) expressed protein (LRR domains) expressed protein (LRR domains)	CATMA5A11170 CATMA3A20560 CATMA1A26520	9.68 10.49 9.59	9.05 9.99 9.26	-0.63 -0.50 -0.33	5.72E-01 1.00 1.00
10		enzyme inhibitors	AT4G25260 AT1G73260 AT1G75830 AT2G02120 AT1G71950 AT2G31980 AT1G73330 AT1G21680 AT3G09090 AT3G51050	pectin methylesterase inhibitor serine protease inhibitor protease inhibitor (gamma-thionin) protease inhibitor (gamma-thionin) protease inhibitor protease inhibitor (cystatin) protease inhibitor (Kunitz) expressed protein (WD40-like Beta Propeller Repeat) expressed protein (FG-GAP domain that may fold into beta-propeller structure) expressed protein (FG-GAP repeat that may fold into beta-propeller structure)	CATMA4A26950 CATMA1A62540 CATMA1A65113 CATMA2A01180 CATMA1A61160 CATMA2A30250 CATMA1A62630 CATMA1A20770 CATMA3A07960 CATMA3A44050	8.67 12.00 8.73 9.87 9.46 10.06 11.73 10.9 9.06 10.10	9.10 12.5 10.53 11.29 10.04 11.32 11.55 11.08 9.16 9.87	0.43 0.50 1.80 1.42 0.58 1.26 -0.18 0.18 0.09 -0.23	1.00 1.00 0.00E+00 0.00E+00 1.00 0.00E+00 1.00 1.00 1.00 1.00
1	Proteins related to lipid metabolism	proteins homologous to lipid transfer proteins	AT2G38540 AT1G27950 AT1G71980	LTP1	CATMA2A36820 CATMA1A26130 CATMA1A61200	11.06 10.02 11.86	11.66 10.29 11.05	0.60 0.27 -0.81	1.00 1.00 1.42E-03
3		proteins homologous to GDSL Lipase/Acylhydrolase	AT3G16370 AT1G28580 AT2G04570 AT5G45670 AT3G04290 AT5G45950 AT5G03610	homologous to glycerophosphoryl diester phosphodiesterase homologous to type I phosphodiesterase homologous to phosphodiesterase homologous to acyltransferase	CATMA3A15780 CATMA1A26790 CATMA3A03550 CATMA5A11660 CATMA3A03260 CATMA5A41950 CATMA5A02790	11.55 10.91 10.23 10.06 9.74 9.42 9.23	11.09 11.02 9.72 10.08 9.50 -0.87 9.19	-0.47 0.11 -0.51 0.03 -0.25 1.16E-04 -0.04	1.00 1.00 1.00 1.00 1.00 1.00 1.00
11		proteins homologous to plant lipid transfer protein/seed storage/trypsin alpha amylase inhibitor	AT1G55260 AT2G15050 AT3G53980 AT4G22490 AT1G12090 AT3G18280 AT3G43720		CATMA1A46320 CATMA2A13890 CATMA3A46940 CATMA4A24210 CATMA1A11135 CATMA3A17820 CATMA3A36600	9.08 8.80 8.99 9.40 12.32 9.93 9.49	9.04 9.20 9.99 -0.29 12.54 11.63 9.86	-0.04 0.40 1.00 0.22 0.22 1.70 0.37	1.00 1.00 3.81E-07 1.00 1.00 0.00E+00 1.00
18		expressed protein (ML domain - MD-2-related lipid recognition domain)	AT3G44100 AT2G16005		CATMA3A36990 CATMA2A14850	9.75 10.13	10.43 9.47	0.68 -0.66	1.03E-01 2.20E-01
21		others	AT5G14180 AT3G51730 AT4G26690 AT4G29700 AT5G34850 AT4G01950	homologous to lipase homologous to saposin homologous to glycerophosphoryl diester phosphodiesterase homologous to type I phosphodiesterase homologous to phosphodiesterase homologous to acyltransferase	CATMA5A12400 CATMA3A44690 CATMA4A28280 CATMA4A31330 CATMA5A29650 CATMA4A02220	8.9 10.82 9.38 9.38 11.84 9.05	10.00 10.63 9.00 8.44 11.39 8.33	1.10 -0.18 -0.38 -0.94 -0.45 -0.72	3.79E-09 1.00 1.00 6.03E-06 1.00 2.97E-02
1	Miscellaneous		AT3G57010 AT3G45040 AT5G10480 AT5G09020 AT1G03880 AT2G22170 AT3G08650 AT5G12200 AT2G34250 AT2G34490 AT3G17520 AT2G41260 AT2G28670 AT1G58170 AT2G17500 AT1G68100 AT1G04560 AT4G09600 AT5G15230 AT3G04720 AT1G75800 AT4G36010	homologous to stricoidine synthase homologous to cytidyltransferase PEPINO/PASTICINO2 putative anti-phosphatase (AtPEP/AtPAS2) COBRA homolog to plant seed storage protein (cupin domain) homologous to Brassica napus dehydration stress-induced protein metal transporter family homolog to amido-hydrolase (metal-dependent hydrolase) homolog to protein transport protein (sec61) cytochrome P450 family proteins homologous to late embryogenesis abundant proteins proteins homologous to dirigent proteins hormone-related proteins pathogenesis-related proteins pathogenesis-related protein (AtPR4) (Carbohydrate Binding-Module family 18 - CBM18) homologous to thaumatin homolog to thaumatin	CATMA3A50000 CATMA3A38960 CATMA5A09210 CATMA5A56600 CATMA1A02730 CATMA2A20720 CATMA3A07470 CATMA5A10970 CATMA2A32380 CATMA2A32630 CATMA3A16945 CATMA2A39610 CATMA2A27080 CATMA1A48300 CATMA2B16180 CATMA1A57475 CATMA1A03410 CATMA4A09633 CATMA5A13515 CATMA3A03743 CATMA1A65090 CATMA4A37660	9.95 9.38 9.06 10.48 11.24 10.57 10.12 10.34 10.57 11.82 9.06 9.45 10.63 8.88 11.69 9.62 7.92 8.11 12.58 9.40 10.52 8.65	9.47 9.20 9.09 10.23 11.39 10.54 9.58 9.66 10.75 11.47 10.84 10.4 10.16 9.05 11.01 9.46 10.17 9.53 13.42 9.30 10.25 9.12	-0.48 -0.18 0.04 -0.26 0.15 -0.03 -0.56 -0.68 0.18 -0.35 1.78 0.95 -0.47 0.16 -0.68 -0.16 2.25 1.43 0.84 -0.10 -0.27 0.46	1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.00 3.44E-06 1.00 1.00 1.01E-01 1.00 0.00E+00 0.00 4.08E-04 1.00 1.00 1.00 1.00

Table S2, supplementary data

23	proteins homologous to phosphate-induced proteins (ph)	AT5G09440		CATMA5A08270	11.87	10.73	-1.14	3.41E-10	1,00
24		AT4G08950		CATMA4A08890	9.90	9.64	-0.26		1,00
25		AT5G51550		CATMA5A47480	9.35	9.77	0.42		1,00
26	proteins homologous to acid phosphatases	AT5G44020		CATMA5A39790	9.62	9.64	0.02		1,00
27		AT2G27190	homologous to purple acid phosphatase	CATMA2A25560	10.67	10.42	-0.25		1,00
28		AT1G25230	AtPAP3 (purple acid phosphatase)	CATMA1A23880	10.42	9.94	-0.48		1,00
29		AT2G01890	AtPAP8 (purple acid phosphatase)	CATMA2A00940	9.99	9.62	-0.36		1,00
30	proteins homologous to ribonucleases	AT1G24450		CATMA1A23980	8.57	9.25	0.67		1,46E-01
31		AT2G39780	RNS2	CATMA2A38006	9.04	9.29	0.25		1,00
1	Unknown function	expressed proteins							
2		AT5G20520		CATMA5A19000	11.32	11.75	0.43		1,00
3		AT3G13410		CATMA3A12540	11.19	10.95	-0.23		1,00
4		AT3G11950		CATMA3A15360	11.07	10.74	-0.33		1,00
5		AT5G52180		CATMA5A48120	10.85	10.70	-0.15		1,00
6		AT3G24160		CATMA3A24093	10.51	10.21	-0.30		1,00
7		AT5G25265		CATMA5A22980	10.32	9.75	-0.57		1,00
8		AT1G78040		CATMA1A67150	10.21	10.23	0.02		1,00
9		AT1G61900		CATMA1A50870	9.99	9.67	-0.33		1,00
10		AT2G12400		CATMA2A10420	9.91	9.22	-0.69	8.75E-02	1,00
11		AT5G58375		CATMA5A54130	9.74	10.32	0.58		1,00
12		AT1G69960		CATMA1A59280	9.69	9.97	0.27		1,00
13		AT4G27120		CATMA4A28730	9.66	9.50	-0.16		1,00
14		AT1G47310		CATMA1A38230	9.56	9.22	-0.34		1,00
15		AT4G39840		CATMA4A41230	9.53	9.82	0.29		1,00
16		AT1G43580		CATMA1A36910	9.47	9.64	0.17		1,00
17		AT2G42840		CATMA2A41257	9.43	9.26	-0.17		1,00
18		AT5G17190		CATMA5A15470	9.41	9.91	0.50		1,00
19		AT5G13140		CATMA5A11350	9.26	9.53	0.27		1,00
20		AT5G61820		CATMA5A57420	9.22	9.28	0.05		1,00
21		AT3G09570		CATMA3A08440	9.15	9.06	-0.09		1,00
22		AT1G29790		CATMA1A27760	9.12	9.37	0.25		1,00
23		AT2G04690		CATMA2A03640	9.11	9.24	0.14		1,00
24		AT1G67785		CATMA1A57160	8.98	9.74	0.76	8.15E-03	1,00
25		AT4G30450		CATMA4A32060	8.97	9.53	0.56		1,00
26		AT5G01075		CATMA5A00100	8.86	9.19	0.34		1,00
27		AT1G48440		CATMA1A39540	8.83	9.15	0.32		1,00
28		AT1G80860		CATMA1A70080	8.71	9.23	0.52		1,00
29		AT2G27385		CATMA2A25790	8.50	9.00	0.50		1,00
30		AT3G27390		CATMA3A27250	8.24	8.34	-0.90	3.05E-05	1,00
31		AT1G16850		CATMA1A15840	8.12	9.01	0.90	3.70E-05	1,00
32	expressed proteins with domains of unknown function (DUF) or uncharacterized protein families (UPF)	AT1G61740	DUF81	CATMA1A50820	9.41	9.41	0.00		1,00
33		AT1G01100	DUF221	CATMA3A00135	9.13	9.04	-0.10		1,00
34		AT1G30360	DUF221	CATMA1A26370	10.46	10.46	-0.91	2.25E-05	1,00
35		AT3G54260	DUF231	CATMA3A47210	7.38	9.00	1.62	0.00E+00	1,00
36		AT1G51630	DUF246	CATMA1A42760	9.65	9.15	-0.50		1,00
37		AT3G23300	DUF248	CATMA3A23300	9.77	9.60	-0.17		1,00
38		AT5G26740	DUF300	CATMA5A24080	10.69	10.24	-0.46		1,00
39		AT5G19590	DUF538	CATMA5A18020	9.37	9.37	0.01		1,00
40		AT5G19860	DUF538	CATMA5A18290	10.34	10.02	-0.32		1,00
41		AT4G02370	DUF538	CATMA4A02670	10.56	10.8	0.24		1,00
42		AT3G07460	DUF538	CATMA3A06690	9.63	9.48	-0.15		1,00
43		AT1G27930	DUF578	CATMA1A26110	9.27	9.58	0.31		1,00
44		AT4G32460	DUF642	CATMA4A34200	9.52	9.02	-0.51		1,00
45		AT5G25460	DUF642	CATMA5A23220	9.81	9.24	-0.57		1,00
46		AT1G29980	DUF642	CATMA1A28000	9.20	9.36	0.15		1,00
47		AT3G08030	DUF642	CATMA3A07290	12.03	11.69	-0.33		1,00
48		AT5G11420	DUF642	CATMA5A10190	11.97	11.68	-0.29		1,00
49		AT3G15480	DUF1218	CATMA3A14890	10.23	10.44	0.21		1,00
50		AT1G27000	DUF1664	CATMA1A25200	9.89	9.81	-0.08		1,00
51		AT4G30660	UPF0057	CATMA4A32270	8.52	9.37	0.85	2.45E-04	1,00
52		AT5G36290	UPF0016	CATMA5A31840	9.26	9.44	0.18		1,00
53	expressed proteins with diverse structural domains	AT1G07080	GILT domain	CATMA1A06150	10.23	10.65	0.42		1,00
54		AT5G25610	BURP domain	CATMA5A23380	11.92	11.36	-0.55		1,00
55		AT3G20370	MATH domain	CATMA3A20050	10.31	10.33	0.01		1,00
56		AT3G22840	cupin domain	CATMA3A22600	9.52	10.52	1.00	4.06E-07	1,00
57		AT2G28490	cupin domain	CATMA2A26920	10.09	10.69	0.59		1,00
58		AT1G53210	calcium-binding EF-hand domains	CATMA1A44240	10.47	9.91	-0.56		1,00
59		AT1G29020	calcium-binding EF-hand domains	CATMA1A26980	9.91	9.19	-0.72	2.90E-02	1,00
60		AT2G20990	calcium dependent membrane targeting domain, CalB protein	CATMA2A19620	10.37	9.89	-0.47		1,00
61		AT1G05500	calcium dependent membrane targeting domain, CalB protein	CATMA1A04400	9.78	9.38	-0.40		1,00
62		AT3G61050	calcium dependent membrane targeting domain, CalB protein	CATMA3A54220	10.31	9.65	-0.66	2.17E-01	1,00
63		AT5G18520	integral membrane protein	CATMA5A16820	11.15	10.90	-0.25		1,00
64		AT5G42090	integral membrane protein	CATMA5A37800	10.65	10.50	-0.15		1,00
65		AT4G39730	membrane or lipid associated protein domain	CATMA4A41080	10.31	10.23	-0.08		1,00
66		AT2G25110	MIR domain	CATMA2A23430	8.70	9.06	0.35		1,00
67		AT5G25100	EMP70 domain	CATMA5A22810	9.10	9.22	0.12		1,00
68		AT1G64670	alpha/beta hydrolase fold	CATMA1A53990	9.87	9.47	-0.40		1,00
69	other expressed proteins	AT4G14420	protein associated to hypersensitive response	CATMA4A14790	8.83	9.44	0.61	9.59E-01	1,00
70		AT2G41470	embryo specific protein	CATMA2A39880	8.96	9.47	0.50		1,00
		AT2G34700	pollen proteins Ole e 1 family	CATMA2A32820	10.14	8.35	-1.80	0.00E+00	1,00



## Additional data file 3

**Table 3. Secretory Pathway Genes (SPGs) with modulated levels of transcripts in 5- and 11-day-old hypocotyls.** SPGs were annotated using bioinformatic softwares for sub-cellular localization and functional domains as described in Methods. The intensity of the signal is expressed as log<sub>2</sub>, the ratio between the levels of transcripts at 11-days and 5-days as well as the p-value are indicated (Bonferroni p-value<5%).

Functional class	Gene family	AGI number	gene name	5-day-old hypocotyls	11-day-old hypocotyls	ratio 11-days / 5-days	p-value
<b>Level of transcripts higher in 5-day-old than in 11-day-old hypocotyls</b>							
1 <b>Proteins involved in polysaccharide synthesis</b>	glycosyl transferase family 2 (cellulose synthases)	AT5G09870	AtCESA5	10.01	9.11	-0.90	3,26E-05
	glycosyl transferase family 48 (callose synthases)	AT1G05570	AtCALS1	9.17	8.14	-1.03	8,89E-08
2							
3 <b>Proteins acting on carbohydrates</b>	glycoside hydrolase family 1 (beta-glucosidase)	AT3G60130		10.84	9.70	-1.14	3.41E-10
	glycoside hydrolase family 3	AT5G49360		12.67	11.94	-0.73	2,20E-02
	glycoside hydrolase family 9	AT4G02990		11.10	10.08	-1.02	1,57E-07
	glycoside hydrolase family 16 (xyloglucan endotransglycosidases/hydrolases)	AT2G06850	AtXTH4	12.98	12.15	-0.83	6,40E-04
		AT4G14130	AtXTH15	12.72	11.98	-0.74	1,62E-02
	glycoside hydrolase family 17	AT3G04010		8.30	9.85	-1.55	0,00E+00
	glycoside hydrolase family 35 (beta-galactosidase)	AT5G56870	AtBGAL4	13.18	11.97	-1.22	5,60E-12
		AT1G45130	AtBGAL5	11.21	10.42	-0.78	3,24E-03
		AT3G52840	AtBGAL2	8.44	7.71	-0.73	2,32E-02
	glycoside hydrolase family 38 (alpha-mannosidase)	AT3G26720		11.33	10.38	-0.95	4,81E-06
13 <b>Proteases</b>	carbohydrate esterase family 8 (pectin methylesterases)	AT3G49220		11.81	10.99	-0.82	8,65E-04
		AT2G43050		12.56	10.76	-1.80	0,00E+00
		AT3G14310	AtPME3	10.79	9.90	-0.90	4,05E-05
	serine proteases (proteins homologous to serine proteases)	AT3G14067	AtSBT1.4	11.85	10.71	-1.15	2,80E-10
		AT5G59090	AtSBT4.12	10.47	9.49	-0.99	7,99E-07
	aspartyl proteases	AT5G10770		12.46	9.97	-2.49	0,00E+00
	peptidases	AT4G32940	asparaginyl peptidase	12.35	10.82	-1.53	0,00E+00
		AT3G01720	metallopeptidase	8.54	9.47	-0.93	9,02E-06
		AT4G38220	metallopeptidases M20/M25/M40	9.76	9.04	-0.72	3,02E-02
22 <b>Proteins possibly involved in signaling</b>	proteins homologous to receptor-kinases (RLKs)	AT1G08590	CLAVATA1	9.22	8.32	-0.90	3,05E-05
		AT2G39180	AtCCR2	9.23	8.30	-0.93	8,33E-06
		AT1G69270	AtRPK1	10.30	9.39	-0.91	2,35E-05
		AT1G07650	LRR VIII-2 subfamily	8.95	8.18	-0.76	7,19E-03
26 <b>Proteins possibly involved in oxido reduction reactions</b>	peroxidases	AT2G37130	AtPrx21	14.28	12.86	-1.42	0,00E+00
		AT4G21960	AtPrx42	13.17	12.40	-0.77	5,29E-03
		AT4G33870	AtPrx48	10.52	9.25	-1.27	0,00E+00
	proteins homologous to berberine bridge enzymes	AT5G44380		8.49	9.93	-1.45	0,00E+00
29 <b>Proteins with interacting domains</b>	proteins homologous to lectins	AT3G15356	legume lectin	10.81	9.96	-0.84	3,38E-04
		AT1G78820	curculin-like, mannose-binding	8.96	8.04	-0.93	1,15E-05
	proteins with leucine-rich repeat (LRR) domains	AT1G78230	expressed protein	8.22	7.48	-0.74	1,41E-02
	proteins with fasciclin domains	AT1G52370	expressed protein	8.80	8.08	-0.73	2,21E-02
34 <b>Proteins related to lipid metabolism</b>	proteins homologous to GDSL Lipase/Acylhydrolase	AT1G71980		11.86	11.05	-0.81	1,42E-03
		AT5G45850		9.42	8.55	-0.87	1,16E-04
		AT2G42990		8.73	7.99	-0.73	1,90E-02
	proteins homologous to acyltransferase	AT4G01950		9.05	8.33	-0.72	2,97E-02
	protein homologous to type I phosphodiesterase	AT4G29700		9.38	8.44	-0.94	6,03E-06
39 <b>Miscellaneous</b>	proteins homologous to phosphate-induced proteins (phi)	AT5G09440		11.87	10.73	-1.14	3,41E-10
40 <b>Unknown function</b>	expressed proteins with domains of unknown function (DUF) or uncharacterized protein families (UPF)	AT1G30360	DUF221	11.37	10.46	-0.91	2,25E-05
	other expressed proteins	AT2G34700	homolog to <i>N. tabacum</i> pistil-specific protein	10.14	8.35	-1.80	0,00E+00
		AT3G27390		9.24	8.34	-0.90	3,05E-05
		AT2G32560		9.34	8.58	-0.76	7,05E-03
<b>Level of transcripts higher in 11-day-old than in 5-day-old hypocotyls</b>							
1 <b>Proteins acting on carbohydrates</b>	glycoside hydrolase family 1	AT3G21360		9.96	11.47	1.51	0,00E+00
	glycoside hydrolase family 16 (xyloglucan endotransglycosidases/hydrolases)	AT4G30270	AtXTH24	10.84	11.64	0.80	1,63E-03
	pectate lyase family 1	AT4G24780		8.40	9.43	1.03	1,03E-07
	expansins	AT4G28250	AtEXPB3	9.07	10.38	1.31	0,00E+00
5 <b>Proteases</b>	cysteine proteases (papain family)	AT3G54940		8.35	10.33	1.98	0,00E+00
		AT1G62290		11.77	12.48	0.72	3,29E-02
	peptidases	AT1G62710	C13 family	8.47	9.19	0.72	2,72E-02
8 <b>Proteins possibly involved in signaling</b>	arabinogalactan proteins (AGPs)	AT3G13520	AtAGP12	10.77	11.51	0.74	1,50E-02
		AT3G61640	AtAGP20	8.69	9.53	0.85	3,14E-04
10 <b>Proteins possibly involved in oxido reduction reactions</b>	peroxidases	AT2G38390	AtPrx23	7.80	8.96	1.15	1,96E-10
11 <b>Structural proteins</b>	glycine-rich proteins (GRPs)	AT2G05440		7.22	11.01	3.79	0,00E+00
		AT1G62240		9.33	10.43	1.10	3,00E-09
13 <b>Proteins with interacting domains</b>	enzyme inhibitors	AT1G75830	protease inhibitor (gamma-thionin)	8.73	10.53	1.80	0,00E+00
		AT2G02120	protease inhibitor (gamma-thionin)	9.87	11.29	1.42	0,00E+00
		AT2G15010	protease inhibitor (gamma-thionin)	7.57	8.70	1.13	5,54E-10
		AT2G43510	protease inhibitor (trypsin inhibitor)	7.99	8.93	0.94	4,99E-06
		AT2G31980	protease inhibitor (cystatin)	10.06	11.32	1.26	0,00E+00

18	Proteins related to lipid metabolism	proteins homologous to plant lipid transfer protein/seed storage/trypsin alpha amylase inhibitor	AT3G53980	8,99	9,99	1,00	3,81E-07
19			AT3G18280	9,93	11,63	1,70	0,00E+00
20		others	AT5G14180 homolog to lipase	8,90	10,00	1,10	3,79E-09
21	Miscellaneous	proteins homologous to late embryogenesis abundant proteins	AT3G17520	9,06	10,84	1,78	0,00E+00
22			AT2G41260 M17	9,45	10,40	0,95	3,44E-06
23		hormone-related proteins	AT1G04560 homolog to AWPMP (ABA-induced) gibberellin regulated protein (GASA4)	7,92	10,17	2,25	0,00E+00
24			AT5G15230 gibberellin regulated protein (GASA3)	12,58	13,42	0,84	4,08E-04
25			AT4G09600	8,11	9,53	1,43	0,00E+00
26		others	AT2G22860 phytosulfokine (AtPSK2)	7,61	8,49	0,88	7,46E-05
27			AT3G49780 phytosulfokine (AtPSK4)	7,64	8,61	0,97	1,97E-06
28	Unknown function	expressed proteins	AT1G67785	8,98	9,74	0,76	8,15E-03
29			AT1G16850	8,12	9,01	0,90	3,70E-05
30			AT1G54860	7,46	8,66	1,20	1,68E-11
31			AT5G40960	6,96	8,07	1,11	1,58E-09
32			AT4G38080	7,16	8,39	1,23	0,00E+00
33			AT5G64510	7,49	8,96	1,48	0,00E+00
34			AT5G53650	7,87	8,96	0,91	2,43E-05
35		expressed proteins with domains of unknown function (DUF) or uncharacterized protein families (UPF)	AT3G54260 DUF231	7,38	9,00	1,62	0,00E+00
36			AT4G30660 UPF0057	8,52	9,37	0,85	2,45E-04
37		expressed proteins with diverse structural domains	AT3G22640 cupin domain	9,52	10,52	1,00	4,06E-07
38			AT1G49320 BURP domain	7,16	8,01	0,86	2,10E-04

## Additional data file 4

**Table 4. Proteins extracted and identified by mass spectrometry from purified cell walls of 5- and 11-day-old etiolated hypocotyls of *A. thaliana*.** All these proteins were annotated using bioinformatic softwares for sub-cellular localization and functional domains as described in Material and methods. Proteomic data are in Irshad *et al.* (2008, submitted for publication). Presence of proteins at 5-days or 11-days is indicated by a ticked box. For transcriptomics, the intensity of the signal is expressed as log<sub>2</sub>, the ratio between the levels of transcripts at 11-days and 5-days as well as the p-value are indicated. Background level for transcriptomics is 6.75 (grey boxes indicate results below background). Genes encoding proteins with predicted GPI anchors or trans-membrane domains are in bold, and in pink boxes respectively.

		proteomic data		transcriptomic data			
accession	annotation	5-days	11-days	5-days	11-days	ratio	p-value
AGI						11-days/5-days	
proteins acting on carbohydrates							
At2g06850	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH4)	✓	✓	12,98	12,15	-0,83	6,40E-04
At5g13870	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH5)	✓		8,17	8,15	-0,02	1,00E+00
At3g44990	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH31)	✓		nd	nd		
At2g36870	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH32)		✓	8,31	7,86	-0,45	1,00E+00
At1g10550	glycoside hydrolase family 16 (endoxylglucan transferase) (AtXTH33)	✓	✓	6,21	6,22	0,01	1,00E+00
At4g16260	glycoside hydrolase family 17		✓	6,56	6,68	0,12	1,00E+00
At4g19810	glycoside hydrolase family 18 (chitinase)	✓		6,88	6,89	0,00	1,00E+00
At3g12500	glycoside hydrolase family 19 (chitinase)	✓	✓	8,08	7,72	-0,36	1,00E+00
	glycoside hydrolase family 20 (beta-hexosaminidase)	✓		9,64	9,13	-0,51	1,00E+00
	glycoside hydrolase family 28 (polygalacturonase)	✓		6,70	6,68	-0,02	1,00E+00
At1g19170	glycoside hydrolase family 28 (polygalacturonase)	✓		6,56	6,46	-0,10	1,00E+00
At2g33160	glycoside hydrolase family 28 (polygalacturonase)	✓					
At3g06770	glycoside hydrolase family 28 (polygalacturonase)		✓	9,01	8,78	-0,23	1,00E+00
At3g16850	glycoside hydrolase family 28 (polygalacturonase)	✓	✓	7,08	6,95	-0,13	1,00E+00
At3g61490	glycoside hydrolase family 28 (polygalacturonase)	✓	✓	8,09	8,55	0,45	1,00E+00
At4g18180	glycoside hydrolase family 28 (polygalacturonase)	✓		6,57	6,55	-0,03	1,00E+00
At1g68560	glycoside hydrolase family 31 (alpha-xylosidase) (AtXYL1)	✓	✓	10,6	10,13	-0,47	1,00E+00
At3g13790	glycoside hydrolase family 32 (beta-fructofuranosidase)	✓	✓	6,89	6,78	-0,11	1,00E+00
At5g34940	glycoside hydrolase family 79 (endo beta-glucuronidase/heparanase)	✓	✓	6,98	6,90	-0,09	1,00E+00
At1g11580	carbohydrate esterase family 8 (pectin methyltransferase)	✓		7,29	7,26	-0,04	1,00E+00
At1g53830	carbohydrate esterase family 8 (pectin methyltransferase) (AtPME2)	✓	✓	8,35	7,61	-0,74	1,61E-02
At3g14310	carbohydrate esterase family 8 (pectin methyltransferase) (AtPME3)	✓	✓	10,79	9,90	-0,90	4,05E-05
At3g43270	carbohydrate esterase family 8 (pectin methyltransferase)	✓	✓	6,30	6,46	0,16	1,00E+00
At4g33220	carbohydrate esterase family 8 (pectin methyltransferase)	✓	✓	8,08	7,75	-0,33	1,00E+00
At5g53370	carbohydrate esterase family 8 (pectin methyltransferase)	✓		7,41	7,30	-0,11	1,00E+00
At4g37950	polysaccharide lyase family 4 (rhamnogalacturonate lyase)		✓	6,68	6,73	0,05	1,00E+00
At1g05570	glycosyl transferase family 48 (callose synthase) (AtCalS1)	✓	✓	6,58	6,58	0,00	1,00E+00
At5g02260	alpha-expansin (AtEXPA9)	✓	✓	12,71	12,44	-0,28	1,00E+00
At1g20190	alpha-expansin (AtEXPA11)	✓	✓	8,76	8,72	-0,03	1,00E+00
At5g39270	alpha expansin (AtEXP22)	✓		nd	nd		
At3g45970	expansin-like A (AtEXLA1)	✓	✓	nd	nd		
At4g38400	expansin-like A (AtEXLA2)	✓	✓	7,98	7,74	-0,24	1,00E+00
At3g45960	expansin-like A (AtEXLA3)	✓		6,97	6,91	-0,06	1,00E+00
oxido-reductases							
At1g71695	peroxidase (AtPrx12)	✓	✓	10,17	9,54	-0,63	4,86E-01
At3g21770	peroxidase (AtPrx30)	✓	✓	9,54	9,77	0,23	1,00E+00
At3g32980	peroxidase (AtPrx32)	✓	✓	nd	nd		
At3g49110	peroxidase (AtPrx33)	✓	✓	6,28	6,37	0,09	1,00E+00
At3g49120	peroxidase (AtPrx34)	✓	✓	nd	nd		
At3g50990	peroxidase (AtPrx36)	✓		6,47	6,58	0,11	1,00E+00
At5g25980	peroxidase (AtPrx43)	✓	✓	6,52	6,54	0,02	1,00E+00
At4g30170	peroxidase (AtPrx45)	✓	✓	7,52	7,65	0,13	1,00E+00
At5g17820	peroxidase (AtPrx57)	✓	✓	6,99	6,76	-0,23	1,00E+00
At5g64100	peroxidase (AtPrx69)	✓	✓	6,72	6,68	-0,05	1,00E+00
At5g66390	peroxidase (AtPrx72)	✓	✓	8,70	8,57	-0,13	1,00E+00
At2g30210	homologous to laccase (AtLAC3)	✓	✓	8,23	8,14	-0,09	1,00E+00
At1g30710	homologous to berberine-bridge enzyme (S)-reticuln:oxygen oxidoreductase	✓		6,32	6,41	0,09	1,00E+00
At4g20860	homologous to berberine-bridge enzyme (S)-reticuln:oxygen oxidoreductase		✓	6,80	6,78	-0,01	1,00E+00
At5g44360	homologous to berberine-bridge enzyme (S)-reticuln:oxygen oxidoreductase	✓	✓	6,56	6,64	0,07	1,00E+00
At5g44410	homologous to berberine-bridge enzyme (S)-reticuln:oxygen oxidoreductase		✓	6,78	6,91	0,12	1,00E+00
At1g01980	homologous to berberine-bridge enzyme (S)-reticuln:oxygen oxidoreductase		✓	6,22	6,34	-0,11	1,00E+00
At5g22140	expressed protein (oxido-reductase domain)	✓		nd	nd		
At5g56490	expressed protein (FAD binding domain)	✓		nd	nd		
	plantacyanin ARPN (blue copper binding protein)		✓	6,84	6,90	0,06	1,00E+00
At2g02850	early nodulin AtEN20 homologous to blue copper binding protein						
At4g12880		✓	✓	10,84	11,19	0,36	1,00E+00

proteins with interacting domains							
At1g53070	homologous to lectin (legume lectin domains)	✓	✓	7,92	7,74	-0,18	1,00E+00
At1g78820	homologous to lectin (curculin-like)	✓		8,96	8,04	-0,93	1,15E-05
At1g78830	homologous to lectin (curculin-like)	✓	✓	9,82	9,22	-0,60	1,00E+00
At1g78850	homologous to lectin (curculin-like)	✓	✓	11,95	11,42	-0,53	1,00E+00
At1g78860	homologous to lectin (curculin-like)	✓	✓	7,05	7,20	0,15	1,00E+00
At5g06860	PGIP1 (LRR domains)	✓	✓	nd	nd		
T23B7.10	homologous to PGIP1 (LRR protein FLR1)	✓	✓	nd	nd		
At5g12940	homologous to <i>Phaseolus vulgaris</i> PGIP2 (LRR domains)	✓	✓	9,68	9,05	-0,63	5,72E-01
At1g33590	expressed protein (LRR domains)	✓	✓	nd	nd		
At2g34930	expressed protein (LRR domains)	✓	✓	7,46	7,68	0,22	1,00E+00
At3g20820	expressed protein (LRR domains)	✓	✓	10,49	9,99	-0,50	1,00E+00
At2g17120	expressed protein (LysM domain)	✓	✓	9,22	9,03	-0,19	1,00E+00
At1g03220	homologous to carrot EDGP and tomato XEGIP	✓	✓	nd	nd		
At1g03230	homologous to carrot EDGP and tomato XEGIP	✓	✓	nd	nd		
At5g19110	homologous to carrot EDGP and tomato XEGIP	✓	✓	8,10	8,09	-0,01	1,00E+00
At1g47710	homologous to serpin (serine protease inhibitor)	✓	✓	8,01	7,97	-0,04	1,00E+00
At1g17860	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	✓	✓	8,78	8,95	0,17	1,00E+00
At1g73260	inhibitor family I3 (Kunitz-P family) (subfamily I3A unassigned peptidase inhibitor homologues, MEROPS)	✓	✓	12,00	12,50	0,50	1,00E+00
At1g47540	inhibitor family I18 (mustard trypsin inhibitor-2 family) (family I18 unassigned peptidase inhibitor homologues, MEROPS)	✓	✓	nd	nd		
At2g40880	inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	✓	✓	nd	nd		
At5g05110	inhibitor family I25 (phytostatin) (cystatin family, I25.014, MEROPS)	✓	✓	7,84	7,93	0,09	1,00E+00
At4g16500	inhibitor family I25 (cystatin family) (subfamily I25B unassigned peptidase inhibitor homologues, MEROPS)	✓	✓	8,08	8,42	0,34	1,00E+00
At4g25260	homologous to plant invertase/ pectin methyltransferase inhibitor (PMEI)	✓	✓	8,67	9,10	0,43	1,00E+00
At5g46940	homologous to plant invertase/pectin methyltransferase inhibitor	✓		6,36	6,33	-0,03	1,00E+00
At5g46960	homologous to plant invertase/pectin methyltransferase inhibitor	✓		nd	nd		
proteases							
At1g09750	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	✓	✓	nd	nd		
At3g02740	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	✓		7,22	7,32	0,10	1,00E+00
At3g52500	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	✓		8,8	8,34	-0,46	1,00E+00
At3g54400	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	✓	✓	9,33	8,92	-0,42	1,00E+00
At5g07030	homologous to aspartic protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS)	✓	✓	nd	nd		
At1g79720	homologous to aspartic protease (CND41 peptidase) (Peptidase family A1, subfamily A1.050, MEROPS)	✓		8,93	8,43	-0,51	1,00E+00
At5g10770	homologous to aspartic protease (CND41 peptidase) (Peptidase family A1, A01.050, MEROPS)	✓	✓	12,46	9,97	-2,49	0,00E+00
At1g47128	homologous to cysteine proteinase (papain family) (RD21A, Peptidase family C1, C01.064 MEROPS)	✓	✓	nd	nd		
At5g43060	homologous to cysteine proteinase (papain family) (RD21 peptidase, Peptidase family C1, C01.064 MEROPS)	✓	✓	nd	nd		
At4g01610	homologous to cysteine proteinase (papain family) (Cathepsin B, Peptidase family C1, C01.049, MEROPS)	✓	✓	12,4	12,38	-0,02	1,00E+00
At4g36880	homologous to cysteine proteinase (papain family) (Peptidase family C1, Brassicain, C01.021 MEROPS)	✓		8,63	8,79	0,16	1,00E+00
At3g02110	homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS)	✓	✓	7,92	7,71	-0,20	1,00E+00
At5g23210	homologous to serine carboxypeptidase (SCPL34) (Peptidase family S10, S10.005, MEROPS)	✓	✓	9,49	8,82	-0,67	1,49E-01
At4g30610	homologous to serine carboxypeptidase (BRS1 - Brassinosteroid-Insensitive BRI suppressor 1) (Peptidase family S10, S10.015 MEROPS)	✓		7,42	7,21	-0,21	1,00E+00
structural proteins							
At1g28290	proline-rich protein	✓	✓	10,97	11,42	0,46	1,00E+00
At5g14920	proline-rich protein	✓	✓	8,11	7,79	-0,31	1,00E+00
At2g05580	glycine-rich protein	✓	✓	8,12	8,40	0,29	1,00E+00
At4g13340	LRR-extensin (AtLRX3)	✓		6,89	6,83	-0,06	1,00E+00
At3g24480	LRR-extensin (AtLRX4)	✓	✓	9,68	9,56	-0,11	1,00E+00
At4g18670	LRR-extensin (AtLRX5)	✓	✓	6,70	6,79	0,09	1,00E+00

<b>signaling</b>								
At4g05200	homologous to receptor kinase (RLK, DUF26-1b subfamily)	✓			6,50	6,51	0,01	1,00E+00
At5g55730	fasciclin-like arabinogalactan protein (AtFLA1)	✓	✓		8,18	7,96	-0,22	1,00E+00
<b>proteins related to lipid metabolism</b>								
At1g29670	homologous to lipase acylhydrolase (GDSL family)	✓	✓	nd	nd			
At1g54010	homologous to lipase/acylhydrolase (GDSL family)	✓		nd	nd			
At1g54030	homologous to lipase/acylhydrolase (GDSL family)	✓	✓		10,10	9,75	-0,35	1,00E+00
At3g48460	homologous to lipase acylhydrolase (GDSL family)	✓	✓		8,05	7,79	-0,26	1,00E+00
At5g15720	homologous to lipase acylhydrolase (GDSL family)		✓		6,85	6,73	-0,13	1,00E+00
At2g38530	non-specific lipid transfer protein 2 (LTP2)	✓	✓	nd	nd			
At5g23820	expressed protein (ML domain - MD-2-related lipid recognition domain)	✓		nd	nd			
At2g16001	expressed protein (lipid recognition domain)	✓		nd	nd			
<b>miscellaneous functions</b>								
At2g27190	homologous to purple acid phosphatase (PAP1)	✓	✓		10,67	10,42	-0,25	1,00E+00
At3g07130	homologous to purple acid phosphatase	✓	✓		6,62	6,52	-0,10	1,00E+00
At5g34850	homologous to purple acid phosphatase	✓			11,84	11,39	-0,45	1,00E+00
At4g29270	homologous to acid phosphatase	✓	✓	nd	nd			
At4g24340	homologous to phosphorylase	✓		nd	nd			
At3g02870	homologous to myo-inositol monophosphatase	✓			7,17	7,38	0,21	1,00E+00
At5g09440	homologous to <i>Nicotiana tabacum</i> phi-I	✓	✓		11,87	10,73	-1,14	3,41E-10
At5g4260	homologous to <i>Nicotiana tabacum</i> phi-I	✓	✓		6,47	6,47	0,00	1,00E+00
At5g66590	pathogenesis-related protein PR1	✓	✓		6,42	6,40	-0,02	1,00E+00
At2g28790	homologous to <i>Lycopersicon esculentum</i> osmotin	✓	✓		8,00	8,45	0,45	1,00E+00
At5g15230	gibberellin-regulated protein (GASA4)	✓			12,58	13,42	0,84	4,08E-04
At4g27110	homologous to COBRA (AtCOBL10)	✓			6,39	6,38	-0,01	1,00E+00
At1g09560	germin (subfamily 2, member 1, GLP5)		✓		10,49	9,99	-0,50	1,00E+00
<b>unknown function</b>								
At3g56750	expressed protein		✓	nd	nd			
At3g22000	expressed protein (DUF26)		✓	nd	nd			
At1g26850	expressed protein (DUF248)		✓	nd	nd			
At1g80240	expressed protein (DUF642)	✓	✓		8,11	7,73	-0,38	1,00E+00
At3g08030	expressed protein (DUF642)	✓	✓		12,03	11,69	-0,33	1,00E+00
At4g32460	expressed protein (DUF642)	✓	✓		9,52	9,02	-0,51	1,00E+00
At5g11420	expressed protein (DUF642)	✓	✓		11,97	11,68	-0,29	1,00E+00
At5g25460	expressed protein (DUF642)	✓	✓		9,81	9,24	-0,57	1,00E+00
At1g78460	expressed protein (SOUL heme binding domain)		✓		6,52	6,54	0,02	1,00E+00
At2g04690	expressed protein (homologous to a human brain CREG protein)	✓	✓		9,11	9,24	0,14	1,00E+00
At2g15220	expressed protein (Plant Basic Secreted Protein domain)	✓	✓	nd	nd			
At2g34700	expressed protein (Ole el allergen domain)	✓	✓		10,14	8,35	-1,80	0,00E+00
At3g20370	expressed protein (MATH domain)	✓		nd	nd			
At2g28490	expressed protein (cupin domain)	✓	✓		10,09	10,69	0,59	1,00E+00
At3g22640	expressed protein (cupin domain)	✓	✓		9,52	10,52	1,00	4,06E-07
At4g36700	expressed protein (cupin domain)	✓			6,72	7,11	0,39	1,00E+00

**Additional data file 5****Table 5. Nucleotide primers used for PCR amplifications**

<b>Accession number</b>	<b>Primer sequence (5'→3', sense orientation)</b>	<b>Primer sequence (5'→3', antisense orientation)</b>	<b>Length of expected fragment (bp)</b>
At1g05570	CACTTCCTCTGATAACAATCAAAGAC	GCTTACACGCTTGTGCAATG	96
At1g10550	CAATTCAAGTAAGATCGCCATTG	TTTGACACCAACCCAGCTC	125
At1g28290	AACCATAAGACTCAAACCCTTC	TGAGGGTGGTGGTGATGAG	95
At1g49240	CACCCGAGAGGAAGTACAGTG	CATACTCTGCCTTAGAGATCCACA	93
At1g66180	ACCACAAGCTCAACAAATGGT	GGAGGAAGCTTTTTACGATGAC	78
At1g68560	GTCGCCGTAAATGTTGTTG	CCCATCTACATTGATGAAATCCT	95
At1g69530	GCATCGCTCAATACAGAGC	GAGTGTCCGTTTATCGTAAACCTT	97
At2g21140	TCTCATCCTTGATAAAGATGC	TCTACGTAGAATTCAACAAAGC	111
At2g28790	GTGCCACTACAACGGAAAA	TGAGCTAGAGAAGCTGGTGGT	100
At3g07130	GGGATCGGCAACCTGATTA	GCCCATGTCTCGTTCTTCAT	82
At3g16850	GTAGCCTTCAGAACCACCAGA	TCAACGCTTGTGAGATCGAC	80
At3g43270	TTCCGATCTGCTACATTTGCT	GACCGCTTGGTGTTTTCC	96
At4g12880	GGCATGAAGCTAGATGTTTTAGTTG	ACATAAAATTGATTTCTTATTGTGCTG	91
At4g18670	CACCACCTCCAATCTACGAAG	GCCCCTTTTGAGAACATTCTG	112
At5g05850	TGGTACCACTGGAGGAAAAGC	TTGCAACTCCATAGCCACAG	95
At5g11420	GTCTCTTCTCTTTACTTTGGTCGTC	AGTCGCCGTTTGGTAACATC	126
At5g44360	AGGAGCTTATTTGAATTACCGAGA	TCGCATCTTCAAAGCTCGTA	74
At5g64100	GCAAGACTTCGCTGCTAAAC	GCCGTTGAAGTTAACGAACC	118
desmin <sup>a</sup>	CAGCCTCAGTCCTCCAAATCACA	TAGGCCTGAGGTCACAGAGGT	

a : Pig desmin RNA was used as an internal control for reverse transcription.

**Muhammad IRSHAD**

**Dynamique des protéines pariétales au cours de l'élongation cellulaire  
dans des hypocotyles étiolés d'*Arabidopsis thaliana* :  
approches protéomique et transcriptomique**

**Résumé**

La paroi cellulaire des végétaux supérieurs est une structure complexe jouant de nombreux rôles au cours du développement ainsi qu'en réponse aux stress. Les protéines pariétales sont notamment importantes au cours de l'élongation cellulaire. Des hypocotyles étiolés d'*Arabidopsis thaliana* ont été choisis comme modèle d'élongation cellulaire parce qu'ils subissent une élongation polarisée et rapide en l'absence de division cellulaire. Deux stades de développement ont été comparés grâce à des analyses protéomique et transcriptomique : pendant leur élongation vs après la fin de leur croissance. Pour rendre l'analyse protéomique efficace, une nouvelle méthode de purification de parois et une stratégie de séparation des protéines pariétales ont été établies. Les protéines ont été identifiées par cartographie peptidique massive en utilisant la spectrométrie de masse de type MALDI-TOF. Cette étude a permis d'identifier 137 protéines parmi lesquelles 51 n'avaient pas été identifiées auparavant. Plusieurs familles de protéines connues pour être impliquées dans l'élongation cellulaire ou son arrêt ont été trouvées par l'une ou l'autre approche (XTH, PG, expansines, peroxydases, laccases). De nouvelles protéines candidates pour jouer des rôles dans l'élongation cellulaire ont été identifiées, telles des protéases, des protéines liées au métabolisme des lipides ou des protéines de fonction inconnue. La comparaison des résultats de protéomique et de transcriptomique ne montre pas de cohérence systématique, suggérant l'existence de mécanismes de régulation post-transcriptionnels des gènes codant les protéines pariétales.

**Mots-clefs :**

*Arabidopsis thaliana*, développement, protéines pariétales, paroi cellulaire végétale, protéomique, transcriptomique

**Dynamics of cell wall proteins during cell elongation in etiolated hypocotyls  
of *Arabidopsis thaliana* as shown by proteomic and transcriptomic surveys**

**Abstract**

The cell wall of higher plants is a complex dynamic entity that performs a variety of functions during growth and development as well as in response to environmental stresses. Cell wall proteins play important roles in cell elongation. Etiolated hypocotyls of *Arabidopsis thaliana* were chosen as elongating organs since they undergo rapid and polar elongation without cell division. Two developmental stages were compared through proteomic and transcriptomic surveys: active elongation (5-day-old hypocotyls) vs after growth arrest (11 day-old hypocotyls). As a prerequisite to proteomic study, an efficient cell wall purification method was established as well as a powerful strategy for protein separation. Proteins were identified by peptide mass mapping using MALDI-TOF mass spectrometry and bioinformatics. This study resulted in identification of 137 proteins among which 51 were not identified before by proteomics, and 36 (resp. 17) were found only at 5-days (resp. 11-days). Overall, many protein families expected to be involved in cell wall elongation or in elongation arrest were found by either survey such as XTHs, PGs, expansins, peroxidases, and laccases. However, some of them were found all along hypocotyl development which suggests other roles in cell walls. In addition, many proteins not described before to play roles during cell wall elongation were identified, such as proteases, proteins related to lipid metabolism and proteins of yet unknown function. Finally, proteomic and transcriptomic results revealed no systematic consistency between them suggesting post-transcriptional regulation steps for genes encoding cell wall proteins.

**Key words:**

*Arabidopsis thaliana*, development, cell wall proteins, plant cell wall, proteomics, transcriptomics