Genome characterization of '*Candidatus* Phytoplasma meliae' (isolate ChTYXIII)

Franco Daniel Fernández[1,2,*], Luis Rogelio Conci[1,2,]

1. Instituto Nacional de Tecnología Agropecuaria (INTA), Centro de Investigaciones Agropecuarias (CIAP), Instituto de Patología Vegetal (IPAVE). Camino 60 cuadras km 5 ½ (X5020ICA), Córdoba. Argentina

2. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Unidad de Fitopatología y Modelización Agrícola (UFYMA). Camino 60 cuadras km 5 ½ (X5020ICA), Córdoba. Argentina

* **Corresponding author:** Franco Daniel Fernández, e-mail: fernandez.franco@inta.gob.ar, tel. +5493514973636

Keywords: Phytoplasma, genome, syntheny, effector protein, chinaberry, MPV, orthologues

Abstract

'*Candidatus* Phytoplasma meliae' (subgroups 16SrXIII-G and XIII-C) has been reported in association to chinaberry yellowing disease in Argentina, Bolivia and Paraguay. In Argentina, this disease constitutes a major phytosanitary problem for chinaberry forestry production. To date, no genome information of this phytoplasma and others from 16SrXIII-group has been published, hindered its characterization at genomic level. Here we analyze the draft genome of '*Candidatus* Phytoplasma meliae' strain ChTYXIII obtained from a chinaberry-infected plant using a metagenomics approach. The draft assembly consists of twenty-one contigs with a total length of 751.949 bp. The annotation contains 669 CDSs, 34tRNA and one set of rRNA operons. Metabolic pathways analysis indicated that the ChTYXIII contains the complete core genes for glycolysis and functional sec system for translocation of proteins. The phylogenetic relationships inferred 132 single copy genes (orthologues core) analysis revealed that '*Ca.* P. meliae' constitutes a clade closely related to the '*Ca*. australiense' and '*Ca*. P. solani'. Thirty-one putative effectors were identified, among which a homologue to SAP11 was found and others that have only been described in this pathogen. This work provides relevant genomic information for '*Ca.* P. meliae' and constitutes the first genome described for the group 16SrXIII (MPV).

1. Introduction

'*Candidatus* Phytoplasma meliae' have been associated with the yellowing disease of China tree (*Melia azedarach*) (Chinatree yellows, ChTY) in Argentina (Fernández et al., 2016) (subgroup16SrXIII-G), Paraguay (Arneodo et al., 2005) (subgroup16SrXIII-G) and Bolivia (Harrison et al., 2003) (subgroup16SrXIII-C). Chinaberry plants affected by this phytoplasma develop characteristic symptoms, such as reduced leaf size and yellowing, witches'-broom and dieback. China tree mortality caused by '*Ca*. P. meliae' and Chinaberry tree decline phytoplasma (subgroup 16SrIII-B) (Galdeano et al., 2004) constitutes a serious phytosanitary problem, mainly in north-east region of Argentina, where this tree species is grown for furniture manufacturing (Arneodo et al., 2007, Fernández et al., 2015, 2016). The 16SrXIII group (*Mexican periwinkle virescence*, MPV) constitute a monophyletic clade (Figure 1) within twelve subgroups have been described so far (last one 16SrXIII-L) (Bongiorno et al., 2020). An interesting aspect of this group of phytoplasmas is their geographical distribution, which seems to be restricted to the American continent (Fernández et al., 2016; Pérez-López et al., 2016). In addition, only few host species have been associated to MPV group, among them strawberry (Jomantiene et al., 1998; Fernández et al., 2015; Pérez-López E and Dumonceaux 2016; Melo et al., 2018; Cui et al., 2019), potato (Santos-Cervantes et al., 2010), periwinkle (Lee et al., 1998), papaya (Melo et al., 2013), broccoli (Pérez-López et al., 2016) in addition to those already mentioned for '*Ca*. P. meliae' above. Thanks to the reduction in the cost of sequencing in the last ten years, phytoplasma sequencing projects have increased notably. Phytoplasmas have a unique biology and genomic knowledge allowed us to understand aspects of pathogenicity and evolution never before described (Oshima et al., 2004; Chung et al., 2013; Orlovskis et al., 2017;Cho et al., 2020; Huang et al., 2021). Currently, no genomic data have been generated in any of the species that make up the MPV group, which makes it difficult to study the mechanisms associated with pathogenicity, evolution or their dispersion mediated by insects. In the present study, we report the draft genome of '*Ca*. Phytoplasma meliae' strain ChTYXIII-Mo (subgroup 16SrXIII-G) obtained from infected chinaberries in Argentina. The goal of this work was to provide basic genomic information, in order to understand fundamental aspects of the evolution and pathogenicity of this phytoplasma and related phytoplasmas from MPV-clade.

2. Materials and Methods

2.1 Plant samples

"Candidatus Phytoplasma meliae" isolate ChTYXIII-Mo (KU850940) (Fernández et al., 2016) was maintained in greenhouses under controlled conditions and propagated in chinaberry tree plantlets by grafting. Total genomic DNA was extracted from infected midribs using DNeasy Plant Mini Kit (Qiagen, Germany) following the manufacturer instructions. Quality and quantity

93 controls were assayed by electrophoresis in 1% agarose gels and spectrophotometry (Nanodrop-
94 1000).

95 **2.2 Library construction and sequencing**

96 Total DNA was used for construction of paired-end libraries (150bp) according to TruSeq™ DNA
97 Nano protocol and sequenced in Illumina Novaseq platform (Macrogen, Korea). Quality of RAW-
98 reads was checked using FastQC (https://github.com/s-andrews/FastQC) and then were trimmed
99 with Trim Galore! (https://github.com/FelixKrueger/TrimGalore) applying default settings.

100 **2.3 Assembling and annotation**

101 A metagenomic-approach was implemented for assembling based on previous pipelines with
102 some modifications (Music et al., 2019). Trimmed reads were assembled using Unicycler (Wick
103 et al., 2017) (Bridging mode:normal, Spades correction and Pilon polish). Assembled contigs
104 belonging to phytoplasmas were identified by BLASTx (E=1e-20, word size=11) against a local
105 database constructed using the 34 phytoplasma genome sequences available from NCBI
106 (txid33926). Trimmed reads were mapped using Bowtie2 v2.3.4.3 (defaults parameters)
107 (Langmead et al., 2012) to phytoplasma-assigned contigs. An iterative process was used until the
108 assembly was completed. Completeness of draft-assembly was estimated by CheckM v1.0.18
109 (Park et al., 2015). The final draft genome was annotated using the NCBI Prokaryotic Genome
110 Annotation Pipeline (Tatusova et al., 2016). KAAS-KEGG Automatic Annotation Server
111 (https://www.genome.jp/kegg/kaas/) was used for functional characterization of protein coding
112 regions and reconstruction of metabolic pathways.

113 **2.4 Identification of putative effector proteins**

114 Putative effector proteins were identified based in previous pipelines (Bai et al., 2006, Fernández
115 et al., 2019, Music et al., 2019). Prediction of signal-peptide was conducted in Signal IP v5.0
116 server (https://services.healthtech.dtu.dk/service.php?SignalP-5.0) and Music et al. (2019)
117 criteria was implemented in order to define positive candidates. Proteins which passed this filter
118 (peptide-signal+) were analyzed in TMHMM - 2.0 server
119 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) and those protein without any
120 transmembrane helices domain after signal-peptide were selected (putative secreted proteins-
121 PSP). PSP were analyzed in the Conserved Domains Database search tool
122 (www.ncbi.nlm.nih.gov/ Structure/cdd/wrpsb.cgi, expect value = 0.01) and nuclear signal
123 prediction was acceded using NLStradamus
124 (http://www.moseslab.csb.utoronto.ca/NLStradamus/). The subcellular localization was accessed
125 using the LOCALIZER (http://localizer.csiro.au/) and ApoplastP (http://apoplastp.csiro.au/)
126 servers. The final set of proteins [signal peptide (+); trans- membrane domains outside SP (−)]

were analyzed by reciprocal BLASTp searches (E-value≤1e-05) against aster yellows witches'-broom (AYWB) phytoplasma proteins (taxid:322098) for identification of SAPs homologs (Bai et al. 2009).

2.5 Orthologues clustering and phylogenetic analyses

Identifications of orthologous protein clusters were conducted using Orthofinder v2.5.2 (https://github.com/davidemms/OrthoFinder). Genomes sequences of representative '*Ca*. Phytoplasmas species were retrieved for Genbank (Table S1). For phylogenetic analyses, alignment of concatenated nucleotide sequences of single-copy core genes or single genes were constructed with MAFFT v7.450 using Geneious R.10 software (Biomatters Ltd., Auckland, New Zealand). Phylogenetic trees were constructed with IQ-TREE (http://www.iqtree.org/) (substitution model: automatic, ultrafast bootstrap=1000).

3. Results and discussion

3.1 Assembly and key features of *Ca*. Phytoplasma meliae draft genome

The genome sequencing of '*Ca*. Phytoplasma meliae' (strain ChTYXIII) generated ~ 3.5Gbp of RAW reads (NCBI accession: PRJNA530090) providing a ~97X-fold coverage of draft genome, representing 21 assembled contigs totaling 751.949 bp (27.31% G+C) (Table 1). Since there are no previous estimations of chromosome size for any phytoplasmas belonging to the 16SrXIII group, we used CheckM software to evaluate the assembly quality based on the presence of conserved marker genes. According to this software, the completeness of this draft was 97.34% and possible contamination of 3.29%. Since, as has been discussed in previous works (Music et al., 2019), estimations are based on a small number of marker genes and there is little representation of phytoplasma genomes in current databases, these estimations have to be considered with caution. The 21 contigs composing the draft assembly ranged between 1.832 bp to 137693 bp (N50= 53.850). In the annotation process, 669 CDSs (full-length coding sequences) were identified, with 472 annotated as proteins with assigned function and 197 as hypothetical proteins, one operon for rRNA genes and 34 tRNAs (Table 1). Functional annotation of CDSs using BlastKOALA (https://www.kegg.jp/blastkoala/), assigned 387 of 669 CDSs (~58%) to orthologues in the KEEG database. From 299 KO categories, 240 were described with only one gene while the remaining 59 presented more than one copies representing 147 genes, ~ 21% of total CDss (49 KO with 2 genes, 4 KOs with 3 genes, 3 KOs with 4 genes, 2 KOs with 5 genes each, and a single KO to which 15 different genes were assigned). The proportion of multicopy genes in the '*Ca*. P. melia' ChTYXIII genome (~ 21%) was higher than the observed in other related phytoplasmas as, '*Ca*. P. solani' SA-1 (18.5%) (Music et al., 2019), '*Ca*. P. asteris' strain AY-WB (10.2%), (Bai et al., 2006), OY-M (14,1%) (Oshima et al., 2004) or 'Ca. P. australiense' PAa (12.1%) (Tran-Nguyen et al.,2008). Multicopy genes in PMU-like regions and genome size

appear to be positively correlated to a broad host range in phytoplasmas (Music et al., 2019). This is in contrast to the fact that '*Ca*. P. meliae' has been only associated with two hosts, chinaberry (Harrison et al., 2003; Fernández et al., 2016) and plum (Bongiorno et al., 2020). However, the total number of hosts may have been underestimated, since there is little information regarding the presence of this phytoplasma in native species, i.e. weeds, and vector insects remain unknown.

3.2 Metabolic pathways

Within the three-major protein families in KEEG database, 184 CDS were assigned to *Metabolism*, 236 CDS to *Genetic Information Processing* and 54 were assigned to *Signaling and Cellular Processes* (Figure 2). Transporter membrane proteins plays fundamental roles in the phytoplasmas metabolism, allowing the incorporation of metabolites and contributing to protein secretion in the host cells cytoplasm. The ATP-binding cassette (ABC) transporters form one of the largest known protein families, and are widespread in bacteria, archaea, and eukaryotes. These proteins are best known for their role in the importation of essential nutrients and the export of toxic molecules, but they can also mediate the transport of many other physiological substrates (Davidson et al, 2008). Twenty eight genes from '*Ca*. P. meliae 'genome have been described as the ABC-transporters (Table S2), including the complete pathway for spermidine/putrescine transport (potA, potB, potC and potD), lysine transport (lysX1, lysX2 and lysXY) and Zinc/Manganse/Iron (II) transport (troA, troC, troD, troB). Regarding to the protein translocation system (Sec system) we identified, *secA* (CHTY_001675), secE (CHTY_0002195), *secY* (CHTY_000200), *ffh* (CHTY_001830), *ftsY* (CHTY_001825) and *yidC* (CHTY_000350), dnaJ (CHTY_000550), dnaK (CHTY_000555) , grpE (CHTY_000560) and groEL (CHTY_001355) which suggests a functional sec systems in the '*Ca*. P. meliae'. Within the carbohydrate metabolism, the core module for glycolysis (genes *gapA*, *pyk*, *pgk*, *eno*, *tpiA* and *gpmI*) and pyruvate oxidation (genes *pdhA*, *pdhB*, *pdhD* and *aceF*) were found, which supports that '*Ca*. P. meliae' could depends on glycolysis for energy generation. This pathway was described as the major energy-yielding pathway for phytoplasmas (Kube et al., 2012). In addition, the complete ORF for gene sucrose phosphorylase (*gtfA*, CHTY_000510) was identified.

3.4 Putative effectors

In the draft genome of '*Ca*. P. meliae' we identified 31 proteins as putative secreted proteins (Table S3). BLASTp search against SAP-protein repertoire of AYWB phytoplasma (accession CP000061) showed the presence of putative orthologs for SAP72 (53.31%), SAP41 (52.29%), SAP21 (38.10%), SAP68 (76.00%), SAP67 (59.67%), SAP21 (38.10%) and SAP11 (50.00%). However, no homologues for SAP54, SAP05 and TENGU factor were identified. The SAP11 homologue consisted of 116 aa (CHTY_003225) with a predicted signal peptide domain (score=

0.98, position 1-32aa) and characteristic SMV signal (pfam12113, position 1-33aa/ E= 5.71e-08). Moreover, nuclear signal localization (position 35-49aa) and coiled-coiled (position 84-106aa) domains were predicted. These features are compatible with those described for the SAP11 homologues in diverse phytoplasma taxons (Bai et al., 2009; Kakisawa et al., 2014; Anabestani et al., 2017; Saccardo et al., 2012; Wang et al., 2018). A phylogenetic analysis based on amino acidic sequence for SAP11-homologues grouped the putative SAP11 protein from '*Ca.* P. meliae' within the SAP11 from '*Ca.* P. mali' despite they are evolutionary distant taxons (Figure 3.a). We also found a conserved synthenia in the genomic context of SAP11 with homologous regions in the chromosome of '*Ca.* P. asteris' AYWB and '*Ca.* P. ziziphi' JWB (Figure 3.b) suggesting possible horizontal transfer. *A. thaliana* transgenic lines that express SAP11 have curly leaves and an increased number of axillary stems that resemble the witches' broom symptoms exhibited by AY-WB phytoplasma (Sugio et al., 2014). In greenhouse chinaberry plantlets infected by '*Ca.* P. melia' shows typical witches' broom symptoms while those infected with '*Ca.* P. pruni' strain ChTDII (16SrIII-B) causes symptoms of yellowing and shortening of internodes but not witch's broom (Figure S1). These could suggest the presence of SAP11 homologue associated mechanism in the generation of witches' broom symptoms since no SAP11 homologues were described for ChTDIII phytoplasma (Fernández et al, 2020). A BLASTp search showed that 10 of 31 putative secreted proteins seem to be unique to '*Ca.* P. meliae' since no homologs were found in the gene databank. For example, the putative protein (CHTY_003505, 226aa) present a domain associated to the TIGR04141 family, which is commonly associated to mobilomes, Others secreted protein with interesting characteristics are the CHTY_001115 (115aa), which presented a domain related to multidrug resistance efflux pump (cl34307) or CHTY_002595 (227aa) with a peptidase related domain. The protein CHTY_002535 (391aa) has a chloroplast predicted localization (score= 0.991, position 105-125) and present a substrate-binding domain of an ABC-type nickel/oligopeptide-like import system (cl01709). These proteins constitute an interesting target of study as they can provide clues regarding the unique pathogenicity mechanisms of this pathogen.

3.5 Phylogenetic analyses

Comparative genomics based on orthologues identification between draft genomes of '*Ca.* P. meliae' ChTYXIII and representative genome sequence of 11 '*Ca.* Phytoplasma species' (Table S1) and *A. palmae* (accession FO681347) reveals the presence of 132 single-copy genes common to all species. The phylogentic tree based on DNA and aa concatenated sequence of 132 single-copy genes showed that ChTYXIII phytoplasma form a particular clade which is more closely related to the phytoplasmas of the group 16SrXII , '*Ca.* P. solani strain SA-1, 284 , 231 and *Ca.* P. australiense' strain PAa and SLY (Figure 4). Similar topology was obtained when *secY* and *tufB* genes (single copy orthologs) were used (Figure S2). These results could suggest that the

ChTYXIII phytoplasmas, and the associated species within the 16SrXIII group (Figure 1), could have suffered an evolutionary divergence driven by their distribution restricted to the Americas.

4. Conclusions

This study describes the draft genome of '*Ca.* Phytoplasma meliae' strain ChTYXIII. Functional analyses reveal the presence of genes related to processing and metabolism highly conserved among phytoplasmas. We have described numerous genes present in multicopy despite the fact that it is a phytoplasma with few cited hosts. Effector proteins have also been identified that could be playing key roles in the regulation of pathogenicity mechanisms. Besides SAP11 homologue, other putative effector proteins with interesting characteristics were identified, at least from the bioinformatics analysis of their domains. The phylogeny inferred from a core of genes has shown that '*Ca.* P. meliae' constitutes a clade closely related to '*Ca.* P. solani' and '*Ca.* P. australiense'. Genomic data obtained here provided new insight into the pathogenic mechanisms and evolutionary history in phytoplasmas from MPV-group.

Funding

This research was supported by INTA, FONCyT PICT2016-0862 and PICT2017-3068. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Data availability

RAW reads were deposited in NCBI Sequence Read Archive (SRA) under the accessions PRJNA638346. The de *novo genome* draft assembly of ChTYXIII was deposited in GenBank under the accession JACAOD000000000.2 (BioProject: PRJNA63834, BioSample: SAMN15186628).

References

Anabestani, A., Izadpanah, K., Abbà, S., Galetto, L., Ghorbani, A., Palmano, S., et al. (2017). Identification of putative effector genes and their transcripts in three strains related to '*Candidatus* Phytoplasma aurantifolia.' Microbiol. Res. 199, 57–66. doi:10.1016/j.micres.2017.03.001.

Andersen, M. T., Liefting, L. W., Havukkala, I., and Beever, R. E. (2013). Comparison of the complete genome sequence of two closely related isolates of "Candidatus Phytoplasma australiense" reveals genome plasticity. BMC Genomics 14, 1. doi:10.1186/1471-2164-14-529.

264 Arneodo, J. D., Galdeano, E., Orrego, A., Stauffer, A., Nome, S. F., and Conci, L. R. (2005).
265 Identification of two phytoplasmas detected in China-trees with decline symptoms in
266 Paraguay. *Australas. Plant Pathol.* 34, 583–585. doi:10.1071/AP05064.

267 Arneodo, J. D., Marini, D. C., Galdeano, E., Meneguzzi, N., Bacci, M., Domecq, C., et al. (2007).
268 Diversity and geographical distribution of phytoplasmas infecting China-tree in Argentina.
269 *J. Phytopathol.* 155, 70–75. doi:10.1111/j.1439-0434.2007.01181.x.

270 Bai, X., Correa, V. R., Toruño, T. Y., Ammar, E. D., Kamoun, S., and Hogenhout, S. A. (2009).
271 AY-WB phytoplasma secretes a protein that targets plant cell nuclei. Mol. Plant-Microbe
272 Interact. 22, 18–30. doi:10.1094/MPMI-22-1-0018.

273 Bai, X., Zhang, J., Ewing, A., Miller, S. A., Radek, A. J., Shevchenko, D. V., et al. (2006). Living
274 with genome instability: The adaptation of phytoplasmas to diverse environments of their
275 insect and plant hosts. J. Bacteriol. 188, 3682–3696. doi:10.1128/JB.188.10.3682-
276 3696.2006.

277 Bongiorno, V., Alessio, F., Curzel, V., Nome, C., Fernández, F. D., and Conci, L. R. (2020). '*Ca*
278 . Phytoplasma pruni ' and ' *Ca* . Phytoplasma meliae ' are affecting plum in Argentina.
279 Australasian Plant Dis. Notes 15, 36. https://doi.org/10.1007/s13314-020-00406-8

280 Cho, S. T., Kung, H. J., Huang, W., Hogenhout, S. A., and Kuo, C. H. (2020). Species Boundaries
281 and Molecular Markers for the Classification of 16SrI Phytoplasmas Inferred by Genome
282 Analysis. Front. Microbiol. 11. doi:10.3389/fmicb.2020.01531.

283 Chung, W. C., Chen, L. L., Lo, W. S., Lin, C. P., and Kuo, C. H. (2013). Comparative Analysis
284 of the Peanut Witches'-Broom Phytoplasma Genome Reveals Horizontal Transfer of
285 Potential Mobile Units and Effectors. PLoS One 8, 1–10. doi:10.1371/journal.pone.0062770.

286 Cui, W., Quiroga, N., Curkovic, S. T., Zamorano, A., and Fiore, N. (2019). Detection and
287 identification of 16SrXIII-F and a novel 16SrXIII phytoplasma subgroups associated with
288 strawberry phyllody in Chile. *Eur. J. Plant Pathol.* 155, 1039–1046. doi:10.1007/s10658-
289 019-01808-w.

290 Davidson, A. L., Dassa, E., Orelle, C., and Chen, J. (2008). Structure, Function, and Evolution of
291 Bacterial ATP-Binding Cassette Systems. Microbiol. Mol. Biol. Rev. 72, 317–364.
292 doi:10.1128/mmbr.00031-07.

293 Fernández FD; Zübert C; Huettel B. Kube M, C. L. (2020). Draft Genome Sequence of
294 "*Candidatus* Phytoplasma pruni" (X-disease Group, Subgroup 16SrIII-B) Strain ChTDIII from
295 Argentina. Microbiol. Resour. Announc. 5, 1–2. doi:https://doi.org/10.1128/ MRA.00792-20.

296 Fernández, F. D., Debat, H. J., and Conci, L. R. (2019). Molecular characterization of effector
297 protein SAP54 in Bellis virescence phytoplasma (16SrIII-J). Trop. Plant Pathol.
298 doi:10.1007/s40858-019-00293-0.

299 Fernández, F. D., Galdeano, E., Kornowski, M. V., Arneodo, J. D., and Conci, L. R. (2016).
300 Description of 'Candidatus Phytoplasma meliae', a phytoplasma associated with Chinaberry
301 (Melia azedarach L.) yellowing in South America. Int. J. Syst. Evol. Microbiol. 66, 5244–
302 5251. doi:10.1099/ijsem.0.001503.

303 Fernández, F. D., Meneguzzi, N. G., Guzmán, F. A., Kirschbaum, D. S., Conci, V. C., Nome, C.
304 F., et al. (2015). Detection and identification of a novel 16SrXIII subgroup phytoplasma
305 associated with strawberry red leaf disease in Argentina. *Int. J. Syst. Evol. Microbiol.* 65,
306 2741–2747. doi:10.1099/ijs.0.000276.

307 Fernández, F.D. (2015). Caracterización molecular y epidemiología de fitoplasmas pertenecientes
308 al grupo 16Sr XIII (Mexican periwinkle virescence group; MPV) presentes en la Argentina.
309 Doctoral Thesis, UNC, Córdoba-Argentina. Repositorio Digital de la UNC: Thesis

310 Galdeano, E., Torres, L. E., Meneguzzi, N., Guzmán, F., Gomez, G. G., Docampo, D. M., et al.
311 (2004). Molecular characterization of 16S ribosomal DNA and phylogenetic analysis of two
312 X-disease group phytoplasmas affecting China-tree (*Melia azedarach* L.) and garlic (*Allium*
313 *sativum* L.) in Argentina. J. Phytopathol. 152, 174–181. doi:10.1111/j.1439-
314 0434.2004.00822.x.

315 Harrison, N. A., Boa, E., and Carpio, M. L. (2003). Characterization of phytoplasmas detected in
316 Chinaberry trees with symptoms of leaf yellowing and decline in Bolivia. *Plant Pathol.*
317 doi:10.1046/j.1365-3059.2003.00818.x.

318 Huang, W., Maclean, A. M., Sugio, A., Maqbool, A., Busscher, M., Cho, S.-T., et al. (2021).
319 Parasite co-opts a ubiquitin receptor to induce a plethora of developmental changes. bioRxiv,
320 2021.02.15.430920. Available at: https://doi.org/10.1101/2021.02.15.430920.

321 Jomantiene, R., Davis, R. E., Maas, J., and Dally, E. L. (1998). Classification of new
322 phytoplasmas associated with diseases of strawberry in Florida, based on analysis of 16S

323 rRNA and ribosomal protein gene operon sequences. *Int. J. Syst. Bacteriol.* 48, 269–277.
324 doi:10.1099/00207713-48-1-269.

325 Kakizawa, S., Makino, A., Ishii, Y., Tanaki, H., and Kamagata, Y. (2014). Draft Genome
326 Sequence of "Candidatus Phytoplasma asteris" Strain OY-V, an Unculturable Plant-
327 Pathogenic Bacterium. Genome Announc. 2, 9–10. doi:10.1128/genomeA.00944-14.

328 Kube, M., Mitrovic, J., Duduk, B., Rabus, R., and Seemüller, E. (2012). Current view on
329 phytoplasma genomes and encoded metabolism. Sci. World J. 2012.
330 doi:10.1100/2012/185942.

331 Kube, M., Schneider,B., Kuhl, H., Dandekar, T., Heitmann,K., Migdoll, A., Reinhard, R., and
332 Seemüller, E. (2008). The linear chromosome of the plant-pathogenic mycoplasma
333 "Candidatus Phytoplasma mali." BMC Genomics. doi:doi:10.1186/1471-2164-9-306 This.

334 Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat.*
335 *Methods* 9, 357–359. doi:10.1038/nmeth.1923.

336 Lee, I., Gundersen-rindal, D. E., Davis, R. E., and Bartoszyk, I. M. (1998). Revised classification
337 scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene
338 sequences between distinct groups were 90 % or below . By including additional groups 16S
339 rDNA sequence data were available to predict restriction . *Int. J. Syst. Bacteriol.*, 1153–1169.

340 Melo, L. de A., Ventura, J. A., Costa, H., Kitajima, E. W., Ferreira, J., and Bedendo, I. P. (2018).
341 Delineation of a novel subgroup 16srxiii-j phytoplasma, a 'candidatus phytoplasma
342 hispanicum'-related strain, based on computer-simulated rflp and phylogenetic analysis. *Int.*
343 *J. Syst. Evol. Microbiol.* 68, 962–966. doi:10.1099/ijsem.0.002547.

344 Melo, L., Silva, E., Flôres, D., Ventura, J., Costa, H., and Bedendo, I. (2013). A phytoplasma
345 representative of a new subgroup, 16SrXIII-E, associated with Papaya apical curl necrosis.
346 *Eur. J. Plant Pathol.* 137, 445–450. doi:10.1007/s10658-013-0267-7.

347 Mitrović, J., Siewert, C., Duduk, B., Hecht, J., Mölling, K., Broecker, F., et al. (2014). Generation
348 and analysis of draft sequences of "stolbur" phytoplasma from multiple displacement
349 amplification templates. *J. Mol. Microbiol. Biotechnol.* 24, 1–11. doi:10.1159/000353904.

350 Music, M. S., Samarzija, I., Hogenhout, S. A., Haryono, M., Cho, S. T., and Kuo, C. H. (2019).
351 The genome of 'Candidatus Phytoplasma solani' strain SA-1 is highly dynamic and prone to

352 adopting foreign sequences. *Syst. Appl. Microbiol.* 42, 117–127.
353 doi:10.1016/j.syapm.2018.10.008.

354 Orlovskis, Z., Canale, M. C., Haryono, M., Lopes, J. R. S., Kuo, C. H., and Hogenhout, S. A.
355 (2017). A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma
356 associate with organ proliferation symptoms of infected maize plants. Ann. Bot. 119, 869–
357 884. doi:10.1093/aob/mcw213.

358 Oshima, K., Kakizawa, S., Nishigawa, H., Jung, H. Y., Wei, W., Suzuki, S., et al. (2004).
359 Reductive evolution suggested from the complete genome sequence of a plant-pathogenic
360 phytoplasma. Nat. Genet. 36, 27–29. doi:10.1038/ng1277.

361 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM:
362 Assessing the quality of microbial genomes recovered from isolates, single cells, and
363 metagenomes. Genome Res. 25, 1043–1055. doi:10.1101/gr.186072.114.

364 Pérez-López, E., and Dumonceaux, T. J. (2016). Detection and identification of the heterogeneous
365 novel subgroup 16SrXIII-(A/I)I phytoplasma associated with strawberry green petal disease
366 and Mexican periwinkle virescence. *Int. J. Syst. Evol. Microbiol.* 66, 4406–4415.
367 doi:10.1099/ijsem.0.001365.

368 Pérez-López, E., Luna-Rodríguez, M., Olivier, C. Y., and Dumonceaux, T. J. (2016). The
369 underestimated diversity of phytoplasmas in Latin America. *Int. J. Syst. Evol. Microbiol.* 66,
370 492–513. doi:10.1099/ijsem.0.000726.

371 Saccardo, F., Martini, M., Palmano, S., Ermacora, P., Scortichini, M., Loi, N., et al. (2012).
372 Genome drafts of four phytoplasma strains of the ribosomal group 16SrIII. Microbiol.
373 (United Kingdom) 158, 2805–2814. doi:10.1099/mic.0.061432-0.

374 Santos-Cervantes, M. E., Chávez-Medina, J. A., Acosta-Pardini, J., Flores-Zamora, G. L.,
375 Méndez-Lozano, J., and Leyva-López, N. E. (2010). Genetic diversity and geographical
376 distribution of phytoplasmas associated with potato purple top disease in Mexico. *Plant Dis.*
377 94, 388–395. doi:10.1094/PDIS-94-4-0388.

378 Sugio, A., Maclean, A. M., and Hogenhout, S. A. (2014). The small phytoplasma virulence
379 effector SAP11 contains distinct domains required for nuclear targeting and CIN-TCP
380 binding and destabilization. New Phytol. 202, 838–848. doi:10.1111/nph.12721.

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44, 6614–6624. doi:10.1093/nar/gkw569.

Tran-Nguyen, L. T. T., Kube, M., Schneider, B., Reinhardt, R., and Gibb, K. S. (2008). Comparative genome analysis of "*Candidatus* Phytoplasma australiense" (subgroup tuf-Australia I; rp-A) and "*Ca*. phytoplasma asteris" strains OY-M and AY-WB. J. Bacteriol. 190, 3979–3991. doi:10.1128/JB.01301-07.

Wang, J., Song, L., Jiao, Q., Yang, S., Gao, R., Lu, X., et al. (2018). Comparative genome analysis of jujube witches'-broom Phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. BMC Genomics 19, 1–12. doi:10.1186/s12864-018-5075-1.

Chen, W., Li, Y., Wang, Q., Wang, N., and Wu, Y. (2014). Comparative genome analysis of wheat blue dwarf phytoplasma, an obligate pathogen that causes wheat blue dwarf disease in China. PLoS One 9, 1–11. doi:10.1371/journal.pone.0096436.

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, 1–22. doi:10.1371/journal.pcbi.1005595.

Figure Legends

Figure 1 Phylogeny of phytoplasmas from group MPV based in the analysis of 16Sr RNA sequence. The tree was inferred using the maximum-likelihood method. *Acholeplasma palmae* was used as an outgroup. The numbers on the branches are bootstrap (confidence >70%) values (expressed as percentages of 1000 replicates). The GenBank accession number for each taxon is given between parentheses and the 16Sr group/subgroup classification was also provided. MPV-clade is boxed with a dotted line and '*Ca*. P. meliae' strain sequenced in this study is in bold. R: reference strain .

Figure 2 Bar chart representing the distribution of KEGG pathways associated with the draft genome of '*Candidatus* Phytoplasma meliae' strain ChTYXIII. The KEGG pathways were assigned by annotating the protein coding genes using the KAAS (KEGG Automatic Annotation Server) web server.

Figure 3 Analysis of '*Ca*. P. meliae' SAP11 homologue. A: phylogeny tree inferred from aa sequence of SAP11 homologues using Maximum-Likelihood algorithm (bootstrap 1000) (scale bar: number of substitutions per site. Predicted domains (Nuclear Signal, Signal Peptide) are highlighted with color boxes in the alignment. B: synthenic organization of contig ChTYXIII-17 containing the SAP11 homologue. The genomic localization (start-end) is given below de '*Ca*. P. specie' identification; SAP homologues are in red. Nucleotide sequence similarities between conserved regions are illustrated by different shades of gray colors

Figure 4 Molecular phylogeny based on nucleotide sequences and amino acid of the core genes. *Acholeplasma palmae* was included as an outgroup to root the tree. The numbers on branches indicate the level of bootstrap support (1000 replicates). The scale bar represents the number of substitutions per site. The heatmaps on the right-hand side are colored based on sequence identity.

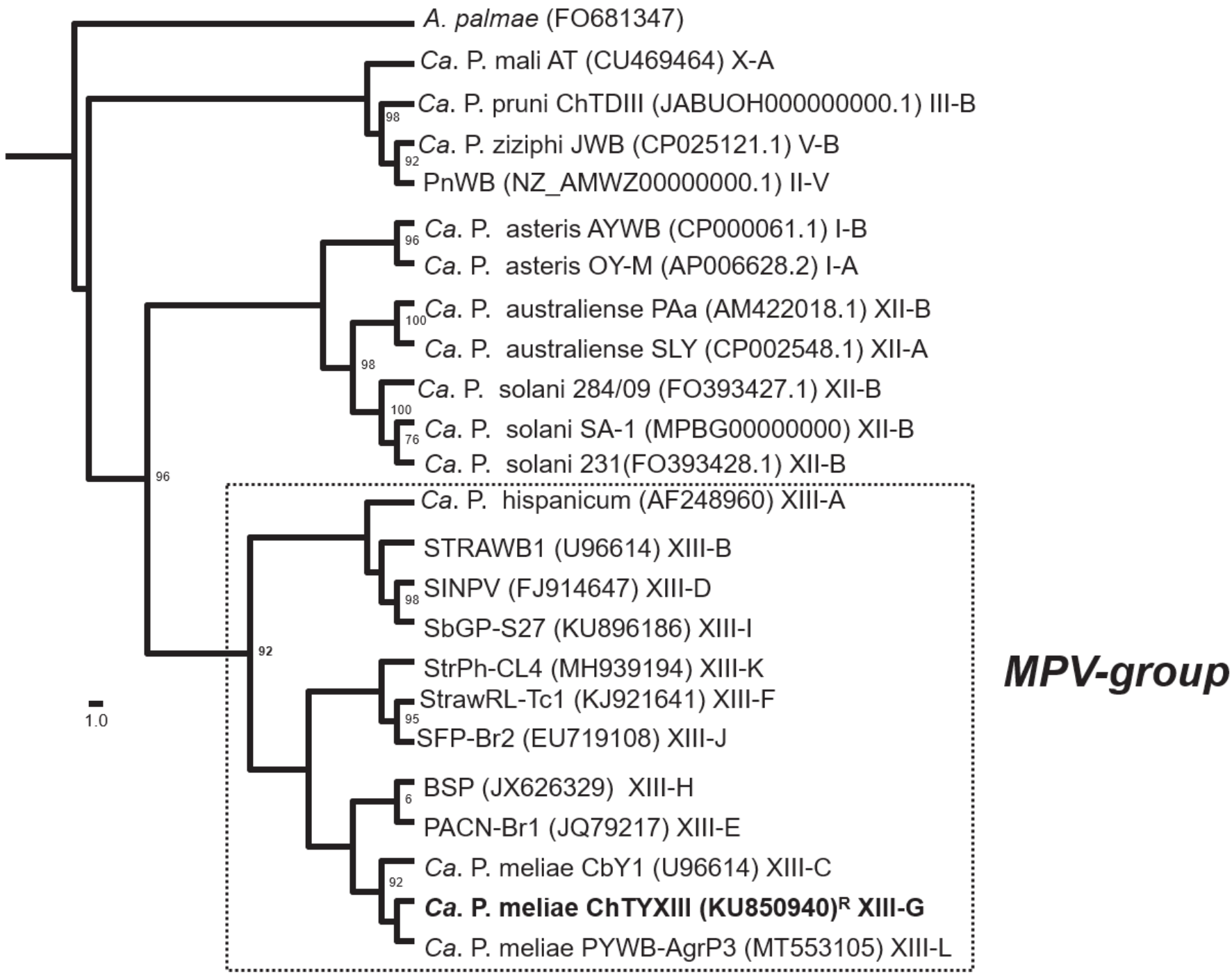
Supplementary Material

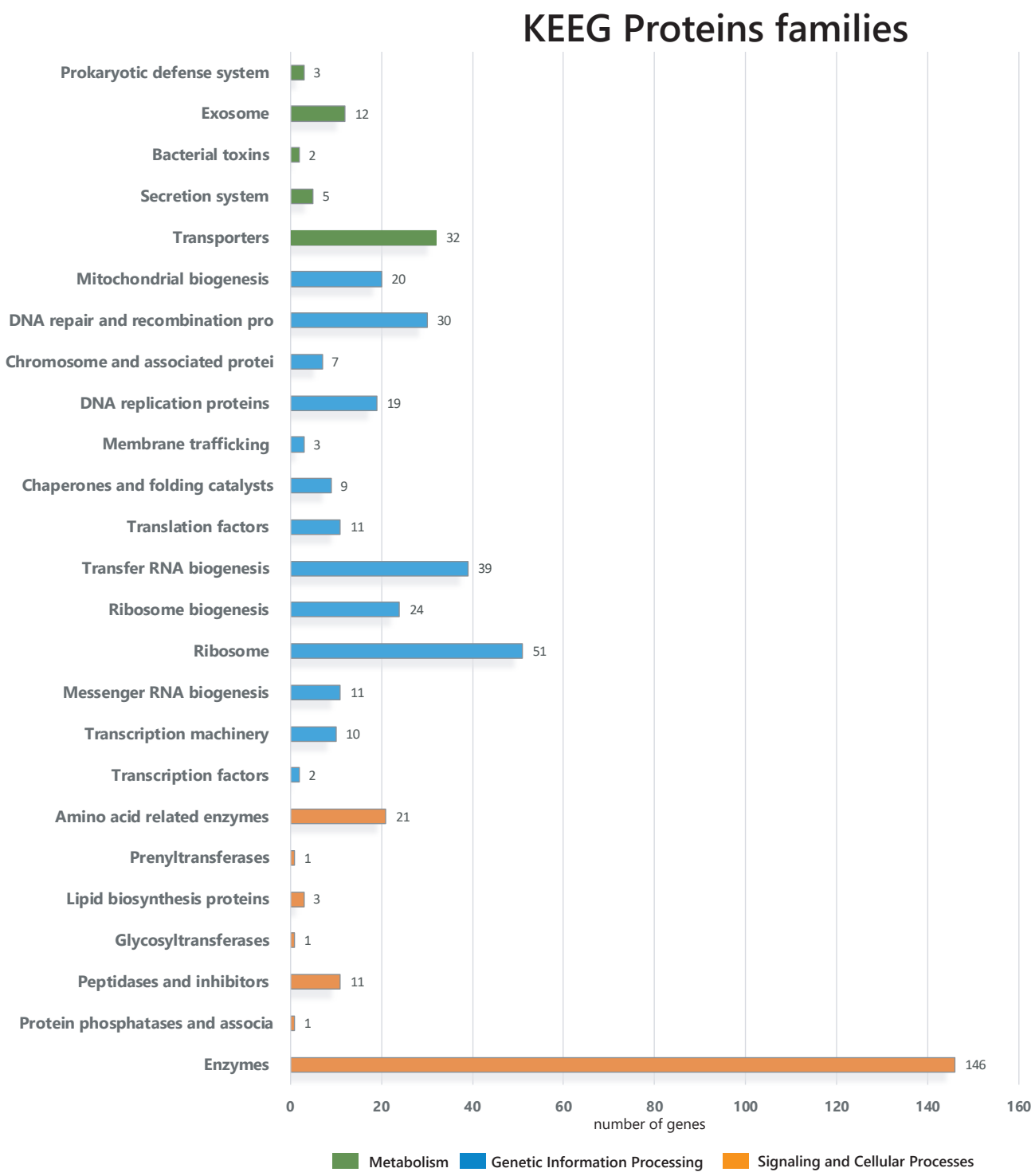
Figure S1 Chinaberry plantlets infected with '*Ca*. P. meliae' (ChTYXIII, 16SrXIII-G) and '*Ca*. P. pruni' (ChTDIII, 16SrIII-B) showing symptoms of witches' broom and leaf yellowing and internode shortening respectively.

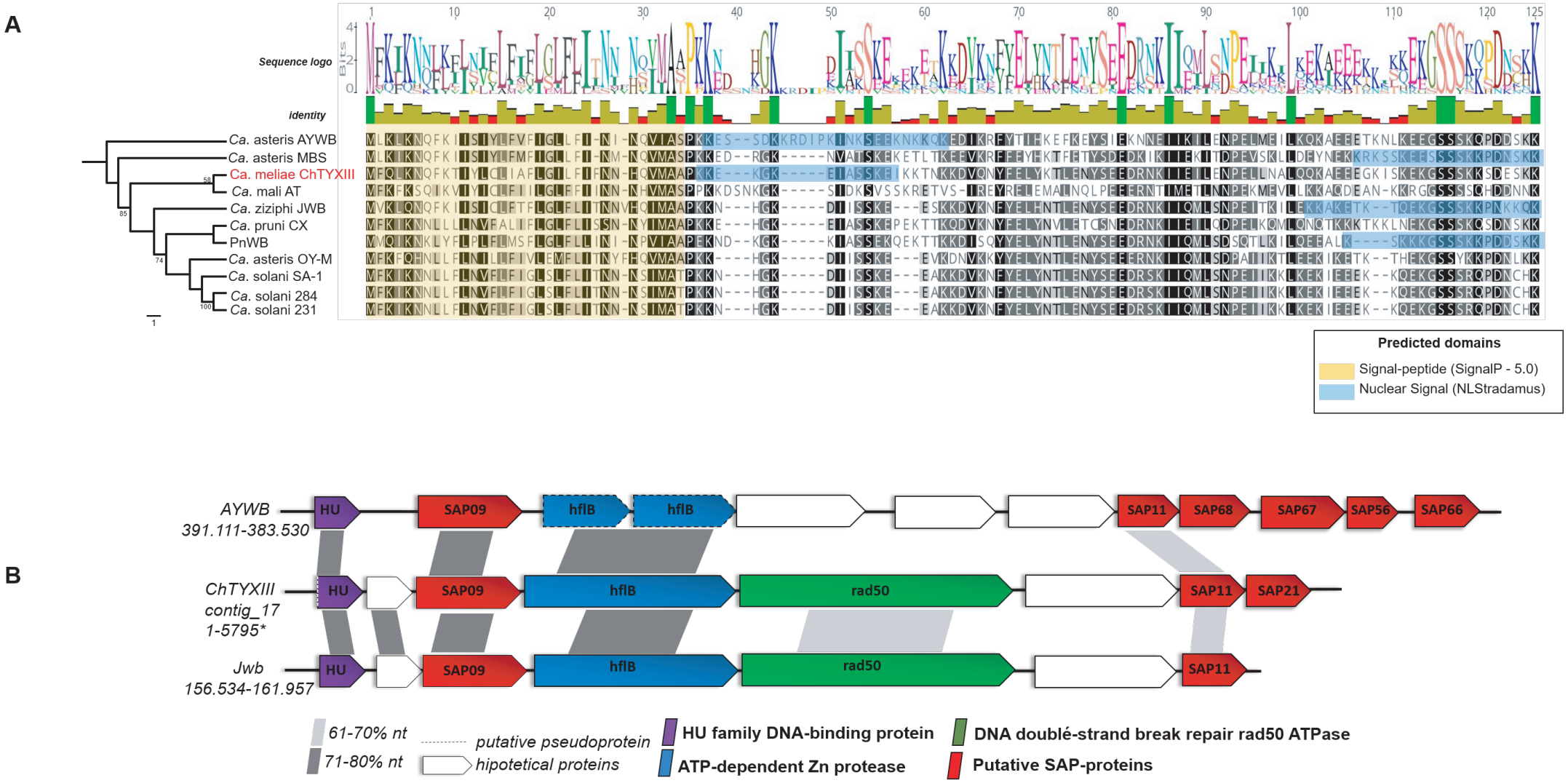
Figure S2 Molecular phylogeny based on nucleotide sequences of *secY* and *tufB* genes. *Acholeplasma palmae* was included as an outgroup to root the tree. The numbers on branches indicate the level of bootstrap support (1000 replicates). The scale bar represents the number of substitutions per site.

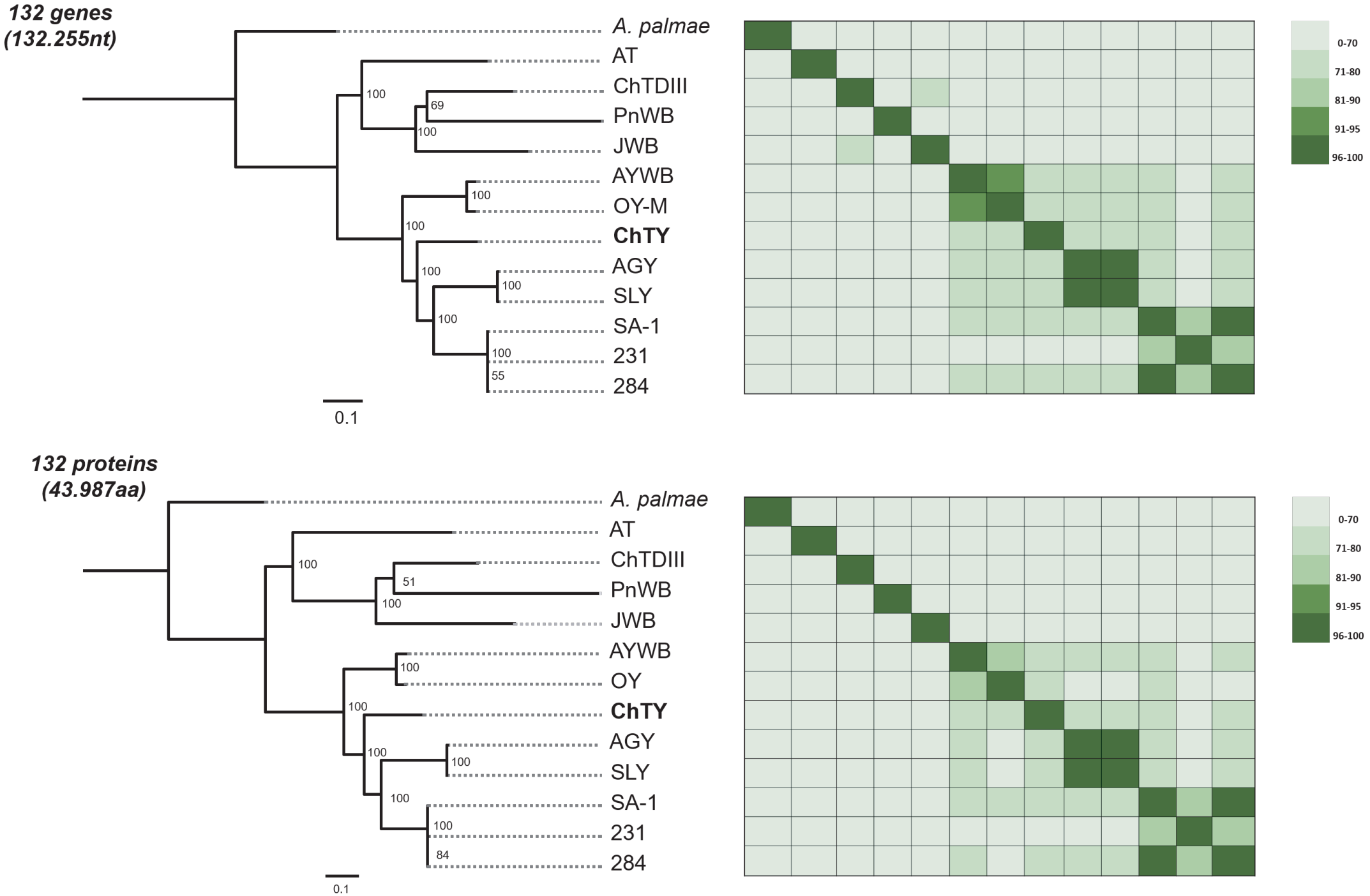
Table S2 List of CDSs associated to carbohydrate metabolism (Glycolysis, Pyruvate oxidation, pyruvate , Starch and sucrose metabolism), ABC transporters and bacterial sec-secretion system

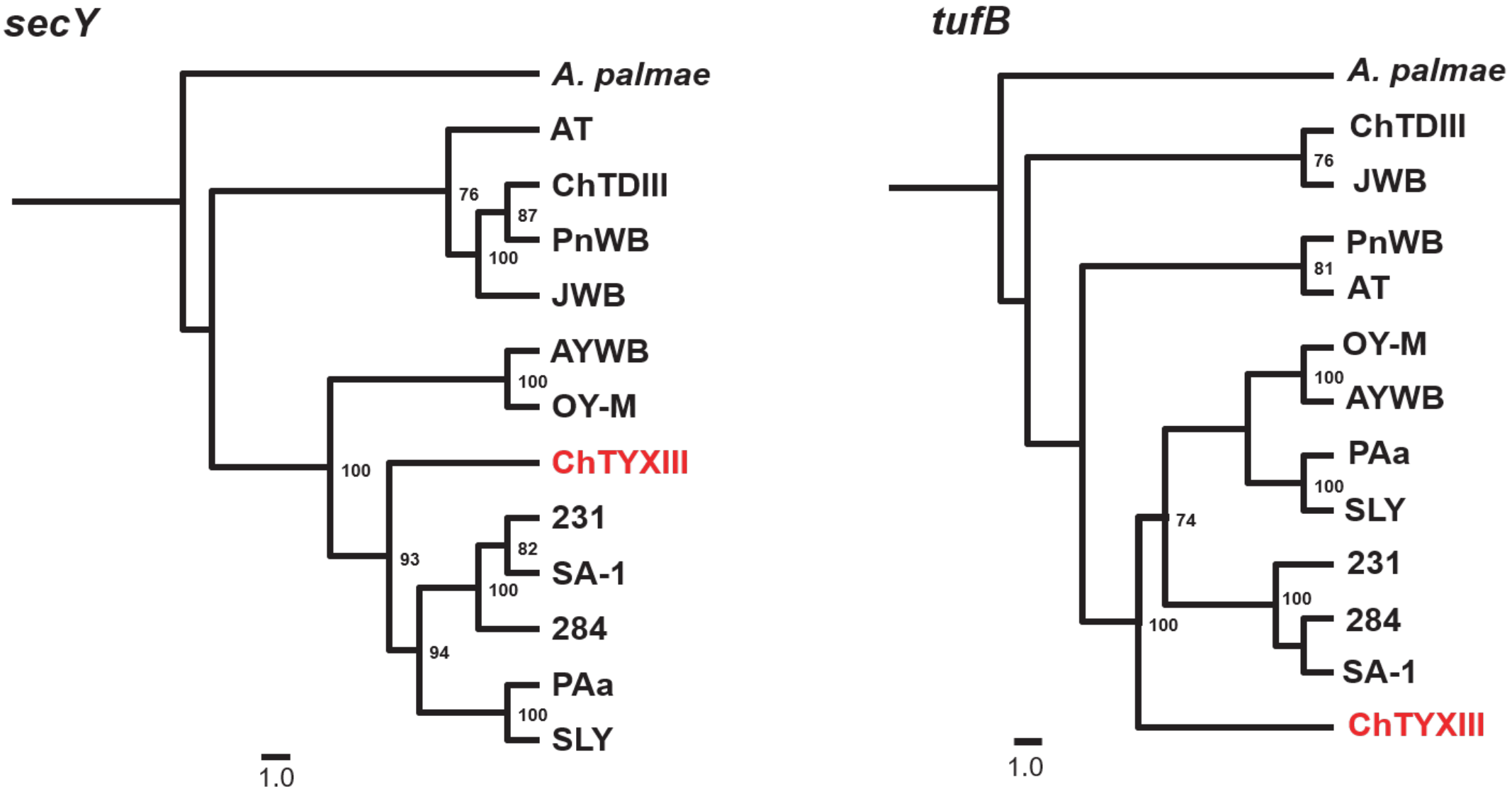
Table S3 List of putative effector proteins in the '*Ca*. P. meliae' draft genome. a: presence of signal peptide (SignalP 5.0); b: localization of proteins to the plant apoplast (ApoplastP); c: subcellular localization prediction in plant cells (LOCALIZER); d: nuclear localization prediction signal (NLstradamus); e: conserved domain prediction (CDD, NCBI); f: annotation based on BLASTp analysis against '*Ca*. Phytoplasma' NCBI database; g: putative SAP homologue assignment based on protein homology against SAP repertoire of AYWB phytoplasma (CP000061.1) (Bai et al., 2009). * putative proteins with no BLASTp-hit


A. palmae (FO681347)
Ca. P. mali AT (CU469464) X-A
Ca. P. pruni ChTDIII (JABUOH000000000.1) III-B
Ca. P. ziziphi JWB (CP025121.1) V-B
PnWB (NZ_AMWZ00000000.1) II-V
Ca. P. asteris AYWB (CP000061.1) I-B
Ca. P. asteris OY-M (AP006628.2) I-A
Ca. P. australiense PAa (AM422018.1) XII-B
Ca. P. australiense SLY (CP002548.1) XII-A
Ca. P. solani 284/09 (FO393427.1) XII-B
Ca. P. solani SA-1 (MPBG00000000) XII-B
Ca. P. solani 231(FO393428.1) XII-B
Ca. P. hispanicum (AF248960) XIII-A
STRAWB1 (U96614) XIII-B
SINPV (FJ914647) XIII-D
SbGP-S27 (KU896186) XIII-I
StrPh-CL4 (MH939194) XIII-K
StrawRL-Tc1 (KJ921641) XIII-F
SFP-Br2 (EU719108) XIII-J
BSP (JX626329) XIII-H
PACN-Br1 (JQ79217) XIII-E
Ca. P. meliae CbY1 (U96614) XIII-C
Ca. P. meliae ChTYXIII (KU850940)R XIII-G
Ca. P. meliae PYWB-AgrP3 (MT553105) XIII-L
98
92
96
100
98
100
76
96
92
98
95
6
92
1.0
MPV-group


KEEG Proteins families
Prokaryotic defense system
3
Exosome
12
Bacterial toxins
2
Secretion system
5
Transporters
32
Mitochondrial biogenesis
20
DNA repair and recombination pro
30
Chromosome and associated protei
7
DNA replication proteins
19
Membrane trafficking
3
Chaperones and folding catalysts
9
Translation factors
11
Transfer RNA biogenesis
39
Ribosome biogenesis
24
Ribosome
51
Messenger RNA biogenesis
11
Transcription machinery
10
Transcription factors
2
Amino acid related enzymes
21
Prenyltransferases
1
Lipid biosynthesis proteins
3
Glycosyltransferases
1
Peptidases and inhibitors
11
Protein phosphatases and associa
1
Enzymes
146
0
20
40
60
80
100
120
140
160
number of genes
Metabolism
Genetic Information Processing
Signaling and Cellular Processes


A
Sequence logo
identity
Ca. asteris AYWB
Ca. asteris MBS
Ca. meliae ChTYXIII
Ca. mali AT
Ca. ziziphi JWB
Ca. pruni CX
PnWB
Ca. asteris OY-M
Ca. solani SA-1
Ca. solani 284
Ca. solani 231
Predicted domains
Signal-peptide (SignalP - 5.0)
Nuclear Signal (NLStradamus)
B
AYWB
391.111-383.530
ChTYXIII
contig_17
1-5795*
Jwb
156.534-161.957
HU
SAP09
hflB
SAP11
SAP68
SAP67
SAP56
SAP66
rad50
SAP21
61-70% nt
71-80% nt
putative pseudoprotein
hipotetical proteins
HU family DNA-binding protein
ATP-dependent Zn protease
DNA doublé-strand break repair rad50 ATPase
Putative SAP-proteins


132 genes
(132.255nt)
A. palmae
AT
ChTDIII
PnWB
JWB
AYWB
OY-M
ChTY
AGY
SLY
SA-1
231
284
100
69
100
100
100
100
100
100
100
55
0.1
0-70
71-80
81-90
91-95
96-100
132 proteins
(43.987aa)
A. palmae
AT
ChTDIII
PnWB
JWB
AYWB
OY
ChTY
AGY
SLY
SA-1
231
284
100
51
100
100
100
100
100
100
100
84
0.1
0-70
71-80
81-90
91-95
96-100


secY
A. palmae
AT
ChTDIII
PnWB
JWB
AYWB
OY-M
ChTYXIII
231
SA-1
284
PAa
SLY
76
87
100
100
100
93
82
100
94
100
1.0
tufB
A. palmae
ChTDIII
JWB
PnWB
AT
OY-M
AYWB
PAa
SLY
231
284
SA-1
ChTYXIII
76
81
100
100
74
100
100
1.0