Sequence/structural analysis of xylem proteome emphasizes pathogenesis-related proteins, chitinases and β-1,3-glucanases as key players in grapevine defense against *Xylella fastidiosa*

Sandeep Chakraborty, Rafael Nascimento, Paulo A. Zaini, Hossein Gouran, Basuthkar J. Rao, Luiz R. Goulart, Abhaya M. Dandekar

Background. *Xylella fastidiosa*, the causative agent of various plant diseases including Pierce's disease in the US, and Citrus Variegated Chlorosis in Brazil, remains a continual source of concern and economic losses, especially since almost all commercial varieties are sensitive to this Gammaproteobacteria. Differential expression of proteins in infected tissue is an established methodology to identify key elements involved in plant defense pathways. **Methods**. In the current work, we developed a methodology named CHURNER that emphasizes relevant protein functions from proteomic data, based on identification of proteins with similar structures that do not necessarily have sequence homology. Such clustering emphasizes protein functions which have multiple copies that are up/down-regulated, and highlights similar proteins which are differentially regulated. As a working example we present proteomic data enumerating differentially expressed proteins in xylem sap from grapevines that were infected with *X. fastidiosa*. **Results**. Analysis of this data by CHURNER highlighted pathogenesis related PR-1 proteins, reinforcing this as the foremost protein function in xylem sap involved in the grapevine defense response to *X. fastidiosa*. β-1,3-glucanase, which has both anti-microbial and anti-fungal activities, is also up-regulated. Simultaneously, chitinases are found to be both up and down-regulated by CHURNER, and thus the net gain of this protein function loses its significance in the defense response. **Discussion**. We demonstrate how structural data can be incorporated in the pipeline of proteomic data analysis prior to making inferences on the importance of individual proteins to plant defense mechanisms. We expect CHURNER to be applicable to any proteomic data set.

1   **Sequence/structural analysis of xylem proteome emphasizes pathogenesis-related proteins,**

2   **chitinases and β-1,3-glucanases as key players in grapevine defense against *Xylella***

3   ***fastidiosa***

5   Sandeep Chakraborty[1,*], Rafael Nascimento[1,2,*], Paulo A. Zaini[2,*], Hossein Gouran[1], Basuthkar J.

6   Rao[3], Luiz R. Goulart[2], Abhaya M. Dandekar[1]

8   1. Department of Plant Sciences, University of California, Davis, USA.

9   2. Institute of Genetics and Biochemistry, Federal University of Uberlândia, Campus Umuarama,

10 Uberlândia, Brazil.

11 3. Department of Biological Sciences, Tata Institute of Fundamental Research, Mumbai, India.

13 * These authors contributed equally to this work.

15 Corresponding author:

16 Abhaya M. Dandekar

17 Dept. of Plant Sciences - University of California

18 One Shields Avenue, Davis, CA 95616, USA

19 E-mail: amdandekar@ucdavis.edu

22 **Abstract**

23 **Background.** *Xylella fastidiosa*, the causative agent of various plant diseases including Pierce's

24 disease in the US, and Citrus Variegated Chlorosis in Brazil, remains a continual source of

25 concern and economic losses, especially since almost all commercial varieties are sensitive to

26 this Gammaproteobacteria. Differential expression of proteins in infected tissue is an established

27 methodology to identify key elements involved in plant defense pathways.

28 **Methods.** In the current work, we developed a methodology named CHURNER that emphasizes

29 relevant protein functions from proteomic data, based on identification of proteins with similar

30 structures that do not necessarily have sequence homology. Such clustering emphasizes protein

31 functions which have multiple copies that are up/down-regulated, and highlights similar proteins

32 which are differentially regulated. As a working example we present proteomic data enumerating

33 differentially expressed proteins in xylem sap from grapevines that were infected with *X.*

34 *fastidiosa*.

35 **Results.** Analysis of this data by CHURNER highlighted pathogenesis related PR-1 proteins,

36 reinforcing this as the foremost protein function in xylem sap involved in the grapevine defense

37 response to *X. fastidiosa*. β-1,3-glucanase, which has both anti-microbial and anti-fungal

38 activities, is also up-regulated. Simultaneously, chitinases are found to be both up and down-

39 regulated by CHURNER, and thus the net gain of this protein function loses its significance in

40 the defense response.

41 **Discussion.** We demonstrate how structural data can be incorporated in the pipeline of proteomic

42 data analysis prior to making inferences on the importance of individual proteins to plant defense

43 mechanisms. We expect CHURNER to be applicable to any proteomic data set.

44

Introduction

Xylella fastidiosa (*X. fastidiosa*) is a xylem-limited pathogen associated with diseases in many economically important plants, including Pierce's Disease of grape (PD) and Citrus Variegated Chlorosis (CVC) (Chatterjee et al. 2008). *X. fastidiosa* lives within the host's water-conducting xylem vessels, where it forms biofilms believed to be responsible for reduced hydraulic conductance caused by clogging of the vessels, and not increased cavitation and embolism of xylem elements (McElrone et al. 2003).

The xylem is composed mainly of lignified vessels that are used for the transportation of water, mineral nutrients and metabolites throughout the vascular system, and in long-distance signaling in response to biotic and abiotic stresses (de Bernonville et al. 2014). Xylem sap contains small molecular weight inorganic compounds, organic substances (Metzner et al. 2010), amino acids and proteins (Biles & Abeles 1991). Recent improvements in genomic and proteomic technologies are accelerating the characterization of these proteins. The xylem sap proteome has been characterized in different plants, which has been shown to contain several protein families such as metabolic enzymes, stress-related proteins and signal transduction proteins (Buhtz et al. 2004; Dafoe & Constabel 2009; Djordjevic et al. 2007; Kehr et al. 2005; Ligat et al. 2011; Rep et al. 2003; Zhang et al. 2015b). These include glycoside hydrolases, peroxidases, chitinases, lipid transfer proteins, proteases, lectins, pathogenesis-related proteins and cell wall structural proteins. The differential accumulation of proteins in xylem sap and apoplast fluid following pathogen infection has been investigated in some pathosystems, clearly indicating that protein composition changes during plant-pathogen interactions, both by the response of the host and by secreted effectors from the pathogen (Floerl et al. 2008; Gawehns et al. 2015; Houterman et al. 2007; Pu et al. 2016; Rep et al. 2002; Subramanian et al. 2009).

Vitis vinifera cv. Chardonnay xylem sap protein composition was previously analyzed by two-dimensional gel electrophoresis, which identified only ten proteins (Agüero et al. 2008). While the role played by xylem proteins in defense against biotic stress has been established in other plant species, the only information available about grapevine xylem sap proteins and their importance to plant response during *X. fastidiosa* pathogenesis came from the pioneering work and Yang and collaborators (2011) and a recent contribution by Katam et al. (2015). While the former showed that thaumatin-like and heat-shock proteins were significantly overexpressed in PD-resistant varieties of grape (Yang, et al. 2011), the latter found several uniquely expressed

76   proteins (β-1, 3-glucanases, 10-deacetyl baccatin III-10-O-acetyl transferase-like, COP9, and

77   aspartyl protease nepenthesin precursor proteins) in PD-tolerant muscadine grape (Katam et al.,

78   2015).The phenolic compounds altered during this plant-pathogen interaction has also been

79   investigated (Wallis & Chen 2012), expanding our understanding of the host molecular response

80   to *X. fastidiosa* infection. Moreover, the comparison of the xylem sap proteome of PD-tolerant

81   and PD-susceptible grapevine species revealed the presence of few proteins that might be

82   directly involved with plant defense against *X. fastidiosa* (Basha et al. 2010). These studies

83   however rely on protein sequence-based approaches for peptide mapping and identification

84   (Altschul et al. 1997; Fenyo & Beavis 2003) , which limits exploring the wealth of information

85   generated in proteomic analysis. Proteins with no sequence homology often possess similar

86   enzymatic capabilities due to convergent evolution (Gherardini et al. 2007) and promiscuity

87   (Chakraborty & Rao 2012; Copley 2003; Jensen 1976); two well-studied phenomena analyzed

88   by considering structural features. As structural data analysis can focus on several properties of

89   target proteins rather than the one-dimensional alignments inherent to sequence-based methods, a

90   structure-based data analysis approach is not well established for proteomics. We present a

91   simple method for classifying protein sets using metrics derived from protein fold which can

92   suggest putative functions to uncharacterized proteins by structural similarity. Our pipeline also

93   performs a more localized perspective and analyzes specific active site residues to determine

94   functional equivalence (Chakraborty et al. 2011; Kleywegt 1999). This approach was applied

95   here to better understand the molecular basis of the interaction between this xylem-colonizing

96   bacterium and grapevines, on data generated by comparing the composition of the xylem sap

97   proteome of infected plants with that of healthy plants. Our analysis pipeline (CHURNER) was

98   able to confirm previous studies cited above and identify novel proteins not previously detected

99   or yet uncharacterized, and is freely available to be used with other proteomic data sets.

100

101   **Materials & methods**

102   **Xylem sap collection and protein precipitation**

103   Xylem sap was collected from six 3-year-old grapevines (*Vitis vinifera* cv. 'Thompson Seedless')

104   located at the University of California Davis (Armstrong field). Three of these plants were

105   mechanically inoculated with *Xylella fastidiosa* Temecula1 12 months prior to sap collection.

106   The presence of *X. fastidiosa* in the xylem sap of infected plants was confirmed using anti- *X.*

107   *fastidiosa* antibodies in a Double Antibody Sandwich ELISA (Agdia, USA) following

108   manufacturer's instructions (Supplemental figure S1). Xylem sap (30-50 mL per plant) was

109   collected overnight in the second week of spring by drip of the cut stem of *X. fastidiosa*-infected

110   and non-infected plants. To initiate sap collection, an apical segment of approximately 10 cm

111   was cut from the stem and the vine terminal introduced into a collection tube sealed with

112   parafilm. Xylem sap was lyophilized followed by protein precipitation using TCA/Acetone

113   (Gorg et al. 2000). The pellets were resuspended in 300 µL of PBS (pH 7.4) and total protein

114   was quantified using BCA Protein Assay Kit (Thermo Fisher Scientific) following

115   manufacturer's instructions for subsequent SDS-PAGE and LC-MS/MS analysis.

116

117   **Protein preparation, mass spectrometry analysis and NMR imaging**

118   Proteins were precipitated using ProteoExtract^{TM} Protein Precipitation kit (Calbiochem) followed

119   by dehydration overnight in a sterile fume hood. The protein pellet was then resuspended in 50

120   mM AmBic (pH 8.0) and 100 µg subjected to an in-solution tryptic digestion. The digested

121   peptides were analyzed using a QExactive mass spectrometer (Thermo Fisher Scientific) coupled

122   with an Easy-LC (Thermo Fisher Scientific) and a nanospray ionization source. One microgram

123   of digested peptides were loaded onto a trap (100 micron, C18 100˚A 5U) and desalted online

124   before separation using a reverse phased column (75 micron, C18 200˚A 3U). The gradient

125   duration for separation of peptides was 60 minutes using 0.1% formic acid and 100% acetonitrile

126   as solvents A and B, respectively. Raw data was analyzed using X!Tandem (Fenyo & Beavis

127   2003) and visualized using Scaffold version 4.4.1 (Proteome Software, OR). Samples were

128   searched against UniProt databases appended with the cRAP database, which recognizes

129   common laboratory contaminants. Reverse decoy databases were also applied to the database

130   prior to the X!Tandem searches. Peptide identifications were accepted if they could be

131   established at greater than 95% probability by the Peptide Prophet algorithm (Keller et al. 2002;

132   Nesvizhskii et al. 2003) with Scaffold delta-mass correction. Protein identifications were

133   accepted if they could be established at greater than 99% probability and contained at least 2

134   identified peptides (see supplemental file S1 for raw data of identified proteins. All supplemental

135   material is available at

136   https://drive.google.com/drive/folders/0BxTKjdWklOUbdmlpdXZQdG00bWc). Proteins that

137   contained similar peptides and could not be differentiated based on MS/MS analysis alone were

138   grouped to satisfy the principles of parsimony. For relative protein quantification of xylem sap

139   from infected and non-infected plants, the QSpec statistical framework (Version

140   2,http://sourceforge.net/pro jects/qspec/) was used to assign significance to differentially

141   regulated proteins, using a Bayes factor > 10 (Choi et al. 2008).

142   Nuclear magnetic resonance imaging ([1]H-MRI) was done in an Avance 400 spectrometer

143   equipped with Bruker DRX console microimaging accessory according to (Dandekar et al.

144   2012). Stem transverse sections of all non-infected and infected plants were collected between

145   internodes located at the top (apical), middle and bottom of the central stem (three cuts per plant)

146   and subjected to MRI. Figure 2 shows representative results.

147

148   **Downstream *In silico* methods**

149   We have written custom programs to automate the extraction of protein sequences, their

150   annotation through the BLAST command line (Camacho 2008), obtaining homologous PDB

151   structures, and getting pairwise structural homology (Konc & Janezic 2010) from proteome data

152   mined with Prophet/Scaffold program (Keller et al. 2002) (see example dataset in supplemental

153   file S1). These programs were integrated in the CHURNER pipeline using freely available

154   BioPerl (Stajich et al. 2002) modules and Emboss (Rice et al. 2000) tools (additional

155   documentation and scripts available as supplemental files S2 and S3). As manual steps, we

156   performed gene ontology of the differentially expressed proteins using the statistical

157   overrepresentation test from the PANTHER protein classification system (Mi et al. 2013), and

158   FATCAT (Ye & Godzik 2004) to supervise and validate the structural homology. All protein

159   structures were rendered by the PyMol Molecular Graphics System, version 1.7.4 Schrödinger,

160   LLC (http://www.pymol.org/). Detection of putative signal peptides and target sub-cellular

161   locations of proteins were done with SignalP 3.0 (Bendtsen et al. 2004) and TargetP 1.1

162   (Emanuelsson et al. 2007), respectively. Protein sequences used in sequence and structural

163   alignments were devoid of signal sequences to better represent mature proteins. Congruence of

164   specific active site residues to determine functional equivalence between proteins was performed

165   with CLASP (Chakraborty et al. 2011). Adaptive Poisson-Boltzmann Solver (APBS) and

166   PDB2PQR packages were used to calculate the electrostatic potentials of all the atoms in the

167   protein (Baker et al. 2001; Dolinsky et al. 2004). The APBS parameters were set as described

168   previously in (Chakraborty et al. 2011). APBS writes out the electrostatic potential in

169 dimensionless units of kT/e where k is Boltzmann's constant, T is the temperature in K and e is

170 the charge of an electron, used to calculate the pairwise potential differences.

171

Results & Discussion

The CHURNER workflow for proteome analysis

174 CHURNER implements tandem analysis of sequence and structure of proteins to highlight

175 potential functional similarities not obvious from simple sequence alignments. It uses simple Perl

176 scripts to obtain the pairwise sequence and structural homology scores from BLAST and ProBiS,

177 respectively (Figure 1). The structural homology is then checked with FATCAT by the FCTSIG

178 significance test (Ye & Godzik 2004). We used these well established methods with different

179 algorithms to detect structural similarity. This helps us corroborate the results between them (PZ

180 scores). FATCAT treats the protein as flexible, allowing twists in the reference protein (akin to a

181 real protein) and minimizes the number of rigid-body movements for the best structural

182 alignment. ProBiS relies on common surface structural patches rather than global conservation in

183 finding structural similarity, with the reasoning that the surface residues are more critical since

184 they determine ligand or protein-protein interactions. The source code, working directory and a

185 README script for the current example is available as supplemental material (see Methods). As

186 a working example we used proteins identified by LC-MS/MS from xylem sap of grapevines

187 infected with *X. fastidiosa* as described below. By using CHURNER we were able to highlight

188 the Pathogenesis-Related proteins (PR-1) as the main proteins of grapevine defense against *X.*

189 *fastidiosa* to be secreted in xylem sap and also to unravel structural similarity between β-D-

190 glucan exo-hydrolase and chitinases, which has no known reference in existing literature. To the

191 best of our knowledge, this is the first attempt to analyze proteomic data for differentially

192 expressed proteins based on structural features. A walkthrough of the steps implemented by

193 CHURNER will be demonstrated next.

194

Our working example: xylem sap proteins identification by LC-MS/MS

196 In this study, proteins from xylem sap of *X. fastidiosa*-infected and non-infected grapevines were

197 lyophilized, precipitated by TCA/acetone and analyzed by SDS-PAGE (Supplemental Figure

198 S2). A total of 91 proteins (Table 1) were identified herein by LC-MS/MS with at least two

199 peptides sequenced per protein. SignalP and TargetP were used to predict the presence of signal

200 peptides and sub-cellular localization, respectively, in all protein sequences. Signal peptides were

201 found in 70 proteins (77%) of which 67 were predicted to be secreted while the others are

202 directed towards an undetermined sub-cellular location.

203

Differentially expressed proteins in xylem sap of infected plants

205 Previous reports investigating differentially expressed transcripts and proteins in xylem sap of *X.*

206 *fastidiosa*-infected plants have provided a wealth of information regarding the plant responses to

207 infection, both in grapevines (Katam et al. 2015; Lin et al. 2007; Yang et al. 2011) and citrus

208 (Rodrigues et al. 2013). These include PR proteins, β-1,3-glucanases, LRR-RLKs and chitinases

209 listed in Table 1. From all proteins detected in xylem sap, those with significant fold changes as

210 indicated by the Prophet/Scaffold program (See supplemental file S1 with protein quantification

211 data and file S4 for multifasta protein sequences) were re-annotated using the BLAST command

212 line interface (Camacho 2008). Twelve of those for which a predicted structure could be

213 assigned in the Protein Data Bank were further analyzed in the CHURNER pipeline (Table 2).

214 Despite some UniProt IDs not having the proper functional annotation, as for example protein

215 UID: F6HBN7, which is not annotated as a pathogenesis-related protein in

216 http://www.uniprot.org/uniprot/F6HBN7, CHURNER uses UniProt IDs as protein identifiers, as

217 it is easily streamlined with other applications such as the PANTHER gene ontology analysis

218 server. The PANTHER overrepresentation test uses the most updated gene ontology database

219 released, and has the option to use Bonferroni correction for multiple testing, which did not alter

220 our results. It also has a wide range of reference genomes to be queried with the user list. Nine of

221 twelve of our differentially expressed proteins could be mapped to ontology terms by

222 PANTHER. The molecular functions with significant overrepresentation were chitinase and

223 glycosyl hydrolase activities (GO:0004568, and GO:0016798, respectively. See complete GO

224 analysis in supplemental file S5). Accordingly, the biological processes overrepresented were

225 catabolism of glucosamines (GO:1901072), chitin (GO:0006032), amino glycans (GO:0006026)

226 and amino sugars (GO:0046348). Interestingly, identification of differentially expressed proteins

227 demonstrated that NtPRp27, which is found in the xylem exudate of non-infected grapevines

228 (Agüero et al. 2008), and is a known pathogenesis-related protein (Okushima et al. 2000) was not

229 over-expressed upon *X. fastidiosa* infection. Comparably, two LRR-RLK (Ciclev10004108m and

230 Ciclev10014130m) found to be up-regulated in Ponkan mandarin infected with *X. fastidiosa*

231 (Rodrigues et al. 2013), were not affected in our data, despite LRR-RLKs being detected in our

232 proteomic analysis (Table 1). Plant receptor-like kinases (RLKs) are a large gene family (~600

233 members in Arabidopsis) (Shiu & Bleecker 2001), consisting of an accelerated evolutionary

234 domain implicated in signal reception through leucine-rich repeat (LRRs) (Afzal et al. 2008).

235 Resistance (R) genes have evolved to counter pathogens that bypass the pathogen-associated

236 molecular patterns (PAMP) mechanism in plants (Nicaise et al. 2009). Most R genes encode

237 proteins comprising of a nucleotide-binding site (NBS) and leucine-rich repeats (LRRs), and

238 recognize and neutralize specialized pathogen avirulence (Avr) proteins, providing plants with

239 resistance (Borhan et al. 2004; Chakraborty et al. 2016; Ernst et al. 2002; Hayashi et al. 2010;

240 Zhang et al. 2010).

241

242 **Up-regulated proteins**

243 The up-regulated protein F6HLL8 listed in Table 2 is a β-1,3-glucanase (GNS), a well-

244 established pathogenesis related protein (Balasubramanian et al. 2012; Shinshi et al. 1988). GNS

245 has strong anti-microbial (Xie et al. 2015) and anti-fungal activity (Su et al. 2013). Expectedly, it

246 is a target of pathogen toxins in the ensuing evolutionary battle (Sanchez-Rangel et al. 2012;

247 Zhang et al. 2015b). GNS, along with chitinase, has been shown to inhibit fungal growth (Mauch

248 et al. 1988; Sela-Buurlage et al. 1993). The presence of two up-regulated chitinases (UIDs:

249 A5BK69, D7T548) induced by *X. fastidiosa* demonstrates that this defense response is similar

250 during both bacterial and fungal attack. Both these chitinases belong to the GH18 sub-family of

251 chitinases (Funkhouser & Aronson 2007). Another up-regulated protein is an expansin-like B1

252 enzyme (UID: A5C594), also a carbohydrate-binding protein involved in cell-wall loosening and

253 restructuring (Zhang et al. 2014). Homologs have previously been link to abiotic stress responses

254 (Han et al. 2014; Nanjo et al. 2013). Two other proteins (UID: F6HBN7, A5BNW5) are the well-

255 established defense PR-1 pathogenesis related proteins (Sudisha et al. 2012). Several functions

256 have previously been attributed to PR1 including, antiviral activity in tobacco (Antoniw & White

257 1980) and anti-herbivory activity in maize (Zhang et al. 2015a). Additional activities suggest

258 protease-mediated programmed cell death pathways in plants (Lu et al. 2013), a symptom

259 commonly seen in leaves of Pierce's diseased grapevines. A lipid-transfer protein of the DIR1

260 type (UID: D7TML8) also ranks among the up-regulated proteins, suggesting a systemic defense

261 response being activated, as previously seen in *Arabidopsis* (Champigny et al. 2013). A blue

262 copper protein-like was also identified as up-regulated. This redox protein family is part of a
263 widespread but yet poorly characterized defense mechanism in plants and/or lignin formation
264 (Cao et al. 2015; Nersissian et al. 1998). It is interesting to note an abundance of cell-wall
265 modification enzymes in xylem sap from infected plants, as this corroborates previous
266 observations in other pathosystems such as *Xanthomonas oryzae* infection of rice (Hilaire et al.
267 2001), among others (van Loon et al. 2006). Since vines infected with *X. fastidiosa* commonly
268 display an increase in stem diameter, we verified if this reflected in an increase of secondary wall
269 deposition by nuclear magnetic resonance imaging (Figure 2). Indeed the MRI showed an
270 increase in dense material (colored in light shades in Figure 2) in infected plant stems, including
271 those used for xylem sap collection. Recently it has also been demonstrated that citrus infected
272 with *X. fastidiosa* display a thickening of secondary cell-walls (Niza et al. 2015).
273

274 **Down-regulated proteins**
275 As mentioned previously the GH19 sub-family chitinase (UID: F6GZC4) is down-regulated. A
276 similar suppression of a chitinase gene in response to mycorrhizal fungus *Glomus intraradices*
277 infection of tobacco roots has been noted previously (David et al. 1998). The GH19 sub-family
278 has members that are sugar-binding proteins but without catalytic activity (Martinez-Caballero et
279 al. 2014). Further experimentation is necessary to verify whether this is the case in *Vitis vinifera*,
280 as it can be part of the cell-wall remodeling in response to the pathogen, as previously suggested
281 by other works (Lin et al. 2007; Rodrigues et al. 2013). A thaumatin-like protein (TLP) (UID:
282 A5AWT7) is also found to be down-regulated upon *X. fastidiosa* infection. TLP's are found in
283 most eukaryotes, and involved in host defense and several developmental processes (Liu et al.
284 2010). While TLP over expression has been shown to enhance resistance to *Alternaria alternata*
285 in tobacco (Safavi et al. 2012), these proteins can also be down-regulated in some cases, as
286 shown in the compilation of transcriptomes of poplar leaf rust infections (Petre et al. 2011). A β-
287 D-xilosidase 4 ortholog (UID: F6I6R4) was also down-regulated in infected vines, again
288 reinforcing the drastic effect on cell-wall remodeling enzymes upon pathogen infection. The
289 other down-regulated protein in Table 2 is a LTP similar to YLS3 (yellow-leaf-specific) which is
290 a marker of leaf senescence, being induced in earlier stages and repressed in later stages in
291 *Arabidopsis* (Yoshida et al. 2001).
292

293   **Sequence homology**

294   Although in our data set only twelve proteins were selected for further analysis through

295   CHURNER, there might be cases when larger data sets render a manual inspection of

296   homologous proteins difficult. Even in the current case, it is difficult to identify D7SLG6 as a

297   lipid transfer protein (LTP) from the BLAST automated annotation (Table 2, value marked in

298   italic). Thus, as the next step in CHURNER, we implemented a pairwise BLAST of all mature

299   proteins (devoid of signal sequences). Table 3 shows the pairwise sequence homology with an E-

300   value cutoff of 0.005. There are several interesting aspects that emerge from this comparison. As

301   expected, the two PR-1 proteins are found to be significantly homologous. The 'YLS3-like'

302   protein (UID: D7SLG6) is found to be quite similar to another LTP (UID: D7TML8).

303   Furthermore, we observe sequence homology between chitinases and expansin (E-value = 4e-

304   04), and much less between chitinases and PR-1 proteins (E-value = 0.002). Interestingly, these

305   similarities are greater than that between the two known chitinases from sub-family GH18 (E-

306   value = 0.003). This raises the interesting question whether these proteins

307   (chitinases/expansin/PR-1) have promiscuous functions (Chakraborty & Rao 2012; Khersonsky

308   & Tawfik 2010), and underlines the problem of depending only on annotation of individual

309   protein sequences, thus providing a more rational alternative to identify proteins with similar

310   functions. Since the structure of a protein is intrinsically related to its function, we implemented

311   structural annotation as the next step in CHURNER.

312

313   **Structural annotation**

314   CHURNER implements an automated search for homologous proteins with known PDB

315   structures (Table 2). As expected, proteins with high similarity and alignment at the sequence

316   level map to the same PDB structure, such as the PR-1 proteins (PDBid: 1CFE, PR-14a protein).

317   Interestingly, this protein was identified as a possible replacement of the human neutrophil

318   elastase component of the chimeric protein (Chakraborty 2012; Chakraborty et al. 2013) that

319   provided enhanced grapevine resistance to *X. fastidiosa* (Dandekar et al. 2012). Moreover, apart

320   from the LTP (UID: D7SLG6, E-value = 0.017), all matches are very significant to their PDB

321   closest model.

322

323   **Structural homology**

324 Using the structures corresponding to the identified proteins, we detected similarities that might

325 have escaped detection in the sequence homology search. Table 4 shows the most significant

326 pairwise structural comparison of the structures (excluding the PR-1 proteins which have the

327 same PDB structure) computed using ProBiS (Konc & Janezic 2010). We subsequently verified

328 the alignment significance using the FATCAT server (Ye & Godzik 2004). The ProBiS Z-score

329 (PZ in Table 4) are standardized alignment scores (Konc & Janezic 2010) which provide

330 statistical and structural significance of local structural alignments. Z-scores >2 are considered

331 highly significant (PDBs: 2RKN/1FK0 and 1HVQ/3AQU), although a Z-score of 1.6 is also

332 significant (PDBs: 3AQU/1EX1), as confirmed by FATCAT (which looks at the global

333 structure). The LTPs (with a sequence homology E-value = 1e-06) are structurally homologous,

334 as expected (Figure 3A). Noteworthy, the chitinases from the GH18 family with a low sequence

335 homology (E-value = 0.003) are structurally homologous (Figure 3B). Finally, in spite of a much

336 lower sequence homology (E-value = 0.44), the chitinase (UID: D7T548) and the β-D-glucan

337 exohydrolase (UID: F6I6R4) are structurally homologous (Figure 3C). These observations

338 highlight the necessity of structural comparison in annotating and grouping proteins based on

339 functionality in proteomic analysis, and points to alternative protein functions that can be tested

340 in subsequent studies.

341 Environmental stimuli (Yang et al. 2012), pathogens (Moy et al. 2004) or disease (Ng et al.

342 2009) induce differential expression of specific genes. Rapid technological advances have helped

343 us identify these genes, and define their roles in defense or pathogenesis. While quantifying

344 transcripts through high-throughput sequencing techniques have revolutionized these efforts

345 (Wang et al. 2009), the correlation between transcriptional and protein abundance remains

346 suspect (Gygi et al. 1999) due to the complexity of the regulatory factors modulating translation

347 (Zhu et al. 2012). Thus, identifying proteins through techniques like mass spectrometry (Witzel

348 et al. 2009), and measuring their relative amounts (Hu et al. 2015), provides a true picture of the

349 genes involved in pathogenesis and defense response, rather than measuring their RNA

350 abundance (Moy et al. 2004). Nevertheless, the value of transcriptomic analysis should not be

351 underestimated as it has proved to be an effective approach to discover genes responsive to

352 infection, as exemplified by sequencing of expressed sequence tags from *X. fastidiosa*-infected

353 grapevines (Lin et al. 2007) and citrus (Rodrigues et al. 2013).

354   Here we showed that several proteins (pathogenesis-related PR-1, chitinases and β-1,3-

355   glucanases) are differentially expressed in the xylem sap of grapevine infected with *X. fastidiosa*,

356   using LC-MS/MS with at least two peptides sequenced per protein, confirming findings of

357   previous investigations. These proteins are recognized as key players in the plant response

358   against pathogen infection, as exemplified by sugarcane infected by *Sporisorium scitamineum*

359   (Su et al. 2013) and other pathosystems reviewed in (Sudisha et al. 2012). Interestingly, we

360   observed that there are two chitinases which are regulated differently (UniProt IDs F6GZC4 and

361   D7T548, Table 4). Chitinases degrade chitin, a component of fungal cell walls (David et al.

362   1998). This observation diminishes the importance of chitinase as a generic defense agent against

363   *X. fastidiosa*. It is also possible that the plant response to the infecting bacterial agent is

364   indifferent to the expression levels of these anti-fungal chitinases, for which we warrant further

365   studies. The pathogenic state can also be characterized by differentially expressed genes within

366   *X. fastidiosa* itself (Shi et al. 2010). Consequently, the analysis of differentially expressed genes

367   and proteins should consider functionally related proteins as a single entity, and not just the

368   expression levels of single genes or proteins. CHURNER allows for such selection based on both

369   sequence and structural homology, as often, the best match for a sequence has an incomplete

370   annotation. Subsequently, structural homology can identify functional relationship among

371   sequences with little sequence homology. For example, the two chitinase homologs have low

372   pairwise sequence homology (E-value = 0.003). However, their significant homologous

373   counterparts in the PDB database are PDBid: 4DWXA (*Secale cereale*, rye) and PDBid: 3AQU

374   (*Arabidopsis thaliana*) have significant structural homology, as computed using ProBiS (Konc &

375   Janezic 2010) and FATCAT (Ye & Godzik 2004).

376

377   **Corroborating the promiscuity of the chitinase and β-D-glucan exohydrolase by active site**

378   **structural homology**

379   The β-D-glucan exohydrolase (PDBid: 1EX1) is critical for hydrolysis of cell walls, containing

380   high levels of 1,3-β-D-glucans, during wall degradation in germinated grain and during wall

381   loosening in elongating coleoptiles (Varghese et al. 1999). Interestingly, the fungal wall is

382   composed of chitin, 1,3-β- and 1,6-β-glucan, mannan and proteins (Adams 2004). Thus, the up-

383   regulation of both chitinases and β-D-glucan exohydrolases is possibly an anti-fungal defense

384   response that has been triggered by *X. fastidiosa*. We have seen that the chitinase and the β-D-

385   glucan exohydrolase have no sequence homology (Table 3), but partial structural homology

386   (Table 4). A detailed analysis of their catalytic residues further strengthens credence of their

387   functional similarity. For β-D-glucan exohydrolase, Asp285 and Glu491 are involved in catalysis

388   (Varghese et al. 1999). In chitinases, the Asp114 and Glu116 are a part of the conserved motif

389   (DXXDXDXE) (Hamid et al. 2013). The active site residues of these proteins demonstrate

390   significant spatial (Figure 4) and electrostatic congruence (Table 5) determined using CLASP.

391   The absence of sequence linearity indicates that this homology arose from convergent evolution.

392   Remarkably, based on the catalytic triad of the barley β-D-glucan exohydrolase (PDBid:

393   1EX1A) provided to CLASP, it was able to pick up the catalytic triad from the Arabidopsis

394   chitinase (PDBid: 3AQU) despite the lack of sequence homology. We later realized this

395   prediction was correct consulting the work from Ohnuma and collaborators on the

396   crystallographic studies of this type V chitinase (2011).

397

398   **Conclusions**

399   A robust proteomic methodology was used to identify novel protein fragments in addition to the

400   proteins identified previously using two-dimensional gel electrophoresis followed by sequencing

401   of protein spots in non-infected xylem exudates collected from *Vitis vinifera* cv. Chardonnay

402   vines (Agüero et al. 2008), and mass spectrometric analysis of *X. fastidiosa*-infected grapevines

403   (Katam et al. 2015; Yang et al. 2011). Thus, here we link the expression pattern of well-studied

404   proteins in grapevines to the pathogen perception response, and present a methodology for

405   assessing their significance by taking into account both their sequence and structural

406   information. Although our data came from a single round of infection containing three plants

407   (and three non-infected control plants), our findings are consistent among the different samples

408   and to previous proteomic studies of *X. fastidiosa* infected grapevines. In conclusion,

409   CHURNER enhances our ability to find functionally-relevant protein candidates that have little

410   or no sequence similarity, and thus would be considered as separate components of a data set.

411   The name "CHURNER" was inspired in the mixing tool used to reach the "cream". We intend to

412   offer a tool to enable for detection of "cream" protein functions, not obvious from simple amino

413   acid sequence alignments. The reduced data set employed in this work was used as a proof of

414   concept, and we encourage readers to use complex data sets with thousands of proteins to find

415   many putative functional relations among proteins that are yet unexplored.

416

Acknowledgements

We thank James Mitch Elmore for helpful suggestions in proteomic analysis.

419

References

421

Adams DJ. 2004. Fungal cell wall chitinases and glucanases. *Microbiology* 150:2029-2035. 10.1099/mic.0.26980-0 150/7/2029 [pii]

Afzal AJ, Wood AJ, and Lightfoot DA. 2008. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Mol Plant Microbe Interact* 21:507-517. 10.1094/MPMI-21-5-0507

Agüero CB, Thorne ET, Ibáñez AM, Gubler WD, and Dandekar AM. 2008. Xylem Sap Proteins from Vitis vinifera L. Chardonnay. *American Journal of Enology and Viticulture* 59:306-311.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402. gka562 [pii]

Antoniw JF, and White RF. 1980. The Effects of Aspirin and Polyacrylic Acid on Soluble Leaf Proteins and Resistance to Virus Infection in Five Cultivars of Tobacco. *Journal of Phytopathology* 98:331-341.

Baker NA, Sept D, Joseph S, Holst MJ, and McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98:10037-10041. 10.1073/pnas.181342398 181342398 [pii]

Balasubramanian V, Vashisht D, Cletus J, and Sakthivel N. 2012. Plant beta-1,3-glucanases: their biological functions and transgenic expression against phytopathogenic fungi. *Biotechnol Lett* 34:1983-1990. 10.1007/s10529-012-1012-6

Basha SM, Mazhar H, and Vasanthaiah HK. 2010. Proteomics approach to identify unique xylem sap proteins in Pierce's disease-tolerant Vitis species. *Appl Biochem Biotechnol* 160:932-944. 10.1007/s12010-009-8620-1

Bendtsen JD, Nielsen H, von Heijne G, and Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783-795. 10.1016/j.jmb.2004.05.028 S0022283604005972 [pii]

Biles CL, and Abeles FB. 1991. Xylem sap proteins. *Plant Physiol* 96:597-601.

Borhan MH, Holub EB, Beynon JL, Rozwadowski K, and Rimmer SR. 2004. The arabidopsis TIR-NB-LRR gene RAC1 confers resistance to Albugo candida (white rust) and is dependent on EDS1 but not PAD4. *Mol Plant Microbe Interact* 17:711-719. 10.1094/MPMI.2004.17.7.711

Camacho C. 2008. BLAST® Command Line Applications User Manual [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2008-. Available from: . *Available at http://www.ncbi.nlm.nih.gov/books/NBK279690/.*

Cao J, Li X, Lv Y, and Ding L. 2015. Comparative analysis of the phytocyanin gene family in 10 plant species: a focus on Zea mays. *Front Plant Sci* 6:515. 10.3389/fpls.2015.00515

Chakraborty S. 2012. An automated flow for directed evolution based on detection of promiscuous scaffolds using spatial and electrostatic properties of catalytic residues. *PLoS One* 7:e40408. 10.1371/journal.pone.0040408 PONE-D-12-13078 [pii]

Chakraborty S, Britton M, Martinez-Garcia PJ, and Dandekar AM. 2016. Deep RNA-Seq profile reveals biodiversity, plant-microbe interactions and a large family of NBS-LRR resistance genes in walnut (Juglans regia) tissues. *AMB Express* 6:12. 10.1186/s13568-016-0182-3 10.1186/s13568-016-0182-3 [pii]

Chakraborty S, Minda R, Salaye L, Bhattacharjee SK, and Rao BJ. 2011. Active site detection by spatial conformity and electrostatic analysis--unravelling a proteolytic function in shrimp alkaline phosphatase. *PLoS One* 6:e28470. 10.1371/journal.pone.0028470 PONE-D-11-14548 [pii]

Chakraborty S, Minda R, Salaye L, Dandekar AM, Bhattacharjee SK, and Rao BJ. 2013. Promiscuity-based enzyme selection for rational directed evolution experiments. *Methods Mol Biol* 978:205-216. 10.1007/978-1-62703-293-3_15

468   Chakraborty S, and Rao BJ. 2012. A measure of the promiscuity of proteins and characteristics of residues in the
469   vicinity of the catalytic site that regulate promiscuity. *PLoS One* 7:e32011. 10.1371/journal.pone.0032011
470   PONE-D-11-20732 [pii]
471   Champigny MJ, Isaacs M, Carella P, Faubert J, Fobert PR, and Cameron RK. 2013. Long distance movement of
472   DIR1 and investigation of the role of DIR1-like during systemic acquired resistance in Arabidopsis. *Front*
473   *Plant Sci* 4:230. 10.3389/fpls.2013.00230
474   Chatterjee S, Almeida RP, and Lindow S. 2008. Living in two worlds: the plant and insect lifestyles of Xylella
475   fastidiosa. *Annu Rev Phytopathol* 46:243-271. 10.1146/annurev.phyto.45.062806.094342
476   Choi H, Fermin D, and Nesvizhskii AI. 2008. Significance analysis of spectral count data in label-free shotgun
477   proteomics. *Mol Cell Proteomics* 7:2373-2385. 10.1074/mcp.M800203-MCP200
478   Copley SD. 2003. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr Opin Chem*
479   *Biol* 7:265-272. S1367593103000322 [pii]
480   Dandekar AM, Gouran H, Ibanez AM, Uratsu SL, Aguero CB, McFarland S, Borhani Y, Feldstein PA, Bruening G,
481   Nascimento R, Goulart LR, Pardington PE, Chaudhary A, Norvell M, Civerolo E, and Gupta G. 2012. An
482   engineered innate immune defense protects grapevines from Pierce disease. *Proc Natl Acad Sci U S A*
483   109:3721-3725. 10.1073/pnas.1116027109
484   David R, Itzhaki H, Ginzberg I, Gafni Y, Galili G, and Kapulnik Y. 1998. Suppression of tobacco basic chitinase
485   gene expression in response to colonization by the arbuscular mycorrhizal fungus Glomus intraradices. *Mol*
486   *Plant Microbe Interact* 11:489-497. 10.1094/MPMI.1998.11.6.489
487   de Bernonville TD, Albenne C, Arlat M, Hoffmann L, Lauber L, and Jamet E. 2014. Xylem sap proteomics.
488   *Methods Mol Biol* 1072:391-405.
489   Dolinsky TJ, Nielsen JE, McCammon JA, and Baker NA. 2004. PDB2PQR: an automated pipeline for the setup of
490   Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32:W665-667. 10.1093/nar/gkh381
491   32/suppl_2/W665 [pii]
492   Emanuelsson O, Brunak S, von Heijne G, and Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP
493   and related tools. *Nat Protoc* 2:953-971. nprot.2007.131 [pii]
494   10.1038/nprot.2007.131
495   Ernst K, Kumar A, Kriseleit D, Kloos DU, Phillips MS, and Ganal MW. 2002. The broad-spectrum potato cyst
496   nematode resistance gene (Hero) from tomato is the only member of a large gene family of NBS-LRR
497   genes with an unusual amino acid repeat in the LRR region. *Plant J* 31:127-136. 1341 [pii]
498   Fenyo D, and Beavis RC. 2003. A method for assessing the statistical significance of mass spectrometry-based
499   protein identifications using general scoring schemes. *Anal Chem* 75:768-774.
500   Floerl S, Druebert C, Majcherczyk A, Karlovsky P, Kues U, and Polle A. 2008. Defence reactions in the apoplastic
501   proteome of oilseed rape (Brassica napus var. napus) attenuate Verticillium longisporum growth but not
502   disease symptoms. *BMC Plant Biol* 8:129. 1471-2229-8-129 [pii]
503   10.1186/1471-2229-8-129
504   Funkhouser JD, and Aronson NN, Jr. 2007. Chitinase family GH18: evolutionary insights from the genomic history
505   of a diverse protein family. *BMC Evol Biol* 7:96. 10.1186/1471-2148-7-96
506   Gawehns F, Ma L, Bruning O, Houterman PM, Boeren S, Cornelissen BJ, Rep M, and Takken FL. 2015. The
507   effector repertoire of Fusarium oxysporum determines the tomato xylem proteome composition following
508   infection. *Front Plant Sci* 6:967. 10.3389/fpls.2015.00967
509   Gherardini PF, Wass MN, Helmer-Citterich M, and Sternberg MJ. 2007. Convergent evolution of enzyme active
510   sites is not a rare phenomenon. *J Mol Biol* 372:817-845. S0022-2836(07)00812-1 [pii]
511   10.1016/j.jmb.2007.06.017
512   Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, and Weiss W. 2000. The current state of two-
513   dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21:1037-1053.
514   10.1002/(SICI)1522-2683(20000401)21:6<1037::AID-ELPS1037>3.0.CO;2-V [pii]
515   10.1002/(SICI)1522-2683(20000401)21:6<1037::AID-ELPS1037>3.0.CO;2-V
516   Gygi SP, Rochon Y, Franza BR, and Aebersold R. 1999. Correlation between protein and mRNA abundance in
517   yeast. *Mol Cell Biol* 19:1720-1730.
518   Hamid R, Khan MA, Ahmad M, Ahmad MM, Abdin MZ, Musarrat J, and Javed S. 2013. Chitinases: An update. *J*
519   *Pharm Bioallied Sci* 5:21-29. 10.4103/0975-7406.106559
520   JPBS-5-21 [pii]
521   Han Y, Chen Y, Yin S, Zhang M, and Wang W. 2014. Over-expression of TaEXPB23, a wheat expansin gene,
522   improves oxidative stress tolerance in transgenic tobacco plants. *J Plant Physiol* 173C:62-71. S0176-
523   1617(14)00264-8 [pii]

524   10.1016/j.jplph.2014.09.007

525   Hayashi N, Inoue H, Kato T, Funao T, Shirota M, Shimizu T, Kanamori H, Yamane H, Hayano-Saito Y, Matsumoto
526   T, Yano M, and Takatsuji H. 2010. Durable panicle blast-resistance gene Pb1 encodes an atypical CC-
527   NBS-LRR protein and was generated by acquiring a promoter through local genome duplication. *Plant J*
528   64:498-510. 10.1111/j.1365-313X.2010.04348.x

529   Hilaire E, Young SA, Willard LH, McGee JD, Sweat T, Chittoor JM, Guikema JA, and Leach JE. 2001. Vascular
530   defense responses in rice: peroxidase accumulation in xylem parenchyma cells and xylem wall thickening.
531   *Mol Plant Microbe Interact* 14:1411-1419. 10.1094/MPMI.2001.14.12.1411

532   Houterman PM, Speijer D, Dekker HL, CG DEK, Cornelissen BJ, and Rep M. 2007. The mixed xylem sap
533   proteome of Fusarium oxysporum-infected tomato plants. *Mol Plant Pathol* 8:215-221. MPP384 [pii]
534   10.1111/j.1364-3703.2007.00384.x

535   Hu J, Rampitsch C, and Bykova NV. 2015. Advances in plant proteomics toward improvement of crop productivity
536   and stress resistancex. *Front Plant Sci* 6:209. 10.3389/fpls.2015.00209

537   Jensen RA. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* 30:409-425.
538   10.1146/annurev.mi.30.100176.002205

539   Katam R, Chibanguza K, Latinwo LM, and Smith D. 2015. Proteome biomarkers in xylem reveal pierce's disease
540   tolerance in grape. *Journal of Proteomics & Bioinformatics* 8:217-224.

541   Keller A, Nesvizhskii AI, Kolker E, and Aebersold R. 2002. Empirical statistical model to estimate the accuracy of
542   peptide identifications made by MS/MS and database search. *Anal Chem* 74:5383-5392.

543   Khersonsky O, and Tawfik DS. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev*
544   *Biochem* 79:471-505. 10.1146/annurev-biochem-030409-143718

545   Kleywegt GJ. 1999. Recognition of spatial motifs in protein structures. *J Mol Biol* 285:1887-1897. S0022-
546   2836(98)92393-2 [pii]
547   10.1006/jmbi.1998.2393

548   Konc J, and Janezic D. 2010. ProBiS algorithm for detection of structurally similar protein binding sites by local
549   structural alignment. *Bioinformatics* 26:1160-1168. 10.1093/bioinformatics/btq100

550   Lin H, Doddapaneni H, Takahashi Y, and Walker MA. 2007. Comparative analysis of ESTs involved in grape
551   responses to Xylella fastidiosa infection. *BMC Plant Biol* 7:8. 1471-2229-7-8 [pii]
552   10.1186/1471-2229-7-8

553   Liu JJ, Sturrock R, and Ekramoddoullah AK. 2010. The superfamily of thaumatin-like proteins: its origin, evolution,
554   and expression towards biological function. *Plant Cell Rep* 29:419-436. 10.1007/s00299-010-0826-8

555   Lu S, Faris JD, Sherwood R, and Edwards MC. 2013. Dimerization and protease resistance: new insight into the
556   function of PR-1. *J Plant Physiol* 170:105-110. S0176-1617(12)00328-8 [pii]
557   10.1016/j.jplph.2012.08.006

558   Martinez-Caballero S, Cano-Sanchez P, Mares-Mejia I, Diaz-Sanchez AG, Macias-Rubalcava ML, Hermoso JA,
559   and Rodriguez-Romero A. 2014. Comparative study of two GH19 chitinase-like proteins from Hevea
560   brasiliensis, one exhibiting a novel carbohydrate-binding domain. *FEBS J* 281:4535-4554.
561   10.1111/febs.12962

562   Mauch F, Mauch-Mani B, and Boller T. 1988. Antifungal Hydrolases in Pea Tissue : II. Inhibition of Fungal Growth
563   by Combinations of Chitinase and beta-1,3-Glucanase. *Plant Physiol* 88:936-942.

564   McElrone AJ, Sherald JL, and Forseth IN. 2003. Interactive effects of water stress and xylem-limited bacterial
565   infection on the water relations of a host vine. *J Exp Bot* 54:419-430.

566   Metzner R, Schneider HU, Breuer U, Thorpe MR, Schurr U, and Schroeder WH. 2010. Tracing cationic nutrients
567   from xylem into stem tissue of French bean by stable isotope tracers and cryo-secondary ion mass
568   spectrometry. *Plant Physiol* 152:1030-1043. pp.109.143776 [pii]
569   10.1104/pp.109.143776

570   Mi H, Muruganujan A, Casagrande JT, and Thomas PD. 2013. Large-scale gene function analysis with the
571   PANTHER classification system. *Nat Protoc* 8:1551-1566. nprot.2013.092 [pii]
572   10.1038/nprot.2013.092

573   Moy P, Qutob D, Chapman BP, Atkinson I, and Gijzen M. 2004. Patterns of gene expression upon infection of
574   soybean plants by Phytophthora sojae. *Mol Plant Microbe Interact* 17:1051-1062.
575   10.1094/MPMI.2004.17.10.1051

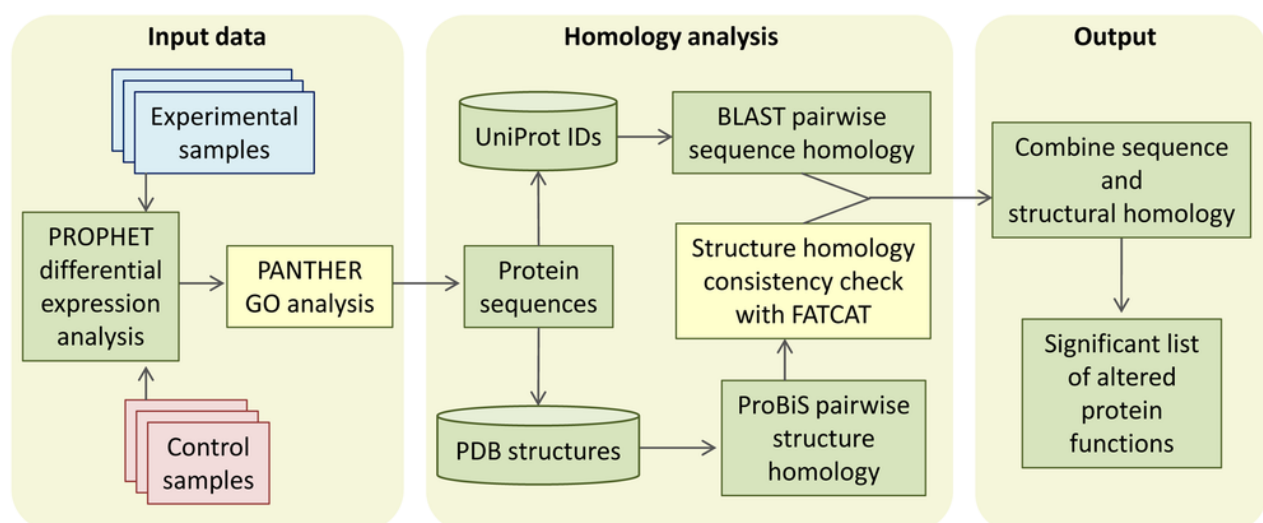
576   Nanjo Y, Nakamura T, and Komatsu S. 2013. Identification of Indicator Proteins Associated with Flooding Injury in
577   Soybean Seedlings Using Label-free Quantitative Proteomics. *Journal of Proteome Research* 12:4785-
578   4798. 10.1021/pr4002349

579   Nersissian AM, Immoos C, Hill MG, Hart PJ, Williams G, Herrmann RG, and Valentine JS. 1998. Uclacyanins,
580   stellacyanins, and plantacyanins are distinct subfamilies of phytocyanins: plant-specific mononuclear blue
581   copper proteins. *Protein Sci* 7:1915-1929. 10.1002/pro.5560070907
582   Nesvizhskii AI, Keller A, Kolker E, and Aebersold R. 2003. A statistical model for identifying proteins by tandem
583   mass spectrometry. *Anal Chem* 75:4646-4658.
584   Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, Poon TC, Ng SS, and Sung JJ. 2009. Differential expression of
585   microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening.
586   *Gut* 58:1375-1381. gut.2008.167817 [pii]
587   10.1136/gut.2008.167817
588   Nicaise V, Roux M, and Zipfel C. 2009. Recent advances in PAMP-triggered immunity against bacteria: pattern
589   recognition receptors watch over and raise the alarm. *Plant Physiol* 150:1638-1647. pp.109.139709 [pii]
590   10.1104/pp.109.139709
591   Niza B, Coletta-Filho HD, Merfa MV, Takita MA, and de Souza AA. 2015. Differential colonization patterns of
592   Xylella fastidiosa infecting citrus genotypes. *Plant Pathology* 64:1259–1269.
593   Ohnuma T, Numata T, Osawa T, Mizuhara M, Lampela O, Juffer AH, Skriver K, and Fukamizo T. 2011. A class V
594   chitinase from Arabidopsis thaliana: gene responses, enzymatic properties, and crystallographic analysis.
595   *Planta* 234:123-137. 10.1007/s00425-011-1390-3
596   Okushima Y, Koizumi N, Kusano T, and Sano H. 2000. Secreted proteins of tobacco cultured BY2 cells:
597   identification of a new member of pathogenesis-related proteins. *Plant Mol Biol* 42:479-488.
598   Petre B, Major I, Rouhier N, and Duplessis S. 2011. Genome-wide analysis of eukaryote thaumatin-like proteins
599   (TLPs) with an emphasis on poplar. *BMC Plant Biol* 11:33. 1471-2229-11-33 [pii]
600   10.1186/1471-2229-11-33
601   Pu Z, Ino Y, Kimura Y, Tago A, Shimizu M, Natsume S, Sano Y, Fujimoto R, Kaneko K, Shea DJ, Fukai E, Fuji S,
602   Hirano H, and Okazaki K. 2016. Changes in the Proteome of Xylem Sap in Brassica oleracea in Response
603   to Fusarium oxysporum Stress. *Front Plant Sci* 7:31. 10.3389/fpls.2016.00031
604   Rep M, Dekker HL, Vossen JH, de Boer AD, Houterman PM, Speijer D, Back JW, de Koster CG, and Cornelissen
605   BJ. 2002. Mass spectrometric identification of isoforms of PR proteins in xylem sap of fungus-infected
606   tomato. *Plant Physiol* 130:904-917. 10.1104/pp.007427
607   Rice P, Longden I, and Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends
608   Genet* 16:276-277. S0168-9525(00)02024-2 [pii]
609   Rodrigues CM, de Souza AA, Takita MA, Kishi LT, and Machado MA. 2013. RNA-Seq analysis of Citrus reticulata
610   in the early stages of Xylella fastidiosa infection reveals auxin-related genes as a defense response. *BMC
611   Genomics* 14:676. 1471-2164-14-676 [pii]
612   10.1186/1471-2164-14-676
613   Safavi K, Zareie R, and Tabatabaei BES. 2012. Constitutive expression of thaumatin-like protein (TLP-3) in
614   transgenic tobacco plants leads to enhance resistance to Alternaria alternata. *Archives of Phytopathology
615   and Plant Protection* 45:161-169. 10.1080/03235408.2010.507947
616   Sanchez-Rangel D, Sanchez-Nieto S, and Plasencia J. 2012. Fumonisin B1, a toxin produced by Fusarium
617   verticillioides, modulates maize beta-1,3-glucanase activities involved in defense response. *Planta
618   235:965-978. 10.1007/s00425-011-1555-0
619   Sela-Buurlage MB, Ponstein AS, Bres-Vloemans SA, Melchers LS, Van Den Elzen P, and Cornelissen B. 1993.
620   Only Specific Tobacco (Nicotiana tabacum) Chitinases and [beta]-1,3-Glucanases Exhibit Antifungal
621   Activity. *Plant Physiol* 101:857-863.
622   Shi X, Bi J, Morse JG, Toscano NC, and Cooksey DA. 2010. Differential expression of genes of Xylella fastidiosa
623   in xylem fluid of citrus and grapevine. *FEMS Microbiol Lett* 304:82-88. FML1885 [pii]
624   10.1111/j.1574-6968.2009.01885.x
625   Shinshi H, Wenzler H, Neuhaus JM, Felix G, Hofsteenge J, and Meins F. 1988. Evidence for N- and C-terminal
626   processing of a plant defense-related enzyme: Primary structure of tobacco prepro-beta-1,3-glucanase. *Proc
627   Natl Acad Sci U S A* 85:5541-5545.
628   Shiu SH, and Bleecker AB. 2001. Receptor-like kinases from Arabidopsis form a monophyletic gene family related
629   to animal receptor kinases. *Proc Natl Acad Sci U S A* 98:10763-10768. 10.1073/pnas.181141598
630   181141598 [pii]
631   Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H,
632   Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD,
633   Stupka E, Wilkinson MD, and Birney E. 2002. The Bioperl toolkit: Perl modules for the life sciences.
634   *Genome Res* 12:1611-1618. 10.1101/gr.361602

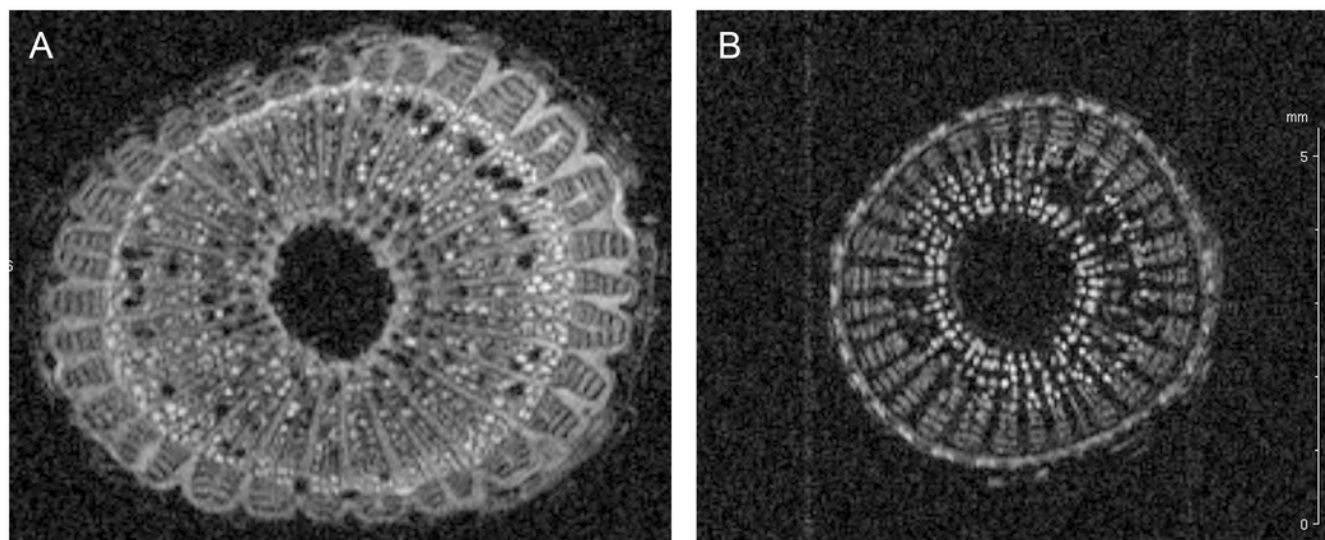
635 Su YC, Xu LP, Xue BT, Wu QB, Guo JL, Wu LG, and Que YX. 2013. Molecular cloning and characterization of
636   two pathogenesis-related beta-1,3-glucanase genes ScGluA1 and ScGluD1 from sugarcane infected by
637   Sporisorium scitamineum. *Plant Cell Rep* 32:1503-1519. 10.1007/s00299-013-1463-9
638 Subramanian S, Cho UH, Keyes C, and Yu O. 2009. Distinct changes in soybean xylem sap proteome in response to
639   pathogenic and symbiotic microbe interactions. *BMC Plant Biol* 9:119. 1471-2229-9-119 [pii]
640 10.1186/1471-2229-9-119
641 Sudisha J, Sharathchandra R, Amruthesh K, Kumar A, and Shetty HS. 2012. Pathogenesis related proteins in plant
642   defense response. . *Plant Defence: Biological Control*: Springer, 379–403.
643 van Loon LC, Rep M, and Pieterse CM. 2006. Significance of inducible defense-related proteins in infected plants.
644   *Annu Rev Phytopathol* 44:135-162. 10.1146/annurev.phyto.44.070505.143425
645 Varghese JN, Hrmova M, and Fincher GB. 1999. Three-dimensional structure of a barley beta-D-glucan
646   exohydrolase, a family 3 glycosyl hydrolase. *Structure* 7:179-190.
647 Wallis CM, and Chen J. 2012. Grapevine phenolic compounds in xylem sap and tissues are significantly altered
648   during infection by Xylella fastidiosa. *Phytopathology* 102:816-826. 10.1094/PHYTO-04-12-0074-R
649 Wang Z, Gerstein M, and Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*
650   10:57-63. nrg2484 [pii]
651 10.1038/nrg2484
652 Witzel K, Weidner A, Surabhi GK, Borner A, and Mock HP. 2009. Salt stress-induced alterations in the root
653   proteome of barley genotypes with contrasting response towards salinity. *J Exp Bot* 60:3545-3557. erp198
654   [pii]
655 10.1093/jxb/erp198
656 Xie YR, Raruang Y, Chen ZY, Brown RL, and Cleveland TE. 2015. ZmGns, a maize class I beta-1,3-glucanase, is
657   induced by biotic stresses and possesses strong antimicrobial activity. *J Integr Plant Biol* 57:271-283.
658   10.1111/jipb.12286
659 Yang L, Lin H, Takahashi Y, Chen F, Walker MA, and Civerolo EL. 2011. Proteomic analysis of grapevine stem in
660   response to Xylella fastidiosa inoculation. *Physiological and Molecular Plant Pathology* 75:90-99.
661 Yang Y, He M, Zhu Z, Li S, Xu Y, Zhang C, Singer SD, and Wang Y. 2012. Identification of the dehydrin gene
662   family from grapevine species and analysis of their responsiveness to various forms of abiotic and biotic
663   stress. *BMC Plant Biol* 12:140. 1471-2229-12-140 [pii]
664 10.1186/1471-2229-12-140
665 Ye Y, and Godzik A. 2004. FATCAT: a web server for flexible structure comparison and structure similarity
666   searching. *Nucleic Acids Res* 32:W582-585. 10.1093/nar/gkh430
667 32/suppl_2/W582 [pii]
668 Yoshida S, Ito M, Nishida I, and Watanabe A. 2001. Isolation and RNA gel blot analysis of genes that could serve
669   as potential molecular markers for leaf senescence in Arabidopsis thaliana. *Plant Cell Physiol* 42:170-178.
670 Zhang J, Li W, Xiang T, Liu Z, Laluk K, Ding X, Zou Y, Gao M, Zhang X, Chen S, Mengiste T, Zhang Y, and
671   Zhou JM. 2010. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune
672   receptors and are targeted by a Pseudomonas syringae effector. *Cell Host Microbe* 7:290-301. S1931-
673   3128(10)00104-6 [pii]
674 10.1016/j.chom.2010.03.007
675 Zhang W, Yan H, Chen W, Liu J, Jiang C, Jiang H, Zhu S, and Cheng B. 2014. Genome-wide identification and
676   characterization of maize expansin genes expressed in endosperm. *Mol Genet Genomics* 289:1061-1074.
677   10.1007/s00438-014-0867-8
678 Zhang YT, Zhang YL, Chen SX, Yin GH, Yang ZZ, Lee S, Liu CG, Zhao DD, Ma YK, Song FQ, Bennett JW, and
679   Yang FS. 2015a. Proteomics of methyl jasmonate induced defense response in maize leaves against Asian
680   corn borer. *BMC Genomics* 16:224. 10.1186/s12864-015-1363-1
681 10.1186/s12864-015-1363-1 [pii]
682 Zhang Z, Xin W, Wang S, Zhang X, Dai H, Sun R, Frazier T, Zhang B, and Wang Q. 2015b. Xylem sap in cotton
683   contains proteins that contribute to environmental stress response and cell wall development. *Funct Integr*
684   *Genomics* 15:17-26. 10.1007/s10142-014-0395-y
685 Zhu J, Sova P, Xu Q, Dombek KM, Xu EY, Vu H, Tu Z, Brem RB, Bumgarner RE, and Schadt EE. 2012. Stitching
686   together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that
687   modulate cell regulation. *PLoS Biol* 10:e1001301. 10.1371/journal.pbio.1001301
688 PBIOLOGY-D-11-03979 [pii]
689
690

1

The CHURNER workflow.

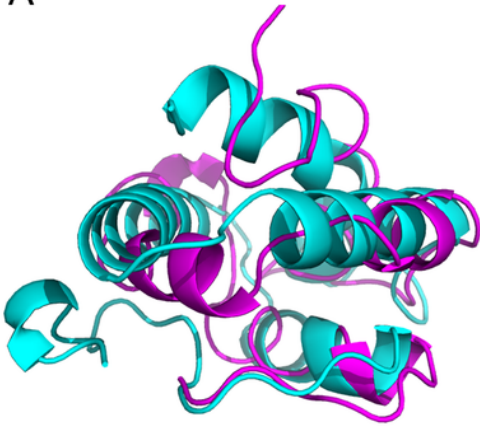
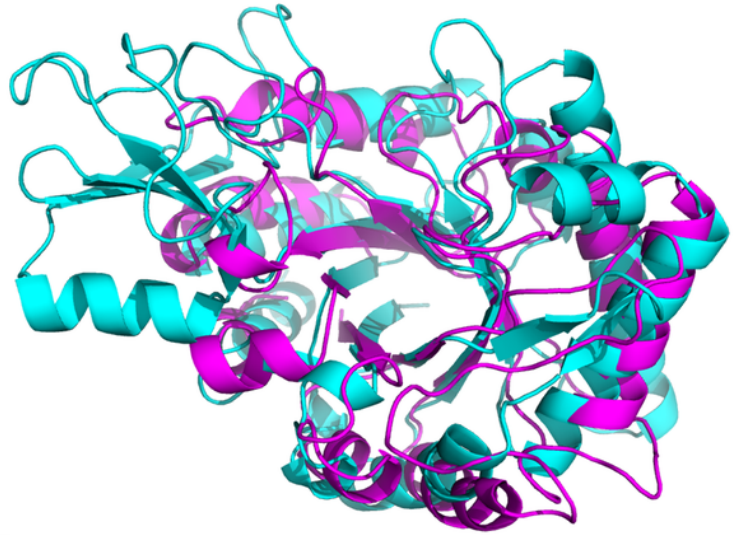
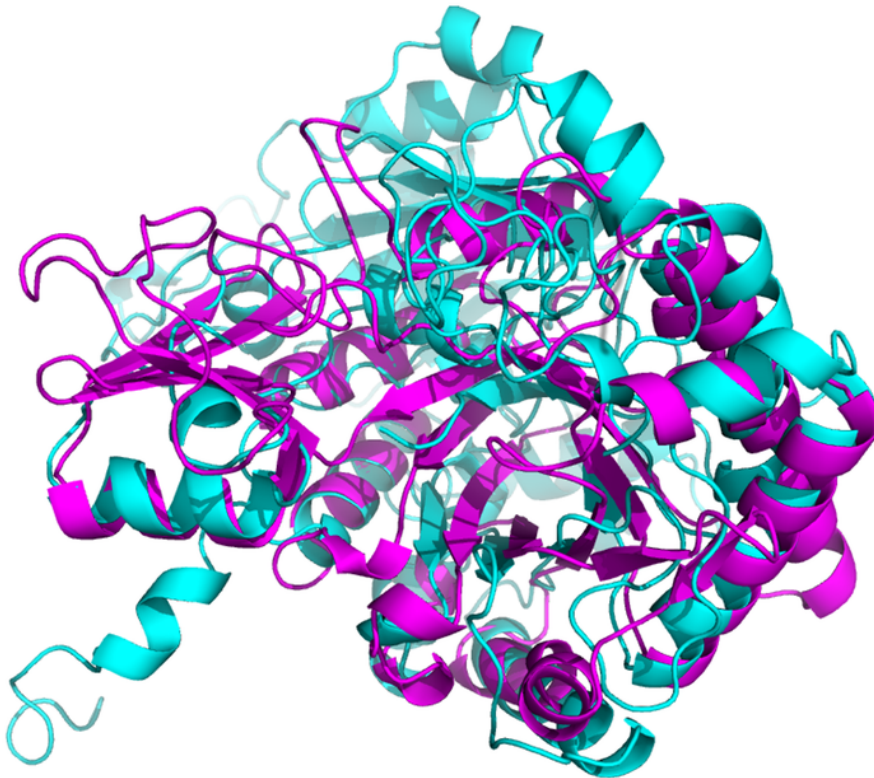
After differentially expressed proteins are selected and grouped by functional analysis, individual protein sequences are used to retrieve UniProt and PDB identificators. All sequences and structures are compared pairwise, and significant structural alignments are then use to reinforce protein functions that are significantly altered in the experiment. Yellow boxes indicate manual steps.

Cell wall thickening of infected grapevines.

Magnetic resonance imaging (1H-MRI) of stems of Xylella fastidiosa (A) infected and (B) non-infected grapevines. Note the brighter contrast (denser material) of secondary xylem and phloem vessels on the infected vine. Both images have the same magnification and scale bar is 5 mm. Images representative of transversal stem cuts obtained near 10 cm from top central stem. Similar thickening of cell walls were also observed in transversal cuts obtained along the vine until the base near the soil.

*Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.

3

Superimposition of proteins that have significant structural homology.

Structural homology has been detected using ProBiS, and confirmed using FATCAT. (A) Lipid transfer proteins: PDBid: 2RKNA (in magenta) and PDBid: 1FK0A (in cyan). (B) Chitinase GH18 proteins: PDBid: 1HVQA (in magenta) and PDBid: 3AQUA (in cyan). Note, that these proteins have low sequence homology (BLAST E-value = 0.003). (C) Chitinase (PDBid: 3AQUA, in magenta) and β-D-Glucan Exohydrolase (PDBid: 1EX1A, in cyan). These proteins have low sequence homology (E-value = 0.44).

A

B

C

4

Active site residues in the chitinase and β-D-glucan exohydrolase.

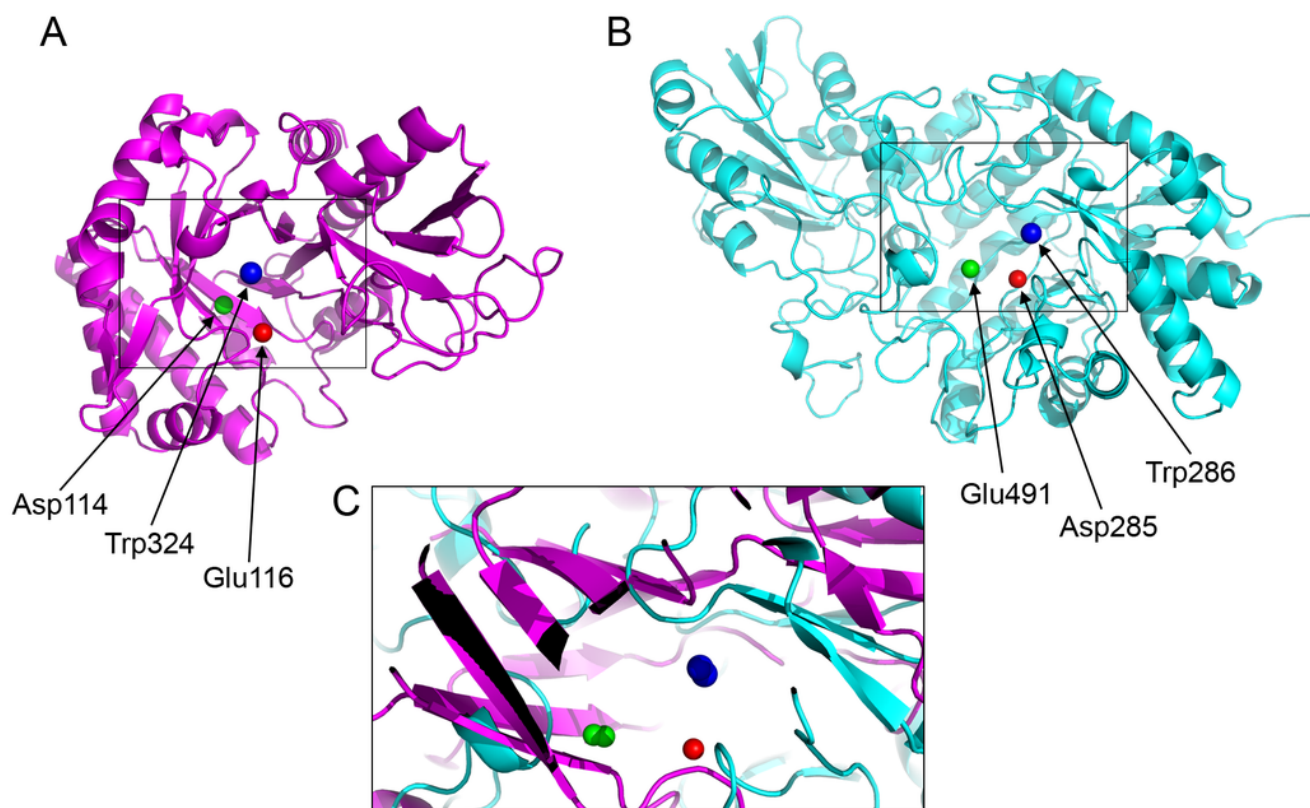
(A) Glu116 (red), Asp114 (green), Trp324 (blue) in chitinase (PDBid:3AQU). (B) Asp285 (red), Glu491 (green), Trp286 (blue) in β-D-glucan exohydrolase (PDBid:1EX1A). (C) Superimposition of the chitinase (in magenta) and β-D-glucan exohydrolase (in cyan).

Proteins identified in grapevine xylem sap by LC-MS/MS[a].

Table 1. Proteins identified in grapevine xylem sap of *Xylella fastidiosa*-infected and healthy plants by LC-MS/MS[a].

UniProt ID	Protein function	*Arabidopsis* best match ID	E-value	MW[b]	MP[c]	Cov. (%)[d]	SignalP	TargetP[e]
Pathogenesis-Related Proteins								
D7TXF5	Pathogenesis-related 4	AT3G04720.1	9E-51	15	6	62	21|22	Sec
F6HVL6	Pathogenesis-related 4	AT3G04720.1	2E-59	20	2	14	21|22	Sec
F6HBN7	Basic pathogenesis-related 1	AT2G14580.1	1E-47	18	4	37	25|26	Sec
A5BNW5	Pathogenesis-related 1	AT2G14610.1	2E-53	17	2	23	19|20	Sec
A5AWT7	Pathogenesis-related thaumatin	AT1G20030.2	8E-83	45	5	17	26|27	Sec
A5AWT9	Osmotin 34	AT4G11650.1	1E-79	24	13	50	24|25	Sec
A5B4P9	Osmotin 34	AT4G11650.1	2E-82	24	6	47	24|25	Sec
Q9M4G7	Osmotin 34	AT4G11650.1	4E-72	20	5	42	-	-
Q9M4G6	Osmotin 34	AT4G11650.1	7E-83	24	7	45	24|25	Sec
A5B2B6	Osmotin 34	AT4G11650.1	8E-80	24	3	35	24|25	Sec
Proteases								
A5AZU5	Aspartyl protease	AT5G07030.1	1E-134	40	5	15	-	-
D7T5Q6	Aspartyl protease	AT4G35880.1	1E-82	33	2	17	-	-
F6H3K5	Aspartyl protease	AT5G10770.1	1E-116	46	3	9	-	-
F6H3K6	Aspartyl protease	AT5G10770.1	1E-152	51	4	11	25|26	Sec
E0CQB3	Subtilisin	AT1G20160.1	0E+00	79	19	36	-	-
F6HNS0	Subtilisin	AT5G59100.1	1E-157	69	9	24	-	-
F6H4J9	Subtilisin	AT5G67090.1	1E-157	79	2	3	-	-
F6HSV1	Subtilase	AT5G67360.1	0E+00	81	18	46	24|25	Sec
F6I357	Subtilase	AT1G01900.1	0E+00	81	18	38	20|21	Sec
A5B179	Subtilase	AT1G01900.1	1E-163	75	6	14	20|21	Sec
A5AIJ0	Serine carboxypeptidase 20	AT4G12910.1	0E+00	54	6	15	26|27	Sec
A5C816	Serine carboxypeptidase 51	AT2G27920.1	1E-163	51	4	16	21|22	Sec
Carbohydrate-Active Enzymes								
F6GZC4	Basic chitinase	AT3G12500.1	3E-89	35	15	68	21|22	Sec
Q9ZTK4	Basic chitinase	AT3G12500.1	2E-85	35	11	52	21|22	TargetP
A5BK69	Chitinase A	AT5G24090.1	1E-109	32	12	48	25|26	Sec
F6H6H7	Chitinase A	AT5G24090.1	1E-108	32	9	37	25|26	Sec
F6HB09	Carrot EP3-3 chitinase	AT3G54420.1	1E-106	29	3	18	29|30	Sec
O24530	Carrot EP3-3 chitinase	AT3G54420.1	5E-90	27	2	16	20|21	Sec
F6HQS7	Alpha-L-arabinofuranosidase	AT3G10740.1	0E+00	84	20	26	28|29	Sec
F6HLL8	Beta-1,3-glucanase 3	AT3G57240.1	1E-109	38	23	69	33|34	Sec
F6HLL9	Beta-1,3-glucanase 3	AT3G57240.1	1E-110	37	4	22	32|33	Sec
A7PQW3	Beta-1,3-glucanase 3	AT3G57240.1	3E-89	37	2	8	29|30	Sec
F6I6R4	Beta-D-xylosidase 4	AT5G64570.1	0E+00	83	23	42	33|34	Sec
D7TXW6	Alpha-galactosidase 2	AT5G08370.1	1E-169	45	11	31	24|25	Sec
F6HGW2	Beta galactosidase 1	AT3G13750.1	0E+00	92	2	2	24|25	Sec
D7SKW9	Beta-galactosidase 8	AT2G28470.2	0E+00	92	2	4	23|24	Sec
D7TPI6	Alpha-amylase-like	AT4G25000.1	1E-161	47	6	20	22|23	Sec

A5C7G0	Glucuronidase 3	AT5G34940.2	0E+00	71	3	8	18|19	Sec
F6GU88	Glycosyl hydrolase	AT5G12950.1	0E+00	97	5	8	24|25	Sec
A5AZM8	Glycosyl hydrolase	AT2G27500.1	1E-132	50	2	5	-	-
F6H158	Glycosyl hydrolase	AT1G58370.1	0E+00	105	5	7	-	-
D7SVH6	Glycosyl hydrolase	AT3G26720.1	0E+00	114	2	3	19|20	Sec
D7TQ09	O-Glycosyl hydrolases	AT4G34480.1	1E-171	52	5	11	24|25	Sec
E0CQB9	O-Glycosyl hydrolases	AT4G34480.1	1E-178	50	3	12	22|23	Sec
D7T828	O-Glycosyl hydrolases	AT5G55180.1	0E+00	50	4	17	20|21	Sec
F6HCL5	Glycosyl hydrolases	AT4G19810.1	1E-107	40	5	24	25|26	Sec
D7T548	Glycosyl hydrolases	AT4G19810.1	1E-118	40	5	21	25|26	Sec
A7PZL3	Pectin lyase-like	AT3G61490.3	0E+00	53	2	10	-	-
F6HUM8	Pectin lyase-like	AT3G61490.3	0E+00	52	7	24	-	-
A5AZD0	Callose-binding protein 3	AT1G18650.1	3E-35	20	3	20	19|20	Sec
D7SI17	Callose-binding protein 3	AT1G18650.1	3E-39	21	2	20	19|20	Sec
A5C594	Expansin-like	AT4G17030.1	2E-40	23	6	42	24|25	Sec

Receptor-like Kinases (RLKs)

F6HIL5	Receptor-like kinase-related	AT3G22060.1	8E-69	27	16	64	24|25	Sec
A5AID0	Receptor-like kinase-related Leucine-rich repeat (LRR)	AT5G48540.1	2E-74	45	3	11	25|26	Sec
D7TPF3	SHV3-like 2	AT4G06744.1	1E-108	49	2	4	28|29	-

Peroxidases

F6GUF3	Peroxidase 2	AT5G06720.1	1E-109	36	2	7	23|24	Sec
F6GUE9	Peroxidase	AT5G19890.1	1E-104	29	15	65	-	-
F6HD61	Peroxidase	AT1G49570.1	1E-110	36	11	35	25|26	Sec
A5BJV9	Peroxidase	AT5G58390.1	3E-77	28	9	49	-	-
F6H776	Peroxidase	AT1G05260.1	8E-77	74	11	23	21|22	Sec
F6HIK4	Peroxidase	AT1G05260.1	1E-135	76	2	3	26|27	Sec
D7TQI6	Peroxidase	AT2G37130.1	1E-137	37	7	29	-	-
F6H3X3	Peroxidase	AT5G14130.1	8E-98	34	5	22	34|35	-
D7SVP1	Peroxidase	AT5G14130.1	2E-20	10	2	43	-	-
F6GXY7	Peroxidase	AT5G05340.1	7E-96	28	2	13	-	-
D7SR21	Peroxidase	AT5G05340.1	3E-98	28	3	19	-	-
F6H0Z1	Peroxidase	AT5G05340.1	1E-113	34	2	13	22|23	Sec
A5B8V0	Peroxidase	AT2G41480.1	6E-91	30	2	12	-	-
F6HH88	Peroxidase	AT2G41480.1	1E-120	70	2	7	24|25	Sec
F6HSU5	Peroxidase	AT5G67400.1	1E-138	36	3	17	27|28	Sec

Others

E0CQL6	Basic blue protein-like	AT2G02850.1	8E-38	19	8	49	-	-
D7TML8	Inhibitor / LTP / seed storage	AT3G53980.2	2E-33	12	11	67	27|28	Sec
D7SLG6	Inhibitor / LTP / seed storage	AT2G44290.1	1E-35	19	3	26	-	-
F6H7X9	Inhibitor / LTP / seed storage	AT4G33550.2	1E-07	12	3	39	29|30	Sec
A5C9S3	Inhibitor / LTP / seed storage	AT4G33550.2	4E-05	12	2	25	21|22	Sec
F6I0G4	Pectin methylesterase inhibitor	AT5G09760.1	0E+00	61	16	33	21|22	Sec

A5BS35	Basic seretory protein	AT2G15220.1	1E-83	25	10	53	23|24	Sec
F6HS61	Glycine-rich protein	AT4G30460.1	7E-03	13	7	67	22|23	Sec
A5AIZ1	Glycine-rich protein	AT4G30460.1	3E-03	13	4	67	22|23	Sec
D7TY88	Protease inhibitor	AT1G17860.1	6E-58	23	5	23	27|28	Sec
D7T293	Cupredoxin	AT4G12420.2	0E+00	66	2	5	23|24	Sec
A5BMY7	Cupredoxin	AT1G72230.1	7E-23	19	2	18	22|23	Sec
D7UBD5	Cupredoxin	AT3G27200.1	4E-40	18	4	41	23|24	Sec
A5BZS1	FAD-binding Berberine	AT4G20840.1	1E-179	59	5	10	30|31	Sec
A5B2E1	Cystatin / Monellin	AT5G47550.1	2E-29	13	2	15	24|25	Sec
A5BH21	PLC-like phosphodiesterase	AT1G66970.1	0E+00	70	3	5	21|22	Sec
D7SVW5	PI-PLC-like	AT4G36945.1	1E-143	45	2	9	27|28	Sec
A5BB66	Fasciclin-like	AT3G60900.1	1E-117	43	3	15	20|21	-
A5B7N6	Fasciclin-like	AT4G12730.1	1E-115	44	2	8	26|27	Sec
D7SXH0	Lamin-like	AT5G15350.1	3E-37	18	2	15	-	-
A5AIY9	Unknown protein	-	-	15	4	59	23|24	Sec

[a]A total of 91 proteins with at least two peptides sequenced per protein were identified and are displayed grouped by functional category. [b] Predicted molecular weight of proteins, in kDa. [c] Matched peptides. [d] Percentage of coverage. [e]TargetP output: Sec = secreted protein, - = undefined.

Differentially expressed proteins and their sequence and structure similarities with reference proteins.

Table 2. Differentially expressed proteins and their sequence and structure similarities with reference proteins.

UniProt ID	Protein function	GenBank / Protein Data Base			Exp.[a]
		ID	Score	E-value	
F6HLL8	β 1-3 glucanase	NP 001268153.1 / 4HPG	696 / 423	0e+00 / 1e-147	Up
A5BK69	Class III chitinase	ACH54087.1 / 1HVQ	592 / 435	0e+00 / 1e-153	Up
D7T548	Chitotriosidase-1	XP .1002270368.1 / 3AQU	743 / 452	0e+00 / 5e-158	Up
F6HBN7	Pathogenesis-related	XP 002274275 / 1CFE	334 / 184	1e-114 / 3e-59	Up
A5BNW5	Pathogenesis-related	XP 002273788.2 / 1CFE	333 / 186	7e-114 / 4e-60	Up
D7TML8	Lipid-transfer protein	XP 002281554.1 / 2RKN	239 / 36.2	9e-79 / 7e-04	Up
A5C594	Expansin-like	XP 002270175.2 / 2HCZ	374 / 89	2e-128 / 5e-21	Up
E0CQL6	Blue copper protein-like	XP 002266573.1 / 2CBP	268 / 158	4e-89 / 3e-49	Up
F6GZC4	Chitinase	CAC14015.1 / 4DWX	660 / 385	0e+00 / 7e-134	Down
A5AWT7	Thaumatin-like 1b	XP 002274137.1 / 3ZS3	518 / 253	1e-180 / 3e-81	Down
F6I6R4	β-xylosidase	XP 002264183.2 / 1EX1	1605 / 154	0e+00 / 2e-39	Down
D7SLG6	YLS3-like, Lipid-transfer protein	XP 002285691.1 / 1FK0	355 / 33.9	3e-122 / *1.7e-02*	Down

[a] Variation of protein level detected in sap from infected grapevine compared to uninfected control.

Pairwise BLAST[a] results of example proteins analyzed with CHURNER.

Table 3. Pairwise BLAST[a] results of example proteins analyzed with CHURNER.

UID[b] #1	UID #2	Identity[c]	E-value	Protein function
F6HBN7	A5BNW5	120/136	8e-90	Pathogenesis-related proteins (PR-1)
D7TML8	D7SLG6	20/78	9e-08	Lipid transfer proteins
A5BK69	A5C594	30/181	1.6e-2	Chitinase or expansin (?)
F6HBN7	A5BK69	16/133	2.4	PR-1 or Chitinase (?)
A5BK69	D7T548	36/272	2e-3	Chitinases GH18

[a] Sequences aligned with bl2seq from NCBI. [b] UniProt identificator. [c] Total number of identical amino acid residues considering the best alignment between the two proteins.

Pairwise superimposition of the PDB structures using ProBiS.

Table 4. Pairwise superimposition of the PDB structures using ProBiS.

PDBid / UID[a] #1	PDBid / UID #2	PZ[b]	Protein function	FCTSIG
2RKN / D7TML8	1FK0 / D7SLG6	2.4	Lipid transfer proteins	Yes
1HVQ / A5BK69	3AQU / D7T548	2	Chitinases GH18	Yes
3AQU / D7T548	1EX1 / F6I6R4	1.6	Chitinase and β-D-Glucan Exohydrolase	Yes
1HVQ / A5BK69	1EX1 / F6I6R4	1.3	Chitinase and β-D-Glucan Exohydrolase	No

[a] UID: UniProt identificator. [b] The results are sorted based on the ProBiS ZScore (PZ). Significance of structural alignment was verified using FATCAT (FCTSIG). Although the two chitinases have no sequence homology, their structural features are conserved. Similarly, we see signifiant structural similarity between a chitinase (UID: D7T548) and a β-D-Glucan Exohydrolase (UID: F6I6R4).

Potential and spatial congruence of the active site residues in proteins chitinase and β-D-glucan exohydrolase detected using CLASP.

Table 5. Potential and spatial congruence of the active site residues in proteins chitinase and β-D-glucan exohydrolase detected using CLASP.

PDB*	Active site atoms (a,b,c)		ab	ac	bc
1EX1A	ASP285 OD1,GLU491 OE1,TRP286 CZ2	**D**	6.4	8.4	9.3
		PD	-24.1	-273.2	-249.1
3AQU	GLU116 OE1,ASP114 OD1,TRP324 CZ2	**D**	7.1	7.9	9.5
		PD	49.8	-296.8	-346.7

* Chitinase: 1EX1A, β-D-glucan exohydrolase: 3AQU. The ability of CLASP to select stereo-chemically equivalent residues (Asp and Glu, both negatively charged residues) is critical to find the homologous active site. D = Pairwise distance in Å. PD = Pairwise potential difference. See Methods section for units of potential.