Molecular remodeling in *Populus PdKOR* RNAi roots profiled using LC-MS/MS proteomics

Paul E. Abraham[1], Anna Matthiadis[2], Robert L. Hettich[1] and Udaya C Kalluri[2*]

[*]Corresponding Author; kalluriudayc@ornl.gov

Author affiliation: Chemical Sciences Division[1] and Biosciences Division[2], Oak Ridge National Laboratory, PO BOX 2008, Oak Ridge, TN 37831-6422

Abstract

Plant endo-β-1,4-glucanases belonging to the Glycoside Hydrolase Family 9 have functional roles in cell wall biosynthesis and remodeling via endohydrolysis of (1→4)-β-D-glucosidic linkages. Modification of cell wall chemistry via RNAi-mediated downregulation of *Populus deltoides KOR1 (PdKOR)*, a endo-β-1,4-glucanase gene, in *Populus deltoides* has been shown to have functional consequences for the composition of secondary metabolome and the ability of modified roots to interact with beneficial microbes. The molecular remodeling that underlies the observed differences at metabolic, physiological, and morphological levels in roots is not well understood. Here we used a LC-MS/MS-based proteome profiling approach to survey the molecular remodeling in root tissues of *PdKOR* and control plants. A total of 14316 peptides were identified and these mapped to 7139 *P. deltoides* proteins. Based on 90% sequence identity, the measured protein accessions represent 1187 functional protein groups. Analysis of GO categories and

specific individual proteins showed differential expression of proteins relevant to plant-microbe interactions, cell wall chemistry, and metabolism. The new proteome dataset serves as a useful resource for deriving new hypotheses and empirical testing pertaining to functional roles of proteins and pathways in differential priming of plant roots to interactions with microbes.

A greater understanding of the functional implications of plant biomass chemistry optimization efforts and the underlying molecular remodeling is critical to closing the knowledge gaps in sustainable bioenergy crop development. We have previously reported that RNAi-mediated down-regulation of the *Populus deltoides KOR* (*PdKOR1*) gene, belonging to the endo-β-1,4-glucanase family, results in altered cell wall, secondary metabolome and associated microbiome composition [1,2]. Greenhouse co-culture studies showed that interactions with beneficial microbes were impacted as a result of the gene modification. Here, we summarize results from the first proteomics characterization of roots from *PdKOR* and control plants grown in the greenhouse.

Greenwood stem cuttings were rooted and grown in ProMix potting soil at 25°C temperature under long day length (16 h) regime in the greenhouse. At the end of five weeks of growth, roots from three biological replicates of *PdKOR* RNAi and control (empty vector transformed) lines were destructively harvested, bulk roots were rinsed in water and blotted dry, and fine roots were collected in 15 mL tubes, flash-frozen in liquid nitrogen, and store at -80°C. To effectively extract and recover proteins from roots, we employed a combination of physical and chemical lysis procedures previously described, albeit with minor modifications [3]. Root samples were ground to powder using a Qiagen TissueLyser II adapter pre-cooled to -80°C at a frequency of 30 for 20 seconds and suspended in lysis buffer (4% SDS in 100 mM NH_4HCO_3), boiled for 5 min, sonically disrupted (30% amplitude, 10 s pulse with 10 s rest, 2 min total pulse time), and then boiled for an additional 5 min. The crude protein extract was pre-cleared via centrifugation and proteins were precipitated using a chloroform/methanol/water extraction procedure. Dried protein pellets were resuspended in 2% SDC (100 mM NH_4HCO_3) and protein amounts were estimated by performing a BCA assay (Pierce biotechnology). For each sample, an aliquot of ~100 ug of protein was digested via two separate additions of sequencing-grade trypsin (Promega, 1:75 [w:w]), once overnight followed by another for 3 hr. at 37 °C. The peptide mixture was adjusted to 1% formic acid to precipitate Sodium deoxycholate (SDC). Hydrated ethyl acetate was added to each sample at a 1:1 [v:v] ratio three times to effectively remove SDC. Samples were then placed in a SpeedVac

Concentrator (Thermo Fischer Scientific) to remove ethyl acetate and further concentrate the sample. The peptide-enriched flow-through was quantified by BCA assay, desalted on RP-C18 stage tips (Pierce Biotechnology), and then stored at −80°C.

Sample peptide mixtures were analyzed on a Q-Exactive Plus mass spectrometer (Thermo Fischer Scientific) coupled with a with a Proxeon EASY-nLC 1200 liquid chromatography (LC) pump (Thermo Fisher Scientific) as previously described with some minor modifications [4]. In brief, peptides were separated on a 75 μm inner diameter microcapillary column packed with 50 cm of Kinetex C18 resin (1.7 μm, 100 Å, Phenomenex) placed in a column heater (Sonation GmbH) at 60°C. For each sample, a 2 μg aliquot was loaded in buffer A (0.1% formic acid, 2% acetonitrile) and eluted with a linear 150 min gradient of 2 – 20% of buffer B (0.1% formic acid, 80% acetonitrile), followed by an increase in buffer B to 30% for 10 min, another increase to 50% buffer for 10 min and concluding with a 10 min wash at 98% buffer A. The flow rate was kept at 200 nL/min. MS data was acquired with the Thermo Xcalibur software version 4.27.19 using the a top10 data-dependent acquisition strategy using a resolution of 70,000 at m/z 200. A 1.6 m/z isolation window and fragmentation of precursor ions was performed by higher-energy C-trap dissociation (HCD) with a normalized collision energy of 30 eV. MS/MS sans were performed at a resolution of 17,500 at m/z 200.

MS raw data files were searched against the *Populus deltoides* reference proteome database (https://phytozome.jgi.doe.gov/pz/portal.html) to which common contaminate proteins had been added. A decoy database, consisting of the reversed sequences of the target database, was appended in order to establish the false-discovery rate (FDR) at the spectral level. For standard database searching, the peptide fragmentation spectra (MS/MS) were analyzed by Proteome Discoverer v2.3. The MS/MS were searched using the MS Amanda v2.0 [5] and was configured to derive fully tryptic peptides using settings for high-high MS/MS data: MS1 mass tolerance of 5 ppm and MS2 mass tolerance of 0.02 Da. A static modification on cysteines (iodoacetamide; +57.0214 Da), a dynamic modification on methionine (oxidation; 15.9949) and aspartate and glutamate (methylation; 14.016) were considered. The results were processed by Percolator [6] to estimate q values. Peptide spectrum matches (PSMs) and peptides were considered identified at a q value <0.01 (Supplemental Table 1).

For label-free quantification, MS1-level precursor intensities (peak areas) were derived from the Minora Feature and Precursor ions quantifier nodes using default parameters. Protein abundance values were Log2-transformed and normalized using InfernoRDN v1.1 [7] by LOESS across biological replicates and median centering adjustments across the global dataset. Missing values were imputed (low abundance resampling method) using Perseus [8]. A Student's t-test and a permutation-based FDR was implemented to characterize protein abundance differences between *PdKOR1* and control roots. Differential protein abundances having a q-value of 0.05 and a Log2 difference of 1 were considered significantly different and were further investigated (Supplemental Table 2). Whole-genome gene ontology (GO) term annotation was performed using Blast2GO [9] with a blastp E-value hit filter of 1×10-5, an annotation cutoff value of 55 and a GO weight of 5. Using ClueGO, [10] observed GO biological processes were subjected to the right-sided hypergeometric enrichment test at default settings and p-value correction was performed using the Holm-Bonferroni step-down method. Minimal reporting of functional groups was achieved by implementing ClueGO GO term fusion and grouping settings were selected to reduce GO term redundancy. The term enriched at the highest level of significance was used as the representative term for each functional cluster. The GO terms at adjusted $p<0.05$ were considered significantly enriched (Supplemental Table 3). Additionally, the *P. deltoides* proteome was annotated by Mercator4 v2.0 to analyze protein function in MapMan v3.6 [11]. A Wilcoxon Rank Sum Test was performed and identified functional bins that exhibit a differential abundance behavior (Supplemental Table 4) that corroborates the GO enrichment analysis.

Overall, LC-MS/MS proteomic characterization of *PdKOR* and control roots (Figure 1a) revealed a comparable total/ average number of proteins identified from each sample group (triplicates of *PdKOR* or control). Principle Component Analysis (PCA) differentiated the replicates into their respective groups. This dataset provides the first insight into significantly differential proteome expression between *PdKOR* and control roots (Figure 1b). Key classes of plant proteins with potential role in mediating plant-microbe interactions; Receptor-like serine/threonine-protein kinases, calcium signaling components, cell wall biosynthesis enzymes and peroxidases, were observed to be differentially expressed at the protein level.

Proteome data analysis based on GO classifications and relative expression levels of specific protein isoforms suggests molecular remodeling in *Populus PdKOR* RNAi roots relative the

control. Our data shows differential expression of homologs of common symbiosis pathway (CSP) factors. CSP has been shown to be core to mediation of rhizobia–legume, actinorhizal and arbuscular mycorrhizal symbiosis and includes calcium signaling factors, such as calcium-dependent protein kinase (CDPK) [12] and Calcium and calmodulin-dependent protein kinase (CCaMK) [13], and in transmembrane signal transduction such as LYK3 receptor-like kinases (Lysin motif receptor-like kinases (LysM RLKs). Homologs in distinct plant species, *Lotus japonicus* [14] *Medicago truncatula* [15], have been shown to be involved in beneficial nodule-forming microbes.

Disease resistance proteins, abscisic acid receptors and heat shock proteins belonging to biotic and abiotic stress responsive pathways, show a 4-5-fold higher expression in *PdKOR* roots. The differential expression of proteins in lignin biosynthesis and phenylpropanoid pathways, such as cinnamoyl-CoA reductase, laccase, UDP-glucose flavonoid 3-O-glucosyltransferase 6 and Flavonol 3-O-glucosyltransferase ascorbate peroxidase, as well as in primary sugar and carbohydrate metabolism pathways as shown by differential GO term enrichment is supported by the differential lignin, secondary metabolite and sugar levels reported previously from *PdKOR* and control plants [1]. Differential expression of specific cell wall biosynthesis pathway genes and disease resistance response proteins has also been reported in a study to identify core transcriptional response to inoculation of three different ectomycorrhizal fungi on oak tree roots [16]. Homolog of one such key core upregulated factor, myo-inositol oxygenase (MIOX) [17], is differentially expressed in *PdKOR* roots.

In summary, the extensive LC-MS/MS proteomics dataset provides insights into global changes in root proteome as a consequence of modification of a cell wall remodeling gene, *PdKOR*, in *P. deltoides*. Previous findings on phenotypic implications of *PdKOR* gene modification on *Populus* cell wall chemistry, secondary metabolome and interactions with microbes are supported [1, 2], in part, by the proteome remodeling shown here by differential proteome analysis. Further functional analysis of targeted proteins, across cellular to whole plant contexts, is needed to validate the hypothesized functional consequences of the differentially expressed proteins.

Acknowledgments

This research was sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research, as part of the Plant Microbe Interfaces Scientific Focus Area at Oak Ridge National Laboratory (http://pmi.ornl.gov). Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

Keywords

Cell wall chemistry, proteome, plant-microbe interaction, *PdKOR*, host, root

Conflicts of Interest

The authors declare no conflicts of interest.

Data deposition in open access database: All spectral data collected in this study was deposited at the ProteomeXchange Consortium via the MASSIVE repository. The project identifier is MSV000085068 and the data can be reviewed under the username "MSV000085068_ reviewer" and password "PdKOR". [is private mode until manuscript is under review]

References

1. U.C. Kalluri, R.S. Payyavula, J. L. Labbé, N. Engle, G. Bali, S.S. Jawdy, R.W. Sykes, M. Davis, A. Ragauskas, G.A. Tuskan, T.J. Tschaplinski, *Front Plant Sci.* **2016,** *7*, 1455.

2. A.M. Veach, D. Yip, N.L. Engle, Z.K. Yang, A. Bible, J. Morrell-Falvey, T.J. Tschaplinski, U.C. Kalluri, C.W. Schadt, *Plant and Soil*. **2018**, *429*, 349.

3. P. Abraham, R.J. Giannone, R.M. Adams, U. Kalluri, G.A. Tuskan, R.L. Hettich, *Mol Cell Proteomics.* **2013,** *12*, 106.

4. M.I. Villalobos Solis, R.J. Giannone, R.L. Hettich, P.E. Abraham, *Anal. Chem.* **2019**, *91*, 7273-7279.

5. V. Dorfer, P. Pichler, T. Stranzl, J. Stadlmann, T. Taus, S. Winkler, K Mechtler, MS Amanda, *J. Proteome Res.* **2014**, *13*, 3679.

6. L. Käll, J.D. Canterbury, J. Weston, W.S. Noble, M.J. MacCoss, *Nat. Methods.* **2007**, *4*, 923.

7. A.D. Polpitiya, W.J. Qian, N.Jaitly, V.A. Petyuk, J.N. Adkins, D.G. Camp II, G.A. Anderson, R.D. Smith, **2008**, *24,* 1556.

8. S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, J. Cox, *Nat Methods* **2016**, *13*, 731.

9. A. Conesa, S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, M. Robles, *Bioinformatics* **2005**, *21*, 3674.

10. G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.H. Fridman, F. Pages, Z. Trajannoski, J. Galon, *Bioinformatics* **2009**, *25*, 1091.

11. R. Schwacke, G. Y. Ponce-Soto, K. Krause, A. M. Bolger, B. Arsova, A. Hallab, K. Gruden, M. Stitt, M. E. Bolger, B. Usadel, *Mol Plant* **2019**, *12*, 879.

12. N. Takeda, T. Maekawa, M. Hayashi, *The Plant Cell*, **2012**, *24*, 810.

13. B.J. Sieberer, M. Chabaud, J. Fournier, A.C.J. Timmers, D.G. Barker, *The Plant J.* **2012**, *69*, 822.

14. E.B. Madsen, L.H. Madsen, S. Radutoiu, M. Olbryt, M. Rakwalska, K. Szczyglowski, S. Sato, T. Kaneko, S. Tabata, N.J. Sandal, J. Stougaard, *Nature.* **2003**, *425*, 637.

15. E. Limpens, C. Franken, P. Smit, J. Willemse, T. Bisseling, R. Geurts. *Science.* **2003**, *302*, 630.

16. M.L. Bouffaud, S. Herrmann, M.T. Tarkka, M. Bonn, L. Feldhahn, and F. Buscot, *BMC Genomics* **2020**, *21*.

17. U. Kanter, B. Usadel, F. Guerineau, Y. Li, M. Pauly, and R. Tenhaken, *Planta* **2005**, *221*, 243.

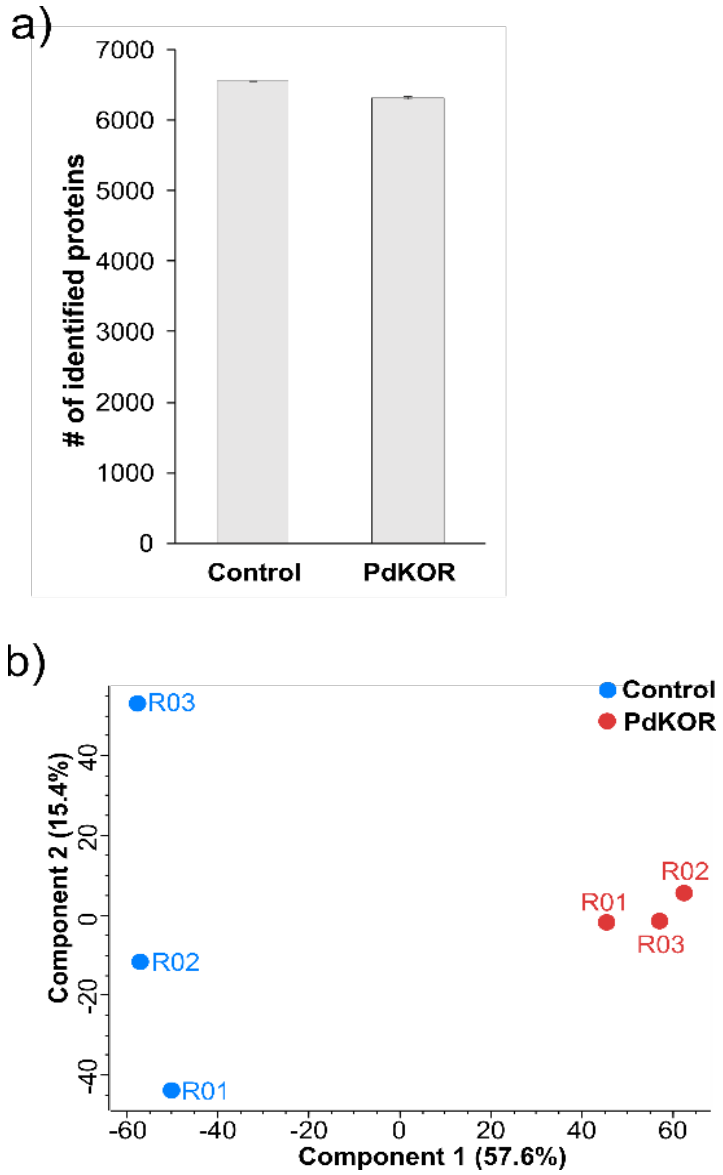

Figure 1. a) Average number of identified proteins for each sample group. Error bars represent the standard deviation across biological triplicates. b) Plot illustrating results from principal component analysis (PCA).

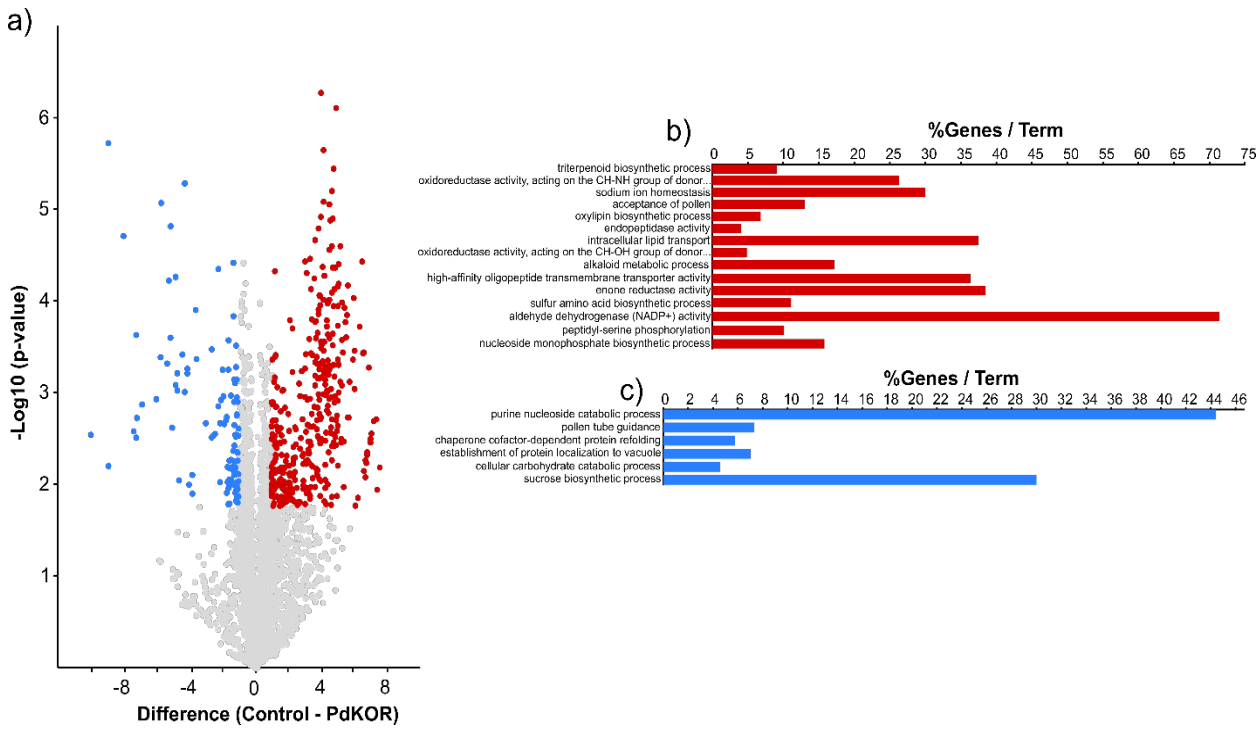

Figure 2: a) Volcano plot illustrating differential Log2 protein abundances. Significant differences were determined by a Student's t-test followed by a permutation-based FDR calculation. Gene ontology (GO) term enrichment was performed (right-sided hypergeometric enrichment test and p-value correction was performed using the Holm-Bonferroni step-down method) to identify biological processes enriched in proteins more abundant in the b) control and c) *PdKOR*. GO term fusion and grouping settings were selected to reduce GO term redundancy and the term enriched at the highest level of significance was used as the representative term for each functional cluster. The GO terms at adjusted $p<0.05$ were considered significantly enriched. Red and blue correspond to an increase or decrease, respectively, in the relative protein abundance in Control when compared to the abundance value in *PdKOR* lines.

Supplemental Table 1. Filtered results (< 1% FDR) of peptides and their associated *P. deltoides* protein accessions.

Supplemental Table 2. List of identified proteins and their relative abundances, which are represented as normalized Log2-transformed values. Missing values were replaced by random numbers drawn from a distribution optimized to simulate a typical abundance region near the limit-of-detection. A Student's t-test was performed, and permutation-based FDR was implemented. Protein abundance differences with p-values > 0.05 and Log2 differences >1 were considered significantly different.

Supplemental Table 3. Summary of results from ClueGO gene ontology enrichment test.

Supplemental Table 4. Summary of results from Wilcoxon Rank sum test performed by MapMan v3.6.