1 **EST mining of yellow potato (*Solanum phureja*) and gene expression profiling of the**

2 **infection by *Phytophthora infestans* using SAGE - SOLEXA**

3

4 Roberto Sierra,[1] Luis Miguel Rodríguez,[1] Adriana González[1], Alejandro Rojas,[1] Andrés

5 Pinzón,[1] Angela María Vargas,[1] Sonia Tinjacá,[2] Carlos Ñústez,[2] Adriana Bernal,[1] and

6 Silvia Restrepo[1]*

7

8 [1]Departamento de Ciencias Biológicas, Cra. 1 No. 18A-10, Universidad de los Andes,

9 Bogotá, Colombia and [2]Facultad de Agronomía, Cra. 45 No. 26-85 Edificio 500,

10 Universidad Nacional, Bogotá, Colombia. *To whom correspondence should be

11 addressed. Tel: +571 3394949 ext. 3748; Fax: +571 3394949 ext. 2817; Email:

12 srestrep@uniandes.edu.co

13

14   **ABSTRACT**

15   The late blight pathogen, *Phytophthora infestans* has a broad host range within the

16   Solanaceae family, including yellow potato (*Solanum phureja*). The disease caused by *P.*

17   *infestans* in *S. phureja* is poorly understood and is a major concern in Colombia.

18   Expressed Sequence Tag (EST) libraries obtained from a normalized library constructed

19   from healthy plant tissue revealed high levels of sequence similarity between *S. phureja*

20   and *S. tuberosum*. Then, utilizing Serial Analysis of Gene Expression and high-

21   throughput sequencing (SAGE-Solexa), we characterized yellow potato gene expression

22   during infection by *P. infestans*. Four-week-old yellow potato plants were inoculated with

23   *P. infestans* and were collected at 12 and 72 hours post inoculation for RNA extraction.

24   We detected differentially expressed genes by comparing inoculated to non-inoculated

25   and resistant to susceptible plants. Results showed that horizontal resistance in *S. phureja*

26   to *P. infestans* is highly dependent on systemic acquired resistance (SAR) and ethylene

27   signaling pathway. The discovery and characterization of the proteins mediating this

28   host–pathogen interaction enable the understanding of this pathosystem and is the key for

29   developing resistant plants.

30

31 **INTRODUCTION**

32   *Solanum phureja* consists of potato landraces widely grown in the Andes from

33 Venezuela to Bolivia (Ghislain, Andrade et al. 2006). It represents an important genetic

34 resource for Colombia. In one hand, its culinary characteristics make this species one of

35 the preferred cultivated potatoes for human consumption. Additionally, it constitutes a

36 breeding stock for developing new varieties (Ghislain, Andrade et al. 2006). Tubers lack

37 dormancy, it is characterized by a diploid ploidy, and it shows short-day adaptation.

38 However, yield losses occur through diseases and pests. One significant foliar and tuber

39 disease is late blight caused by the oomycete *Phytophthora infestans.*

40   Late Blight is the most important disease of potato (*Solanum tuberosum* and *S.*

41 *phureja*) worldwide. Colombia is the third largest potato producer in South America

42 (FAO 2007) and late blight is a disease of great concern. The disease was first officially

43 reported in Colombia at the beginning of the 20th century by Toro in 1927 (Toro 1927),

44 Chardon in 1928 (Chardon 1928) and Chardon and Toro (Chardon and Toro 1930) in

45 1930. Since official reports of the disease in Colombia have been scarce its impact may

46 be underestimated but growers have recognized late blight as severely limiting for the

47 crop. The most desirable method of control is the use of resistant cultivars, and the search

48 for resistance sources is a priority. Since the incorporation of major genes into

49 commercial varieties has lead to non-durable resistance due to the ability of the pathogen

50 to overcome the disease, breeders are interested in more stable type of resistance. Some

51 cultivars of *S. phureja* show horizontal or field resistance, a race non-specific resistance

52 thought to be more durable (Cañizares and Forbes 1995; Costanzo, Simko et al. 2005).

53 Previous studies using diploid mapping have improved the understanding of the genetic

54   control of field resistance (Leonards-Schippers, Gieffers et al. 1994; Collins, Milbourne

55   et al. 1999; Oberhagemann, Chatot-Balandras et al. 1999; Ewing, Simko et al. 2000;

56   Sandbrink, Colon et al. 2000; Ghislain, Trognitz et al. 2001; Costanzo, Simko et al.

57   2005). However, although some candidate genes have been associated with QTLs and *S.*

58   *phureja* genes have been identified and characterized upon *P. infestans* challenge

59   (Trognitz, Manosalva et al. 2002; Evers, Ghislain et al. 2006; Evers, Schweitzer et al.

60   2006) no high-throughput analysis of gene expression has been conducted to dissect the

61   molecular basis of quantitative resistance.

62   Among the methods for measuring gene expression, the Serial Analysis of Gene

63   Expression (SAGE) (Vesculescu, Zhang et al. 1995) shows a series of advantages. It

64   allows for the global profiling of an mRNA population regardless of whether the

65   transcripts have been previously identified in the same species or a related species. It is

66   also highly quantitative and SAGE profiles from different tissues are readily compared.

67   However, for the SAGE strategy to be successful, tags should be easily mapped to

68   previously curated transcripts. By October 27th, 2008 only 36 nucleotide sequences, 1

69   EST and 17 proteins were available for *S. phureja.* ESTs libraries provide a way to gain

70   insight into genomes at the DNA and RNA levels. While analyses of ESTs libraries allow

71   us to determine the level of similarity at the DNA level among well characterized

72   *Solanum* species and an unexplored one, data mining them could reveal unique protein

73   sequences of the yellow potato. Additionally, normalization of libraries improves the

74   probability to discover genes (Soares, Bonaldo et al. 1994).

75   We generated expressed sequenced tags (ESTs) to verify if *S. phureja* tags could

76   be mapped to *S. tuberosum* and *S. lycopersicum* transcripts. Here we report the

77 application of SAGE coupled to a high – throughput sequencing technology (Solexa,

78 www.illumina.com) to characterize the gene expression profile of horizontal resistance in

79 the infection of yellow potato (*S. phureja*) by *P. infestans* using common potato (*S.*

80 *tuberosum*) and tomato (*S. lycopersicum*) transcripts as templates for annotating the

81 SAGE tags. This study shows how this strategy could be successfully applied for an

82 economically and agronomically important crop that has not been well studied at the

83 molecular level.

84

85 **MATERIALS AND METHODS**

86 **Plants, pathogen isolates, and inoculations.**

87 Four-week-old *S. phureja* resistant (cv. Col3) and susceptible (cv. Col2) plants

88 were selected for inoculations. One day prior to inoculation, all plants were placed into a

89 humid chamber with a 12-h light period, 100% relative humidity, and an average

90 temperature of 15ºC. Susceptible cultivars of *S. phureja* (cv. Col2) and *S. tuberosum*

91 (Tuquerreña and Diacol Capiro) were used as inoculation controls and were left for

92 observation for 5 days after samples were collected to confirm that the inoculations were

93 successful.

94 Plants were inoculated with a mixture of *P. infestans* isolates collected from *S.*

95 *phureja* and *S. tuberosum* diseased tissues. Isolates 2011, 2021, 2031, 2041 and 2061

96 were isolated from *S. phureja* and previously described (Vargas, Quesada-Ocampo et al.

97 2008) and isolates A07, A13 and A15 (isolated from *S. tuberosum*), A08 and A09

98 (isolated from *S. phureja*) were characterized in this study. Inoculations were made as

99 previously described (Smart et al. 1998), spraying plants until run-off, using a sporangial

100   suspension at 32,000/ml. An equal number of plants were mock inoculated with water.

101   All leaflets from three plants per group were collected at three timepoints after

102   inoculation (0, 12, and 72 h). Tissue from the three plants was pooled, flash frozen in

103   liquid nitrogen, and stored at −80°C. Additional plants that had suffered the same

104   treatment were kept in the humid chamber for additional 7 days to ensure that inoculated

105   control susceptible plants became fully diseased, while mock-inoculated ones remained

106   healthy. The entire experiment was repeated three times.

107   *P. infestans* isolates were characterized as previously described (Vargas, Quesada-

108   Ocampo et al. 2008). Briefly, the sporangial width and length of each isolate were

109   determined. Other morphological characteristics specific to *P infestans* were also

110   considered for species identification including the shape of sporangia, ramification of the

111   sporangiophores and the presence of papillae (Erwin and Ribeiro 1996). Mating type of

112   isolates was determined according to the methodology described by (Forbes, Escobar et

113   al. 1997).

114   Additionally, the ITS of the rDNA of all *P. infestans* isolates was amplified using

115   PCR according to the methodology described by (Cooke, Drenth et al. 2000) and using

116   the ITS primers ITS 5 (5'-GGAAGTAAAAGTCGTAACAAGG) and ITS 4 (5'-

117   TCCTCCGCTTATTGATATGC) amplified the ITS region I between the 18S and 5.8S

118   rDNAs (White, Bruns et al. 1990). All the *P. infestans* isolates yielded a 914-bp product.

119   This product was then sequenced and sequences were compared to GenBank databases

120   using Blastn. Finally, all sequences were aligned (using the software CLC Free

121   Workbench, CLC bio, Cambridge, MA, USA). The *P. infestans* isolates were analyzed

122   for mitochondrial haplotypes using the following primer pairs for specific mitochondrial

123 DNA regions: P2 (P2F: 5'-TTCCCTTTGTCCTCTACCGAT-3'; P2R:5'-

124 TTACGGCGGTTTAGCACATACA-3'), and P4 (P4F: 5'-

125 TGGTCATCCAGAGGTTTATGTT-3'; P4R: 5'-CCGATACCGATACCAGCACCAA-

126 3') (Griffith and Shaw 1998).

127

128 **Potato RNA isolation, cDNA libraries and SAGE.**

129 Frozen plant tissues were ground, using a cold mortar and a pestle. Total potato

130 leaf RNA was extracted using the SV RNA Total Isolation kit (Promega, WI, USA). The

131 RNA was precipitated and equal amounts of RNA were pooled from samples collected at

132 same time points for either resistant or susceptible. They were stored at $-80°C$. For

133 cDNA libraries and EST analysis, RNA was pooled from *S. phureja* Col2 and Col3

134 cultivars. A cDNA library was constructed from these non-inoculated samples in

135 Evrogen, Moscow. Briefly, total RNA sample was used for dsDNA synthesis using

136 SMART approach (Zhu, Machleder et al. 2001). The resulting amplified cDNA was then

137 normalized using DSN normalization method (Zhulidov, Bogdanova et al. 2004).

138 Normalization included cDNA denaturation/reassociation, treatment by duplex-specific

139 nuclease (DSN, (Shagin, Rebrikov et al. 2002) and amplification of the normalized

140 fraction by PCR. The resulting normalized cDNA was used for library preparation.

141 Amplified normalized cDNA was purified using QIAquick PCR Purification Kit

142 (QIAGEN, CA, USA), and then cDNA was digested with restriction enzyme *Sfi*1.

143 Modified pAL 17.3 vector was digested with restriction enzyme *Sfi*1. Digested and

144 purified cDNAs (BD CHROMA SPINTM – 1000column) were ligated into modified

145 pAL 17.3 vector and used for *E. coli* transformation with the BioRad Micropulser (Bio-

146 Rad). The normalized library was arrayed in five 25x25 cm plates (approximately 10^0—

147 10^5 colonies per plate) and approximately 5.0×10^5 independent clones were obtained.

148 Colonies were resuspended in LB/Amp medium. The library was stored at $-70°C$.

149   For SAGE libraries, total RNA was obtained from both Col2 and Col3 cultivars

150 and for each of the three timepoints. RNA obtained from each of the three biological

151 replicates was pooled to obtain a single RNA sample for each time point for each cultivar

152 combination. A total of six different SAGE libraries were thus obtained. For all libraries,

153 Illumina sequencing was performed at Canada´s Michael Smith Genome Sciences Centre

154 (http://www.bcgsc.ca/) using Solexa/Illumina technology.

155

156 **Validation of gene expression results by real-time quantitative (rtq)-PCR**

157   Two genes that were identified as highly induced or suppressed during the *P.*

158 *infestans– S. phureja* interaction were chosen to validate the differential expression

159 obtained using SAGE. The genes included are endo-1,3-beta-D-glucanase, pathogenesis

160 related protein isoform b1 (PR-1b1), and elongation-factor 1-alpha as a housekeeping

161 gene. Total RNA (1 µg), from resistant (cv. Col3) and susceptible (cv. Col2) interactions

162 was used to generate cDNA, and total transcript levels were determined by rtq-PCR using

163 a two-step approach. The iScript Select cDNA Synthesis Kit (BioRad, Hercules, CA) was

164 used for reverse transcription following the manufacturer's instructions to a final volume

165 of 20 µl. Reverse transcription reaction was carried out for 90 min at 42ºC followed by 5

166 min at 85ºC. Following reverse transcription step, 1 µl of the resulting cDNA was used as

167 a template for rtq-PCR with iQ SYBR Green Supermix (BioRad, Hercules, CA)

168 following the manufacturer's instructions. The PCRc onsisted of an initial denaturation

169   step at 95ºC for 10 min, the amplification occurred in a three-step procedure: 15 s at

170   95°C, 1 min at 58°C, and 30 s at 72°C, with a total of 45 cycles. Controls lacking reverse

171   transcriptase were included to check for DNA contamination in the cDNA generation

172   step, and controls lacking template were also included in the PCR. The primers used for

173   rtq-PCR are shown in Table 1. Transcript levels were calculated using the Pfaffl method

174   (Pfaffl 2001). Standard curves were generated by performing rtq-PCR on serial dilutions

175   of a PCR amplification product of the target gene. Standard curves were constructed by

176   plotting the threshold cycle versus the logarithm of the concentrations. The relative

177   quantity of the product in each sample was calculated ($\Delta\Delta C_t$) from the standard curves

178   and was normalized against the potato elongation factor 1-alpha gene as an internal

179   control. Reverse transcription rtq-PCR product generated from RNA isolated from mock-

180   inoculated tissue at each time point was used as a reference sample. The reference sample

181   was considered to be the 1X expression level, and results were expressed as the fold

182   increase of mRNA level over the reference sample.

183

184   **Computational Methods**

185   EST sequences were trimmed to eliminate vector and low-quality sequences using

186   the SeqClean software (parameters by default). Sequences sharing greater than 90%

187   identity over 30 or more contiguous bases were placed into clusters. Clusters were

188   assembled using the Cap3 Assembler (http://seq.cs.iastate.edu; (Huang and Madan 1999))

189   to produce tentative cluster (TC) sequences. Sequences not clustered into a TC were

190   named singletons.

191   The TCs and singletons were searched against the nr nucleotide database (word

192 size = 15 and e-value threshold = e^{-20}). Also, they were searched against and a nr protein

193 database (word size = 3, matrix = BLOSUM62) to provide a putative function. The TCs

194 and singletons were also annotated using the 12 algorithms of InterProScan (Zdobnov

195 and Apweiler 2001) against the InterPro database (Apweiler, Attwood et al. 2001). Gene

196 Ontology analysis was also performed since InterPro hits are automatically linked to GO

197 annotations.

198   The SAGE tags were counted, and catalogued using SAGE2000

199 (http://www.invitrogen.com/sage). The significance of the differential expression

200 between two libraries was calculated according to the formula of Audic and Claverie

201 (Audic and Claverie 1997). The condition for significant differences between two

202 libraries was set at a *p*-value of <0.01. To estimate the extent of similarity between two

203 libraries or to compare the change of expression of particular tags in each library, we

204 normalized the tag frequencies by scaling them all up to the same total number of tag

205 copies (4,000,000 tags / library) (Porter, Krop et al. 2001; Sander, Ng et al. 2005).

206   SAGE tags were identified in two ways. First we used Identitag (http://pbil.univ-

207 lyon1.fr/software/identitag/) (Keime, Damiola et al. 2004) to identify tags from our

208 SAGE libraries using the TCs and singletons we obtained from *S. phureja* and using the

209 *S. tuberosum* and *S. lycopersicum* TIGR TCs and singletons (release April, 2008),

210 (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=potato   and

211 http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=tomato). Tags that could

212 not be identified using ours or TIGR transcripts were mapped to a virtual Tag library

213 constructed   from   the   NCBI   tomato

214 (http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=4081)   and   potato

215 (http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=4113)   UniGene   build

216 database (release April, 2008).

217

218 **RESULTS**

219 **Potatoes challenged to *P. infestans***

220   No symptoms of infection were observed on any of the potato plants (resistant or

221 susceptible cultivars) in the first 48 h following inoculation. Several small black lesions

222 were seen on plants inoculated with *P. infestans* isolates three days after inoculation. By

223 five days after inoculation, severe symptoms and pathogen sporulation were observed on

224 all susceptible controls of *S. tuberosum* as well as in the Col2 *S. phureja* susceptible

225 cultivar left as inoculation control. Likewise, no symptoms were observed on potato

226 plants mock – inoculated with water (healthy control plants).

227   All isolates fit the morphological description of *P. infestans* with deciduous and

228 semi – papillate sporangia (Erwin and Ribeiro 1996). Sporangia had a mean length of

229 39.4 ± 1.1 µm and a mean width of 26.1 ± 1.9 µm falling within the range reported for *P.*

230 *infestans* (Erwin and Ribeiro 1996). The rDNA ITS1 characterization also supported the

231 hypothesis that all isolates are *P. infestans*. In the analysis using Blastn, all sequences

232 showed at least 97% similarity (with e value threshold equal to 0) to sequences in the

233 databases reported as *P. infestans*. All isolates show the A1 mating type and were

234 grouped into the mitochondrial haplotype IIa, as was expected for these strains isolated in

235 Colombia according to previous characterization of strains in this area (Vargas, 2008.

236 Article in press)

237

EST sequence analysis and functional characterization

238

239   To determine sequence similarity of expressed genes in *S. phureja*, we sequenced

240 2112 ESTs. These sequences were screened for quality and a total of 2098 good

241 sequences with an average edited length of 840 bases were generated (GenBank

242 accession numbers FG645638 – FG647735). 488 ESTs were assembled into 215 TCs

243 while 1610 sequences remained as singlet's (Figure 1). This corresponds to a redundancy

244 of 26,7% (number of ESTs assembled in TCs/total number of ESTs), which is very low,

245 showing the high efficiency of the normalization.

246   Using Blastn against the nr database we found a hit for 1814 sequences with an

247 average of 640bp high scoring segment pair (HSP) setting an e-value threshold of e^{-20}.

248 Among these, 97% of the sequences showed significant similarities with the Solanaceae

249 (74% showed similarity with *S. tuberosum* and 23% with other species), 2% with other

250 botanical families and 1% with sequences of unknown origin. 433 sequences showed

251 99% identity to *S. tuberosum* sequences, 219 showed 98% and 105 showed 97% identity

252 to this species. Also, 11 sequences showed 100% identity with *S. lycopersicum* sequences

253 and 256 showed identity comprised between 99 and 95% with sequences of this species.

254 These results showed that SAGE tags could be mapped to virtual tags extracted from

255 transcripts of the Solanaceae with high reliability.

256   To determine the percentage of sequence identity at the protein level between

257 *S. phureja* and members of the Solanaceae family or to any protein in the database that

258 could facilitate SAGE tag identification and to identify potential homologs to the yellow

259 potato, the Blastx algorithm was used to translate each TC and singleton sequence into

260 the six possible reading frames. These were then compared against the nr NCBI protein

261 sequences database. For the 1825 TCs and singletons, we obtained 1710 hits or matches.

262 Among these we only considered those with an HSP equal or longer than 100 amino

263 acids. Among these 1417 sequences, 43% showed similarity to sequences from *Vitis*

264 *vinifera*, 13% to sequences from *S. tuberosum*, 7% to sequences from *S. lycopersicum*

265 and 34% to sequences from several different species. Also, from these 1417, a total of

266 1367 had an identity equal or greater than 40% with sequences in the nr database.

267 Matches against the nr NCBI protein database were divided into highly significant (P

268 values $\leq 10^{-20}$), and moderately significant (10^{-5} to 10^{-19}). Among the 1825 TCs and

269 singletons, 1368 were classified as highly significant, and 49 as moderately significant.

270   The 1825 TCs and singletons were also annotated using the InterPro protein motif

271 database with the 12 algorithms implemented in InterProScan. 1367 TCs and singletons

272 did show a match in at least one of the databases searched by InterPro Scan. The most

273 represented protein domains in the *S. phureja* EST TCs and singletons are summarized in

274 Table 2. Molecular functional categories as determined by the GO Consortium appear in

275 Figure 2 for 916 sequences. The InterPro Scan tool permitted to assign a putative function

276 to 148 sequences that did not show a hit using Blastx. This result is particularly relevant

277 because in this way *S. phureja* unique or specific sequences can be detected. Several

278 enzymes involved in different metabolic processes could be identified, ribosomal

279 proteins, enzymes participating in the carbohydrate metabolic process among others.

280 Additionally, among these 148 sequences some defense genes could be identified which

281 could be useful to identify SAGE tags in the transcript profiling study. These included

282 protein inhibitor I9, Barwin-related endoglucanase (Biological Process: defense response

283 to fungus (GO:0050832)), metallothionein.

284

285 **SAGE**

286   To build gene expression profiles of the horizontal resistance in *S. phureja* to *P.*

287 *infestans,* six SAGE libraries were constructed from RNAs harvested from leaves of two

288 cultivars, one resistant (cv. Col3) and one susceptible (cv. Col2) and at 0h, 12h and 72h

289 post inoculation (hpi). From now on, libraries will be named 0HR, 0HS, 12HR, 12HS,

290 72HR and 72HS (Accession numbers: GSM298192 (72HS), GSM298191 (72HR),

291 GSM298190 (12HS), GSM298189 (12HR), GSM298188 (0HS), GSM298187 (0HR)).

292 From the six libraries, 38,461,463 SAGE clones 17-bp in length were sequenced. From

293 these tags we eliminated the low-complexity tags, those having more than 50% of their

294 length composed of a stretch of the same nucleotide. Then, 38,090,130 tags were

295 analyzed that yielded 7,544,496 different (or unique species of) SAGE tags. Data for each

296 library appear in Table 3 and an overview of the abundance of the different libraries is

297 presented in Figure 3.

298   To assign the tag sequences to the tomato and potato genes, we used Identitag

299 (Keime, Damiola et al. 2004) and Unigene Build from NCBI (release April, 2008).

300 Identitag is based on a relational database structure relating observed tags with a virtual

301 tag library and with a database that stores the Blastx results of the transcripts used to

302 construct the virtual library. For this study in identitag we used *S. tuberosum* and *S.*

303 *lycopersicum* TIGR Gene indices and the TCs and singletons described and annotated.

304   In this study, we were interested in differentially expressed genes in the resistant

305 cultivar compared to the mock inoculated plants or to the susceptible cultivar and

306 proceeded to identify statistically significant differences in the tag frequency of the up

307 regulated and down regulated for different libraries combinations (Table 4). Among the

308 differentially expressed tags (significant at $p < 0.01$), the induced and repressed genes for

309 each library comparison were chosen for analysis. From these genes we selected the ones

310 that had a $p = 0$ and had a blast hit with a cut-off of 1e0 against the *S. phureja* ESTs from

311 this study or the potato TIGR TAs (The Institute for Genomic Research Transcript

312 Assemblies) (Table 4). In the case of a resistant vs. susceptible comparison, we were

313 interested in the down- and up regulated genes in the resistant library. To find markers

314 associated to resistance or susceptibility we analyzed the differentially expressed genes

315 by pairwise comparisons of the six SAGE libraries. The complete SAGE data and lists of

316 differentially   expressed   genes   are   available   at   our   database

317 http://bioinf.uniandes.edu.co/soldb/. For further analyses we selected genes responsive to

318 the jasmonic acid (JA), salicylic acid (SA), and ethylene (ET) signaling pathway and

319 genes related with defense in the resistant cultivar (Table 1 and 4). Finally, for

320 comparisons of different time points from the same cultivar we reported the down- and

321 up-regulated genes in later time points. We found previously reported genes in the

322 interaction of S. phureja – P. infestans such as pathogenesis-related protein (pr1b1),

323 osmotin-like protein (PR-5), Phenylalanine ammonia-lyase (PAL), cytochrome P450

324 (Trognitz, Manosalva et al. 2002; Castillo Ruiz, Herrera et al. 2005; Evers, Schweitzer et

325 al. 2006). For a complete list see supplementary tables.

326

327 **Real-time quantitative (rtq) PCR Analysis and SAGE validation**

328   To validate the results from pooled samples obtained in the SAGE experiment and

329 to determine if differences in gene induction and repression present in the SAGE-Solexa

330 experiment from pooled RNA was due to particular samples (replicate 1, 2 or 3),

331 individual replicates and pooled samples were used as templates in rtq-PCR experiments

332 for each gene. Previously reported genes expressed in compatible and incompatible

333 reaction in Solanaceae challenged with *P. infestans* with vertical resistance were chosen

334 from the 14 different paired comparisons (Shah and Klessig 1996; Jia and Martin 1999;

335 Smart, Myers et al. 2003; Restrepo, Myers et al. 2005) for a total of 13 genes (Table 1).

336 Rtq-PCR for the basic glucanase gene (GluB) appeared to be slighty repressed in

337 resistant plants (cv. Col3) down regulating 0.3-fold at 72 hpi compared to the healthy

338 control, while it expression was up regulated 0.6-fold at 72 hpi in susceptible plants. This

339 gene was found up regulated in SAGE analyses when comparing 0HS vs. 72HS. Basic

340 PR-1 (pr1b1) was up regulated in SAGE analyses for 0HS vs. 12HS and 0HR vs. 72HR.

341 Similarly, in rtq-PCR assays the expression ratios obtained for pr1b1 were 1.14-fold

342 upregulated at 72 hpi in the resistant cv. Col3 and 1.6-fold at 72 hpi for cv. Col2. These

343 results agree with the SAGE results validating the gene expression.

344

345 **DISCUSSION**

346 This study represents the first high-throughput profiling of gene expression on

347 plants with horizontal resistance in the *Phytophthora – Solanum* pathosystem. We were

348 able to successfully extend the knowledge in the field of horizontal resistance from the

349 interaction of the late blight pathogen and yellow potato, which has been poorly studied,

350 with minimum sequence information. We demonstrate that SAGE coupled with a large-

351 scale sequencing technology can be used in poorly studied organisms applying

352 comparative genomics tools. Our results demonstrate that horizontal resistance in *S.*

353 *phureja* to *P. infestans* is highly dependent on systemic acquired resistance (SAR) and

354 the ethylene signaling pathway.

355   *S. phureja* has been described as partially resistant (Cañizares and Forbes 1995)

356 and it has been used in breeding programs to characterize QTLs with the final goal to

357 introgress resistance into the tetraploid *S. tuberosum* genetic backgrounds (Costanzo,

358 Simko et al. 2005). In Colombia, yellow potato itself is an important crop for human

359 consumption. However, little is known about this important resource, which is threatened

360 by the late blight pathogen *P. infestans*. Some efforts have been made to characterize the

361 molecular basis of field resistance in *S. phureja* to *P. infestans* but with a limited number

362 of genes using the candidate gene approach (Trognitz, Manosalva et al. 2002; Evers,

363 Schweitzer et al. 2006) or using differential display (Evers, Ghislain et al. 2003; Evers,

364 Ghislain et al. 2006) but not high-throughput or quantitative techniques. Additionally

365 hybrid plants have been traditionally used and no time-course expression has been

366 exploited to follow the expression of defense genes in both resistant and susceptible

367 cultivars.

368   The similarity of *S. phureja* with other members of its family had not been studied

369 before. We found high nucleotide similarity in *S. phureja* with Solanaceae (97% of the

370 1825 unigenes) mainly with potato and tomato. Most of the ESTs showed over 97%

371 identity with potato and tomato. Blastx results showed that 97% of the ESTs had a hit

372 with known proteins in the nr protein database. The EST annotation with GO by most

373 common domains and the distribution of molecular function are shown in Table 2 and

374 Figure 2. Before obtaining these results, it was not clear if it was possible to conduct a

375 large-scale genomics study with this agriculturally important potato due to the lack of

376   sequence information. These sequences and its similarity results allowed an efficient

377   mapping of the tags obtained. They also open the possibility to conduct studies with *S.*

378   *phureja* on comparative genomics that will benefit the knowledge on the crop itself and

379   extrapolate results with *S. tuberosum*. Our contribution of new ESTs of *S. phureja* is of

380   great interest for a data mining study based on bioinformatics in which new genes and

381   polymorphisms of this species can be found and thus present new insights.

382   A recurrent problem in SAGE studies is the correct assignment of tags to genes or

383   the tag-to-gene mapping. For our pathosystem this problem was particularly critical

384   because very few sequences for *S. phureja* were available in public databases. In

385   consequence we took advantage of the overall sequence similarity between members of

386   the Solanaceae and we demonstrated the high levels of transcripts similarity between *S.*

387   *phureja* and tomato and potato and we even used the *S. phureja* ESTs to map the SAGE

388   tags along with potato and tomato unigenes. SAGE provides an economical way to study

389   transcript profiles and quantify expression of plants under different stress conditions.

390   Besides helping with the tags annotations, the EST collection constitutes the first high-

391   throughput effort to characterize the yellow potato transcriptome and a source of markers

392   for mapping and breeding.

393   We suggest as markers for mapping or breeding the up-regulated genes in the

394   pairwise comparison of 0 hpi vs. 72 hpi in the resistant cultivar. At 72 hpi small lesions

395   were observed in leaf tissue in the susceptible cultivar indicating that resistant plants

396   triggered defense. Many of the genes we obtained have been previously reported in the

397   ethylene or salicylic acid signaling pathway and in systemic acquired resistance. From the

398   ethylene signaling pathway, osmotin-like (PR-5 or thaumatin-like) was highly expressed

399 at 12HR and 72HR and down-regulated at 72HS. It has been previously reported that

400 osmotin delays the late blight symptoms in *S.tuberosum*?. Another PR protein highly

401 expressed in the resistant cultivar 72 hpi and down-regulated in the susceptible cultivar is

402 the basic PR-1 which is also down-regulated when comparing resistant vs. susceptible

403 (12HR vs. 12HS) (Liu, Raghothama et al. 1994; Zhu, Chen et al. 1995; Smart, Myers et

404 al. 2003). The chitinases (PR-3, -4, -8, -11) such as class I, class II chitinases, and

405 endochitinases followed the same expression pattern and showed to be up-regulated when

406 either comparing 0HR-72HR or 12HR-72HR and repressed in the susceptible cultivar

407 (72HS).

408   All of the above genes have been shown to be involved in systemic acquired

409 resistance (SAR) and especially PR-1 and -5 proteins have been used as SAR markers

410 because of the correlation between timing of the PR gene expression and the onset and

411 duration of SAR (Ryals, Neuenschwander et al. 1996; Takemoto, Furuse et al. 1997;

412 Vleeshouwers, Van Dooijeweert et al. 2000). Although salicylic acid (SA) is related to

413 SAR (Durrant and Dong 2004) our study does not demonstrate the induction of SA in

414 either resistant or susceptible cultivars. However, salicylic acid-binding protein 2

415 (SABP2) with lipase activity was induced in resistant plants (72 hpi) and have been

416 suggested to be a receptor of SA during development of SAR and required for the plant

417 immune response (Kumar and Klessig 2003). Some of these genes may appear to be up-

418 regulated in the susceptible cultivar at 12 hpi but then are down-regulated at 72 hpi. This

419 might be due to pathogen's ability to suppress defenses during colonization of the host.

420   Carbonic anhydrase (CA) has been related to HR in tobacco, tomato, and potato

421 (Kumar and Klessig 2003; Smart, Myers et al. 2003; Restrepo, Myers et al. 2005).

422   Resistant plants exhibited induced expression at 12 and 72 hpi while susceptible plants

423   repressed CA from 0 to 72 hpi indicating a possible general role of pathogen-induced

424   gene suppression in host susceptibility. The cysteine protease inhibitor 1 (CPI1) was up-

425   regulated in resistant plants and, contrarily to what was expected, it has been reported to

426   block HR or programmed cell death (PCD) triggered by an avirulent strain or oxidative

427   stress (Solomon, Belenghi et al. 1999; Avrova, Taleb et al. 2004; Tian, Huitema et al.

428   2004; Rooney, Van't Klooster et al. 2005; Tian and Kamoun 2005; Tian, Win et al. 2007).

429   It is possible that this induction of CPI1 in the resistant plants is part of the modulation of

430   HR as a response to the concentrated inoculum used in this study. A full list of these

431   genes and other expressed genes with a $p=0$ and a hit against the nr protein database is

432   available in the supplementary material.

433   Real-time quantitative PCR complemented the results obtained from SAGE-

434   Solexa analyses not only by validating the results but also by providing the means to

435   quantify the expression of the selected genes in a horizontal resistance – based

436   pathosystem. The fold expression of genes did not vary dramatically from one time point

437   to another and showed slight changes in expression when comparing susceptible to

438   resistant (i.e. basic PR-1), as was expected. In this study, it is important to focus on the

439   broad range of genes being induced and repressed at a given time point since the basis of

440   qualitative resistance relies in many minor genes that regulate and confer field resistance

441   and not on abrupt changes in the expression of a particular gene.

442   In conclusion, the horizontal resistance studied here is strongly supported on a

443   large group of genes related to SAR and the ET and SA signaling pathways. Small

444   variations in the levels of expression of all these genes modulate the resistance of *S.*

445 *phureja* to *P. infestans*. *S. phureja* is not only a good model to study this interaction

446 because of its qualitative resistance against the pathogen but because it is agriculturally

447 important. We have greatly contributed to the understanding of the horizontal resistance

448 in the *Solanum – Phytophthora* pathosystem by coupling large-scale cDNA analysis, high

449 throughput sequencing technology, comparative genomics, and bioinformatic analyses.

450 Bioinformatic studies can be conducted with the available SAGE-Solexa results to

451 understand the pathogen expression profile when challenged to a host with a complex

452 resistance such as yellow potato. Finally, it is important to include the suggested markers

453 in further studies and breeding programs to validate and incorporate field resistant genes

454 to *S. tuberosum* aiming to breed plants with longer lasting resistance.

455

460

461 **REFERENCES**

462

463 Apweiler, R., T. K. Attwood, et al. (2001). "The InterPro database, an integrated

464 documentation resource for protein families, domains and functional sites."

465 Nucleic Acids Res **29**(1): 37-40.

466 Audic, S. and J. M. Claverie (1997). "The significance of digital gene expression

467 profiles." Genome Res **7**(10): 986-95.

468 Avrova, A. O., N. Taleb, et al. (2004). "Potato oxysterol binding protein and cathepsin B
469   are rapidly up-regulated in independent defence pathways that distinguish *R* gene-
470   mediated and field resistances to *Phytopthora infestans*." <u>Molecular Plant</u>
471   <u>Pathology</u> **5**(1): 45-56.

472 Cañizares, C. A. and G. A. Forbes (1995). "Foliage resistance to Phytophthora infestans
473   (Mont.) de Bary in the Ecuadorian national collection of Solanum phureja ssp.
474   phureja Juz. and Buk." <u>Potato Research</u> **38**: 3-10.

475 Castillo Ruiz, R. A., C. Herrera, et al. (2005). "Organization of phenylalanine ammonia
476   lyase (PAL), acidic PR-5 and osmotin-like (OSM) defence-response gene families
477   in the potato genome." <u>Mol Genet Genomics</u> **274**(2): 168-79.

478 Chardon, C. E. (1928). "Contribución al estudio de la flora micológica de Colombia."
479   <u>Bol. Real Soc. Esp. Hist. Nat.</u> **28**: 111-124.

480 Chardon, C. E. and R. A. Toro (1930). "Mycological explorations of Colombia." <u>J- Dep.</u>
481   <u>Agric. Porto Rico</u> **14**: 195-357.

482 Collins, A., D. Milbourne, et al. (1999). "QTL for field resistance to late blight in potato
483   are strongly correlated with maturity and vigour." <u>Molecular Breeding</u> **5**: 387-398.

484 Cooke, D. E. L., A. Drenth, et al. (2000). "A Molecular phylogeny of *Phytophthora* and
485   related oomycetes." <u>Fungal Genetics and Biology</u> **30**: 17-32.

486 Costanzo, S., I. Simko, et al. (2005). "QTL analysis of late blight resistance in a diploid
487   potato family of Solanum phureja x S. stenotomum." <u>Theor Appl Genet</u> **111**(3):
488   609-17.

489 Durrant, W. E. and X. Dong (2004). "Systemic acquired resistance." <u>Annu Rev</u>
490   <u>Phytopathol</u> **42**: 185-209.

491   Erwin, D. C. and O. K. Ribeiro (1996). <u>Phytophthora diseases worldwide</u>. St. Paul, APS

492   Press.

493   Evers, D., M. Ghislain, et al. (2003). "Differential gene expression in two potato lines

494   differing in their resistance to Phytophthora infestans." <u>J Plant Physiol</u> **160**(6):

495   709-12.

496   Evers, D., M. Ghislain, et al. (2006). "A late blight resistant potato plant overexpresses a

497   gene coding for α-galactosidase upon infection by *Phytophthora infestans* "

498   <u>Biologia plantarum</u> **50**(2): 265-271.

499   Evers, D., C. Schweitzer, et al. (2006). "Two PR-1 loci detected in the native cultivated

500   potato Solanum phureja

501   appear differentially expressed upon challenge by late blight " <u>Physiological and</u>

502   <u>Molecular Plant Pathology</u> **67**: 155-163.

503   Evers, D., C. Schweitzer, et al. (2006). "Two PR-1 loci detected in the native cultivated

504   potato Solanum phureja appear differentially expressed upon challenge by late

505   blight." <u>Physiological and Molecular Plant Pathology</u> **67**(3-5): 155-163.

506   Ewing, E. E., I. Simko, et al. (2000). "Genetic mapping from field tests of qualitative and

507   quantitative resistance to *Phytophthora infestans* in a population derived from

508   *Solanum tuberosum* and *Solanum berthaultii*." <u>Molecular Breeding</u> **6**: 25-36.

509   FAO. (2007). "FAO agricultural statistics."   Retrieved September 9th, 2007, from

510   <u>http://faostat.fao.org/default.aspx</u>.

511   Forbes, G. A., X. C. Escobar, et al. (1997). "Population genetic structure of *Phyophthora*

512   *infestans* in Ecuador." <u>Phytopathology</u> **87**: 375-380.

513   Ghislain, M., D. Andrade, et al. (2006). "Genetic analysis of the cultivated potato

514   Solanum tuberosum L. Phureja Group using RAPDs and nuclear SSRs." <u>Theor</u>

515   <u>Appl Genet</u> **113**(8): 1515-27.

516   Ghislain, M., B. Trognitz, et al. (2001). "Genetic loci associated with field resistance to

517   late blight in offspring of *Solanum phureja* and *S. tuberosum* grown under short-

518   day conditions." <u>Theor. Appl. Genet.</u> **103**: 433-442.

519   Griffith, G. W. and D. S. Shaw (1998). "Polymorphisms in *Phytophthora infestans* : four

520   mitochondrial haplotypes are detected after PCR amplification of DNA from pure

521   cultures or from host lesions." <u>Appl. Environ. Microbiol.</u> **64**(10): 4007-4014.

522   Huang, X. and A. Madan (1999). "CAP3: A DNA sequence assembly program." <u>Genome</u>

523   <u>Res</u> **9**(9): 868-77.

524   Jia, Y. and G. B. Martin (1999). "Rapid transcript accumulation of pathogenesis-related

525   genes during an incompatible interaction in bacterial speck disease-resistant

526   tomato plants." <u>Plant Mol Biol</u> **40**(3): 455-65.

527   Keime, C., F. Damiola, et al. (2004). "Identitag, a relational database for SAGE tag

528   identification and interspecies comparison of SAGE libraries." <u>BMC</u>

529   <u>Bioinformatics</u> **5**: 143.

530   Kumar, D. and D. F. Klessig (2003). "High-affinity salicylic acid-binding protein 2 is

531   required for plant innate immunity and has salicylic acid-stimulated lipase

532   activity." <u>Proc Natl Acad Sci U S A</u> **100**(26): 16101-6.

533   Leonards-Schippers, C., W. Gieffers, et al. (1994). "Quantitative resistance to

534   Phytophthora infestans in potato: a case study for QTL mapping in an allogamous

535   plant species." <u>Genetics</u> **137**(1): 67-77.

536   Liu, D., K. G. Raghothama, et al. (1994). "Osmotin overexpression in potato delays
537   development of disease symptoms." Proc Natl Acad Sci U S A **91**(5): 1888-92.

538   Oberhagemann, P., C. Chatot-Balandras, et al. (1999). "A genetic analysis of quantitative
539   resistance to late blight in potato: towards marker-assisted selection." Molecular
540   Breeding(5): 399-415.

541   Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time
542   RT-PCR." Nucleic Acids Research **29**(9).

543   Porter, D. A., I. E. Krop, et al. (2001). "A SAGE (serial analysis of gene expression) view
544   of breast tumor progression." Cancer Res **61**(15): 5697-702.

545   Restrepo, S., K. L. Myers, et al. (2005). "Gene profiling of a compatible interaction
546   between *Phytophthora infestans* and *Solanum tuberosum* suggests a role for
547   carbonic anhydrase." Mol Plant Microbe Interact **18**(9): 913-22.

548   Rooney, H. C., J. W. Van't Klooster, et al. (2005). "Cladosporium Avr2 inhibits tomato
549   Rcr3 protease required for Cf-2-dependent disease resistance." Science
550   **308**(5729): 1783-6.

551   Ryals, J. A., U. H. Neuenschwander, et al. (1996). "Systemic Acquired Resistance." Plant
552   Cell **8**(10): 1809-1819.

553   Sandbrink, J. M., L. T. Colon, et al. (2000). "Two related genotypes of *Solanum
554   microdontu*m carry di erent segregating alleles for field resistance to
555   *Phytophthora infestans*." Molecular Breeding **6**: 215-225.

556   Sander, J., R. T. Ng, et al. (2005). "A Methodology for Analyzing SAGE Libraries for
557   Cancer Profiling." ACM Transactions on Information Systems **23**: 35-60.

558 Shagin, D. A., D. V. Rebrikov, et al. (2002). "A novel method for SNP detection using a
559   new duplex-specific nuclease from crab hepatopancreas." Genome Res **12**(12):
560   1935-42.
561 Shah, J. and D. F. Klessig (1996). "Identification of a salicylic acid-responsive element in
562   the promoter of the tobacco pathogenesis-related beta-1,3-glucanase gene, PR-
563   2d." Plant J **10**(6): 1089-101.
564 Smart, C. D., K. L. Myers, et al. (2003). "Partial resistance of tomato to Phytophthora
565   infestans is not dependent upon ethylene, jasmonic acid, or salicylic acid signaling
566   pathways." Mol Plant Microbe Interact **16**(2): 141-8.
567 Soares, M. B., M. d. F. Bonaldo, et al. (1994). "Construction and characterization of a
568   normalized cDNA library." Proc. Natl. Acad. Sci. USA **91**: 9228-9232.
569 Solomon, M., B. Belenghi, et al. (1999). "The involvement of cysteine proteases and
570   protease inhibitor genes in the regulation of programmed cell death in plants."
571   Plant Cell **11**(3): 431-44.
572 Takemoto, D., K. Furuse, et al. (1997). "Identification of chitinase and osmotin-like
573   protein as actin-binding proteins in suspension-cultured potato cells." Plant Cell
574   Physiol **38**(4): 441-8.
575 Tian, M., E. Huitema, et al. (2004). "A Kazal-like extracellular serine protease inhibitor
576   from Phytophthora infestans targets the tomato pathogenesis-related protease
577   P69B." J Biol Chem **279**(25): 26370-7.
578 Tian, M. and S. Kamoun (2005). "A two disulfide bridge Kazal domain from
579   Phytophthora exhibits stable inhibitory activity against serine proteases of the
580   subtilisin family." BMC Biochem **6**: 15.

581   Tian, M., J. Win, et al. (2007). "A Phytophthora infestans cystatin-like protein targets a

582   novel tomato papain-like apoplastic protease." <u>Plant Physiol</u> **143**(1): 364-77.

583   Toro, R. A. (1927). "Enfermedades y plagas de las plantas." <u>Esc. Agric. Vet. Medellín</u>

584   **Circ. No 4**.

585   Trognitz, F., P. Manosalva, et al. (2002). "Plant defense genes associated with

586   quantitative resistance to potato late blight in Solanum phureja x dihaploid S.

587   tuberosum hybrids." <u>Mol Plant Microbe Interact</u> **15**(6): 587-97.

588   Trognitz, F., P. Manosalva, et al. (2002). "Plant defense genes associated with

589   quantitative resistance to potato late blight in *Solanum phureja* x dihaploid *S.*

590   *tuberosum* hybrids." <u>Molecular Plant Microbe Interactions</u> **15**(6): 587-597.

591   Vargas, A. M., L. M. Quesada-Ocampo, et al. (2008). "Characterization of *Phytophthora*

592   *infestans* Populations in Colombia: First Report of the A2 Mating Type."

593   Vesculescu, V. E., L. Zhang, et al. (1995). "Serial analysis of gene expression." <u>Science</u>

594   **270**: 484-487.

595   Vleeshouwers, V. G. A. A., W. Van Dooijeweert, et al. (2000). "Does basal PR gene

596   expression in Solanum species contribute to non-specific resistance to

597   *Phytophthora infestans* ?" <u>Physiological and Molecular Plant Pathology</u> **57**(1):

598   35-42.

599   White, T. J., T. Bruns, et al. (1990). Amplification and direct sequencing of fungal

600   ribosomal RNA genes for phylogenetics. <u>PCR Protocols</u>. M. A. Innis, D. H.

601   Gelfand, J. J. Sninsky and T. J. White. San Diego, Academic Press.

602   Zdobnov, E. M. and R. Apweiler (2001). "InterProScan--an integration platform for the

603   signature-recognition methods in InterPro." <u>Bioinformatics</u> **17**(9): 847-8.

604   Zhu, B., T. H. Chen, et al. (1995). "Activation of two osmotin-like protein genes by

605   abiotic stimuli and fungal pathogen in transgenic potato plants." Plant Physiol

606   **108**(3): 929-37.

607   Zhu, Y. Y., E. M. Machleder, et al. (2001). "Reverse transcriptase template switching: a

608   SMART approach for full-length cDNA library construction." Biotechniques

609   **30**(4): 892-7.

610   Zhulidov, P. A., E. A. Bogdanova, et al. (2004). "Simple cDNA normalization using

611   kamchatka crab duplex-specific nuclease." Nucleic Acids Res **32**(3): e37.

612

613

614

615 **Table 1**: Primer sequences used for real-time quantitative PCR

Gene Name	Primer Sequence (5' to 3')
Elongation factor 1-alpha	Fwd ACC ACT GGT GGT TTT GAA GC
	Rev ACG ACC AAC AGG GAC AGT TC
Pathogenesis related protein isoform b1	Fwd TCA CTC TTG TGA TGC CCA AA
	Revev AAT GAA CCA CCA TCC GTT GT
Endo-1,3-beta-D-glucanase	Fwd GAA GCT GGT TTG GGA AAT GA
	Rev CCA TCG CAG CAT AAA CAG AA
Cysteine protease inhibitor 1 precursor	Fwd CCC GTC AAG TTC CTT CGT AA
	Rev CTT TAA AGT TGC CGC CGA TA
Carbonic Anhydrase	Fwd CCA CAG TGC TTG TGG AGG TA
	Rev CAA GTC CCC ACA GCT CAA AT
Catalase isozyme 2	Fwd GCG GTT TTG CTG TCA AGT TT
	Rev TGT GCT TTC CCC TCT TTG TT
Ethylene-forming enzyme	Fwd GCA ATG GAT CGA TGT TCC TC
	Rev ATC TTG GCT CTT TGG CTT GA
Double WRKY type transfactor	Fwd GCA GAG AGG ACT GGT TCT GG
	Rev GCA GCT GTC CCT ACT TGA GG
salicylic acid-binding protein 2	Fwd ATG CCT TCC ACA AGA GGA GA
	Rev CCT CAG GAG GGC ATA ACT GA
Pectate lyase	Fwd TGA TTG ATG CAA TTC GTG GT
	Rev CAC TTT CTG CTG CAT CCT CA
Glutathione S-transferase	Fwd GGC AGC AGT GTG GAA AAG TT
	Rev GCT CGG TAA CGG ATA AGC AA
Kunitz-type protease inhibitor	Fwd GGG GGA AAA CAC AAG CAA TA
	Rev AAT ACG CCA TCT GGA CAA GG
Light harvesting chlorophyll a /b binding protein of PSII	Fwd TCT GCC AGA TTC ACC ATG AG
	Rev AAG ATT TGG GCT CCA GCT TT
Phenylalanine ammonia-lyase; n=1	Fwd AGC TTT CGA GGA CGA ATT GA
	Rev GCA GAT TGG AAG AGG AGC AC

616

617

618 **Table 2:** Most represented protein domains in the *Solanum phureja* EST, TCs, and

619 singletons

ID	GO description	Number of sequences
IPR002048	Calcium-binding EF-hand	179
IPR001680	WD40 repeat	111
IPR001128	Cytochrome P450	90
IPR000719	Protein kinase, core	73
IPR016040	NAD(P)-binding	68
IPR000425	Major intrinsic protein	62
IPR000308	14-3-3 protein	61
IPR000626	Ubiquitin	61
IPR000504	RNA recognition motif, RNP-1	54
IPR002130	Peptidyl-prolylcis-transisomerase, cyclophilin-type	48
IPR000164	Histone H3	47
IPR001993	Mitochondrial substrate carrier	44
IPR002198	Short-chain dehydrogenase/reductase SDR	42
IPR011009	Protein kinase-like	42
IPR002119	Histone H2A	41
IPR001806	RasGTPase	39
IPR001623	Heat shock protein DnaJ, N-terminal	38
IPR002401	Cytochrome P450, E-class, group I	38
IPR002016	Haemperoxidase, plant/fungal/bacterial	37
IPR002213	UDP-glucuronosyl/UDP-glucosyltransferase	37
IPR000608	Ubiquitin-conjugating enzyme, E2	36
IPR001848	Ribosomal protein S10	36
IPR001245	Tyrosine protein kinase	35

IPR002347	Glucose/ribitoldehydrogenase	35
IPR001878	Zinc finger, CCHC-type	34
IPR006689	ARF/SAR superfamily	32
IPR012336	Thioredoxin-like fold	32
IPR000823	Plant peroxidase	31
IPR001344	Chlorophyll A-B binding protein	30
IPR009072	Histone-fold	30

620

621

622 **Table 3:** Characteristics of the six SAGE libraries (R= resistant cultivar Col3; S=

623 susceptible cultivar Col2; 0, 12 and 72 hours post inoculation).

Library	No. of unique tags	No. of total tags	No. of singletons	Average (copies/unique tag)
0H_R	2710061	9254605	2425492	3.4149
0H_S	1342555	6521400	1182069	4.8575
12H_R	1584586	7287128	1384990	4.5988
12H_S	1084391	4941366	943921	4.5568
72H_R	1140741	4919073	991243	4.3122
72H_S	1225281	5166558	1064863	4.2166

624 **Table 4:** Genes up regulated and down regulated for different libraries comparisons ($p < 0.01$)

Pairwise Comparison Regulation[a]	Gene [b]	Hit against *S. phureja* ESTs	Hit against TIGR TA
0R-72R Up-regulated	Osmotin-like protein	-	TA24474_4113
	Endochitinase	-	TA28339_4113
	Cysteine protease inhibitor 1 precursor [*Solanum tuberosum* (Potato)]	Contig49	-
	Class I chitinase	-	TA24202_4113
	Pathogenesis related protein isoform b1 precursor [*Solanum phureja*]	Contig52	-
	Salicylic acid-binding protein 2 [*Nicotiana tabacum*]	-	TA29339_4113
	Carbonic Anhydrase	Contig187	-
0R-72R down-regulated	Asparagine synthetase [glutamine-hydrolyzing] [*Triphysaria versicolor* (Yellow owl's clover)]	Contig104	-
	Catalase isozyme 2 [*Solanum tuberosum* (Potato)]	-	TA23780_4113
0R-12R Up-regulated	Dof zinc finger protein DOF5.5 (AtDOF5.5) [*Arabidopsis thaliana* (thale cress)]	-	-
	Endochitinase 2 precursor [*Solanum tuberosum* (Potato)]. *Solanum tuberosum* chitinase (chtB3)	CSRLCD001_16-H01-PAL17d	-
	Ethylene-forming enzyme [*Lycopersicon esculentum* (Tomato)]	-	TA25537_4113
	Trypsin proteinase inhibitor [*Nicotiana rustica* (Aztec tobacco)]	-	TA41208_4113
0R-12R down-regulated	Double WRKY type transfactor - Cluster: WRKY-like transcription factor; n=1; *Solanum peruvianum*|Rep: WRKY-like transcription factor - *Solanum peruvianum* (Peruvian tomato) (*Lycopersicon peruvianum*)	-	TA25252_4113
	MAPKK - MAPKK [*Lycopersicon esculentum* (Tomato)]	-	TA29661_4113

12R-72R up-regulated	14-3-3 protein [*Solanum tuberosum* (Potato)]	-	TA25421_4113
	Peroxidase precursor [*Euphorbia characias* (Spurge)]	-	-
	Chalcone synthase 1B [*Solanum tuberosum* (Potato)]	-	TA30979_4113
	Catalase isozyme 1 *Solanum lycopersicum* (Tomato) (*Lycopersicon esculentum*)	-	TA23780_4113
	Cytosolic ascorbate peroxidase 1; *Solanum lycopersicum* (Tomato) (*Lycopersicon esculentum*)	-	TA24254_4113
	Zinc finger protein CONSTANS-LIKE 13; *Arabidopsis thaliana* (Mouse-ear cress)	-	-
12R-72R down-regulated	Salicylic acid-binding protein 2 [*Nicotiana tabacum*]	-	TA29339_4113
	Cytochrome P450 71D6 [*Solanum chacoense* (Chaco potato)]	-	BQ517565 [S. CHACOENSE]
	Pectate lyase [*Gossypium hirsutum* (Upland cotton)]	-	TA33277_4113
0S-72S up-regulated	Chalcone synthase 2 [*Solanum tuberosum* (Potato)]	-	CF802508
	Glucan endo-1	CSRLCD001_17-G04-PAL17d	-
	Peroxidase [*Lycopersicon esculentum* (Tomato)]	CSRLCD001_18-C11-PAL17d	-
	Probable glutathione S-transferase [*Solanum tuberosum* (Potato)]	CSRLCD001_02-A09-pAL17d	-
0S-72S down-regulated	Superoxide dismutase [*Lycopersicon esculentum* (Tomato)]	-	TA24918_4113
	Kunitz-type protease inhibitor [*Solanum tuberosum* (Potato)]	CSRLCD001_09-D09-pAL17d	
0S-12S up-regulated	Ascorbate peroxidase [*Solanum tuberosum* (Potato)]	-	TA24256_4113
	Basic PR-1 protein precursor [*Capsicum annuum* (Bell pepper)]	-	TA25631_4113
	Catalase isozyme 2 [*Lycopersicon esculentum* (Tomato)]	-	TA25243_4113

	CBL-interacting serine/threonine-protein kinase 15 (SOS2-like protein kinase PKS3) (SNF1-related kinase 3.1) (SOS-interacting protein 2) (Serine/threonine-protein kinase ATPK10) [*Arabidopsis thaliana* (thale cress)]	CSRLCD001_08-B09-pAL17d	-
	Ethylene-responsive proteinase inhibitor 1 precursor [*Lycopersicon esculentum* (Tomato)]	-	TA33995_4113
	Wound-inducible proteinase inhibitor I [*Solanum tuberosum* (Potato)]	-	*S. lycopersicum*
0S-12S down-regulated	Kunitz-type protease inhibitor [*Solanum tuberosum* (Potato)]	-	TA35635_4113
	Light harvesting chlorophyll a /b binding protein of PSII [*Euglena gracilis*]	-	TA23048_4113
12S-72S up-regulated	Catalase isozyme 2 [*Lycopersicon esculentum* (Tomato)]	-	TA25243_4113
	Patatin [*Solanum tuberosum* (Potato)]	-	TA23325_4113
12R-12S up-regulated	Phenylalanine ammonia-lyase; *Capsicum chinense* (Scotch bonnet) (Bonnet pepper)	-	TA24485_4113
	Endochitinase 2 precursor [*Solanum tuberosum* (Potato)]	CSRLCD001_16-H01-PAL17d	-
72R-72S up-regulated	Ethylene-responsive proteinase inhibitor 1 precursor [*Lycopersicon esculentum* (Tomato)].		TA33995_4113
72R-72S down-regulated	Carbonic anhydrase [*Lycopersicon esculentum* (Tomato)]	-	TA24302_4113
	Light harvesting chlorophyll a /b binding protein of PSII [*Euglena gracilis*]	-	TA23445_4113

625   [a] Pairwise comparison results correspond to the second SAGE library relative to the first shown in the column.

626   [b] Genes are presented based on a Blastn result of e=0 and p=0.

627

628

629 **Figure 1**: Distribution of *Solanum phureja* ESTs. EST redundancy among the 1825 TCs

630 and singletons obtained from a normalized cDNA library of the yellow potato *Solanum*
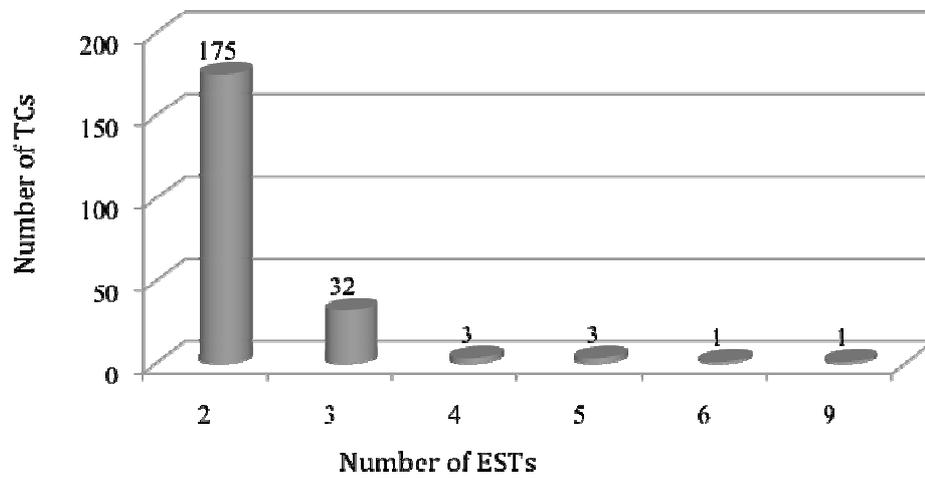
631 *phureja*

632

633 **Figure 2:** Molecular function categories of *Solanum phureja* sequences. Distribution of

634 the 916 TCs and singletons having a significant hit into molecular function class

635 according to the Gene Ontology classification

636

637 **Figure 3**: Number of unique tags in relation to tag copy number for the six libraries (R =

638 resistant cultivar Col3; S = susceptible cultivar Col2; 0, 12 and 72 hours after

639 inoculation). Only tags with a copy number of 2 or higher were plotted on the graph.
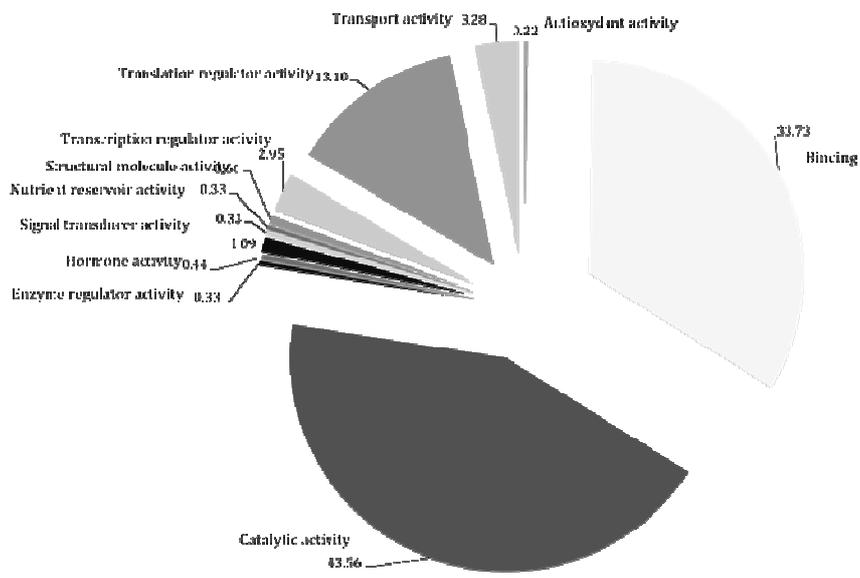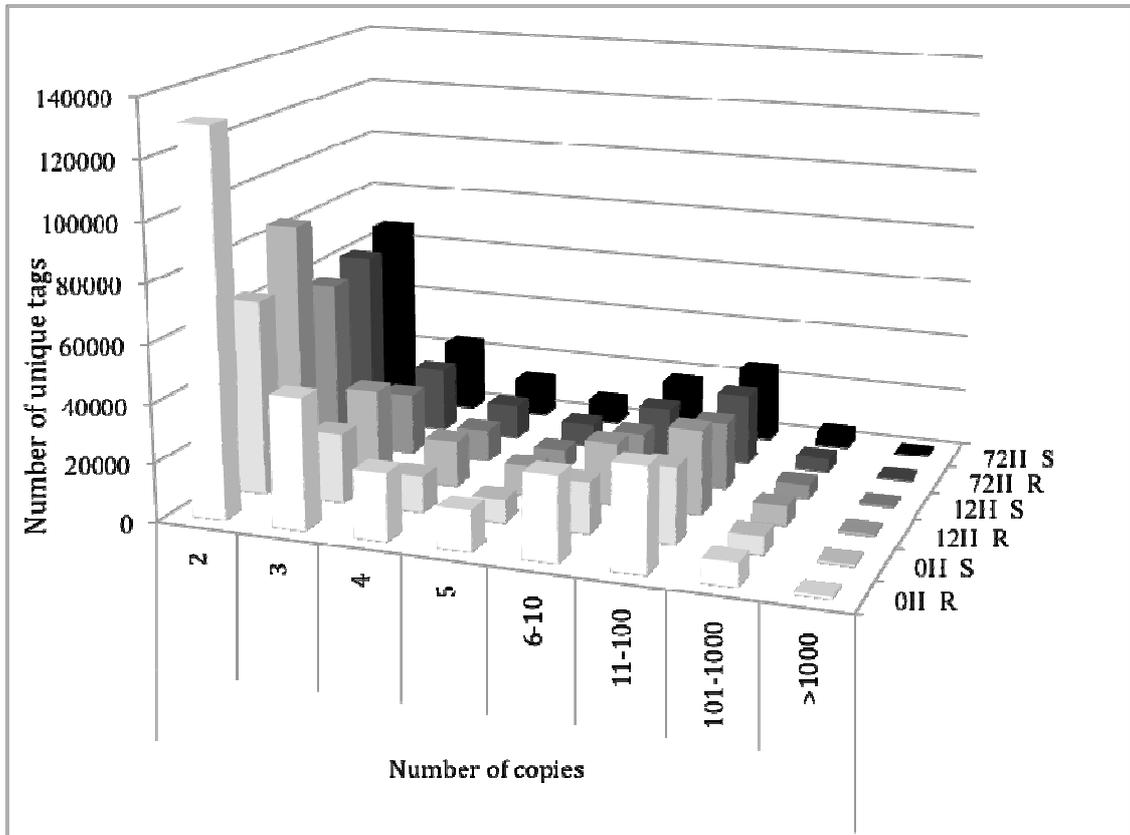
640

641

642

643

644