# Analysis of the xylem sap proteome of Brassica oleracea reveals a high content in secreted proteins.

Laetitia Ligat, Emmanuelle Lauber, Cécile Albenne, Hélène San Clemente, Benoît Valot, Michel Zivy, Rafael Pont-Lezica, Matthieu Arlat, Elisabeth Jamet

Published in Proteomics (2011) 11 : 1798-1813

# Analysis of the xylem sap proteome of *Brassica oleracea* reveals a high content in secreted proteins

*Laetitia Ligat[1,2]*, Emmanuelle Lauber[3]*, Cécile Albenne[1,2], Hélène San Clemente[1,2], Benoît Valot[4], Michel Zivy[5], Rafael Pont-Lezica[1,2], Matthieu Arlat[3,6] and Elisabeth Jamet[1,2]*

* These authors equally contributed to the work.

[1.] Université de Toulouse; UPS; UMR 5546, Surfaces Cellulaires et Signalisation chez les Végétaux; BP 42617, F-31326 Castanet-Tolosan, France
[2.] CNRS; UMR 5546; BP 42617, F-31326 Castanet-Tolosan, France
[3.] Laboratoire des Interactions Plantes Microorganismes (LIPM); UMR CNRS-INRA 2594-441; F-31320 Castanet-Tolosan, France.
[4.] CNRS, PAPPSO, UMR 0320 / UMR 8120 Génétique Végétale, F-91190 Gif sur Yvette, France
[5.] INRA, PAPPSO, UMR 0320 / UMR 8120 Génétique Végétale, F-91190 Gif sur Yvette, France
[6.] Université de Toulouse; UPS; 118 route de Narbonne; F-31062 Toulouse, France.

**Correspondence**: Dr Elisabeth Jamet, UMR 5546 UPS/CNRS, Surfaces Cellulaires et Signalisation chez les Végétaux; BP 42617, F-31326 Castanet-Tolosan, France
**E-mail**: jamet@lrsv.ups-tlse.fr
**Fax**: +33(0)534 32 38 02

**Keywords**: *Arabidopsis thaliana* / Bioinformatics / *Brassica oleracea* / Cell wall protein / Glycoproteome / Xylem sap

**Abbreviations**: AGP, arabinogalactan protein; FLA, fasciclin AGP; GH, glycoside hydrolase; GRP, glycine-rich protein; LTP, lipid transfer protein; XTH, endotransglucosylase-hydrolase

## Abstract

Xylem plays a major role in plant development, and is considered part of the apoplast. Here we studied the proteome of *Brassica oleracea cv* Bartolo and compared it to the plant cell wall proteome of another *Brassicaceae*, the model plant *Arabidopsis thaliana*. *B. oleracea* was chosen because it is technically difficult to harvest enough *A. thaliana* xylem sap for proteomic analysis. We studied the whole proteome and an *N*-glycoproteome obtained after Concanavalin A affinity chromatography. Altogether, 189 proteins were identified by LC-MS/MS using *Brassica* EST and cDNA sequences. A predicted signal peptide was found in 164 proteins suggesting that most proteins of the xylem sap are secreted. Eighty one proteins were identified in the *N*-glycoproteome, with 25 of them specific of this fraction, suggesting that they were concentrated during the chromatography step. All the protein families identified in this study were found in the cell wall proteomes. However proteases and oxido-reductases were more numerous in the xylem sap proteome, whereas enzyme inhibitors were rare. The origin of xylem sap proteins is discussed. All the experimental data including the MS/MS data were made available in the *WallProtDB* cell wall proteomic database.

## 1 Introduction

Xylem tissue is a major component of the vascular system of plants with a critical role in the transport of water, minerals and nutrients [1]. It is composed of tracheary elements, parenchyma cells, and fibers. During their differentiation, tracheary elements lose their nuclei and cellular content. A lignified secondary wall is formed. At the end of the process, cell death occurs, leaving a hollow tube which becomes a part of a vessel [2]. Xylem sap was shown to contain small molecular weight inorganic compounds and organic substances including hormones, amino acids, sugars, oligo- and polysaccharides, and proteins (for a review, see [3]). The question of the origin of xylem sap proteins was discussed and it was assumed that they could be breakdown products occurring during xylem formation or

protein [11]. On the contrary, a major Cys-rich protein of the healthy tomato xylem sap was shown to disappear after infection by *Fusarium oxysporum* [14].

A description of the xylem sap proteome of the model plant *Arabidopsis thaliana* should help us to better characterize this important compartment of the plant and might allow the characterization of determinants limiting vascular pathogen infection or facilitating their growth. However, such an analysis is still missing, probably because efficient xylem sap harvesting in *A. thaliana* is technically difficult. Such an analysis should complete proteomic studies on this model plant [18-27]. In this study, we took advantage of the close genetic vicinity between *A. thaliana* and another *Brassicaceae*, *Brassica oleracea* to identify proteins of the xylem sap. *B. oleracea* has two main advantages: (i) the diploid *Brassica* species are descended from an hexaploid ancestor and the genome of *A. thaliana* is similar to each of their hypothetical diploid progenitors and (ii) it is one (2n=18, CC genome) of the two ancestors of the *B. napus* amphidiploid (2n=38, AACC genome) [28] for which a systematic program of EST sequencing has been developed (http://compbio.dfci.harvard.edu/tgi/gi/bngi/GenInfo.html). Two different proteomes of *B. oleracea* xylem sap were studied: a xylem sap proteome and the xylem sap *N*-glycoproteome because *N*-glycosylation was assumed to be a major post-translational modification (PTM) that occurs in secreted proteins. The identified proteins are homologous to previously described cell wall proteins, except that no structural protein was identified [29]. The origin of xylem sap proteins is discussed using *A. thaliana* root transcriptomic data available online.

## 2 Material and methods

### 2.1 Xylem sap harvesting

Xylem sap harvesting method from the cultivated plant Bartolo cabbage (*Brassica oleracea* var. *capitata*) was adapted from [30]. Harvesting was found to be optimal for 6-8 week-old plants. Briefly, stems were cut with a razor blade 2-3 cm above the cotyledons and below the first leaves. Before sampling from the remaining stem on the root side, the cut surface was washed with water to remove the content of cut cells and the phloem sap which exudes after cutting, and gently dried with a paper towel. Aliquots of xylem sap was collected in a tube using a micropipette (Supporting information Fig. 1) and stored at –20°C immediately after harvesting. All the xylem sap aliquots were pooled prior to further analysis. Before use, the samples were filtered using 0.45 µm Millipore filters (Carrigtwohill, Ireland), to discard soil

particles, microbial cells or tissue remnants. After 8 h-sampling, we obtained from 0.3 to 0.7 mL of xylem sap from one plant. The experiment was performed twice.

### 2.2 Preparation of the protein samples for LC-MS/MS analysis

After harvesting the *B. oleracea* xylem sap, the sample was dialyzed against buffer 1 (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$/$MnCl_2$/$CaCl_2$ ) in a *Mega GeBAflex-tube* (MWCO 12-14 kDa, GeBA, Yavne, Israel). Half of this sample, the "xylem sap proteome", was desalted using Econo-Pac 10 DG columns (Bio-Rad, Hercules, CA) and lyophilized. The second half of the sample used to get the "xylem sap *N*-glycoproteome" was directly separated by affinity chromatography on Concanavalin A (ConA) (Sigma, St Louis, MO). ConA lectin affinity chromatography is specific for Man residues and allows specific capture of *N*-glycoproteins [19]. Briefly, the resin was pre-washed with 20X volume of buffer 2 (20 mM Tris-HCl pH 7.4, 1 M NaCl, 3.3 mM $MgCl_2$/$MnCl_2$/$CaCl_2$) and equilibrated with 10X volume of buffer 1. The dialyzed xylem sap (10 mL) was mixed with the matrix (0.6 mL) in batch for 1 h at 4°C. After flow-through removal, the resin was washed three times with 1.5 mL of buffer 1. Proteins were eluted with 3X 1.5 mL of buffer 1 supplemented with 1 M methyl-α-D-glucopyranose (Sigma). The first and second eluted fractions were combined and, after desalting using Econo-Pac 10 DG columns (Bio-Rad) and lyophilization, employed to get the xylem sap *N*-glycoproteome.

### 2.3 Separation of proteins by SDS-PAGE

The two protein samples (xylem sap proteome and xylem sap *N*-glycoproteome) were suspended in 300 µL and 100 µL of UHQ water respectively. Fifty µL of each sample were loaded on 10 x 12 x 0.15 cm SDS-polyacrylamide gels with a concentration of 12.50%/ 0.33% of acrylamide/bisacrylamide. Separation was performed as previously described [31]. The gel staining was carried out with Coomassie Brilliant Blue (CBB) [32], silver nitrate [32], or with the β-glucosyl Yariv reagent [33]. The rest of the samples were dried under vacuum prior to LC-MS/MS analysis.

### 2.4 LC-MS/MS analyses

Prior to analysis, proteins were briefly separated by SDS-PAGE to get three samples in order to increase the efficiency of tryptic digestion. In-gel digestion was performed as previously described [34]. Separation of tryptic peptides was performed by HPLC on a NanoLC-Ultra system (Eksigent, Dublin, CA). A 4 µL sample was loaded at 7.5 $\mu L.min^{-1}$ on a precolumn

cartridge (stationary phase: C18 PepMap 100, 5 µm; column: 100 µm inner diameter, 1 cm in length; Dionex, Voisins le Bretonneux, France) and desalted with 0.1% HCOOH. After 3 min, the precolumn cartridge was connected to the separating PepMap C18 column (stationary phase: C18 PepMap 100, 3 µm; column: 75 µm inner diameter, 150 mm in length; Dionex). Buffers were 0.1% HCOOH in water (A) and 0.1% HCOOH in ACN (B). The peptide separation was achieved with a linear gradient from 5 to 30% B for 28 min at 300 $nL.min^{-1}$. One run took 45 min including the regeneration step at 95% B and the equilibration step at 95% A.

Eluted peptides were analyzed on-line with a LTQ XL ion trap (Thermo Electron, Thermo Fisher Scientific Inc, Courtaboeuf, France) using a nano electrospray interface as previously described with slight modifications detailed below [35]. Ionization (1.5 kV ionization potential) was performed with liquid junction and a noncoated capillary probe (10 µm inner diameter; New Objective). Peptide ions were analyzed using Xcalibur 2.07 (Thermo Fisher Scientific Inc) with the following data-dependent acquisition steps: (1) full MS scan (mass-to-charge ratio (m/z) 300 to 1400, centroid mode); and (2) MS/MS (qz = 0.25, activation time = 30 ms, and collision energy = 35%; centroid mode). Step 2 was repeated for the three major ions detected in step 1. Dynamic exclusion was set to 30 s.

A database search was performed with X! Tandem (version 2010.01.01.4) (http://www.thegpm.org/tandem/) using parameters and protein identification specifications previously described [35]. Three databases were used: (i) a *Brassica napus* EST database (Compbio, http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=oilseed_rape); (ii) the Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org/) database (32825 entries, version 8); and (iii) a contaminant database (trypsin, keratins). To take redundancy into account, proteins with at least one peptide in common were grouped. Within each group, proteins with at least one specific peptide relatively to other members of the group were reported as sub-groups.

### 2.5 Bioinformatics

Two databases were used to analyze the cell wall proteome of *B. oleracea*: *ProtAnnDB* (http://www.polebio.scsv.ups-tlse.fr/ProtAnnDB/) for the annotation of *A. thaliana* proteins [36]; and *WallProtDB* which collects *A. thaliana* cell wall proteomes (http://www.polebio.scsv.ups-tlse.fr/WallProtDB/) [37].The *Brassica* proteins were annotated

as previously described for *A. thaliana* proteins [36]. Several available software were used to predict sub-cellular localization and functional domains of proteins: TargetP (http://www.cbs.dtu.dk/services/TargetP/) and SignalP (http://www.cbs.dtu.dk/services/SignalP/) for sub-cellular localization and InterProScan for prediction of functional domains (http://www.ebi.ac.uk/Tools/InterProScan/). The results were combined to improve the quality of the predictions and to propose a structural and a functional annotation. The PROSITE software was used to predict *N*-glycosylation sites (http://www.expasy.org/prosite/). The AREX database was used to look for the root pattern of expression of *A. thaliana* genes homologous to those of *B. oleracea* encoding xylem sap proteins (http://www.arexdb.org/) [38].

### 2.6 Implementation of MS/MS data in *WallProtDB*

This *WallProtDB* knowledgebase was developed with PHP5/AJAX/MySQL5. It presently contains CWPs (476 from *A. thaliana*, 263 from *Oryza sativa*) and ESTs (162 from *B. oleracea*) which were classified as described [29]. *B. oleracea* sequences are linked to their closest homologues in *A. thaliana* as inferred from BLASTX searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi, [39]). For *B. oleracea*, the proteomic data are linked to the MS data allowing protein identification. The spectra files are stored on a GPM server and usable with X! Tandem *via* a simple hyperlink. *WallProtDB* can be queried *via* an html form with various criteria including plant species, organ, and experimental conditions. User can refine their selection if necessary and export the result in a tab delimited text, or export the sequences of interest in the FASTA format.

## 3. Results

### 3.1 Harvesting of xylem sap and separation of proteins by SDS-PAGE

The xylem was collected during 8 h from cut stems of *B. oleracea* (Supporting information Fig. 1). To avoid contamination by phloem sap and intracellular proteins of cut cells, the cut surface was rinsed with water. It was not possible to quantify the amount of proteins with the Bradford reagent probably because of their low concentration [40]. However, CBB stained bands were clearly visible after separation of proteins of the total extract (xylem sap proteome) by SDS-PAGE (Fig. 1, lane A). After lectin affinity chromatography on ConA to separate *N*-glycoproteins (xylem sap *N*-glycoproteome), a distinct profile of proteins was obtained showing enrichment in certain proteins and depletion in others (Fig. 1, lane B). Again, it was not possible to get reliable quantification of proteins after ConA affinity

chromatography. In both cases, the profiles do not suggest any protein degradation. This was confirmed by a good distribution of the peptides allowing protein identification by LC-MS/MS all over the protein amino acid sequences (Supporting information Tables S1 and S2). A staining with silver nitrate showed additional bands of lower molecular masses (Fig.1, lanes C and D). The total extract was also submitted to β-glucosyl Yariv reagent staining to reveal arabinogalactan proteins (AGPs) which are poorly stained by CBB and silver nitrate because of their high degree of glycosylation. A smear was observed at the top of the gel showing the presence of AGPs in the sample (Fig. 1, lane E). Both the total extract and the *N*-glycoproteins retained on ConA were then analyzed by LC-MS/MS.

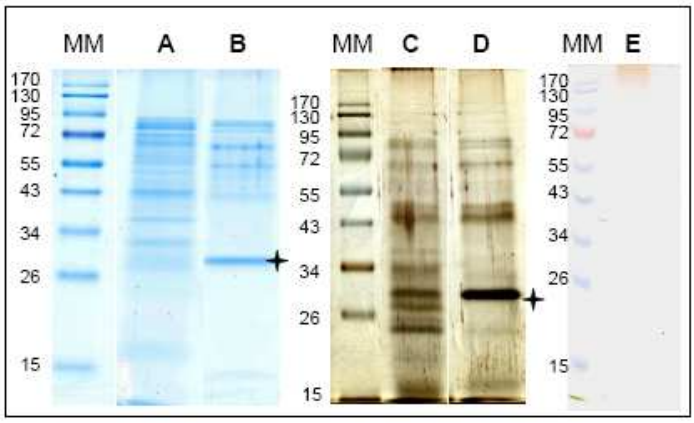


**Figure 1.** Separation of xylem sap proteins by SDS-PAGE.

Proteins from the xylem sap were separated by SDS-PAGE either directly after sampling and dialysis (lanes A and C, xylem sap proteome), or after an additional step of affinity chromatography on ConA (lanes B and D, xylem sap *N*-glycoproteome). The gel was stained with CBB (lanes A and B) or silver nitrate (lanes C and D). Bands shown by a star correspond to the ConA protein leaking from the column. The total extract was also stained by the -glucosyl Yariv reagent (lane E). MM are molecular mass markers (kDa).

### 3.2 Identification by LC-MS/MS of proteins present in the xylem sap proteome and in the xylem sap *N*-glycoproteome trapped on ConA

Most of the previous studies on xylem sap relied on separation of proteins by 2D-electrophoresis prior to MS analysis. However, 2D-electrophoresis can be limiting in the case of secreted proteins which are mostly basic glycoproteins [41]. In our study, proteins were not separated before digestion and liquid chromatography tandem MS (LC-MS/MS). This technique also allowed working with limiting amount of proteins. *B. napus* cDNA and EST sequences were used for proteins identification. In all cases, it was possible to identify

proteins using *Brassica* EST or cDNA sequences. This was a great advantage since it allowed precise identification of the genes encoding the proteins especially in multigene families.

Altogether, 189 proteins were identified by LC-MS/MS with at least two peptides sequenced per protein (Table I, Supporting information Tables S1 and S2). One hundred and sixty four proteins were identified in the xylem sap proteome, whereas 81 proteins were identified in the xylem sap *N*-glycoproteome (Supporting information Table S3). Fifty six proteins were common to both proteomes. Twenty five proteins were only found in the xylem sap *N*-glycoproteome. This fraction was probably enriched in these proteins after selection by the ConA affinity chromatography. Identification of proteins in previously published *B. napus* and *B. oleracea* xylem sap proteomes were done by comparison to heterologous sequences, mainly from *A. thaliana* [4, 7, 8]. We performed a new TBLASTN analysis against *B. napus* ESTs with the peptide sequences and obtained the identification of 45 different proteins. Then it was possible to compare these results to our data. Twenty out of these 45 proteins were also found in our study (see Table I). On the contrary, 25 proteins present in one of these proteomes were not found in ours probably because of different culture conditions.

**Table I. Proteins identified by LC-MS/MS in the xylem sap of *B. oleracea***

| protein number [a] | *B. napus* EST, cDNA or protein accession number [b] | homologue in *A. thaliana* | predicted functional domain [c] | predicted sub-cellular localization [d] |
|---|---|---|---|---|
| **Proteins acting on carbohydrates (48)** | | | | |
| **71** | TC38976 | At3g18080 | GH1 (ß-glucosidase, ß-mannosidase) (AtBGLU44 homologue) | chloroplast: 0.265 / SignalP-NN: 1-26, 0.679 |
| **125-1** | TC20974/TC45939 | At5g44640 | GH1 (ß-glucosidase) (AtBGLU3 homologue) | N-ter missing (secretory pathway: 1-22, 0.811) |
| **125-2** | TC50854 | At2g44450 | GH1 (ß-glucosidase) (AtBGLU15 homologue) | N-ter missing (secretory pathway: 1-22, 0.853) |
| **19-1** | TC36871 | At5g64570 | GH3 (ß-D-xylosidase) (AtXYL4 homologue) | N-ter missing (secretory pathway: 1-38, 0.773) |
| **19-2** | TC20792/CD814082 | At5g64570 | GH3 (ß-D-xylosidase) (AtXYL4 homologue) | mitochondrion: 0.706 / SignalP-NN: 1-33, 0.797 |
| **40** | TC20222/EE464798 | At5g64570 | GH3 (ß-D-xylosidase) (AtXYL4 homologue) | chloroplast: 0.698 / SignalP-NN: 1-44, 0.644 |
| 83 | TC32271 | At5g10560 | GH3 (ß-D-xylosidase) | N-ter missing (secretory pathway: 1-18, 0.311) |
| **102** | TC49601 | At4g33820 | GH10 | N-ter missing (secretory pathway: 1-23, 0.966) |
| 23 | TC32833 | At4g15210 | GH14 | secretory pathway: 1-32, 0.977 |
| 52-1 | TC52964 | At4g25810 | GH16 (xyloglucan:xyloglucosyltransferase) (AtXTH23 homologue) | N-ter missing (secretory pathway: 1-24, 0.988) |
| **52-2** | TC23545 | At4g30270 | GH16 (xyloglucan:xyloglucosyltransferase) (AtXTH24 homologue) | secretory pathway: 1-22, 0.971 |
| 3-1 | TC38745 (*B. juncea*) **[4, 8]** | At3g57240 | GH17 (ß-1,3-glucanase) | secretory pathway: 1-34, 0.960 |
| 3-2 | Q2HZ53 (*B. juncea*) | At3g57240 | GH17 (ß-1,3-glucanase) | N-ter missing (secretory pathway: 1-33, 0.980) |
| 47-1 | TC26578 | At5g42720 | GH17 (ß-1,3-glucanase) | secretory pathway: 1-22, 0.911 |
| 47-2 | TC24955 | At4g34480 | GH17 (ß-1,3-glucanase) | secretory pathway: 1-22, 0.942 |
| 66 | TC32141 | At4g34480 | GH17 (ß-1,3-glucanase) | secretory pathway: 1-22, 0.931 |
| 95 | TC30288 | At5g55180 | GH17 (ß-1,3-glucanase) | secretory pathway: 1-22, 0.829 |
| 128 | Q2VT22 (*B. rapa*) | At3g57260 | GH17 (ß-1,3-glucanase) | secretory pathway: 1-37, 0.713 |
| 159 | TC49994 | At3g55430 | GH17 (ß-1,3-glucanase) | N-ter missing (secretory pathway: 1-28, 0.973) |
| **28** | TC17043 | At5g24090 | GH18 (chitinase) | secretory pathway: 1-27, 0.985 |
| 88 | TC54505 | At4g19810 | GH18 (chitinase) | N-ter missing (secretory pathway: 1-24, 0.944) |
| 8 | TC58115 **[4]** | At2g43590 | GH19 (chitinase) | secretory pathway: 1-33, 0.938 |
| 110 | TC15913 | At3g12500 | GH19 (chitinase) | secretory pathway: 1-20, 0.989 |
| **116** | TC25879 | At3g55260 | GH20 (*N*-acetyl-hexosaminidase) | N-ter missing (secretory pathway: 1-20, 0.805) |
| **56** | TC37303 | At3g56310 | GH27 | secretory pathway: 1-23, 0.689 |
| 144 | EV101247 | At5g08370 | GH27 | N-ter missing (secretory pathway: 1-25, 0.909) |
| **26-1** | TC25741/TC46151 **[4]** | At3g61490 | GH28 (polygalacturonase) | N-ter missing (secretory pathway: 1-23, 0.765) |

| | | | | |
|---|---|---|---|---|
| 26-2 | EV178995 | At3g61490 | GH28 (polygalacturonase) | N-ter missing (secretory pathway: 1-23, 0.765) |
| **26-3** | EV121625 | At3g23500 | GH28 (polygalacturonase) | N-ter missing (secretory pathway: 1-22, 0.948) |
| 27 | TC23229 | At4g23500 | GH28 (polygalacturonase) | secretory pathway: 1-25, 0.613 |
| **81-1** | TC21367 | At5g49215 | GH28 (polygalacturonase) | N-ter missing (secretory pathway: 1-21, 0.825) |
| 81-2 | TC28593 | At3g62110 | GH28 (polygalacturonase) | N-ter missing (secretory pathway: 1-21, 0.946) |
| 107 | EE453748 | At5g41870 | GH28 (polygalacturonase) | N-ter missing (secretory pathway: 1-21, 0.989) |
| **97** | TC60884 | At5g11720 | GH31 (α-glucosidase) (AtGLU1 homologue) | N ter missing (secretory pathway: 1-23, 0.976) |
| **65** | TC31486/TC29838 | At2g28470 | GH35 (ß-D-galactosidase) (AtBGAL8 homologue) | secretory pathway: 1-23, 0.991 |
| 35 | TC47393 | At5g13980 | GH38 ( -D-mannosidase) | secretory pathway: 1-23, 0.955 |
| 112 | TC17346 | At5g13980 | GH38 ( -D-mannosidase) | N-ter missing (secretory pathway: 1-21, 0.953) |
| **2-1** | TC43276 | At3g10740 | GH51 ( -L-arabinofuranosidase/ß-D-xylosidase) (AtASD1 homologue) | N-ter missing (secretory pathway: 1-33, 0.962) |
| **2-2** | TC43393 | At3g10740 | GH51 ( -L-arabinofuranosidase/ß-D-xylosidase) (AtASD1 homologue) | N-ter missing (secretory pathway: 1-33, 0.962) |
| **7-1** | TC53916 | At3g10740 | GH51 ( -L-arabinofuranosidase/ß-D-xylosidase) (AtASD1 homologue) | N-ter missing (secretory pathway: 1-33, 0.962) |
| 7-2 | TC61736 | At3g10740 | GH51 ( -L-arabinofuranosidase/ß-D-xylosidase) (AtASD1 homologue) | secretory pathway: 1-33, 0.943 |
| **25** | TC31356 | At5g34940 | GH79 (ß-D-glucuronidase) | secretory pathway: 1-21, 0.988 |
| 122 | TC41425 | At5g09760 | pectin methylesterase (CE8) | N-ter missing (SignalP-NN: 1-20, 0.803) |
| 145 | AY036606 | At2g47550 | pectin methylesterase (CE8) | secretory pathway: 1-23, 0.860 |
| 136 | CD814593 | At3g04910 | pectin acylesterase (CE17) | N-ter missing (SignalP-NN: 1-44, 0.417) |
| 91 | TC16095 | At5g45280 | pectin acylesterase (CE23) | secretory pathway: 1-23, 0.989 |
| 51 | TC16080 | At4g17030 | expansin-like B (AtEXLB1 homologue) | secretory pathway: 1-24, 0.988 |
| **138** | DY023690 | At5g02260 | alpha-expansin (AtEXPA9) | N-ter missing (secretory pathway: 1-21, 0.949) |
| **Oxido-reductases (39)** | | | | |
| **58** | TC47947 | At1g05260 | peroxidase (AtPrx03 homologue) | secretory pathway: 1-24, 0.970 |
| **99** | TC33540 | At1g71695 | peroxidase (AtPrx12 homologue) | secretory pathway: 1-31, 0.856 |
| 64 | TC25610 | At2g22420 | peroxidase (AtPrx17 homologue) | secretory pathway: 1-20, 0.762 |
| 43 | TC36205 **[4, 8]** | At2g37130 | peroxidase (AtPrx21 homologue) | secretory pathway: 1-29, 0.981 |
| **20-1** | TC30566 | At4g08770 | peroxidase (AtPrx37 homologue) | secretory pathway: 1-29, 0.514 |
| **20-2** | TC27727 | At3g32980 | peroxidase (AtPrx32 homologue) | chloroplast: 1-54, 0.309/SignalP-NN: 1-30, 0.684 |
| **20-3** | TC25612 | At3g49120 | peroxidase (AtPrx34 homologue) | secretory pathway: 1-32, 0.820 |
| **20-4** | NP1447930 | At3g49120 | peroxidase (AtPrx34 homologue) | secretory pathway: 1-31, 0.525 |
| **77** | TC18276 | At4g11290 | peroxidase (AtPrx39 homologue) | secretory pathway: 1-23, 0.962 |
| 49-1 | TC58934 **[4]** | At4g33420 | peroxidase (AtPrx47 homologue) | secretory pathway: 1-30, 0.772 |
| 49-2 | TC61099 **[4]** | At4g33420 | peroxidase (AtPrx47 homologue) | secretory pathway: 1-25, 0.920 |
| **13-1** | TC35223 **[7]** | At5g05340 | peroxidase (AtPrx52 homologue) | secretory pathway: 1-21, 0.796 |

| | | | | |
|---|---|---|---|---|
| **13-2** | TC45523 **[4]** | At5g05340 | peroxidase (AtPrx52 homologue) | N-term missing (secretory pathway: 1-20, 0.916) |
| 13-3 | P00434 (*B. rapa*) | At5g05340 | peroxidase (AtPrx52 homologue) | N-term missing (secretory pathway: 1-20, 0.916) |
| **11-1** | TC55805 **[4]** | At5g19890 | peroxidase (AtPrx59 homologue) | secretory pathway: 1-25, 0.739 |
| **11-2** | EE424445 **[7]** | At5g39580 | peroxidase (AtPrx62 homologue) | secretory pathway: 1-22, 0.962 |
| 11-3 | TC16466 | At5g64120 | peroxidase (AtPrx71 homologue) | secretory pathway: 1-23, 0.719 |
| **59** | TC48654 | At4g12420 | multicopper oxidase (AtSKU5 homologue) | N-ter missing (secretory pathway: 1-20, 0.989) |
| **111** | TC16200 | At4g12420 | multicopper oxidase (AtSKU5 homologue) | secretory pathway: 1-20, 0.990 |
| **22-1** | TC16404 | At1g41830 | multicopper oxidase (AtSKS6 homologue) | secretory pathway: 1-24, 0.970 |
| **22-2** | TC38876 | At1g76160 | multicopper oxidase (AtSKS5 homologue) | secretory pathway: 1-23, 0.991 |
| **22-3** | TC37557 | At1g76160 | multicopper oxidase (AtSKS5 homologue) | N-ter missing (secretory pathway: 1-23, 0.984) |
| **22-4** | TC19179 | At1g76160 | multicopper oxidase (AtSKS5 homologue) | secretory pathway: 1-23, 0.989 |
| 60-1 | TC39082 | At1g72230 | uclacyanin (AtUCC8 homologue) | secretory pathway: 1-25, 0.972 |
| 60-2 | TC20750 | At1g72230 | uclacyanin (AtUCC8 homologue) | secretory pathway: 1-24, 0.985 |
| 117 | TC26850 | At2g02850 | plantacyanin (AtPNC homologue) | secretory pathway: 1-33, 0.825 |
| **38** | TC29181 | At2g25060 | plastocyanin (AtEN7 homologue) | secretory pathway: 1-23, 0.988 |
| 39-1 | TC32771 | At4g27520 | plastocyanin (AtEN12 homologue) | secretory pathway: 1-27, 0.975 |
| 39-2 | TC47186 | At4g27520 | plastocyanin (AtEN12 homologue) | secretory pathway: 1-28, 0.967 |
| 63-1 | TC21582 | At4g31840 | plastocyanin (AtEN13 homologue) | secretory pathway: 1-22, 0.989 |
| 63-2 | TC17946 | At4g31840 | plastocyanin (AtEN13 homologue) | secretory pathway: 1-21, 0.986 |
| 63-3 | TC23721 | At4g31840 | plastocyanin (AtEN13 homologue) | secretory pathway: 1-25, 0.991 |
| 130 | TC21566 | At5g15350 | plastocyanin (AtEN22 homologue) | secretory pathway: 1-26, 0.970 |
| **153** | TC25008 | At5g15350 | plastocyanin (AtEN22 homologue) | secretory pathway: 1-24, 0.981 |
| **5** | TC18996 | At4g20840 | berberine-bridge oxido-reductase | secretory pathway: 1-26, 0.872 |
| 15-1 | TC17262 | At1g30760 | berberine-bridge oxido-reductase | secretory pathway: 1-29, 0.979 |
| **15-2** | TC38473 | At2g34790 | berberine-bridge oxido-reductase | N-ter missing (secretory pathway: 1-26, 0.989) |
| 142 | ES903641 | At4g20840 | berberine-bridge oxido-reductase | secretory pathway: 1-23, 0.803 |
| 143 | EV087794 | At5g44390 | berberine-bridge oxido-reductase | N-ter missing (secretory pathway: 1-30, 0.825) |
| **Proteases (28)** | | | | |
| **1-1** | TC18683/TC28749/ TC22282 **[4]** | At5g67360 | Ser protease (ATSBT1.7, ARA12 homologue) | secretory pathway: 1-29, 0.977 |
| **1-2** | TC19435 | At5g67360 | Ser protease (ATSBT1.7, ARA12 homologue) | N-ter missing (secretory pathway: 1-24, 0.973) |
| **12/33** | ES901600/TC56831/ TC46249 **[4]** | At1g20160 | Ser protease (ATSBT5.2 homologue) | secretory pathway: 1-29, 0.991 |
| **29-1** | TC27270 | At4g36195 | Ser carboxypeptidase | secretory pathway: 1-21, 0.704 |
| **29-2** | TC25039 | At4g36190 | Ser carboxypeptidase | secretory pathway: 1-25, 0.595 |
| **75** | TC27534 | At4g12910 | Ser carboxypeptidase (AtSCPL20 homologue) | secretory pathway: 1-23, 0.988 |
| 30-1 | TC34104 | At4g30810 | Ser carboxypeptidase (AtSCPL29 homologue) | secretory pathway: 1-27, 0.992 |
| 30-2 | TC47262 | At4g30810 | Ser carboxypeptidase (AtSCPL29 homologue) | secretory pathway: 1-25, 0.984 |

| | | | | |
|---|---|---|---|---|
| **69** | EV124833 **[7]** | At1g11080 | Ser carboxypeptidase (AtSCPL31 homologue) | N-ter missing (secretory pathway: 1-30, 0.902) |
| **74** | EV124923 | At1g11080 | Ser carboxypeptidase (AtSCPL31 homologue) | N-ter missing (secretory pathway: 1-30, 0.902) |
| **14** | TC45864 | At3g45010 | Ser carboxypeptidase (AtSCPL48 homologue) | secretory pathway: 1-25, 0.603 |
| **62** | TC23586/EV048732 | At2g27920 | Ser carboxypeptidase (AtSCPL51 homologue) | secretory pathway: 1-20 0.810 |
| **18-1** | TC21318 | At5g10770 | Asp protease | N-ter missing (secretory pathway: 1-25, 0.276) |
| 18-2 | Q8LK82 | At5g10770 | Asp protease | N-ter missing (secretory pathway: 1-25, 0.276) |
| 18-3 | ES904551 | At5g10770 | Asp protease | N-ter missing (secretory pathway: 1-25, 0.276) |
| 18-4 | EE486131 | At5g10770 | Asp protease | N-ter missing (secretory pathway: 1-25, 0.276) |
| **44** | EV196557/TC46757 | At5g10770 | Asp protease | N-ter missing (secretory pathway: 1-25, 0.276) |
| 93-1 | TC21632 | At3g18490 | Asp protease | N-ter missing (SignalP-NN: 1-24, 0.895) |
| 93-2 | TC23324 | At3g18490 | Asp protease | chloroplast: 0.342 / SignalP-NN: 1-24, 0.833 |
| 93-3 | TC33329 | At3g18490 | Asp protease | secretory pathway: 1-26, 0.456 |
| **100** | TC17031 | At5g07030 | Asp protease | secretory pathway: 1-19, 0.805 |
| **115** | TC23722 | At3g54400 | Asp protease | secretory pathway: 1-19, 0.294 |
| **120** | TC33693 | At1g79720 | Asp protease | secretory pathway: 1-28, 0.967 |
| 96-1 | B1Q3A2 | At1g47128 | Cys protease (RESPONSIVE TO DEHYDRATION, RD21 homologue) | N-ter missing (secretory pathway: 1-21, 0.993) |
| 96-2 | TC59035 | At5g43060 | Cys protease | secretory pathway: 1-20, 0.987 |
| **104** | TC53234 | At5g60360 | Cys protease (ALEURAIN LIKE PROTEASE, AALP homologue) | secretory pathway: 1-22, 0.976 |
| 119 | TC33308 **[4]** | At4g35350 | Cys protease (XYLEM CYSTEINE PEPTIDASE 1, XCP1 homologue) | secretory pathway: 1-29, 0.986 |
| 134 | TC26010 | At4g01610 | Cys protease | secretory pathway: 1-29, 0.979 |
| **Lipid metabolism (8)** | | | | |
| 21 | TC36034 **[7]** | At3g53980 | lipid transfer protein | secretory pathway: 1-23, 0.936 |
| 132 | TC19947 | At2g44300 | lipid transfer protein | secretory pathway: 1-23, 0.937 |
| 57 | TC45634 | At5G05960 | lipid transfer protein | secretory pathway: 1-25, 0.720 |
| 98 | O82582 (*B. oleracea*) | At2g38540 | lipid transfer protein (LTP1) | secretory pathway: 1-25, 0.970 |
| 68 | Q9S9F9 | At2g38530 | lipid transfer protein (LTP2) | N-ter missing (secretory pathway: 1-23, 0.967) |
| 73 | EG019134 | At5g59310 | lipid transfer protein (LTP4) | N-ter missing (secretory pathway: 1-23, 0.976) |
| **139** | DY024942 | At1g09390 | lipase acylhydrolase (GDSL family) | secretory pathway: 1-27, 0.983 |
| 160 | TC50035 | At4g26690 | glycerophosphoryl diester phosphodiesterase (MORPHOGENESIS OF ROOT 5, MHR5 homologue) (GPI-anchored protein) | N-ter missing (SignalP-NN : 1-27, 0.761) |
| **Signaling (8)** | | | | |
| **16-1** | TC35372 | At5g55730 | FLA (AtFLA1 homologue) | secretory pathway: 1-26, 0.799 |
| **16-2** | TC31355 | At5g55730 | FLA (AtFLA1 homologue) | secretory pathway: 1-26, 0.799 |

| | | | | |
|---|---|---|---|---|
| **46** | TC39357 | At4g12730 | FLA (AtFLA2 homologue) | secretory pathway: 1-23, 0.817 |
| **37** | TC21253 | At2g04780 | FLA (AtFLA7 homologue) | secretory pathway: 1-23, 0.882 |
| **24-1** | TC21238 | At2g45470 | FLA (AtFLA8 homologue) | N-ter missing (secretory pathway: 1-25, 0.936) |
| 24-2 | TC30763 | At3g60900 | FLA (AtFLA10 homologue) | N-ter missing (secretory pathway: 1-25, 0.869) |
| **61-1** | TC32307 | At1g03870 | FLA (AtFLA9 homologue) | secretory pathway: 1-25, 0.912 |
| **61-2** | TC23584 | At1g03870 | FLA (AtFLA9 homologue) | secretory pathway: 1-25, 0.912 |
| **45** | TC26664 | At5g03170 | FLA (AtFLA11 homologue) | secretory pathway: 1-24, 0.815 |
| **Interacting domains (8)** | | | | |
| 54 | ES901495 | At1g17860 | protease inhibitor | N-ter missing (secretory pathway: 1-19, 0.910) |
| 163 | B2KSD1 (*B. oleracea*) | At1g73260 | protease inhibitor I3 (Kunitz legume) | N-ter missing (secretory pathway: 1-28, 0.985) |
| 113 | TC20389 **[4]** | At1g47960 | pectinesterase inhibitor (PMEI) | secretory pathway: 1-20, 0.974 |
| **17** | TC40783 | At1g03220 | xylanase inhibitor I (EDGP) | secretory pathway: 1-23, 0.917 |
| **53** | TC32327 **[4]** | At1g78850 | lectin (curculin-like, mannose binding) | secretory pathway: 1-22, 0.990 |
| 67-105 | TC53636 **[4]** | At1g78830 | lectin (curculin-like, mannose binding) | secretory pathway: 1-22, 0.962 |
| 123 | TC41889 | At1g26450 | unknown function (X8 domain) | secretory pathway: 1-19, 0.809 |
| 114 | TC22255 | At1g21880 | unknown function (LysM domain) | secretory pathway: 1-27, 0.775 |
| **Miscellaneous (14)** | | | | |
| **36** | TC59434 | At4g23690 | dirigent protein | secretory pathway: 1-29, 0.994 |
| **146** | TC16085 | At1g13900 | purple acid phosphatase (AtPAP2 homologue) | N-ter missing (secretory pathway: 1-19, 0.983) |
| 42 | TC32617 | At3g07130 | purple acid phosphatase (AtPAP15 homologue) | N-ter missing (secretory pathway: 1-19, 0.859) |
| **72** | TC62786 | At3g07130 | purple acid phosphatase (AtPAP15 homologue) | N-ter missing (secretory pathway: 1-19, 0.859) |
| **10-1** | TC21902 | At5g34850 | purple acid phosphatase (AtPAP26 homologue) | secretory pathway: 1-27, 0.991 |
| **10-2** | TC50714 | At5g34850 | purple acid phosphatase (AtPAP26 homologue) | secretory pathway: 1-22, 0.772 |
| 55 | TC25081 | At3g19320 | thaumatin | secretory pathway: 1-26, 0.974 |
| 82 | TC28594 | At2g14610 | cysteine-rich secretory protein (SCP) (PR1 homologue) | secretory pathway: 1-26, 0.798 |
| 86 | TC38503 **[4]** | At4g34180 | cyclase | secretory pathway: 1-23, 0.922 |
| **106** | TC55340 | At3g62020 | germin (GLP10 homologue) | N-ter missing (secretory pathway: 1-21, 0.970) |
| 126 | TC54095 **[4]** | At4g11650 | osmotin (AtOSM34 homologue) | secretory pathway: 1-26, 0.991 |
| 41-1 | TC22018 | At2g02990 | ribonuclease T2 (AtRNS1 homologue) | secretory pathway: 1-26, 0.995 |
| 41-2 | TC22953 | At1g26820 | ribonuclease T2 (AtRNS3 homologue) | secretory pathway: 1-19, 0.983 |
| 154 | TC33600 | At1g78680 | homologous to gamma-glutamyl hydrolase | secretory pathway: 1-24, 0.672 |
| **Unknown function (10)** | | | | |
| **4** | TC27548 | At5g48540 | unknown function (DUF26) | secretory pathway: 1-20, 0.930 |
| **32** | EV115188 | At3g22060 | unknown function (DUF26) | secretory pathway: 1-27, 0.880 |

| | | | |
|---|---|---|---|
| **48** | TC55540 | At1g07390 | unknown function (DUF568) (AIR12, AUXIN-INDUCED IN |

**a.** Protein numbers refer to Supporting information Tables S1, S2, and S3. Protein numbers in bold refer to proteins identified in the xylem sap *N*-glycoproteome.
**b.** Nucleotide or amino acid sequences can be found either in the CompBio (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=oilseed_rape) or the NCBI (http://www.ncbi.nlm.nih.gov/) dabases. Sequences are from *B. napus* otherwise stated. Numbers between brackets refer to previously published xylem sap proteomes.
**c.** Functional domains were found as described in Material and methods. GH were annotated according to CAZy (http://www.cazy.org/) and [59].
**d.** Prediction of sub-cellular localization was done with TargetP (http://www.cbs.dtu.dk/services/TargetP/) and SignalP (http://www.cbs.dtu.dk/services/SignalP/). When both predictions are consistent, only the TargetP result is shown (size of the predicted signal peptide, score). When it is not the case, both predictions are shown. When the *B. oleracea* protein sequence is not complete at its N-terminus (N-ter), the prediction for the closest homologue in *A. thaliana* is shown.

### 3.3 General features of proteins identified in the *B. oleracea* xylem sap

All the *Brassica* protein sequences were analyzed with bioinformatics software to predict their sub-cellular localization and the presence of functional domains. The same work was done for *A. thaliana* proteins homologous to *Brassica* sequences. The proteins could then be classified in (i) intracellular when they were devoid of signal peptide or of any signal targeting them to an intracellular compartment, and (ii) secreted proteins (Table I, Supporting information Table S3). Twenty five proteins (13%) were predicted to be intracellular whereas 164 proteins (87%) were predicted to be secreted. The latter proteins could be distributed in eight of the nine functional classes previously defined for *A. thaliana* cell wall proteins [29]: proteins acting on carbohydrates (29.2%), oxido-reductases (23.8%), proteases (17.1%), proteins related to lipid metabolism (4.9%), proteins involved in signaling (5.5%), proteins with domains interacting with carbohydrates or proteins (4.9%), miscellaneous proteins having diverse functions (8.5%), and proteins with yet unknown function (6.1%). No structural protein was identified in these analyses. Among the 48 proteins acting on carbohydrates, two glycoside hydrolase (GH) families were well represented: eight proteins belonged to GH17 (β-1,3-glucosidases), and seven to GH28 (polygalacturonases). Nearly half of the oxido-reductases were peroxidases (17), the others being proteins homologous to plastocyanins (11), multicopper oxidases (6), and berberine bridge enzymes (5). Proteases were mainly predicted as Ser proteases (12) and Asp proteases (11). Most of the proteins included in the lipid metabolism functional class were lipid transfer proteins (LTPs). Finally, all the proteins included in the signaling functional class were fasciclin AGPs (FLAs). Finding such proteins was expected because of the positive Yariv staining observed after analysis of xylem sap proteins by SDS-PAGE (Fig. 1).

All the proteins of the xylem sap proteome were also analyzed for the presence of putative *N*-glycosylation sites. Proteins homologous to LTPs, ribonuclease T2, and osmotin had no predicted *N*-glycosylation sites and were only found in the xylem sap proteome. All the other proteins had predicted *N*-glycosylation sites. However, 70 proteins predicted to be secreted and having *N*-glycosylation sites were only found in the xylem sap proteome. One cannot exclude the possibility that they were poorly or not *N*-glycosylated consistent with the fact that they were lost during the affinity chromatography step. Interestingly, although 17 out of 25 proteins predicted to be intracellular had predicted *N*-glycosylation sites, none was retained on the ConA affinity column. It suggested that the *N*-glycosylation sites of these proteins were not occupied.

LC-MS/MS data allow getting some information on the abundance of proteins relying on spectral counting. The rationale is that the number of spectra collected for a protein in a MS/MS run is correlated to the abundance of the protein (Supporting information Tables S4). Among the proteins predicted to be intracellular, two proteins homologous to methionine synthase (proteins 9-1, 9-2) were identified with 35 and 24 spectra respectively, a protein homologous to jacalin (protein 31) with 11 spectra, and a protein homologous to peptidase M1 (protein 34) with 10 spectra. Ten out of the remaining 18 proteins predicted to be intracellular were identified with only 2 spectra which meant that they were probably present at a low level. In two independent experiments, the most abundant proteins predicted to be secreted were the following: two Ser proteases (proteins 1-1 and 12 identified with 74 and 25 spectra respectively in the reported experiment); an Asp protease (protein 18-1; 20 spectra); two α-arabinofuranosidases (proteins 2-1 and 7-1; 43 and 29 spectra respectively); a protein homologous to peroxidase AtPrx59 (protein 11-1; 23 spectra); three berberine-bridge oxido-reductases (proteins 5, 15-1 and 15-2; 36, 23 and 20 spectra respectively); a protein homologous to AtFLA1 (protein 16-1; 20 spectra); a protein homologous to purple acid phosphatase AtPAP26 (protein 10; 24 spectra); a protein of yet unknown function with a DUF26 domain (protein 4; 32 spectra). Altogether, these results show that the proteins predicted to be intracellular are mostly minor components of the xylem sap proteome.

### 3.4 Extension and screening of *WallProtDB* with the *B. oleracea* xylem sap proteome

As mentioned above, *WallProtDB* was built up to collect plant cell wall proteomic data [37]. It was initially devoted to *A. thaliana* and *O. sativa* because (i) their genomes are fully sequenced thus allowing the precise identification of proteins by MS and bioinformatics, and (ii) many proteomic studies were published on different organs. *WallProtDB* displays information about experimental conditions, and cell wall proteomes previously analyzed. Each protein accession number in *WallProtDB* has been linked to the *ProtAnnDB* database including bioinformatics predictions about subcellular localization, functional domains, and gene networking [36]. Literature dealing with plant cell wall proteomics has been included.

Two novelties were introduced to describe the *B. oleracea* xylem sap proteome and to allow comparisons with the *A. thaliana* cell wall proteomes characterized in different organs (Supporting information Fig. 2). First, considering the closeness of both genomes, each *B. oleracea* EST is linked to the closest *A. thaliana* gene as determined by BLASTX analysis. In many cases, two or more *B. oleracea* proteins corresponded to a single *A. thaliana* gene were

found thanks to the presence of proteotypic peptides. This is consistent with the evolution of the *Brassica* genomes [42]. Second, LC-MS/MS peptide sequencing data allowing identification of *B. oleracea* proteins encoded by ESTs are provided using the X! Tandem software. The position of sequenced peptides on the translated EST or cDNA sequence can be visualized as well as MS/MS spectra and data (Supporting information Fig. 3). All this information provides a comprehensive view of the identification process using MS data.

A search was performed in *WallProtDB*, a cell wall proteomic database, to look for the *A. thaliana* proteins homologous to the *B. oleracea* xylem sap proteins. Eighty six proteins were already identified in at least one cell wall proteome. For example, 30 proteins were found in the leaf apoplastic proteome [18], 23 in the stem cell wall *N*-glycoproteome [19], 28 and 27 in the 5- and 11-day-old etiolated cell wall proteomes respectively [20]. Seventy eight proteins were new proteins never found in previously characterized cell wall proteomes.

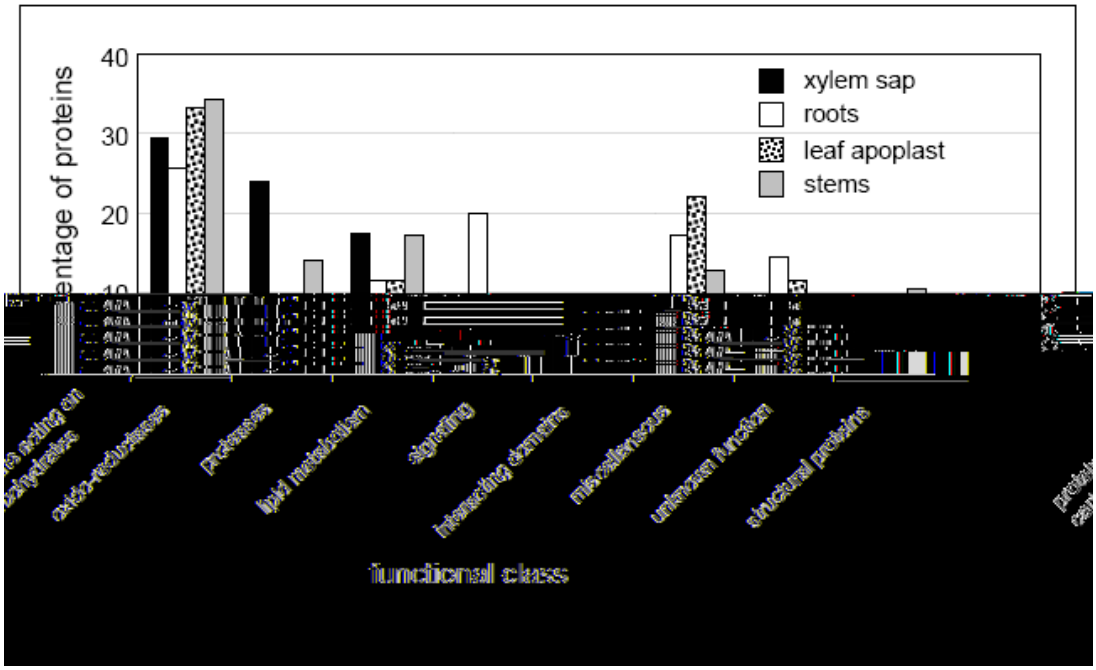


**Figure 2.** Comparison between the *B. oleracea* xylem sap proteome and previously characterized *A. thaliana* cell wall proteomes.

*A. thaliana* cell wall proteomes were found in the *WallProtDB* database (http://www.polebio.scsv.ups-tlse.fr/WallProtDB/). As in *WallProtDB*, *B. oleracea* proteins were annotated with regard to predicted functional domains and distributed in functional classes accordingly [29]. The distribution of the xylem sap proteins (xylem sap proteome, black bars) in these functional classes was compared to that of proteins previously identified in *A. thaliana* cell wall proteomes: roots (white bars) [43], rosette leaves (punctuated bars) [18], and stems (grey bars) [19]. Only proteins predicted to be secreted are considered in all cases, *i.e.* proteins having a predicted signal peptide and no known targeting signal in any cell compartment.

Using *WallProtDB*, the whole xylem sap proteome was compared to *A. thaliana* cell wall proteomes of roots [43], rosette leaves [18], and stems [19]. In all cases only proteins predicted to be secreted were considered. Fig. 2 highlights several differences. There were two times more oxido-reductases in the xylem sap proteome compared to the other cell wall proteomes. The proportion of proteases was higher in the xylem sap and the stem proteomes than in the rosette leaves and root proteomes. On the contrary, there were fewer proteins with interacting domains in the xylem sap than in the other proteomes, with only two proteins homologous to lectins, two to protease inhibitors, one to xylanase inhibitors, and one to pectin methylesterase inhibitors. Structural proteins were missing as in the rosette leaves and stem proteomes.

### 3.5 Expression of *A. thaliana* genes homologous to genes encoding proteins identified in the xylem sap of *B. oleracea* in roots

The presence of proteins in the xylem sap raises the question of the origin of these proteins. To better understand it, we looked at the pattern of expression of *A. thaliana* genes homologous to *B. oleracea* genes encoding xylem sap proteins in the AREX database which collects *A. thaliana* genes patterns of expression based on transcriptomics data [38]. The root is characterized by different developmental zones and tissues. We focused our attention on the root tip where xylem vessels are formed and where root absorption occurs. From outside to inside, there are different cell layers, namely epidermis, cortical cells, endodermis, pericycle and the stele comprising the xylem and the phloem vessels. Expression data could be found for 90% of the *A. thaliana* genes corresponding to proteins identified in the xylem sap and all those genes were found to be transcribed even at low level in root tips. Most relevant patterns were the following (Supporting information Fig. 4): expression in all root tissues except in stele parenchyma cells (pattern 1: 17% of the genes including one third of the genes encoding proteins predicted to be intracellular); expression mostly in epidermis including or not root hairs, or cortical cells (pattern 2: 10%); expression mostly in cortical cells, and eventually in endodermis, pericycle or stele (pattern 3: 16%); expression in stele parenchyma stele including or not vessels (pattern 4: 37%); expression mostly in phloem cells (pattern 5: 8%); expression mostly in xylem cells (pattern 6: 12%). Very few genes had transcripts neither in the stele, the pericycle or the endodermis.

## 4. Discussion

Previous studies showed the presence of proteins in the xylem sap of different species [4, 6, 8-17]. However, with the exception of the xylem sap proteome of an hybrid poplar [13], identification of proteins was done against heterologous sequences because of the lack of genomic or EST sequences for all the plants studied. The most complete xylem sap proteomes are those of *B. napus* [5], an hybrid poplar [13] and *Z. mays* [17] with 69, 97, and 154 proteins identified respectively. However, a detailed examination of the results indicates that several of the identified proteins show homology to different parts of the same protein and/or are identified with identical peptides. Such proteins can be present in different spots of 2D-gels or in different bands of 1D-gels, thus indicating the presence of isoforms of the same gene product as a consequence of PTMs or resulting from protein degradation. As a consequence, the number of proteins in each proteome is certainly lower. This was discussed in the case of the *Z. mays* xylem sap proteome, thus leading to the conclusion that only 59 different proteins were identified instead of 154 [17]. To our knowledge, our proteomic study provides the characterization of the largest xylem sap proteome with 189 different proteins identified. When compared to previous *B. napus* xylem sap proteomes [4, 7, 8], only 20 proteins out of the 189 identified in this study were already found. In addition to data obtained by MS analysis of the *B. oleracea* xylem sap, this study provides information on the xylem sap *N*-glycoproteome. About half of the proteins (81 proteins) predicted to be secreted identified in the *B. oleracea* xylem sap were retained on the ConA column, showing that they are *N*-glycosylated and confirming that *N*-glycosylation is a major PTM of extracellular proteins. The high proportion of secreted *N*-glycosylated proteins was expected because they pass through the endoplasmic reticulum where *N*-glycosylation occurs [44]. As expected, the thirteen proteins predicted to be secreted and devoid of *N*-glycosylation sites were only found in the xylem sap proteome. For the remaining 70 proteins not retained on ConA, it is assumed that their *N*-glycans were removed in the xylem sap. Indeed all the GH families possibly involved in *N*-glycan degradation were found, namely β-D-xylosidases (GH3), N-acetyl-hexosaminidases (GH19), β-D-galactosidases (GH35), and α-D-mannosidases (GH38) [45]. Besides, ConA affinity chromatography was assumed to enrich the protein mixture in glycoproteins present in low amount, thus allowing to increase the coverage of the xylem sap proteome.

The proportion of proteins predicted to be intracellular (13%) is rather low in this study. All except three of the proteins predicted to be intracellular have predicted *N*-glycosylation sites.

However, none of them was retained on the ConA column, suggesting that their *N*-glycosylation sites are not occupied. All these proteins are present at a low level apart from a protein homologous to methionine synthase. Sixty six percent of the proteins identified in hybrid poplar xylem proteome are devoid of predicted signal peptide [13]. The current hypothesis to explain the presence of intracellular proteins in xylem sap is that such proteins originate from differentiating xylem cells, and that they are released in xylem sap after cell death [13]. In the case of perennial plants such as poplar, secondary wall formation and xylem growth are more intensive than in annual plants. This would explain why there are more types of intracellular proteins in the poplar xylem sap. On the contrary, the *Z. mays* xylem proteome only contained proteins predicted to be secreted [17].

A major difficulty encountered in xylem sap proteome analysis relies in the harvesting step. The plants have to be decapitated and the harvesting can last for several hours. With regard to the harvesting duration, a detailed study performed in *G. max* showed that the 1D-electrophoresis pattern of xylem sap proteins was constant over a 28 h-period of harvesting

homologous to At5g07630, was predicted to be secreted. Altogether, the xylem and phloem sap proteomes thus appear to be very different, as expected from their different physiological roles.

All the proteomic data were included in *WallProtDB*, a database dedicated to cell wall proteomics. *WallProtDB* is a tool complementary to existing databases since it allows direct comparison between cell wall proteomes of various organs of *A. thaliana* and *O. sativa*. Only data from plants with genomes completely sequenced or large collection of ESTs are included since unequivocal identification of proteins by peptide mass fingerprinting or peptide sequencing can be done. Other proteomic databases such as the Plant Proteome Database (PPDB, http://ppdb.tc.cornell.edu/) and the Atproteome database (http://fgcz-atproteome.unizh.ch/) are built in a different way. For each gene, the latter databases give information on the conditions in which the proteins were identified as well as MS data when available. Being devoted to cell wall proteomes only, *WallProtDB* resembles AT_CHLORO (http://www.grenoble.prabi.fr/at_chloro/) which describes the chloroplast proteome. In this new version, MS data are also included for the *B. oleracea* xylem sap proteome. The next step would be to crosslink all the plant proteomic databases to get all the information at the same place as was done in the Human Proteomic reference database (http://www.hprd.org/). It would give the plant community a great tool to better understand protein structure and gene regulation.

An interesting outcome of the new version of *WallProtDB* is the comparison between the *B. oleracea* xylem sap proteome and previously characterized cell wall proteomes of *A. thaliana*. Indeed, the fact that xylem sap is considered as part of the apoplast and the closeness of the two species allowed this comparison. Three features distinguish the *B. oleracea* proteome from previously characterized cell wall proteomes [29]: there are more oxido-reductases, more proteases, and less enzyme inhibitors. As discussed above, because of the technical constrains to collect xylem sap, we cannot exclude that some of the proteins related to stress response may change during the harvesting period. Among oxido-reductases, peroxidases represent one tenth of the predicted secreted proteins identified. Peroxidases were previously described as important proteins in xylem sap proteomes. They were associated either to lignin biosynthesis in xylem vessels undergoing differentiation or to plant defense [4, 7, 12, 13]. Proteins homologous to multicopper oxidases were also found to be numerous in *A. thaliana* stems at late flowering stage when lignification occurs [19]. SKU5 (SKEWED 5) was shown

to be involved in root growth and SKS6 (SKU5-SIMILAR 6) to contribute to cotyledon vascular patterning [50, 51]. Blue copper binding proteins were abundant in *A. thaliana* cell suspension cultures. Although their exact role in cell walls is not known, they have been associated to redox processes as electron transfer proteins with small molecular weight compounds [52]. A great proportion of proteases is the second feature of the *B. oleracea* xylem sap proteome. It is the first time that so many proteases are identified. Different specificities could be predicted such as Ser proteases (subtilases), Ser carboxypeptidases, Cys proteases, and Asp proteases. Proteases are assumed to play roles in maturation of enzymes, signaling, protein turnover, and defense against pathogens [53]. It was previously shown that maturation of enzymes occur in the cell wall [54]. Over-expression of CDR1 (CONSTITUTIVE DISEASE RESISTANT 1) encoding an *A. thaliana* Asp protease causes dwarfism and resistance to virulent *Pseudomonas syringae* [55]. Finally, in all previously characterized plant extracellular proteomes, enzymes and the corresponding inhibitors are present, probably allowing fine regulation of enzymatic activities [20]. In the *B. oleracea* xylem sap, there are only a few enzyme inhibitors (two protease inhibitors, a pectin methyl esterase inhibitor, and a xylanase inhibitor). It suggests that the enzymes are fully active. Some of them may play roles in defense reactions against pathogens invading the xylem vessels.

Three additional protein functional classes deserve comments. The main one comprises proteins acting on carbohydrates. Forty two proteins having GH domains were identified. Several of them could play roles in hydrolysis of PTMs of *N*-glycoproteins as discussed above. Others are assumed to contribute to defense reactions, such as β-1,3-glucanases (GH17, 6 proteins) and chitinases/lyzozymes (GH18-19, 5 proteins) [45]. More puzzling are the roles of polygalacturonases and α-L-arabinofuranosidases/β-D-xylosidases. Polygalacturonases (GH28, 7 proteins) are assumed to play roles in the organization of pectins and in their modification in response to pathogen attack [45]. However, the contribution of pectins to secondary walls is very low. α-L-arabinofuranosidases/β-D-xylosidases (GH51, 4 proteins) are major proteins in xylem sap as estimated from MS data. Such proteins were also identified in the poplar xylem sap [13]. Their preferred substrates in cell wall are assumed to be arabinoxylan, and arabinan as inferred from *in vitro* tests [56]. The second functional class to be mentioned is that of proteins possibly involved in signaling. It is the first time the importance of FLAs in xylem sap can be stressed. Indeed, seven proteins homologous to *A. thaliana* FLAs (AtFLA1, 2, 7-10) were identified. Their roles in cell walls

are not yet understood, but they were found to accumulate at the inner side of the G-layer of the xylem of poplar tension wood. They were assumed to have a specific function in the building of this cell wall layer [57]. Finally no structural protein could be identified in the *B. oleracea* xylem sap although a glycine-rich protein (GRP) was previously found in the *B. napus* xylem sap proteome [4]. Such proteins were shown to be present in cucumber xylem sap and to accumulate in the walls of cucumber root metaxylem cells [6]. A bean GRP was shown to be synthesized by living protoxylem cells and xylem parenchyma cells, and to be transported from xylem parenchyma cells to the protoxylem wall after cell death [58].

Several authors have already discussed the origin of xylem sap proteins [3, 10, 12]. In the root, the endodermis constitutes a barrier preventing the movement of organic substances and even water from the epidermis and cortical cells to the stele through the apoplast. It is assumed that proteins present in the xylem sap are synthesized in the stele cells. They would then be delivered apoplastically to the xylem sap thus considered as part of the apoplast. In this study, we looked at the pattern of expression of the *A. thaliana* genes homologous to the *B. oleracea* xylem sap proteins using the AREX database. All the genes were found to be expressed in the root tip, and only a few of them were expressed neither in the stele, the pericycle, or the endodermis. It suggests that the proteins were synthesized in the root tip, and then loaded into the xylem sap. These data are consistent with previous experimental data. A cucumber GRP was shown to be synthesized in the vascular tissues of the root and assumed to be transported over long distance *via* the xylem sap to vessels and sclerenchyma of aboveground organs [6]. Xylem sap proteins as diverse as a cucumber lectin [5], a cucumber peroxidase [12], and a tomato Cys-rich protein with structural similarity to LTPs [14] were found in roots and the genes encoding the lectin and the Cys-rich protein were shown to be transcribed in roots. Altogether, these data strengthen the hypothesis of the production of xylem sap proteins in the root stele, and their further loading into the xylem sap. The water flow would then ensure their long distance transport to aboveground organs. The composition of the xylem sap proteome emsapyov . 3.74(p)-0.295585( )-.295585(0.29585bbw)-0.295585(

[16] Houterman, P., Speijer, D., Dekker, H., de Koster, C. *et al.*, The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Molec Plant Pathol* 2007, *8*, 215-221.
[17] Alvarez, S., Goodger, J. Q., Marsh, E. L., Chen, S. *et al.*, Characterization of the maize xylem sap proteome. *J. Proteome Res.* 2006, *5*, 963-72.
[18] Boudart, G., Jamet, E., Rossignol, M., Lafitte, C. *et al.*, Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics. *Proteomics* 2005, *5*, 212-221.
[19] Minic, Z., Jamet, E., Negroni, L., der Garabedian, P. A. *et al.*, A sub-proteome of *Arabidopsis thaliana* trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. *J. Exp. Bot.* 2007, *58*, 2503-2512.
[20] Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R. *et al.*, A new picture of cell wall protein dynamics in elongating cells of Arabidopsis thaliana: Confirmed actors and newcomers. *BMC Plant Biol.* 2008, *8*, 94.
[21] 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.
[22] Bayer, E. M., Bottrill, A. R., Walshaw, J., Vigouroux, M. *et al.*, *Arabidopsis* cell wall proteome defined using multidimensional protein identification technology. *Proteomics* 2006, *6*, 301-11.
[23] Schultz, C. J., Ferguson, K. L., Lahnstein, J., Bacic, A., Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. *J. Biol. Chem.* 2004, *279*, 455103-45511.
[24] Borderies, G., Jamet, E., Lafitte, C., Rossignol, M. *et al.*, Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: A critical analysis. *Electrophoresis* 2003, *24*, 3421-3432.
[25] Chivasa, S., Ndimba, B. K., Simon, W. J., Robertson, D. *et al.*, Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* 2002, *23*, 1754-1765.
[26] Kwon, H.-K., Yokoyama, R., Nishitani, K., A proteomic approach to apoplastic proteins involved in cell wall regeneration in protoplasts of *Arabidopsis* suspension-cultured cells. *Plant Cell Physiol.* 2005, *46*, 843-857.
[27] Robertson, D., Mitchell, G. P., Gilroy, J. S., Gerrish, C. *et al.*, Differential extraction and protein sequencing reveals major differences in patterns of primary cell wall proteins from plants. *J. Biol. Chem.* 1997, *272*, 15841-15848.
[28] Prakash, S., Bhat, S., Quiros, C., Kirti, P. *et al.*, in J. Janick (Ed.), *Plant Breeding Reviews, vol 31*, John Wiley & Sons, Inc, Hoboken, NJ, 2009, pp. 21-188.
[29] Jamet, E., Albenne, C., Boudart, G., Irshad, M. *et al.*, Recent advances in plant cell wall proteomics. *Proteomics* 2008, *8*, 893-908.
[30] Kehr, J., Rep, M., Protein extraction from xylem and phloem sap. *Methods Mol Biol* 2007, *355*, 27-35.
[31] Laemmli, U. K., Cleavage of the structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, *227*, 680-685.
[32] Scheler, C., Lamer, S., Pan, Z., Li, X. P. *et al.*, Peptide mass fingerprinting 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.
[33] Willats, W. G., Knox, J. P., A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of β-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. *Plant J.* 1996, *9*, 919-925.

[34] Martin, A., J, L., Kichey, T., Gerentes, D. *et al.*, Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. *Plant Cell* 2006, *18*, 3252-3274.
[35] Frottin, F., Espagne, C., Traverso, J., Mauve, C. *et al.*, Cotranslational proteolysis dominates glutathione homeostasis to support proper growth and development. *Plant Cell* 2009, *21*, 3296-3314.
[36] San Clemente, H., Pont-Lezica, R., Jamet, E., Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case. *Bioinform. Biol. Insights* 2009, *3*, 15-28.
[37] Pont-Lezica, R., Minic, Z., San Clemente, H., Roujol, D. *et al.*, in M. A. Osborne (Ed.), *Advances in Genetics Research*, Nova Science Publishers, Inc., Hauppauge, NY, 2010.
[38] Birnbaum, K., Shasha, D., Wang, J., Jung, J. *et al.*, A gene expression map of the Arabidopsis root. *Science* 2003, *302*, 801-806.
[39] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. *et al.*, Basic local alignement search tool. *J. Mol. Biol.* 1990, *215*, 403-410.
[40] Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 1976, *72*, 248-254.
[41] Jamet, E., Canut, H., Boudart, G., Pont-Lezica, R., Cell wall proteins: a new insight through proteomics. *Trends Plant Sci.* 2006, *11*, 33-39.
[42] Lagerkrantz, U., Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 1998, *150*, 1217-1228.
[43] Basu, U., Francis, J. L., Whittal, R. W., Stephens, J. L. *et al.*, Extracellular proteomes of *Arabidopsis thaliana* and *Brassica napus* roots: analysis and comparison by MUdPIT and LC-MS/MS. *Plant Soil* 2006, *286*, 357-376.
[44] Faye, L., Boulaflous, A., Benchabane, M., Gomord, V. *et al.*, Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine* 2005, *23*, 1770-1778.
[45] Minic, Z., Physiological roles of plant glycoside hydrolases. *Planta* 2008, *227*, 723-740.
[46] Thompson, G., Schulz, A., Macromolecular trafficking in the phloem. *Trends Plant Sci* 1999, *4*, 354-360.
[47] Dinant, S., Clark, A., Zhu, Y., Vilaine, F. *et al.*, Diversity of the superfamily of phloem lectins (Phloem Protein 2) in angiosperms. *Plant Physiol* 2003, *131*, 114-128.
[48] Cho, W., Chen, X.-Y., Rim, Y., Chu, H. *et al.*, Proteome study of the phloem sap of pumpkin using multidimensional protein identification technology. *J Plant Physiol* 2010, *167*, 771-778.
[49] Lin, M.-K., Lee, Y.-J., Lough, T., Phinney, B. *et al.*, Analysis of the pumpkin phloem proteome provides insight into angiosperm sieve tube function. *Mol Cell Proteomics* 2009, *8*, 343-356.
[50] Jacobs, J., Roe, J. L., *SKS6*, a multicopper oxidase-like gene, participates in cotyledon vascular patterning during *Arabidopsis thaliana* development. *Planta* 2005, *222*, 652-666.
[51] Sedbrook, J. C., Carroll, K. L., Hung, K. F., Masson, P. H. *et al.*, The *Arabidopsis SKU5* gene encodes an extracellular glycosyl phosphatidylinositol-anchored glycoprotein involved in directional root growth. *Plant Cell* 2002, *14*, 1635-1648.

[52] Nersissian, A. M., Shipp, E. L., Blue copper-binding domains. *Adv. Protein Chem.* 2002, *60*, 271-340.
[53] Schaller, A., A cut above the rest: the regulatory function of plant proteases. *Planta* 2004, *220*, 183-97.
[54] Albenne, C., Canut H, Boudart, G., Zhang, Y. *et al.*, Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships. *Mol. Plant* 2009, *2*, 977-989.
[55] Xia, Y., Suzuki, H., Borevitz, J., Blount, J. *et al.*, An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. *EMBO J* 2004, *23*, 980-8.
[56] Minic, Z., Rihouey, C., Do, C., Lerouge, P. *et al.*, Purification and characterization of enzymes exhibiting beta-D-xylosidase activities in stem tissues of Arabidopsis. *Plant Physiol* 2004, *135*, 867-878.
[57] Lafarguette, F., Leple, J. C., Dejardin, A., Laurans, F. *et al.*, Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol.* 2004, *164*, 107-121.
[58] Ringli, C., Keller, B., Ryser, U., Glycine-rich proteins as structural components of plant cell walls. *Cell. Mol. Life Sci.* 2001, *58*, 1430-41.
[59] Minic, Z., Jouanin, L., Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol Biochem* 2006, *44*, 435-449.

**Supporting information**

**Supporting information Table S1**. Identification by LC-MS/MS of proteins present in the *B. oleracea* xylem sap (xylem sap proteome).

**Supporting information Table S2.** Identification by LC-MS/MS of proteins of *B. oleracea* retained by ConA affinity chromatography (xylem sap *N*-glycoproteome).

**Supporting information Table S3.** Description of the xylem sap proteome and of the xylem sap *N*-glycoproteome of *B. oleracea*.

**Supporting information Table S4.** Spectral counting data for the most abundant proteins identified in the B. oleracea xylem sap proteome and for the proteins predicted to be intracellular.

**Supporting information Fig. 1.** Xylem sap sampling from *B. oleracea* cut stems.

**Supporting information Fig. 2.** Content of *WallProtDB.*

**Supporting information Fig. 3.** LC-MS/MS data as shown in *WallProtDB.*

**Supporting information Fig. 4.** Root pattern of expression of *A. thaliana* genes encoding proteins homologous to *B. oleracea* xylem sap proteins.