Complete genome sequence of *Xylella taiwanensis* and comparative analysis of virulence gene content with *Xylella fastidiosa*

Running Head: Genome analysis of *Xylella taiwanensis*

Ling-Wei Weng[1#], Yu-Chen Lin[2#], Chiou-Chu Su[3], Ching-Ting Huang[2], Shu-Ting Cho[2], Ai-Ping Chen[2], Shu-Jen Chou[2], Chi-Wei Tsai[1*], Chih-Horng Kuo[2*]

[1] Department of Entomology, National Taiwan University, Taipei, Taiwan

[2] Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

[3] Division of Pesticide Application, Taiwan Agricultural Chemicals and Toxic Substances Research Institute, Taichung, Taiwan

[#] Equal contribution

* Authors for Correspondence:

Chi-Wei Tsai; chiwei@ntu.edu.tw

Chih-Horng Kuo; chk@gate.sinica.edu.tw

Keywords: *Xylella*, Xanthomonadaceae, plant pathogens, pear leaf scorch, genome, virulence

Abstract

The bacterial genus *Xylella* contains plant pathogens that are major threats to agriculture in America and Europe. Although extensive research was conducted to characterize different subspecies of *Xylella fastidiosa* (*Xf*), comparative analysis at above-species levels were lacking due to the unavailability of appropriate data sets. Recently, a bacterium that causes pear leaf scorch (PLS) in Taiwan was described as the second *Xylella* species (i.e., *Xylella taiwanensis*; *Xt*). In this work, we report the complete genome sequence of *Xt* type strain PLS229^{T}. The genome-scale phylogeny provided strong support that *Xf* subspecies *pauca* (*Xfp*) is the basal lineage of this species and *Xylella* was derived from the paraphyletic genus *Xanthomonas*. Quantification of genomic divergence indicated that different *Xf* subspecies share ~87-95% of their chromosomal segments, while the two *Xylella* species share only ~66-70%. Analysis of overall gene content suggested that *Xt* is most similar to *Xf* subspecies *sandyi* (*Xfs*). Based on the existing knowledge of *Xf* virulence genes, the homolog distribution among 28 *Xylella* representatives was examined. Among the 11 functional categories, those involved in secretion and metabolism are the most conserved ones with no copy number variation. In contrast, several genes related to adhesins, hydrolytic enzymes, and toxin-antitoxin systems are highly variable in their copy numbers. Those virulence genes with high levels of conservation or variation may be promising candidates for future studies. In summary, the new genome sequence and analysis reported in this work contributed to the study of several important pathogens in the family Xanthomonadaceae.

45 **Contribution to the Field**

46 *Xylella fastidiosa* is a plant-pathogenic bacterium with multiple subspecies
47 that are major threats to agriculture in America and Europe. Although extensive
48 research has been conducted, comparative analysis of this species with other
49 bacteria is lacking due to the unavailability of known close relatives. In this work, we
50 report the complete genome sequence of *Xylella taiwanensis*, a newly described
51 species within the same genus. This new data set and our focused analysis helped
52 to better understand the evolutionary relationships among different *Xylella* lineages
53 and their genomic diversity. Moreover, detailed examination of their virulence genes
54 identified those that are either highly conserved or variable, providing promising
55 candidates for future studies to further investigate the molecular mechanisms of
56 *Xylella* virulence.
57

58 **Introduction**

59 The gammaproteobacterium *Xylella fastidiosa* (*Xf*) is an insect-vectored plant
60 pathogen that resides in plant xylem and is fastidious (Wells et al., 1987). To date,
61 more than 563 plant species in 82 families have been reported as hosts for *Xf*
62 (European Food Safety Authority, 2018). *Xf* could be classified into at least five
63 subspecies; some of the notable examples include *Xf* subspecies *fastidiosa* (*Xff*) that
64 causes Pierce's disease (PD) of grapevine, *Xf* subspecies *pauca* (*Xfp*) that causes
65 citrus variegated chlorosis (CVC) and olive quick decline syndrome (OQDS), and *Xf*
66 subspecies *sandyi* (*Xfs*) that causes oleander leaf scorch (OLS). Because of their
67 economic and ecological impacts, substantial resources have been devoted to
68 related research. Notably, several large-scale studies were conducted to investigate
69 the genomic diversity and evolution of *Xf* (Denancé et al., 2019; Potnis et al., 2019;
70 Vanhove et al., 2019). Based on a comparison of 72 strains, the five *Xf* subspecies
71 harbor high levels of genetic diversity (Vanhove et al., 2019). With an average gene
72 content of ~2,150 per strain, the core genome (i.e., genes shared by >95% of the
73 strains) contains only ~900 genes, while the pangenome contains ~10,000 genes.
74 Moreover, although certain patterns of sequence divergence were found among
75 those subspecies (Denancé et al., 2019), extensive recombination occurred at the
76 levels of within- and between-subspecies (Potnis et al., 2019).

77 In contrast to the extensive genomic research at within-species level,
78 comparative studies of *Xf* at higher taxonomic levels are lacking. Under the current
79 taxonomy, *Xylella* belongs to the family Xanthomonadaceae and is most closely
80 related to *Xanthomonas* (Rodriguez-R et al., 2012; Anderson et al., 2013). However,
81 the genomic divergence between *Xylella* and *Xanthomonas* is very high in terms of
82 chromosomal organization, gene content, and sequence variation. Thus, extracting

83 biological insights from such comparisons is difficult. At within-genus level, *Xf* was
84 largely considered as the only species within this genus since it was formally
85 described in 1987 (Wells et al., 1987), which made between-species comparison
86 infeasible. Intriguingly, a *Xylella* lineage that causes pear leaf scorch (PLS) in
87 Taiwan was found to exhibit a slightly lower level of 16S rRNA gene sequence
88 identity at 97.8-98.6% when compared to different subspecies of *Xf* (Su et al., 2012).
89 In 2016, this PLS *Xylella* was formally reclassified as a novel species *Xylella*
90 *taiwanensis* (*Xt*) based on a polyphasic approach (Su et al., 2016). Although a draft
91 genome sequence of *Xt* was published earlier (Su et al., 2014), that draft assembly
92 was produced with only ~20-fold coverage of Roche/454 GS-FLX reads and is highly
93 fragmented (i.e., 85 contigs; N50 = 121 kb). Moreover, no comparative analysis of
94 gene content between *Xt* and *Xf* has been conducted.
95 To fill this gap, we determined the complete genome sequence of the type
96 strain of *Xt* (i.e., PLS229^{T}) for comparative analysis with its relatives. In addition to
97 providing a genome-level overview of their diversity and evolution, we utilized the
98 existing knowledge of *Xff* virulence genes and conducted detailed comparisons of
99 virulence gene content among different *Xylella* lineages.
100
101 **Materials and Methods**
102 The strain was acquired from the Bioresource Collection and Research
103 Centre (BCRC) in Taiwan (accession 80915). The procedures for genome
104 sequencing and comparative analysis were based on those described in our
105 previous studies (Lo et al., 2013; Lo et al., 2018; Cho et al., 2020). All bioinformatics
106 tools were used with the default settings unless stated otherwise.

107 Briefly, the strain was cultivated on PD2 medium as described (Su et al.,
108 2016) for DNA extraction using Wizard Genomic DNA Purification Kit (A1120;
109 Promega, USA). For Illumina sequencing, a paired-end library with a target insert
110 size of 550-bp was prepared using KAPA LTP Library Preparation Kit (KK8232;
111 Roche, Switzerland) without amplification, then sequenced using MiSeq Reagent
112 Nano Kit v2 (MS-103-1003; Illumina, USA) to obtain ~50X coverage. For Oxford
113 Nanopore Technologies (ONT) sequencing, the library was prepared using ONT
114 Ligation Kit (SQK-LSK109) and sequenced using MinION (FLO-MIN106; R9.4
115 chemistry and MinKNOW Core v3.6.0) to obtain ~228X coverage; Guppy v3.4.5 was
116 used for basecalling. The raw reads were combined for *de novo* assembly by using
117 Unicycler v0.4.8-beta (Wick et al., 2017). For validation, the Illumina and ONT raw
118 reads were mapped to the assembly using BWA v0.7.12 (Li and Durbin, 2009) and
119 Minimap2 v2.15 (Li, 2018), respectively. The results were programmatically checked
120 using SAMtools v1.2 (Li et al., 2009) and manually inspected using IGV v2.3.57
121 (Robinson et al., 2011). The finalized assembly was submitted to the National Center
122 for Biotechnology Information (NCBI) and annotated using their Prokaryotic Genome
123 Annotation Pipeline (PGAP) (Tatusova et al., 2016).
124 A total of 40 genomes, including 27 from *Xf*, were used for comparative
125 analysis (**Table 1**). Our taxon sampling mainly focused on the strains that could
126 represent the known *Xylella* diversity (Vanhove et al., 2019). Two other
127 Xanthomonadaceae genera were also included. For the closely-related
128 *Xanthomonas*, 10 species were selected to represent the key lineages (Parkinson et
129 al., 2009; Rodriguez-R et al., 2012). For the distantly-related *Pseudoxanthomonas*,
130 only two species were sampled.

131 Chromosomal level comparisons of nucleotide sequences were conducted
132 using fastANI v1.1 (Jain et al., 2018). Homologous gene clusters were identified
133 based on protein sequences using BLASTP v2.10.0+ (Camacho et al., 2009) and
134 OrthoMCL v1.3 (Li et al., 2003). For gene content comparisons, the homolog
135 clustering result was converted into a matrix of genomes by homolog clusters with
136 the value in each cell corresponding to the copy number. This matrix was converted
137 into a Jaccard distance matrix among genomes using VEGAN package v2.5-6 in R,
138 then processed using the principal coordinates analysis function in APE v5.4
139 (Paradis and Schliep, 2019) and visualized using ggplot2 v3.3.2 (Wickham, 2016).
140 For phylogenetic analysis, homologous sequences were aligned using MUSCLE
141 v3.8.31 (Edgar, 2004). The maximum likelihood inference was performed using
142 PhyML v.3.3.20180214 (Guindon and Gascuel, 2003); the proportion of invariable
143 sites and the gamma distribution parameter were estimated from the data set and
144 the number of substitute rate categories was set to four. The PROTDIST program of
145 PHYLIP v3.697 (Felsenstein, 1989) was used to calculate sequence similarities.
146
147 **Results and Discussion**
148 Genome Characteristics
149 Strain *Xt* PLS229^{T} has one 2,824,877-bp circular chromosome with 53.3%
150 G+C content; no plasmid was found. The annotation contains two complete sets of
151 16S-23S-5S rRNA genes, 49 tRNA genes, and 2,176 intact coding sequences
152 (CDSs). This genome size is near the upper range of those *Xf* representatives
153 (median: 2.54 Mb; range: 2.39-2.88 Mb) and much smaller compared to
154 *Xanthomonas* spp. (median: 5.09 Mb; range: 3.76-5.35 Mb) (**Table 1**). Among all 40
155 representative Xanthomonadaceae genomes, the genome sizes and the numbers of

156 intact CDSs have a correlation coefficient of 0.989 ($p < 2.2e^{-16}$). Compared to those
157 *Xf* representatives with similar genome sizes (i.e., ~2.73-2.88 Mb), such as those
158 five *Xfp* strains from Brazil or the *Xfs* strain from the USA, the *Xt* PLS229^{T} genome
159 has fewer intact CDSs (i.e., 2,273-2,456 vs. 2,173) and fewer pseudogenes (i.e.,
160 153-242 vs. 132). It is unclear if these observations were caused by annotation
161 artifacts or have biological meaning.
162
163 Molecular Phylogeny and Genome Divergence
164 A total of 779 single-copy protein-coding genes were found to be shared by
165 the 40 Xanthomonadaceae genomes compared (**Table 1**). Based on the
166 concatenated alignment of the protein sequences derived from these genes, a robust
167 maximum likelihood phylogeny was inferred (**Figure 1**). The availability of this *Xt*
168 genome sequence provided a more appropriate outgroup to root the *Xf* phylogeny
169 and further supported that *Xfp* is the basal lineage (Denancé et al., 2019; Potnis et
170 al., 2019; Vanhove et al., 2019).
171 The genus *Xanthomonas* was known to be paraphyletic but the relationships
172 of its two major clades (i.e., represented by *Xanthomonas albilineans* and
173 *Xanthomonas campestris*, respectively) with *Xylella* were controversial (Pieretti et al.,
174 2009; Rodriguez-R et al., 2012). With our genome-scale phylogeny, it is clear that
175 *Xylella* is more closely related to *X. campestris* (**Figure 1**) and has experienced
176 genome reduction since their divergence (**Table 1**).
177 When the genetic divergence was measured by overall nucleotide sequence
178 conservation using fastANI (Jain et al., 2018), comparisons within each of the five *Xf*
179 subspecies found that 88.8-99.8% of the chromosomal segments are shared and
180 those segments have 98.5-100% average nucleotide identity (ANI) (**Figure 2**). For

181 between-subspecies comparisons, 86.6-94.8% of the chromosomal segments are
182 shared and those segments have 96.3-98.8% ANI. When those *Xf* subspecies were
183 compared to *Xt*, only 66.4-70.3% of the chromosomal segments are shared and
184 those segments have 82.9-83.4% ANI. These results are consistent with previous
185 findings (Su et al., 2016; Denancé et al., 2019) and provide further support to the
186 current taxonomy based on the 95% ANI threshold recommended for delineating
187 bacterial species (Jain et al., 2018).

188 Because the ANI approach provides low resolutions when the nucleotide
189 sequence identity drops to ~80% (Jain et al., 2018) and may not be appropriate for
190 cross-genus comparisons, we also evaluated divergence based on the protein
191 sequences of those 779 Xanthomonadaceae core genes. The two *Xylella* species
192 have ~88.8-89.1% protein sequence similarity, which is lower than the values
193 observed in the comparisons among those eight *X. campestris* clade representatives
194 (median: 93.8%; range: 92.6-97.2%), comparable to the *X. albilineans-X. hyacinthi*
195 comparison (88.6%), and higher than the *P. mexicana-P. spadix* comparison
196 (75.4%).

197 In addition to analysis of sequence divergence based on those 779 core
198 genes (**Figure 1**) and the entire chromosomes (**Figure 2**), the divergence in gene
199 content was also examined. The gene content comparisons were based on principal
200 coordinates analysis that examines copy number variation among all homologous
201 gene clusters in the entire pangenome and does not consider sequence divergence
202 within each homologous gene cluster. When all 40 Xanthomonadaceae genomes
203 were compared together based on their 11,455 homologous gene clusters, the
204 grouping patterns (**Figure 3A**) are consistent with the phylogenetic clades inferred
205 based on sequence divergence of the 779 core genes (**Figure 1A**). All 27 *Xf*

206 genomes form a tight cluster (**Figure 3A**) despite their differences in the number of
207 intact CDSs (range:1,911-2,456; av.± std. dev.: 2,156 ± 144) (**Table 1**). In contrast,
208 although the *Xt* genome has 2,176 intact CDSs, which is close to the average
209 observed among those 27 *Xf* representatives, it does not fall into the *Xf* cluster
210 (**Figure 3A**). This result indicates that the gene content divergence between these
211 two *Xylella* species is much higher than the divergence among *Xf* subspecies. For
212 the within-*Xylella* comparison based on 5,395 homologous gene clusters, the
213 grouping patterns are consistent with the taxonomic assignments and *Xt* is most
214 similar to *Xfs* (**Figure 3B**). It is interesting that *Xt* and *Xfs* are similar in having a
215 narrow host range (i.e., *Xt* is restricted to pear and *Xfs* is mostly known for oleander
216 infections), while other *Xf* subspecies can infect a wide range of hosts (Baldi and
217 Porta, 2017; European Food Safety Authority, 2018; Rapicavoli et al., 2018).
218 However, it is also important to note that the host range information may be limited
219 by sampling and experimental efforts. As more research results become available,
220 this information may be updated. For example, *Xfs*-related strains have been
221 reported to infect coffee (Jacques et al., 2016) and whether *Xt* can infect a wider
222 range of plants remains to be investigated.
223
224 Virulence Genes and Pathogenicity Factors
225 Based on the current knowledge of putative virulence genes and
226 pathogenicity factors identified in *Xf*, type IV pili (T4P) are important for twitching
227 motility and movements within infected plants (Meng et al., 2005; Li et al., 2007;
228 Burdman et al., 2011; Cursino et al., 2011; Hao et al., 2017). The *Xylella* T4P genes
229 are organized into four major gene clusters with 25 homologs and highly conserved
230 among the 28 representative genomes examined (**Figure 4**). Among these four T4P

231 gene clusters, cluster II that corresponds to the *pil-chp* operon (Cursino et al., 2011;
232 Hao et al., 2017) is the most conserved cluster with only two putative gene losses
233 (i.e., *chpB* in *Xfp* Salento-1 and *chpC* in *Xt*). A notable gene absence is PD1925 in
234 cluster IV, which encodes a hypothetical protein and is absent in all *Xfp* strains and
235 *Xt*. The only gene duplication observed involves a tandem duplication of the cluster
236 IV *pilA2* homolog in *Xt*. Intriguingly, *Xt* PLS229^{T} is the only strain that has two copies
237 of *pilA2* and no *pilA1*. These two type IV pilin paralogs were shown to have different
238 functions in *Xff*, with *pilA1* affecting pilus number and location while *pilA2* is required
239 for twitching (Kandel et al., 2018). It is unclear if this *pilA2* duplication in *Xt* PLS229^{T}
240 can complement its lack of *pilA1*.

241 In addition to the T4P genes, many other *Xylella* pathogenicity factors have
242 been identified and these additional virulence genes may be classified into 10 major
243 functional categories (Lee et al., 2014; Merfa et al., 2016; Chen et al., 2017;
244 Rapicavoli et al., 2018; Ge et al., 2021). Among these categories, secretion systems
245 and metabolism are the most conserved ones with no variation in gene copy number
246 across all *Xylella* representatives (**Figure 5**). Additionally, those genes involved in
247 regulatory systems are also highly conserved. In contrast, several genes related to
248 adhesins, hydrolytic enzymes, and toxin-antitoxin systems are highly variable in copy
249 numbers.

250 For more detailed examination, these putative virulence genes were classified
251 into 38 homologous gene clusters and six are absent in the *Xt* genome (**Figure 5**).
252 These include the genes that encode a putative adhesin (PD0986, hemagglutinin-
253 like protein), two hydrolytic enzymes (PD0956, serine protease; PD1485,
254 polygalacturonase), one pair of toxin-antitoxin (PD0370, motility quorum sensing
255 regulator MqsR ribonuclease; PD0371, MqsA antitoxin), and another separate toxin

256 (PD1100, endoribonuclease). Notably, the *mqsR-mqsA* toxin-antitoxin system genes
257 (Lee et al., 2014; Merfa et al., 2016) are differentially distributed among those *Xf*
258 subspecies. Homologs of these two genes are entirely conserved in all *Xff* and *Xf*
259 subspecies *morus* strains, present in five out of the nine *Xfp* strains, and completely
260 absent in *Xfs* and *Xf* subspecies *multiplex*.

261 Based on previous studies that characterized mutant phenotypes, PD0956
262 (Gouran et al., 2016) and PD1100 (Burbank and Stenger, 2017) are both
263 antivirulence factors and the loss of either one resulted in hypervirulence of *Xff* in
264 grapevines. Similarly, PD0370 is another antivirulence factor that reduces the
265 virulence of *Xfp* against citrus when overexpressed (Merfa et al., 2016). In contrast,
266 both PD0986 and PD1485 are critical for *Xff* virulence in grapevines. For PD0986,
267 this gene is absent in a *Xf* biocontrol strain EB92-1 that can infect and persist in
268 grapevines but causes only very slight symptoms. When PD0986 is cloned into
269 EB92-1, the transformant induces significantly increased symptoms that are
270 characteristic of PD (Zhang et al., 2015). For PD1485, the knockout mutant was
271 avirulent due to the loss of ability to systemically colonize grapevines (Roper et al.,
272 2007).

273 Two gene families appeared to have experienced copy number expansion in
274 the *Xt* genome. The first family includes homologs of PD1792 and PD2118, which
275 encode hemagglutinins. These adhesins are antivirulence factors that restrict *in*
276 *planta* movement by promoting self-aggregation; transposon-insertion mutants of *Xff*
277 PD1792 and PD2118 both exhibit hypervirulence in grapevines (Guilhabert and
278 Kirkpatrick, 2005). Among the representative *Xf* and *Xff* genomes, the median copy
279 numbers of this family are 3 and 8, respectively. In comparison, *Xt* has 12 copies. It
280 remains to be investigated if the copy number variation is linked to protein

281 expression level and virulence. The second family includes a Zot-like toxin
282 (PD0928). Similar to PD0986 (hemagglutinin-like protein), the biocontrol strain EB92-
283 1 lacks the homolog of PD0928 and the transformant that expresses this gene is
284 virulent (Zhang et al., 2015).
285
286 Conclusions
287 In conclusion, this work reported the complete genome sequence of an
288 important plant-pathogenic bacterium that is endemic to Taiwan. In addition to
289 providing the genomic resource that contributes to the study of this pathogen, this
290 species is the only known sister of *Xf*, which has extensive genetic variations and
291 devastating effects on agriculture worldwide. The availability of this new *Xt* genome
292 sequence provides critical genomic information of a key lineage that may improve
293 the study of *Xylella* evolution and the inference of *Xf* ancestral states. At above-
294 genus level, our genome-scale phylogenetic inference resolved the relationships
295 between *Xylella* and *Xanthomonas*, which are some of the key plant pathogens in
296 the family Xanthomonadaceae.
297 For gene content analysis, our comparison of the putative virulence genes
298 and pathogenicity factors among representative *Xylella* strains identified the genes
299 that exhibit high levels of conservation or diversity (**Figures 4 and 5**). These genes
300 are promising candidates for future functional studies to investigate the molecular
301 mechanisms of *Xylella* virulence. Previous characterizations of single-gene mutants,
302 particularly those conducted in *Xff*, have provided a strong foundation (Burdman et
303 al., 2011; Rapicavoli et al., 2018). However, it is important to note that the current
304 knowledge of *Xylella* virulence genes is mostly derived from to those strains that are
305 relatively easy to culture and transform, such that only limited diversity has been

306 investigated in molecular genetics studies. Moreover, infection experiments for the
307 investigation of gene functions were limited to a small number of plant species. For
308 further improvements, experimental studies that examine more diverse *Xylella*
309 lineages and plant hosts, as well as the combined effects of multiple virulence genes
310 will be critical.
311
312

313 **Data Availability**

314 The complete genome sequence of *Xylella taiwanensis* PLS229^{T} has been
315 deposited in GenBank/ENA/DDBJ under the accession CP053627. The raw reads
316 have been deposited at the NCBI Sequence Read Archive under the accession
317 numbers SRR11805344 and SRR11805345.

318

319 **Funding**

320 The funding was provided by the Ministry of Science and Technology of
321 Taiwan (MOST 106-2923-B-002-005-MY3) to CWT and Academia Sinica to CHK.
322 The funders had no role in study design, data collection and interpretation, or the
323 decision to submit the work for publication.

324

325 **Acknowledgments**

326 The Illumina and Oxford Nanopore sequencing library preparation service was
327 provided by the Genomic Technology Core (Institute of Plant and Microbial Biology,
328 Academia Sinica). The Illumina MiSeq sequencing service was provided by the
329 Genomics Core (Institute of Molecular Biology, Academia Sinica).

330

331 **Author Contribution**

332 Conceptualization: CWT, CHK
333 Funding acquisition: CWT, CHK
334 Investigation: LWW, YCL, CTH
335 Methodology: CCS, STC, APC, SJC, CHK
336 Project administration: CWT, CHK
337 Resources: CCS, CWT, CHK

Supervision: CWT, CHK

Validation: LWW, YCL, CTH

Visualization: LWW, YCL, CTH

Writing – original draft: LWW, CHK

Writing – review & editing: LWW, YCL, CCS, CTH, STC, APC, SJC, CWT, CHK

Conflict of Interest

The authors declare no conflict of interest.

348 **References**

349 Anderson, I., Teshima, H., Nolan, M., Lapidus, A., Tice, H., Rio, T. G. D., et al.
350 (2013). Genome sequence of *Frateuria aurantia* type strain (Kondô 67^{T}), a
351 xanthomonade isolated from *Lilium auratium* Lindl. *Stand Genomic Sci* 9, 83.
352 doi:10.4056/sigs.4338002.

353 Baldi, P., and La Porta, N. (2017). *Xylella fastidiosa*: host range and advance in
354 molecular identification techniques. *Front Plant Sci* 8, 944.
355 doi:10.3389/fpls.2017.00944.

356 Burbank, L. P., and Stenger, D. C. (2017). The DinJ/RelE toxin-antitoxin system
357 suppresses bacterial proliferation and virulence of *Xylella fastidiosa* in
358 grapevine. *Phytopathology* 107, 388–394. doi:10.1094/PHYTO-10-16-0374-R.

359 Burdman, S., Bahar, O., Parker, J. K., and De La Fuente, L. (2011). Involvement of
360 type IV pili in pathogenicity of plant pathogenic bacteria. Genes 2, 706–735.
361 doi:10.3390/genes2040706.

362 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et
363 al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10,
364 421. doi:10.1186/1471-2105-10-421.

365 Chen, H., Kandel, P. P., Cruz, L. F., Cobine, P. A., and De La Fuente, L. (2017). The
366 major outer membrane protein MopB Is required for twitching movement and
367 affects biofilm formation and virulence in two *Xylella fastidiosa* strains. Mol
368 Plant Microbe Interact 30, 896–905. doi:10.1094/MPMI-07-17-0161-R.

369 Cho, S.-T., Zwolińska, A., Huang, W., Wouters, R. H. M., Mugford, S. T., Hogenhout,
370 S. A., et al. (2020). Complete genome sequence of "*Candidatus* Phytoplasma
371 asteris" RP166, a plant pathogen associated with rapeseed phyllody disease
372 in Poland. *Microbiol Resour Announc* 9. doi:10.1128/MRA.00760-20.

373 Cursino, L., Galvani, C. D., Athinuwat, D., Zaini, P. A., Li, Y., De La Fuente, L., et al.
374 (2011). Identification of an operon, Pil-Chp, that controls twitching motility and
375 virulence in *Xylella fastidiosa*. Mol Plant Microbe Interact 24, 1198–1206.
376 doi:10.1094/MPMI-10-10-0252.
377 Denancé, N., Briand, M., Gaborieau, R., Gaillard, S., and Jacques, M.-A. (2019).
378 Identification of genetic relationships and subspecies signatures in *Xylella*
379 *fastidiosa*. *BMC Genomics* 20, 239. doi:10.1186/s12864-019-5565-9.
380 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and
381 high throughput. *Nucl Acids Res* 32, 1792–1797. doi:10.1093/nar/gkh340.
382 European Food Safety Authority (2018). Update of the *Xylella* spp. host plant
383 database. *EFSA Journal* 16, e05408. doi:10.2903/j.efsa.2018.5408.
384 Felsenstein, J. (1989). PHYLIP - phylogeny inference package (version 3.2).
385 *Cladistics* 5, 164–166.
386 Ge, Q., Cobine, P., and De La Fuente, L. (2021). The influence of copper
387 homeostasis genes *copA* and *copB* on *Xylella fastidiosa* virulence is affected
388 by sap copper concentration. Phytopathology. doi:10.1094/PHYTO-12-20-
389 0531-R.
390 Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to
391 estimate large phylogenies by maximum likelihood. *Syst Biol* 52, 696–704.
392 doi:10.1080/10635150390235520.
393 Gouran, H., Gillespie, H., Nascimento, R., Chakraborty, S., Zaini, P. A., Jacobson,
394 A., et al. (2016). The secreted protease PrtA controls cell growth, biofilm
395 formation and pathogenicity in *Xylella fastidiosa*. *Sci Rep* 6, 31098.
396 doi:10.1038/srep31098.

397 Guilhabert, M. R., and Kirkpatrick, B. C. (2005). Identification of *Xylella fastidiosa*
398 antivirulence genes: hemagglutinin adhesins contribute to *X. fastidiosa* biofilm
399 maturation and colonization and attenuate virulence. *Mol Plant Microbe*
400 *Interact* 18, 856–868. doi:10.1094/MPMI-18-0856.
401 Hao, L., Athinuwat, D., Johnson, K., Cursino, L., Burr, T. J., and Mowery, P. (2017).
402 *Xylella fastidiosa pil-chp* operon is involved in regulating key structural genes
403 of both type I and IV pili. Vitis 56, 55–62. doi:10.5073/vitis.2017.56.55-62.
404 Jacques, M.-A., Denancé, N., Legendre, B., Morel, E., Briand, M., Mississipi, S., et
405 al. (2016). New coffee plant-infecting *Xylella fastidiosa* variants derived via
406 homologous recombination. Appl Environ Microbiol 82, 1556–1568.
407 doi:10.1128/AEM.03299-15.
408 Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S.
409 (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals
410 clear species boundaries. *Nat Commun* 9, 5114. doi:10.1038/s41467-018-
411 07641-9.
412 Kandel, P. P., Chen, H., and De La Fuente, L. (2018). A short protocol for gene
413 knockout and complementation in *Xylella fastidiosa* shows that one of the type
414 IV pilin paralogs (PD1926) is needed for twitching while another (PD1924)
415 affects pilus number and location. Appl Environ Microbiol 84, e01167-18.
416 doi:10.1128/AEM.01167-18.
417 Lee, M. W., Tan, C. C., Rogers, E. E., and Stenger, D. C. (2014). Toxin-antitoxin
418 systems *mqsR/ygiT* and *dinJ/relE* of *Xylella fastidiosa*. Physiol Mol Plant
419 Pathol 87, 59–68. doi:10.1016/j.pmpp.2014.07.001.
420 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*
421 34, 3094–3100. doi:10.1093/bioinformatics/bty191.

422 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–
423 Wheeler transform. *Bioinformatics* 25, 1754–1760.
424 doi:10.1093/bioinformatics/btp324.
425 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009).
426 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25,
427 2078–2079. doi:10.1093/bioinformatics/btp352.
428 Li, Yaxin., Hao, Guixia., Galvani, C. D., Meng, Y., De La Fuente, L., Hoch, H. C., et
429 al. (2007). Type I and type IV pili of *Xylella fastidiosa* affect twitching motility,
430 biofilm formation and cell–cell aggregation. Microbiology 153, 719–726.
431 doi:10.1099/mic.0.2006/002311-0.
432 Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog
433 groups for eukaryotic genomes. *Genome Res* 13, 2178–2189.
434 doi:10.1101/gr.1224503.
435 Lo, W.-S., Chen, L.-L., Chung, W.-C., Gasparich, G. E., and Kuo, C.-H. (2013).
436 Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a
437 honeybee-associated bacterium. *BMC Genomics* 14, 22. doi:10.1186/1471-
438 2164-14-22.
439 Lo, W.-S., Gasparich, G. E., and Kuo, C.-H. (2018). Convergent evolution among
440 ruminant-pathogenic *Mycoplasma* involved extensive gene content changes.
441 *Genome Biol Evol* 10, 2130–2139. doi:10.1093/gbe/evy172.
442 Meng, Y., Li, Y., Galvani, C. D., Hao, G., Turner, J. N., Burr, T. J., et al. (2005).
443 Upstream migration of *Xylella fastidiosa* via pilus-driven twitching motility. J
444 Bacteriol 187, 5560–5567. doi:10.1128/JB.187.16.5560-5567.2005.
445 Merfa, M. V., Niza, B., Takita, M. A., and De Souza, A. A. (2016). The MqsRA toxin-
446 antitoxin system from *Xylella fastidiosa* plays a key role in bacterial fitness,

447 pathogenicity, and persister cell formation. Front Microbiol 7, 904.
448 doi:10.3389/fmicb.2016.00904.
449 Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern
450 phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528.
451 doi:10.1093/bioinformatics/bty633.
452 Parkinson, N., Cowie, C., Heeney, J., and Stead, D. (2009). Phylogenetic structure
453 of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int J Syst*
454 *Evol Microbiol* 59, 264–274. doi:10.1099/ijs.0.65825-0.
455 Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., et al.
456 (2009). The complete genome sequence of *Xanthomonas albilineans*
457 provides new insights into the reductive genome evolution of the xylem-limited
458 Xanthomonadaceae. *BMC Genomics* 10, 616. doi:10.1186/1471-2164-10-
459 616.
460 Potnis, N., Kandel, P. P., Merfa, M. V., Retchless, A. C., Parker, J. K., Stenger, D.
461 C., et al. (2019). Patterns of inter- and intrasubspecific homologous
462 recombination inform eco-evolutionary dynamics of *Xylella fastidiosa*. *ISME J*
463 13, 2319–2333. doi:10.1038/s41396-019-0423-y.
464 Rapicavoli, J., Ingel, B., Blanco-Ulate, B., Cantu, D., and Roper, C. (2018). *Xylella*
465 *fastidiosa*: an examination of a re-emerging plant pathogen. *Mol Plant Pathol*
466 19, 786–800. doi:https://doi.org/10.1111/mpp.12585.
467 Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz,
468 G., et al. (2011). Integrative genomics viewer. *Nat Biotech* 29, 24–26.
469 doi:10.1038/nbt.1754.

470 Rodriguez-R, L. M., Grajales, A., Arrieta-Ortiz, M. L., Salazar, C., Restrepo, S., and
471 Bernal, A. (2012). Genomes-based phylogeny of the genus *Xanthomonas*.
472 *BMC Microbiol* 12, 43. doi:10.1186/1471-2180-12-43.
473 Roper, M. C., Greve, L. C., Warren, J. G., Labavitch, J. M., and Kirkpatrick, B. C.
474 (2007). *Xylella fastidiosa* requires polygalacturonase for colonization and
475 pathogenicity in *Vitis vinifera* grapevines. *Mol Plant Microbe Interact* 20, 411–
476 419. doi:10.1094/MPMI-20-4-0411.
477 Su, C.-C., Chang, C.-J., Yang, W.-J., Hsu, S.-T., Tzeng, K.-C., Jan, F.-J., et al.
478 (2012). Specific characters of 16S rRNA gene and 16S–23S rRNA internal
479 transcribed spacer sequences of *Xylella fastidiosa* pear leaf scorch strains.
480 *Eur J Plant Pathol* 132, 203–216. doi:10.1007/s10658-011-9863-6.
481 Su, C.-C., Deng, W.-L., Jan, F.-J., Chang, C.-J., Huang, H., and Chen, J. (2014).
482 Draft genome sequence of *Xylella fastidiosa* pear leaf scorch strain in Taiwan.
483 *Genome Announc.* 2, e00166-14. doi:10.1128/genomeA.00166-14.
484 Su, C.-C., Deng, W.-L., Jan, F.-J., Chang, C.-J., Huang, H., Shih, H.-T., et al. (2016).
485 *Xylella taiwanensis* sp. nov., causing pear leaf scorch disease. *Int J Syst Evol*
486 *Microbiol* 66, 4766–4771. doi:10.1099/ijsem.0.001426.
487 Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P.,
488 Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline.
489 *Nucleic Acids Res* 44, 6614–6624. doi:10.1093/nar/gkw569.
490 Vanhove, M., Retchless, A. C., Sicard, A., Rieux, A., Coletta-Filho, H. D., De La
491 Fuente, L., et al. (2019). Genomic diversity and recombination among *Xylella*
492 *fastidiosa* subspecies. *Appl Environ Microbiol* 85, e02972-18.
493 doi:10.1128/AEM.02972-18.

494 Wells, J. M., Raju, B. C., Hung, H.-Y., Weisburg, W. G., Mandelco-Paul, L., and
495 Brenner, D. J. (1987). *Xylella fastidiosa* gen. nov., sp. nov: gram-negative,
496 xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. *Int J Syst*
497 *Bacteriol* 37, 136–143. doi:10.1099/00207713-37-2-136.
498 Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving
499 bacterial genome assemblies from short and long sequencing reads. *PLOS*
500 *Comput Biol* 13, e1005595. doi:10.1371/journal.pcbi.1005595.
501 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY:
502 Springer-Verlag New York doi:10.1007/978-0-387-98141-3.
503 Zhang, S., Chakrabarty, P. K., Fleites, L. A., Rayside, P. A., Hopkins, D. L., and
504 Gabriel, D. W. (2015). Three new Pierce's Disease pathogenicity effectors
505 identified using *Xylella fastidiosa* biocontrol strain EB92-1. *PLOS ONE* 10,
506 e0133796. doi:10.1371/journal.pone.0133796.
507
508

509 **Figure Legends**

510 **Figure 1.** Molecular phylogeny of *Xylella* and related species in the family
511 Xanthomonadaceae. The maximum likelihood phylogeny was based on 779 shared
512 single-copy genes, the concatenated alignment contains 252,319 aligned amino acid
513 sites. The genus *Pseudoxanthomonas* was included as the outgroup. (A) A
514 phylogram for illustrating the relationships among all 40 genomes analyzed. Each of
515 the *X. fastidiosa* subspecies was collapsed into a triangle for simplified visualization.
516 All internal nodes illustrated in this phylogram received >95% bootstrap support
517 based on 1,000 replicates. (B) A cladogram for illustrating the relationships among
518 those 28 *Xylella* genomes analyzed. Internal nodes with bootstrap values lower than
519 100% were labeled.

520

521 **Figure 2.** Genome similarity among the representative *Xylella* strains. The pairwise
522 comparisons were classified into three categories: (1) within the same *X. fastidiosa*
523 subspecies, (2) between different *X. fastidiosa* subspecies, and (3) between *X.*
524 *fastidiosa* and *X. taiwanensis*.

525

526 **Figure 3.** Principal coordinates analysis of gene content dissimilarity. The %
527 variance explained by each coordinate was provided in parentheses. (A) Based on
528 the 11,455 homologous gene clusters found among all 40 Xanthomonadaceae
529 genomes analyzed in this work. (B) Based on the 5,395 homologous gene clusters
530 found among the 28 *Xylella* genomes analyzed in this work.

531

532 **Figure 4.** Distribution of the type IV pili genes found among representative *Xylella*
533 genomes. *Xanthomonas oryzae* was included as the outgroup. The genes were

534 identified by the PD numbers based on the annotation of Temecula1 genome; those
535 with adjacent PD numbers are located in neighboring regions on the chromosome. A
536 total of four gene clusters (labelled as I-IV) located in syntenic regions were found to
537 be conserved. Gene names were provided when available; '*hypo*' indicates those
538 annotated as encoding hypothetical proteins. Patterns of gene presence and
539 absence were illustrated in the format of a heatmap. For gene absence, those with
540 identifiable pseudogenes were labelled accordingly. One case of tandem gene
541 duplication was observed for the PD_1926 (*pilA2*) homolog in the *Xylella taiwanensis*
542 genome.
543
544 **Figure 5.** Distribution of putative virulence genes and pathogenicity factors among
545 representative *Xylella* genomes. *Xanthomonas oryzae* was included as the outgroup.
546 The homologous gene clusters were identified by the PD numbers based on the
547 annotation of Temecula1 genome; gene names were provided when available. Gene
548 copy numbers were illustrated in the format of a heatmap; values higher than two
549 were labeled with the exact numbers. Two adhesin genes (i.e., PD1792 and
550 PD2118) were assigned to the same homologous gene cluster and were combined
551 for copy number calculation.
552

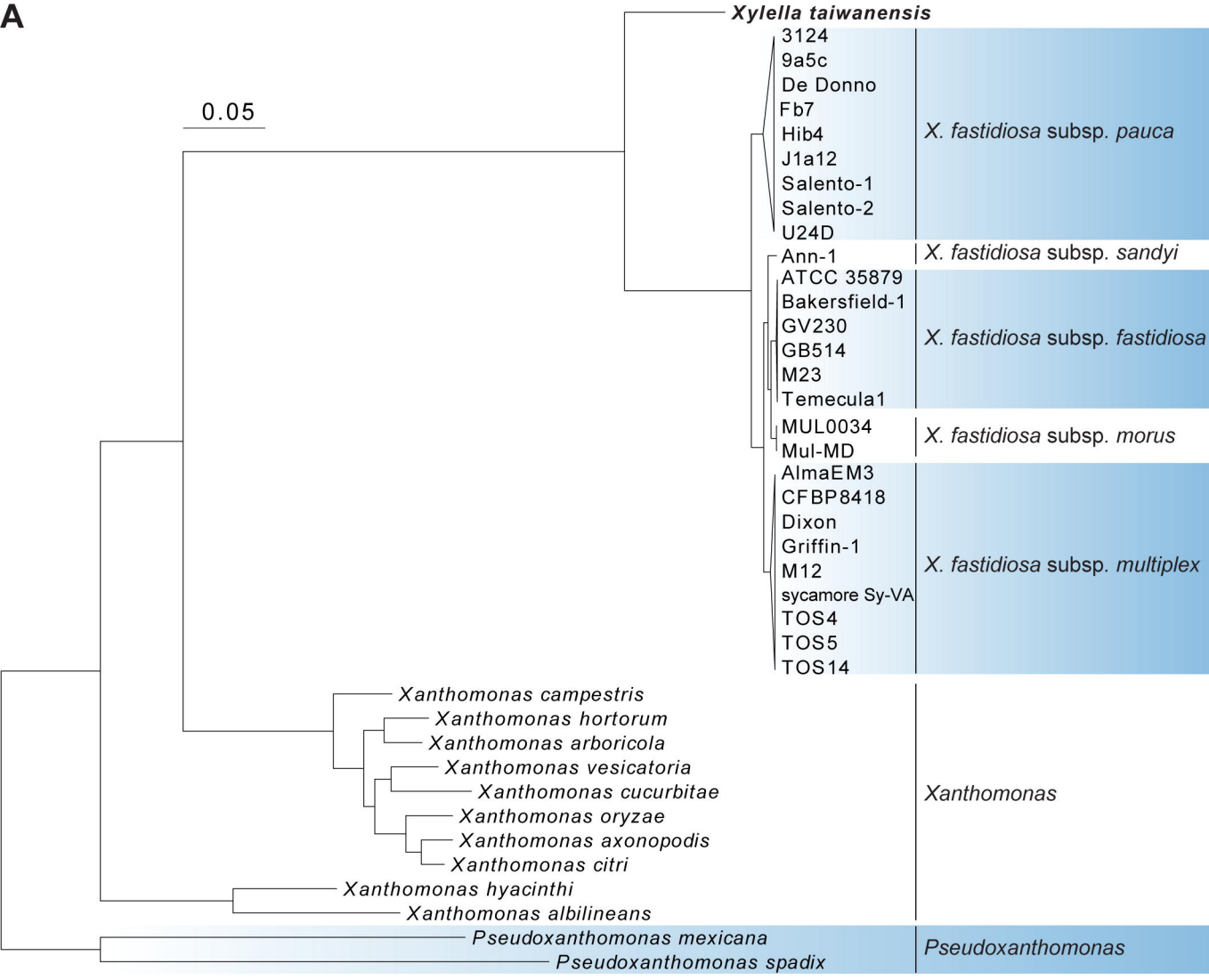
553 **Table**

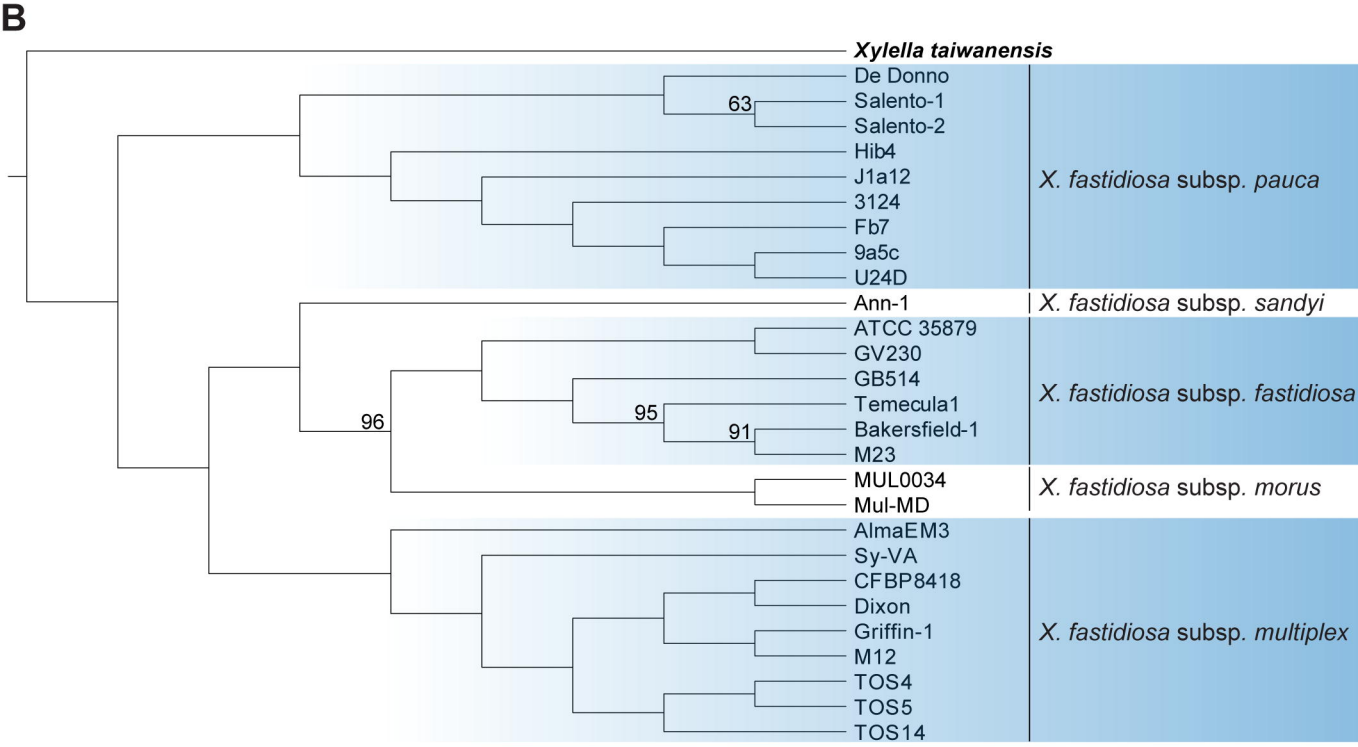
554 **Table 1.** List of the genome sequences analyzed.

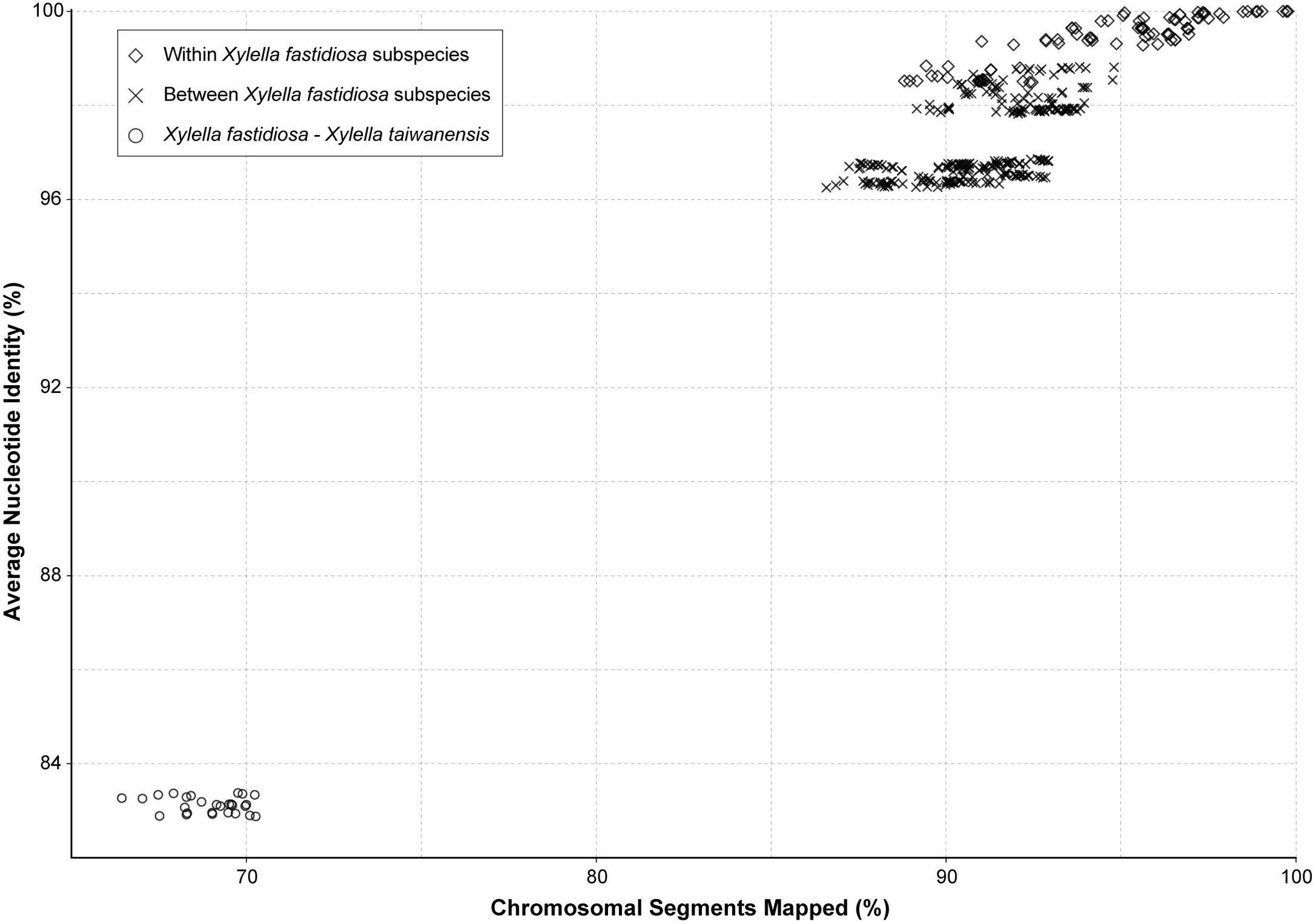
555

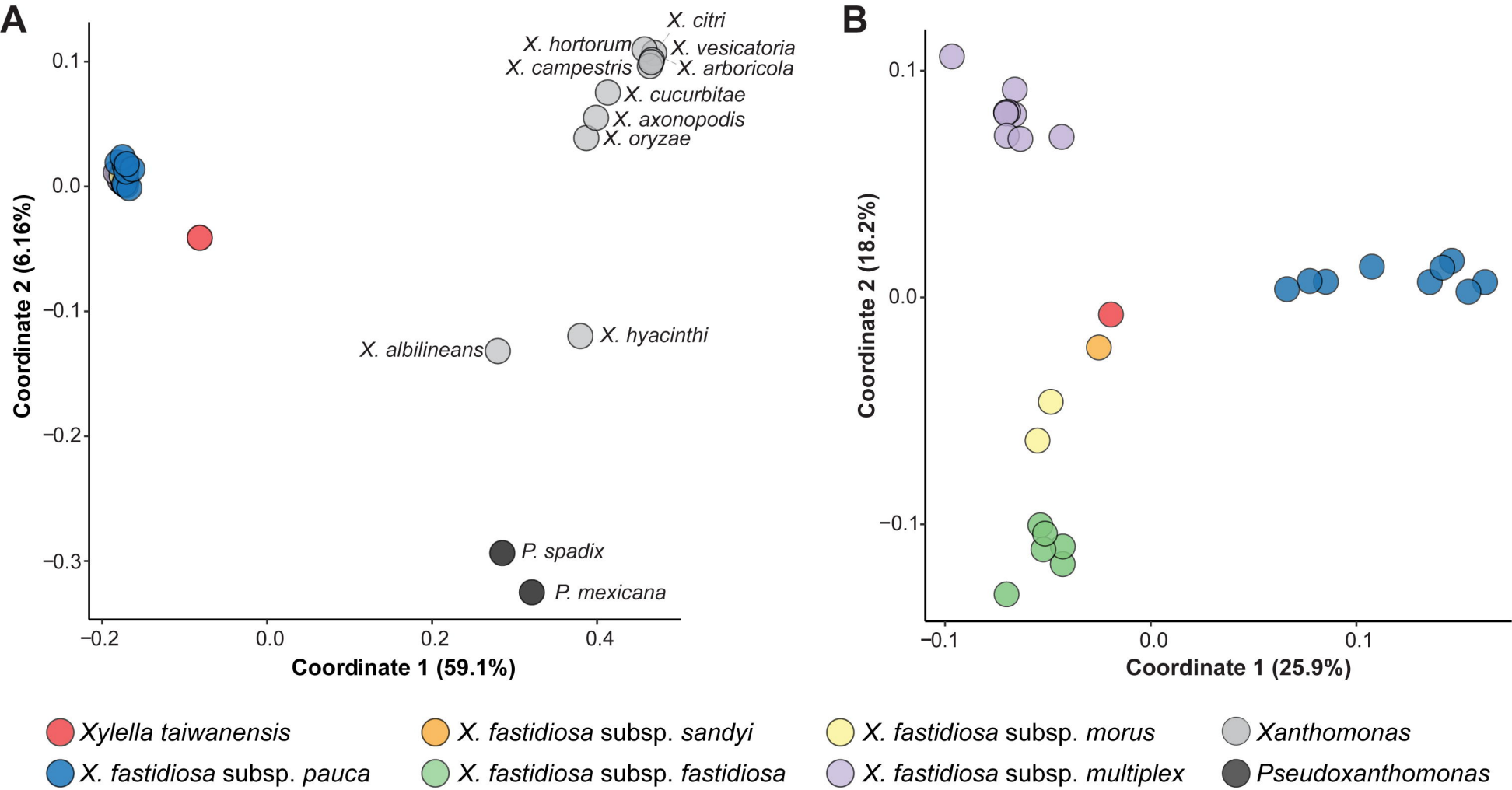
Species	Strain	Location	Accession	Assembly	Genome Size (bp)	CDS (Intact)	CDS (Pseudo)	CDS (All)
Xylella taiwanensis	PLS229	Taiwan: Houli, Taichung	GCA_013177435.1	Complete	2,824,877	2,176	132	2,308
Xylella fastidiosa subsp. *fastidiosa*	ATCC35879	USA: Florida	GCA_011801475.1	Complete	2,607,257	2,189	133	2,322
Xylella fastidiosa subsp. *fastidiosa*	Bakersfield-1	USA: Bakersfield, California	GCA_009664125.2	Complete	2,575,627	2,198	70	2,268
Xylella fastidiosa subsp. *fastidiosa*	GB514	USA: Texas	GCA_000148405.1	Complete	2,517,383	1,998	197	2,195
Xylella fastidiosa subsp. *fastidiosa*	GV230	Taiwan: Waipu, Taichung	GCA_014249995.1	Complete	2,514,993	2,092	61	2,153
Xylella fastidiosa subsp. *fastidiosa*	M23	USA: California	GCA_000019765.1	Complete	2,573,987	2,235	118	2,353
Xylella fastidiosa subsp. *fastidiosa*	Temecula1	USA: California	GCA_000007245.1	Complete	2,521,148	2,107	64	2,171
Xylella fastidiosa subsp. *morus*	MUL0034	USA: California	GCA_000698825.1	Complete	2,666,577	2,266	151	2,417
Xylella fastidiosa subsp. *morus*	Mul-MD	USA: Maryland	GCA_000567985.1	101 contigs	2,520,555	2,121	155	2,276
Xylella fastidiosa subsp. multiplex	AlmaEM3	USA: Georgia	GCA_006369915.1	30 contigs	2,479,954	2,049	141	2,190
Xylella fastidiosa subsp. multiplex	CFBP8418	France: Corse, Alata	GCA_001971465.1	271 contigs	2,513,969	2,199	298	2,497
Xylella fastidiosa subsp. multiplex	Dixon	USA: California	GCA_000166835.1	32 contigs	2,622,328	2,272	123	2,395
Xylella fastidiosa subsp. multiplex	Griffin-1	USA: Georgia	GCA_000466025.1	84 contigs	2,387,314	1,911	303	2,214
Xylella fastidiosa subsp. multiplex	M12	USA: California	GCA_000019325.1	Complete	2,475,130	2,041	116	2,157
Xylella fastidiosa subsp. multiplex	sycamore Sy-VA	USA: Virginia	GCA_000732705.1	128 contigs	2,475,880	2,068	178	2,246
Xylella fastidiosa subsp. multiplex	TOS14	Italy: Tuscany	GCA_007713995.1	80 contigs	2,445,518	2,013	159	2,172
Xylella fastidiosa subsp. multiplex	TOS4	Italy: Tuscany	GCA_007713905.1	77 contigs	2,445,114	2,017	158	2,175
Xylella fastidiosa subsp. multiplex	TOS5	Italy: Tuscany	GCA_007713945.1	72 contigs	2,443,867	2,017	152	2,169
Xylella fastidiosa subsp. *pauca*	3124	Brazil: Matao, Sao Paulo	GCA_001456195.1	Complete	2,748,594	2,273	179	2,452
Xylella fastidiosa subsp. *pauca*	9a5c	Brazil: Macaubal, Sao Paulo	GCA_000006725.1	Complete	2,731,750	2,333	153	2,486
Xylella fastidiosa subsp. *pauca*	De Donno	Italy: Apulia	GCA_002117875.1	Complete	2,543,738	2,092	152	2,244
Xylella fastidiosa subsp. *pauca*	Fb7	Argentina: Corrientes	GCA_001456335.3	Complete	2,699,320	2,178	286	2,464
Xylella fastidiosa subsp. *pauca*	Hib4	Brazil: Jarinu, Sao Paulo	GCA_001456315.1	Complete	2,877,548	2,456	185	2,641
Xylella fastidiosa subsp. *pauca*	J1a12	Brazil: Jales, Sao Paulo	GCA_001456235.1	Complete	2,867,237	2,421	242	2,663
Xylella fastidiosa subsp. *pauca*	Salento-1	Italy: Taviano, Lecce, Apulia	GCA_002954185.1	Complete	2,543,366	1,989	260	2,249
Xylella fastidiosa subsp. *pauca*	Salento-2	Italy: Ugento, Lecce, Apulia	GCA_002954205.1	Complete	2,543,566	2,033	212	2,245
Xylella fastidiosa subsp. *pauca*	U24D	Brazil: Ubarana, Sao Paulo	GCA_001456275.1	Complete	2,732,490	2,274	178	2,452
Xylella fastidiosa subsp. *sandyi*	Ann-1	USA: California	GCA_000698805.1	Complete	2,780,908	2,379	179	2,558
Xanthomonas albilineans	Xa-FJ1	China: Fujian	GCA_009931595.1	Complete	3,756,117	2,968	114	3,082
Xanthomonas arboricola	17	China: Jiangsu	GCA_000972745.1	Complete	5,254,865	4,330	162	4,492
Xanthomonas axonopodis	NCPPB 796	Mauritius	GCA_013177355.1	Complete	4,886,779	3,514	683	4,197
Xanthomonas campestris	MAFF106181	Japan: Aomori	GCA_013388375.1	Complete	4,942,039	4,041	75	4,116
Xanthomonas citri	GD3	China: Guangdong	GCA_000961335.1	Complete	5,223,748	4,219	141	4,360
Xanthomonas cucurbitae	ATCC 23378	USA: New York	GCA_009883735.1	Complete	4,615,492	3,702	234	3,936
Xanthomonas hortorum	B07-007	Canada: Monteregie, Quebec	GCA_002285515.1	Complete	5,250,904	4,241	218	4,459
Xanthomonas hyacinthi	CFBP 1156	Netherlands	GCA_009769165.1	Complete	4,963,026	4,025	282	4,307
Xanthomonas oryzae	PXO99A	Philippines	GCA_000019585.2	Complete	5,238,555	3,952	736	4,688
Xanthomonas vesicatoria	LMG911	New Zealand	GCA_001908725.1	Complete	5,349,905	4,340	159	4,499
Pseudoxanthomonas mexicana	GTZY	China: Beijing	GCA_014211895.1	Complete	3,936,186	3,575	37	3,612
Pseudoxanthomonas spadix	BD-a59	South Korea	GCA_000233915.4	Complete	3,452,554	3,048	85	3,133

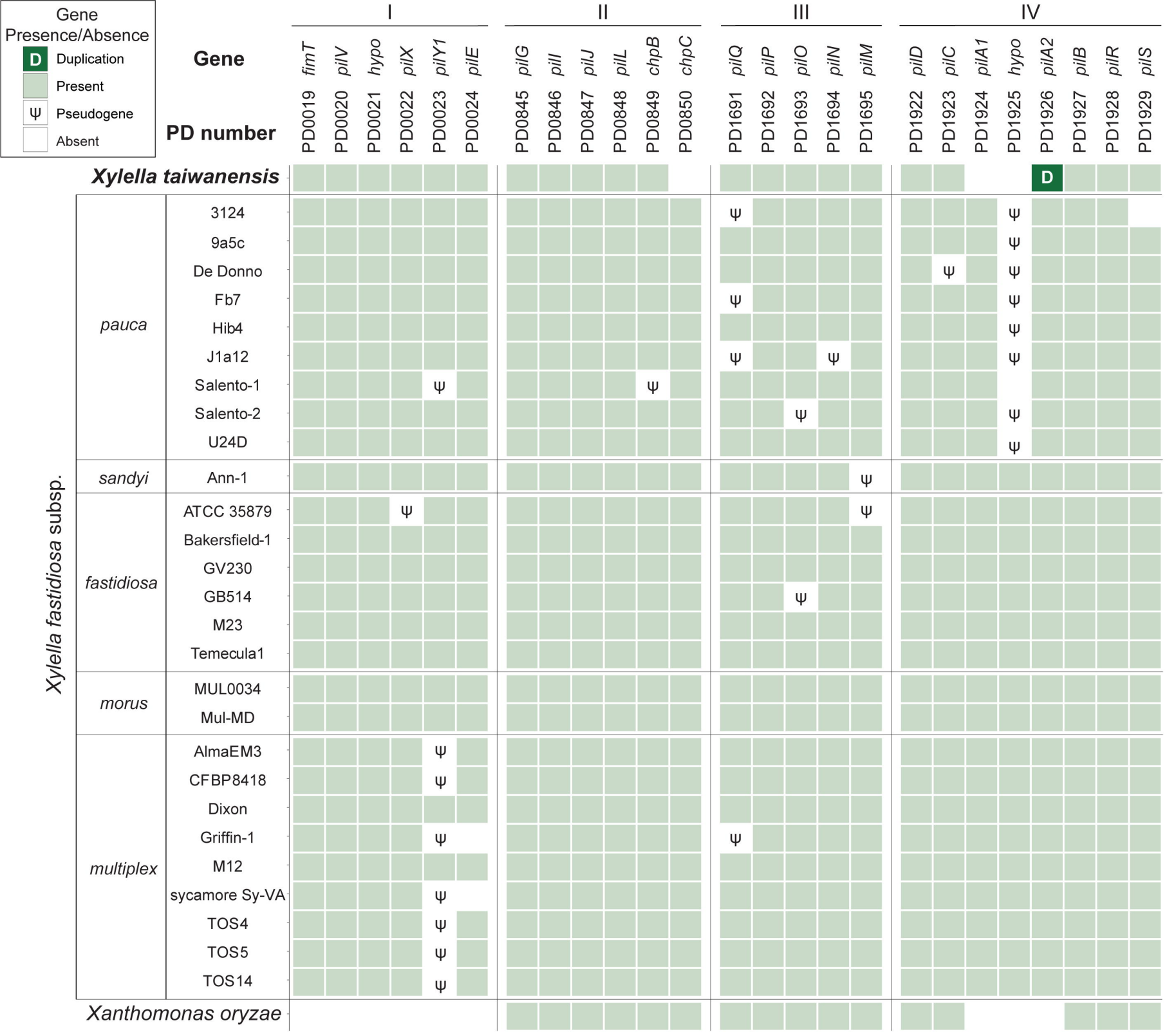
556

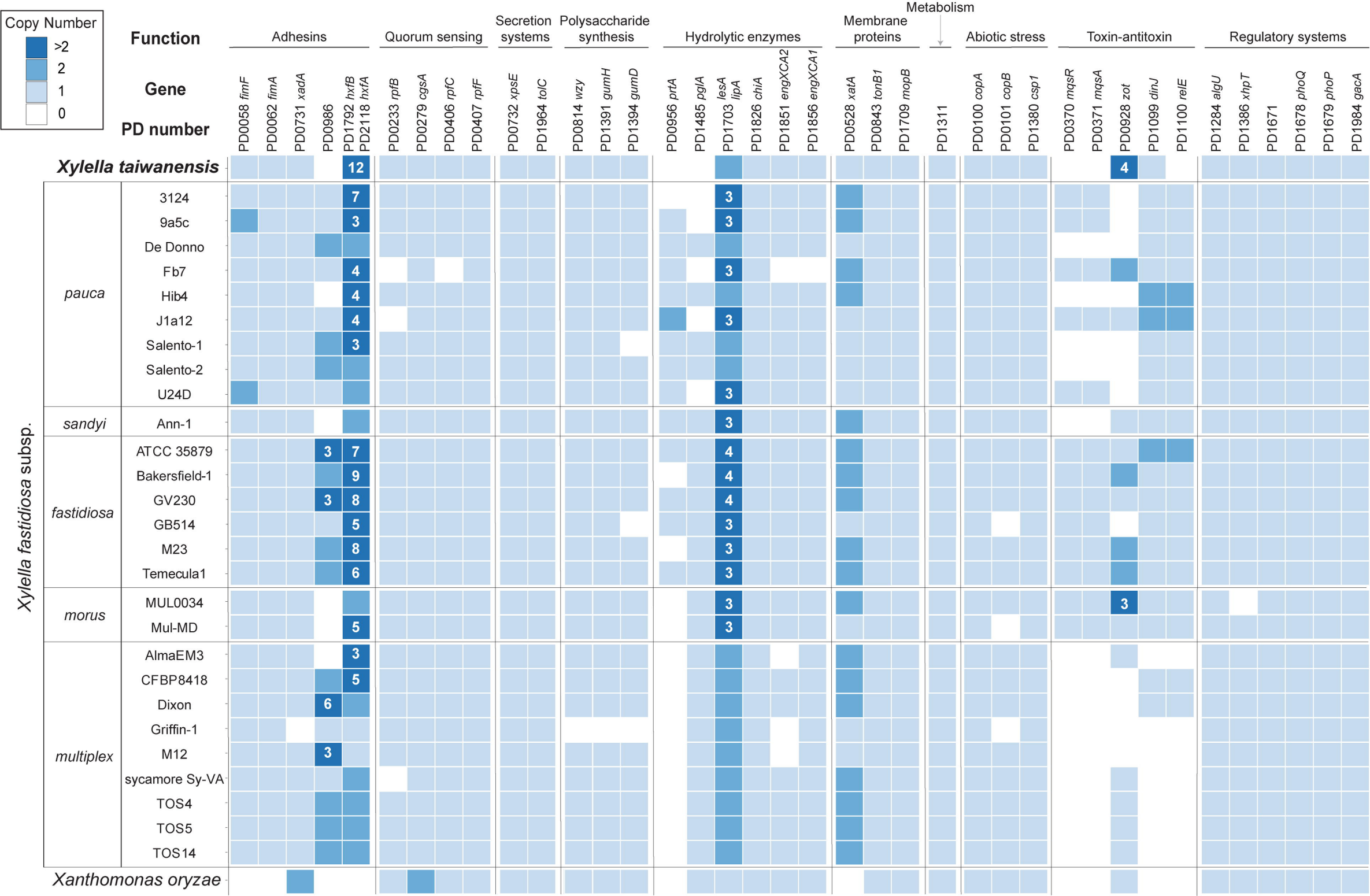

A
0.05
Xylella taiwanensis
3124
9a5c
De Donno
Fb7
Hib4
J1a12
Salento-1
Salento-2
U24D
X. fastidiosa subsp. pauca
Ann-1
X. fastidiosa subsp. sandyi
ATCC 35879
Bakersfield-1
GV230
GB514
M23
Temecula1
X. fastidiosa subsp. fastidiosa
MUL0034
Mul-MD
X. fastidiosa subsp. morus
AlmaEM3
CFBP8418
Dixon
Griffin-1
M12
sycamore Sy-VA
TOS4
TOS5
TOS14
X. fastidiosa subsp. multiplex
Xanthomonas campestris
Xanthomonas hortorum
Xanthomonas arboricola
Xanthomonas vesicatoria
Xanthomonas cucurbitae
Xanthomonas oryzae
Xanthomonas axonopodis
Xanthomonas citri
Xanthomonas hyacinthi
Xanthomonas albilineans
Xanthomonas
Pseudoxanthomonas mexicana
Pseudoxanthomonas spadix
Pseudoxanthomonas


B
Xylella taiwanensis
De Donno
63
Salento-1
Salento-2
Hib4
J1a12
3124
Fb7
9a5c
U24D
X. fastidiosa subsp. pauca
Ann-1
X. fastidiosa subsp. sandyi
ATCC 35879
GV230
GB514
Temecula1
95
91
Bakersfield-1
M23
96
X. fastidiosa subsp. fastidiosa
MUL0034
Mul-MD
X. fastidiosa subsp. morus
AlmaEM3
Sy-VA
CFBP8418
Dixon
Griffin-1
M12
TOS4
TOS5
TOS14
X. fastidiosa subsp. multiplex


Within *Xylella fastidiosa* subspecies
Between *Xylella fastidiosa* subspecies
Xylella fastidiosa - *Xylella taiwanensis*
Average Nucleotide Identity (%)
100
96
92
88
84
Chromosomal Segments Mapped (%)
70
80
90
100


A
X. citri
X. hortorum
X. vesicatoria
X. campestris
X. arboricola
X. cucurbitae
X. axonopodis
X. oryzae
X. albilineans
X. hyacinthi
P. spadix
P. mexicana
Coordinate 2 (6.16%)
Coordinate 1 (59.1%)
0.1
0.0
−0.1
−0.2
−0.3
−0.2
0.0
0.2
0.4
B
Coordinate 2 (18.2%)
Coordinate 1 (25.9%)
0.1
0.0
−0.1
−0.1
0.0
0.1
Xylella taiwanensis
X. fastidiosa subsp. pauca
X. fastidiosa subsp. sandyi
X. fastidiosa subsp. fastidiosa
X. fastidiosa subsp. morus
X. fastidiosa subsp. multiplex
Xanthomonas
Pseudoxanthomonas


Gene Presence/Absence
D Duplication
Present
Ψ Pseudogene
Absent
Gene
PD number
I
II
III
IV
fimT PD0019
pilV PD0020
hypo PD0021
pilX PD0022
pilY1 PD0023
pilE PD0024
pilG PD0845
pilI PD0846
pilJ PD0847
pilL PD0848
chpB PD0849
chpC PD0850
pilQ PD1691
pilP PD1692
pilO PD1693
pilN PD1694
pilM PD1695
pilD PD1922
pilC PD1923
pilA1 PD1924
hypo PD1925
pilA2 PD1926
pilB PD1927
pilR PD1928
pilS PD1929
Xylella taiwanensis
Xylella fastidiosa subsp.
pauca
3124
9a5c
De Donno
Fb7
Hib4
J1a12
Salento-1
Salento-2
U24D
sandyi
Ann-1
fastidiosa
ATCC 35879
Bakersfield-1
GV230
GB514
M23
Temecula1
morus
MUL0034
Mul-MD
multiplex
AlmaEM3
CFBP8418
Dixon
Griffin-1
M12
sycamore Sy-VA
TOS4
TOS5
TOS14
Xanthomonas oryzae


Copy Number
>2
2
1
0
Function
Gene
PD number
Adhesins
Quorum sensing
Secretion systems
Polysaccharide synthesis
Hydrolytic enzymes
Membrane proteins
Metabolism
Abiotic stress
Toxin-antitoxin
Regulatory systems
PD0058 fimF
PD0062 fimA
PD0731 xadA
PD0986
PD1792 hxfB
PD2118 hxfA
PD0233 rpfB
PD0279 cgsA
PD0406 rpfC
PD0407 rpfF
PD0732 xpsE
PD1964 tolC
PD0814 wzy
PD1391 gumH
PD1394 gumD
PD0956 prtA
PD1485 pglA
PD1703 lesA lipA
PD1826 chiA
PD1851 engXCA2
PD1856 engXCA1
PD0528 xatA
PD0843 tonB1
PD1709 mopB
PD1311
PD0100 copA
PD0101 copB
PD1380 csp1
PD0370 mqsR
PD0371 mqsA
PD0928 zot
PD1099 dinJ
PD1100 relE
PD1284 algU
PD1386 xhpT
PD1671
PD1678 phoQ
PD1679 phoP
PD1984 gacA
Xylella taiwanensis
Xylella fastidiosa subsp.
pauca
3124
9a5c
De Donno
Fb7
Hib4
J1a12
Salento-1
Salento-2
U24D
sandyi
Ann-1
fastidiosa
ATCC 35879
Bakersfield-1
GV230
GB514
M23
Temecula1
morus
MUL0034
Mul-MD
multiplex
AlmaEM3
CFBP8418
Dixon
Griffin-1
M12
sycamore Sy-VA
TOS4
TOS5
TOS14
Xanthomonas oryzae