Deep learning neural network prediction method improves proteome profiling of vascular sap of grapevines during Pierce's disease development

Cíntia H. D. Sagawa[1], Paulo A. Zaini[1], Renata de A. B. Assis[1,2], Houston Saxe[1], Michelle Salemi[3], Aaron Jacobson[1], Brett S. Phinney[3], Abhaya M. Dandekar[1]*

[1] Department of Plant Sciences, University of California, Davis, CA 95616, USA;

[2] Departamento de Ciências Biológicas, Instituto de Ciências Exatas e Biológicas, Núcleo de Pesquisas em Ciências Biológicas, Universidade Federal de Ouro Preto, Ouro Preto, MG, 35400-000, Brazil;

[3] Proteomics Core Facility, University of California, Davis, CA 95616, USA;

*Author to whom correspondence should be addressed: amdandekar@ucdavis.edu

Abstract

Plant secretome studies have shown the importance of plant defense proteins in the vascular system against pathogens. Studies on Pierce's disease of grapevines caused by the xylem-limited bacteria *Xylella fastidiosa* (*Xf*) have detected proteins and pathways associated to its pathobiology. Despite the biological importance of the secreted proteins in the extracellular space to plant survival and development, proteome studies are scarce due to technical and technological challenges. Deep learning neural network prediction methods can provide powerful tools for improving proteome profiling by data-independent acquisition (DIA). We aimed to explore the potential of this strategy by combining it with *in silico* spectral library prediction tool, Prosit, to analyze the proteome of vascular leaf sap of grapevines with Pierce's disease. The results demonstrate that the combination of DIA and Prosit increased the total number of identified proteins from 145 to 360 for grapevines and 18 to 90 for *Xf*. The new proteins increased the range of molecular weight, assisted on the identification of more exclusive peptides per protein, and increased the identification of low abundance proteins. These increases allowed the identification of new functional pathways associated with cellular responses to oxidative stress to be further investigated.

Keywords: predicted spectral library; quantitative proteomics; Prosit; apoplast; xylem sap; grapevine; Pierce's Disease

1. Introduction

The vascular system is essential for the exchange of information and resource allocation throughout the plant, from roots to aerial tissues. It is composed of two types of vascular tissues:

phloem and xylem. The phloem sap contains photoassimilates and other macromolecules that move throughout the plant from areas of synthesis or excess (source) to areas of use (sink) and storage [1]. The xylem sap transports water and nutrients from roots to aerial tissues, driven by a difference in water potential due to transpiration (Tanner and Beevers, 2001). Recent studies have shown that the xylem can also contain a wide range of proteins involved in various biological processes involved in growth regulation, protection against environmental stress, homeostasis, gas exchanges, cell to cell adhesions, and plant defense against pathogens [3]. These processes are dependent on vesicular trafficking of proteins to the extracellular space, which can either follow conventional or unconventional secretion routes in plant cells. The conventional secretion in plants requires signal peptides in the N-terminus or proper recognition signals to direct them to the endomembrane system pathway, while proteins that follow the unconventional secretion route lack these signals [4]. Plant secretome studies have shown that proteins that follow unconventional secretion can allow plants to respond to a wider range of extracellular stresses and stimuli, facilitating defense responses under stress [4], [5]. Despite the biological importance of the secreted proteins in the extracellular space to plant survival and development, proteome studies are scarce due to technical and technological challenges.

Studies on the role of vascular sap have helped to better understand plant responses to vascular plant diseases (Yadeta and Thomma, 2013). The Gram-negative gammaproteobacteria *Xylella fastidiosa* (*Xf*) is a xylem-limited pathogen that colonizes several economically important crops worldwide causing deadly diseases such as Pierce's disease in grapevines (PD) (Davis et al., 1978), Citrus Variegated Chlorosis (CVC) [8] and most recently Olive Quick Decline Syndrome (OQDS) in Europe (Martelli, 2016). Due to the significant economic impact on the production of citrus in Brazil, *X. fastidiosa* was the first plant pathogen to have its genome sequence determined [10]. The genomic landscape provided an initial description of potential virulence factors and revealed the absence of a type III secretion system commonly employed by plant pathogens to deliver virulence effectors inside plant cells. Molecular and cellular studies followed proposing that the mechanism of disease symptoms would be associated with biofilm formation and xylem blockage triggering the observed disease symptoms [11]–[15]. Additionally, genomics and proteomics have shown the importance of virulence factors secreted by the type II secretion system and outer membrane vesicles for symptom development (Nascimento et al. 2016; Gouran et al. 2016; Santiago et al. 2016; Cianciotto and White 2017;

Feitosa-Junior et al. 2019). These studies highlighted the molecular complexity of the plant-pathogen interaction that takes place in the vascular system.

The first study on xylem sap proteomics in grapevines was performed in sap bleedings from the cultivar Chardonnay (Agüero et al. 2008), which revealed only ten proteins from two-dimensional (2D) gel electrophoresis analysis. As new technologies and proteomic approaches became more sensitive, more proteins were found in the vascular sap of grapevines, increasing the number of identified proteins to 200 varying from 20 to 75 kDa showing differences among resistant and susceptible cultivars to PD (Delaunois et al. 2013). The importance of proteins in the plant response to *X. fastidiosa* was initially shown by Yang et al. (2011) in a proteomic study of stems from infected grapevines. This study revealed thaumatin-like, pathogenesis-related protein 10 and three heat shock proteins were significantly overexpressed in PD-resistant varieties of grapes (Yang et al. 2011). Another study also conducted on the stem of infected grapevines of PD-tolerant and susceptible cultivars identified more than 200 proteins associated with disease resistance, energy metabolism, protein processing and degradation, biosynthesis, stress-related functions, cell wall biogenesis, signal transduction, and ROS detoxification among others [23]. The most recent published study conducted on sap bleeding of infected grapevines highlighted 91 proteins. The novelty of this study was the incorporation of structural data into the proteomic data analysis to enhance the identification of functionally relevant protein candidates that would not be detected from simple amino acid sequence alignments. This study highlighted pathogenesis-related proteins, chitinases, and β-1, 3-glucanases as crucial players in the defense against *X. fastidiosa* [25]. These studies greatly enhanced our understanding of xylem sap physiology; however, they were restricted to more abundant proteins which we have learned to be only a small fraction of xylem sap complexity.

The standard approach in proteomic studies was 2D gel electrophoresis for many years due to its robustness and compatibility with bottom-up (shotgun) proteomics in which the crude protein extract is digested directly for analysis. However, the limitations regarding reproducibility and narrow dynamic range of high abundance proteins masked low abundant counterpart, limiting those analyses [26]. Electrophoresis gels can now be replaced by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS), which has become the most used method to measure the different states and abundance of proteins, lipids and other metabolites [27].

One of the acquisition schemes of tandem mass spectrometry is called data-independent acquisition (DIA) which is based on the acquisition of fragment-ion information for all precursor ions until the desired mass range has been covered, as demonstrated by the sequential window acquisition of all the theoretical mass spectra (SWATH) approach [28]. DIA has been used to identify and quantify thousands of proteins without performing fractionation, increasing reproducibility, and requiring a small amount of protein [27], [29], [30]. Although it improves protein detection with higher reproducibility, the lack of accurate predictive models for fragment ion intensities has impaired its full potential. DIA analysis often uses peptide physiochemical properties stored in spectral libraries or chromatogram libraries. These properties can include information on peptide retention time, product ion m/z, product ion intensity and ion mobility among others [31], [32]. Using this information can ensure confident peptide identification and quantification. Two methods exist to obtain this information, one is experimental and the other is predictive. An example of a predictive method is the deep learning architecture termed Prosit which was created to take advantage of a large number of synthetic peptides and tandem mass spectra generated within the ProteomeTools project to predict with high quality both chromatographic retention time and fragment ion intensity of any peptide [33]. Here we demonstrate the improved performance of integrating Prosit into the DIA pipeline. By reanalyzing our DIA data of the vascular leaf sap of grapevines infected by *X. fastidiosa* compared with healthy plants, we increased the number of identified proteins depicting a deeper description of this plant pathogen interface and generated spectral libraries for DIA analysis of *Vitis vinifera* and *Xylella fastidiosa* that can be incorporated in future proteome studies.

2. Material and methods

2.1. Plant material and *X. fastidiosa* inoculation

Clonal grapevine plants (*Vitis vinifera* L. cv. 'Thompson Seedless') were generated from cuttings using green canes from the current season’s growth. Each cutting was approximately 6 inches long and contained two nodes, with a petiole originating from the top node that supported approximately one square inch of leaf area to maintain minimal photosynthesis during rooting. These prepared cuttings were placed into an EZ-Clone aeroponic cloning system that circulates water purified by reverse osmosis. Roots begin to self-generate after two weeks, and the rooted

cuttings were potted after three-weeks and grown in a greenhouse. New plant growths was trained to a single cane by removing any lateral shoots that emerged. The single cane plants were topped at the height of 1 meter, and additional lateral shoots were removed as they emerged during the experiment. After ten-weeks, the grapevines were infected at 8–12 cm above soil level by punching with a needle gauge to inoculate 20 μL of cultured cells of *Xylella fastidiosa* Temecula1 (*Xf*; ATCC 700964) into the stem as described by Nascimento et al. (2016). The bacterial culture was grown on PD3 medium at $2x10^8$ cells/mL incubated with aeration (120 rpm) at 28°C. After inoculation, plants were placed in the greenhouse in a randomized block design and monitored for 12 weeks post inoculation until leaf symptoms developed.

2.3. Vascular sap extraction and *X. fastidiosa* quantification

Vascular leaf sap was collected from ten leaves above the inoculation point using a pressure chamber (Soil Moisture Equipment Corp., Santa Barbra, CA, USA). Pressure was applied to each leaf blade and the sap collected from the end of the petiole. The leaf blade was placed inside the pressurized chamber leaving only the cut surface of the petiole exposed to release the vascular content, which was collected using a micropipette and stored in a tube on ice during harvest. Pools of about ten leaves above the inoculation point from one plant made one sample (500 uL - 1000 uL). Before processing with the sample preparation for proteomics analysis, an aliquot of 25 uL was reserved from each sample for extraction of DNA with MasterPure™ kit (Epicentre) and bacterial cell count was measured using qPCR (TaqMan™). The primers used were HL5 and HL6 described by Francis et al. (2006). A standard curve was used based on a known serial dilution of *Xf* cells measured by OD_{600}.

2.4. Protein digestion of vascular leaf sap

Up to one milliliter of vascular leaf sap was collected from each plant (pooled from 10 leaves) and a total of three plants per group (Healthy and Diseased) were used. Samples were centrifuged at 5,000 rcf for 5 min at 4°C. The supernatant containing the vascular leaf sap was transferred to a new tube. Total protein content was quantified by Qubit™ Protein Assay Kit (Thermo Fisher Scientific). Sap containing 100 ug of protein was freeze-dried and resuspended in 5% SDS and 50mM triethylammonium bicarbonate (TEAB) at pH 7.55 to a concentration of 0.5 ug/uL. Digestions with trypsin followed the S-Trap™ Micro Spin Column Digestion

Protocol with few modifications. Initially, 10 mM dithiothreitol (DTT) was added and incubated at 50°C for 10 min and rested at room temperature for 10 min. Next, 5 mM iodoacetamide (IAA) was added and incubated at room temperature for 30 min in the dark. The samples were acidified with 12% phosphoric acid followed by the addition of 2.348 mL of freshly made S-trap buffer (90% methanol, 100 mM TEAB, pH 7.1) and mixed immediately by inversion. The entire acidified lysate/St-buffer mix was transferred to the S-trap spin column (650 uL at a time) and centrifuged at 3,000 rcf for 1 min or until all the solution passed through the column. Columns were washed with 400 uL of S-trap buffer and centrifuged at 4,000 rcf until dry. Columns were transferred to a clean elution tube. Trypsin enzyme digest buffer was carefully added (1:25 enzyme: total protein in 121 uL 50mM TEAB, pH 8.0) to the column and followed by incubation at 37°C overnight. After the first hour, the trypsin digestion step was repeated. Peptide elution steps included 80 uL of 50 mM TEAB (pH 8.0) followed by centrifugation at 1,000 rcf for 1 min, 80 uL of 0.5% formic acid followed by centrifugation at 1,000 rcf for 1 min, 80 uL of the solution containing 50% acetonitrile and 0.5% formic acid followed by centrifugation at 4,000 rcf for 1 min. The final pooled elution was dried down in a speed-vacuum. Peptides were resuspended in 0.1% TFA 2% ACN and quantified using Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Fisher Scientific). Equal portions of all samples were mixed together to make a reference sample to be run multiple times for chromatogram library runs.

2.5. Liquid chromatography tandem mass spectrometry

The next steps were processed at the UC Davis Proteomics Core Facility. Peptides were trapped on a Thermo PepMap trap and separated on an Easy-spray 100 um x 25 cm C18 column using a Dionex Ultimate 3000 nUPLC at 200 nl/min. Solvent A= 0.1% formic acid, Solvent B = 100% Acetonitrile 0.1% formic acid. Gradient conditions = 2%B to 50%B over 60 minutes, followed by a 50%-99% B in 6 minutes and then held for 3 minutes than 99%B to 2%B in 2 minutes and total run time of 90 minutes using Thermo Scientific Fusion Lumos mass spectrometer running in Data Independent Acquisition (DIA) mode.

2.6. Chromatogram library creation

Six-gas phase fractionated (GFP) chromatogram library injections were made using staggered 4 Da isolation widows. GFP1 = 400-500 m/z, GFP2 = 500-600 m/z, GFP3 = 600-700

m/z, GFP4 = 700-800 m/z, GFP5 = 800-900 m/z, GFP6 = 900-1000 m/z, mass spectra were acquired using a collision energy of 35, resolution of 30 K, maximum inject time of 54 ms and a AGC target of 50K. Each individual sample was run in DIA mode with staggered isolation windows of 12 Da in the range 400-1000 m/z.

2.7. Analytic samples, data analysis and raw data processing

Each individual sample was run in DIA mode using the same settings as the chromatogram library runs except using staggered isolation windows of 12 Da in the m/z range 400-1000 m/z. DIA data was analyzed using Scaffold DIA v.2.0.0 (Proteome Software, Portland, OR, USA). Raw data files were converted to mzML format using ProteoWizard v.3.0.11748 [35].

2.8. Spectral library search

The Reference Spectral Library was created by EncyclopeDIA v.0.9.2. Chromatogram library samples were individually searched against Prosit predicted databases created using Prosit online server (https://www.proteomicsdb.org/prosit/) and converted for ScaffoldDIA using the Encyclopedia tools [32]. The input for the Prosit prediction consisted of UniProt proteome UP000009183 (*Vitis vinifera*, Grape), UniProt proteome UP000000812 (*Xylella fastidiosa*) and 114 common laboratory contaminants (https://www.thegpm.org/crap/) with a peptide mass tolerance of 10.0 ppm and a fragment mass tolerance of 10.0 ppm. Variable modifications considered were oxidation of methionine and static modifications were carbamidomethyl of cysteine. The digestion enzyme was assumed to be Trypsin with a maximum of 1 missed cleavage site(s) allowed. Only peptides with charges in the range [2□3] and length in the range [6□30] were considered. Peptides identified in each search were filtered by Percolator (3.01.nightly-13-655e4c7-dirty) [36]–[38] to achieve a maximum FDR of 0.01. Individual search results were combined, and peptides were again filtered to an FDR threshold of 0.01 for inclusion in the reference library. A summary of the workflow is presented in Figure 1.

2.9. Quantification and criteria for protein identification

Peptide quantification was performed by EncyclopeDIA v. 0.9.2. For each peptide, the five highest quality fragment ions were selected for quantitation. Proteins that contained similar

peptides and could not be differentiated based on MS/MS analysis were grouped to satisfy the principles of parsimony. Proteins with a minimum of 2 identified peptides were thresholder to achieve a protein FDR threshold of 1.0%.

2.10. Functional enrichment analysis

The functional analysis of proteomics of vascular leaf sap of grapevines was performed by the online software Metascape [39] using the express analysis settings. The up and downregulated *Vitis vinifera* protein IDs of diseased samples were converted into the corresponding *Arabidopsis* homolog protein IDs and analyzed independently. The *Arabidopsis* homologs were identified in TAIR using Protein Basic Local Alignment Search Tool (BLASTP). Metascape identified pathways and process enrichment analysis defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG). P-value was adjusted by the method of Benjamin-Hochberg to control the false discovery rate (FDR).

3. Results

3.1. Creating a DIA library and improving the datamining of xylem proteome data.

In this study, we compared the proteome of vascular leaf sap from healthy grapevines to those developing PD symptoms due to *X. fastidiosa* (*Xf*) infection. Infection was confirmed by qPCR that quantified a high number of bacterial cells $1.5x10^9$ cells/mL present in the diseased samples (Table S1). The vascular system is particularly crucial for this pathosystem as *Xf* cells are restricted to this microenvironment within plants. Thus, much of its interaction with the host occurs on the surface of xylem cells. As proteomic methods and equipment are rapidly evolving, we investigated the effect of a new deep neural proteome prediction method, Prosit, to identify proteins from mass spec data applied on Data Independent Acquisition (DIA) currently in use.

Figure 2a shows the proteomics results from vascular leaf sap of grapevines *Vitis vinifera* (VIT) at 12 weeks post-inoculation with *Xf*. DIA analysis identified 145 and 18 proteins for VIT (Table S2) and *Xf* (Table S3), respectively. After integrating Prosit into database search pipelines, the number of proteins increased by more than 148% for VIT and 400% for *Xf*, to a final total of 360 and 90 proteins (Tables S4 and S5). Only six VIT proteins were identified exclusively without Prosit and 221 only by integrating Prosit (DIA+Prosit), with 139 detected in

either approach for VIT (Fig.2b, Table S6). Among the six VIT proteins identified by DIA only, four are peroxidases (VIT_01s0010g01950, VIT_01s0010g01960, VIT_01s0010g02000, VIT_01s0010g02010), an uncharacterized protein with serine-type endopeptidase activity (VIT_16s0098g01160), and a Glyco_hydro_18 domain-containing protein (VIT_16s0050g02220). Nevertheless, the proteins detected exclusively by Prosit were associated with many more molecular functions, including cell adhesion molecules, scaffold/adaptors proteins, chaperones, translational proteins, transporters, and nucleic acid-binding proteins. Regarding the *Xf* bacterial proteins, 18 proteins were identified by both methods; however, DIA+Prosit allowed the detection of an additional 72 proteins that were not present in the DIA data (Table S7).

The application of Prosit to our data substantially increased the number of proteins with a molecular weight below 100 kDa. The range of molecular weight varied from 12 kDa to 217 kDa in DIA data and 8 kDa to 217 kDa in DIA+Prosit data. A breakdown of identified proteins on both methods by molecular weight and the number of mapped peptides is shown in Figure 3a. The smallest proteins predicted by DIA are AAI domain-containing proteins (VIT_02s0236g00020 and VIT_02s0236g00030) with 12 kDa, both upregulated in diseased plants. In addition to identifying more proteins, DIA+Prosit also increased the number of peptides identified for each protein. This is a significant advancement since we set a minimum of two mapped peptides per protein for it to be considered, considerably increasing the confidence and reducing false discoveries. The maximum of peptides identified per protein for DIA was 22, and for DIA+Prosit was 31 peptides. Most of the proteins identified after Prosit integration showed 2 to 10 peptides per protein (Fig.3b). In DIA+Prosit data, 8 kDa was the smallest protein detected, identified as BBE domain-containing protein (VIT_10s0003g05430) with a signal peptide targeting mitochondria (mTP) according to TargetP (Fig.4). Both AAI domain-containing proteins detected by DIA were also present with DIA+Prosit, and a third AAI domain-containing protein (VIT_16s0013g00070) was also detected. This is yet another important improvement as protein families with multiple members represented in a dataset gain higher scores in functional analyses such as gene ontology or pathway mapping.

The analyzed material is an enriched vascular leaf sap; thus, we determined the proportions of proteins predicted to be secreted (Fig.4). The percentage of secreted proteins with a predicted signal peptide within the total proteins predicted for DIA was 68% (99/145), and for

DIA+Prosit was 57% (205/360), according to SignalP. By using TargetP to analyze the same data sets, we found similar results: 72% and 59% for DIA and DIA+Prosit, respectively. The remaining were classified as non-secretory targeting the mitochondria (1-2%), chloroplast (3-4%), or other (23-34%). By performing the same analysis using the prediction tool ChloroP, we showed that actually, 16% of the proteins in both data sets would target chloroplasts; therefore, their presence in the xylem sap possibly reflects some degree of cellular content contamination of the samples during vacuum-assisted sap extraction or alternatively products of natural cellular and organellar degradation.

3.2. Regulation of proteins secreted to the xylem during Pierce's disease

We used the MetaboAnalyst v.4.0 (https://www.metaboanalyst.ca) to visualize both proteome data sets and examined the variation between the groups and samples [40]. The variability was examined by the unsupervised principal component analysis (PCA), which showed a distinct separation between groups in both data sets, DIA and DIA+Prosit (Fig.5). In this case, the intense response to *Xf* proliferation is so marking that Prosit was unnecessary to efficiently cluster the samples by type; however, we cannot exclude the possibility Prosit would be decisive in more attenuated differences. Healthy and Diseased groups showed 85.6% variation in PC1 for DIA (Fig.5a) and 68% variation in PC1 for DIA+Prosit (Fig.5b). These results suggest the effect of *Xf* cells in the plant stress response in the proteome of the vascular leaf sap. The variation among samples explained by PC2 was 7.8% for DIA. Prosit increased the variation among samples to 16.3%, explained by PC2. For this clustering analysis, the third sample of the Healthy grapevines was discarded due to bad MS/MS data quality; therefore, a virtual sample was created using an average of the other two samples (W5 and W6; W5_W6). The protein levels in Healthy and Diseased group samples were distinct, independent of the method (Fig.S1). To further analyze the differences between methods, we analyzed the ratio-intensity of Healthy and Diseased groups and compared them to the protein abundance in both proteome data sets. The fold change of protein detection between Diseased and Healthy plants presented similar results for DIA and DIA+Prosit data (Fig.6a and 6b). However, the implementation of Prosit increased the detection of the proteins that were in low abundance, as shown by the x-axis in Figures 6a and 6b. The correlation of results obtained by both methods was significant and had

an R^2 of 0.8795 (Fig.6c), showing that the increase of protein prediction power by Prosit correlates well with the observed data without introducing bias in differential expression.

To visualize proteins that are significantly either up or downregulated in the Diseased group, we examined volcano plots of both data sets (Fig.7). The comparison of the $\log_2$ fold change of the data sets and their adj. P-values by false discovery rates show a similar profile; however, the integration of Prosit allowed the identification of additional proteins that were significantly up and downregulated in the Diseased group. We observed that DIA without Prosit was more restrictive, and the maximum fold changes between Diseased and Healthy plants were not as high. The three most upregulated proteins identified by DIA+Prosit were chitinase A (VIT_16s0050g02230), Cupredoxin superfamily protein (VIT_18s0001g11180), and beta-1,3-glucanase 3 (VIT_08s0007g06060 - PR-2 family of pathogenesis-related proteins). The most downregulated proteins were Plant invertase/pectin methylesterase inhibitor superfamily (VIT_07s0005g00720), Glyco_hydro_18 domain-containing protein (VIT_06s0004g03840), and FAD-binding berberine family protein (VIT_10s0003g05470).

For a balanced comparison between both methods, we used partial least squares - discriminant analysis (PLS-DA) of the 139 proteins that were detected by both methods. The VIP score (a metric that identifies which variables are most responsible for the differences between the classes in the analysis) was higher in DIA compared to DIA+Prosit. Among the top 25 proteins contributing to the variations among the two sample groups, we can highlight the pathogenesis-related proteins (PR1, PR2, PR3, PR4) that are upregulated in Diseased plants independent of the chosen method. Only five proteins among the top 25 in DIA were not in DIA+Prosit, and seven proteins were in DIA+Prosit and not DIA. The PLS-DA plots and the cross-validation test result are shown in Fig.S2.

3.3. Pathway regulation in grapevine vascular leaf sap

Representation of known enzyme pathways or protein complexes in vascular leaf sap proteome assists in the functional characterization of the plant response to infection and virulence strategies by the pathogen. The results showed that Prosit provides the identification of more pathways involved in defense during Pierce's disease symptom development. Proteins that were up or downregulated in the Diseased group were analyzed separately to detect enriched pathways in each condition. Figures 9 and 10 show the up and downregulated proteins in both

methods considering all the detected proteins for each (145 for DIA and 360 for DIA+Prosit). Most of the enriched pathways identified using DIA datasets were also present in DIA+Prosit. However, for DIA+Prosit, due to the higher number of proteins, more pathways significantly affected were revealed.

The proteins identified by DIA that were upregulated in Diseased samples were involved in aminoglycan catabolic process, response to bacterium, cell wall organization or biogenesis, innate immune response, gluconeogenesis, hydrogen peroxide catabolic process, and response to cadmium ion. The most statistically significant pathways (involved in aminoglycan catabolic process, response to bacterium, cell wall organization or biogenesis, and innate immune response) were upregulated in both DIA and DIA+Prosit data. The latter revealed a higher number of proteins thus higher coverage and also showed other pathways such as response to ion and carbon fixation in photosynthetic organisms as more significantly enriched (lower p-values).

The analysis of the downregulated proteins in the Diseased plants showed that except for the arabinan catabolic process, all the other identified pathways were significantly enriched in the DIA+Prosit approach, which revealed galactose metabolism, hexose metabolic process, reductive pentose phosphate cycle and response to cadmium ion.

4. Discussion

This was the first DIA study of vascular sap of grapevines using Prosit [33]. We used a pressure chamber to extract the vascular leaf sap from grapevines comparing healthy and diseased plants and submitted samples for proteome analysis. Previous studies of the grapevine xylem proteome have provided important clues regarding the plant responses to infection; however, they have also faced several technical challenges in extracting enough material to adequately describe the complexity of this pathosystem. The focus of this study was to show the application of DIA in combination with Prosit to improve protein prediction and quantification in the vascular sap of leaves infected with *Xylella fastidiosa*. Our results suggest that incorporating a deep learning architecture approach like Prosit to DIA data could help researchers identify more protein candidates in response to pathogenesis and other biological phenomena. Prosit significantly increased the number of proteins, especially in low abundance detected in both from *Vitis* and *Xylella*, contributing to a more detailed picture of this plant-pathogen interaction.

4.1. A new proteomic approach for vascular sap studies

The implementation of Prosit to the DIA data increased detected proteins from 145 to 360 for grapevines and from 18 to 90 proteins detected for *Xylella fastidiosa*. Proteomics studies from vascular plant sap have always faced technical challenges due to the low protein concentration present in this plant organ. Previous studies identified differently expressed transcripts and proteins in grapevines by 2D-PAGE for protein isolation and further detection by MS/MS. The maximum resolution for these sample types was around 100 proteins with molecular weights from 20 kDa-75 kDa, with a majority higher than 40 kDa [41]. The most recent proteomic study related to Pierce's disease detected 91 proteins by LC-MS/MS that ranged from 12 kDa-114 kDa. That study demonstrated that structural data could be incorporated in the pipeline of proteomic data analysis using CHURNER [25]. The number of identified peptides from these 91 proteins also ranged from 2-23 peptides. Combining DIA+Prosit with these complementary functional approaches might provide yet a deeper comprehension of the relevant processes taking place during infection and the molecular functions that could be targeted with priority for increased plant defense.

By using DIA and Prosit, the number of proteins increased as well as the sensitivity of the detection. The number of proteins in low abundance were mostly predicted by Prosit. That's because the intensity prediction improved the quality of peptide identification by data searching [33]. The molecular weight of proteins from our study ranged from 8 kDa - 217 kDa, significantly broader than in previous studies. The smallest protein predicted by DIA+Prosit was BBE domain-containing protein (VIT_10s0003g05430) with 8 kDa predicted only by DIA+Prosit with six exclusive peptides. This protein has been previously described as necessary in the plant-pathogen interaction of Vitis and *Botrytis cinerea*. BBE-like enzymes inactivate oligo galacturonides (OGs) accumulated as intermediate reaction products of the inhibition of polygalacturonases (PGs) by PG-inhibiting proteins (PGIPs) [42], [43]. By oxidizing OGs, those became less active as defense inducers and less susceptible to hydrolysis of the pathogen's PGs. The accumulation of OGs can compromise plant growth and resistance through cell death induction. Therefore, the downregulation of BBE-like enzymes in grapevines infected with *Xf* contribute to the plant's susceptibility. This is the first report of detection of this protein in grape xylem sap, only achieved with Prosit.

403 The largest protein was a member of the subtilase family (VIT_16s0098g00970), with 217
404 kDa detected by DIA and DIA+Prosit. These proteins control the establishment of systemic
405 induced resistance and immune priming by the detection of the biotic stimulus [44]. This protein
406 was not detected in Healthy plants in the DIA data, only in the Diseased plants with seven
407 identified peptides. In the DIA+Prosit data, the number of peptides increased and were then
408 detected in Healthy samples as well, but at lower levels compared to the Diseased. Prosit also
409 increased the number of detected peptides to nine. This result exemplifies the increase in
410 sensitivity by implementing Prosit to DIA data.
411
412 **4.2. Plant response to *X. fastidiosa* infection as assessed by the vascular sap**
413 Although the number of studies investigating expressed transcripts and proteins in the
414 xylem sap of plants infected with *Xf* is small, they have provided valuable information regarding
415 plant responses to infection [23], [25], [41]. By accurately evaluating the vascular leaf sap of
416 infected plants with *Xf* using a more sensitive and reproducible proteomic approach, our study
417 confirmed the presence of secreted proteins associated with pathogenesis-related (PR) proteins,
418 chitinases, and β-1-3-glucanases as the key players in mediating the defense response upon
419 pathogen infection [25]. Our study specifically revealed β-1-3-glucanase 3
420 (VIT_08s0007g06060) as the vital protein contributing to the variance between Healthy and
421 Diseased plants in the DIA+Prosit data (VIP = 3.1) and the second most important in the DIA
422 data (VIP = 2.7). A total of five β-1-3-glucanases proteins were also detected in the DIA+Prosit,
423 and only four were in the DIA data. Except for one protein (VIT_06s0061g00100) that was
424 slightly downregulated in the diseased plants, all the others in both data sets were upregulated. β-
425 1-3-glucanases belong to the PR2 class, and their expression is induced by several pathogens
426 including fungi, oomycetes and most recently shown to be induced by a bacterial infection [25],
427 [41], [45], [46]. Other PR proteins (including PR1), proteases, chitinases, and peroxidases were
428 also confirmed in our study but in a higher number of proteins. Chakraborty et al. (2016) were
429 able to detect 15 peroxidases, and our DIA and DIA+Prosit increased this number to 20. This
430 could be due to the Prosit predictions being generalized to non-tryptic peptides increasing
431 peptide predictions [33].
432
433 **5. Conclusions**

This study demonstrated a successful example of using the DIA approach combined with deep learning neural network Prosit for analysis of proteomic data. A total of 360 proteins were identified and quantified from the grapevines subjected to *Xf* inoculation. We also identified different sets of proteins regulated upon infection that were previously shown in other proteomic studies and highlighted new low molecular weight and low abundance proteins previously undetected. This is especially useful in samples with a lower protein abundance and diversity, providing more functional clues of significant players.

6. Acknowledgments and Funding

We thank Samuel Metcalf for sharing his expertise with the pressure chamber and Ken A. Shackel for allowing us to use his equipment. This work was supported by grants obtained from the California Department of Food and Agriculture Pierce's Disease Board (CDFA-PD Board). LC-MS was supported by a NIH shared instrumentation grant S10OD021801. C.H.D.S was supported by a Coordination for the Improvement of Higher Level Personnel grant (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brazil, No. 99999.013202/2013-08).

7. Author Contributions

CHDS and AMD conceived and designed the experiments; CHDS coordinated and performed experiments, functional analysis and wrote of the manuscript edited by PAZ and AMD; RABA contributed in the data analysis and discussions; HS helped data analysis using MetaboAnlyst; MS helped with protein digestion and performed LC-MS/MS and chromatogram library creation; AJ helped plant material and inoculations, BSP conducted proteome data analysis, raw data processing and spectral library search; AMD and all others revised the final manuscript.

8. Conflicts of Interest

The authors declare no conflict of interest. The sponsors had no role in the design, execution, interpretation, or writing of the study

9. References

[1] V. De Schepper, T. De Swaef, I. Bauweraerts, and K. Steppe, "Phloem transport: A review of mechanisms and controls," *Journal of Experimental Botany*, vol. 64, no. 16. pp. 4839–4850, 2013, doi: 10.1093/jxb/ert302.

[2] W. Tanner and H. Beevers, “Transpiration, a prerequisite for long-distance transport of minerals in plants?,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 16, pp. 9443–9447, 2001, doi: 10.1073/pnas.161279898.

[3] P. Carella, D. C. Wilson, C. J. Kempthorne, and R. K. Cameron, “Vascular Sap Proteomics: Providing Insight into Long-Distance Signaling during Stress,” *Front. Plant Sci.*, vol. 7, no. MAY2016, p. 651, May 2016, doi: 10.3389/fpls.2016.00651.

[4] G. Drakakaki and A. Dandekar, “Protein secretion: How many secretory routes does a plant cell have?,” *Plant Science*, vol. 203–204. pp. 74–78, Apr-2013, doi: 10.1016/j.plantsci.2012.12.017.

[5] G. K. Agrawal, N. S. Jwa, M. H. Lebrun, D. Job, and R. Rakwal, “Plant secretome: Unlocking secrets of the secreted proteins,” *Proteomics*, vol. 10, no. 4. pp. 799–827, Feb-2010, doi: 10.1002/pmic.200900514.

[6] K. A. Yadeta and B. P. H. J. Thomma, “The xylem as battleground for plant hosts and vascular wilt pathogens,” *Frontiers in Plant Science*, vol. 4, no. APR. Frontiers Research Foundation, 23-Apr-2013, doi: 10.3389/fpls.2013.00097.

[7] M. J. Davis, A. H. Purcell, and S. V. Thomson, “Pierce’s Disease of Grapevines: Isolation of the Causal Bacterium,” *Science (80-.).*, vol. 199, no. 4324, pp. 75–77, 1978.

[8] V. Rossetti *et al.*, “Occurrence of xylem-restricted bacteria in sweet orange trees affected by chlorotic variegation, a new citrus disease in Brazil.,” *Comptes Rendus l’Academie des Sci. Ser. 3, Sci. la Vie*, vol. 310, no. 8, pp. 345–349, 1990.

[9] G. P. Martelli, “The current status of the quick decline syndrome of olive in southern Italy,” *Phytoparasitica*, vol. 44, no. 1, pp. 1–10, 2016, doi: 10.1007/s12600-015-0498-6.

[10] A. J. G. Simpson *et al.*, “The genome sequence of the plant pathogen Xylella fastidiosa,” *Nature*, vol. 406, pp. 151–157, 2000, doi: 10.1038/35018003.

[11] E. T. Thorne, J. F. Stevenson, T. L. Rost, J. M. Labavitch, and M. A. Matthews, “Pierce’s disease symptoms: Comparison with symptoms of water deficit and the impact of water deficits,” *Am. J. Enol. Vitic.*, vol. 57, no. 1, pp. 1–11, 2006.

[12] S. Chatterjee, C. Wistrom, and S. E. Lindow, “A cell-cell signaling sensor is required for virulence and insect transmission of [i]Xylella fastidiosa[/i] .,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 7, pp. 2670–5, 2008, doi: 10.1073/pnas.0712236105.

[13] S. Chatterjee, R. P. Almeida, and S. Lindow, “Living in two worlds: the plant and insect lifestyles of Xylella fastidiosa,” *Annu Rev Phytopathol*, vol. 46, pp. 243–271, 2008.

[14] M. Ionescu, C. Baccari, A. M. Da Silva, A. Garcia, K. Yokota, and S. E. Lindow, “Diffusible signal factor (DSF) synthase RpfF of Xylella fastidiosa is a multifunction protein also required for response to DSF,” *J. Bacteriol.*, vol. 195, no. 23, pp. 5273–5284, Dec. 2013, doi: 10.1128/JB.00713-13.

[15] Q. Sun, Y. Sun, M. Andrew Walker, and J. M. Labavitch, “Vascular occlusions in grapevines with Pierce’s disease make disease symptom development worse,” *Plant Physiol.*, vol. 161, no. 3, pp. 1529–1541, Mar. 2013, doi: 10.1104/pp.112.208157.

[16] R. Nascimento *et al.*, “The Type II Secreted Lipase/Esterase LesA is a Key Virulence

Factor Required for Xylella fastidiosa Pathogenesis in Grapevines," *Sci. Rep.*, vol. 6, no. 1, pp. 1–17, Jan. 2016, doi: 10.1038/srep18598.

[17] O. R. Feitosa-Junior *et al.*, "Proteomic and Metabolomic Analyses of *Xylella fastidiosa* OMV-Enriched Fractions Reveal Association with Virulence Factors and Signaling Molecules of the DSF Family," *Phytopathology*, vol. 109, no. 8, pp. 1344–1353, Aug. 2019, doi: 10.1094/PHYTO-03-19-0083-R.

[18] H. Gouran *et al.*, "The Secreted Protease PrtA Controls Cell Growth, Biofilm Formation and Pathogenicity in Xylella fastidiosa," *Sci. Rep.*, vol. 6, no. April, p. 31098, 2016, doi: 10.1038/srep31098.

[19] N. P. Cianciotto and R. C. White, "Expanding Role of Type II Secretion in Bacterial Pathogenesis and Beyond," 2017, doi: 10.1128/IAI.00014-17.

[20] A. da S. Santiago *et al.*, "The Antitoxin Protein of a Toxin-Antitoxin System from Xylella fastidiosa Is Secreted via Outer Membrane Vesicles," *Front. Microbiol.*, vol. 7, no. DEC, p. 2030, Dec. 2016, doi: 10.3389/fmicb.2016.02030.

[21] C. B. Agüero, E. T. Thorne, A. M. Ibáñez, W. D. Gubler, and A. M. Dandekar, "Xylem sap proteins from Vitis vinifera L. Chardonnay," *Am. J. Enol. Vitic.*, vol. 59, no. 3, pp. 306–311, 2008.

[22] B. Delaunois *et al.*, "Large-scale proteomic analysis of the grapevine leaf apoplastic fluid reveals mainly stress-related proteins and cell wall modifying enzymes," *BMC Plant Biol.*, vol. 13, no. 1, pp. 1–15, Feb. 2013, doi: 10.1186/1471-2229-13-24.

[23] R. Katam, K. Chibanguza, L. M. Latinwo, and D. Smith, "Proteome Biomarkers in Xylem Reveal Pierce's Disease Tolerance in Grape," 2015, doi: 10.4172/jpb.1000372.

[24] L. Yang, H. Lin, Y. Takahashi, F. Chen, M. A. Walker, and E. L. Civerolo, "Proteomic analysis of grapevine stem in response to Xylella fastidiosa inoculation," *Physiol. Mol. Plant Pathol.*, vol. 75, no. 3, pp. 90–99, Jan. 2011, doi: 10.1016/j.pmpp.2010.11.002.

[25] S. Chakraborty *et al.*, "Sequence/structural analysis of xylem proteome emphasizes pathogenesis-related proteins, chitinases an β-1, 3-glucanases as key players in grapevine defense against Xylella fastidiosa," *PeerJ*, vol. 2016, no. 5, p. e2007, May 2016, doi: 10.7717/peerj.2007.

[26] S. Magdeldin *et al.*, "Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis," *Clin. Proteomics*, vol. 11, no. 1, pp. 1–10, 2014, doi: 10.1186/1559-0275-11-16.

[27] F. Zhang, W. Ge, G. Ruan, X. Cai, and T. Guo, "Data-Independent Acquisition Mass Spectrometry-based Proteomics and Software Tools: A Glimpse in 2020," *Proteomics*, vol. 1900276, p. 1900276, 2020, doi: 10.1002/pmic.201900276.

[28] J. D. Venable, M. Q. Dong, J. Wohlschlegel, A. Dillin, and J. R. Yates, "Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra," *Nat. Methods*, vol. 1, no. 1, pp. 39–45, 2004, doi: 10.1038/nmeth705.

[29] D. Martins-de-Souza, V. M. Faça, and F. C. Gozzo, "DIA is not a new mass spectrometry acquisition method," *Proteomics*, vol. 17, no. 7, Apr. 2017, doi: 10.1002/pmic.201700017.

[30] Y. Yang, X. Liu, C. Shen, Y. Lin, P. Yang, and L. Qiao, "In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics," *Nat. Commun.*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-019-13866-z.

[31] B. C. Searle *et al.*, "Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry," *Nature Communications*, vol. 9, no. 1, 2018.

[32] B. C. Searle *et al.*, "Generating high quality libraries for DIA MS with empirically corrected peptide predictions," *Nat. Commun.*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-15346-1.

[33] S. Gessulat *et al.*, "Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning," *Nat. Methods*, vol. 16, no. 6, pp. 509–518, Jun. 2019, doi: 10.1038/s41592-019-0426-7.

[34] M. Francis, H. Lin, J. C. La Rosa, H. Doddapaneni, and E. L. Civerolo, "Genome-based PCR primers for specific and sensitive detection and quantification of Xylella fastidiosa," *Eur. J. Plant Pathol.*, vol. 115, no. 2, pp. 203–213, 2006, doi: 10.1007/s10658-006-9009-4.

[35] M. C. Chambers *et al.*, "A cross-platform toolkit for mass spectrometry and proteomics," *Nature Biotechnology*, vol. 30, no. 10. Nature Publishing Group, pp. 918–920, 10-Oct-2012, doi: 10.1038/nbt.2377.

[36] L. Käll, J. D. Canterbury, J. Weston, W. S. Noble, and M. J. MacCoss, "Semi-supervised learning for peptide identification from shotgun proteomics datasets," *Nat. Methods*, vol. 4, no. 11, pp. 923–925, Nov. 2007, doi: 10.1038/nmeth1113.

[37] L. Käll, J. D. Storey, M. J. MacCoss, and W. S. Noble, "Assigning significance to peptides identified by tandem mass spectrometry using decoy databases," *Journal of Proteome Research*, vol. 7, no. 1. American Chemical Society, pp. 29–34, Jan-2008, doi: 10.1021/pr700600n.

[38] L. Käll, J. D. Storey, and W. S. Noble, "Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry," in *Bioinformatics*, 2008, vol. 24, no. 16, pp. 42–48, doi: 10.1093/bioinformatics/btn294.

[39] Y. Zhou *et al.*, "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nat. Commun.*, vol. 10, no. 1, 2019, doi: 10.1038/s41467-019-09234-6.

[40] J. Chong, D. S. Wishart, and J. Xia, "Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis," *Curr. Protoc. Bioinforma.*, vol. 68, no. 1, Dec. 2019, doi: 10.1002/cpbi.86.

[41] S. M. Basha, H. Mazhar, and H. K. N. Vasanthaiah, "Proteomics approach to identify unique xylem sap proteins in pierce's disease-tolerant vitis species," *Appl. Biochem. Biotechnol.*, vol. 160, no. 3, pp. 932–944, Feb. 2010, doi: 10.1007/s12010-009-8620-1.

[42] R. M. Kalunke, S. Tundo, M. Benedetti, F. Cervone, G. De Lorenzo, and R. D'Ovidio, "An update on polygalacturonase-inhibiting protein (PGIP), aleucine-rich repeat protein

that protects crop plants against pathogens," *Frontiers in Plant Science*, vol. 6, no. MAR. Frontiers Media S.A., p. 146, 20-Mar-2015, doi: 10.3389/fpls.2015.00146.

[43] M. C. Héloir *et al.*, "Recognition of Elicitors in Grapevine: From MAMP and DAMP Perception to Induced Resistance," *Frontiers in Plant Science*, vol. 10. Frontiers Media S.A., p. 1117, 18-Sep-2019, doi: 10.3389/fpls.2019.01117.

[44] A. Figueiredo, F. Monteiro, and M. Sebastiana, "Subtilisin-like proteases in plant-pathogen recognition and immune priming: a perspective," *Front. Plant Sci.*, vol. 5, no. DEC, p. 739, Dec. 2014, doi: 10.3389/fpls.2014.00739.

[45] C. Giannakis, C. S. Bucheli, K. G. M. Skene, S. P. Robinson, and N. Steele Scott, "Chitinase and β-1,3-glucanase in grapevine leaves: A possible defence against powdery mildew infection," *Aust. J. Grape Wine Res.*, vol. 4, no. 1, pp. 14–22, 1998, doi: 10.1111/j.1755-0238.1998.tb00130.x.

[46] P. Mestre *et al.*, "Identification of a Vitis vinifera endo-β-1,3-glucanase with antimicrobial activity against Plasmopara viticola," *Mol. Plant Pathol.*, vol. 18, no. 5, pp. 708–719, Jun. 2017, doi: 10.1111/mpp.12431.

602 **10. Tables**
603
604
605 **Table 1.** Overview of Proteomics studies of vascular sap of grapevines.
606

Vitis sp. variety	Biological material	*X. fastidiosa* inoculation	Method	Peptide spectra Analysis	Total proteins	Protein size (kDa)	Matched peptides	Signal peptide detection	Reference
Chardonnay	Xylem sap	No	2D-PAGE MALDI-TOF MS/MS	GPM	10	25-150	1	No	Agüero et al., 2008
PD tolerant and susceptible varienties	Xylem sap	No	2D-PAGE LC-MS/MS	Mascot	~100	20-75	1-4	No	Bascha et al., 2010
PD tolerant and susceptible varienties	Stem	Yes	2D-PAGE nano-LC-MS/MS	Bioworks	~200	14.4-45	2-32	No	Yang et al., 2010
Chardonnay	Leaf and Apoplastic fluid	No	2D-PAGE MALDI-TOF MS/MS	Mascot	227 and 89	15-120	-	No	Delaunois et al., 2013
PD tolerant and susceptible varienties	Xylem tissue	No	2D-PAGE MALDI-TOF MS/MS	Mascot	~200	20-75	-	No	Katam et al., 2015
Thompson Seedless	Xylem sap	Yes	LC-MS/MS	Scaffold	91	10-114	2-23	Yes	Chakraborty et al., 2016
Thompson Seedless	Vascular leaf sap	Yes	LC-MS/MS	ScaffoldDIA	145	12-217	2-22	Yes	*This study*
Thompson Seedless	Vascular leaf sap	Yes	LC-MS/MS	ScaffoldDIA + Prosit	360	8-217	2-31	Yes	*This study*

607
608

11. Figures

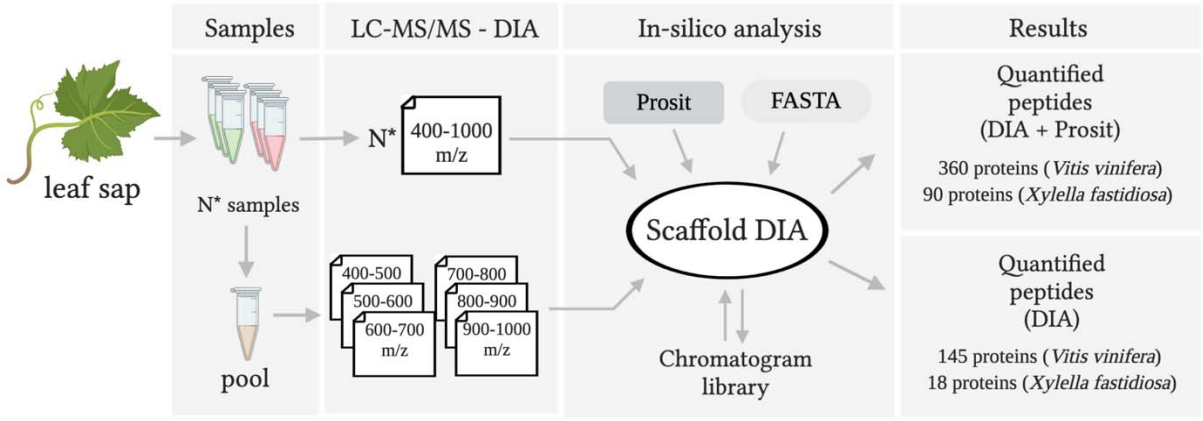

Figure 1. Quantification of peptides with chromatogram libraries workflow. The chromatogram library generation was based on Searle et al. (2018). In summary, each quantitative replicate (analytic samples) for each group was measured by wide-window DIA experiment (400-1000 m/z) besides the collection of several staggered narrow-window DIA experiment from the pooled sample of all samples. Afterwards, these narrow-window experiments have 2 m/z precursor isolation targeting every peptide between 400 and 1000 m/z. The peptides anchors were detected using ScaffoldDIA. Chromatographic data about each peptide was stored in a chromatogram library with retention times, peak shape, fragment ion intensities, and known interferences tuned specifically for the LC-MS/MS setting. ScaffoldDIA uses these precise coordinates for m/z, time, and intensity to detect peptides in the quantitative samples generating the DIA results box. Alternatively, in addition to the chromatogram library generated, a predicted library created using Prosit and FASTA information was added to determine quantified peptides and generated the DIA + Prosit results box. Created with BioRender.com.

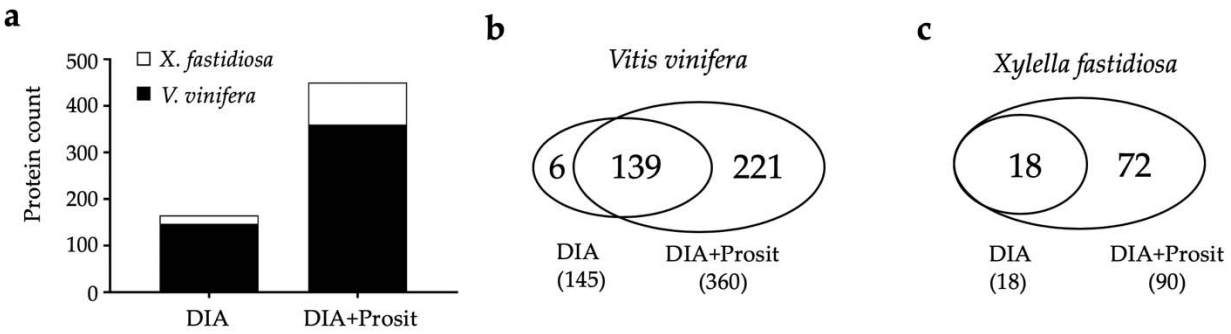

Figure 2. Proteomic analysis of *Vitis vinifera* and *Xylella fastidiosa*: a) total proteins identified by data-independent acquisition (DIA) and DIA+Prosit; b) Venn diagram of the number of proteins identified by each method for *V. vinifera*; and c) for *X. fastidiosa*.

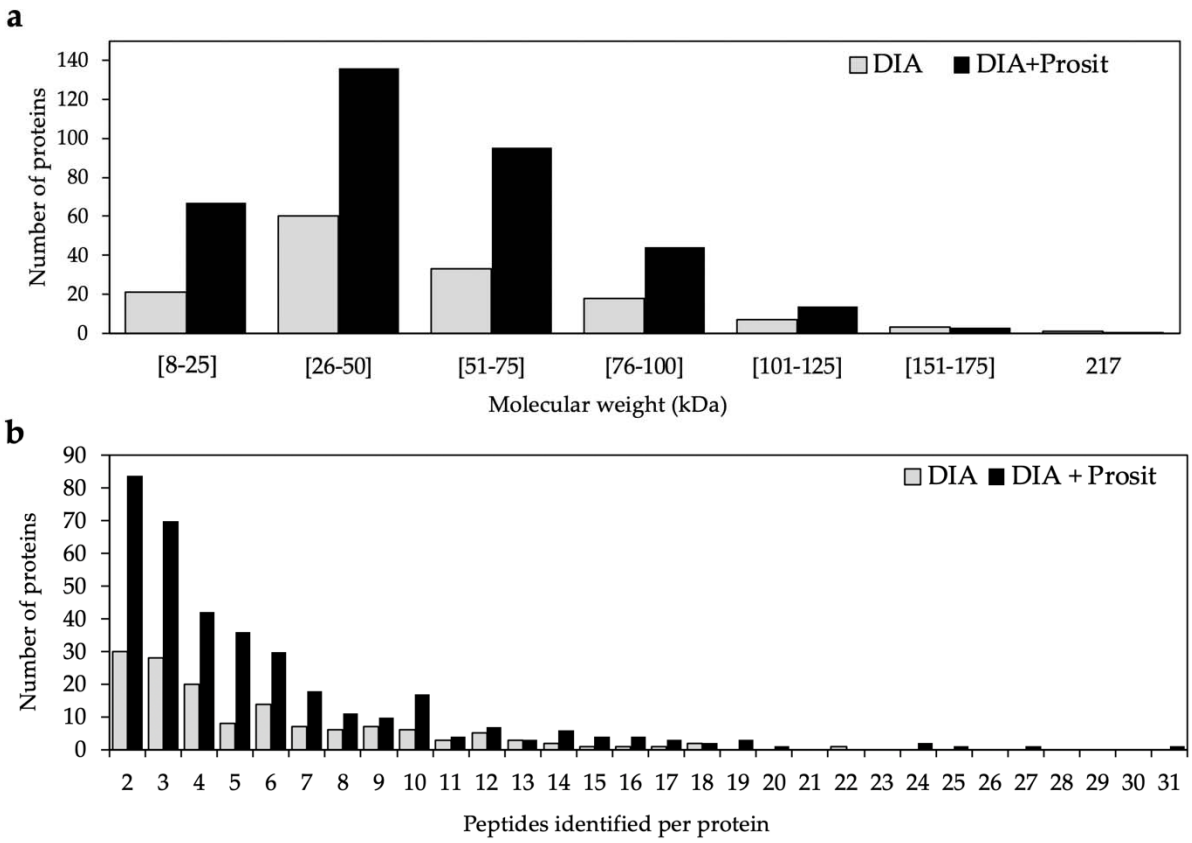

Figure 3. Distribution of the total number of proteins of *V. vinifera* identified by DIA and DIA+Prosit by a) molecular weight (kDa) ranging from 8 to 217 kDa. b) Identified peptides varying from 2 to 31 peptides per protein. Predicted proteins with only one peptide were discarded.

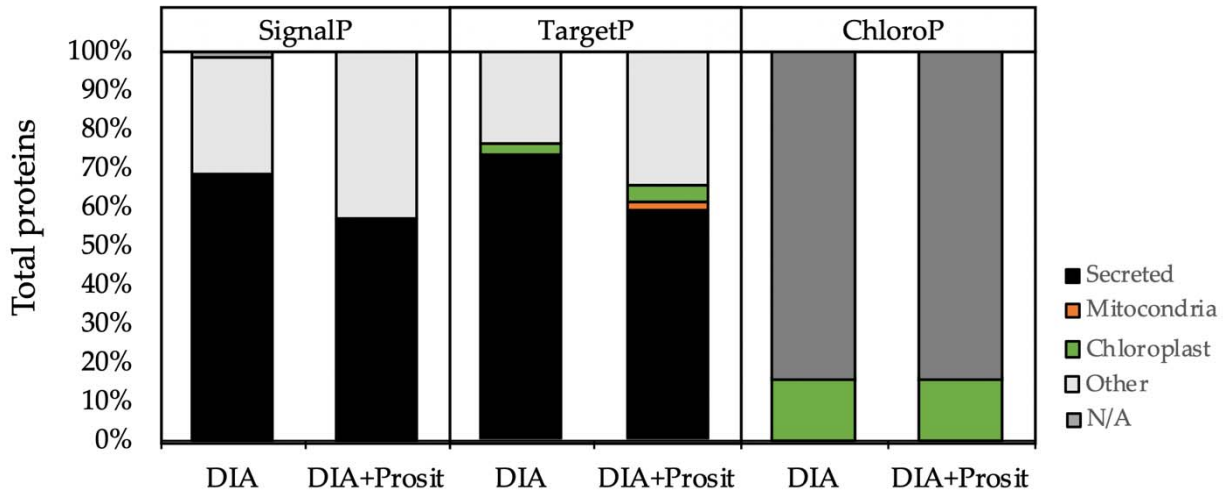

Figure 4. Subcellular localization prediction analysis and comparisons between DIA and DIA+Prosit data using SignalP, TargetP, and ChloroP servers. More than 50% of the total proteins identified were predicted having a signal peptide, according to SignalP and TargetP-SP. TargetP output revealed less than 3% of total proteins containing a mitochondrial targeting peptide (mTP) and less than 5% of proteins containing a chloroplast transit peptide (cTP). ChloroP predicted 16% of the collected vascular sap targeting the chloroplast by both methods. DIA considered a total of 145 proteins and DIA+Prosit, a total of 360 proteins for *V. vinifera*.

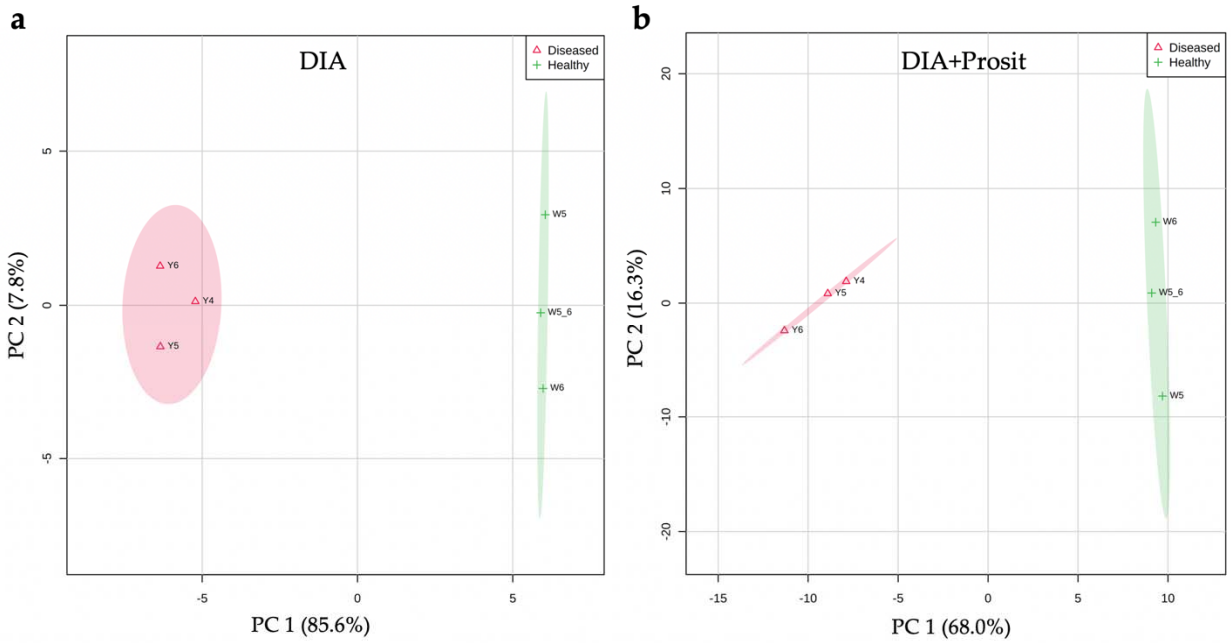

Figure 5. Principal Component Analysis (PCA) scores plots between PC1 and PC2 and explained variances are shown. The clear distinction between Diseased vs. Healthy proteomic data for *V. vinifera* at 12 weeks post-inoculation in both methods DIA and DIA+Prosit. W5_6 is a virtual sample made of the average of data for plants W5 and W6 (Healthy plants), and Y4, Y5, and Y5 were individual Diseased plants.

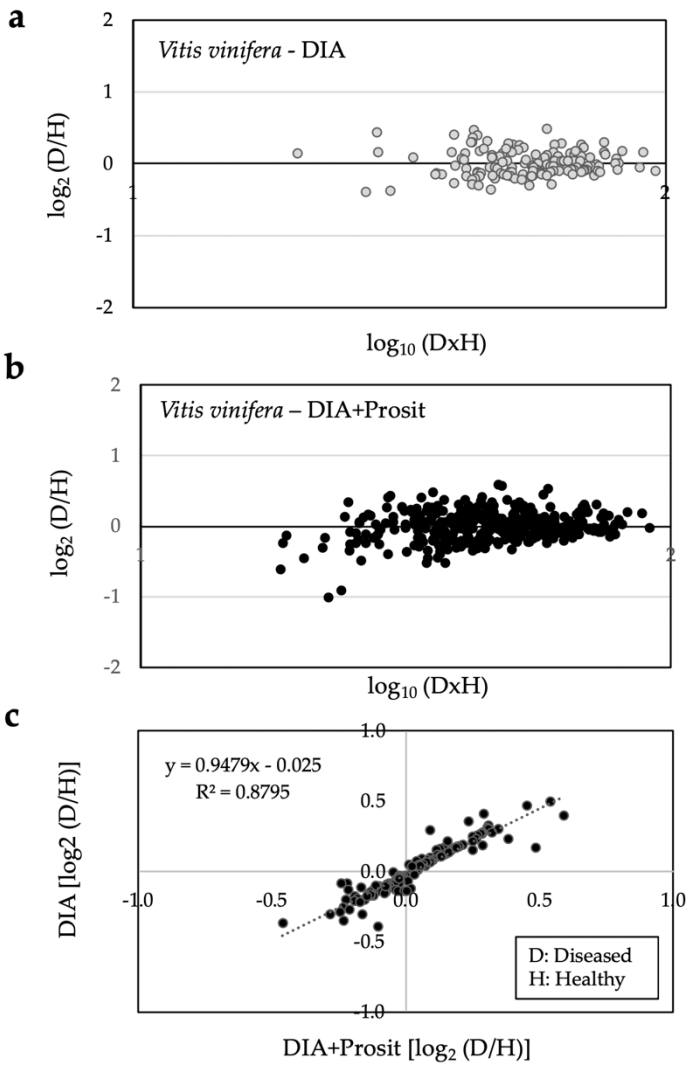

Figure 6. Overview of the plant response to *Xf* in Diseased samples in both data sets. Analysis of ratio-intensity plots displaying the $\log_2$ D/H fold-change ratio of Diseased over Healthy plants for each protein as a function of the abundance by $\log_{10}$ DxH product intensities: a) 145 proteins identified using DIA and b) 360 proteins identified by DIA+Prosit; c) Correlation between the ratios obtained from both analyses from the proteins detected in both analysis (139) with R^2 = 0.8795 show that the incorporation of Prosit maintained provided similar results but with higher quality and expanded the detection. D: Diseased and H: Healthy plants. The $\log_{10}$ exclusive intensity data for each protein using an FDR>1% was used for both analyses.

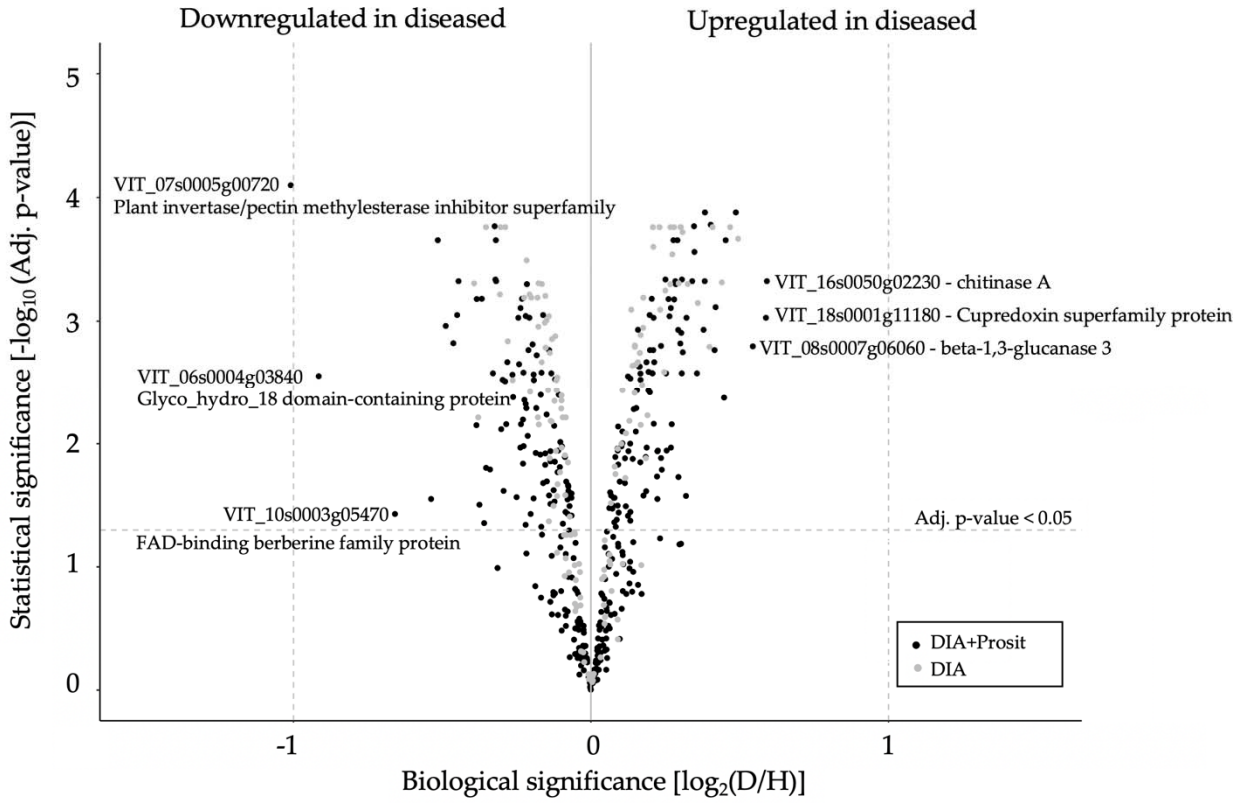

Figure 7. Proteome response of *V. vinifera* to *Xf* infection. Volcano plot analysis of Diseased (D) and Healthy (H) plants data identified by DIA and DIA+Prosit overlapped. Proteins identified by DIA are represented in grey dots and identified by DIA+Prosit in black dots. Adj. p-value calculated by Benjamin-Hochberg's false discovery rate greater than or equal to 0.05 were considered significant.

a

DIA				
VIP	H	D	Accession ID	Protein name
5.2			VIT_16s0098g00970	Subtilase family protein
2.7			VIT_08s0007g06060	Beta1,3-glucanase 3 (PR2)
2.3			VIT_14s0066g00620	Glyco_hydro_18 domain containing protein*
2.2			VIT_03s0088g00810	Pathogenesis related gene 1 (PR1)
1.8			VIT_03s0091g00160	Plant basic secretory protein BSP family protein
1.8			VIT_16s0050g02230	Chitinase A
1.8			VIT_07s0005g03140	Eukaryotic aspartyl protease family protein
1.7			VIT_04s0008g01150	Glycosyl hydrolases family 32 protein
1.6			VIT_03s0088g00700	Cysteine rich secretory proteins (CAP)
1.6			VIT_12s0059g02420	Peroxidase superfamily protein*
1.6			VIT_02s0025g04330	Osmotin 34
1.6			VIT_04s0008g00120	Basic chitinase (PR3)
1.5			VIT_03s0063g02490	O-Glycosyl hydrolases family 17 protein
1.5			VIT_13s0019g02490	Subtilase family protein*
1.5			VIT_08s0007g06040	Beta1,3-glucanase 3 (PR2)
1.5			VIT_02s0025g04250	Osmotin 34*
1.4			VIT_10s0071g01130	Alpha-galactosidase 2
1.4			VIT_11s0206g00030	Glycosyl hydrolase family protein
1.4			VIT_01s0127g00560	DC1 domain-containing protein
1.4			VIT_04s0008g01140	Beta-fructofuranosidase 5
1.4			VIT_14s0081g00030	Pathogenesis related 4 (PR4)
1.4			VIT_06s0061g00120	Beta1,3-glucanase 3 (PR2)
1.3			VIT_02s0025g02170	Trehalase 1*
1.3			VIT_07s0005g06000	Uncharacterized prot. related to pathogenesis
1.2			VIT_04s0008g07080	Eukaryotic aspartyl protease family protein

b

DIA+Prosit				
VIP	H	D	Accession ID	Protein name
3.1			VIT_08s0007g06060	Beta-1,3-glucanase 3 (PR2)
3.1			VIT_16s0050g02230	Chitinase A
2.6			VIT_03s0088g00700	Cysteine-rich secretory protein (CAP)
2.6			VIT_03s0088g00810	Pathogenesis-related gene 1 (PR1)
2.2			VIT_00s0357g00080	Cellulase glycosyl hydrolase family 5 protein*
2.1			VIT_10s0071g01130	Alpha-galactosidase 2
2.0			VIT_04s0008g00140	Chitin-binding type-1 domain-containing prot.*
2.0			VIT_04s0008g00120	Basic chitinase
2.0			VIT_03s0091g00160	Plant basic secretory protein BSP family prot.*
1.8			VIT_04s0008g01140	Beta-fructofuranosidase 5
1.8			VIT_02s0025g04330	Osmotin 34
1.7			VIT_01s0127g00560	DC1 domain-containing protein
1.7			VIT_03s0063g02490	O-Glycosyl hydrolases family 17 protein
1.7			VIT_06s0061g00120	Beta-1,3-glucanase 3 (PR2)
1.6			VIT_08s0007g06040	Beta-1,3-glucanase 3 (PR2)
1.6			VIT_14s0081g00030	Pathogenesis-related 4 (PR4)
1.5			VIT_07s0005g03140	Eukaryotic aspartyl protease family protein*
1.5			VIT_07s0005g06000	Uncharacterized prot. related to pathogenesis
1.4			VIT_12s0028g00610	Eukaryotic aspartyl protease family protein
1.4			VIT_11s0206g00030	Glycosyl hydrolase family protein
1.4			VIT_04s0008g07080	Eukaryotic aspartyl protease family protein
1.4			VIT_03s0063g00400	Alpha-amylase-like*
1.3			VIT_04s0008g01150	Glycosyl hydrolases family 32 protein*
1.3			VIT_02s0025g04270	Osmotin 34
1.3			VIT_16s0098g00970	Subtilase family protein

Figure 8. Top 25 proteins of *V. vinifera* contributing to the variance between the groups observed by PLS-DA. The plot shows the variable importance in projection (VIP) scores, and the colored boxes indicate the relative intensity detected by DIA and DIA+Prosit of the corresponding protein in Diseased and Healthy plants. Red represents high and green, low exclusive intensity detected. Proteins marked with (*) are exclusive among the top 25 of the respective method.

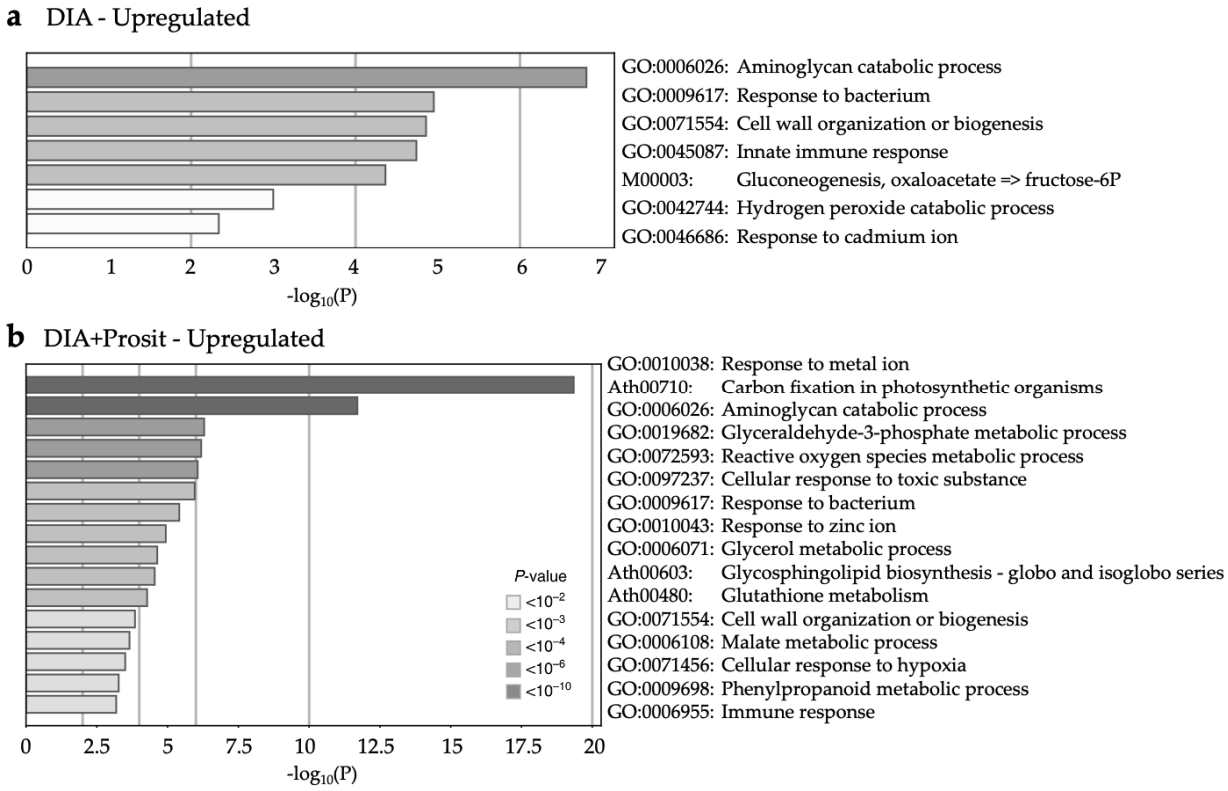

Figure 9. Upregulated pathways during *Xf* infection in *V. vinifera*. Non-redundant enriched ontology clusters of significantly expressed proteins upregulated during *Xf* infection ($p<0.05$) in a) DIA and b) DIA+Prosit data sets. DIA+Prosit allows the identification of a higher number of pathways likely involved with plant response to bacteria.

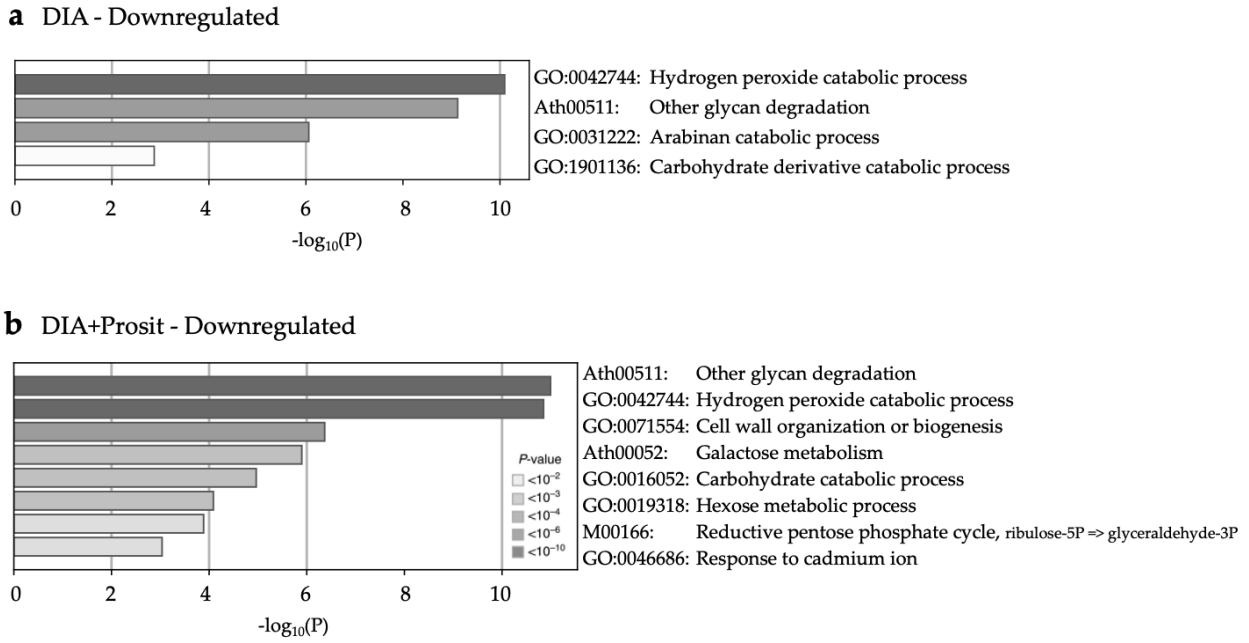

Figure 10. Downregulated pathways during *Xf* infection in *V. vinifera*. Non-redundant enriched ontology clusters of significantly expressed proteins downregulated during *Xf* infection ($p<0.05$) in a) DIA and b) DIA+Prosit data sets. Similarly to Fig.8, DIA+Prosit allowed the identification of a higher number of pathways likely involved with plant response to infection.

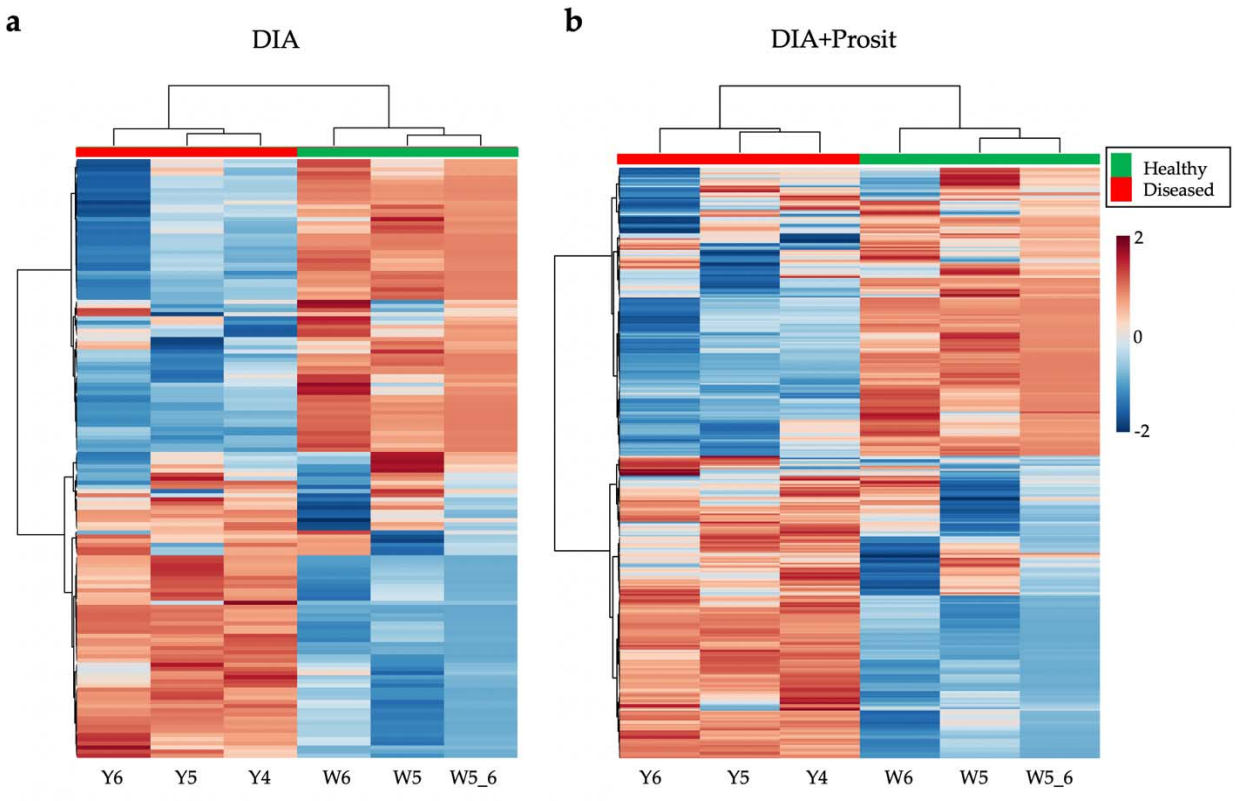

Figure S1. Heat map visualization of the effects of *X. fastidiosa* in grapevines through proteomic analysis of vascular sap of leaves. Hierarchical clustering using Euclidean distance and Ward's linkage for the clustering algorithm. Samples Y4, Y5, Y6 from Diseased plants and W5 and W6 from Healthy plants. Sample W5_6 is a virtual sample made from the average of W5 and W6. Log_{10} of the exclusive intensity was used from a) DIA and b) DIA+Prosit data sets.

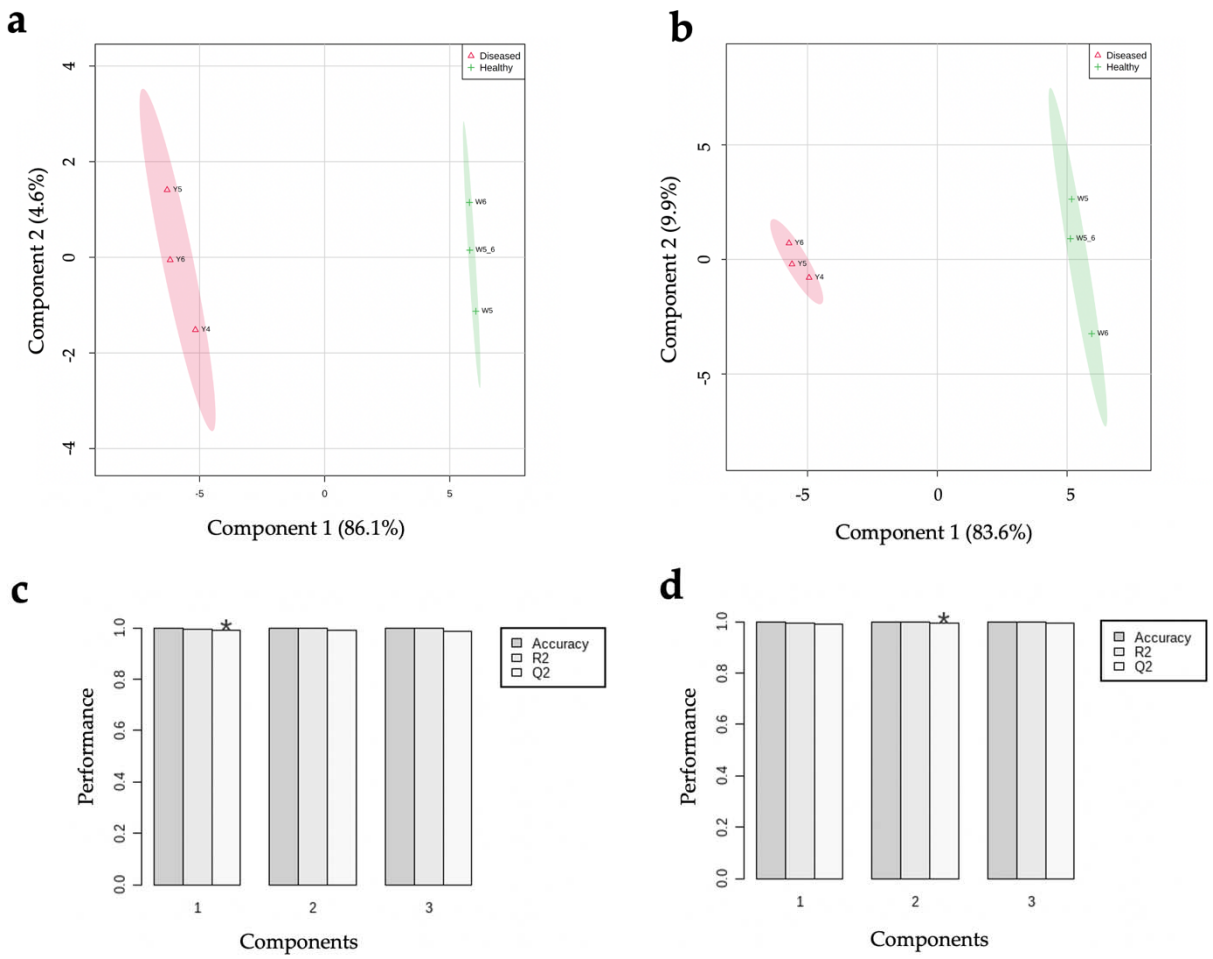

Figure S2. PLS-DA plots and cross-validation of *V. vinifera* samples of Diseased and Healthy plants using a) for DIA data and b) for DIA+Prosit data set. Validation of both models shown by R2 (the sum of squares captured by the model) and Q2 (cross□validation of R2) for the first three components for c) DIA and d) DIA+Prosit. By using Q2, the star indicates the best number of components for the model.