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, Department of Entomology, National Taiwan University, Taipei, Taiwan. Phone: +886-2-33665576; E-mail: chiwei@ntu.edu.tw

Chih-Horng Kuo, Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan. Phone: +886-2-27871127; E-mail: chk@gate.sinica.edu.tw

Author Contribution

Conceptualization: CWT, CHK

Funding acquisition: CWT, CHK

Investigation: LWW, YCL, CTH

Methodology: CCS, STC, APC, SJC, CHK

Project administration: CWT, CHK

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

Abstract

The bacterial genus *Xylella* contains plant pathogens that are major threats to agriculture worldwide. 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 genomic 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 10 functional categories, those involved in secretion, metabolism, and stress response 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.

Keywords

Xylella, Xanthomonadaceae, plant pathogens, pear leaf scorch, genome, virulence

Introduction

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

In contrast to the extensive genomic research at within-species level, comparative studies of *Xf* at higher taxonomic levels were lacking. Under the current taxonomy, *Xylella* belongs to the family Xanthomonadaceae and is most closely related to *Xanthomonas* (Rodriguez-R et al. 2012; Anderson et al. 2013). However, the genomic divergence between *Xylella* and *Xanthomonas* is very high in terms of chromosomal organization, gene content, and sequence variation. Thus, extracting biological insights from such comparisons is difficult. At within-genus level, *Xf* was

largely considered as the only species within this genus since it was formally described in 1987 (Wells et al. 1987), which made between-species comparison infeasible. Intriguingly, a *Xylella* lineage that causes pear leaf scorch (PLS) in Taiwan was found to exhibit a slightly lower level of 16S rRNA gene sequence identity at 97.8-98.6% when compared to different subspecies of *Xf* (Su et al. 2012). In 2016, this PLS *Xylella* was formally reclassified as a novel species *Xylella taiwanensis* (*Xt*) based on a polyphasic approach (Su et al. 2016). Although a draft genome sequence of *Xt* was published earlier (Su et al. 2014), it was highly fragmented and no comparative analysis of gene content between *Xt* and *Xf* has been conducted.

To fill this gap, we determined the complete genome sequence of the type strain of *Xt* (i.e., PLS229^{T}) for comparative analysis with its relatives. In addition to providing a genome-level overview of their diversity and evolution, we utilized the existing knowledge of *Xff* virulence genes (Rapicavoli et al. 2018) and conducted detailed comparisons of virulence gene content among different *Xylella* lineages.

Materials and Methods

The strain was acquired from the Bioresource Collection and Research Centre (BCRC) in Taiwan (accession 80915). The procedures for genome sequencing and comparative analysis were based on those described in our previous studies (Lo et al. 2013; Lo et al. 2018; Cho et al. 2020). All bioinformatics tools were used with the default settings unless stated otherwise.

Briefly, the strain was cultivated on PD2 medium as described (Su et al. 2016) for DNA extraction using Wizard Genomic DNA Purification Kit (A1120; Promega, USA). For Illumina sequencing, a paired-end library with a target insert size of 550-

bp was prepared using KAPA LTP Library Preparation Kit (KK8232; Roche, Switzerland) without amplification, then sequenced using MiSeq Reagent Nano Kit v2 (MS-103-1003; Illumina, USA) to obtain ~50X coverage. For Oxford Nanopore Technologies (ONT) sequencing, the library was prepared using ONT Ligation Kit (SQK-LSK109) and sequenced using MinION (FLO-MIN106; R9.4 chemistry and MinKNOW Core v3.6.0) to obtain ~228X coverage; Guppy v3.4.5 was used for basecalling. The raw reads were combined for *de novo* assembly by using Unicycler v0.4.8-beta (Wick et al. 2017). For validation, the Illumina and ONT raw reads were mapped to the assembly using BWA v0.7.12 (Li and Durbin 2009) and Minimap2 v2.15 (Li 2018), respectively. The results were programmatically checked using SAMtools v1.2 (Li et al. 2009) and manually inspected using IGV v2.3.57 (Robinson et al. 2011). The finalized assembly was submitted to the National Center for Biotechnology Information (NCBI) and annotated using their Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al. 2016).

A total of 40 genomes (**supplementary table S1**) were used for comparative analysis. Our taxon sampling mainly focused on the strains that could represent the known *Xylella* diversity (Vanhove et al. 2019). Two other genera were included for comparative analysis. For the closely-related *Xanthomonas*, 10 species were selected to represent the key lineages (Parkinson et al. 2009; Rodriguez-R et al. 2012). For the distantly-related *Pseudoxanthomonas*, only two species were sampled.

Genome-wide comparisons were conducted using fastANI v1.1 (Jain et al. 2018). Homologous gene clusters were identified using BLASTP v2.10.0+ (Camacho et al. 2009) and OrthoMCL v1.3 (Li et al. 2003). For gene content comparisons, the homolog clustering result was converted into a matrix of 40 genomes by 11,455

clusters with the value in each cell corresponding to the copy number. This matrix was converted into a Jaccard distance matrix among genomes using VEGAN package v2.5-6 in R, then processed using the principal coordinates analysis function in APE v5.4 (Paradis and Schliep 2019) and visualized using ggplot2 v3.3.2 (Wickham 2016). For phylogenetic analysis, homologous sequences were aligned using MUSCLE v3.8.31 (Edgar 2004). The maximum likelihood inference was performed using PhyML v.3.3.20180214 (Guindon and Gascuel 2003); the proportion of invariable sites and the gamma distribution parameter were estimated from the data set and the number of substitute rate categories was set to four. The PROTDIST program of PHYLIP v3.697 (Felsenstein 1989) was used to calculate sequence similarities.

Results and Discussion

Genome Characteristics

Strain *Xt* PLS229^{T} has one 2,824,877-bp circular chromosome with 53.3% G+C content; no plasmid was found. The annotation contains two complete sets of 16S-23S-5S rRNA genes, 49 tRNA genes, and 2,176 protein-coding genes. This genome size is near the upper range of those *Xf* representatives (median: 2.54 Mb; range: 2.39-2.88 Mb) and much smaller compared to *Xanthomonas* spp. (median: 5.09 Mb; range: 3.76-5.35 Mb) (**supplementary table S1**). The genome sizes and the numbers of protein-coding genes have a correlation coefficient of 0.989 ($p < 2.2e^{-16}$).

Molecular Phylogeny and Genome Divergence

A total of 779 single-copy genes were found to be shared by the 40 Xanthomonadaceae genomes compared (**supplementary table S1**). Based on the concatenated alignment of these genes, a robust maximum likelihood phylogeny was inferred (**fig. 1A**). The availability of this *Xt* genome sequence provided a more appropriate outgroup to root the *Xf* phylogeny and further supported that *Xfp* is the basal lineage (Denancé et al. 2019; Potnis et al. 2019; Vanhove et al. 2019).

The genus *Xanthomonas* was known to be paraphyletic but the relationships of its two major clades (i.e., represented by *Xanthomonas albilineans* and *Xanthomonas campestris*, respectively) with *Xylella* were controversial (Pieretti et al. 2009; Rodriguez-R et al. 2012). With our genome-scale phylogeny, it is clear that *Xylella* is more closely related to *X. campestris* and has experienced genome reduction since their divergence (**fig. 1A**).

When the genetic divergence was measured by sequence conservation, comparisons within each of the five *Xf* subspecies found that 88.8-99.8% of the genomic segments are shared and those segments have 98.5-100% average nucleotide identity (ANI) (**supplementary fig. S1**). For between-subspecies comparisons, 86.6-94.8% of the genomic segments are shared and those segments have 96.3-98.8% ANI. When those *Xf* subspecies were compared to *Xt*, only 66.4-70.3% of the genomic segments are shared and those segments have 82.9-83.4% ANI. These results are consistent with previous findings (Su et al. 2016; Denancé et al. 2019) and provide further support to the current taxonomy based on the 95% ANI threshold recommended for delineating bacterial species (Jain et al. 2018).

Because the ANI approach provides low resolutions when the values drop to ~80% (Jain et al. 2018) and may not be appropriate for cross-genus comparisons, we also evaluated divergence based on the protein sequences of those 779

Xanthomonadaceae core genes. The two *Xylella* species have ~88.8-89.1% sequence similarity, which is lower than the values observed in the comparisons among those eight *X. campestris* clade representatives (median = 93.8%; range = 92.6-97.2%), comparable to the *X. albilineans-X. hyacinthi* comparison (88.6%), and higher than the *P. mexicana-P. spadix* comparison (75.4%).

In addition to sequences, the divergence in gene content was also examined. When all 40 genomes were compared together, the grouping patterns (**fig. 1B**) are consistent with the phylogenetic relationships (**fig. 1A**). Despite the similarities in gene counts (**supplementary table S1**), *Xt* is distinct from all *Xf* subspecies based on the gene content. For within-*Xylella* comparisons, genomes are clustered based on the taxonomic assignments and *Xt* is most similar to *Xfs* (**fig. 1C**). Intriguingly, *Xfs* is similar to *Xt* in terms of having a narrow host range (i.e., restricted to oleander and pear, respectively), while other *Xf* subspecies can infect a wide range of hosts (Baldi and Porta 2017; Rapicavoli et al. 2018).

Virulence Genes and Pathogenicity Factors

Based on the current knowledge of putative virulence genes and pathogenicity factors identified in *Xff* (Rapicavoli et al. 2018), these genes may be classified into 10 major functional categories (**fig. 2**). Among these, secretion systems, metabolism, and stress response are the most conserved categories with no variation in gene copy number across all *Xylella* representatives. In contrast, several genes related to adhesins, hydrolytic enzymes, and toxin-antitoxin systems are highly variable in copy numbers.

For more detailed examination, these putative virulence genes were classified into 34 homologous gene clusters and four are absent in the *Xt* genome. These

include the genes that encode a putative adhesin (PD0986, hemagglutinin-like protein), two hydrolytic enzymes (PD0956, serine protease; PD1485, polygalacturonase), and a toxin (PD1100, endoribonuclease). Based on previous studies that characterized mutant phenotypes, PD0956 (Gouran et al. 2016) and PD1100 (Burbank and Stenger 2017) are both antivirulence factors and the loss of either one resulted in hypervirulence of *Xff* in grapevines. In contrast, both PD0986 and PD1485 are critical for *Xff* virulence in grapevines. For PD0986, this gene is absent in a *Xf* biocontrol strain EB92-1 that can infect and persist in grapevines but causes only very slight symptoms. When PD0986 is cloned into EB92-1, the transformant induces significantly increased symptoms that are characteristic of PD (Zhang et al. 2015). For PD1485, the knockout mutant was avirulent due to the loss of ability to systemically colonize grapevines (Roper et al. 2007).

Two gene families appeared to have experienced copy number expansion in the *Xt* genome. The first family includes homologs of PD1792 and PD2118, which encode hemagglutinins. These adhesins are antivirulence factors that restrict *in planta* movement by promoting self-aggregation; transposon-insertion mutants of *Xff* PD1792 and PD2118 both exhibit hypervirulence in grapevines (Guilhabert and Kirkpatrick 2005). Among the representative *Xf* and *Xff* genomes, the median copy numbers of this family are 3 and 8, respectively. In comparison, *Xt* has 12 copies. It remains to be investigated if the copy number variation is linked to protein expression level and virulence. The second family includes a Zot-like toxin (PD0928). Similar to PD0986 (hemagglutinin-like protein), the biocontrol strain EB92-1 lacks the homolog of PD0928 and the transformant that expresses this gene is virulent (Zhang et al. 2015).

Conclusions

In conclusion, this work reported the complete genome sequence of an important plant-pathogenic bacterium that is endemic to Taiwan. In addition to providing the genomic resource that contributes to the study of this pathogen, this species is the only known sister of *Xf*, which has extensive genetic variations and devastating effects on agriculture worldwide. The availability of this new *Xt* genome sequence provides critical genomic information of a key lineage that may improve the study of *Xylella* evolution and the inference of *Xf* ancestral states. At above-genus level, our genome-scale phylogenetic inference resolved the relationships between *Xylella* and *Xanthomonas*, which are some of the key plant pathogens in the family Xanthomonadaceae.

For gene content analysis, our comparison of the putative virulence genes and pathogenicity factors among representative *Xylella* strains identified the genes that exhibit high levels of conservation or diversity (**fig. 2**). These genes are promising candidates for future functional studies to investigate the molecular mechanisms of *Xylella* virulence. Previous characterizations of single-gene mutants, particularly those conducted in *Xff*, have provided a strong foundation (Rapicavoli et al. 2018). For further improvements, experimental studies that examine the combined effects of multiple virulence genes and extension to other *Xylella* lineages will be critical.

Data Availability

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

Acknowledgments

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

Funding

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

Conflict of Interest

The authors declare no conflict of interest.

Literature Cited

Anderson I et al. 2013. Genome sequence of *Frateuria aurantia* type strain (Kondô 67^T), a xanthomonade isolated from *Lilium auratium* Lindl. Stand Genomic Sci. 9:83. doi: 10.4056/sigs.4338002.

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

Burbank LP, Stenger DC. 2017. The DinJ/RelE toxin-antitoxin system suppresses bacterial proliferation and virulence of Xylella fastidiosa in grapevine. Phytopathology. 107:388–394. doi: 10.1094/PHYTO-10-16-0374-R.

Camacho C et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421. doi: 10.1186/1471-2105-10-421.

Cho S-T et al. 2020. Complete genome sequence of "*Candidatus* Phytoplasma asteris" RP166, a plant pathogen associated with rapeseed phyllody disease in Poland. Microbiol Resour Announc. 9. doi: 10.1128/MRA.00760-20.

Denancé N, Briand M, Gaborieau R, Gaillard S, Jacques M-A. 2019. Identification of genetic relationships and subspecies signatures in *Xylella fastidiosa*. BMC Genomics. 20:239. doi: 10.1186/s12864-019-5565-9.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res. 32:1792–1797. doi: 10.1093/nar/gkh340.

European Food Safety Authority. 2018. Update of the *Xylella* spp. host plant database. EFSA Journal. 16:e05408. doi: 10.2903/j.efsa.2018.5408.

Felsenstein J. 1989. PHYLIP - phylogeny inference package (version 3.2). Cladistics. 5:164–166.

307 Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate
308 large phylogenies by maximum likelihood. Syst Biol. 52:696–704. doi:
309 10.1080/10635150390235520.
310 Gouran H et al. 2016. The secreted protease PrtA controls cell growth, biofilm
311 formation and pathogenicity in *Xylella fastidiosa*. Sci Rep. 6:31098. doi:
312 10.1038/srep31098.
313 Guilhabert MR, Kirkpatrick BC. 2005. Identification of *Xylella fastidiosa* antivirulence
314 genes: hemagglutinin adhesins contribute to *X. fastidiosa* biofilm maturation
315 and colonization and attenuate virulence. Mol Plant Microbe Interact. 18:856–
316 868. doi: 10.1094/MPMI-18-0856.
317 Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High
318 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
319 boundaries. Nat Commun. 9:5114. doi: 10.1038/s41467-018-07641-9.
320 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics.
321 34:3094–3100. doi: 10.1093/bioinformatics/bty191.
322 Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics.
323 25:2078–2079. doi: 10.1093/bioinformatics/btp352.
324 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
325 transform. Bioinformatics. 25:1754–1760. doi: 10.1093/bioinformatics/btp324.
326 Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for
327 eukaryotic genomes. Genome Res. 13:2178–2189. doi: 10.1101/gr.1224503.
328 Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. 2013. Comparative
329 genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated
330 bacterium. BMC Genomics. 14:22. doi: 10.1186/1471-2164-14-22.

331 Lo W-S, Gasparich GE, Kuo C-H. 2018. Convergent evolution among ruminant-
332 pathogenic *Mycoplasma* involved extensive gene content changes. Genome
333 Biol Evol. 10:2130–2139. doi: 10.1093/gbe/evy172.
334 Parkinson N, Cowie C, Heeney J, Stead D. 2009. Phylogenetic structure of
335 *Xanthomonas* determined by comparison of *gyrB* sequences. Int J Syst Evol
336 Microbiol. 59:264–274. doi: 10.1099/ijs.0.65825-0.
337 Pieretti I et al. 2009. The complete genome sequence of *Xanthomonas albilineans*
338 provides new insights into the reductive genome evolution of the xylem-limited
339 Xanthomonadaceae. BMC Genomics. 10:616. doi: 10.1186/1471-2164-10-
340 616.
341 Potnis N et al. 2019. Patterns of inter- and intrasubspecific homologous
342 recombination inform eco-evolutionary dynamics of *Xylella fastidiosa*. ISME J.
343 13:2319–2333. doi: 10.1038/s41396-019-0423-y.
344 Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and
345 evolutionary analyses in R. Bioinformatics. 35:526–528. doi:
346 10.1093/bioinformatics/bty633.
347 Rapicavoli J, Ingel B, Blanco-Ulate B, Cantu D, Roper C. 2018. *Xylella fastidiosa*: an
348 examination of a re-emerging plant pathogen. Mol Plant Pathol. 19:786–800.
349 doi: https://doi.org/10.1111/mpp.12585.
350 Robinson JT et al. 2011. Integrative genomics viewer. Nat Biotech. 29:24–26. doi:
351 10.1038/nbt.1754.
352 Rodriguez-R LM et al. 2012. Genomes-based phylogeny of the genus *Xanthomonas*.
353 BMC Microbiol. 12:43. doi: 10.1186/1471-2180-12-43.
354 Roper MC, Greve LC, Warren JG, Labavitch JM, Kirkpatrick BC. 2007. *Xylella*
355 *fastidiosa* requires polygalacturonase for colonization and pathogenicity in

356 *Vitis vinifera* grapevines. Mol Plant Microbe Interact. 20:411–419. doi:
357 10.1094/MPMI-20-4-0411.
358 Su C-C et al. 2012. Specific characters of 16S rRNA gene and 16S–23S rRNA
359 internal transcribed spacer sequences of *Xylella fastidiosa* pear leaf scorch
360 strains. Eur J Plant Pathol. 132:203–216. doi: 10.1007/s10658-011-9863-6.
361 Su C-C et al. 2014. Draft genome sequence of *Xylella fastidiosa* pear leaf scorch
362 strain in Taiwan. Genome Announc. 2:e00166-14. doi:
363 10.1128/genomeA.00166-14.
364 Su C-C et al. 2016. *Xylella taiwanensis* sp. nov., causing pear leaf scorch disease.
365 Int J Syst Evol Microbiol. 66:4766–4771. doi: 10.1099/ijsem.0.001426.
366 Tatusova T et al. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids
367 Res. 44:6614–6624. doi: 10.1093/nar/gkw569.
368 Vanhove M et al. 2019. Genomic diversity and recombination among *Xylella*
369 *fastidiosa* subspecies. Appl Environ Microbiol. 85:e02972-18. doi:
370 10.1128/AEM.02972-18.
371 Wells JM et al. 1987. *Xylella fastidiosa* gen. nov., sp. nov: gram-negative, xylem-
372 limited, fastidious plant bacteria related to *Xanthomonas* spp. Int J Syst
373 Bacteriol. 37:136–143. doi: 10.1099/00207713-37-2-136.
374 Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome
375 assemblies from short and long sequencing reads. PLOS Comput Biol.
376 13:e1005595. doi: 10.1371/journal.pcbi.1005595.
377 Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag
378 New York: New York, NY doi: 10.1007/978-0-387-98141-3.

379 Zhang S et al. 2015. Three new Pierce's Disease pathogenicity effectors identified
380 using *Xylella fastidiosa* biocontrol strain EB92-1. PLOS ONE. 10:e0133796.
381 doi: 10.1371/journal.pone.0133796.
382

Figure Legends

Fig. 1. Evolutionary relationships between *Xylella taiwanensis* and its relatives in the family Xanthomonadaceae. (A) Maximum likelihood phylogeny based on 779 shared single-copy genes (252,319 aligned amino acid sites). The internal nodes illustrated in this tree all received >80% bootstrap support based on 1,000 replicates, including all nodes that represent the most recent common ancestors for each of the *X. fastidiosa* subspecies. Some of the relationships at the within-subspecies level were less well-supported; those parts were collapsed into triangles for simplified visualization. (B) Principal coordinates analysis of gene content dissimilarity. All 40 Xanthomonadaceae genomes are included. The % variance explained by each coordinate is provided in parentheses. (C) Gene content dissimilarity among the 28 *Xylella* genomes.

Fig. 2. Distribution of putative virulence genes and pathogenicity factors among representative *Xylella* genomes. *Xanthomonas oryzae* is included as the outgroup. The homologous gene clusters are identified by the PD numbers based on the annotation of Temecula1 genome (Rapicavoli et al. 2018); gene names are provided when available. Gene copy numbers are illustrated in the format of a heatmap; values higher than two are labeled with the exact numbers. Two adhesin genes (i.e., PD1792 and PD2118) were assigned to the same homologous gene cluster and were combined for copy number calculation.

406 **Supplementary Materials**

407 **Supplementary table S1.** List of the genome sequences analyzed. The accession

408 numbers, strain information, and genome characteristics are provided.

409

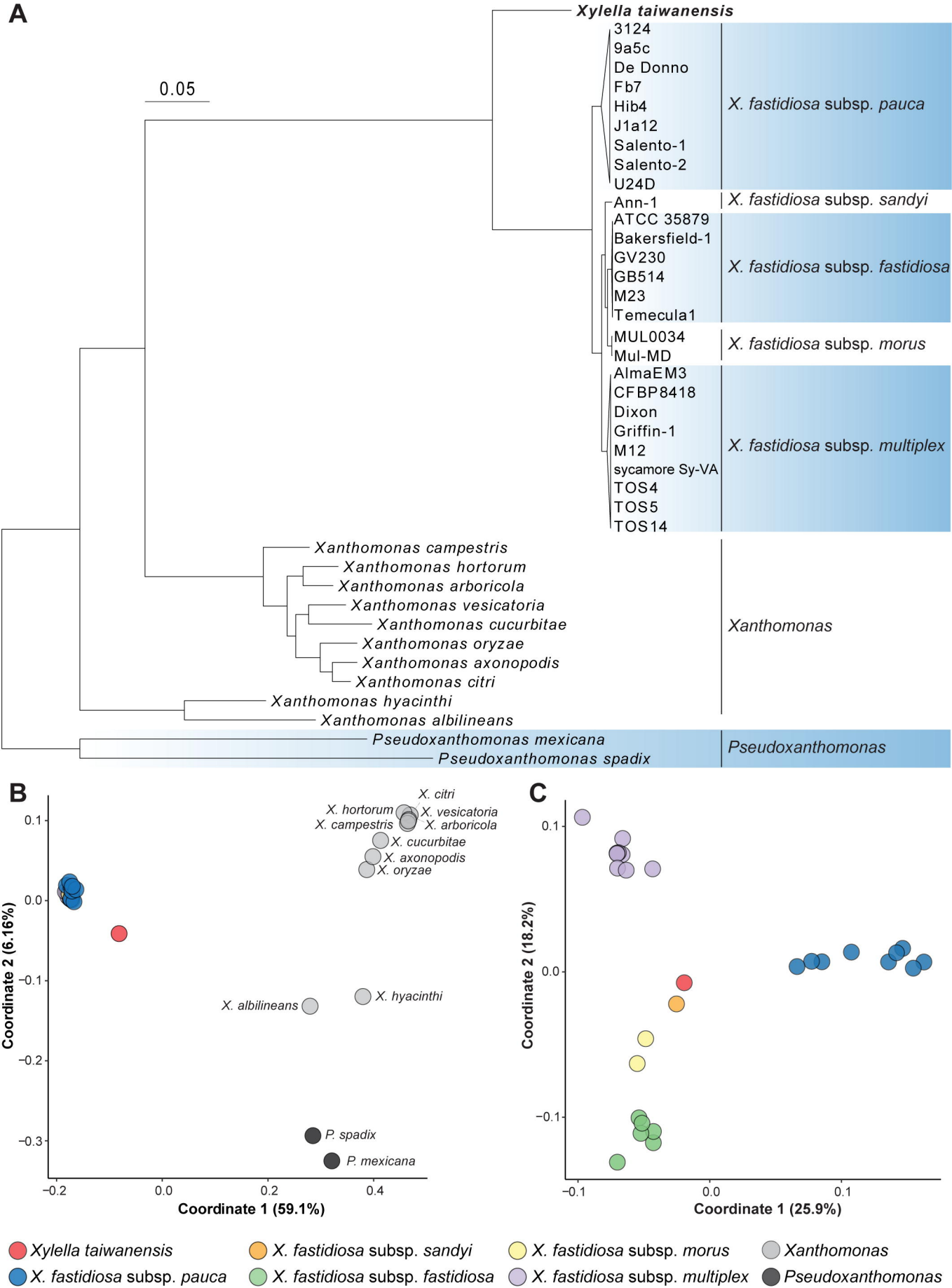
410 **Supplementary fig. S1.** Genome similarity among the representative *Xylella* strains.

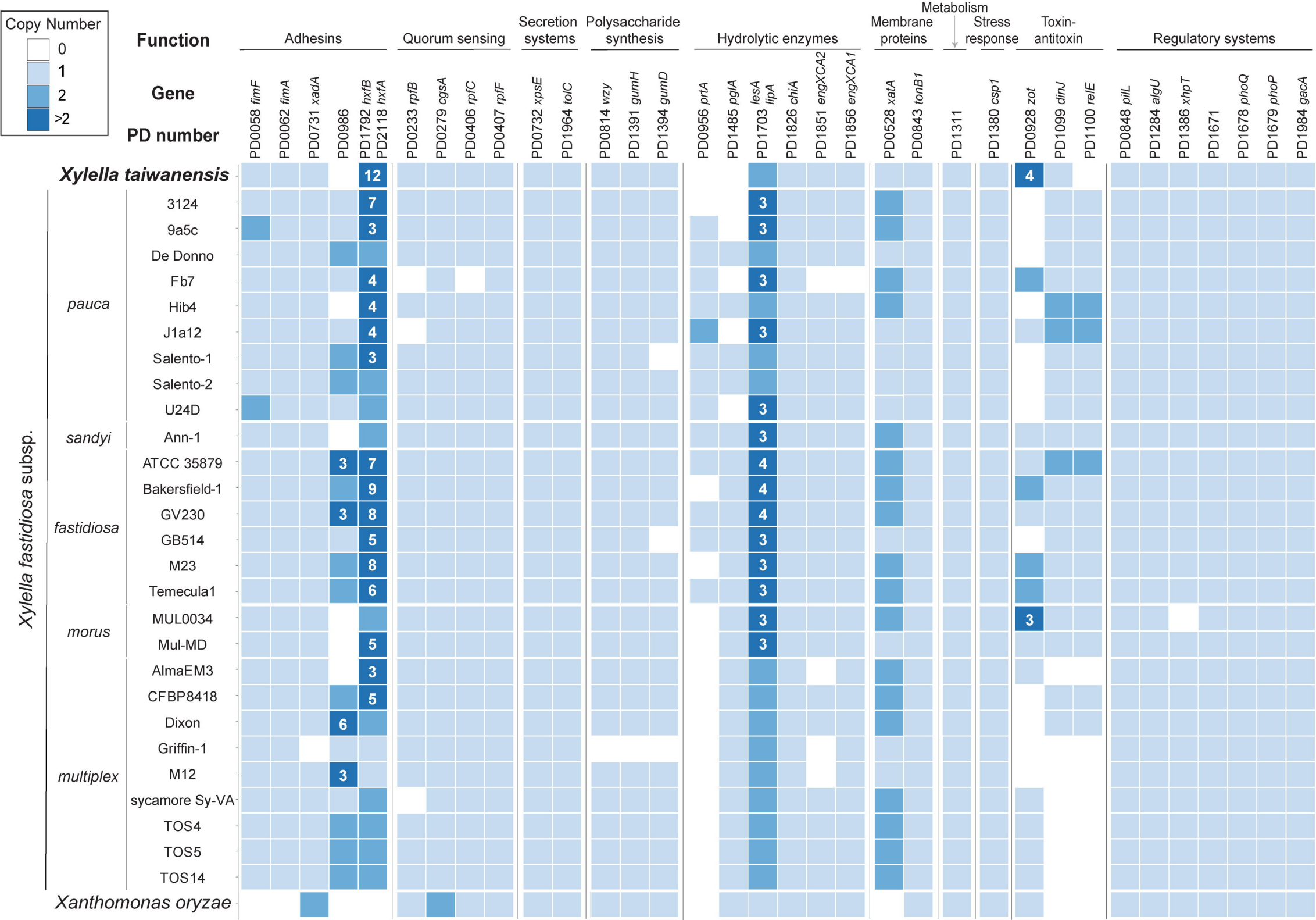
411 The pairwise comparisons are classified into three categories: (1) within the same *X.*

412 *fastidiosa* subspecies, (2) between different *X. fastidiosa* subspecies, and (3)

413 between *X. fastidiosa* and *X. taiwanensis*.

414


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
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%)
C
Coordinate 2 (18.2%)
Coordinate 1 (25.9%)
Xylella taiwanensis
X. fastidiosa subsp. pauca
X. fastidiosa subsp. sandyi
X. fastidiosa subsp. fastidiosa
X. fastidiosa subsp. morus
X. fastidiosa subsp. multiplex
Xanthomonas
Pseudoxanthomonas


Copy Number
0
1
2
>2
Function
Gene
PD number
Adhesins
Quorum sensing
Secretion systems
Polysaccharide synthesis
Hydrolytic enzymes
Membrane proteins
Metabolism
Stress response
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
PD1311
PD1380 csp1
PD0928 zot
PD1099 dinJ
PD1100 relE
PD0848 pilL
PD1284 algU
PD1386 xhpT
PD1671
PD1678 phoQ
PD1679 phoP
PD1984 gacA
Xylella taiwanensis
Xylella fastidiosa subsp.
pauca
sandyi
fastidiosa
morus
multiplex
3124
9a5c
De Donno
Fb7
Hib4
J1a12
Salento-1
Salento-2
U24D
Ann-1
ATCC 35879
Bakersfield-1
GV230
GB514
M23
Temecula1
MUL0034
Mul-MD
AlmaEM3
CFBP8418
Dixon
Griffin-1
M12
sycamore Sy-VA
TOS4
TOS5
TOS14
Xanthomonas oryzae